

What Is a Mashup?

Darlene Fichter
University of Saskatchewan Library

A mashup is a web application that uses content from more than one source to create a single new service displayed in a single graphical interface. For example, you could combine the addresses and photographs of your library branches with a Google map to create a map mashup (Figure 1.1). The term *mashup* originally comes from pop music, where people seamlessly combine music from one song with the vocal track from another—thereby mashing them together to create something new.

Mashups have recently exploded on the web, for two main reasons. First, many of the major internet companies, such as Yahoo! (www.yahoo.com), Google (www.google.com), and Amazon (www.amazon.com), have opened up their data to be used with other data sources without a lengthy licensing negotiation. In just a minute or two, you can set up and use the data resources they make available. The other reason for this rapid growth is the advent of new tools that make creating mashups easy for anyone, regardless of their technical know-how.

Popular Mashups

The most popular type of mashup is a map mashup. Map mashups make up 36 percent of the mashups tracked by ProgrammableWeb (www.programmableweb.com), the most comprehensive listing of mashups. Figure 1.2 shows the distribution of mashups by type.

Other types of popular mashups mix video and photos. For example, the Viral Video Chart (viralvideochart.unrulymedia.com) site tracks

4 Library Mashups

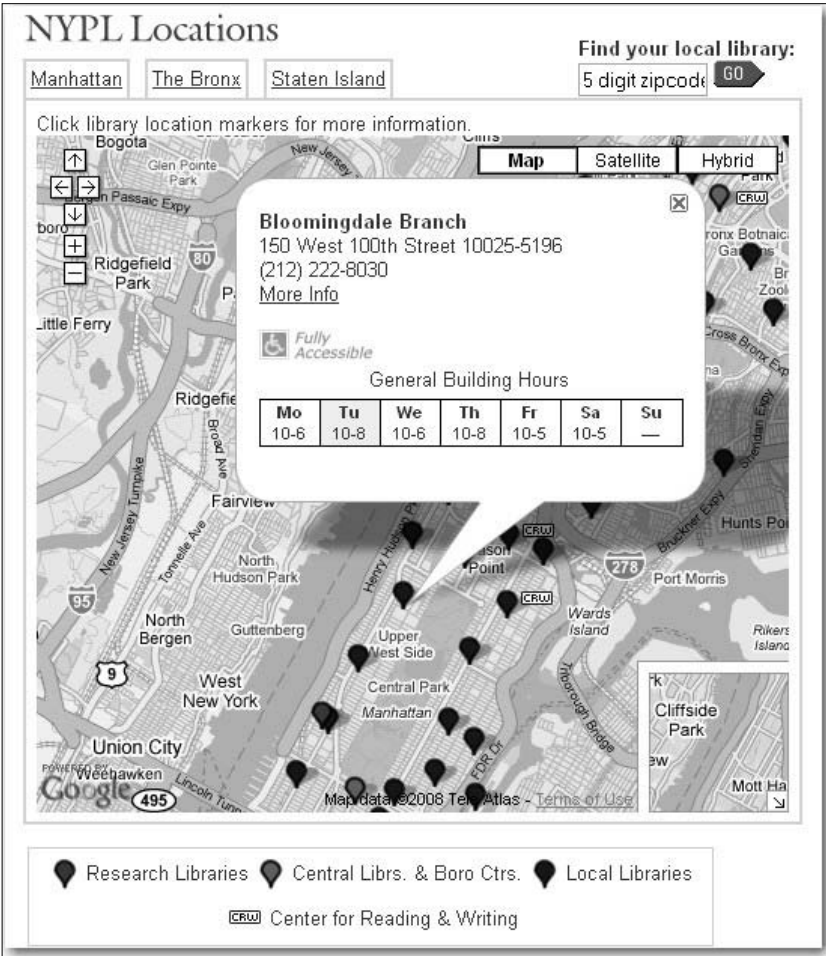


Figure 1.1 Google Maps mashup of New York Public Library branches

YouTube (www.youtube.com), MySpace (www.myspace.com), and Google Video (video.google.com) in order to identify the most talked about new clips, overall and by category (Figure 1.3).

