University of Washington School of Law

UW Law Digital Commons

2014

In-House Digital Libraries for Law Firms

Tanya Livshits

Recommended Citation

In-House Digital Libraries for Law Firms

By: Tanya Livshits

Submitted to:
Professor Penny A. Hazelton
to fulfill course requirements for Current Issues in Law Librarianship, LIS
595, and to fulfill the graduation requirement of the Culminating Experience
Project for MLIS University of Washington Information School Seattle,
Washington

May 13, 2014

In-House Digital Libraries for Law Firms

Table of Contents

In the current legal research and knowledge management environment, digital libraries and in-house knowledge management systems are not only a growing field but also are a necessity for a law firm's continued survival. For a law firm to compete in the present world, a knowledge management system, which maintains a usable and searchable database of the firms work product and shared knowledge, is an absolute must to a collaborative work environment and client relations.

This paper will focus on creating and maintaining such knowledge management databases, i.e. digital libraries, in law firms. First, this paper will define digital libraries. Second, it will discuss the history of digital libraries in general. Third it will discuss the current trends in knowledge management of law firms. Fourth, it will shift into a discussion of various database platform software options and what is necessary for selecting one, with a focus on selecting a vendor maintained software versus on open-source one. Fifth the paper will look at licensing and copyright concerns. Sixth, it will discuss metadata principles. Seventh, this paper will look at the problem of updating information. Lastly, this paper will discuss a land use database created using the principles studied for this paper.

Digital Libraries Defined

This paper will focus on the born digital libraries. That is, libraries designed entirely in the digital space. This paper uses the terms digital library and database interchangeably. More specifically the paper will focus on libraries that are designed around a single topic or tailored to a specific user group, i.e. knowledge management systems. On a grand scale, Westlaw and Lexis are digital libraries. However, the focus of this paper will be more specifically on small in-house knowledge management systems used by law firms.

Digital libraries are defined as a "systems and services, often openly available, that

> (a) support the advancement of knowledge and culture;
> (b) contain managed collections of digital content (objects or links to objects, annotations and metadata) intended to serve the needs of defined communities;
> (c) often use an architecture that first emerged in the computer and information science/library domain and that typically feature a repository, mechanisms supporting search and other services, recourse identifies and user interfaces (human and machine)."[1]

The highest priority of the digital library, like any library, is to service the needs of its users.

The digital library also serves as an information gateway and as such is a "quality-controlled information service that offers

> (1) online links to other Internet sites or documents;
> (2) selection of resources via an intellectual process, within a pre-defined collection scope;
> (3) intellectually-produced content descriptions, preferably with keywords or controlled terminology;
> (4) an intellectually-constructed structure for browsing; and
> (5) at least partially manually-created metadata for individual resources."[2]

Like a traditional brick and mortar library, the information gateway helps the users access relevant information quickly and effectively. A good digital library will combine the points of these definitions into a single user-friendly system.

---

[1] KAREN CALHOUN, EXPLORING DIGITAL LIBRARIES, FOUNDATIONS, PRACTICE, PROSPECTS 18 (2014).
[2] Karen Calhoun, *From Information gateway to digital library management system: a case analysis*, 26 LIBRARY COLLECTIONS, ACQUISITIONS, & TECHNICAL SERVICES 141, 142 (2002).

History of digital libraries

The term digital library was first coined in 1991 at the National Science Foundation workshops on making digital libraries a reality.[3] However, the idea was created decades before the term came into play. Over a 25 year period beginning in 1965, "the technologies needed to build digital libraries became not only available but affordable – for example, digital storage, processors, connectivity, natural language processing, text formatting and scanning, optical character recognition (OCR), indexing and more".[4] This was also accompanied by a rapid growth of databases from 1970s to 1990s. Databases started "in the form of online host services that mounted databases and software from which subscribers could retrieve information using first, dedicated terminals and later, personal computers."[5] The growth during this period was exponential. "The supply of online content was already relatively large by the early 1990s; online databases grew from around 300 in 1979 to nearly 5200 in 1993."[6]

Libraries were early adapters of online information systems serving as intermediaries, searching databases that were not user friendly and not designed for end-user searching.[7] Libraries were at the forefront of metadata, the main organizational tool of digital libraries with the MARC records; which were created by the Library of Congress in 1968. "Over the ensuing years MARC had a transformative influence on libraries, as did the founding of the first shared computerized cataloging system based on MARC, the Ohio College Library Center."[8] The Ohio College Library Center survives at the forefront of library cataloging to this day, now just known as OCLC. "MARC made it possible to aggregate large structured datasets to underpin the conversion from printed to online catalogs of library holdings; the first generation of robust automated systems for libraries; and many new services in libraries."[9]

UK eLIb Programme, which started in 1995, is one of the earliest digital library projects. It involved over seventy different projects.[10] The UK eLib Programme was created in response to the Follett report which reviewed United Kingdom's academic libraries in light of the huge expansion of undergraduate populations, rising costs of library materials and opportunities of new forms of information storage, access and retrieval over networks and recommended that the problems be addressed through the use of information technology.[11]

---

[3] CALHOUN, *Supra* note 1, at 1.
[4] *Id.* at 2.
[5] *Id.* at 3.
[6] *Id.* at 4.
[7] *Id.*
[8] *Id.*
[9] *Id.*
[10] *Id.* at 15.
[11] *Id.*

At the same time, in the United States the DLI-1 project was begun. It started in 1994 and was sponsored by the National Science Foundation (NSF), the National Aeronautics and Space Administration (NASA) and the Defense Advanced Research Projects Agency (DARPA).[12] The DLI-1 project emphasized the technical aspects of digital libraries and largely ignored behavioral, social and economic issues.[13] The user experience was not considered or taken into account when DLI-1 was created.

