Ground-up law: Open access, Source quality, and the CFR

Sara S. Frug[*]
[*]*Legal Information Institute, Cornell Law School*

Abstract. In late 2009, the LII began a joint study with the Federal Depository Library program of the United States Library of Congress, the Government Printing Office, and the Office of the Federal Register to work through technical issues in producing a web version of the Code of Federal Regulations.. This paper discusses some of the challenges we faced and explores some of the implications for open access initiatives.

Keywords: Quality Assurance, Quality Control, Regulations, Standards, XML

The LII spent the better part of 2010 developing a web version of the Code of Federal Regulations (CFR). The Federal Depository Library Program (FDLP) of the U.S. Government Printing Office (GPO) approached the LII because of our experience processing government-standard typesetting code and our position within Cornell Law School, a Federal Depository Library. We entered into a joint study with the FDLP, the GPO, and the Office of the Federal Register (OFR) to analyze and work through technical issues in the typesetting-code-to-XML process. During the early stages of the joint study, the GPO made publicly available bulk XML source data through its FDsys program, and we began to use that XML as our source data.

Part 1 of this paper discusses what we built and why it was hard to build it. Part 2 provides some context for the project, notes some of the warning signs we might have heeded, and reflects upon the systemic nature of the problems we encountered.

1. Building a Fully-Featured, Open-Access CFR

1.1. THE PRODUCT

The Code of Federal Regulations is the annually-updated codification of United States federal regulations. Its text is derived from the central publication of final rules in the Federal Register, the official publication established by Federal Register Act of 1935 (44 U.S.C., Chap. 15). The CFR text is prepared by the Office of the Federal Register (OFR) and printed by the United States Government Printing Office (GPO). In print, it comprises over 200 volumes. In addition to the rule text, the OFR also publishes indexes, authorizing statutes in the Parallel Table of Authorities, and, to bridge the gap between the daily changes to the regulations and the annual revision cycle of the CFR, updating information in the List of Sections Affected.

The LII CFR provides a full-text version of the Code of Federal Regulations, one section per page. It displays the CFR's nested paragraph structure and pinpoint-linked cross-references. It supplies navigation - links to previous and next pages, breadcrumbs, and tables of contents. It allows full-text search and a citation-based search. It also provides a linked Parallel Table of Authorities and features related to updating. The architecture supports the development of future features such as persistent crowdsourced metadata.

Behind the scenes, the source XML is stored in an XML database, eXist, and presented through a content management system (CMS), Drupal. Pre-processing software retrieves the original XML from a government server. For each CFR Title , it compiles a single table of contents from the

multi-volume original. It marks up cross-references and nested paragraph structure. For each Title , it puts the table of contents and an XML file for each CFR part into the XML database.

The content management system generates HTML from the XML source and turns it into web pages. The first time a section is requested, the CMS queries the XML database (using XQuery with XSLT), which returns metadata and the section content, which the CMS processes, puts into its own database, then turns into a web page. At this point, the search engine, Solr, indexes the content. The CMS can then self-sufficiently fulfill subsequent requests for the same page.

The initial prototype for the CFR, based on the original typesetting code, was built in under two months. It took another year to launch an experimental beta using the bulk XML.

1.2. WHY IT TOOK A YEAR

If there had been only one problem with the data, it would not have taken so long. In reality, each problem in the data compounded the others.

1.1.1. Artifacts of print production

The CFR is logically organized into 50 Titles. In the print publication, each Title comprises one or more volumes. The XML source files for the CFR correspond to the print volumes. Therefore, in order to extract a per-Title table of contents, it is necessary to amalgamate the structure of all of each Title 's volumes' tables of contents.

Amalgamating the tables of contents would have been a straightforward engineering task, had the XML source files accurately reflected the structure of the CFR. Our initial attempt to combine the tables of contents involved walking up the tree from the section level, so that we'd find the whole structure even if a part spanned multiple volumes. But this approach required us to be confident that the hierarchy was correct. It turned out we could not.

