

# Preserving State Government Digital Information Minnesota Historical Society

## Mashups Using Government Data: A White Paper

---

### *Abstract*

*Mashups are a combination of two or more data sets presented via a single tool, usually a web application, and viewed through a web browser. Third parties (i.e., someone other than the data creators) often take existing data and combine it in ways that add meaning to the existing data or that may create a completely new meaning not available from the individual data sets themselves. This white paper offers background on mashups, examines why government entities may want to make public data available for mashups, and how government data may be enhanced by mashups.*

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

Nancy Hoffman  
Project Analyst  
Minnesota Historical Society  
nancy.hoffman@mnhs.org / (651) 259-3367

---

### **Introduction**

The term “mashup” came into the popular lexicon to refer to new music created by combining two or more existing ones. A data mashup has come to describe the practice of combining two or more sets of data electronically to enhance current meaning or create new meaning, where the original data and the resulting product are all usually available on the Internet. For example, Google makes their map data available for reuse and as a result, a plethora of mashups have sprung up that combine many types of geographically referenced data with Google maps. The quantity, variety, and intrinsic value of government data make it suitable for use in mashups. Citizens and government entities have begun to ask for direct access to data, especially to the critical information that can improve decision making inside and outside of government.

The Programmable web serves as a clearinghouse for a wide variety of individuals and organizations that share data and reuse it in mashups. A portion of their website is dedicated to government data mashups<sup>1</sup>. In January 2009, only 36 of the 3630 of the registered mashups included government data. As Vivek Kundra, the Chief Technology Officer (CTO) of the

---

<sup>1</sup> <http://www.programmableweb.com/government> [accessed 2/20/2009]

District of Columbia Government points out, “the government has largely been absent in the trend toward worldwide exchange of data.”<sup>2</sup> This absence is, in part, due to the fact that much government data are produced in formats that do not allow for its use in mashups.

Public information created by government entities has customarily been provided in a variety of formats with print being the most widely used method of dissemination for many years. With the advent of Internet access, citizens increasingly have expected to find government information on the web. Posting the same information online that previously appeared in print may have initially appeared to be the most straightforward and efficient way to update delivery. Digital information delivery differs fundamentally from print media distribution. Because of the flexible way in which digital information can be analyzed and presented, information creators such as government entities do not, indeed should not need to, imagine the most useful ways for users to look at that information.<sup>3</sup> Releasing access to data itself rather than a single view of the data allows multiple presentations to be made, depending upon the needs or interests of the user. Mashups have been used to improve accessibility to Internet users with disabilities, for example.<sup>4</sup> Making raw data accessible does not preclude a government entity from continuing to present the data as it has done in the past, but it does mean that additional value need not represent additional expense to the public. The US Office of Management and Budget (OMB) recognized this fact in 2004 when it first asked federal agencies to make data available in open formats as much as possible so that citizens could manipulate and analyze the data as needed.<sup>5</sup> Analysis that combines various types of data sets is not new, doing it on the web facilitated by open access to data is.

## Data Sources, Standards, and Tools

Data reuse depends upon access to data in a machine readable format. This can be accomplished in several different ways. The principle techniques for gaining access to online data for reuse are through the release of Application Program Interfaces (APIs) by the data creators, the release of RSS feeds by the data creators, or by a process called “screen scraping” that takes data presented in a human readable format and converts it into computer readable format.

### *Data Sources*

#### APIs

Large quantities of data are usually managed in software applications such as databases. These applications, even when they are available on the web, cannot necessarily interact with other applications unless Application Program Interfaces are provided. Anyone who has searched a database for information has essentially performed the same operation as one type of API. This

---

<sup>2</sup> <http://www.appsfordemocracy.org/building-the-digital-public-square/> [accessed 2/20/2009]

<sup>3</sup> <http://www.mercatus.org/PublicationDetails.aspx?id=21066> [accessed 2/20/2009]

<sup>4</sup> <http://services.alphaworks.ibm.com/socialaccessibility/> [accessed 2/20/2009]

<sup>5</sup> [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1138083](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1138083) [accessed 2/20/2009]

type of API allows programmers to get results from a database and then combine those results with another application, to create a mashup.