There are many wonderful mashup services built using photos. Here are three examples to spark your imagination:

1. Colr Pickr (krazydad.com/colrpickr), developed by Jim Bumgardner, lets you search Flickr (www.flickr.com) photos by color. Flickr is a widely used photo-sharing site (offering both free and professional accounts) that attracts many libraries and librarians. Use Colr Pickr to click on a

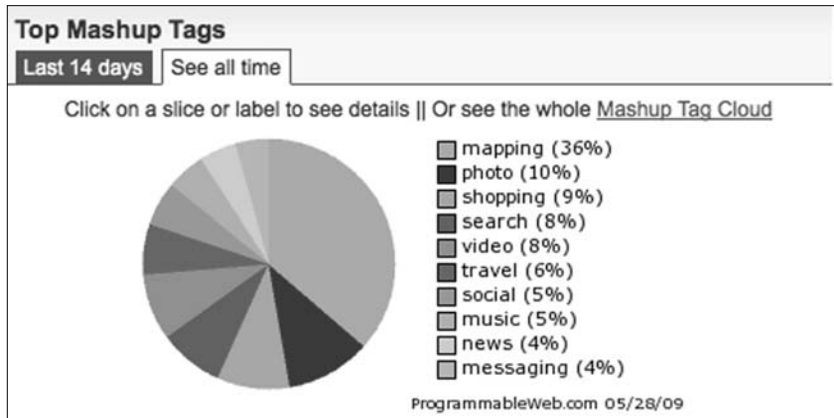


Figure 1.2 Pie chart showing popularity of different types of mashups

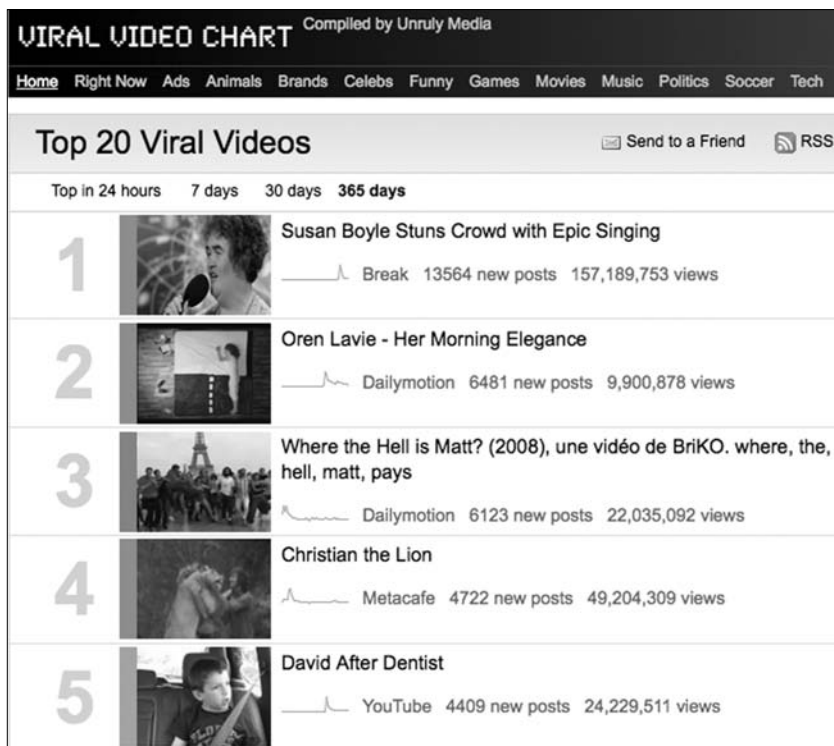


Figure 1.3 Most talked-about videos for the past 365 days

6 Library Mashups

color in a photo or color wheel, and it will retrieve photos that have a large concentration of that color.