In 1998, the DLI-1 was followed up DLI-2, which was largely concerned with the social, behavioral and economic aspects of digital libraries.[14] The DLI-2 was still funded by the NSF, but it also received additional funding from several agencies, including national libraries and the Institute of Museum and Library Services (IMLS).[15]

Perhaps the most notable digitization project was the American Memory project, started by the Library of Congress in 1995.[16] The goal of the project was to digitize five million items and make them available on the web within five years.[17] The technical building blocks for the American Memory project came from the National Digital Library Project (NDLP) at the Library of Congress.[18]

As well as the above-mentioned publically funded projects, the early years also had several private projects of note. The most notable of which and that still survives today is Project Gutenberg. Project Gutenberg started in 1971, when Michael Hart, then a student at the University of Illinois, manually typed up the Declaration of Independence and unsuccessfully tried sending it out to the entire university network.[19] "[Project] Gutenberg's goal has been to provide public domain e-texts a short time after they enter the public domain, for free, using only volunteers and donations to get the work done."[20]

Recognizing the need for preservation of information and the fluidity and unstable nature of the Internet, the Internet Archive was created in 1995. One of its best-known projects is the Wayback Machine.[21] The Internet Archive also founded a co-operative project called the Open Content Alliance in order to build and preserve a massive digital library of multilingual digitized text and multimedia content.[22]

---

[12] *Id.*
[13] *Id.*
[14] *Id.* at 16.
[15] *Id.*
[16] *Id.*
[17] *Id.*
[18] *Id.*
[19] *Id.* at 17.
[20] *Id.*
[21] *Id.*
[22] *Id.*

Knowledge Management in the Modern Law Firm

Law firms sell legal knowledge and expertise. In the past, they were easily able to bill by the hour for fairly standard work performed over several hours, for example the crafting of standard transactional agreements. With the advent of the Internet, also came the advent of free or cheap legal advice through services like legalzoom, an "easy-to-use, online service that helped people create their own legal documents."[23] After the economic collapse in 2008, clients began questioning the traditional billing structures of law firms and finding value in collaborative arrangements. "Clients are becoming more and more aware of the collaboration advantage. There are more and more legal collaboration portals available, such as the Association of Corporate Counsel that provides templates and other legal documents to its members, and Legal OnRamp, a collaboration system for in-house counsel and invited outside lawyers and third party service providers. Another interesting collaboration example is Pfizer Legal Alliance, a collaboration program for Pfizer's outside counsel, which makes them work more closely and collaboratively both with Pfizer and with each other using standardized fixed-fee billing arrangements."[24] Given the new client awareness, the traditional model of legal service is no longer viable. Clients now demand flat-fee structures, which are only available through efficient knowledge management.

"Law firms have to review the value of their services and the use of technology to streamline processes and take better advantage of a firm's accumulated knowledge to ensure better service than their competitors. **For the first time in legal history, there is now a true incentive for law firms to deliver results faster, through the right combination of internal and external resources and the better use of IT as a competitive edge.**"[25] Law firms need to reevaluate the value of legal documents and realize that the documents themselves only serve as a basis for legal services. Legal documents in general are easily available to clients.[26] "Instead, they have to look closer at how to better share knowledge from their experience, better re-use documents they have developed, standardize more routine work, and to analyze their most valuable knowledge in order to leverage it to fully support their clients."[27]

"Legal [Knowledge Management] has its roots in helping attorneys practice more efficiently and effectively, by drawing on colleagues' prior work product and

---

[23] LEGALZOOM, https://www.legalzoom.com/about-us (last visited May 13, 2014).
[24] Helena Hallgarn and Ann Björk, *Future of Legal Services and the Development of Legal Knowledge Management*, VOXPOPULII BLOG, (Aug. 30, 2013, 9:35 AM), http://blog.law.cornell.edu/voxpop/2013/08/30/future-of-legal-services-and-the-development-of-legal-knowledge-management/
[25] *Id.*
[26] *Id.*
[27] *Id.*

through sharing information, expertise, and documents within the firm."[28] Historically, this sharing happened naturally without colleagues realizing it—KM was at work behind the scenes finding and organizing resources created by individual attorneys and providing searchable, efficient access to that product to all attorneys.[29] But when "combining strategic development of template or master resources with document automation, KM can shift attorneys from the ancient practice of search/save as/edit to web-based questionnaires that generate a customized "best practice" final document, at a fraction of the time and cost it would take to start from scratch and without the propensity for errors inherent in editing an older document."[30]

The attorney's real value lies not in re-inventing the document every time a client needs a similar document, but in the attorney's knowledge, experience and the ability to understand what the client needs and to satisfy that need as efficiently, cheaply and accurately as possible.