## RSS

RSS is well known as method for delivering online news feeds. RSS provides updates or feeds from data sources that have frequent additions or changes. By using the RSS “publish once and read by many” model, data creators can provide current information in a standardized file format, usually XML that can easily be combined with other data. RSS essentially performs a query for new data in the way that APIs allow for a variety of calls to be made on a data source. This is ideal for mashups that seek to interpret current rather than historical trends.

## Screen Scraping

When a data creator only provides data in a presentation-oriented format, or does not provide an interface such as an API for their data, a third party that wishes to use the data for another purpose must resort to a program that takes presentation-oriented data and converts it to a format suitable as an input for another program. This method, called screen scraping can introduce errors because unstructured information intended for display often contains inconsistencies that are carried into the attempt to convert it into a regular, systematic form.

## *Mashups and Open Source Software*

Open source software plays a key role in data sharing initiatives because applications built on open source software can not only be used, but reused and improved by others. An open government data standards group has articulated a set of principles to ensure equal access that include a recommendation for the use of open source software.<sup>6</sup> The degree to which proprietary data formats inhibit interoperability varies. Tim O’Reilly points out that,

“Google Maps isn’t open source by any means, but it was open enough (considerably more so than any preceding web mapping service) and so it became a key component of a whole generation of new applications that no longer needed to do their own mapping. A quick look at programmableweb.com shows Google maps with about 90% share of mapping mashups. Google Maps is proprietary, but it is reusable. A key test of whether an API is open is whether it is used to enable services that are not hosted by the API provider, and are distributed across the web.”<sup>7</sup>

So proprietary formats are not the issue per se, but data access that is severely limited by format or restrictions on use can easily undo the benefits of providing access to the data in the first place.

---

<sup>6</sup> <http://wiki.opengovdata.org/index.php/OpenDataPrinciples> [accessed 2/20/2009]

<sup>7</sup> <http://radar.oreilly.com/2008/07/open-source-and-cloud-computing.html> [accessed 2/20/2009]

## *Tools for Creating Mashups*

As more data sources have become available, more user-friendly tools have appeared to create mashups with that data. Widgets are simple pieces of portable code. Widgets themselves cannot technically be considered mashups because they present data from a single source. However, because they are intended to be added to an existing web page presenting a second type of data, they serve as a constituent of a simple mashup. Similarly, web portals and aggregators that bring two or more information sources together side by side, without any overlap in the data could also be considered a very basic type of mashup, because they “combine data from multiple sources in a single tool.”

Truly integrated data sets or applications that utilize data visualization tools demonstrate the full capacity of mashups to create a new understanding from existing knowledge. Map tools such as ZeeMaps,<sup>8</sup> Wayfaring,<sup>9</sup> and MapBuilder<sup>10</sup> have simple graphical interfaces that allow individuals to put together a mashup without any programming skills. Yahoo! Pipes offers a simple visual editor to combine data feeds to create mashups. Yahoo! Pipes also serves as a community space for mashups application code so that existing mashups can be remixed with new data sources.<sup>11</sup>

Despite the recent increase in offerings, user-friendly tools do not exist for all applications. Knowledge of computer programming is needed to create most types of mashups. API providers usually aim to make it possible for mashups to be written in a variety of programming languages, so that a developer does not need to learn a new code just to work with a new data set.

Most mashups also require the application creator to host the application on a server. Some simple mashups can be designed so that the data come together in the client’s browser window, but client-side mashups can pose security problems for users.<sup>12</sup> The Programmableweb has excellent information explaining how to use the APIs and applications available on the site.<sup>13</sup>

## *Additional Tools for Mashups Creation*

Community for open source software that lists many applications to build mashups  
<http://sourceforge.net/>

IBM Many Eyes  
<http://manyeyes.alphaworks.ibm.com/manyeyes/>

sMash <http://www-01.ibm.com/software/webservers/smash/>

---

<sup>8</sup> <http://www.zeemaps.com/> [accessed 2/20/2009]