2. There are also mashup tools that use photographs to create books, posters, magazine covers, and so on. Most of us have received photograph booklets comprised of stunning images with short captions. One way to create these photograph books is to use a mashup tool called Bookr (www.pimpumpum.net/bookr). Start by searching Flickr to locate images. Then, add these photos to a booklet page and type in your caption. Save your work, and email the book to friends or publish it on your blog or library website.
3. A mashup tool called Ad Generator (theadgenerator.org), created by Alexis Lloyd for his MFA thesis project, is intended to inform, enlighten, and entertain site visitors. The mashup randomly combines slogans and images to explore the relationship among language, manipulation, and images.

Mashups come in all shapes and sizes, from the very simple to the complex. Some mix and mash up search results, others introduce interesting visualizations, and still others aggregate and combine newsfeeds. Libraries have lots of opportunities to use mashups to help liven up their websites, deliver new and interesting services, or entertain website visitors. We'll explore some of these options later in this chapter and throughout this book and hopefully spark some new ideas for mashups at your own library.

Mashup Ecosystem

The mashup ecosystem contains some wonderful ingredients to make a rich and fertile environment; Figure 1.4 identifies the major players in the mashup ecosystem. The fundamental component of the mashup ecosystem is the “open” data providers operating in the internet “cloud.” Companies and organizations such as Yahoo!, Google, Technorati (technorati.com), EVDB (eventful.com), Flickr, government departments, academic research units, and think tanks offer up social, economic, and scientific data. The beauty of the internet is that it's easy to open up data sources. No one needs permission. An organization simply makes a data source available, describes how to access it, and announces its availability, normally setting out the

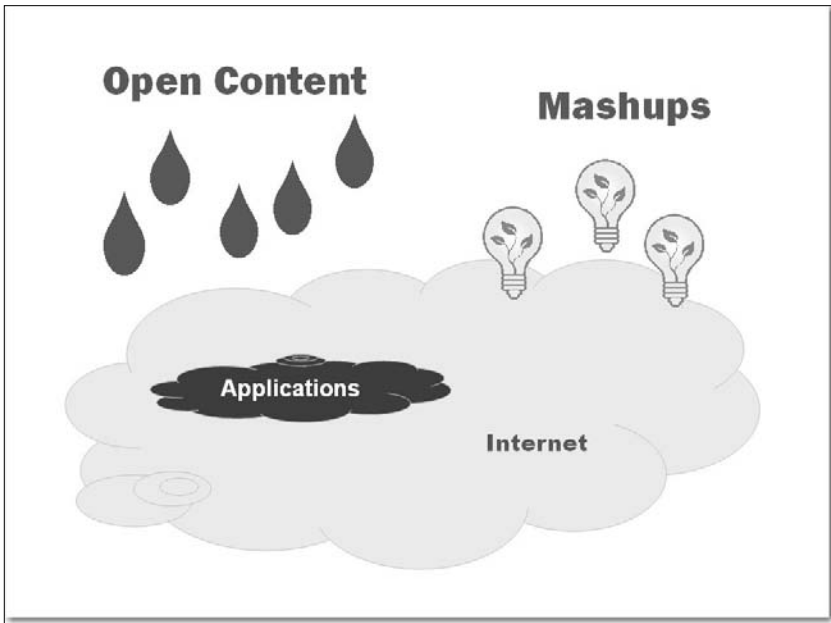


Figure 1.4 Mashup ecosystem

acceptable uses for that data source. Many companies permit free use of their data sources for noncommercial applications; sometimes there are caps on the number of requests per day. Commercial licensing options are also available for some data sources so that corporations can mash up the data source inside the enterprise or use it to build an application for consumers. Most libraries fall into the non-profit use of data, so it's easy to find data sources to use.

