

Preserving State Government Digital Information Minnesota Historical Society

Sunlight Foundation's NDIIPP projects

Summary

The following report was written by the Sunlight Foundation to describe their involvement with the NDIIPP project. Sunlight developed a OpenGovernment page using Minnesota data as well as a mobile application using OpenState data; both of these show how accessible government information can be used by outside parties.

Any comments, corrections, or recommendations may be sent to the project team, care of:

Carol Kussmann
Collections Assistant, State Archives
Minnesota Historical Society
carol.kussmann@mnhs.org / 651.259.3262

Summary

During the past year Sunlight Foundation has created two products serving the purpose of increasing public access to state legislative data: the Open Government website (OpenGovernment.org) and the Open States iOS app (to be publicly released Q1 2012). Both of these sites are powered by the Open States API, a data service that provides a common interface to state legislative data including bills, legislators, roll call votes, and committees.

OpenGovernment.org

OpenGovernment.org is the name given to the public-facing website built around Sunlight's Open States data. The site presents a common, easy to use interface to state legislative information with additional context pulled from news aggregation services, social media, and partner organizations such as Project Vote Smart for legislator contact information and National Institute on Money in State Politics for campaign finance.

When a user visits a state page (such as <http://mn.opengovernment.org>) they are presented with an overview of data for that state, including popular bills, legislators, and issues (determined by online mentions). A selection of state specific resources such as official state sites and relevant political blogs is also provided for each state.

It is possible to browse bills by their date of introduction, most recent action, news mentions, times viewed on the site, or subject area. If a user is looking for a specific bill it is also possible to search by keywords contained in the bill title. When a bill is selected it brings a user to a bill page (e.g. <http://mn.opengovernment.org/sessions/2011-2012/bills/sf-86>) that serves as a permanent reference to all actions taken on a bill. Included are any reported actions out of either chamber, a full listing of sponsors with links to their site profiles, an indication of the current

status of how and how far along the bill is in the legislative process. Votes are also listed when present, providing a visual indicator of how the vote total broke down. By clicking on a vote it is possible to get a detailed view of how individual legislators voted. Using Google News, news stories that appear to be about the bill in question are linked and each bill is also given a Twitter hashtag that users can use to have their comments on Twitter pulled in for the bill page. Like all data-driven pages on OpenGovernment.org, a link to the primary source for the bill on the official state website is given.

Legislators are also searchable or browsable by their district, chamber, or party. Like bills, legislators have distinct permanent pages (e.g. <http://mn.opengovernment.org/people/2164-richard-cohen>). Candidate pages include basic biographical details, contact information, a link to outside resources such as Wikipedia, and relevant news stories. All roles that a legislator is known to have held, including committee memberships and previous legislative sessions are also linked from the legislator page. Recent votes and sponsored bills are also shown, with links to the relevant sections of the site. Additionally, when possible the legislator pages also feature campaign contributions and interest group ratings, using data from National Institute on Money in State Politics and Project Vote Smart respectively.

Additional sections of the site provide alternate ways at looking at campaign finance or bill information, grouped by subject area instead of as it associates with a single legislator. Functionality also exists on the site for users to use their location to look up their representatives and also connect with them via email where an email address is available.

Open States iOS Application

The Open States iOS application is targeted at putting the same legislative information collected by the Open States project and utilized in the OpenGovernment.org site in the hands of mobile users. This application targets the iOS platform and runs on iPhone, iPod Touch, and iPad devices. Future plans also include building an application to run on the Android platform, most likely based on Sunlight's popular Congress for Android application that serves a similar purpose at the federal level.

The iOS application provides a legislative directory of all members of a state's legislature. For each legislator it includes a photo, basic biographical details, an interactive district map, committee memberships, and contact information (including, when available, information such as phone number and email by which the legislator can be directly contacted).

The application also makes it possible to track specific legislation; each bill page includes all actions that have been recorded by the legislature on the bill. Sponsors, roll call votes, and the current stage of the bill in the legislative process are included in this bill view.

Given the large degree of overlap between these two projects, the goals and challenges described in the remainder of this document apply to both unless otherwise noted.

Purpose

The fundamental purpose behind each of these applications is the same: to increase the ability of citizens to be aware what is going on in their state capitol and increase the potential for them to become involved in the process. Both OpenGovernment.org and the iOS application attempt to provide the information in an easier to understand and more aesthetically pleasing way than most state websites are able to do, while also linking to the state website when verification or additional information is required. Additionally, both sites provide a great deal of contact information for legislators so that individuals can contact their state representative.

