

Libraries and Cloud Computing

Matthew Goldner, MSLS, MAR
Product & Technology Advocate, OCLC

Andrew Pace, MSLS
Executive Director – Network Library Services, OCLC

Last February walking through O’Hare Airport, it was slightly surprising to see large banners hanging that advertised a cloud-computing solution. But it was *very* surprising to see signs in the concourse for yet more cloud-computing solutions. The term has become ubiquitous and yet there still remains confusion about exactly what cloud computing is and what it offers libraries. This prompted us to take a survey of different types of librarians as to their use of cloud-computing solutions. The purpose was to ascertain whether librarians have begun to seriously consider the changes cloud-computing solutions can bring to our community. We would argue that it is not the technology that is important, but rather the opportunities it affords us for efficiency, savings and cooperation.

For the purposes of this paper we use two widely quoted definitions of cloud computing. The Gartner Group defines cloud computing as “a style of computing in which massively scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies.”¹ In various presentations, KPMG breaks this into essentially four different types of cloud computing: infrastructure, platform, applications and services. To put this in more concrete terms, examples of each can be:

Table 1

Type	What it is	Examples
Infrastructure	Buying space / time on external servers	Amazon A3 Bungee
Platform	An existing software platform to build your own applications on	Facebook
Applications	Software applications accessed with a Web browser	Google Docs Salesforce.com
Services	Ready-to-use services accessed with a Web browser	ADP Mint.com

Table 1 illustrates why there are varying definitions of cloud computing. Many cloud services actually incorporate two or more of these types. For example, Google Docs provides infrastructure as well as applications. But what does this new style of computing mean for libraries?

As Chris Anderson, Editor of *Wired* magazine and the author of “The Long Tail,” has pointed out, “The Web is all about scale, finding ways to attract the most users for centralized resources, spreading those costs over larger and larger audiences as the technology gets more and more capable. It’s not about the cost of the equipment in the racks at the data center; it’s about what that equipment can do.”ⁱⁱⁱ This gets to the crux of cloud computing in the library environment. While it is true we should be able to recognize IT efficiencies and savings by not replicating infrastructures repeatedly across our libraries, the outcome of aggregating the library community and library data into a more cohesive whole in a cloud environment is what can change our future.

Let’s first look at what we have learned about librarians’ current use of cloud-computing services and then consider what types of uses should be considered and acted upon.

Results of cloud-computing survey

In March 2011 we conducted a survey sent to 2,700 librarians in academic, public, government and corporate libraries, and in various roles across the library. More than 300 librarians responded. We purposely did not put forth long definitions of cloud computing, wanting to get a feel of the level of knowledge about cloud services at the same time we were gathering data on uses of cloud services. The survey was structured around the types of cloud computing, asking questions on uses of cloud infrastructure, cloud applications and developing and deploying applications into a cloud environment. We received several direct e-mail responses as well as comments in several surveys that the person did not know what cloud computing was so he or she could not accurately respond to the survey. However compared to the number of responses, these were negligible.

The major results of the survey told us the following:

1. Only 5 percent have started to use cloud-based infrastructure such as Amazon EC2, or VMWare vCloud Express.ⁱⁱⁱ
2. 13 percent used cloud-based storage or back-up services such as Dropbox, Barracuda or MS Windows Azure. Of interest was the statement by several respondents that they considered having their ILS or institutional repository hosted by the supplier as cloud-based infrastructure. While this could technically be considered cloud-based, we will discuss later why this fails to take full advantage of the new opportunities that cloud solutions provide to libraries.
3. 55 percent were using cloud applications, with the leaders being Google Docs, Google Calendar and Gmail.
4. 66 percent were using cloud-based sharing applications, with the leaders being Facebook, Twitter, Flickr and YouTube.
5. 19 percent (57 individuals) responded that they had created a library-related application using a cloud-based service, but when asked where these applications had been deployed only 5 respondents stated that they had put an app in either the Apple Store or the Android Market. The remaining 52 cited either their Facebook page or an app placed on their library’s Web site. Since

the question did not delve deeper, it is not possible to know if these were actually applications put on the Facebook platform.