Once the ecosystem is populated with a rich array of data sources, an easy means to access or query the data source is needed so that you can combine the data on your website or in your mashup application. For example, if you would like to add pushpins for library locations to the map of your city or town, you need a mechanism to request a local map zoomed into the appropriate scale to show library locations. Typically, data providers permit access to their information, either as an RSS feed or other XML (eXtensible Markup Language) marked-up format, or via an application programming interface (API). An API spells out how to formulate a query for the data. Usually these queries are written by programmers in languages such as PHP, JavaScript, Perl, Java, .NET, or Python. Programmers are very familiar with APIs and can easily write a few lines of code to collect the data. What if you're

8 Library Mashups

not a programmer? No worries—there are lots of web-based applications that let you make use of APIs by pointing, clicking, and pasting a snippet of generated code into your website or blog.

Check out some of the data sources that you can remix into mashups at ProgrammableWeb, Roy Tennant's list of Library Application Program Interfaces (techessence.info/apis), and the JISC Information Environment Service Registry (iesr.ac.uk).

Once the ecosystem is well supplied with raw ingredients, the environment is set to support various “life forms,” or new creations. These raw materials need an engaged group of creators and consumers. The creators are the people with a “spark.” They can see how two or more things can be combined to make something new, richer, or better. And of course, creators need an audience, or consumers, eager to explore and make use of their work. Mashups have been very successful because they allow the end user to be a creator and because there is an eager audience for these new creations.

Where do libraries fit into this ecosystem? Everywhere. We can be data providers, allowing our customers and mashup developers to remix our data, including acquisition lists, most popular titles, catalog records, event and program information, digital collections, and so forth. Libraries are also mashup creators, mixing open data sources with each other or with in-house data sources. And last but not least, we are consumers. Libraries benefit from mashups for internal use and as information sources for our patrons.

Library-Created Mashup Tour

Let's dive in and take a look at how some libraries are using mashups to create new services and features for their websites.

1. Library Locations

Let's start with a couple of examples of the most popular type of mashup, the map mashup. The simplest map mashup can be created by adding a map link to your library locations page. The link calls up Google Maps (maps.google.com) with a pushpin for the location of your library and an overlay window that provides some basic information, such as library hours. For an example of this approach, look at Cambridge Libraries and Galleries' location page (cambridge-libraries.ca/library.cfm?subsection=locations).

On the Hillsborough County Public Library Cooperative (HCPLC) website, a Google Maps mashup shows the locations of the libraries and partner agencies (www.hcplc.org/hcplc/liblocales) (Figure 1.5). Another page uses color-coded pushpins to group different kinds of service locations (www.hcplc.org/hcplc/liblocales/locationsallmap.html). One of the nice features of a Google Maps mashup is the “get directions” capability.

2. Sweetening Up the Library Catalog

The University of Texas tries to provide its users with more from the OPAC by using Google Books Preview. Whenever a visitor looks at a page in the catalog, a script runs behind the scenes to check whether the book is available in Google Book Search. If it is, a hyper-link for Preview appears. Clicking this link will open an overlay window for browsing the full text. Figure 1.6 shows the Preview window overlaying the catalog search results for the title *Reworking the Student Departure Puzzle*.