1.1.2. Non-uniform authoring

The CFR presents the facade of a uniform document. Most of it contains standardized hierarchy and numbering conventions.

> The standard organization consists of the following structural units:
> (a) Titles, which are numbered consecutively in Arabic throughout the Code;
> (b) Subtitles, which are lettered consecutively in capitals throughout the title;
> (c) Chapters, which are numbered consecutively in Roman capitals throughout each title;
> (d) Subchapters, which are lettered consecutively in capitals throughout the chapter;
> (e) Parts, which are numbered in Arabic throughout each title;
> (f) Subparts, which are lettered in capitals;
> (g) Sections, which are numbered in Arabic throughout each part. A section number includes the number of the part followed by a period and the number of the section. For example, the section number for section 15 of part 21 is " § 21.15 "; and
> (h) Paragraphs, which are designated as follows:
> level 1(a), (b), (c), etc.
> level 2(1), (2), (3), etc.
> level 3(i), (ii), (iii), etc.
> level 4(A), (B), (C), etc.
> level 5(1), (2), (3), etc.

level 6(i), (ii), (iii), etc.

(1 CFR 21.11)

Yet, the regulations themselves provide a mechanism for agencies to use alternate numbering (1 CFR § 21.14), which, in many places is used.

Where we naively hoped uniformly to have numbering like:

"Title 24 › Subtitle B › Chapter I › Subchapter A › Part 110 › Subpart B › Section 110.25"

We had, e.g.,

"Title 41 › Subtitle B › Chapter 50 › Part 50-201 › Section 50-201.102"

and

"Title 48 › Chapter 8 › Subchapter G › Part 846 › Subpart 846.4"

which in turn comprised sections:
846.408,
846.408-70,
846.408-71,
846.470,
846.471,
846.472,
846.472-1, and
846.472-2.

So there was not an easy way to get correct collation of section numbers; nor was it easy to use the numbering system to check quality. Also, our original software used an adapted version of the unique identifiers we had used for the US Code; we had to abandon those, develop a robust system for generating new ones, and rewrite the parsing to accommodate them.

The lack of uniform numbering would pose little problem if (again) the XML were structured correctly.

1.1.3. Source processing glitches

As we worked with the data, we discovered that the XML source files did not, in fact, always accurately reflect the CFR's structure. In one case, e.g., Title > Subtitle > Subchapter[1], would be represented in the source XML as Title > Subtitle > Chapter > Subchapter[2].

Occasionally, an element would close prematurely, and the rest of the volume's text would be left outside its proper hierarchy. For example, in Title 1, Part 17 closed before its constituent subparts; the subchapter, chapter, and title elements closed immediately after the part:

<PART>

[1] See, e.g., 45 CFR Subtitle A, which comprises no chapters, only Subchapters A through D.
[2] See http://www.gpo.gov/FDsys/pkg/CFR-2010-title45-vol1/xml/CFR-2010-title45-vol1.xml.

```
            <EAR>Pt. 17</EAR>
            <HD SOURCE="HED">PART 17—FILING FOR PUBLIC INSPECTION
AND PUBLICATION SCHEDULES</HD>
            <CONTENTS/>
            <EAR>Pt. 17</EAR>
          </PART>
        </SUBCHAP>
      </CHAPTER>
    </TITLE>
    <SECHD>Sec.</SECHD>
    <SUBPART>
      <HD SOURCE="HED">Subpart A—Receipt and Processing</HD>
      <SECTNO>17.1</SECTNO>
      <SUBJECT>Receipt and processing.</SUBJECT>
    </SUBPART>
```

(1 CFR 17)[3]

With neither a reliable XML structure nor a completely systematic authored numbering system, it was difficult to find guideposts for automated error correction. Given the unreliability of the markup, the truncation incidents required us to refer to multiple sources in order to be confident of the true structure of the CFR: the authored table of contents in the volume file, the interpretation of the authored table of contents reflected in the metadata (MODS[4]) file, and the eCFR, the OFR's daily-updated CFR version.