<sup>9</sup> <http://www.wayfaring.com/> [accessed 2/20/2009]

<sup>10</sup> <http://www.mapbuilder.net/> [accessed 2/20/2009]

<sup>11</sup> <http://pipes.yahoo.com/pipes/docs> [accessed 2/20/2009]

<sup>12</sup> <http://en.wikipedia.org/wiki/JavaScript> [accessed 2/20/2009]

<sup>13</sup> <http://www.programmableweb.com/howto> [accessed 2/20/2009]

IBM's mashups Center with IBM Lotus Works that provides version control <http://www-01.ibm.com/software/info/mashup-center/>

Mozilla Ubiquity includes functionality to create client-side mashups  
<http://labs.mozilla.com/projects/ubiquity/>

Google Mashups editor, still in testing, is comparable to Yahoo! Pipes  
<http://editor.googlemashups.com/>

Information visualization wiki <http://www.wikiviz.org/wiki/Tools>  
Screen Scraping tool <http://www.crummy.com/software/BeautifulSoup/>

## Who is Creating Government Data Mashups and Why

Rhode Island,<sup>14</sup> Massachusetts,<sup>15</sup> and the District of Columbia<sup>16</sup> are amongst the government entities that have decided to offer access via APIs to their public data that is stored in web-accessible databases or “data warehouses.” Washington DC partner, iStrategyLabs<sup>17</sup> explains that, “The Citywide Data Warehouse mission is to democratize government data by providing a centralized access point for enterprise-wide data with a focus on providing real-time operational data from multiple agencies and sources that enables decision support and government transparency.”<sup>18</sup> Transparency, civic engagement, and efficiency are the most commonly reasons stated for enabling data sharing for mashups.

In Massachusetts, the new OpenMass.org, modeled on the OpenCongress.org site, offers information about elected officials, bills, bill status, and earmarks, combined with news stories, and blog postings about the legislators. Several ways to contact government officials are offered and links to FollowtheMoney.org allows citizens to track campaign contributions for any individual in state government. The site offers this description of its intent:

“OpenMass brings together official government data with news and blog coverage to give you the real story behind what's happening in the Massachusetts Legislature...Our goal at OpenMass.org is to help empower civic engagement by providing easy to access information...At OpenMass.org you can track which Bills, Representatives and Senators are most viewed, monitor public hearings, keep updated on public issues that matter to you on the Hearings Page. OpenMass tracks bills signed into law by the Governor.”

---

<sup>14</sup> <http://www.ri.gov/rex/feeds.php> [accessed 2/20/2009]

<sup>15</sup> <http://www.openmass.org/> [accessed 2/20/2009]

<sup>16</sup> <http://dcstat.octo.dc.gov/dcstat/site/default.asp> [accessed 2/20/2009]

<sup>17</sup> <http://www.istrategylabs.com/about> [accessed 2/20/2009]

<sup>18</sup> <http://www.istrategylabs.com/government-20-the-rise-of-citizen-innovation-through-open-data/> [accessed 2/20/2009]

OpenMass.org does not seek to replace the official legislative website and in fact credits the state employees who gather the original data for the mass.gov/legis site.

Third parties give many of the same reasons as government for direct data access and reuse. The New York Times, for example, has created their own mashup, called “Represent,” by combining news articles and mapping tools with government data. New York City residents can enter their address to discover the names of their elected officials and related news about those individuals. This mashup also allows users to contact their government representatives directly.

“Represent figures out which political districts you live in and who represents you at different levels of government. It draws maps that show how where you live fits into the political geography of the city. And using information collected from around the web, it presents a customized activity stream that tracks what the people who represent you are doing... Represent crawls a collection of New York Times stories and City Room blog posts, looking for references to public officials. It also draws from official data sources — currently, Congressional roll-call votes, which we collect by parsing feeds and scraping government websites. It evaluates each article, blog post and vote to find the stories most relevant to you.”<sup>19</sup>

