FDL Conference and Fall DLC Meeting

AALL Authentication Report and National Summit

Mary Alice Baish and Sarah G. Holterhoff American Association of Law Libraries October 17, 2007

AALL Reports on the Electronic Life Cycle

 2003 AALL 50-State Report on Permanent Public Access to Electronic Government Information (PPA).

 2007 AALL 50-State Report on Authentication of Online Legal Resources.

2006 State Authentication Survey

- Goal: to determine which states, if any, have adopted website versions of primary legal resources as official and/or authentic.
- Six online state sources checked:
 - administrative code
 - administrative register
 - session laws
 - statutory code

- state Supreme Court
 - intermediate appellate court information, including opinions

What is *Official?*

- Mandated or approved by statute or rule.
- An online official legal resource has the same status as a print official legal resource.
- The fixed nature of print, plus multiple copies and wide distribution, ensure that the print official legal resource is an authentic resource.

GPO's Definition of Authentic

- Content verified by a government entity to be complete and unaltered when compared with the version approved or published by the content originator.
- Authentic text will typically bear a certificate or mark that conveys information as to its certification.
- Encryption, digital signatures, PKI & "chain of custody" planned for GPO's Future Digital System (FDsys).
- New GPO Pilot: Authenticated Public and Private Laws of the 110th Congress.

- States are discontinuing print official resources and substituting online official sources.
- This trend will continue.
- State agencies believe they are serving citizens best by providing online access to legal sources.
- State agencies save print and distribution costs by publishing only online.

Online Versions Are the Sole *Official* Resource

- 5 states Alaska, Indiana, New Mexico,
 Tennessee and Utah give official status to an online legal resources and have eliminated print.
- None are capable of being authenticated and only Utah requires PPA.
- The disappearance of print official legal resources without an authentic online substitute threatens trustworthiness of the resources.

- Ten states & D.C. have designated as official one or more of their online primary legal resources.
- Alaska, Indiana, Maryland, Michigan, Minnesota, New Mexico, New York, Tennessee, Utah and Virginia & the District of Columbia.
- 23 sources of law (primarily regulatory) are available in online repositories that are considered official.

- Eight states have "official traits," but evidence as to the actual status of the resources is conflicting.
- The word "official" is not always used on a resource; or,
- "Official" is there, but the online resource isn't considered <u>as</u> official as the print version; or,
- The *official* status is noted with no information as to why it is *official* (e.g., statutory authority or court rule).

- States have not been sufficiently deliberate in their policies and practices.
- States have not acknowledged important needs of citizens and researchers who seek government information.
- States have not considered the issues raised when their only legal resources are not authenticated.
- New ARJD Principles for "Official" On-line Documents parallel our findings.

Our Recommendations

- Provide a clear statement of the official status and source of authority for same.
- Identify the print official sources.
- Identify source of data, its "chain of custody" and relevant processes.
- Prominently display any representations and disclaimers, and provide the specific scope of such.
- Address official status, authentication and PPA for online resources.

- No state's online primary legal resources are authenticated or afford ready authentication by standard methods.
- Minnesota, Ohio, Vermont, and Virginia are beginning to address the problem.
- Eight other states Alabama, Arkansas, Connecticut, Maryland, Montana, South Carolina, and Tennessee perceive authentication as a specific concern that warrants attention.

- Since our 2003 report, nine states have provided for PPA for one or more of their online primary legal resources.
- AK, CA, IL, IN, MN, MT, OH, PA, TX and UT
- Typical language: "The state library shall coordinate with state agencies and depository libraries to ensure permanent public access to state publications."

Conclusions

- Online legal resources are increasingly the sole official published source.
- Official status requires authentication procedures (encryption, digital signatures, PKI, "chain of custody" information).
- The goal is that online legal resources will be as trustworthy as print.



Authentic Legal Information in the Digital Age: A National Summit

Renaissance Schaumburg Hotel & Convention Center Schaumburg, IL April 20-21, 2007

National Summit on Authentication of Online Legal Resources

- Convened by AALL, April 20-21, 2007.
- 50 delegates from ABA, NCSL, NCCUSL, state courts, state archives, state legislative IT services & federal officials.
- Sessions on Authentication Report and findings, technological and legal challenges.
- Breakout sessions on challenges and on building alliances.
- A great success, but just the beginning!