A much more pervasive problem was that in certain cases, a section was amended, and the OFR would include the revised text ("REVTXT") within the effective date note. However, the <REVTXT> element wouldn't close properly and would swallow subsequent sections, even, potentially, the rest of the Title 's text.[5]

The upshot of the structural problems in the source data was that the markup gave us bad information a lot of the time. At one point, more than one third of the CFR text was not accessible via automatically generated tables of contents. We could be confident that where a <SECTION> element appeared, inside it would first be the text of the section. That was important because it meant that we could provide the full text of the regulations. We could walk the tree down from the top as well as up from the bottom in order to try to correct for incorrect nesting caused by unclosed tags. (That part took a few weeks to refine in its own right; at each level of <REVTEXT>, there were many possible configurations of elements.) In several places, we had to write custom code to account for specific combinations of elements.

1.1.4. Unreliable metadata

We actually had two streams of data. One of the beautifully designed features of the FDsys CFR was that it provided both XML volume files and separate metadata (MODS) files. The FDsys engineers recognized the flaws in the structural markup and used an enormous amount of ingenuity in pulling the structure from every scrap of data in each volume's XML (Wu, et al. 2008). So the MODS file for each volume typically reflected the CFR structure much more

[3] 1 CFR, http://www.gpo.gov/FDsys/pkg/CFR-2010-title1-vol1/xml/CFR-2010-title1-vol1.xml.
[4] http://www.loc.gov/standards/mods/
[5] See, e.g., 45 CFR § 303.3

accurately than the source XML markup within the volume. Originally, we thought that where the volume XML was inaccurate, or where markup was ambiguous, we could cross-check against the MODS files, and the MODS files did prove to be extremely useful in providing us with leads on places where our XML might be flawed. However, because the MODS files were generated by parsing the XML, although they were useful in correcting some errors, they introduced some of their own. For instance, in Title 41, the MODS represented several parts with the same part number, presumably because of the non-standard chapter numbering.[6]

1.1.5 A policy problem that became a technical problem

We also had an administrative problem: GPO policy was to place suspect volumes into "quarantine" until the relevant bug could be fixed and the text could be regenerated. However, it was very difficult for them to regenerate volumes between quarterly revision cycles, so in practice the text would disappear altogether until at the earliest the next quarter. (At one point, over 5% of the CFR volumes were unavailable in quarantine.) Worse, when the quarantined volume was the last in a Title , the last published volume would be presented as if it were the final volume of the Title . So the only way to be certain of what was supposed to be in a CFR Title was to refer to an earlier printed edition or the eCFR.[7]

2. Reflections

2.1. Why did we bother?

2.1.1. Regulations are important

The breadth of federal regulations and the amount of industry-specific detail contained within them makes them potentially the largest interface between the general public and the law. And in the United States, the federal rulemaking process very deliberately and explicitly includes the public. Under the Administrative Procedure Act of 1946, federal agencies must give public notice in the Federal Register of proposed rules; they must take comments from the public; and, when they publish a final rule, they must respond to the comments they've received. The APA also requires that regulations be published in the Federal Register in order for them to be enforceable. But although federal regulations must be and are made available, they weren't and still aren't easily accessible from the government.

2.1.2. Strategic opportunity

This joint study represented a strategic opportunity for us. In exchange for volunteering our expertise, we were offered access to a data source, the bulk locator code that would otherwise have been inaccessible to us. We had built a considerable base of knowledge, from our work on the U.S. Code, in turning GPO typesetting code into XML. At the time, there was no easily navigable, up-to-date version of the CFR available online. From a sustainability perspective, the CFR would double the number of web pages on our site and potentially garner as much traffic as our most popular resource, the U.S. Code. As an operation partially supported by advertising, we found the prospect of building a new, important, potentially self-funding collection to be highly appealing.

---

[6] See http://www.gpo.gov/FDsys/pkg/CFR-2010-title41-vol1/mods.xml (MODS for 41 CFR vol. 1).

[7] Each volume listed the range of parts within it. The terminal volume would add ":End" to its part range. When the actual terminal volume was in quarantine, the last available volume would show ":End" in its part range.