Lexis and Westlaw, the big legal database publishers and database providers, have also entered the knowledge management field.

Westlaw offers a product, West KM, which is a searchable work product repository for law firms. Per West KM's marketing materials, West KM contains "proprietary algorithms developed from our partnerships with leading law firms that tag documents by firm, court, jurisdiction, document type and title, and extract them from your document management system (DMS) or other document store. So even if your documents are formatted inconsistently or incompletely, you still find what you need. It's "intelligent searching" no DMS or other knowledge management provider offers you – for maximum productivity across your organization."[31]

Unlike West's single access point to knowledge management in either litigation or transactional law, Lexis offers a series of products that can be bought together or separately. Lexis's mission in knowledge management is to "make it easier for customers to harness their internal store of knowledge and to enable them to utilize it fully."[32]

---

[28] *Id.*

[29] *Id.*

[30] *Id.* (citing Corinn Jackson and Karen Sundermier, *The Evolving Outward-Facing Role of Knowledge Management*, ILTS KM BLOG, (June 17, 2013), http://km.iltanet.org/2013/06/17/the-evolving-outward-facing-role-of-knowledge-management-part-1-of-2/).

[31] *See* THOMSON REUTERS, *West km/Legal Knowledge Management/Legal Solutions,* http://legalsolutions.thomsonreuters.com/law-products/solutions/westkm (last visited Apr. 9, 2014).

[32] *See* LEXISNEXIS, *Products & Services for Knowledge management,* http://www.lexisnexis.com/km/ (last visited Apr. 9, 2014).

An in-house digital library in a law firm is a very efficient form of a knowledge management system. Not only can it be used to store work product in a searchable, organized manner, but it is also used to store general knowledge. For instance, in a firm that practices land use, the database can be used to store and keep updated all local statutes related to land use, all the necessary processes, environmental laws, contact information for all important government officials and whatever else is necessary for a successful practice as well as zoning variance proposals, construction contracts and other useful work product.

Database Platform Software

When creating an in-house digital library, it is important to first choose the best database software to meet the needs of the particular firm and the particular subject matter. "Which database to choose really depends on the following criteria:

> 1. Who you are;
> 2. What you're trying to achieve (business/functionality requirements, performance/reliability/scalability/availability requirements);
> 3. How much data you want to store in the current databases before archiving them;
> 4. Which OS/Language platform you want to choose for the application;
> 5. How much money you can budget for it; and
> 6. Whether you want/need to build a data warehouse, a BI or decision supporting system on top of it eventually, etc."[33]

Vendor supported platform v. open-source

When deciding on the database to use, budget will play a great role in the choice. Thus the first decision to be made is whether to use a vendor or an open source system. Each has its costs and benefits. Vendor databases often come with a support package, at least initially. "Application vendors usually start by supporting one database platform, and have a longer track record shaking out bugs and increasing performance on this database. Adding additional supported databases later usually means support for the new database platforms is not as mature, at least initially."[34] That is, a vendor will focus their efforts on developing a single database and working out all the bugs. When a new database is added, it will often not have the kinks worked out. Furthermore, it is important

---

[33] Bo Chen, *Choosing a Database Platform,* DATABASE JOURNAL (Mar. 15, 2010), http://www.databasejournal.com/sqletc/article.php/3869736/Choosing-a-Database-Platform.htm (last visited Apr. 9, 2014)
[34] UC BERKLEY IST, *Selecting Your Database Platform* (Mar. 21, 2012), http://ist.berkeley.edu/is/database/services/selecting_a_platform (last visited Apr. 9, 2014).

to review the terms of the vendor contract to determine exactly the level and kind of support to be provided and what the expiration is. When purchasing a vendor supported software, one must be sure that the vendor will not stop producing upgrades and will not stop troubleshooting the software after a certain amount of time has lapsed.

Open source means that the database code is free. Initially this may seem like the economical idea. However, support for the system is generally lacking so the employment of an Internet Technology (IT) professional is required. The man-hours spent working out bugs may also come at a substantial cost. However, support for open-source software is usually available online via user communities. "Determining the best-supported database for a given open source application is usually easy. Rich sources for this information include Google, the open source project web page, popular support forums or the mailing list for the software package."[35]

Oracle and SQL Server are the best-known vendor database platforms and MySQL and PostgreSQL are the best open-source database platforms. Each has its positives and negatives.

