



METAARCHIVE & CLOUD COMPUTING

Bill Robbins
System Administrator
MetaArchive Cooperative

Overview for the Library of Congress

MetaArchive in a Nutshell



- MetaArchive uses the free open source LOCKSS archiving software to operate a network of preservation servers.
- All content is stored in multiple copies on multiple caches at geographically dispersed locations.

MetaArchive Properties Server



- Common needs for the MetaArchive caches are hosted on what we call the Properties Server
- Along with the Public Web Site, there are other Central functions & tools

Technology Concerns for the MetaArchive Cooperative



- Three Year Technology Refresh
- Data Integrity Provided by LOCKSS Architecture
- Challenge: Network Bandwidth

Why Cloud Computing



□ **Unattached Central Servers**

The central properties server of the cooperative should belong to the whole, not be property of a single member. The Cooperative is a virtual entity. The cloud would be perfect to host the central/properties servers.

Cloud Service Provider - Amazon EC2

Findings After First Several Months



- Amazon EC2 Strengths
 - ▣ Consistent Build Outs
 - ▣ Predictable Costs
- Amazon EC2 Drawbacks
 - ▣ Storage Capability & Workability (Not a capacity issue)

Summary



- Cloud Computing Is Working well for MetaArchive for the central properties servers
- Vendors need to do better in offering a full set of tools for server operations

□ Contact Info: <http://www.metaarchive.org>

Bill Robbins

MetaArchive Cooperative

1230 Peachtree Street, Suite 1900

Atlanta, GA 30309

bill.robbins@metaarchive.org or katherine.skinner@metaarchive.org

Side Note

The Next Level for Open Source Software

- Amazon EC2 can be used as a way to bring Open Source projects to the next level.
- A public server image can be made available to run software developed for LC, other government agencies or non-profit institutions.
- There would be no software to download, and / or build and no additional server configuration needed.