The privately run WashingtonWatch.com states that the site “is maintained by Jim Harper, Director of Information Policy Studies at the Cato Institute, in his spare time, as a public service.”<sup>20</sup> This site combines government calculations about the costs or savings from proposed changes to government spending, taxation, and regulation and combines this data with Harper’s own calculation of the “net present value.” Each piece of proposed legislation has a wiki page that contains an overview, summary of pros and cons, the bill status, and user input which include “Vote on this Bill,” “Write Your Representative in Congress,” “Alert Your Friends and Colleagues,” functions, a comment section, and links to del.icio.us, Dig, Face book, Google, Reedit, and Yahoo! The site also incorporates a blog. This entry from the blog explains where the data used on the website come from:

“The Data We Bring You . . . comes from government websites. Much of it we scrape using clumsy web-page parsing software, and some of it we have to enter by hand. We could do a lot more - meaning you would know a lot more - if the government would release raw data in usable formats.”<sup>21</sup>

The non-profit Sunlight Foundation is a third-party user of government data that “supports, develops and deploys new Internet technologies to make information about Congress and the federal government more accessible to the American people... to create greater political transparency and to foster more openness and accountability in government.”<sup>22</sup> To promote these goals, Sunlight combines federal data from at least ten different sources and releases API access to them to make it possible for anyone to create a mashup from this data set or combine it with

---

<sup>19</sup> <http://blog.programmableweb.com/2008/12/27/the-new-york-times-rolls-its-own-map-mashup-represent/> [accessed 2/20/2009]

<sup>20</sup> <http://www.washingtonwatch.com/about/> [accessed 2/20/2009]

<sup>21</sup> <http://www.cato.org/event.php?eventid=5475> [accessed 2/20/2009]

<sup>22</sup> <http://www.sunlightfoundation.com/mission/> [accessed 2/20/2009]

another API.<sup>23</sup> Sunlight held a congressional data mashups contest in 2007. The winning entry produced a web program that generates interactive network maps of state level political contribution data.<sup>24</sup> They recently opened a second data mashups contest. Sunlight has also recently instigated the Sunlight Labs Names Project, an effort to standardize name references for politicians to help create more meaningful data by ensuring that different versions of the same name all point to information about the same person.

## Pros and Cons of Government Data Sharing and Mashups

### *Benefits Related to External Partnerships*

W. David Stephenson suggests that open electronic access to government data can result in, “more informed policy debate, grounded in fact, rather than rhetoric, greater transparency and less corruption, optimizing program efficiency and reducing costs, [and] new perspectives, especially when “the wisdom of crowds” emerges.”<sup>25</sup>

Increased debate – Government data in mashups foster greater civic participation when citizens can comment and take action as a result of improved data access that incorporates web 2.0 tools. A number of websites such as WashingtonWatch.com (see above) and PublicMarkup.org<sup>26</sup> promote civic engagement by soliciting comments on new federal legislation. The new OpenMass.org (see above) site also includes a place for comments on proposed state legislation. This trend will broaden as the new federal administration seeks to translate public comment tools used on the campaign website into the official White House website.

Transparency – The Sunlight Foundation is one of the organizations dedicated to improving transparency in government by ensure that public data is easy to access and reuse. Sunlight focuses on, “methods for obtaining basic information on Members of Congress, legislator IDs used by various websites, and lookups between places and the politicians that represent them.”<sup>27</sup>

Less corruption – Since 2005, MapLight.org has sought to illuminate the relationship between money and political influence by releasing public campaign donation data and correlating it with votes taken by government representative on legislation so that voters may examine the data for any patterns that may exist.<sup>28</sup>

Reducing Costs – Washington DC has demonstrated that a larger number and variety of applications can be developed for less money by releasing access to data sources than when

---

<sup>23</sup> <http://services.sunlightlabs.com/api/> [accessed 2/20/2009]

<sup>24</sup> <http://www.sunlightfoundation.com/mashup/> [accessed 2/20/2009]

<sup>25</sup> [http://www.huffingtonpost.com/w-david-stephenson/let-my-data-go-how-activi\\_b\\_112870.html](http://www.huffingtonpost.com/w-david-stephenson/let-my-data-go-how-activi_b_112870.html) [accessed 2/20/2009]

