

Digital Repository Infrastructure: Should you rent or buy?

Robin Dean, Repository Director, Colorado Alliance of Research Libraries

robin@coalliance.org

Ed Fugikawa, Repository Technical Lead, Colorado Alliance of Research Libraries

ed@coalliance.org

OR 2013 Conference Paper Proposal

In an environment of successful repository growth, financial challenges, and the obsolescence of our original hardware, the Alliance Digital Repository launched an evaluation of digital repository infrastructure options.

Background

Founded in 2007 and launched in 2008, the Alliance Digital Repository was designed and funded in a different financial and technical environment than the one that exists in 2013. When the service was first beginning, it made sense to purchase and maintain hardware for the digital repository.

As the repository grew in size and complexity, so did the hardware. In 2009, the infrastructure had expanded such that the Alliance's commercial office space could no longer meet its power, cooling, and connectivity requirements. An agreement with one of our member institutions to provide physical collocation for the hardware allowed the repository to continue to run on locally owned hardware.

By 2013, we were facing high maintenance/licensing costs as well as large one-time expenditures to purchase replacement hardware (which had to be faster and larger than the hardware it replaced). Meanwhile, advancements in virtualization, cloud-based hosting, and infrastructure-as-a-service were changing the digital repository infrastructure landscape. When it came time to refresh our infrastructure, we decided to take another look at fully hosted and virtualized solutions.

Infrastructure Options

We examined the following options:

1. Managing our own hardware, but with cost reductions
2. Hosting with a third-party commercial hosting solution with a nearby physical location
3. Hosting by a member university that could provide "infrastructure-as-a-service"

Another option we considered but eliminated due to cost and management concerns was hosting the repository entirely with a remote cloud provider such as Amazon Web Services.

Evaluation Criteria

We evaluated the infrastructure solutions for sustainability according to six criteria:

- Cost
- Reliability
- Scalability (Storage Capacity)
- Openness
- Performance
- Disaster Recovery

For each category, we used a rubric of criteria to assign one of three rankings:

- Not Sustainable
- Acceptable
- Sustainable

Once we had ranked the solutions in each of those categories, we could compare which solutions were more sustainable and determine whether any solutions failed to live up to the sustainability goals of the repository.

Our criteria were informed by discussions about risk and levels of service with our members, and by guidance from the CCSDS "Audit and Certification of Trustworthy Digital Repositories" Recommended Practice (September 2011) and DRAMBORA.

The complete hardware evaluation matrix and example evaluations will be included in the presentation materials.

Conclusions

At the time of this conference proposal, data is still being gathered about some of the infrastructure options, so a decision has not yet been made. We expect to make a decision according to the above criteria by June 2013 and would share the process, outcome, and "lessons learned" at OR 2013.