Where Do We Go From Here?

- Education and outreach articles, programs.
- Technology initiatives develop standards;
 possible state initiatives in CT, DE, MN, NM, WA.
- Legal initiatives NCCUSL study committee; changes to court rule; "best practices" manual.
- Advocacy build alliances at state, national and international levels.
- More on Summit & follow-up at: http://www.aallnet.org/summit/

Member Log in

Join AALL



SEARCH

HOME PAGE

ABOUT AALL

ANNUAL MEETING

CALENDAR OF EVENTS

CAUCUSES

CHAPTERS

COMMITTEES

EDUCATION & CAREERS

EXECUTIVE BOARD

JOB HOTLINE

MEMBERS ONLY SECTION

NEWS

ONLINE DISCUSSION

PLANNED GIVING

PROFESSIONAL EDUCATION

PUBLICATIONS

REPRESENTATIVES

SPEAKERS DIRECTORY

SPECIAL INTEREST SECTIONS

STRATEGIC DIRECTIONS

WASHINGTON OFFICE

CONTACT AALL STAFF

Authentic Legal Information in the Digital Age: AALL National Summit

Renaissance Schaumburg Hotel & Convention Center Schaumburg, IL April 20-21, 2007

National Summit Agenda

- Print vs. Digital Assumptions and Differences Professor Robert Berring
 - o "Losing the Law: a Call To Arms," 10 Green Bag 2d 279 (2007)
- Session One: AALL's State-By-State Report on Authentication of Online Legal Resources
 - o <u>State-by-State Report on Authentication of Online Legal</u> Resources
 - o Authentication of Online Legal Resources: Key Findings
- · Session Two: The Technological Solutions for Best Practices
 - o The Technological Solutions for Best Practices
 - o U.S. Government Printing Office: FDsys Content Authentication
 - O What Needs Authentication?
- · Session Three: The Legal Solutions
 - The Legal Solutions
- Authentication of Digital Legal Information—Where Do We Go From Here?

Your Help Is Needed!

- To create a robust alliance in your state to educate, promote solutions and provide grassroots support.
- To work together to overcome technological obstacles in your state and promote solutions.
- To seek legislative changes to ensure that digital authentication and permanent public access of online government information become a reality in your state.

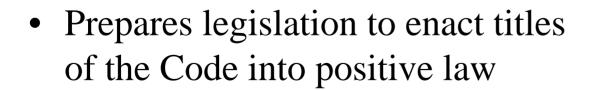
The U.S. Code & Authentication

Peter G. LeFevre, Law Revision Counsel

Office of the Law Revision Counsel

Nonpartisan office in the U.S.
 House of Representatives

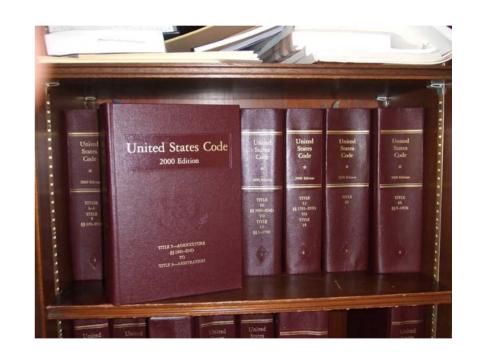
Prepares and publishes the U.S.
 Code





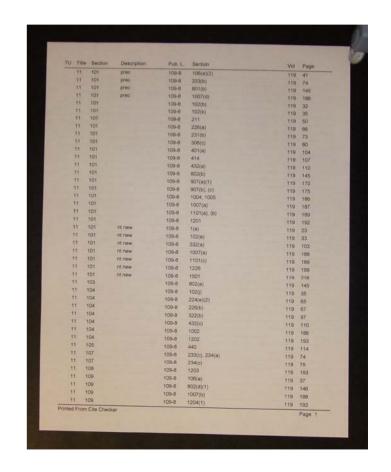
What is the U.S. Code?

