Content and Quality of Legal Information and Data on the Internet with a Special Focus on the United States

Claire M. Germain

University of Florida Levin College of Law, germain@law.ufl.edu

Follow this and additional works at: http://scholarship.law.ufl.edu/facultypub

Part of the Legal Writing and Research Commons

Recommended Citation

Claire M. Germain, Content and Quality of Legal Information and Data on the Internet with a Special Focus on the United States, 27 Int'l J. Legal Info. 289 (1999), available at http://scholarship.law.ufl.edu/facultypub/159

This Article is brought to you for free and open access by the Faculty Scholarship at UF Law Scholarship Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of UF Law Scholarship Repository. For more information, please contact outlier@law.ufl.edu.
Content and Quality of Legal Information and Data on the Internet with a Special Focus on the United States

CLAIRE M. GERMAIN*

SUMMARY

In the United States today, digital versions of current decisions, bills, statutes, and regulations issued by federal and state governments are widely available on publicly accessible Web sites. Worldwide, official (defined as "authoritative," or "the official" word of the law) legal information issued by international organizations and foreign governments is also becoming available on the Web. However, there are currently no standards for the production and authentication of digital documents. Moreover, the information is sometimes available only for a short time and then disappears from the site. No guidelines exist either to promote a uniform way to cite to digital legal materials.

This article examines the contents of legal data and information on the Internet, with a special focus on the United States. It then evaluates the quality of the data, its impact on legal research and access to legal information, and addresses some issues raised by the digital medium, such as its reliability and permanent access concerns.

* Claire M. Germain is the Edward Cornell Law Librarian and Professor of Law at Cornell Law School, Ithaca, New York, U.S.A. She was the Chair of the American Association of Law Libraries (AALL)/Law Library of Congress Task Force on Preservation and Access to Digital Legal Information, a work continued by the AALL Committee on Authentication and Preservation of Digital Law. The author can be reached at: cmg13@cornell.edu.

The paper was originally presented at "Computerized Legal Data and Information: Contents, Access and Circulation," Association pour le développement de l'informatique juridique, Paris, October 23, 1998. A version of it (with active links to documents and web sites referenced in the paper) can be found through the author's homepage at http://www.lawschool.cornell.edu/faculty/.
TABLE OF CONTENTS

I. CONTENTS OF LEGAL DATA AND INFORMATION ON THE INTERNET

A. LEGAL DATA PROVIDERS

1. Federal and State Official/Quasi-Official Legal Information
   a. Federal and State Governmental Sites
   b. Universities and Non Profit Organizations

2. Commercial Databases
   a. Lexis and Westlaw
   b. Lois and Versuslaw

B. EVALUATION OF LEGAL DATA

1. Quality of Data
   a. Internet/Commercial/Print
   b. Impact of Legal Information on the Internet on Legal Research/Access to Legal Information
   c. Impact on Legal Research

2. Access to Legal Information
   a. Search Engines
   b. Uniform Citation System

II. ISSUES RAISED BY DIGITAL MEDIUM

A. AUTHENTICITY AND RELIABILITY

1. Authenticity Issue

B. LONG TERM ACCESS TO DIGITAL LEGAL INFORMATION

1. Fragility of Digital Medium
2. Responsibility to Archive and Preserve for Future
I. CONTENTS OF LEGAL DATA AND INFORMATION ON THE INTERNET

Much legal information is currently available over the Internet, both in the form of raw data and edited data, from governmental and private sources, non-profit and commercial, and directed toward the public and legal professionals. In the United States, most government information is available free of charge, from official web sites, as well as from a number of non-profit organizations and universities. Sometimes, the information is presented as raw data, but often, it is enhanced, with search engines, some form of indexing, etc. It is therefore hard to contrast raw data and edited data. It is still true, nevertheless, that when it really matters, legal data coming from commercial publishers and accessible for a fee are usually more reliable and favored by legal professionals. Most of that digital information provides only a right of access, and no ownership, or control over the data, unless it is downloaded on a server, or stored on a CD-ROM. This is a matter of concern with regard to long term access to digital legal information.

