

UC San Diego

Articles

Title

Remixing the Government: Mashups and Government Information

Permalink

<https://escholarship.org/uc/item/7v17s2xs>

Journal

DttP: Documents to the People, 36(2)

Author

Sklar, Annelise

Publication Date

2008-07-01

In This Issue

- Docs 2.0
- Social Bookmarking and Folksonomies
- Remixing the Government
- The “Facebook” Effect
- Mashing Congress

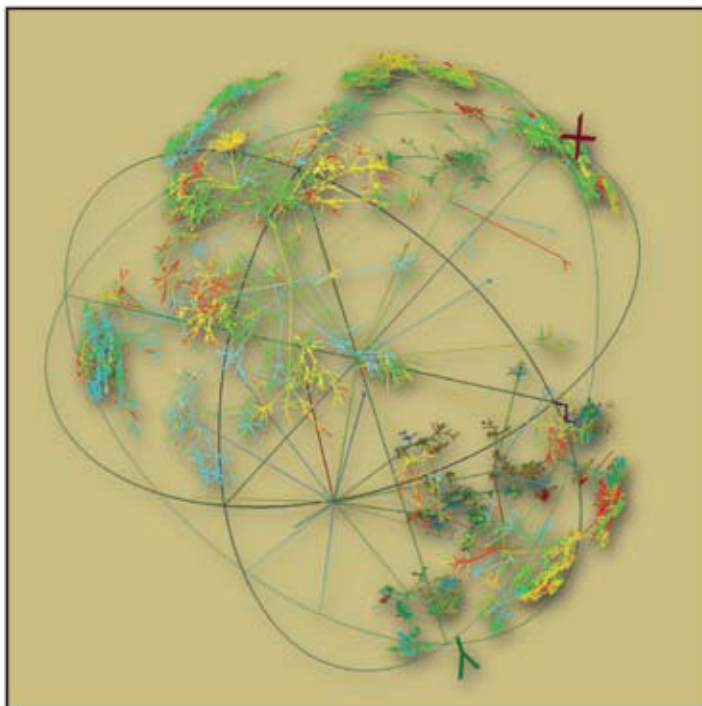
DttP

Documents to the People

Summer 2008 | Volume 36, No. 2 | ISSN 0091-2085



CYBERINFRASTRUCTURE VISION FOR 21ST CENTURY DISCOVERY



National Science Foundation
Cyberinfrastructure Council
March 2007

DttP: Documents to the People (ISSN: 0091-2085) is the official publication of the Government Documents Round Table (GODORT) of the American Library Association (ALA). *DttP* is published quarterly in spring, summer, fall, and winter.

DttP features articles on local, state, national, and international government information and government activities and documents the professional activities of GODORT. The opinions expressed by its contributors are their own and do not necessarily represent those of GODORT.

Editorial Staff

Please see the web site for full contact information: www.ala.org/ala/godort/dtpp/aboutdtpp.htm.

Lead Editor: Andrea Severson, 18 Wheatherstone, Lake Oswego, OR 97035; (301) 951-4749; dtpp.editor@live.com

Editors:

Jackie Druery, Queen's University; (613) 533-6401; drueryj@post.queensu.ca

Chuck Eckman, UC Berkeley; (510) 642-7999; fax (510) 643-8179; ceckman@library.berkeley.edu

Julie Linden, Yale University, (203) 432-3310; fax: (203) 432-3214; julie.linden@yale.edu

Helen M. Sheehy, Pennsylvania State University Libraries; (814) 863-1347; fax: (814) 863-1403 hms2@psu.edu

Contributions Editor: Ben Amata, CSU Sacramento; (916) 278-5672; fax: (916) 278-7089; bamata@csus.edu

Reviews Editor: Susan Tulis, Southern Illinois University, Morris Library, Mailcode 6632, 605 Agriculture Dr., Carbondale, IL 62901, (618) 453-2522, fax: (618) 453-3440; stulis@lib.siu.edu

Advertising Editor: Sherry DeDecker, Davidson Library, University of California, Santa Barbara, CA 93106-9010, 805-893-3713, fax: 805-893-4676, dedecker@library.ucsb.edu

Advertising: Inquiries about advertising may be addressed to the Advertising Editor. Please see www.ala.org/ala/godort/dtpp/adrates for rates and publication schedule. *DttP* accepts advertising to inform readers of products and services. *DttP* will adhere to all ethical and commonly accepted advertising practices and will make every effort to ensure that its practices in relation to advertising are consistent with those of the other Association publications. *DttP* reserves the right to reject any advertisement deemed not relevant or consistent to the above or to the aims and policies of ALA.