6. Those 70 percent who responded that they were using at least one type of cloud service were asked the main reasons for using these services. The breakout of the reasons from which they could choose is as follows:
 - a. Increased efficiencies: 21%
 - b. Opportunities for collaboration: 16%
 - c. Decreased need for in-house technical expertise: 9%
 - d. Cost savings: 7%
 - e. More timely access to latest IT functionality: 5%
7. All respondents were asked to indicate from the following list of concerns what their main concern for using cloud-based services was, and results broke out as follows:
 - a. Data security: 35%
 - b. Long-term stability of the service provider: 31%
 - c. Privacy: 13%
 - d. Data ownership: 7%

This was not intended as a scientific survey, but rather as a method to get the pulse of the library community in relation to cloud services. Given this disclaimer, some high-level observations are still warranted by the data.

First, the main use of cloud services by libraries is either taking advantage of freely available applications for internal use in the library or for social networking purposes within their own community. It did not reveal a move to use the cloud for building the larger library community into a force on the Web. Secondly, librarians have begun to consider the advantages of cloud computing for efficiency and collaboration, but the types of services being used imply this is only efficiency and collaboration within their own libraries, again, not taking advantage to reach out across the community to build system-wide efficiencies and collaboration. Finally, librarians are expressing the same concerns about cloud services that have been expressed in other industries, with data security and long-term stability of the service provider as the most widely expressed concerns.

What should we be looking at in relation to cloud-based services and what can we learn from the success of commercial ventures in fully exploiting the potential of these services? We would say that the library community can apply the concept of cloud computing to amplify the power of cooperation and to build a significant, unified presence on the Web. This new, unified presence has the potential to give libraries significant scale and impact on the Web in a manner similar to major search engines, online booksellers and social networking sites. Such improvements can be grouped into three basic areas: technology, data and community.

Technology

The first advantage libraries can get from cloud-computing services is to work with system suppliers to change the way systems are built and how they open opportunities for technology sharing.

The Web has opened up a way for building systems that can take real advantage of service-oriented architecture. Rather than applications isolated in their own silos, we need applications that are built to call shared services. By way of example, libraries around the world are maintaining knowledge bases of their electronic resources. But these data stores are used only by local applications that access and act upon the data directly. Why isn't this method replaced by having shared universal knowledge base data stores with services built to use that data across libraries and across multiple applications?

With this type of architecture the opportunity exists to take a shared infrastructure to a new level. Why did Facebook succeed and MySpace failed? One reason is Facebook created an open platform where anyone could create an application and deploy it. As these applications multiplied they attracted more and more users to the core services of Facebook.^{iv} As a result, Facebook aggregated enormous amounts of data that could be used in multiple ways, creating a circle of increasing value to anyone using the platform. This will be discussed further below.

But how do libraries truly take advantage of cloud computing? By learning from Facebook the value of an open platform that allows innovation by anyone and not just the staff of Facebook. A great deal of attention has been given to creating open-source library management systems in the last five years, but these systems simply replicate the status quo with a new business model for creating, distributing and maintaining the core software of a system. Any innovation that comes with these efforts must still be distributed through outdated technologies.

With cloud-based computing, work done in one place echoes through the rest of the system and is available to all.

Consider the Apple AppStore and the Android Market Place. These make it simple to share innovation—just click, download, use. And they accommodate varying motivations for innovation. People may create apps simply for the perceived social good they can cause, or they may do it because they believe they can monetize the apps. It is the platform the apps are shared on that makes it possible to succeed in sharing the innovation, no matter the original motivation of the developer.

So too, librarians need to be considering how we can better share innovation with each other. At the April OCLC Global Council meeting Robin Murray, Vice President Product Management, made the statement, "We have shared records, we have shared resources, it is time to share innovation." Murray went on to say:

"When Apple debuted the iPhone ... even when they opened the App Store ... they didn't envision anyone using their product as an astronomical device. But now there isn't just one, but a dozen or so iPhone apps you can use to chart and map the nighttime sky. Hold your phone up and the various sensors for position, time, facing direction and angle of tilt will show you the outline and names of the constellations you're 'seeing' through the phone's screen. It's incredibly cool, makes perfect sense, and couldn't have been imagined by the people who created the tools that make it possible. That illustrates one benefit of shared data and a cooperative platform. It encourages innovation and pools the creativity of many participants."

In 2002, Amazon was one of the first companies to open up its technology for other organizations to use and build services on. Today, through its cloud services, Amazon shares its content and e-commerce tools with software developers and website owners, allowing them to leverage Amazon data and functionality to power their own services. As a result, about 1 million businesses with more than 150,000 developers are building innovative applications by mixing their software applications with Amazon's.