<sup>26</sup> <http://www.publicmarkup.org/> [accessed 2/20/2009]

<sup>27</sup> <http://services.sunlightlabs.com/api/> [accessed 2/20/2009]

<sup>28</sup> <http://www.maplight.org> [accessed 2/20/2009]

government entities try to determine the most useful applications and develop them in-house. Washington DC's Office of the Chief Technology Officer (OCTO) already had a catalog of reusable data, but the Chief Technical Officer, Vivek Kundra, wanted to make it more useful not only for citizens, but also visitors, businesses, and government. Using the same model as the Sunlight foundation, (see above) he launched a mashups contest.<sup>29</sup> "The Apps for Democracy contest featured 60 cash prizes from \$2000 to \$100 dollars for a total of \$20,000 in prizes. Developers and designers competed to create web applications, widgets, Google Maps mash-ups, phone apps, Face book apps, and other digital utilities that visualize OCTO's Data Catalog,<sup>30</sup> which provides real-time data from multiple agencies to citizens — a catalyst ensuring agencies operate as more responsive, better performing organizations."<sup>31</sup> According to Washington DC's partner iStrategyLabs, during the contest, "47 apps were created, representing \$2,000,000 in value for DC.gov, and at a cost of \$50,000 the program represented a 4000% return on investment."<sup>32</sup>

Moreover, Dion Hinchcliffe writing for ZDNet points out, "mashups have enormous potential to allow more rapid and much less expensive development of online applications by emphasizing assembly over development, economies of scale by enabling high levels of reuse, and the consequent ability to rapidly get software solutions with the right data in the right place at the right time."<sup>33</sup>

Why should government data creators go to the trouble of making public data open and accessible for reuse if third parties already provide that service by screen-scraping existing web pages? As noted above, screen scraping can introduce errors. Subsequent reuse of that data can magnify errors many times over. If government entities wish to ensure that public data seen in third party applications is complete and accurate, releasing an API can avoid potentially embarrassing inaccuracies.<sup>34</sup>

As more and more people seek access to data, for a wide variety of reasons, government entities need to consider if it might not be more prudent to stay ahead of the curve by providing reusable access to data sources. A proactive approach could enable governments to get more value out of their own data and meet public expectations at the same time, instead of waiting for lawsuits to pop up that demand access.<sup>35</sup>

### *Benefits Related to Internal Applications*

---

<sup>29</sup> <http://www.appsfordemocracy.org/building-the-digital-public-square/> [accessed 2/20/2009]

<sup>30</sup> <http://data.octo.dc.gov> [accessed 2/20/2009]

<sup>31</sup> <http://www.istrategylabs.com/government-20-the-rise-of-citizen-innovation-through-open-data/> [accessed 2/20/2009]

<sup>32</sup> <http://www.scribd.com/doc/8678520/Government-20-Open-Data-Apps-for-Democracy-> [accessed 2/20/2009]

<sup>33</sup> <http://blogs.zdnet.com/Hinchcliffe/?p=162> [accessed 2/20/2009]

<sup>34</sup> <http://www.mercatus.org/PublicationDetails.aspx?id=21066> p. 36 [accessed 2/20/2009]

<sup>35</sup> <http://publicrecords.onlinesearches.com/public-records-blog/2008/12/05/california-groups-seek-to-expand-access-to-legislative-records/> [accessed 2/20/2009]

Compelling examples of the value of mashups inside government also come from work in the District of Columbia. Because geo-referenced data makes spatial correlations easy to identify and act on, Washington DC used spatially referenced data to implement a computer roll-out to public school employees in seven weeks instead of the 52 weeks anticipated in the original work plan that did not recognize proximity relationships when creating the schedule.<sup>36</sup> RSS access to data allows timely information delivery and can produce cost savings when it is used to monitor project data for efficiencies. Washington DC has used RSS feeds of project reporting data to determine resource allocation for Information technology (IT) initiatives based upon each initiative's performance in much the same way a stock portfolio is managed.<sup>37</sup>

### *Pro or Con?*

Although many expect reusable data will minimize rhetoric, evidence already suggests that it will not do away with it entirely. A widget created by the California governor's office is meant to bring pressure to bear on state legislators by counting the days running since the governor sent his proposed budget to the legislature and providing a running total of the state budget deficit<sup>38</sup> While seeming to promote transparency, this tool also might be viewed as an instrument to forward a political agenda.