- Contains the permanent and general Federal laws
- Organized into 50 broad subject matter titles
- Positive law titles are enacted into law as titles of Code
- Non-positive law titles are editorial arrangements of sections from hundreds of different statutes



Classification of Laws

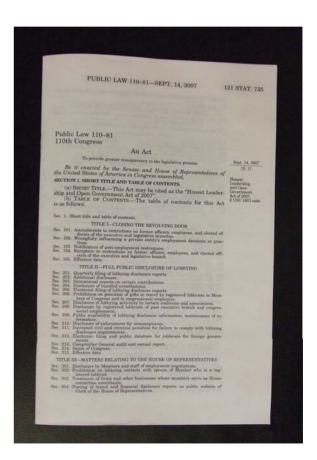
- A team of lawyers in OLRC decides what laws go into the Code.
- Permanent and general laws are included, but temporary and special provisions are not.
- New provisions are located in the Code based on subject and content.
- Classifications are listed in the classification tables on the OLRC website.



Classification Table

Updating Laws

- Electronic copy of slip law is obtained from GPO.
- OLRC process sorts classified provisions.
- Text editing program used to make changes in text of Code.
- Editorial aids prepared –
 source credits,
 amendment notes, etc.



Slip Law

Volume and Nature of Federal Legislation



Statutes At Large

• 5,000 to 7,500 pages of new law per Congress

Plethora of amendments,`cut and bite' amendments

 Code sections reflect law as amended

Accuracy of the U.S. Code is paramount

- Potential for error is high because of number of updates.
- Even small errors can be significant.
- Code is relied on as authoritative source of Federal law.



Legal Requirements for Code publication

- There must be an annual update of the Code.
- Each main edition and supplement must be <u>printed</u> by GPO.
- The printed main edition and supplement of the Code are evidence of the law in all State and Federal courts and public offices.

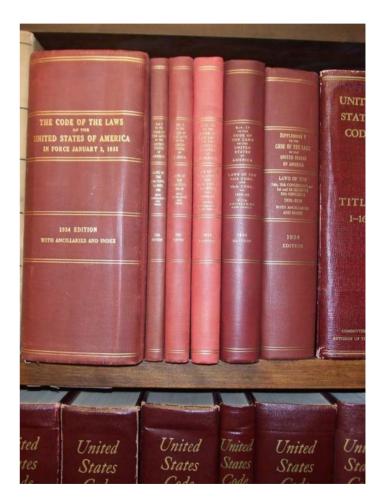




Supreme Court

How is the Code published?

- Since 1934, there has been a new main edition every six years and a cumulative supplement every year in between.
- Within a supplement or main edition, the Code is updated on a title by title basis.
- Volumes are released for printing as completed.



1934 Edition

Electronic database

- Code database was created in 1976 and has been maintained by OLRC ever since.
- Database is used for all of the following:
 - by GPO to print the U.S. Code.
 - OLRC website
 - GPO Access Code website
 - Cornell Legal Information
 Institute Code website



Features of OLRC website

- Searchable version of the current Code—specific section and word searches possible
- Searchable versions of past editions of the Code
- Downloadable versions of the Code in ASCII text, PDF, and GPO locator formats
- Classification tables linked to search results to provide alerts of new laws affecting sections



http://uscode.house.gov/

Difference between Code versions

- Same database is used for printing main edition and supplements.
- Supplements contain only new material and parts of existing sections that have changed.
- Website is "virtual main" and usually a composite of two supplements.
- Formatting results in differences in appearance of text in different versions.
- Every page of the printed version is checked by 3 proofreaders--not so for the virtual main.



Authentication and status of online Code

- The online U.S. Code is the type of document for which authentication is especially important.
- Federal laws apply to everyone in the U.S., and the online Code is relied on for legal research.
- An authenticated online Code should supplement, not supplant, the printed Code.
- The printed version should remain the official version.





Peter LeFevre

Oct. 17, 2007

Intro and Office

Good morning, it's a pleasure to be here in front of this audience. I've been a big fan of libraries ever since I got my first library card in Chicago at age 4. You folks perform a wonderful public service.

The Office that I work for is a nonpartisan office in the U.S. House of Representatives responsible to the Speaker of the House. It was created in 1975 and given two main functions:

- -- (1) to prepare and publish new editions and supplements of the United States Code: and
- -- (2) to prepare legislation to enact titles of the Code into positive law.