Oracle was "once described as 'the aircraft carrier of databases.'"[36] Oracle has the richest feature set and the largest market share of any other databases. It can do most everything that other databases can do.[37] Oracle is designed for a wide array of operating systems, both for clients and servers, and most programming languages work with Oracle.[38] "Oracle provides a free client for development and deployment. Server-side development programming is rich. Oracle performance is first rate. Most vendor packages support Oracle, and when open source packages support a commercial database engine, often it is Oracle."[39] However, Oracle is very costly to purchase and complex to manage due to its wide array of features.[40]

SQL Server is "an engine that is very sophisticated, approaching that of Oracle in features and depth. The cost is comparatively modest. Performance is excellent, and the optimizer is advanced."[41] However, its platform is limited. Unless the database is being developed in Java, a Microsoft server-side tech stack, a set of server software tools that is open-source and designed to work together, is a requirement.[42]

---

[35] *Id.*
[36] *Id.*
[37] *Id.*
[38] *Id.*
[39] *Id.*
[40] *Id.*
[41] *Id.*
[42] *Id.*

"Originally designed as a very lightweight database-like engine, MySQL did not begin with the intent to be a general purpose RDBMS.[43] As such, in versions 3.x and 4.x, it lacked many features considered essential for a relational database. Even so, it was useful as a persistent data engine for purposes that didn't require standard RDBMS features. Not supporting core RDBMS features did mean MySQL was simpler to setup and run."[44] Recently MySQL has been turned into a full-featured RDBMS. MySQL is not very mature and "it is not uncommon for MySQL to reverse itself on design decisions, even in a minor release."[45] However, even given the negatives, the top ten busiest websites in the world utilize MySQL as their backend.[46]

PostgreSQL is the main rival of MySQL. It has a reputation for being complex.[47] "The PostgreSQL slogan is "the world's most advanced open source database". The slogan is well applied. This engine does not suffer at all from MySQL's 'conceptual integrity problems'. The core RDBMS functions are very mature, and the optimizer is far more intelligent than MySQL's. PostgreSQL runs on a very wide array of platforms, and has bindings[48] for many popular languages. Server-side programming is dramatically richer than MySQL, and performance scales better. In many ways, PostgreSQL is Oracle-lite."[49] PostgreSQL's biggest negative is it's comparatively small user base.[50] Since the user base makes up a great deal of a product's support in an open-source system, this is problematic. Working with your own IT person is essential to choosing a database platform that works best for the proposed database. Everything must be taken into account. The main features of the various platforms are largely the same, thus the support, updates and user communities must be closely paid attention too. At the end of the day, a decision should be based on budget and the particular needs of the database and the technological resources available in-house.

Database v. Dataspace

In the current database scholarship, dataspace is being proposed as an alternative. Because, "the design of a digital library not only should meet its key business requirements, but also has to deal with the cross-domain, heterogeneity, and uncertainty of data resources … digital library development is costly and some

---

[43] RDBMS stands for Relational database management system, a database management system that stores data in the form of related tables.
[44] *Id.*
[45] *Id.*
[46] *Id.*
[47] *Id.*
[48] Binding refers to mapping one piece of software to another. Binding is the glue that connects the library to another programming language. *See* http://www.cairographics.org/manual/language-bindings.html
[49] UC BERKLEY IST, *Supra* note 34.
[50] *Id.*

work at data management level is repeated."[51] Dataspace serves as an efficient alternative to the traditional model. Below is a table comparing the typical features of a database with those proposed for a dataspace model. Each of the components will be discussed in this section.

	Dataspace	Database
Construction approach	Pay-as-you-go	Pay-before-you-go
Data schema	Data first, schema later or never	Schema first, data later
Data model	Network model	Relational model
Data storage	Data integration	Data coexisting
Data control	Incomplete	Complete

[52]

"Dataspace can be defined as a set of participants and a set of relationships among them. The participants provide it with their data resources and computing services. Dataspace should not only integrate the resources from participants, notably cross domain, heterogeneous, uncertain ones, but also manage relationships between them, including overlapping, conflicting, inheriting, homogeneous, matching or mapping."[53] A typical system will include a catalog, local storage and index, search and query, discovery, enhancement and administration among other features.[54] This allows for keeping records of participants and their relationships as well as finding participants and mining relationships between them, running unified, structured, metadata queries, extending the data management capabilities of participants, and managing all of preceding parts.[55]

Dataspace is based on five principles: pay-as-you-go; data first, schema later or never; data networking; data coexisting; and incomplete control. This is in complete contradiction to a typical database. As the chart above shows, there is little to no flexibility in a database, whereas a dataspace is based entirely on the idea of flexibility.

In a typical database, the design is established prior to its use and in a commercial setting payment is received before the database platform and software is provided to the user. "Once the design is accepted, it will remain stable for a long period of time. This conventional construction approach results in two acute shortcomings of traditional databases applications. One is they have to amass large amounts of seldom used or useless information in order merely to consider the possibility of use in the future. The other is lack of flexibility. When there are changes in users'

---

[51] Chaolemen Borjigin, Yong Zhang, Chunxiao Xing, Chao Lan and Jian Zhang, *Dataspace and its application in digital libraries*, 31 THE ELECTRONIC LIBRARY 688, 689 (2013).
[52] *Id.* at 690.
[53] *Id.* at 689.
[54] *Id.*
[55] *Id.*

needs or application contexts, it is difficult for a traditional application to alter its design to keep pace with these changes. On the contrary, dataspace abides by an alternative construction principle called pay-as-you-go and its design could evolve with its use."[56] Thus, in a database, the what if I need it later mentality is applied creating a database of a greater volume and requirement a greater amount of storage and maintenance. In a dataspace however, the design is not static and can be easily changed with need for an additional fee. In fact a dataspace is designed with continued change and development in mind.