### *Drawbacks of Providing Reusable Public Data*

The initial investment in restructuring of electronic data to make it accessible via an API cannot be considered insignificant, particularly if it does not fall within the scope of a project that would derive immediate benefit from the improved access. Data that is not currently available in electronic format could require substantial effort to convert, but API access can be built in during conversion.

John Palfrey and Urs Gasser discuss the roles of data providers and data users in the creation of mashups from a contractual standpoint. They observed,

“the current successes of web mashups have largely depended on players on the database/API side embracing the web 2.0 business model and allowing free access, at least for the time being... The ethos of web 2.0, at least at present, is very strong: the system on which users have come to participate in creating meaning is grounded in sharing data, code, and information. It would be very hard for a company to offer APIs for others to use and then abruptly shut them off, for fear of the consequences of violating online social norms. This factor is yet stronger when the use of the API is for the public interest.”<sup>39</sup>

---

<sup>36</sup><http://www.techpresident.com/blog/entry/33270/district-of-columbia-gives-obama-model-for-effective-transparency-strategy> [accessed 2/20/2009]

<sup>37</sup> <http://www.bloomberg.com/apps/news?pid=20601109&sid=a8q7UONag9nA&refer=home> [accessed 2/20/2009]

<sup>38</sup> <http://mashable.com/2008/12/30/california-budget-crisis/> [accessed 2/20/2009]

<sup>39</sup> <http://cyber.law.harvard.edu/interop/pdfs/interop-mashups.pdf> [accessed 2/20/2009]

Maintenance of data access therefore also needs to be considered. In fact, data access should be a consideration in an overall records management system. The Federal web Managers Council issued a white paper that specifically mentioned the need to evaluate usage and manage on-line content as a key issue that must be addressed in a more systematic fashion saying there is, “too much content to categorize, search, and manage effectively, and there is no comprehensive system for removing or archiving old or underused content.”<sup>40</sup>

Licenses and terms of agreement for data sharing represent another significant area of concern. Any kind of formal agreement can only improve the quality and stability of applications because both the data provider and the data user will have a clear understanding of the limits and obligations entailed in the partnership. The lack of clearly established practices and legal precedence can leave these essentially collaborative relationships open to court challenges when some form of agreement is not in place.<sup>41</sup>

The Washington DC OCTO makes a number of provisos in their terms of use. The first, prohibiting unlawful uses seems almost too obvious to mention. Promotion or misrepresentation of the data or the origin of the data, solicitation, advertising, profit, or collecting personal information is also prohibited. Washington DC's OCTC further stipulates that it cannot be held liable for problems relating to data quality and availability. Washington DC reserves the right to monitor use and holds no liability for data that is incorrect, unavailable, or unsuitable for a particular purpose.<sup>42</sup> Some of these disclaimers run counter to the standards promoted by open government data advocates. Every government entity will need to determine what kind of a balance to strike between access and risk of misuse. Washington DC, like many data providers requires users to register and it controls access by issuing keys. The key can be revoked if any misuse comes to light. This implies some commitment to monitoring should be taken into account when considering use of this method to manage risk.

Data users could benefit from consistent agreements because, as Palfrey and Gasser suggest, “Where these contracts are inconsistent or poorly compatible, innovation can be stifled or, perhaps worse, can continue only to be suddenly thrown off by expensive lawsuits. A useful response to this issue would be to adopt a model akin to Creative Commons<sup>43</sup> of standardized terms of service and other licensing terms for mashups.” They further suggest that, “Such standards could also be communicated through metadata, as with some Creative Commons works, or even as part of an API. If a programmer can simply check if the “useable for profit” attribute is set to “true,” she does not have to worry about license conflicts or changes at all”<sup>44</sup>