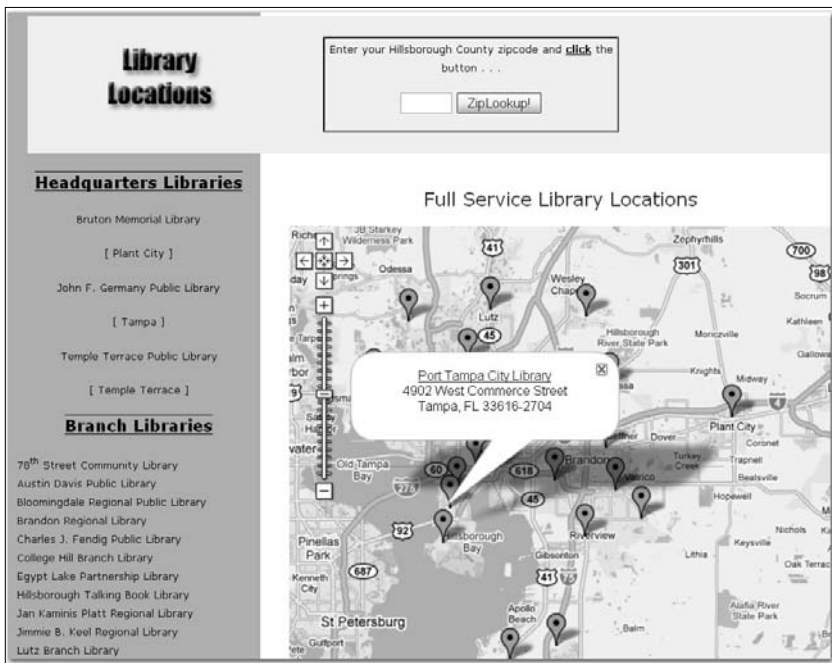


Figure 1.5 Google Maps mashup of Hillsborough County Public Library Cooperative locations

10 Library Mashups

Title: Reworking the student departure puzzle / edited by John M. Braxton.
Edition: 1st ed.
Imprint: Nashville [Tenn.] : Vanderbilt University Press, 2000.

Location: EC1245
Call No.: LC148 .L5.B48 2000
Current Status: DUE 01-22-09

Description: vi, 288 p. : ill. ; 24 cm.
Bibliography: Includes bibliographical references and index.
Series: Vanderbilt issues in higher education
Subject: College dropouts -- United States.
College attendance -- United States.
Info, Vincent.
Added author: Braxton, John M.
ISBN: 0526513085
OCLC number: 44681704

from Google Book Search

Go Book Search

Contents search

Reworking the Student Departure Puzzle

Buy this book More about this book

Figure 1.6 Google Books Preview integrated into library catalog result page

The developers also display book cover images that are mashed in using the Amazon API (aws.amazon.com), as well as tags and reviews from LibraryThing (www.librarything.com).

3. Creating Discovery Tools for Collections

A popular type of library mashup that enhances discovery of collections and resources is the book cover carousel. Just like the “new book” or “recommended book” shelves in libraries, library websites can offer a dynamic display of book covers by combining book lists and book covers. We’ll take a look at one of these later, in the user-created mashups section.

Figure 1.7 shows an ideal use of a map mashup for discovery. McMaster University Library had print indexes to its aerial photographs collection. By putting these online with a Google Maps mashup, the library allowed researchers to see the photos available for a particular location and/or a particular year. It’s incredibly fast and easy to browse what’s available.

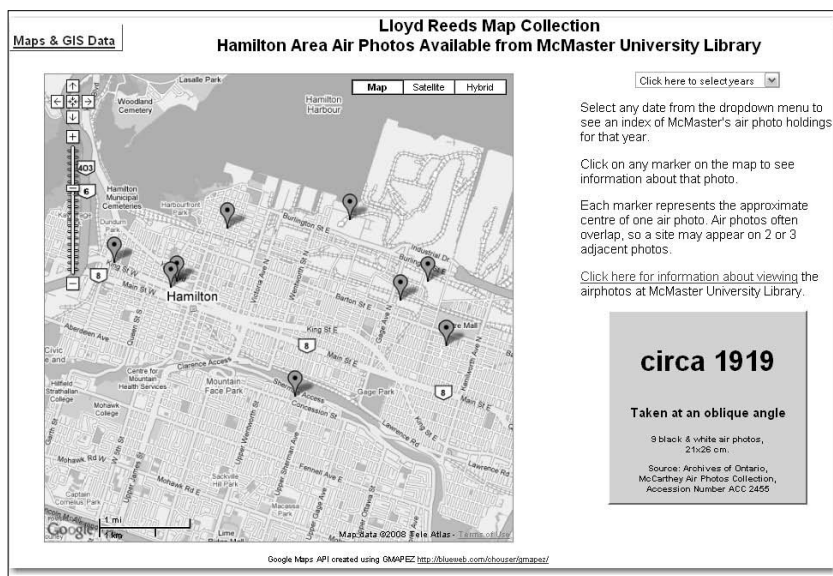


Figure 1.7 Google Maps mashup of Hamilton area air photos circa 1919

User-Created Library Mashup Tour

Next let's see how some libraries have become data providers and how some library users are mixing up that data to provide interesting, exciting, and occasionally wacky mashups.