A. LEGAL DATA PROVIDERS

1. Federal and State Official/Quasi-Official Legal Information

   a. Federal and State Governmental Sites

   In the United States, because of the American principle that citizens should have free access to government information, most federal and state publications are in the public domain, without being subject to copyright, and many official texts and documents are now available on the Internet. A study of the availability was conducted in 1996 for the first Association pour le développement de l'informatique juridique (ADIJ) conference, Paris, Oct. 23, 1998. Since 1996, the government sites have greatly expanded. Today, most current U.S. legislative, regulatory, and case law texts are available from federal and state government web sites, and steadily expanding. They

---


2 Some good commercial or non-profit gateway sites to governmental sites include: Washburn Law Library (http://lawlib.wuacc.edu/washlaw/reflaw/reflaw.html), Cornell Law Library's Legal Research Encyclopedia, which includes both full text electronic sources and references to print sources (http://www.law.cornell.edu/library/states/us.html), Findlaw (http://www.findlaw.com), National Center for State
are free, although some other government information is available only for a fee, e.g., Stat-USA, which contains statistical information. Http://www.stat-usa.gov.

b. Universities and Non Profit Organizations

The governmental web sites are completed by university (many of the federal court of appeals decisions) and other private or public non-profit web sites, as well as commercial ones. They are, in general, reliable sites, currently accessible free of charge. Activities from bar associations are growing, both the ABA and state bar association, and some have started including full text editions of newsletters and other documents. The trend continues to put law journals and law reviews online. The trade off is to get increased exposure and be read worldwide, at the risk of losing possible royalties. Many law school web sites offer the tables of contents of law review articles, which can then be requested for a fee.³

Universities took the initiative to make federal court decisions available over the Internet several years ago. For a while, federal appellate court decisions were only available thanks to university web sites. Now, five fed courts of appeals have their own web sites, the fifth (took the job over from the Univ. of Texas), Ninth, Tenth, D.C., and Armed Forces circuits. The others are still served by law schools. Today, there are about 35 US district court web sites, 30 bankruptcy court sites. Most states today have a web site and distribute their bills, statutes, court decisions and regulations over the Internet.

Non profit organizations sometimes take the official text, and improve upon it by providing a useful search engine, and linking to other sources, e.g., at Cornell, the Legal Information Institute collaborates with the U.S. Supreme Court for the delivery of U.S. Supreme Court decisions, and has improved the indexing and delivery of the official U.S. Code version provided by the U.S. government.⁴

---

³ A good gateway site to law reviews is the one of the University of Southern California Law School Library (http://www.usc.edu/dept/law-lib/frcontent.htm).

Information specialists and other legal researchers use the Internet more and more to get documents from various organizations, such as the ABA (American Bar Association)\(^5\) and the ALI (American Law Institute).\(^6\) A few years ago, sites contained mainly addresses and directory type information. Now full text articles and documents are widely available, and much experimenting is being conducted.

The Australian Legal Information Institute presents a model of a non profit organization, created expressly for the purpose of facilitating free access to Australian primary legal materials.\(^7\) AustLII has been most successful in obtaining free public access to data from government and court sources that often previously could not or would not provide the data.

2. Commercial Databases

The Internet has levelled the playing field, but the information available on the commercial databases is still more trustworthy. Many more commercial sites have blossomed, with much valuable information. Some are free, and serve as gateway sites to entire areas of the law, though links to a variety of web sites (e.g., Hieros Gamos,\(^8\) Findlaw,\(^9\) etc.). They rely on advertising to cover their costs. Some are fee-based.

\textit{a. Lexis and Westlaw}

Lexis and Westlaw are getting stronger, benefitting from increased mergers and consolidation in the legal publishing industry.\(^10\) They offer a greater degree of trust, and are still the tool of choice of legal professionals who can afford the high rates. They offer special rates for smaller law firms. However, legal professionals only get a right of access to the data, and no ownership of the legal information. U.S. law schools benefit from academic subscriptions, to encourage the new graduates to use the services at

---

\(^5\) http://www.aba.org.  
\(^6\) http://www.ali.org.  
\(^7\) The way in which this has been carried out is discussed in detail in "The Politics of Public Legal Information," at: http://ltc.law.warwick.ac.uk/jilt/LegInfo/97_2gree/paper4.htm.  
\(^8\) http://www.hg.com.  
\(^10\) For an up-to-date list of "who owns whom," check the site of the University of Colorado Law Library: http://www.Colorado.EDU/Law/lawlib/ts/legpub.htm.
commercial rates upon graduation. Some of the materials are not part of the subscription.

