Setting the Standards for Open Government Data

June 17, 2014 (https://www.opengovfoundation.org/setting-the-standards-for-open-government-data/)Blog (https://www.opengovfoundation.org/category/blog/)admin (https://www.opengovfoundation.org/author/admin/)

Transforming US Civic Data Into a "What You See Is What You Mean" Format

By Bill Hunt (twitter.com/krues8dr), lead OpenGov developer for America Decoded (AmericaDecoded.org)

City and state governments across America are adopting Open Data policies at a fantastic clip. Technology hotbed San Francisco has one (http://govfresh.com/2013/03/san-francisco-makes-open-data-city-policy/), but places far from Silicon Valley – like Pittsburgh, PA (http://triblive.com/news/adminpage/5745435-74/government-legislation-data#axzz34u82Pq6w) – are also joining the municipal open data movement. The increasing availability of free tools and enthusiastic volunteer software developers has opened the door for a vast amount of new government data to be made publicly available in digital form. But merely putting this data out there on the Internet is only the first step.

(http://opengovfoundation.org/assets/No-

Standards.jpg)Much of this city data is released under assumption that a government agency must publish something – anything – and fast. In this rush to demonstrate progress, little thought is given to the *how*. But the citizens who care about this data – and are actually building websites and applications with it – need to access it in machine-readable, accessible, and standards-compliant formats, as the Sunlight Foundation explains here (http://sunlightfoundation.com/policy/documents/ten-open-data-principles/). This explains why most city open data sets aren't seen or used. There is a vast difference between merely opening data, and publishing Open Data.

By publishing data in good formats that adhere to modern, widely-accepted standards, users of the data may reuse existing software to manipulate and display the data in a variety of ways, without having to start from scratch. Moreover, it allows easy comparison between data from different sources. If every city and every organization chooses to adopt their own standard for data output, this task becomes absolutely insurmountable – the data will grow faster than anyone on the outside can possibly keep up.

Most Government "Open Data" Is A Useless Mess

Take, for example, the mess that is data.gov (http://data.gov/). Lots of data is available – but most of these datasets are windows-only self-extracting zip archives of Excel files without headings that are nearly useless. This is not what the community at large means by "Open Data" – there are lots of closed formats along the way.

Similarly, data which is released with its own schema, rather than adopting a common standard, is just as problematic. If you're forcing your users to learn an entirely new data schema – essentially, a brand new language – and to write entirely new parsing software – a brand new translator – just to interact with your data, you've added a considerable barrier to entry that undercuts openness and accessibility.

A good standard lets anyone who wants to interact with the data do so easily, without having to learn anything new or build anything new. Standard programming libraries can be built, so that it's as simple as opening a webpage for everyone. This means that in most programming languages, using a standards-based data source can be as simple as it is to interact with the web, import httplib and you're done.

Evaluating Existing Standards

 (http://opengovfoundation.org/assets/Existing-

Standards.jpg)Every day at The OpenGov Foundation, I work with legal and legislative data. Laws, legal codes, legislation, regulations, and judicial opinions are a few examples. What standard do we use? Well, let's look at the most common standards available for publishing legal data on the Internet:

- **Akoma Ntoso (http://www.akomantoso.org/)** – a well-known XML-based format that is very verbose. The level of complexity presents a high barrier to entry for most users, and has prevented its wide adoption.

- **United States Legislative Markup (USLM) (http://uscode.house.gov/download/resources/USLM-User-Guide.pdf)** – another XML-based format used by the US House of Representatives. It has the advantage of being not very verbose, extensible, and easy to use.

- **State Decoded XML** – the format used by The State Decoded (http://www.statedecoded.com/) project. Currently, this only support law code data, and is not widely adopted outside of this project.

- **JSON** – JSON is not actually a standard, but a general format well suited to relational and tabular data, and chunks of plain text. A variant is JSON-LD which has all of the same properties, but is better for relational data. It is commonly used for transferring data on the web, but it is not practical for annotated or marked-up data.

None of these are ideal. But if I had to pick a single option to move forward, the **USLM** standard is the most attractive for several reasons:

- It is older, established, and has good documentation
- It is easily implemented and used
- It is extensible, but not especially verbose
- It is designed to handle inline markup and annotation, such as tables, mathematical formulas, and images

It also acts as a very good "greatest common factor" as a primary format – it can be translated easily into common formats such as HTML, Microsoft Word, plain text, and even JSON – but does not add any superfluous complexity to address most common needs (e.g., tables or annotations) that other formats require.

Setting the Standard for Open Law & Legislative Data

(http://opengovfoundation.org/assets/Building-A-

Standard.jpg)Moving forward, the next step beyond simply exporting USLM data from existing data sources would be to to have end-to-end solutions that speak USLM natively. Instead of editing Word or WordPerfect documents to craft legislation, lawmakers could write bills in new tools that look and feel like Word, but are actually crafting well-formatted USLM XML behind the scenes, instead of a closed-source, locked-in format. This is what we call "What You See Is What You Mean" – or WYSIWYM.

Here at The OpenGov Foundation, we believe in a rich, standards-based data economy, and we are our actively doing our part to contribute. Our open-source online policymaking platform – Madison (http://bit.ly/MadisonV2)– already consumes USLM, and we are actively working on an WYSIWYM editor to make it easier to create and modify policy documents in Madison. We are also investigating USLM support for The State Decoded (http://americadecoded.org)– both as an input and output format. Hopefully, other software projects will actively follow suit – creating an interoperable ecosystem of legal data for everyone in the United States.

What do you think? Send me an email (mailto:%20bill@opengovfoundation.org)or find me on Twitter (http://twitter.com/krues8dr).

Post navigation

← East and West Coast Legislators Set National Free Law Founders Challenge: Free and Open Source Platform for Governing Civic Hackers Accept One Year Challenge to Build Democracy 2.0 Platform (https://www.opengovfoundation.org/east-and-west-coast-legislators-set-national-free-law-founders-challenge-free-and-open-source-platform-for-governing-civic-hackers-accept-one-year-challenge-to-build-democracy-2-0-platform/) Join DC Council Committee Hearing Online with MadisonDC → (https://www.opengovfoundation.org/new-chance-to-weigh-in-on-dc-council-committee-hearing-with-madisondc/)

Get Involved

There's lots of ways to be involved with OpenGov, from helping work on a particular set of laws to attending hackathons and working on code and design. Let us know what you want to do. (/contact)