This can be done in a scalable sustainable manner for libraries only when we work together to create platforms where innovative development ideas can be shared. It also requires the aggregation of accessible, reusable data stores.

Data

Currently libraries have most of their data stored in individual silos, making reuse of data difficult if not impossible. And unfortunately the often-held view about opening data for reuse is to simply harvest it to be placed in yet another silo. Multiple small silos of data, however useful, are far less likely to attract a large community of users. This is why dominant services grow for different classes of data. eBay is successful because it has the most data about items that individuals are auctioning off to the highest bidder. Therefore it attracts this most buyers for the sellers meaning more sellers will want to put their data into eBay about items they are selling. Add to this each buyer adding reviews and ratings of the sellers. Like Facebook, this causes a circle of increasing value for every individual using eBay.

We as librarians must learn from this success, so we are driven to explore the possibilities of aggregating our data for the benefit of both the librarian community, but more importantly for the community of library users and information seekers. This hearkens back to the need for a change in our technology from applications working directly with data to a service-oriented architecture with open, reusable services that work with data. This needs to extend beyond shared bibliographic data to include other types of library data such as knowledge base data, license data, vendor data, etc. It is this shift that can really start to take advantage of cloud-computing technology to make libraries more efficient and build large collaborative communities of librarians and library users.

Community

Libraries already form a community of shared needs and shared values. Historically it has often proved difficult or costly to collaborate in useful ways. An early leader in library collaboration was the inception of interlibrary loan. But before computers allowed automating the interlibrary loan process it was a labor intensive manual hit-and-miss task. However, as national libraries and large cooperatives began to aggregate the data of library holdings it became possible to create a "cloud service" before the Internet existed. Resource sharing became a possibility to even the smallest libraries. Technology made it possible but a shared need made it happen.

Now that there is a “cloud,” and the Internet and technology have advanced to allow anyone to take advantage of this cloud, where should we as librarians be looking to build more communities of common need or common values? There are already examples of how this can be done.

When Google approached large research libraries to digitize their collections, the agreements allowed the libraries to have a digital copy of each book or journal. But each library was then faced with the task of creating a repository and access service for these digital objects. Instead, several of the libraries came together to form the HathiTrust so they could have a single repository for all of these items and a single point of access to them. This is a real-world example of using cloud-based services to enable the technology, data aggregation and community of shared needs and values to create an incredibly useful service for library users.

Before this collaboration, the collections in each library existed in isolation. HathiTrust brought them together, pooling resources and eliminating redundancies, and producing a valuable research tool. The representation of these resources in digital form provides expanded opportunities for innovative use in research, teaching and learning, and enables each university to provide its students and scholars with access to one of the most significant digital collections ever assembled. HathiTrust also makes its content and records widely available to other clouds through data and bibliographic APIs and an OAI feed, maximizing the exposure of this valuable library content.

When cooperation takes place at the platform level, participants can reinvent workflows, dramatically increase the utility of shared data, reuse each other’s applications and combine any number of activities. These aspects of cloud computing hold the promise of more than reduced overhead through cost sharing. They can enable another computer revolution in how we do business.

Concluding questions

Given the results of the survey, looking at the commercial world’s use of cloud-based services and considering the nature of the library community as a community of shared needs and values some questions are warranted. Have we as librarians truly considered the enabling power of cloud computing to change our future? Areas to consider would include:

1. Sharing more types of data for increased efficiency, cooperative intelligence and cost savings
2. Sharing computing infrastructure to free up staff from managing IT infrastructure to working on needed public services and innovation
3. Building larger communities of collaboration for creating new pools of shared data, collection development/sharing, digitization and preservation decisions, and sharing innovative services
4. Creating a more unified Web presence for libraries to give information seekers a valuable alternative to generally used online resources.

We do have the opportunity to change what our future will be and a responsibility to see how cloud-based services can be employed to bring about that change.

Endnotes

ⁱ Gartner Group: <http://www.gartner.com/it/page.jsp?id=1035013>, September 23, 2010.

ⁱⁱ Anderson, Chris: http://www.wired.com/techbiz/it/magazine/16-03/ff_free?currentPage=all, April 13, 2011.

ⁱⁱⁱ The question on cloud-computing systems was defined as “vendor-provided software stack in place with default configurations that can be easily turned on for use by the library to deploy applications.”

^{iv} Babu, Ganesh : <http://www.axleration.com/facebook-vs-myspace/>, April 12, 2011.