What is the Code?

Most of you, no doubt, have some familiarity with the U.S. Code. But I'll talk a little bit about what is in the Code and how it is put together.

When a law is enacted by Congress, it is initially printed as a slip law and then later as part of the Statutes at Large. The Statutes at Large includes all of the laws as originally enacted in chronological order.

The Code is more selective and organized differently. It contains only the permanent and general laws, and it arranges them into 50 broad subject matter titles. 24 of the 50 titles have been enacted into positive law. The remaining titles, referred to as non-positive law titles, consist of sections from many different laws that have been editorially organized into the titles. The Code is now well over 40,000 pages and growing every year.

Classification

Only Congress can create and determine what goes into a positive law title, but the arrangement of laws in the non-positive law titles is a function of the Law Revision Counsel's Office. As soon as a bill is passed by both the House and Senate, a copy of it is sent to our office. We have a team of attorneys who carefully read through the bill and decide which provisions should go in the Code,

and if so, where they will be placed. This process is called ``classification''. Any amendment to a section already in the Code will be classified to that section. For new provisions, we make a judgment as to whether or not they are permanent and general. If they are, we will try to find the best place to fit them in the Code. Some laws will be picked up in their entirety and made into new chapters. Many others contain a mixture of new provisions and amendments and cover more than one subject area. These laws have to be split up into dozens, and some times even hundreds of pieces, which are dispersed throughout the Code. We keep track of all these pieces with our classification tables which are posted on our website. These tables provide the most up-to-date information available on the codification of new laws. They can be used to identify all the sections of the Code affected by a particular law and also whether a particular section of the Code is affected by any new laws.

Updating laws

electronic copy

After a new law is classified, the next step is to incorporate it into the existing Code. We begin this step by obtaining an electronic copy of the slip law from the Government Printing Office. We take that electronic copy and feed it into a process we have which extracts the classified provisions and sorts them out according to what they affect in the Code.

By law, the printed slip law is considered competent evidence of the law, but what we actually use is an electronic copy of the slip law that we download from a GPO file transfer site. We have to rely on GPO to give us the correct electronic copy because we do not have the time to verify that every word we take from that copy is identical to the printed slip law. Has there ever been a problem with this? Unfortunately, the answer is yes. There has been at least one instance where the electronic copy of a law made available to us was not the right one. We discovered this during the review process when an editor noticed that the text of a new section we were adding to the Code did not match the slip law. It turned out that GPO had temporarily placed the wrong file on their transfer site. There have also been a few cases where a slip law has differed from the

enrolled bill. In these cases, GPO has had to print a corrected slip law, known as a star print. This is one of the reasons why I am very interested in the work being done by GPO and the American Association of Law Libraries to advance authentication of electronic documents.

Once our office has sorted out the provisions of a law, we do the actual text editing and preparation of editorial material that explains and documents each change in the Code. Because of the volume and nature of Federal legislation, this editing process requires a great deal of work. On average, a two-year Congress enacts between 5,000 and 7,500 pages of new law. Much of it is amendatory, and Congress, unlike many State legislatures, tends to favor the use of ``cut and bite" amendments. Rather than replacing an old section with a new one, these amendments make changes by specific word and phrase substitutions, and there can be a lot of these. For example, the recent Bankruptcy Reform law made 46 separate amendments to the first section of the Bankruptcy title. The rest of that 200 page bill made similar types of amendments to more than 160 other sections of that title.

It is not unusual for positive law titles and major acts in the Code to be amended hundreds and even thousands of times over the years. The Social Security Act, as originally enacted in 1935, started out with 87 sections and took up 28 pages in the Statutes at Large. Today, the Social Security Act has more than 450 sections, and some of those sections are more than twice as long as the entire Act was in 1935. If a judge or lawyer had to reconstruct one of those sections from scratch today using the Statutes at Large, it might take a couple of days to do it. Fortunately, that isn't necessary because the Code shows the Social Security Act as it currently reads and provides a source credit and an explanation for each amendment that has affected the Act since 1935.