1. Finding Short Books

Let's start with a bit of a wacky mashup—the LazyLibrary (www.lazylibrary.com). Every school and public librarian has heard a request from reluctant or time-pressed readers for a short book about “topic X.” According to the LazyLibrary website, “this mashup pulls in book data from Amazon and filters out anything with more than 200 pages. Add to that an exceptional user interface and you have one solid service for anyone looking to get more out of reading less” (Figure 1.8).

2. Visualizing Your Bookshelf

Edward Vielmetti, also known as Superpatron (vielmetti.typepad.com/superpatron), was an early pioneer in showing how library users can remix library data streams. He took an RSS feed from the Ann Arbor District Library's catalog (www.aadl.org/catalog) and created a

12 Library Mashups

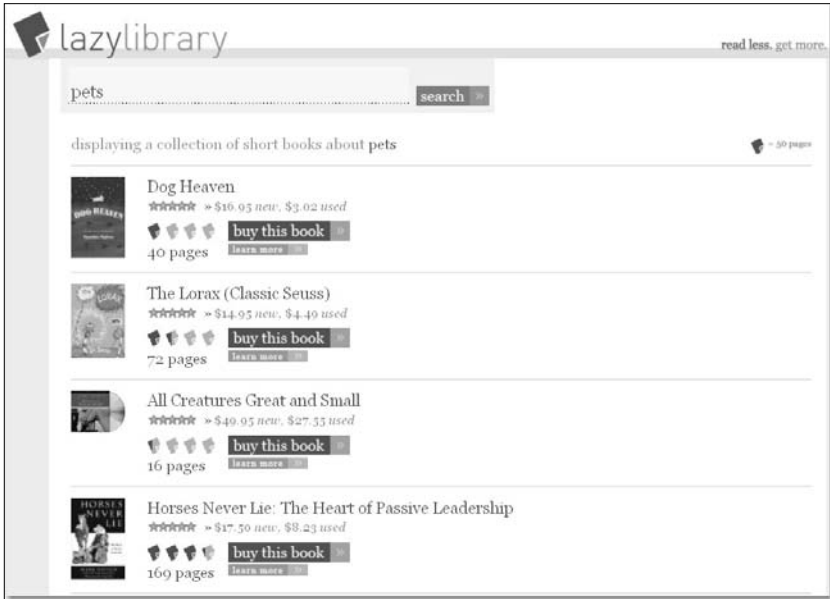


Figure 1.8 “Short Books” about pets from LazyLibrary



Figure 1.9 Wall of Knitting Book Covers created from the RSS feed from the Ann Arbor District Library

wall of book covers showing new nonfiction books, new children's books, and new knitting books (Figure 1.9).

The book covers were from Syndetic Solutions, and the RSS feed was generated by a catalog search. Vielmetti did this as a proof-of-concept exercise, and his work inspired Kate, an avid knitter who blogs at *Four Obsessions: Reading, Writing, Cooking and Crafting* (4obsessions.blogspot.com), to post the new knitting books and their covers on her blog (Figure 1.10). She also added the RSS feed to her reader to get the jump on requesting new items. Kate has another area on her blog where she displays covers of books she's checked out of the library.

FOUR OBSESSIONS: READING, WRITING, COOKING AND CRAFTING

A PLACE FOR FRIENDS AND FELLOW OBSESSORS TO GATHER

SATURDAY, SEPTEMBER 06, 2008

A return to fiber

Now that there is some cool air moving back in to my surroundings, I'm willing to immerse myself up to my armpits in woolly goodness again. My serving-as-a-cat-bed for the summer Urban Aran Cardigan has been retrieved from the felines in time for the restarting of my son's piano lessons (also known as my guaranteed one-hour of knitting time.)