Depending upon how data access and use is handled, it could affect standards meant to ensure fair government procurement procedures. Andrea DiMaio writing for Gartner observes,

---

<sup>40</sup> [http://www.usa.gov/webcontent/documents/Federal\\_Web\\_Managers\\_WhitePaper.pdf](http://www.usa.gov/webcontent/documents/Federal_Web_Managers_WhitePaper.pdf) [accessed 2/20/2009]

<sup>41</sup> <http://cyber.law.harvard.edu/interop/pdfs/interop-mashups.pdf> [accessed 2/20/2009]

<sup>42</sup> <http://dc.gov/terms.asp> [accessed 2/20/2009]

<sup>43</sup> <http://creativecommons.org/licenses/by-nc-sa/3.0/> [accessed 2/20/2009]

<sup>44</sup> <http://cyber.law.harvard.edu/interop/pdfs/interop-mashups.pdf> [accessed 2/20/2009]

"Electronic information and service provision so far has been mostly limited to individual agency websites or government-wide portals, whose implementation and operation is either managed internally or tendered out to vendors through traditional public procurement mechanisms. Deciding to reach out to specific communities and channels implies making a choice, and this is going to create issues for conservative government executives. For instance, the increasingly popular "mashup contests" such as those promoted by [www.showusabetterway.com](http://www.showusabetterway.com) or [www.appsfordemocracy.org](http://www.appsfordemocracy.org), while starting with very open premises, may end up making choices about which application is going to use which data. Of course making data available in a mashable format will allow many others to leverage those, but the fact remains that some application developers, websites, communities will get an advantage."

"This is opening two new fronts in the whole web 2.0 discussion. The first one concerns how government organizations may circumvent public procurement by invoking the "power of information": what if mashup contests and similar initiatives become the conduit to procure entire systems in a completely different way, breaking them into smaller chunks that stay below the threshold for competitive tendering? Or are mashup contests a new form of competitive bidding and, if so, should they be further regulated? The second front is how to ensure a smooth functioning of the market created by such "power of information". Players in this market include IT vendors, associations,<sup>45</sup> professional services organizations, online communities, and government entities. What are the competition rules? Which traits would define a dominant player? How to ensure market neutrality when publishing a new piece of public information? While it may look still too early to pose these questions, they will need to be tackled sooner than many think."<sup>46</sup>

When governments create mashups that incorporate citizen comments, another set of problems surface. Efforts to moderate discussion forums or edit annotations and other web 2.0 tools may open government entities to the possibility of suites alleging first amendment violations. This scenario argues for the value in providing access to data without, at the same time trying to develop a suite of related applications. Third party actors may prove more effective hosts for these kinds of social web tools, in most cases.

Concerns about the protection of private data will pose security challenges as public data is made accessible for reuse. Preventing accidental public release of private data held by government agents is already an issue and a comprehensive data management program is the key, regardless of the ultimate disposition of this data.

---

<sup>46</sup> [http://blogs.gartner.com/andrea\\_dimaio/2008/10/28/government-mashups-and-market-neutrality/](http://blogs.gartner.com/andrea_dimaio/2008/10/28/government-mashups-and-market-neutrality/) [accessed 2/20/2009]

Over the long-term, some care will also need to be taken so that any laws intended to make public data more accessible electronically do not become a complex set of contradictory specifications that unintentionally inhibit access.<sup>47</sup>

## Long-term Implications

Information that can be analyzed across the Internet has become increasingly valuable as a result of a fundamental change being brought about by the open source software movement. A general shift in focus has been occurring from the value provided by software tools used to process information to the information itself. Mike Mathieu of frontseat.org predicts that, "Increasingly, data rather than applications will offer the most value to web users."<sup>48</sup> Government entities are seen as a prime source for this kind of information since they typically collect and hold large amounts of data. Mathieu expresses the position of many data users when he says that, "Pressing government to change is necessary but is not the only bet we should place."<sup>49</sup> The fact that users will continue to screen scrape the data they want as well as pressing for direct access through legal channels or in the court of public opinion is inevitable in light of these trends.