b. Lois and Versuslaw

Lois\textsuperscript{11} and Versuslaw\textsuperscript{12} offer lower priced alternatives to the two giants, Lexis and Westlaw. These two new entrants have commercialized the information provided free through the Internet and are making it into attractive packages geared toward legal professionals. They are lower-priced than Lexis and Westlaw, and offer claims of great reliability, including ownership of CD-ROM with the text of the documents. They are smaller databases, limited to primary legal information.

B. Evaluation of Legal Data

1. Quality of Data

a. Internet/Commercial/Print

When it really matters, there is still a great degree of reliance on the "official" word of the law. Serious people still go back to "official" documents, and print publications. Student-edited law reviews prevalent in law schools are still under the obligation to cite to the original print source when they do cite checking of sources. For instance, they find it on Lexis, but have to verify the source by examining the actual print source. Courts have been slower to make the transition to a reliance on the electronic text. For some purposes, the legibility of the electronic text still leaves to be desired, e.g., Lexis still has footnotes of law review articles in the middle of the text. Because of the lack of regulation of the Internet, many people are experimenting. Because of the interconnectivity of the Internet, readers write to the Web site to explain if something is wrong.

The contents of legal data are improved by the use of a reader, such as Adobe, so that the result is better than the raw text, because the document looks like the print product. This gives a better guarantee of authenticity, and is also convenient, and aesthetically pleasing. The PDF (Portable Document Format) Adobe Acrobat format is now the standard for federal government

\textsuperscript{11} lois www.pita.com.
\textsuperscript{12} versuslaw www.versuslaw.com.
publications, even though it is criticized by some because of its proprietary nature. It reproduces the physical appearance of a page much better than technologies such as HTML.

b. Impact of Legal Information on the Internet on Legal Research/Access to Legal Information

Some solid research skills are lost with the transition to the digital format. In the print world, there was a distinct series of paper publications, with corrections in-between. In the electronic world, cases are not systematically updated or corrected after they are put online. Is the citizen better served? Also, there are some limitations to getting the plain text of the law. How much can one understand the law by looking at a text? If no context is provided, it may be harder to understand the issues, the procedure, etc., which are provided in a commercial system such as West, with headnotes, annotations, etc.? The greatest danger is for non professionals who get the letter of the law, but not the context.

The digital medium has opened up new fields of legal research, e.g., empirical studies. You can take statistical data and manipulate them, to study social indicators, and the like, e.g., studies on jury verdicts conducted by Professors Ted Eisenberg and Kevin Clermont at Cornell Law School on their federal statistical web site.\(^3\)

c. Impact on Legal Research

The reliance on Internet search engines leads to the loss of a lot of sophisticated indexing tools, such as subject and digest keyword indexing, the elaborate system created by West since the end of the 19th century.\(^4\) The Internet makes legal information much more accessible to the public. But, it is not clear that the greater accessibility makes the law more understandable, because it may lack a context. People can misinterpret the text of the law, unless there are disclaimers. It may also put a greater burden on legal profession to explain the law.

\(^{13}\) http://teddy.law.cornell.edu:8090/questata.htm.

2. Access to Legal Information

   a. Search Engines

   A great deal of progress in finding legal information on the Web comes from the increased sophistication of search engines. Search engines are essentially of two kinds, human-mediated "intellectual" indexes and "robot" or automated indexes. In the intellectual indexes, individual web sites are classified by hand according to various classificatory schemes, such as the popular Yahoo!, albeit not law-specific. "Robot" or automated indexes use programs that download every page of the web, so that every word on every page can be indexed by a remotely located search engine, e.g., Alta Vista, and LawCrawler.

   An April 1998 study by the journal Science concludes that search engines are not thorough in finding relevant documents, because they each only index a fraction of the total documents available on the web. The lesson is not to rely on just one engine. One option is to use a meta index, which combines the results of several search engines, such as MetaCrawler. The slight downside is that the search capabilities are limited.