And I figure I really should get going on it since there are a bevy of new knitting books at the library that I've requested and will be receiving in the near future. Thanks to Ed, who taught me how to create an RSS feed for any new knitting book that my library enters into its catalog, I get first notice on acquisitions and can get my request in at the front of the line. I usually get the books within a month or two of them being added to the collection.

Coming my way in the near future are:



BOUTIQUE KNITS

Boutique Knits



Closely Knit

Closely Knit

ABOUT ME

KATE

[VIEW MY COMPLETE PROFILE](#)

FLICKR

www.flickr.com



what is this?

LABELS

- beverages (17)
- Book Group (18)
- Books (104)
- Crafts (16)
- food thoughts (120)
- kitchen equipment (10)
- knitting (73)
- Parenting (85)
- poems (4)

Figure 1.10 Four Obsessions' display of knitting book covers

14 Library Mashups

3. The Portable Library

Jesse Andrews is a self-proclaimed book lover and developer of Book Burro (www.bookburro.org), a client-side mashup. Book Burro is an extension for the Firefox and Flock web browsers. With Book Burro, any time you are surfing and encounter a page with book information, a small panel will overlay the page with options to click for more information, such as library holdings and online bookstore prices. Figure 1.11 shows the activation of the small panel and the search options.

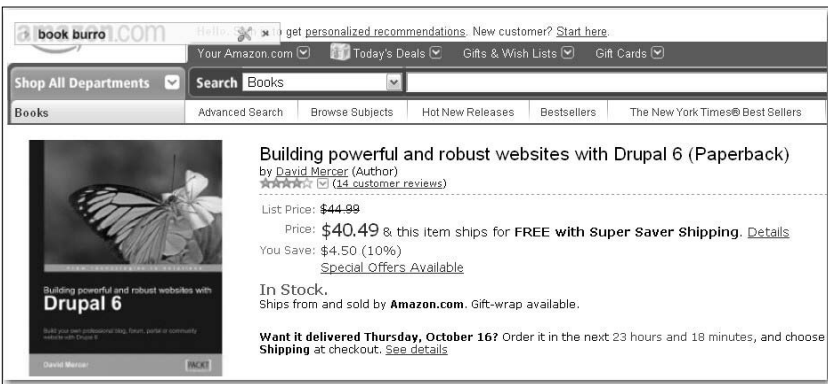


Figure 1.11 Pale yellow Book Burro overlay on Amazon book page

Library Mashups: The Good, the Bad, and the Ugly

Hopefully this brief tour of library-related mashups has given you a peek at the diversity, creativity, and utility of library mashups. Throughout the rest of this book, other library-related mashups will be explored and described in detail. By creating and remixing publicly available data, libraries can offer richer and more dynamic services to library patrons and empower them to develop some wonderful services.

Mashups are not perfect, however. There are some downsides and some caveats to creating and using mashups. Some of the first ones that pop to mind are privacy, rights management, reliability, and provenance.

First and foremost, when creating a mashup, respect copyright and rights management terms. Not all data sources available via an API or XML data stream can be remixed, and it's critical that copyright and the uses allowed by the copyright holder be respected.

Second, privacy is definitely a factor for any provider, particularly for libraries that are vigilant in protecting the privacy of our users and their borrowing habits. The issue of privacy extends beyond just one data provider, however. For anyone involved with releasing survey data and maintaining user confidentiality, the risk of disclosure of an individual's identity is not new. It is a bit of a new realm, however, for some of the web companies that are releasing data and finding themselves caught on the wrong foot. Two well-known examples of privacy violation are the AOL release search query data and Google Street View (maps.google.com/help/maps/streetview) sightings.¹