2.2. Did we over-engineer?

Yes and no. We were using some experimental-to-us technology. The basic toolkit – XML source, XSLT, PHP, MySQL – was very familiar. The CMS, Drupal, we'd used before, although this was one of our first sophisticated add-on modules and the first time we'd used it for a large-scale corpus. The eXist XML database, however, was brand new to us, as was the Solr search engine. Each of the new components was interesting as a tool to start using for the rest of the LII's collections, and each would have cost time on whatever large-scale project we tried it first.

In retrospect, however, it was hubris to take on so many unfamiliar components while simultaneously trying to plumb the depths of a large new text corpus. When we ran into performance problems, we had a large number of unfamiliar interacting components that were difficult to isolate. The truth is that, had we to do it over again, we probably would have made the same choices about platform and architecture, but we would also have made more realistic allowances in the schedule to reflect the increased difficulty.

2.3. Should we have tried to build a less ambitious product?

In the sense that we could have offered the CFR text with minimal value-added features, you could call what we did over-engineering. Once GPO's FDsys came online, however, there was little demand for what would essentially have been a replica of their product, differing only in having been created from the typesetting code data – and in having much less sophisticated search. When FDsys became available, our options seemed to be either to abandon the project or to build a CFR product comparable to our U.S. Code offering.

2.3.1. No one would stand for a comparable product in print

Imagine, for a moment, the OFR decided to publish just the text of the CFR – no indexes, no authorities, no updating information. Just the text. Would we consider that to be an adequate publication of the regulations?

Actually, the OFR couldn't get rid of the indexes. Or the authorities. Or the updating information. In its publication of the CFR, the OFR has its own regulations with which to comply. Not surprising, but somewhat intriguing, are the regulations on materials it is required to publish with the CFR. For instance, "A subject index to the entire Code shall be annually revised and separately published." (1 CFR § 8.4 Indexes). And,

> The Code shall provide, among others, the following-described finding aids:
> (a) Parallel tables of statutory authorities and rules. In the Code of Federal Regulations Index or at such other place as the Director of the Federal Register considers appropriate, numerical lists of all sections of the current edition of the United States Code (except section 301 of title 5) which are cited by issuing agencies as rulemaking authority for currently effective regulations in the Code of Federal Regulations. The lists shall be arranged in the order of the titles and sections of the United States Code with parallel citations to the pertinent titles and parts of the Code of Federal Regulations.
> (b) Parallel tables of Presidential documents and agency rules. In the Code of Federal Regulations Index, or at such other place as the Director of the Federal Register considers appropriate, tables of proclamations, Executive orders, and similar Presidential documents which are cited as rulemaking authority in currently effective regulations in the Code of Federal Regulations.

> (c) List of CFR sections affected. Following the text of each Code of Federal Regulations volume, a numerical list of sections which are affected by documents published in the Federal Register. (Separate volumes, "List of Sections Affected, 1949–1963" and "List of CFR Sections Affected, 1964–1972", list all sections of the Code which have been affected by documents published during the period January 1, 1949, to December 31, 1963, and January 1, 1964, to December 31, 1972, respectively.)[1] Listings shall refer to Federal Register pages and shall be designed to enable the user of the Code to find the precise text that was in effect on a given date in the period covered.
>
> (1 CFR § 8.5 Ancillaries.)

The value-added features necessary to afford functional use of a 100,000-plus page regulatory corpus are well-enough recognized that they have themselves been mandated as regulations. In print, it would be impossible to find the text one required without the ancillaries. When the text becomes electronic and therefore searchable, it is much more difficult to pin down which features are truly indispensable.

Which textual features is it worthwhile to mark up? Is it sufficient, or even useful, to provide data that can be formatted as a table of contents, or does there need to be correct semantic markup of the document's component structural units? Should authorities be marked up, or should that task be left to aftermarket parsers? Does a formatted list of sections affected by new regulation make any sense? Is there any need for an index?