"The overall design of a database is called database schema and a database has several schemas such as physical schema, logical schema, and sub schema. Of these, the logical schema is by far the most important, in terms of its effect on application programs, since programmers construct applications by using the logical schema."[57] In a database, the schema is designed in advance and the data is captured according to that schema. This provides a lot of accuracy at the expense of flexibility and breadth of information. Since the data is captured according to a predefined schema, some variables or methods, which are not allowed in the schema, will be refused. At the same time, some data or services have to be adapted in order to follow the schema of the database. This sometimes results in loss of information or distortion of data processing."[58]

Current databases most commonly use the RDBMS model; which "employs relational algebra as its theoretical foundation"[59] "However, the cross-domain, heterogeneous, massive, and uncertain feature of data in a database makes it impossible to model the data using relational data model. Compared to relational model, network model is more suitable for modeling data in a dataspace."[60] When all the data is networked together, it can be more easily navigated and browsed and has a better opportunity for completeness.

Data does not stand-alone. Instead it coexists with other data in a greater context. "Dataspace Support System (DSSP) can provide keyword search over all of its data sources, similar to that provided by existing desktop search systems. When more sophisticated operations are required, such as relational-style queries, data mining, or monitoring over certain sources, then additional effort can be applied to more closely integrate those sources in an incremental, pay as you go fashion"[61] The full text searching feature can be easily used to access complete information, but also creates a danger of returning too many and irrelevant results. However, the pay-as-you-go design capability can customize the search feature to meet the exact needs of the user community and the subject matter of the data.

---

[56] *Id.* at 690.
[57] *Id.*
[58] *Id.*
[59] *Id.*
[60] *Id.*
[61] *Id.* at 691.

"In database technologies, data are fully managed by Database Management Systems (DBMS) and they have full control over the data stored in them"[62] "Dataspace Management System (DSMS), which has similar roles with DBMS, has only incomplete control over the data in it. Data in a dataspace may be controlled not only by its DSMS, but also by its providers."[63] Incomplete control allows greater contribution to the database but at the same time, a list of standards should be developed for how and when new information should be added, of course with some flexibility to allow for the addition of unforeseen data.

In a law firm knowledge management system, where the primary goal is to capitalize on the knowledge of the firms' attorneys and prior work product, a more flexible system like the proposed dataspace holds a lot of promise. When each individual attorney can create and upload data and that data is organized according to the firms' design and readily accessible through full text search, even when the metadata may be incorrect, the law firm work flow can be greatly streamlined. However, the accuracy created through the logical schema database standards may also be lost even while the greater flexibility of a dataspace allows for more universal inclusion of data instead of limiting data according to particular standards. Each firm must individually conduct a cost-benefit analysis and decide their own comfort with the level of risk that a lack of rigid standards creates independently.

Licensing and Copyright

"Subject to some limitations, a copyright is the exclusive right to make use of an original literary, musical, or artistic work for a specific period of time."[64] The foundation for copyright protection can be found in the United States Constitution, Article I, Section 8, Clause 8, which states that Congress may "promote the progress of science and the useful arts by securing for a limited time to authors and inventors right to their writings and discoveries."[65] Pursuant to the authorization in the Constitution, Congress has passed The Copyright Act of 1976, which was the first complete revision of the federal copyright statute since 1909.[66]

"Congress recognized the needs of educators, scholars, and librarians in the 1976 Act, although not always to their satisfaction. Teaching, scholarship, and research are specifically mentioned in section 107, the fair use provision. Library copying is addressed in section 108."[67]

---

[62] *Id.*

[63] *Id.*

[64] JAMES S. HELLER, PAUL HELLYER AND BENJAMIN J. KEELE, THE LIBRARIAN'S COPYRIGHT COMPANION 2ND 2 (2012).

[65] *Id.* at 3.

[66] *Id.* (citing Pub. L. No. 94-553, 90 Stat. 2541 (1976)).

[67] J *Id.* at p. 4.

The 1976 Act also placed copyright law within the purview of federal law and federal law alone. Therefore, when copyright research is conducted only federal law and the federal courts need to be consulted.[68]

 "Copyright does not place an author's work in a lockbox. The primary purpose of copyright is not to compensate creators. The U.S. Supreme Court has stated, many times, that copyright is a means to a greater social end: the dissemination and promotion of knowledge. As librarians, we promote the dissemination of knowledge. With this in mind, when there is a close call whether a certain use is or is not allowed, we tend to resolve the answer in favor of the library or the user, rather than the copyright owner.[69]

There is a recent trend toward license agreements gradually displacing copyright law. "Users and owners of copyrighted material have always been free to alter their copyright right and responsibilities by mutual agreement. Digital publishers are compelled to rely on license agreements, partly because their products are more vulnerable to copying and other misuse, and partly because their users sometimes need rights that copyright law doesn't provide."[70] As will be shown in the sections below, in-house digital libraries are hard to fit into copyright exceptions and licensing agreements should serve as a valid alternative to allow the use of legal research materials for new digital libraries.

