

August 18, 2000

Proposal Submitted to the Red Hat Center by Cornell Law School's Legal Information Institute

I. Relationship of Legal Information Institute and This Proposal to Open Source Values

Since 1992 the Legal Information Institute of Cornell Law School (LII) has led the movement to improve public access to law in the U.S. by placing key legal materials on the Internet in non-proprietary format, structured in ways that facilitate unrestricted reuse by others. The Institute's ultimate goal -- free and direct access to information of the most critical sort, namely law -- corresponds closely to the core values of the open source movement. Our means are similar, as well. The LII uses open standards like XML and open source tools like linux, perl, and Apache to convert public-but-semi-proprietary legal documents into formats that are physically and intellectually accessible to a broad public. Through the use of open, platform-independent technologies, we remove barriers that would prevent others from making effective use of the electronic output of courts and legislatures. In that sense we render the law open not only for reading, but for reuse. We have, from the start, presented law data within an open but structured architecture designed to enable others to incorporate our data within their own Web-based materials. For this reason the LII site is the most heavily linked to in the field of law.

II. Public Law Data That Is Not Truly Open

While governmental bodies at all levels in the U.S. have begun to use the Net to fulfill their fundamental obligation to make the law accessible to all, many, including some of the most important, have failed to do so effectively. Openness in this context calls for formats and data structures that allow others to select, reorganize, add commentary, and in other ways reuse such foundational legal materials as the Copyright Act, U.S. immigration and voting rights laws, the Internal Revenue Code, decisions of the Supreme Court, or federal regulations. The digital version of the Copyright Regulations available from the U.S. Copyright Office Web site illustrates the phenomenon. While free (i.e., downloadable without charge), the regulations are offered solely in large files, in a

proprietary format (PDF), optimized for print. (For example, the thirty-five sections that comprise Part I are lumped without navigable structure into a 79 page PDF file.) As a consequence, this law is open only in a minimal sense. It cannot readily be excerpted for inclusion in educational materials, placed in a searchable database with other related documents, or linked to with any precision.

III. The LII's On-Line U.S. Code as a Model of Open Architecture

By contrast, the version of the full U.S. Code (Copyright Act, immigration and voting rights laws, Internal Revenue Code and all the rest) mounted at the LII site is open for reuse down to the section level. As a consequence it is integrated through links with legal and other material of astonishing diversity across the Web. Providing the foundation lawyer of federal law, in an open architecture, it is the most heavily used legal resource at the LII site.

IV. Our Proposal

We respectfully request a grant from the Red Hat Center to enable the LII to realize major improvements in this central legal resource, improvements that will demonstrate even more clearly than our efforts to date what open access to law can mean. We would hope in time to induce the relevant public bodies to distribute law from source in so open a manner, leading by example.

Our plan for the U.S. Code has several elements. The LII's current version is designed for on-line browsing. An important set of users are frustrated at their present inability to download from this collection in units larger than a section (e.g., title, chapter, subchapter) for local printing, incorporation in a compilation of some sort, or off-line storage and use. In addition, the current version holds limited metadata, is too tightly bound to HTML rather than XML, and lacks historic depth. When a new compilation of a particular title of the Code is released by the Office of Law Revision of the House of Representatives our software detects the fact, downloads the file, reformats, parses, and indexes the data, and replaces the prior version of that title at the LII site with the new. Using XML, the wide variety of open-source tools available for use with it, and a database approach developed at the Royal Melbourne Institute of Technology <<http://www.thelaw.tas.gov.au/>> we would like to build a system capable of generating a

"dynamically generated point in time" view of any portion of the U.S. Code. The system would replace the standard "single most recent point in time" view we currently offer. The same system should be capable of allowing users to extract large ranges of sections in any one of several data formats suitable for rendering on devices ranging from workstations to personal digital devices to printers.

V. Details

With a grant of \$100,000 from Red Hat Center, we would be able to pursue these related plans for the U.S. Code by:

1. Developing XML DTDs appropriate for markup of the Code;
2. Developing appropriate conversion software for
 - a. Converting the Code as it is presently delivered by the House site into XML conforming to those DTDs
 - b. Converting historic versions of the Code to XML, reaching back to approximately 1986;
3. Developing appropriate delivery mechanisms (centered around use of the Apache-XML/Cocoon technologies) compatible with a wide variety of browsing and rendering tools, including access by the disabled;
4. Developing "point-in-time" retrospective systems to permit historical analysis of the Code.

Not accidentally, development of this underlying open-code base for the U.S. Code would also permit parallel development of improved resource-location and metadata tools that would permit linking of the Code to relevant court cases, regulations, and explanatory material.