2.3.2. Doesn't search substitute for all of these features?

Search can substitute for some of these features (it doesn't substitute for updating). In fact, the FDsys CFR includes a purpose-built search application with just such features in mind. The FDsys engineers used their knowledge of the corpus to parse authorities and citations (see Appendix B), and they also parsed general information to support metadata-based searches by organization, person, location.

Yet, FDsys still supplies tables of contents for its online version. CFR users expect it, for one. Also, tables of contents provide context that search does not – the view of all items at a particular level of aggregation affords an opportunity to grasp context very quickly. Although one could, of course, design a search application to generate the same output, that would be an unexpected result for a search engine to return.

2.4. Could we have been more prudent?

2.4.1. When we made the joint-study agreement

The availability of source data was so useful to us that we overlooked some technical and administrative details that would have given us a more realistic picture of what it would be like to work with the data.

In October 2008, the FDsys technical team had given a presentation on the system architecture. We attended to statements like this one:

> "-If quality errors exist or metadata is missing, notify metadata correction group to review and edit metadata
> -Users can search for package with low quality metadata and update metadata values appropriately"

Had we paid more attention to steps in the process like:

> "Create Preservation Rendition (XML)
> - Generate xml rendition from pdf content files using Adobe LiveCycle"

we might have been better prepared for the state of the XML source data.

Similarly, the dotted line labeled "Firewall" in the system architecture diagram (Appendix A), might have alerted us that, as an administrative matter, it would be extremely difficult to get back-end access to the system.

2.4.2 When the FDsys released its XML

When the GPO released the CFR-XML in FDsys, the schema it published was described by our XML expert as a "bag of tags". The schema did not reflect the large-scale hierarchical CFR structure; rather, it listed elements that could be part of a CFR XML document. This element listing did not provide enough information to be useful in catching bugs.

The GPO's description of the schema in its guide to the FDsys CFR-XML should have been of greater concern to us:

> "The schema being produced for this effort describes the data as it actually occurs from the OFR. Documents are not being cleaned up because they do not match the schema; instead, the schema was selectively relaxed. Such an approach maintains 100% fidelity to the original data, and eliminates any errors that might occur in schema interpretation or further data manipulation."
>
> (USGPO PMO, 2009)

Although we did not know everything there was to know about the structure of the CFR, from the outset, we could have constructed a bare-bones schema to use for quality assessment purposes. Specifically, although the structure of the CFR is complicated and uneven, certain characteristics, such as the ancestor-descendant hierarchy, are reliable. In other words, even if we weren't sure whether a subchapter had to be the child of a chapter (it turns out that it doesn't), we knew that it must always be a descendant and not an ancestor of a title. Although we developed assessment tools as we went (a useful one was a count of how many elements that should have been nested within the <TITLE> element in fact resided outside it), we could have generated a basic but comprehensive analysis earlier in the process.

2.4.3 When we started running into trouble

It took us longer than it should have to develop basic quality analysis tools for our own work. We invested a bit too much time, for example, in trying to work out the logic of the numbering system, when a better solution was to be much more agnostic about what an enumerator should look like. We took a bit too much of the low-hanging fruit approach, relying on pervasive non-functionality to provide a roadmap for repair when we could have come up with efficient but comprehensive tools for assessing the state of the corpus as a whole.

2.5. What would we like to change?

The LII CFR project has been the occasion for a certain amount of good-natured and not-so-good-natured griping about why the data was so difficult. Having worked extensively with the corpus, we gained a great amount of admiration for the engineering achievements of the government contractors who were working around the same data problems we saw. But quality control is a band-aid, not a substitute for re-engineering.

2.5.1 The government from which we require source data is not monolithic

Open-access legal information publishers rely upon legal information originators for their source material. But it isn't "the government" that originates the data, nor is it "the government" that publishes the data. Where does the data originate? In the regulatory process, regulations originate in federal agencies or with lobbyists. Each agency submits the text of a final rule to the Office of the Federal Register, which prepares and publishes it in the daily Federal Register and then in the collected Code of Federal Regulations. The OFR prepares and publishes, while the GPO does the composition and printing. So the government "publisher" is aggregating raw material from a variety of sources, conforming it, and making it available to the public, including to republishers.