Reproduction

"The most common copyright infringement is copying, (or, as it's referred to in the Copyright Act, "reproducing").[71] It is no longer easy to know when one is copying. "Modern technology has changed what was once a straightforward issue into a somewhat difficult one, so today's users need to be especially alert when it comes to copying. Generally, any action that transfers a file from one electronic device to another involves making a copy, even if you intend to erase that copy at some point in the future."[72]

However, "sharing material doesn't have to involve copying, sharing a link to material on the Web is not a form of copying, nor is sharing an existing physical copy. Although some types of copying are permitted without the owner's permission, the easiest way to avoid copyright infringement is not to copy when you don't have to."[73]

---

[68] *Id.*
[69] *Id.*
[70] *Id.* at 20.
[71] *Id.* at 22.
[72] *Id.*
[73] *Id.* at 22-23.

In building a digital library, it will often be necessary to refer to other works available on the Internet, such as local codes and forms. While reproducing these in the digital library may not necessarily constitute copyright infringement, providing a link to the webpage where the content can be found directly, will protect the creator of the digital library from inadvertent copyright infringement. The life of links, however is tenuous at best. A study conducted at Harvard Law School found that 49% of all links cited to by the Supreme Court are dead.[74] Therefore a dead-link scanner would also need to be coded into the database to keep it as up to date as possible.

Fair Use and Library Exemption

Fair use is covered under section 107 and "provides the broadest scope of protection for those who use copyrighted works."[75] The origin of fair use is generally traced back to *Folsom v. Marsh,*[76] a case in which Reverend Charles Upham used pages from an edited set of George Washington's letters in his biography of the man.[77] "Justice Joseph Story decreed that the court had to look at three things (1) the nature and objects of the selection, (2) the quantity and value of the materials used, and (3) the degree which the use may prejudice the work, diminish the author's profits, or supersede the objects of the original work."[78] Justice Story determined that Upham's use was not fair.[79]

When Congress codified fair use in the Copyright Act of 1976, it set out the four factors that courts must consider when determining if the work is fair use. "In determining whether the fair use made of a work in any particular case is a fair use the factors to be considered shall include – (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantially of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use on the potential market for or the value of the copyrighted work.[80] Given the commercial nature of an in-house digital library in a law firm, the library increases the firm's worth to the clients, and given that law firms are the primary consumers of commercial databases, an in-house digital library would be hard pressed to justify a fair use exception when facing an infringement suit.

---

[74] Jonathan Zittrain, *Perma: Scoping and addressing the problem f "link rot"*, THE FUTURE OF THE INTERNET AND HOW TO STOP IT BLOG (Sept. 22, 2013) http://blogs.law.harvard.edu/futureoftheinternet/2013/09/22/perma/
[75] HELLER, *Supra* note 64 at 41.
[76] 2 Story 100, 9 F. Cas. 342 (C.C. Mass 1841)
[77] HELLER, *Supra* note 64 at 41.
[78] *Id.* at 41-42.
[79] *Id.* at 42.
[80] *Id.* at 43-44 (quoting 17 U.S.C. § 107 (2006)).

Section 108 of the Copyright Act of 1976 created a library exemption. "Section 108(a) does make it clear that not every instance of copying by libraries qualifies for the section 108 exemption. To qualify for the library exemption:

- The library or archives' collection must be open to the public or to researchers;
- Copying or distribution must be made without any purpose of direct or indirect commercial advantage; and
- The copy must include a notice of copyright."[81]

There requirements are much more stringent then those for fair use. Thus a digital library must be careful, especially, if the digital library is in house and is not open for use by the public or researchers and is used for profit making activities. The law is written in such a way that it would be difficult for a law firm digital library to justify a section 108 exemption.

Remedies and Damages

"Remedies and damages are governed by section 504 of the Copyright Act. In a nutshell, a copyright owner may seek actual or statutory damages, and also try to prohibit the infringing activity."[82] "Actual damages are measured by what was lost as a result of the infringement. Statutory damages can range from $750 to $30,000 per infringing event, and usually will exceed actual damages. If the infringement was willful – if the defendant engaged in the infringing activity knowing that his or her conduct was infringing, or recklessly disregarded the copyright owner's rights – statutory damages can be as much as $150,000 per infringing act."[83] Thus, it is best to be careful and not infringe, especially not willfully. A disregard of copyright rights in intellectual property can be a costly proposition.

Can the digital library created be copyrighted?

When creating an in-house digital library, it is also important to consider whether or not the library should itself be copyrighted. Is it a copyrightable work? And if it is a copyrightable work, should it in fact be copyrighted?