Distribution Manager: ALA Subscription Department, 50 E. Huron St., Chicago, IL 60611. 1-800-545-2433, press 5; fax: (312) 280-1538; subscriptions@ala.org

Subscriptions/Claims: *DttP* is sent free to ALA/GODORT members on a per volume (annual) basis. For subscriptions, prepayment is required in the amount of \$35 in North America, \$45 elsewhere. Checks or money orders should be made payable to "ALA/GODORT" and sent to the Distribution Manager. Changes of address and claims six months of the date of issue should be sent to the Distribution Manager. To purchase back issues, write to: UMI, 300 North Zeeb Rd., Ann Arbor, MI 48106. POSTMASTER: Send address changes to *DttP*, 50 E. Huron St., Chicago, IL 60611. Periodicals postage is pending at Chicago, Illinois, and additional mailing offices.

Contributions: Articles, news items, letters, and other information intended for publication in *DttP* should be submitted to the Lead Editor. All submitted material is subject to editorial review. Please see the web site for additional information, above.

Indexing: Indexed in Library Literature 19, no. 1 (1991), CSA Worldwide Political Science Abstracts 33, no. 1 (2005), and selectively in PAIS 33, no 1 (2005).

Editorial Production: ALA Production Services—Troy D. Linker, Karen Sheets, Chris Keech, and Tim Clifford.

Columnists:

On the Range

Brian Rossmann
Montana State Univ.
brossmann@montana.edu

Geospatial News

Marcy M. Allen
Pennsylvania State Univ.
mallen@psu.edu

International Documents Roundup

James Church
Univ. of California, Berkeley
jchurch@library.berkeley.edu

Washington Report

Kevin McClure
Chicago-Kent College of Law
kmclure@kentlaw.edu

State and Local Documents Roundup

Kris Kasianovitz
Univ. of California, Los Angeles
krisk@library.ucla.edu

By the Numbers

Stephen Woods
Pennsylvania State Univ.
swoods@psu.edu

Tech Watch

Valerie Glenn
Univ. of Alabama
vglenn@ua.edu
Amy West
Univ. of Minnesota
westx045@umn.edu

News from the North

Michael McCafferty
Univ. of Toronto
michael.mccafferty@utoronto.ca

DttP

Documents to the People

Summer 2008 | Volume 36, No. 2 | ISSN 0091-2085

Columns

- 4 Editor's Corner
- 5 From the Chair
- 7 Washington Report
- 9 News from the North
- 12 State and Local Documents Roundup
- 16 Tech Watch

Articles

- 19 Docs 2.0
An Introduction
Valerie Glenn and Amy West
- 21 Social Bookmarking and Folksonomies
Possibilities for Government Information?
Tea Rokolj
- 25 Remixing the Government
Mashups and Government Information
Annelise Sklar
- 29 The "Facebook" Effect on FDLP Communication
Using Social Networking Tools in a New Context: A Pennsylvania-based Approach
Tom Adamich
- 33 Mashing Congress
Using Web 2.0 Tools to Teach Congressional Research
Karen Munro and Jesse Silva
- 39 International Information Sources
LIS Education for Global Environments
Debbie L. Rabina

Round the Table

- 44 The Interview
Cass Hartnett, Incoming GODORT Chair (2008–09)
- 44 GODORT Awards Nominations
- 45 Research and Scholarship Applications
- 45 Index to Advertisers

Copyright © 2008 ALA/GODORT

About the Cover: *Cyberinfrastructure Vision for 21st Century Discovery* (purl.access.gpo.gov/GPO/LPS80410). The visualization on the cover depicts a single cycle of delay measurements made by a CAIDA Internet performance monitor. The graph was created using the Walrus graph visualization tool, designed for interactively visualizing large directed graphs in three-dimensional space. For more information see www.caida.org/tools/visualization/walrus.