2.5.2. Government publishers face the same problems as republishers

When the source data is "bad", the natural reaction is to complain to the "publisher". And when we ran into problems, it became very clear what the publishers were facing.

From the FDsys description of the XML schema, it is clear that the GPO was getting source data of uncertain quality from OFR. It could not make authoring judgments. So it had to pass the problem data along. They were just as frustrated as we were.

In 2008, the GPO issued a requirements document for a composition replacement system. In the introduction, it had this to say about its current publication system.

> "The United States Government Printing Office's (GPO) current composition system is based on a 30-year-old batch composition engine, developed and maintained by GPO, called Microcomp….
>
> In response to Congress' adoption of XML as a data standard, GPO attempted to retrofit Microcomp to compose Standardized General Markup Language (SGML) / Extensible Markup Language (XML), with limited success. Subsequently, Congress sought to improve workflow processes for content creators as well as search and retrieval of data by end users. As a near-term solution, the retrofitting of Microcomp was suspended and GPO decided instead to translate XML files into locator codes for printed publications."
>
> (USGPO Requirements Document for Composition Replacement System (CSR), 2008)

When we had problems, we'd complain to GPO. But the GPO doesn't have authority over the content, the OFR does. So we'd talk to the OFR. But the OFR can only process what it gets from its sources. And its sources are diverse, individually powerful, and often intransigent.

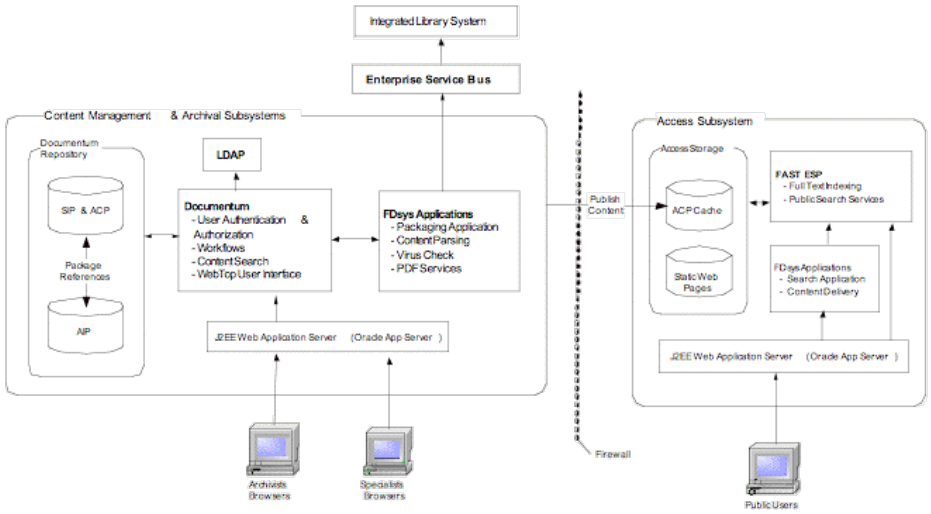
2.5.3 The government should have adequate infrastructure to originate accurately marked-up text.

Yet, without a systemic approach that addresses upstream quality, we're left with a situation that duplicates – and even magnifies – costs downstream, as each republisher struggles with the same data defects.

3. Conclusion

Anyone starting up or working in an LII needs the support of government in order to gain and maintain access to the necessary source material. Relationships with government publishers are indispensable, and there will be a lot of pressure to undertake these sorts of projects. Given the configuration of internal and external incentives, it will perhaps be tempting to overlook the real costs and risks involved: either the source data is in adequate condition, or it will be a great deal more difficult and expensive to publish than data in adequate condition. In our role as informal or formal advisors or consultants to governments, LIIs can provide domain-appropriate technical expertise. In our role as service providers to the public, we are possibly uniquely positioned to advocate for a more sensible strategy for infrastructure investments.