"In order to qualify for copyright protection, the selection or arrangement of the contents of the database must be original and it would only be original if that selection or arrangement constituted the author's own intellectual creation."[84] "A database may be eligible for protection if the compiler exercised sufficient skill and judgment in selecting, organizing, and arranging the data."[85] "Compilation

---

[81] *Id.* at 70.

[82] *Id.* at 33.

[83] *Id.*

[84] PAUL PADLEY, THE E-COPYRIGHT HANDBOOK 19 (2012).

[85] HELLER, *Supra* note 64 at 114.

copyright extends only to the material contributed by the author, not the underlying materials that are complied."[86] Thus, any data gathered in the database is not copyrightable and can be used by a third party. However, if the organizational scheme is replicated, then a copyright violation has occurred.

In deciding whether a work is copyrighted and infringement has occurred and should be pursued through the courts, a cost-benefit analysis must be conducted of whether the database in question will be found to be an original work and the likelihood that that finding will be enough to justify the cost of litigation.

Metadata

"One central technical problem of the digital library is providing effective access to heterogeneous, distributed, digital content. If content remains distributed at the point of its creation, we must have tools to search and retrieve content automatically in all its possible technical formats and present it seamlessly to the end user."[87] The solution to this problem is metadata. Metadata allows us to classify documents according to their facets making them easily searchable via keyword queries.

"Metadata is a term used to refer to a particular kind of data or information. It is data or information that is about other data or information resources, such as a book, an audio file, a scientific data set, or a digital image. Metadata is data or information that enables people to perform certain functions in relation to the information resources that the metadata is about. Metadata is information that is distinct from the resource which it is about, even when the metadata is embedded within a digital resource."[88] Librarians use metadata to describe, manage, and disseminate information about books and journals via library catalogs. Other examples of metadata use include personal research collections, i.e. a set of slides, databases, Web sites, for example a department home page, and even university courses.[89] "Good metadata makes collections and objects discoverable, accessible, manageable, and usable."[90]

Metadata is very diverse and user, industry, and/or item centric. However, as diverse as metadata can be, it all contains several things in common. "First, [it] all consist[s] of a set of properties (elements or fields) and a set of values for each property."[91] "Second, the metadata stands apart from the information resources it

---

[86] *Id.*

[87] David Bearman, *Digital Libraries*, *in* 41 ANNUAL REVIEW OF INFORMATION SCIENCE AND TECHNOLOGY 233 (2008).

[88] STEVEN J. MILLER, METADATA FOR DIGITAL COLLECTIONS: A HOW-TO-DO-IT MANUAL 1(2011).

[89] Calhoun, *Supra* note 2 at 144.

[90] *Id.*

[91] Miller, *Supra* note 88 at 4.

is about. It is something extra, in addition to, and logically separate from the data or information it is about."[92] "Third, the metadata properties and values for each information resource are grouped together in what is traditionally called a record, each record representing selected attributes or properties about one information resource."[93]

"Creators and managers of digital resources usually create different types of metadata to accomplish different kids of functions. Descriptive metadata serves the function of providing users with intellectual access to digital resources by searching and browsing."[94] Other kinds of metadata purposes include: management, administration and preservation of digital resources and internal structuring of complex digital objects.[95]

"If metadata is going to function as intended for users, it needs to be created and structured in a consistent way."[96] In order to be computer processable, metadata needs to be structured into a standardized set of fields, even in a local database, used by a single organization.[97] Metadata and the standardization of it are especially important in the case of interoperability between databases. It is highly unlikely that any database will be entirely contained from within. It will rely on information from other sources; sometime it will provide mere links to avoid running afoul of copyright laws. Thus to make documents findable, it is important to have very good metadata that is set up according to very exact standards. A database is useless to an end user if it is not searchable in an easy and efficient manner, a task completely unmanageable without good metadata.

The only alternative to having good metadata is a dataspace scheme, which allows for full text searching of all documents. Metadata can be very rigorous in its standards and leave the potential for a document to be left out if it does not fit exactly into the rigorous standards. However, metadata does assure a great deal of accuracy and uniformity allowing for easy integration with other sources and continued future use.

Updating information

The biggest challenge to creating any sort of database or digital library is assuring that the information included in the database remains up to date. There is no easy solution to this problem. The literature is widely silent on the subject matter. Even the large legal database providers, such as Lexis and Westlaw, employ an army of people whose job it is to keep those databases updated. For an in-house

---

[92] *Id.* at 5.
[93] *Id.*
[94] *Id.* at 10-11.
[95] *Id.* at 11.
[96] *Id.* at 12.
[97] *Id.*

library, it is not as feasible to devote the manpower necessary to keep a database updated manually.