Remixing the Government

Mashups and Government Information

Annelise Sklar

Library and information science (LIS) scholarly communication is rife with discussion of web two-point-this and library two-point-that and how we can dramatically change the library paradigm with new technology. As trendy—and (two-point-)oh-so-easy to poke fun at—as the new jargon is, Web 2.0 is fundamentally about web technology giving users individual choices in how they access and use information. This is embodied in interactive tools like blogs (users publishing information themselves), wikis (users creating and editing information they access), personalized portals and widgets (users customizing their online services), and social networking (users sharing their content with others who are somehow like them).

Then there are mashups, where users repurpose existing information and tools. A mashup is a web-based application that combines or reformats into a single interface one or more sets of data or online services. The term is borrowed from a musical genre where parts of one song are overlaid onto another to form a new song.¹ On the web, a mashup is a program that combines elements of previously unconnected web sites or data sources (such as blogs, wikis, portals, widgets, social networks, or databases) into one (hopefully) more useful or user-friendly interface, often with functions not found in any of the original resources. For example, Libraries411 (www.libraries411.com) takes public library location data from the National Center for Education Statistics and maps it using a Google, Yahoo!, or Microsoft Virtual Earth map. Another mashup, Libworm (www.libworm.com), compiles RSS (Really Simple Syndication, a subset of general markup language XML and the means by which blogs are published) feeds from multiple library blogs and repackages the posts by area of interest.²

Mashup data is usually derived from other web sites or programs in the form of APIs (application programming interfaces), data feeds, or screen-scraping. An API is the set of functions a program needs in order to “talk” directly with another program. These days, most key Web 2.0 players—Google,

Yahoo!, Amazon, eBay, and the like—open their APIs to developers. Other information producers publish their data as a feed. RSS is one well-known data feed type. Screen-scraping is a method of data extraction used when data is not neatly packaged in a usable format. As the name suggests, a screen-scraping queries another program and then captures the display data for other use.

Naturally, mashups have received some LIS attention. They are discussed in various publications, and UK-based vendor Talis sponsored a contest for library-themed mashups.³ Not surprisingly, mashups can also make government information more accessible and user-friendly by blending government-produced data with resources from other government agencies or non-government sources into new services. Some govdoc’ers are already in the know: *Government Computer News (GCN)* features a guide to map mashups of government-produced data; Karen Huffman and Dan Newman presented on government information mashups at the 2007 Special Libraries Association Annual Meeting; and Laura Gordon-Murnane included mashups in her examples of Web 2.0 uses of federal information.⁴ Building on this momentum, this article discusses finding, evaluating, and making mashups, while highlighting some of my favorites that use federal, state, and international government information.

Mashups can combine any kind of data or tools, but the most commonly seen piece is a map, usually Google Maps (maps.google.com), Yahoo! Maps (maps.yahoo.com), or Microsoft Virtual Earth (www.microsoft.com/virtualearth). Maps mashups are popular because they simply illustrate a point and are cheap (usually free to the end user), quick, and easy to use, and many of them feature practical, everyday

Discuss this article on the wiki:
<http://wikis.ala.org/godort/index.php/DttP>

information. Popular government-produced information topics such as public transit, crime, weather, natural phenomena, environmental concerns, boundaries, and statistics all lend themselves well to maps mashups. With maps mashups, I can often quickly see what's going on out there all at once, from how many sex offenders live near me using US-SEX-OFFENDERS.COM, to which celebrities are giving money to which candidates with Money Track at Political Base (www.politicalbase.com/money/search).⁵ I can look at Census 2000 demographic data for the country or a subset of it with gCensus (gcensus.com), track volcanic activity with Active Volcanos of the World (www.geocodezip.com/v2_activeVolcanos.asp), or get train schedules from sites like BART Station Maps and Timetables (bart.barelyconnected.net).⁶