How will citizens, indeed government itself, make sense of the massive amounts of data that will become available? Edward Tufte,<sup>50</sup> a leading statistician and information designer, has argued that the most effective way to look at information is in great detail, rather than just in summary form, because summaries can unintentionally alter the original meaning of the data when it is translated into a compressed view, especially the kind of format offered in applications such as PowerPoint. Instead, Tufte asserts that side comparison of data sets allows interpretation of complex information and, "often the most effective way to describe, explore and summarize a set of numbers -- even a very large set -- is to look at pictures of those numbers."<sup>51</sup> Mashups provide an excellent tool for bringing data sets together in a graphical representation in a way that makes an increasingly deep and complex universe of data intelligible and allows new insights and creative solutions to emerge.

W. David Stephenson points out that, "Already, individuals worldwide and groups such as the Sunlight Foundation use these sites [for data visualization tools and mashups] to illustrate and debate issues ranging from US government spending to how nations rank regarding privacy safeguard... The issue of data feeds and visualizations is much more than making pretty pictures: if more government agencies begin to release data and we find imaginative ways to use them to illustrate critical issues and potential solutions."<sup>52</sup>

---

<sup>47</sup> [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1138083](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1138083) [accessed 2/20/2009]

<sup>48</sup> <http://radar.oreilly.com/2008/12/software-for-civic-life-an-interview-with-mike-mathieu.html> [accessed 2/20/2009]

<sup>49</sup> Ibid.

<sup>50</sup> [http://en.wikipedia.org/wiki/Edward\\_Tufte](http://en.wikipedia.org/wiki/Edward_Tufte) [accessed 2/20/2009]

<sup>51</sup> Tufte, E.R. 2001. *The Visual Display of Quantitative Information*, 2nd ed. Graphics Press, Cheshire, CT.

<sup>52</sup> [http://www.huffingtonpost.com/w-david-stephenson/let-my-data-go-how-activi\\_b\\_112870.html](http://www.huffingtonpost.com/w-david-stephenson/let-my-data-go-how-activi_b_112870.html) [accessed 2/20/2009]

Regardless of whether government entities choose to participate in data reuse on the Internet, government data will be included in one form or other, and data creators will feel the pull from this sea-change. A white paper written by Federal web Managers Council Set out a number of recommendations for the new president concerning on-line government, saying “The new Administration should develop government-wide guidelines for disseminating content in universally accessible formats (data formats, news feeds, mobile, video, podcasts, etc.), and on non-government sites such as YouTube, Wikipedia, and SecondLife...Having guidelines will ensure that we’re part of the larger “online information ecosystem,” without compromising the integrity of government information.”<sup>53</sup>

### **Additional Examples of Government Data Mashups and APIs**

Neighborhood Knowledge Los Angeles (NKLA) was developed and is maintained by the UCLA Advanced Policy Institute. <http://nkla.ucla.edu/>

List of mashups created during the Apps for Democracy Contest.  
<http://www.appsfordemocracy.org/application-directory/>

EveryBlock filters an assortment of local news by location keep track of what’s happening on each block, and neighborhood all over the cities of Boston, Charlotte, Chicago, Los Angeles, Miami-Dade, New York, Philadelphia, San Francisco, San Jose, Seattle and Washington, DC.  
<http://www.everyblock.com/about/>

Michigan bill status website that tracks state legislator’s votes and hosts public comments run by Mackinac Center for Public Policy. [http:// www.michiganvotes.org](http://www.michiganvotes.org)

OpenCongress brings together official government data with news articles, blog coverage, and public comments. Jointly run by the Sunlight Foundation and the Participatory Politics Foundation. [http:// www.opencongress.org](http://www.opencongress.org)

The UK government data mashups contest website. [http:// www.showusabetterway.com](http://www.showusabetterway.com)

---

<sup>53</sup> [http://www.usa.gov/webcontent/documents/Federal\\_Web\\_Managers\\_WhitePaper.pdf](http://www.usa.gov/webcontent/documents/Federal_Web_Managers_WhitePaper.pdf) [accessed 2/20/2009]