Appendix A: FDsys Application Architecture



(Wu, et al. 2008, p.7)

Appendix B: MODS support for ancillary features

```
<extension>
        <searchTitle>Part 8; CODE OF FEDERAL REGULATIONS</searchTitle>
        <chapterHeading>Chapter I</chapterHeading>
        <chapterTitle>ADMINISTRATIVE COMMITTEE OF THE FEDERAL
REGISTER</chapterTitle>
        <granuleClass>NODE</granuleClass>
        <accessId>CFR-2010-title1-vol1-part8</accessId>
        <sequenceNumber>42</sequenceNumber>
        <heading>Part 8</heading>
        <granuleLabel>part</granuleLabel>
        <granuleNumber>8 </granuleNumber>
        <partRange from="8" to="8"/>
        <leafRange from="8.1" to="8.9" type="section"/>
        <authority>44 U.S.C. 1506, 1510; sec. 6, E.O. 10530, 19 FR 2709, 3 CFR, 1954-1958
Comp., p. 189.</authority>
        <source>37 FR 23605, Nov. 4, 1972, unless otherwise noted.</source>
        <graphicsInPDF>false</graphicsInPDF>
        <USCode title="44">
           <section number="1506"/>
           <section number="1510"/>
        </USCode>
        <fr context="AUTHORITY" volume="19">
           <pages pages="2709"/>
        </fr>
        <fr context="SOURCE" volume="37">
           <pages pages="23605"/>
        </fr>
     </extension>
<relatedItem type="isReferencedBy">
        <titleInfo>
           <title>United States Code</title>
           <partNumber>Title 44 Section 1506</partNumber>
           <partNumber>Title 44 Section 1510</partNumber>
        </titleInfo>
        <identifier type="USC citation">44 U.S.C. 1506</identifier>
        <identifier type="USC citation">44 U.S.C. 1510</identifier>
     </relatedItem>
     <relatedItem type="isReferencedBy">
        <titleInfo>
           <title>Federal Register</title>
           <partNumber>Volume 19 Page 2709</partNumber>
        </titleInfo>
        <identifier type="FR citation">19 FR.          2709</identifier>
     </relatedItem>
     <relatedItem type="isReferencedBy">
        <titleInfo>
           <title>Federal Register</title>
           <partNumber>Volume 37 Page 23605</partNumber>
        </titleInfo>
```

<identifier type="FR citation">37 FR. 23605</identifier>
</relatedItem>

Acknowledgements

This paper is about a project of the LII, which consumed the better part of a year of the working life of the people who built it. LII director Thomas Bruce designed the architecture, wrote the CMS code for the prototype, and developed the Parallel Table of Authorities feature. David Shetland harnessed his deep expertise in XML and the US Code corpus to build the XML back end for the initial prototype, develop new approaches to detecting citations and nested paragraph structure, and refine endless xqueries to address data oddies. Wayne Weibel adopted the CMS code and wrestled cross-volume super-section hierarchy and <REVTXT> to the mat. Weekly. Brian Hughes painstakingly tuned the search engine and chased down innumerable slow queries. Dan Nagy fought for realistic system resources, stalked eXist developers, and otherwise kept the systems beautifully invisible to the programmers. Together, they built a highly useful, usable CFR product. In the process of writing this paper, I benefited greatly from the ongoing dialogue with Tom, Dave, Wayne, Dan, and Brian, as well as feedback from LII affiliated researchers Núria Casellas, Joan-Josep Vallbé, and co-director emeritus Peter Martin. Thanks are also due to our government partners in GPO, OFR, and the FDLP.

References

U.S. Government Printing Office, Program Management Office (2009), *U.S. Government Printing Office Federal Digital System User Guide Document Code of Federal Regulations XML Rendition.* U.S. Government Printing Office, Washington, D.C.

U.S. Government Printing Office (2010), *FDsys Program Review*. U.S. Government Printing Office, Washington, D.C.

U.S. Government Printing Office (2008), *Requirements Document for Composition Replacement System (CSR).*

Wu, D., Nelson, P., and Gee, J. (2008), *GPO Federal Digital System Architecture and Design* (PowerPoint presentation).