Both products also use an innovative geolocation method to help users find their state legislative districts. This method combines address geocoding as seen in services like Google maps with publicly available legislative district boundary shapefiles to provide more accurate results than a traditional zipcode approach allows. Merely providing a common way for citizens to find this information represents a notable improvement over the status quo for many states.

Issues Encountered

Both OpenGovernment.org and the Open States iOS application are primarily powered by the data collected by Sunlight's Open States API (<http://openstates.org>). Open States is a project that aims to collect and standardize state legislative information including metadata pertaining to bills, legislators, votes, committees, and schedules.

Most states, including Minnesota, do not currently provide any form of bulk access or API to their legislative data. As a result it is necessary to crawl the official state legislature's website and extract all relevant information. This task provides a large number of disadvantages including timeliness, potential inaccuracies, increased load on state websites, and potentially incomplete data.

The lack of bulk data presents one of the largest obstacles for these efforts. If a legislature has perhaps five thousand bills each session and each bill has its information spread across five pages (perhaps one each for general information, versions, sponsors, actions, and votes) this requires approximately 25,000 pages to be visited to index an entire session. Because few if any states provide a reliable way to tell which of these pages have changed, to avoid incomplete information it is necessary to check all of these pages on each update. This means an update running at 1 page/sec will take nearly 7 hours to complete. Here practice agrees with theory with our updates taking between one hour and twelve depending on the complexity of the state website.

The next most serious challenge is in the fact that collecting information from a webpage is a potentially delicate process: a slight change to page layout may result in crawlers missing important information until it is noticed and the crawler can be corrected. In many cases state websites are old and somewhat fragile, so it is not uncommon to see HTML that does not comply with standards and is thus all the more difficult to reliably parse. At times key legislative information such as votes is only available via PDF which presents an even more difficult challenge. PDF files are extremely difficult to programmatically parse, extracting information contained within them can be challenging or sometimes impossible.

These two challenges: the availability of bulk data and the availability of machine readable data are perhaps the most vital but by no means the only challenges faced. The best outline of the challenges we face and their potential solutions is presented in the whitepaper previously authored in collaboration with MNHS “Best Practice Principles for Opening Up Government Information”

(http://www.mnhs.org/preserve/records/legislativerecords/docs_pdfs/BestPracticePrinciplesOpenGovtMarch2011_000.pdf) which addresses such vital issues as that of completeness, timeliness, and permanence.

Recommendations

In order to facilitate the development of sites like OpenGovernment.org and applications like the Open States iOS application, governments should take it upon themselves to ensure timely and permanent machine readable legislative records. All too often when a website is redesigned resources break or disappear entirely. URLs should be designed to be permanent so that third parties can always link back to the original source.

Adopting a machine-readable format is a good first step for legislatures to take. XML or JSON provide an ideal starting point, though other formats such as CSV, Excel, or DBF have been seen in states and provide acceptable levels of accessibility. When compiling these formats it should be the position of the implementers to err on the side of disclosure: all too often fields present on the website are omitted in the electronic record, forcing individuals to choose either to omit certain data or favor the slower, less reliable crawling method of access.

While PDF is a fine format for bill text, metadata should never be exclusively presented in that format. A list of roll call votes or table of actions is a perfect fit for one of the aforementioned data formats. Extraction from a PDF is time consuming and subject to errors, rendering the data inaccessible to many kinds of future use.

When available, a bulk download capability can greatly reduce the amount of load placed by users of the data on state servers. Some states have shown it is possible to provide one or two archive files updated nightly that result in a single request by Open States and similar projects. This single archive replaces tens of thousands of requests that would otherwise be made.

Finally, it is strongly suggested that state legislative information systems give and make public unique identifiers for legislators. Names are not guaranteed to be unique, especially across multiple sessions and can also change more frequently than many people assume. All too often states publish a vote where it is impossible to determine which “J. Smith” voted yea on a controversial bill.

Evaluation

The intended audience for these products has been primarily the citizens of the states in which they are available, with an emphasis on engaging citizens that have not typically been aware of the activities in their statehouse.

The OpenGovernment.org site has received just over 100,000 unique views in the last 6 months and 250,000 page views total. The average user spends over two minutes on the site and visits 2.5 pages. Compared to similar sites this shows a great deal of “stickiness” showing that users are coming and actually spending time looking around.

The Open States iOS app has not yet gone through Apple’s app store approval process so usage numbers are not yet available. The app does however derive from an existing Texas-specific application which has proven of tremendous use to those wishing to visit their statehouse and discuss issues with legislators directly and as such has grown features related to the needs of those users such as statehouse maps where available.