Mashups are useful because they serve as all-in-one research tools but, like other databases and database-driven tools, they are something of a pain collection development-wise: the only way to “collect” them is to provide links on the library web site, in the catalog, in subject guides, and so forth, like we already do with other online resources. They also pose preservation problems, as most online archiving tools struggle with dynamic pages. Since depository coverage will be spotty at best—and many of the tools are made by non-government entities—government information specialists must continue being proactive in maintaining guides to government information tools. However, because the beauty of well-done mashups is that users will notice nothing special about them except that data searches are streamlined into one search in one place, mashups are just as likely to be listed as useful sites on whatever their topic as they are “mashups.” There's no single place to find them all, and more appear daily, but Programmable Web's Government APIs and Mashups Dashboard (www.programmableweb.com/government), Google Maps Mania (googlemapsmania.blogspot.com), and the Sunlight Foundation's Insanely Useful Sites (www.sunlightfoundation.com/node/2) are good starting places.

As librarians we're naturally inclined to approach mashups cautiously because they're trendy, and as a profession we've been burned by new formats before. Nonetheless, once we've found them, we should evaluate mashups with the same collection development rubrics we use with other sources, looking at authority, accuracy, currency, scope, audience, ease of use, and so forth. Most Internet users don't really care where or who their information comes from, but the part that will throw most librarians, especially documents librarians, is authority: by their very nature, mashups contain data from multiple sources and often it must be tweaked to be usable. Thus, mashups of government information are not always official government resources; they can be third-party applications using gov-

ernment data, sometimes provided by a secondary source.

In theory, government-produced mashups have the same authority as any other government information: it's generally preexisting data displayed in a more user-friendly way. For example, the World Bank's Geo.worldbank.org is a Google map with clickable county points linking to statistical information, news, and project information. Business Planet (rru.worldbank.org/businessplanet) is a similar maps mashup with information from Doing Business and the Enterprise Surveys. The United Nations Environment Programme revamped the print *One Planet, Many People: Atlas of Our Changing Environment* as *Atlas of Our Changing Environment* (na.unep.net/unep-atlas.php) on both Google Earth and Google Maps. At the federal level, many agencies are starting to provide Google Maps views, such as U.S. Geological Survey's WaterWatch (water.usgs.gov/waterwatch/?m=real&cw=gmap), which plots streamflow conditions. Reportedly, even the Defense Intelligence Agency uses an internal mashup combining human intelligence with public Internet information into a single analysis tool.⁷

“Official” mashups from government and IGOs will probably become more prevalent over the next few years, but the process is slow because bureaucracies must change infrastructure in order to adopt new technology. For example, the Environmental Protection Agency (EPA) demonstrated an interest in mashups with its November 2007 Office of Environmental Information Symposium, which included a “mashup camp” structured around a wiki project for the Puget Sound Leadership Council, and Chief Information Officer Molly O'Neill told *GCN* that “We [the EPA] definitely have our toes dangling in the pond. We are trying to figure out the policies and procedures for using Web 2.0 externally.” However, the EPA and other agencies still must decide how to moderate content, write new job descriptions, determine the best ways to meet high information demands, and update old data systems to work with Web 2.0 technologies before they can fully jump in.⁸

In the meantime, the gap is filled by nongovernment mashup creators, who are generally individuals, groups, or companies that see an information need and decide to fill it themselves. A prime example is GovTrack (www.govtrack.us), which was created by University of Pennsylvania linguistics graduate student Joshua Tauberer in 2004 and has pretty much revolutionized legislative history research, providing access to bills by number, subject, and keyword and combining in one page the status information, voting records, cost analysis, and links to multiple versions of the bill text, analysis, related legislation, and like information.⁹

