

Reprinted with permission from the May 6, 2014 edition of Law Technology News website © 2014 ALM Media Properties, LLC. All rights reserved. Further duplication without permission is prohibited. For information, contact 877-257-3382, reprints@alm.com or visit www.almprints.com.

Can Open Source Transform Democracy?

Stanford panelists say free and unfettered access to government documents would jumpstart justice for all Americans.

Mark Michels, Law Technology News

May 6, 2014

"What are the benefits of a more open data ecosystem for the law?" The Codex FutureLaw Conference 2014 at Stanford Law School grappled with this question during a panel discussion entitled "Forging an Open Legal Document Ecosystem." Moderated by Law Technology News' Editor-in-Chief Monica Bay, the panel included Brian Carver, assistant professor at the U.C. Berkeley School of Information ([Free Law Project](#)); [Thomas Bruce](#), director and co-founder of the Legal Information Institute at the Cornell University Law School; attorney [Julio Avalos](#), general counsel at GitHub, and [Paul Sawaya](#), creator of Restatement, an open source software toolkit for legal technologists.

The four panelists contended that 1) open legal data can produce innumerable societal and economic benefits, and 2) there are still significant roadblocks impeding that vision.

One major theme that emerged from the panel: While there are a growing number of government entities publishing statutes, codes, regulations and judicial decisions online, the format in which these documents are published limit their utility. For example, documents in PDF are easy for humans to read, but are almost impossible for computers to digest. PDF documents (and many other file types) must go through a conversion procedure to make the documents "machine readable" so they can be processed by computers. Only when documents are in a format that computers can read will legal technologists be able to use them systematically and efficiently, panelists observed.

Carver described the problem this way—as an intellectual property professor, he wanted to know straightaway when courts published decisions related to issues of interest to him. One approach was that he could scan the Supreme Court and circuit courts' opinions daily, but that would be inefficient and time consuming. Furthermore, there was no effective way to have a computer scan the court websites because the computers could not read the human-readable files.

To solve this problem he and his colleague, Michael Lissner, developed a process using a tool called a "scraper" to collect the published cases from the U.S. Supreme Court and circuit courts on a daily basis. After collecting the cases, the documents were converted

("parsed") into a format that a computer could read—in this case into an XML format. Once in this machine readable format a computer query ran against the new data to identify and report cases of interest.

The Free Law Project emerged from Carver and Lissner's work and now provides free access to its CourtListener repository which, according to Carver, contains more than 2.5 million decisions from about 350 court websites. The project uses its "Juriscraper" tool to collect decisions from the U.S. Supreme Court, federal courts of appeal, selected federal district courts and numerous state courts. The Free Law Project also created an application programming interface that makes the data on CourtListener available at no cost, and allows others to access the data and download decisions in bulk.

Carver provided one example of how this open legal data ecosystem operates. [State Decoded](#) is an organization that created an open platform to display (among other things) state legal codes, to make them more accessible and understandable. State Decoded linked some of its state code pages to CourtListner cases that cite the specific state codes. This user-friendly data set of statutes and case law is available to the public at no cost.

In true Silicon Valley fashion, panelist and software developer Paul Sawaya said he was inspired by last year's Stanford CodeX program and started a project to "computerize the law" by developing tools to "parse legal documents into a machine-readable format that is more semantic."

Many legal technologists face the same challenges Carver encountered when he had to transform court opinions into a machine readable format, said Sawaya. He developed an open source toolkit called "[Restatement](#)" which provides structured parsing and web publishing for legal text. Developers like Sawaya play an important role by providing legal technologists the tools they need to operate in an open document ecosystem.

Bruce has been a leader in the open legal document movement since he co-founded the Legal Information Institute at Cornell Law School in 1992. LII's website is the "most trafficked open access legal website," he noted. The United States Code and Supreme Court decisions have long been posted on the LII website, and his team recently undertook a challenging effort to get the Federal Code of Federal Regulations accessible on it as well.

His current focus is on fostering a "semantic [machine-readable] web representations of regulations," said Bruce, who noted that regulations are the "largest contact surface people have with the law." There is considerable business value in making regulations accessible to the public, he argue, which often facilitating banking growth.

An open legal document ecosystem is not all about case law, statutes and regulations. GitHub's Avalos said he envisions tremendous value emerging from placing transactional legal documents in an open legal document ecosystem. GitHub describes itself as the world's largest host for open source computer code. Not surprisingly,

Avalos pointed to the open source community as a model for legal document collaboration. As an example, he cited the work of [Docracy](#), a site that contains an open collection of contracts and agreements.

Avalos shared his vision of how posting corporate legal forms on the Web can provide opportunities for corporate counsel to get timely and value-added legal advice from attorneys who may not otherwise have access to corporate counsel. For example, imagine that a California corporate counsel, whose company has offices throughout the U.S., posts her standard employment agreement on the Web in a machine readable format (e.g., XML). Also imagine that a solo attorney in Maine is a specialist in national employment law matters and was aware of the California counsel's agreement, and observed a defect in a non-compete provision. With a very simple search query, the Maine attorney could find the California counsel's document on the Web (and others like it) and propose amending the specific agreement terms.

The scenario illustrates the value to the corporate counsel, who now has a better agreement, and to the solo, who now has a new corporate counsel relationship. According to Avalos, an open legal document system can provide low cost, value-added advice to the client, and provide lower barriers of entry for attorneys to get access to clients.

FOREFRONT

The panelists are at the forefront of a growing open legal document movement, but for the public to truly benefit and for emerging legal technologies to flourish, the legal data must be more accessible and usable, they argued. The panel's consensus was that government publishers should make public domain documents available in a way that fosters broader access to enable legal technology innovation.

First, and foremost, legal documents should be published in machine readable format, they agreed. Indeed, the Federal Government has taken steps in that direction—the Office of Management and Budget's "[Open Government Directive](#)" states that in order to "increase accountability, promote informed participation by the public, and create economic opportunity, each agency shall take prompt steps to expand access to information by making it available online in open formats. . . . An open format is one that is platform independent, machine readable, and made available to the public without restrictions that would impede the re-use of that information."

In addition to publishing data in a machine-readable format, allowing data to be downloaded in bulk could also facilitate users' ability to use and commercialize legal technology, argued Carver. Publishing judicial decisions, statutes, regulations, legislation and the like in a machine readable format and making those documents downloadable in bulk can provide legal technologists and entrepreneurs with opportunities to provide legal services in the future in creative and innovative ways, the panelists concluded.

Mark Michels is a director in Deloitte Transactions & Business Analytics, based in Silicon Valley.