   Other new indexing tools include search engines created on gateway sites, such as GILS Government Information Locator Service, GPO Access, an online search tool that enables the researcher to search through many databases of government documents, and search engines developed to search large topical sites, such as the one of the United Nations.

---

15 The following is adapted from an excellent article by Graham Greenleaf, Professor of Law and Co-Director of AustLII, University of New South Wales, "Moving access to law into the 21st century," http://www2.austlii.edu.au/~graham/AALS/Restatement-A.html.
19 A synopsis of the article is available at http://www.sciencemag.org. Information about the authors can be found at NEC Research Institute: http://www.neci.nj.nec.com.
b. Uniform Citation System

Encouragement of the development of uniform standards for the citation of cases and other legal documents is important in ensuring uniformity. In the United States, the detailed proposal of the American Association of Law Libraries (AALL) for a national standard, was adopted by the American Bar Association Special Committee on Citation Issues (Resolution of August 6, 1996). The Resolution recommends a generic citation format where each decision is given a sequential serial number and internal paragraph numbering assigned by the court. This system is to be equally applicable to printed and electronic case reports, and therefore medium neutral.

Example: Smith v. Jones, 1996 5 Cir 15, par. 18, 22 F.3d 955

1996 is the year of the decision, 5Cir refers to the United States Court of Appeals for the 5th Circuit; 15 indicates that it is the 15th decision rendered in that year. 18 is the paragraph number in the opinion, and the rest is the parallel cite to the volume and page in the printed case report.

Both the AALL and the American Bar Association have approved the Committee’s recommended format, and have urged federal and state courts to adopt the vendor-neutral and medium-neutral citation standards. The Judicial Conference of the United States declined to adopt the ABA’s proposal, but that may change in the future.

II. ISSUES RAISED BY DIGITAL MEDIUM

A. AUTHENTICITY AND RELIABILITY

1. Authenticity issue

The authenticity of digital information is an issue, in a network which has been called “the Net of a Million of lies.” In the print world, the medium of paper automatically authenticates the content. In the digital world, there

---

23 http://www.aallnet.org/committee/citation.
24 See http://www.abanet.org/citation for history, text of resolution, and updates on which courts have adopted it.
25 The section below draws on the papers written for the Canadian Conference, The Official Version..., listed in the Bibliography.
26 Phrase used by Verner Vinge in Fire upon the Deep (1992).
is a disintermediation with the medium, and special care needs to be taken in the production of the information. Everyone can be a web author, and create a fake web site. The Internet is decentralized and unregulated. Can it become self-regulated? Anybody can be an author. A web whiz can put legal information online, without any need for legal training. On the other hand, authenticity problems are more likely to result from error rather than deliberate attempt to corrupt. It is a question of risk management, and whether the risks vary with the type of record. Some of the ways to provide assurances about authenticity have to do with the reputation of the source, including commercial and government publishers. Some of the technological solutions include the technique of the digital signature which uses public key cryptography to insure the integrity of the record—that it has not been altered—and the source of the record and digital watermarking. It is important to develop a partnership with the Information technology industry.

B. LONG TERM ACCESS TO DIGITAL LEGAL INFORMATION

1. Fragility of Digital Medium

Documents are going digital for good reasons, including easy distribution and access over the Internet, as well as hypertext and multimedia capabilities. The process is irreversible politically and economically. However, digital information is characterized by fragility and rapid technological obsolescence. Under good conditions, books printed on acid-free paper, e.g., official state reports and codes, will last for centuries. The lifespan of a CD or disk is estimated at 10 to 30 years, but its lifespan is further limited by the hardware and software needed to read it. This means that digital information may become obsolete within five years unless it can be refreshed or migrated to a newer technology. Refreshing data (copying it periodically to more stable media) cannot solve the long-term problem. It can save simple ASCII files, but anything more complex may lose functionality that was built into it. Migration means moving files to a new system. It has risks, too, such as loss or change of information in the translation. Emulation consists of designing hardware and software that emulate the old system. Much research needs to be done on solving these technological issues.