Mashups from nongovernmental sources require evaluation by librarians before we recommend them, though most I've looked at seem like useful tools like Opensecrets' Travel Database (www.opensecrets.org/travel) or Environmental Working Group's U.S. Mining Database (www.ewg.org/sites/mining_google/US). Nonprofit organizations often use mashed up government data to highlight for the public what the government is doing, where governmental players are doing it, and who's paying for it. These are often the same watchdog groups that make FOIA requests, raise a stink when publications disappear from agency web sites or library shelves, and generally try to hold the government accountable for its actions. One of the most notable nonprofits is the Sunlight Foundation, whose Sunlight Labs (www.sunlightlabs.org) has several mashup projects including a government information mashup contest (www.sunlightfoundation.com/mashup) and an API (sunlightlabs.com/api) featuring basic information about current members of Congress. Sunlight Labs also created several mashups of its own, most notably LOUIS, the Library of Unified Information Sources (www.louisdb.org), which provides unified searching of executive and congressional documents available through *GPO Access*.¹⁰

On the other hand, it's easy to drum up outrage without a full explanation of the data or a discussion of what "normal" spending, pollution, and so forth, is. A site could easily be a front for a dishonest political action committee or an adware company or some other perpetrators of evil. I haven't heard any reports of such behavior yet, but it will probably happen at some point as we all become more creative in using the Internet. However, data can always be manipulated to further an agenda—be it liberal, conservative, or toward a product—and many users consider governmental agencies biased, too. As librarians, our job is to read the background information and know who's behind the data and how it's being manipulated, and then recommend and use the tools accordingly.

We can also develop our own mashups, just as we create pathfinders, bibliographies, indexes, reference books, and custom search engines. We don't need to plot out everything we encounter with Google Maps just to do it, but as information mediators we should use or create whatever tools best meet our and our patrons' information needs. And, after all, because we're information professionals with no agenda but leading users to information, the tools we create would, theoretically, be free of the bias and errors we suspect in others. For beginners, Programmable Web has a how-to page (www.programmableweb.com/howto) that includes links to popular APIs and tutorials in using them. Sunlight Labs also has a tutorial with an example mashup of their API (sunlightlabs.com/api/

example/explanation.php). Tools like Yahoo! Pipes (pipes.yahoo.com), Microsoft Popfly (www.popfly.com), Google Mashup (editor.googlemashups.com), and IBM's QEDwiki (services.alphaworks.ibm.com/qedwiki) can aid in connecting and manipulating data feeds and APIs, though none of them are as quick and easy as I'd hoped: even with their help, one still needs to understand how to manipulate variables, format output, and create a pleasing interface, and, at least right now, that takes programming skill and time to experiment.

Still, the hardest part in planning a mashup may be finding usable data, as it is hidden all over the web. Just to see what's out there, one might Google the data type and limit to the .gov domain (for example, `kml site:.gov`; `xml site:.gov`; Google Earth `site:.gov`) or search through USA.gov. Many agencies provide data web sites like the EPA Databases and Software page (www.epa.gov/epahome/Data.html) or the live feeds from the District of Columbia Center for Innovation and Reform (cir.oca.dc.gov/cir/site/default.asp). USA.gov's Data and Statistics—General Reference Resources (www.usa.gov/Topics/Reference_Shelf/Data.shtml) is a good place to start, as is Geodata.gov (gos2.geodata.gov/wps/portal/gos). As it's hard enough to hunt down government data in the first place, it's probably easiest to build a mashup around data with which one is somewhat familiar.

Keep in mind, both as creator and user that mashups are only as good as the data they use, and sometimes the only data available is old or incomplete. Likewise, if any data source changes (say, if an agency moves its data or changes its output type), the mashup will probably cease working until it too is updated. Ditto if the agency takes the data down, say for Homeland Security or copyright concerns. Then there's the longevity issue: if a mashup is created by a solo creator in her spare time, all it takes is one life change and the mashup might be abandoned. I lost count of the mashups that disappeared in the months between my starting and finishing this article. While really amazing mashups may be adopted by another interested party, such as those now maintained by the Sunlight Labs, or purchased like Weather Bonk (www.weatherbonk.com), which was bought by Weather Channel Interactive in September 2007, many just fade away, as, like with open access software, a stable resource usually isn't sustainable without some sort of financial backing or strong community efforts.¹¹