Accuracy important

As you would expect, accuracy and precision are essential in maintaining the Code. Certainly, failing to execute an amendment would be a problem, but even putting a comma in the wrong place can change the meaning. Cases have been decided by the Supreme Court on the basis of the punctuation used in a

law. My staff and I understand that the official U.S. Code is what lawyers, judges, agencies, and the public ultimately have to rely on as the authoritative source for much of the federal law, so ensuring the accuracy of the Code is our highest priority.

Legal Requirements

Those are the basic steps we follow in producing the Code, but what does the law say about the publication and official status of the Code. There are six sections in the Code that deal with this. Basically, they require that--

- (1) there be an annual update of the Code, incorporating the laws enacted in the latest session of Congress,
 - (2) the Code and supplements must be printed by GPO, and
- (3) in all State and Federal courts and public offices, the then current **printed** main edition and supplement shall be considered evidence of the laws they contain.

As a practical matter, how is Code published?

Since 1934, there has been a new main edition published every six years and a cumulative supplement published each year in between. Each main or supplement cumulatively incorporates the laws enacted during one session of Congress and is current through the last law enacted in the covered session.

Within a supplement or main edition, the Code is updated on a title by title basis. This means that we start with title 1 and work our way through the 50 titles in order. When a title is finished we release it on our website, and as we complete work on enough titles to fill a volume of the Code, we send the volume to GPO for printing. Under this system, we must wait until the end of a session before we can release even the first title, because the last law enacted in a session might affect something in that first title.

Electronic database

For many years the actual printing of the Code was done using the hot metal type-setting process. In 1976, the first edition of the Code was printed using a computer database with GPO photo composition codes. To make that possible, every word of the 1970 Code and fifth supplement was double keyed to

create the Code database. Since then our office has maintained and updated that database, and that is the database that is used today to print the Code and for our online Code. It is also the database that is used by GPO for their U.S. Code on GPO Access and by Cornell Legal Information Institute for their Code website.

Website

Although, our primary focus is on the print version of the Code, our website is very important to us. The main feature of the website is the searchable version of our most current titles of the Code. This version allows a user to call up a specific section or research the law by doing word searches.

Our website also has searchable versions of 14 past editions of the Code and downloadable versions of current and past editions of the Code in various formats including ASCII text, PDF, and GPO locator. We also carry current and past sets of our classification tables with the most recent tables linked to our searchable Code, so that when you search for a specific section of the Code, you will be alerted of any new laws that affect the section.

Although, the same database is used for both the print and online versions of the Code, there are differences between them. Everything in the database is tagged as ``Main only", ``Main and Supplement", or ``Supplement only". When printing, the tags make it possible to include everything in a main edition but only new material and changed text in a supplement. That is why people using the books will need both the main and supplement to get the entire law. On the website, there is what we refer to as a ``virtual main edition" published for each year. We have programmatically merged the main and supplement data, so that the reader is looking at the entire Code as if there was a new main edition every year. However, most of the time, this virtual main is actually a composite of titles incorporating new laws from two supplements. This is because we update the Code on a title by title basis, and we start loading new titles on the website as we finish them. So at any given time, the searchable U.S. Code on our website will likely be a combination of two supplements. For example, on October 1, we

updated our online database to supplement V for titles 1 through 41, but titles 42 to 50 were still based on supplement IV.

Two other differences between the online and print versions are worth mentioning. Because different formatting is used in the digital and print versions, the fonts, capitalization, headings, indentations, and some characters are not the same in both versions. The printed version is more true to the statutes enacted by Congress. And finally, the printed version undergoes more checks than the online versions. Every page that appears in print is reviewed by three proofreaders, but when a virtual main edition is generated for our website, no one actually looks at each section. We take it on faith that the computer programs will properly merge the main and supplement data into an accurate rendition of each section. Because of these differences and because there is always a chance that the wrong version could be loaded on our site or that our database could be corrupted, we carry a disclaimer on our website which cautions people using our database for legal research to verify their results against the printed version of the U.S. Code.

Status of digital and printed version of the Code

Now, I'll finish up with a few comments about authentication and the Code.

The Code database is being accessed hundreds of thousands of times each month. Some percentage of the hits is by people doing legal research and perhaps relying on the site to make important decisions. Those users should be able to know exactly what they are looking at, where it comes from, and that it has not been altered. I believe that there is a compelling need for authentication of the online Code database.

