



The Great Server in the Sky: Cloud Computing and Government Information

Amy West

Note: For this issue, the data column has moved to Tech Watch.

GPO's PURL server completely failed on August 24, 2009, and as of this writing has yet to be fully restored. The GPO PURL server resolves PURLs, i.e., persistent URLs assigned by GPO to publications from agencies across the federal government in order to manage the volatile nature of web addresses. Libraries depend on PURLs to get users from catalog records to government publications on the web. When the system works, it works pretty well; but when it fails, it fails badly, as librarians trying to get to government information via PURLs can attest. While it's the PURL server that has been having problems lately, there's nothing particularly unusual about GPO—similar failures could and probably will occur with government information resources at all levels. As GPO works to restore service, we and they should be considering advantages and disadvantages of the “cloud computing” technological infrastructure.

The hallmarks of cloud computing include:

- a shift away from local to remote management;
- more reliable and more powerful computing resulting from economies of scale; and
- an extension of the trust relationship to the new service providers.¹

Many methods of delivering cloud computing exist, but the most common include:

- fully outsourced technology infrastructure to an external company;
- virtualized software running across multiple servers to manage spikes in traffic and failure of individual servers; and
- distributed networks of storage using peer-to-peer sharing systems.²

In an e-mail to Govdoc-1, GPO staff described the PURL problem this way: “the PURL server suffered a significant hardware failure.”³ They further said that “many institutions have automated URL checkers that run against the PURL server. Please be aware that the PURL restoration process is severely slowed by checkers repeatedly hitting the PURL

server.” Restoring service has been complicated by old software that must be patched onto new hardware and limited agency resources overall. Given the claims listed above, what might some of the issues be if GPO employed cloud computing?

GPO could decide to simply outsource their technical infrastructure. By paying an external entity for technical support, GPO could benefit from economies of scale that would preclude lengthy downtime. Google and Amazon both sell cloud services, and in early September 2009 the General Services Administration (GSA) introduced Apps.Gov, which GSA describes as a provider of cost-effective cloud computing. Were GPO to outsource its technological infrastructure, accountability, privacy, and service guarantees would become paramount. There are plenty of examples of commercial services and government agencies experiencing data breaches, but best practices are only beginning to emerge.

Another approach, which in theory could be managed within GPO, would be to create a virtual version of the PURL server that could be run across however many physical pieces of hardware are needed and available at any given time in order to avoid the kind of failure GPO had in August. Virtualized servers should also remain unaffected by automated link-checking systems since the links aren't being checked on specific pieces of hardware. In this scenario, the key is that the server becomes the “server,” i.e., a virtual machine rather than a physical one. Such a scenario might or might not be more cost effective than GPO's current methods.

James R. Jacobs, on the Free Government Information blog, proposes a third scenario.⁴ Jacobs suggests that the existing network of FDLP libraries should become a peer-to-peer network in order to help manage spikes in interest in publications—say, a new health care bill is released to the public—or as alternative servers when one or more others fail. In this case, the cloud aspect comes not from virtualizing the idea of one whole database across multiple pieces of hardware, but by creating an inherently decentralized network of collections and services in which storage, while still specific to particular servers, is redundant and distributed. For the user, the result would ideally be the same—the publication desired is available when needed—but the structure behind the scenes is quite different. An added advantage to a peer-to-peer system like this is that it puts libraries, traditionally politically disinterested parties, in a position of acting on that impartiality for everyone's benefit.

The costs to GPO relative to existing technology costs would be unknown without a much more detailed plan. Costs for some participating libraries might increase as they'd be taking on new responsibilities, but other libraries may not incur extra costs since they may already have technical infrastructure in place capable of handling the new content.

The problem is simple: it is very common for government information to come to us via a limited number of paths. When those paths are blocked, everyone is affected. The most recent case has been with the GPO PURL server. It probably won't be the last. As GPO examines its options for improved technological infrastructure, we should all be thinking about cloud-based solutions and what they might mean for reliable access to government information. We should also consider whether we wish to remain exclusively clients in a client-server

relationship with government information providers, as would be the case in most cloud-based solutions, or whether we wish to become part of the cloud ourselves.

References

1. "Cloud computing," Wikipedia, en.wikipedia.org/wiki/Cloud_computing.
2. Mark Hopkins, "Can We Please Define Cloud Computing?" Mashable Blog, Aug. 19, 2008, mashable.com/2008/08/19/cloud-computing-defined.
3. John S. Dowgiallo, "PURL Server Update 2," e-mail to Govdoc-I, Sept. 1, 2009.
4. James R. Jacobs, "Critical GPO Systems and the FDLP Cloud," Free Government Information Blog, August 30, 2009, freegovinfo.info/node/2704.

Give to the Rozkuszka Scholarship

The W. David Rozkuszka Scholarship provides financial assistance to an individual who is currently working with government documents in a library and is trying to complete a master's degree in library science. This award, established in 1994, is named after W. David Rozkuszka, former documents librarian at Stanford University. The award winner receives \$3,000.

If you would like to assist in raising the amount of money in the endowment fund, please make your check out to ALA/GODORT. In the memo field please note: Rozkuszka Endowment.

Send your check to GODORT Treasurer: John Hernandez, Coordinator for Social Sciences, Northwestern University Library, 1970 Campus Drive, Evanston, IL 60208-2300.

More information about the scholarship and past recipients can be found on the GODORT Awards Committee wiki (wikis.ala.org/godort/index.php/awards).