2. Responsibility to Archive and Preserve for Future

Beyond the technical problems related to the fragility of the digital medium, there are a number of financial, legal, and policy issues at stake with digital legal information. Who will decide what to preserve and pay for preserving digital information? Who is going to be responsible for continued and long-term access to authoritative digital primary legal information sources? This is especially important in a democracy where free access to government information is supposed to be a right. It is assumed here that most primary legal information, at least in the US, is in the public domain. However, copyright may be an issue for some digital records in the US and for many sources abroad because of different governmental policies.

With the advent of the paperless documents, there is a new issue of the obligation of the government to maintain electronic records. A current controversy in the United States involves concerned citizens and the National Archives. The National Archives and Records Administration advised its agencies that they could delete and destroy their e-mail and other electronic records as long as they printed out copies and saves the paper copies. A group of researchers, librarians, historians and journalists have sued NARA charging that vital data could be lost, (Public Citizen v. Carlin, text available on Public Citizen web site) and that electronic government records need to be made available to researchers. The issue is currently under federal litigation. A federal district court judge declared the policy "null and void." The judgment is currently appealed by NARA, arguing that the government has no system capable of storing the voluminous electronic output of a bureaucracy in the electronic age. Both groups are using their respective web sites to buttress their respective positions.

The previous example relates to electronic mail. It can be safely assumed that official primary legal information in digital form needs to be accessible far into the future. However, no one has yet taken the responsibility to archive digital information for long-term public access. There is therefore a risk of loss of information. The U.S. Government Printing Office — which


has traditionally published on paper the highest federal court decisions and federal legislative materials, such as *U.S. Supreme Court* decisions, *U.S. Code, Federal Register, Code of Federal Regulations*, and much else, and distributed them to the elaborate system of 1400 depository libraries (including most law schools) — has made a commitment to disseminate its publications in electronic form. But it has not committed to serve as an archive, because it is not in its legislative mandate. It may have the willingness, but not the budgetary appropriation.

Several prospective models have emerged as potential archival sites for legal information.\(^\text{30}\)

1. **Individual Researchers.** Individual researchers, universities who publish legal information) and issuing agencies (e.g., legislatures and courts) currently often archive their own materials. This may cause problems when the research project stops, or the database exceeds the server capacity of the issuing body.

2. **Federal, State, Local, and Foreign Governments.** Governments will or should take responsibility to preserve its own digital publications. Or they should provide funding for others to carry out digital preservation. The question will be the money available. The government involvement may depend on whether preservation of digital information is seen as a national policy issue.

3. **Publishers.** Publishers are already providing archives of digital legal information to their subscribers. Will they be available in perpetuity? What if the publisher goes out of business or the maintenance of the archive becomes unprofitable. Also, should citizens rely on commercial entities, or even the government, to find the official word of the law, or should there be a reliable, neutral source, in addition to others available, similar to the current print official state court reports (medium of print authenticates content, not so with digital information).

4. **National Libraries.** National libraries may also play the role of archiving information through legal deposit programs. Depository legislation for electronic information varies widely among nations that have deposit laws.

---

Copyright may be a problem in some countries. Continuous budget appropriations may be needed to implement.

5. Professional Library Groups. Other groups, such as commercial entities (OCLC, RLG, www.rlg.org) are ready to take on a role. Should they? Several library organizations are also involved in the archiving of digital information. The Digital Library Federation (DLF), the Coalition for Networked Information (CNI, http://www.cni.org), the Association of Research Libraries (ARL, http://www.arl.org/), the Council on Library and Information Resources (formerly Council for Preservation and Access (CLIR, formerly CPA, http://www.arl.org/), the Inter-university Consortium for Political and Social Research (ICPSR), located at the University of Michigan, the National Archives (NARA), etc. RLG and CLIR (formerly CPA) have made a proposal to certify digital preservation operations and create fail-safe preservation and storage. That may be worth studying as a possible model.

6. Consortia of Law Libraries (academic, court, state and county, etc.). Over the centuries, libraries have played the role of preserving and making information available to present and future generations of scholars. One proposal to consider, in addition to other preservation measures taken by courts, legislatures, and publishers, would be for libraries to form consortia agreements. Each would become responsible for digital preservation of part of the corpus of official primary legal information, and rely on others for other parts. The scheme would extend access to that information far into the future. AALL and the Law Library of Congress could play a vital role in this venture, by coordinating efforts with the major stakeholders.