However, I would not like to see a digital version of the Code replace the printed Code as the official version, and I would not favor reducing the number of printed copies that are distributed. For one thing, only the Code books printed by GPO are considered evidence of the law in court. But in addition, as convenient and useful as the online Code is for research and access, it does not have many of the advantages of the printed Code. Browsing the books often gives the reader a better perspective on the law and flipping through pages in a book can often be

a better way figure out the law than scrolling through hundreds computer screens. Plus, it is my impression that there are still a lot of library patrons out there who prefer the books to the online Code.

But there is another reason why I think we need to keep the printed Code as the official version. As I've described it, the Code database is very large and complex, and continuously being changed. Yet it sets out the national laws that apply to everyone in the country, and it must be true to the enactments of Congress down to the smallest detail. Even with authentication technology, a computer glitch, a website failure, or some kind of tampering or corruption of computer files could undermine the integrity of an online Code.

Thinking about this issue caused me to remember George Orwell's novel Animal Farm. The animals took over a farm and one of the first things they did was agree on seven commandments which they painted in big white letters on a wall. Later in the story, one of the pigs, Napoleon, began to take over the farm and establish a dictatorship. As he broke each of the commandments, he would secretly change how it was written on the wall. He added or erased words so as to do away with the rights and protections of the other animals. Some of the animals were confused because the commandments seemed to be different from what they remembered, but in the end, they trusted what was written on the wall more than their own memories.

If a few words in the Code mysteriously appeared or disappeared, how many people would even realize it. Like the animals, we would be more likely to rely on the written word than our memories. By this, I'm not suggesting that anyone would try to subvert a digital official version of the Code, but there is some comfort in knowing that in every courthouse and in many libraries through out the country, there is a set of U.S. Code books which are permanent and which cannot be altered.

A world class system for managing official Government content, which will verify and track versions, track versions, assure authenticity, preserve content, and provide assure authenticity, preserve content, and provide permanent access.

access. A world class system for managing official Government content, which versions, assure authenticity, preserve content, and provide permanent access.

an integrated digital content management system

ass system for managing official **Government content**, which will

U.S. Government Printing Office

FDsys Content Authentication

DLC Session

October 17, 2007





GPO Mission

- To provide the three branches of the Federal Government with expert publishing and printing services.
- To provide perpetual, free, and ready public access to the printed and electronic information published by the Federal Government, in partnership with Federal Depository Libraries.
- To distribute, on a cost recovery basis, printed and electronic copies of information published by the Federal Government.





Challenges to meeting the Mission

- Access to government published information is now widely expected to be electronic.
- Digital information needs to be authentic and verified to be the correct version.
- Digital information needs to be available for access almost immediately.
- Information needs to be preserved, making it available for generations to come.





Content-centric Solution – FDsys

- FDsys will automate the collection and dissemination of electronic information from all three branches of government.
- Information will be:
 - permanently available in electronic format
 - authenticated and versioned
 - accessible for Web searching, viewing, downloading and printing
 - available for conventional and on-demand printing





FDsys and Document Authentication

- How can a user be sure that the document they have is the one intended?
 - Need a process to verify electronic document attribution and integrity
- FDsys will use content authentication methods to ensure the following:
 - Content has been approved by, contributed by, or harvested from an official source; and
 - Content has been verified by GPO to be complete and unaltered





FDsys Authentication Workflow

- Content Originator (CO) deposits content with FDsys
 - FDsys authenticates user based on user role
 - CO enters metadata and deposits content
 - Content is digitally signed
- End User downloads content from FDsys
 - Integrity mark conveyed and displayed with content
 - Signature properties available for viewing
 - Provenance information available for viewing





Current Status

- Beta testing is underway
 - Test includes Public & Private laws from the110th Congress
 - GPO staff are manually signing the PDF files
 - Approach allows for testing of technology and user feedback before full release
- Future Plans
 - Automated signing is being developed now
 - Anticipate the capability to include all Public & Private laws on GPO Access
 - Available in early 2008





Additional Information

www.gpo.gov/projects/fdsys.htm

www.gpoaccess.gov/authentication/