Let's look at an example of how mashing up from multiple data sources can have surprising and downright unpleasant consequences. Tom Owad was one of the first people to show some of the "Big Brother" effects of open data and mashups. He searched a common name on Amazon and captured a sample of 250,000 people with that name who had wish lists. (Wish lists can list your first and last name, city, and state, if you choose to do so.) Next, he winnowed the list down to people who requested certain "controversial" books, such as George Orwell's *1984*. Once he had that result set, he used the Yahoo! People Search (people.yahoo.com) and was able to find the exact street address and phone number for some of these people. With Google Maps, he was able to pull up a satellite view of their homes. He sent one of them a copy of *1984* as a gift in the mail and blogged about his discoveries. You can read about the steps he took in more detail on his site, Applefritter (www.applefritter.com/banned-books).

Before you leap to the conclusion that the internet has to return to a set of balkanized websites to preserve privacy, pause for a moment and consider how we might find a balance between being totally open and totally closed. Sharing data streams and combining them via mashups and web services is in its infancy online, but not so in the world of academic research, statistical agencies, and organizations that have for decades created public-use microdata files that are anonymized to allow reuse. A bit of thoughtful planning can minimize the risk of disclosure. The media coverage of the missteps by high-profile companies has helped sensitize us and raise awareness about disclosure risks with data providers.

Fortunately, libraries already have a high level of awareness of the risks of disclosure, and we simply need to apply what we know when opening up library data sources. For example, we may be able to randomly show books returned during the past 24 hours, but we probably

16 Library Mashups

don't want to show this information in real time because doing so would create the possibility of someone identifying an individual borrower by watching the library's circulation desk and website at the same time. The lesson to learn from the Amazon wish list example is the need to release library data carefully for remixing, keeping in mind that it will be combined with other data sources.

One major drawback of mashups is their long-term reliability. The mashup ecosystem is still only a few hundred days old; it's in its infancy. Although a company or organization may provide a data stream via an API today, will it continue to do so in 6 months—or 6 years? When you build a mashup on open data, in many cases you are relying on the kindness of strangers and the future health of the organization offering the service. You need to consider carefully who is offering the resource and that entity's capacity and commitment to supporting it over time. You also need to consider plan B—what if the data stream goes away? Is there another source you can use? The bottom line is: How much should you invest in building a discovery tool or mashup site? This will really depend on the source of the data, the benefits, and the resources involved. Another important factor is scale. If there is a cap on the number of queries per day, the mashup may work fine at the outset, but what if it really takes off? Some providers, such as Google Maps, offer a licensing arrangement for high-volume or commercial applications.

Last but not least, in terms of major considerations when thinking about creating or using a mashup, is the provenance and authority of the mashup resource. The casual viewer often cannot easily discern who provided what piece of data in a mashup, making it difficult to assess credibility and authority. Some mashup developers state the sources of their data streams; others do not. Even if they do state their source, have they manipulated the data—either deliberately, or inadvertently through a miscalculation?

Conclusion

Each month more organizations are opening up data sources, and the mashup ecosystem becomes richer. Mashup code libraries and hosted mashup sites targeting nonprogrammers, such as Microsoft Popfly (www.popfly.com) and Yahoo! Pipes (pipes.yahoo.com), add more features and services. Put on your creative thinking cap and start dreaming about new services and features that would delight,

entertain, inform, and promote libraries. Think about ways to allow your library users to remix library data. There are many exciting opportunities for libraries and users to create interesting mashups.

Endnote

1. In August 2006, AOL released search data logs for more than half a million anonymized users from a 3-month period for academic researchers to analyze. Although the logs were anonymized, the search queries themselves included private information. Within just a few days, there was an outcry about the privacy breach; AOL took down the data set and apologized. In 2007, Google launched its new service—Street View—with street-level photographs of North American cities in which you could clearly see faces of people caught on camera. Some of these included benign but unsettling photographs, such as a crisp photograph of Mary Kalin-Casey's cat sunning itself inside her living room. Other photographs were not so benign, including burglars breaking into apartments, nude sunbathers, and arrests.