For an example, a land use database is being created for a wireless tower siting consulting firm. The database consists of local laws, forms, processes and contact information for local government officials dealing with all aspects of zoning and permitting of wireless towers. The database is an aggregation of all individual information gathered by land use auditors over the past three years and will serve as a tool for all future zoning and permitting specialists. Until the database is complete, every time an auditor or a consultant needs to permit or zone a new tower, that person needs to conduct his or her own audit of the requirements. This wastes a great deal of manpower replicating a task that could very well have been done before. A database will alleviate the need for replication and will streamline workflows and provide a product faster. It can potentially also be monetized through sales to wireless carriers.

However, monetization and even usefulness are questionable long term because the database is a snapshot of requirements at a particular point in time and there is not a easy or efficient way to keep the database updated. The database is designed in such a way that the users, who can add up to date information every time they go back to the jurisdiction, can edit it. Essentially, the database will be kept up to date through crowd-sourcing the company's employees. This is not ideal, even if it is time saving overall to not start anew every time. The biggest problem with this method of keeping information up to date is that it will be virtually impossible to monetize the database long term.

Another solution would be to install a dead-link scanner to find any links to municipal codes and forms in the database that need to be updated. This would give the opportunity for an administrator to fix the link as soon as practicable. Code can also be written for the scanner to update the links automatically. However, as with anything else, the scanner is not a perfect solution. While it can find most of the broken links, there is no guarantee that all broken links are found. To make the scanner really work, the code must be maintained and an administrator must quality control its results. Manpower however will be saved from a purely manual system and quality control can be done by spot-checking to assure that the scanner is working correctly.

The best solution, although the most costly, is to install Spider software. Google uses spider to search and index the entire Internet into its search engine. On a much smaller scale, the Spider software can be used to keep linked and networked data in a database up to date.

Land Use Case Study

I am currently in the process of applying the principles outlined in this paper to a database of my own. I did not have a choice in my database platform. Instead, I

closely worked with the IT department to make clear what I needed from the database and the IT department then proceeded to select a platform and construct a database. This required several meetings and a lot of follow up. As it turned out, the IT department was not familiar with what is best from a knowledge management and organization perspective. What made sense technologically to IT did not always translate to database design best practices and vice versa.

Metadata played a major role in the database design and creation. For each new cell tower site, a record is created. When accessing the database, the user will be presented with a search function, whereby they can first find out if the jurisdiction is already represented in the database. It is searchable via a controlled vocabulary, by zip code, city or county. If the jurisdiction is not already present in the database, the crowd-sourcing idea comes into play. The user will then be given the opportunity to expand the database by adding the new jurisdiction. The user is presented with a detailed questionnaire to be completed for the jurisdiction, which covers building permit, zoning, and environmental and regulatory requirements. The user will then also upload all the necessary forms for the jurisdiction into the database and link to the jurisdiction's municipal code, if it is available electronically.

If the jurisdiction is already present in the database, the user will use the information provided. However, the user is required to contact the jurisdiction and confirm that the requirements remain unchanged. If there has been a change, the user will update the database. This significantly decreases the time for obtaining zoning and building permits. A follow up phone call or email to a designated contact person decreases a two-week process down to a day or two.

This also allows the employees to share their knowledge, streamline the information gathering process and in the long term save men hours and decrease the company overhead. It also greatly helps client relations. An answer to a client question is easily and efficiently accessed and can be provided within minutes.

Because of copyright concerns and given that most municipal codes are published online by a particular publisher, it is possible to run afoul of copyright. Thus, instead of downloading and including the municipal codes directly into the database, the database will link to the external sources, which are commonly free for users to access. A dead-link scanner is included in the database software to keep these links up to date. The quality control will occur when the database is referenced for a particular jurisdiction and all information is double-checked. Spider software was considered but proved to costly and difficult to maintain to make it a workable solution on our small scale.

Conclusion

Digital libraries are becoming more commonplace and databases even more so. In the past 35 years growth has been exponential. The legal landscape is also rapidly changing and knowledge management is quickly becoming necessary for the survival of law firms. The real value of a law firm is no longer drafting standard forms, but is instead in the knowledge and experience of the attorneys. Working collaboratively is what clients want and adds value to the bottom line. In a landscape where keeping clients is key, an in-house digital library is the law firm's most valuable tool.

However, going into a building a digital library blindly is not wise. It is important to consider the database platform choice and whether to select one that is sold by a vendor or one readily available online through open source. Closely working with IT will be absolutely necessary no matter what platform is chosen.

Copyright laws must be closely reviewed and strictly followed. Copyright exceptions most likely do not apply to in-house digital libraries. As history has shown, copyright infringement can carry a hefty price. If there is any question as to whether copyright infringement is occurring, a copyright attorney should be consulted and the suspected infringement should stop immediately.

Most important of all, when building a database, metadata must be created according to strictly created standards. Without good metadata, any digital library or databases created will be rendered absolutely useless to the users for whose use they are intended.

Lastly, a scheme to keep the database updated must be devised and maintained. A database is a snapshot in time and the law is ever evolving. Everything in the database must be updated before it can be used and close attention must be paid to any and all changes in the law.

While there are many things to consider when creating a digital library and such libraries do not come without apparent risk, the cost of not having such a library to client relations is far costlier. Every law firm should look into creating a digital library as a knowledge management base.