A few libraries are moving in that direction in other fields. One notable example is the Mann library at Cornell University, which has taken responsibility for archiving research publications in digital form in the field of agriculture, in cooperation with the National Library of Agriculture and landgrant university libraries. No mechanisms exist yet for national and international coordination in the legal field.

These important issues are being studied both in Canada and in the U.S. Following in the footsteps of the Canadian Association of Law Libraries (CALL), the American Association of Law Libraries (AALL) and the Law Library of Congress are currently starting a process of identifying the stakeholders considered to be producers, keepers, and consumers of digital legal information, and organizing a National Summit Conference. The Conference will include judges, legislators, regulators, members of the
practicing bar, academics, librarians and archivists, court administrators, official printers, public and private publishers, Internet providers, webmasters, and computer science engineers. Judy Meadows, 1997-98 AALL President, has been talking with various groups to determine interested parties to involve in the Summit and she has appointed a Task Force on Preservation and Access to Digital Legal Information. There was an exploratory program at the AALL annual meeting in Anaheim, California, in July 1998, followed by a Planning Meeting at Cornell Law School, Ithaca, New York, in August 1998. That work is now continued by the AALL Committee on Authentication and Preservation of Digital Law, which is meeting at the U.S. National Center for State Courts in December 1999.

CONCLUSION

The contents of the Internet are becoming richer every day. One issue of particular importance that has emerged now is the need to have access to the permanent digital records far into the future. Current efforts have the same goal, to make sure that in a paperless world there will be a permanent record of the law in its many forms, and that the document will be authentic. They demonstrate the importance to work with partners on joint problem solving, including the legal information publishing industry, the information technology industry, computer scientists, and other interested stakeholders. What is at stake is the transmission of official documents, "the word of the law," to future generations.
SELECTIVE BIBLIOGRAPHY

Newsletters


Useful bibliography of articles on the Internet of interest to legal researchers.


Excellent review of web sites.


Evaluates and rates web sites.

Background Reports


Good summary of current issues. Discusses some strategies and models currently explored, as well as specific initiatives from libraries and professional associations.


Co-sponsored with the Commission on Preservation and Access.


These two publication of the UK Arts & Humanities Data Service (AHDS) provide useful references and links to organizations and digital library initiatives in the US and the United Kingdom.


Organizations


Comprises 121 large North American research institutions, both in the U.S. and Canada. ARL at: 21 Dupont Circle, Suite 800, Washington, DC 20036, voice: 202-296-2296. Fax: 202-872-0884; arlhq@arl.org. Executive Director: Duane E. Webster.


CNI is an organization to advance the transformative promise of networked information technology for the advancement of scholarly communication and the enrichment of intellectual productivity. Founded in 1990 by the Association of Research Libraries, Educom, and CAUSE, CNI is supported by the members of an institutional Task Force representing higher education, publishing, network and telecommunications, information technology, and libraries and library organizations.

Digital Library Federation (DLF).

Founded in 1995 to establish the conditions for creating, maintaining, expanding, and preserving a distributed collection of digital materials accessible to scholars, students, and a wider public. The Federation is a leadership organization operating under the umbrella of the Council on Library and Information Resources. It is composed of participants who manage and operate digital libraries. A consortium of fifteen of the US's largest libraries and archives cooperating to ensure access to digitized materials. The site provides links to member organizations with a range of digital library access and preservation projects. Contact: Donald J. Waters, Director, Digital Library Federation, Council on Library and Information Resources, 205 Church Street, Third Floor, New Haven, CT 06510-1805. Phone: +1-203-498-6076; Fax: +1-203-498-6078; Email: dwaters@clir.org.


NARA currently does not, in general, post any electronic records on the Internet.

Contact: Ricky Erway, Member Services Officer, Digital Initiatives, RLG, 650-691-2228 bl.rle@rlg.org.


Independent agency within the executive branch, created in 1970, to advise the US president on national policy issues relating to libraries and informational needs of the country. 1110 Vermont Avenue, N.W. Suite 820, Washington, DC 20005-3522. Telephone: 202-606-9200; Fax: 202-606-9203. Jeanne Hurley Simon, Chairperson; jhsimon@siu.edu. Robert S. Willard, Executive Director; rw_nclis@inet.ed.gov.