As librarians, our job is to match patrons with information they can use, and more and more, that will be through a mashup. Librarians should incorporate mashups into our collection development of online resources and keep up with them, since they may change or disappear at will. As we select them for our patrons, we must be vigilant in our evaluation

of their authority, aware of shortcomings in the data used, and, at times, verify the information using traditional sources. Mashing data into a usable tool is time-consuming, but in doing so, librarians can create the (free) online resources we'd like to see the way we'd like them to be. Librarians without the time or inclination to create and maintain tools could always lobby our techie peers or government agencies to make them instead. Though the process will be slow and success rates will vary, these agencies are, after all, supposed to make information available for the public good, and mashups can help them be more efficient: the easier the data is to access, the easier it is to do more with it themselves or in collaboration with external partners. Even with their potential drawbacks of possible data manipulation, functional longevity issues, and the Section 508 accessibility issues that most dynamic web sites have, the user-friendly interfaces of mashups will appeal to many users who might not otherwise access government information. And they might do so in ways we haven't thought of yet.

Annelise Sklar, Librarian for Political Science, Law and Society, and State, Local, and International Government Documents, University of California–San Diego, asklar@ucsd.edu.

References and Notes

1. The most well-known example is DJ Danger Mouse's overlaying of Jay-Z's *The Black Album* on The Beatles' *White Album* to make *The Grey Album*.
2. For other examples of library-related mashups, check out the entries in the Talis Mashing Up the Library competition (www.talis.com/tdn/forum/84).
3. Mashups are often mentioned in articles on Web 2.0, but a few have dealt specifically with mashups. For example, Darlene Fitcher, "Doing the Monster Mashup," *Online* 30, no. 4 (July/Aug. 2006): 48–50; Paul Miller, "What Happens When We Mash the Library?" *Ariadne* 50 (2007), www.ariadne.ac.uk/issue50/miller.
4. Joab Jackson, "A Quick Guide to Citizen-Mapped Agency Data Sites," *GCN* (Apr. 16, 2007), www.gcn.com/print/26_08/43502-1.html; Cassandra Harper, "Mashups and Remixes for Government Information," Special Libraries Association Government Information Division, units.sla.org/division/dgi/print/print_mashups.html; Laura Gordon-Murnane, "Mashups, Blogs, Wikis Go Federal: Creative Uses of Government and Public Data," *Searcher* 15, no. 3 (March 2007): 33–39.
5. For more crime mashups, see the crime sections on Google Maps Mania (googlemapsmania.blogspot.com/#crime) and Programmable Web (www.programmableweb.com/tag/crime).
6. For more transit mashups, see the transit sections on Google Maps Mania (www.googlemapsmania.blogspot.com/#transit) and Programmable Web (www.programmableweb.com/tag/transit).
7. Heather Havenstein, "Top Secret: DIA Embraces Web 2.0: Analysts Are Turning to Wikis, Blogs, RSS Feeds and Enterprise 'Mashups,'" *Computerworld* (Feb. 23, 2007), www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=9011671.
8. Joab Jackson, "Molly O'Neill | EPA the Web 2.0 Way: GCN Interview," *GCN*, Feb. 4, 2008, www.gcn.com/print/27_3/45741-1.html?topic=&CMP=OTC-RSS.
9. For an in-depth review, see Peggy Garvin, "The Government Domain: GovTrack and OpenCongress Go Beyond THOMAS," LLRX.com (Mar. 18, 2007), www.llrx.com/columns/govdomain25.htm.
10. See also Peggy Garvin, "The Government Domain: 'Insanely Useful' Legislative Sites," LLRX.com (July 27, 2007), www.llrx.com/columns/govdomain28.htm.
11. "Mashup Acquired by the Weather Channel," Programmable Web (Sept. 12, 2007), <http://blog.programmableweb.com/2007/09/12/mashup-acquired-by-the-weather-channel>.

DttP Online!

www.ala.org/ala/godort/dttp/dttponline

Check out the new and the old! The digital archive, hosted by Stanford University Libraries & Academic Information Resources, contains all issues of the journal published from its inception in 1972 through 1998 (volumes 1–26). The contemporary material, 2003 (volume 31) to present, is hosted on the ALA/GODORT server.