Using the Virtual-Private Cloud Model to Serve and Preserve Historical Collections: A Case Study (Based on Islandora)

Gail Truman
Truman Technologies, LLC
4096 Piedmont Avenue, Ste. 217
Oakland, CA 94611
+1510-502-6497
gail@trumantechnologies.com

Jaime Henderson
California Historical Society
678 Mission Street
San Francisco, CA 94105
+1415-357-1848 X214
jhenderson@calhist.org

ABSTRACT
This poster session describes the selection criteria and process used for evaluating three repository software offerings and cloud platforms, with pros and cons. It describes implementation of workflows, representations of PREMIS metadata for objects in the repository, documenting fixity checks performed on datastreams, mapping of "rights" elements in DC datastreams to PREMIS "rightsExtension" elements, and more.

General Terms
Infrastructure opportunities and challenges

Keywords
storage cloud, Islandora, digital repository, SOAR®, preservation[1]

1. INTRODUCTION
The California Historical Society sought to implement a digital asset management and repository system to help preserve and showcase two terabytes of digitized materials. Faced with aging on-premise servers and storage, the society decided to remove the financial and resource burden of technology migration and local IT staffing and move from capital expense to an operation expense model– one based on a virtual-private, secure cloud.

2. PHILOSOPHY APPLIED
When evaluating and recommending approaches to the long-term protection of digital assets, we apply the following "big rules" or philosophies

- *Keep it simple*: Digital repository systems should be easy to implement, understand, and support.
- *Don't overbuild*: If you try to anticipate every "what if" scenario, you will a) overspend, b) be late to deploy, c) probably not need everything that was included.
- *Lots of copies keep stuff safe*: Ideally store 3 copies of all data in 3 different geographies, stored 3 different ways.
- *Have an exit strategy*: Standards-based open software and SLAs provide for vendor liquidation or end of services.

3. REQUIREMENTS FROM CALIFORNIA HISTORICAL SOCIETY
The historical society had a high-level list of requirements coming into the projects that are shown below. After a detailed evaluation of the three options, additional and more granular, requirements were identified. These requirements were given a weight for their importance and then given a score.

- Bulk ingest
- Fixity check
- Metadata standards
- Cloud based
- Stable URLs
- Bit preservation
- Exit strategy
- Rights and restrictions
- Public interface
- Exposed to Google
- Offers staff training
- Includes support
- Open source
- Cost
- Others are using it
- User friendly viewers for books, manuscripts, photos, and other formats
- Streaming
- OCR
- Online exhibitions
- Restrictions/embargos
- Analytics/stats
- Multi-lingual

4. OUR SOLUTION
Truman Technologies, LLC (TTL) utilizes Islandora software and the open-source SOAR® (Scalable Online Archive & Repository)* stack to recommend the best possible digital repository solution for organizations weighing their digital options. For the California Historical Society, TTL recommended the Islandora DAM repository software as offered and serviced by Discovery Garden Inc. (DGI), hosted by the secure private cloud (Infrastructure as a Service) vendor KomodoCloud. TTL also recommended that another copy of all digital assets (historical collections) be hosted at DuraCloud.

iPres 2015 conference proceedings will be made available under a Creative Commons license.
With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license.
Authorship of this work must be attributed. View a copy of this license.

5. BENEFITS OF ISLANDORA

Aside from meeting California Historical Society's functional requirements, there were several preservation features that were drove the decision to move forward with Islandora:

- Preservation File Formats and Normalization: a single workflow can ingest a preservation/archival file and convert it into a customized preferred format as well as an access format.
- Versioning[2]: a record of how digital content has changed allows clients to preserve the data as well as the look and feel of a digital object, since its original dissemination mechanism and all subsequent changes are stored and linked to the original content.
- Interoperability, Reusability, and Bitstream/ Object Replication[3]: all data from a digital object are stored in a format that can be exported to future systems or shared between applications.
- Fixity, File Format Identification, and Data Integrity: Islandora FITS, Checksum, and Checksum Checker modules add functionality to the Islandora solution pack by adding technical metadata extraction, enabling checksum generation, and adding a PREMIS "fixity check" entry to an object's audit log[4].
- Preservation Metadata: the PREMIS module[5] produces XML and HTML representations of PREMIS metadata for objects in a repository; its current features include documenting all fixity checks performed on datastreams, including "agent" entries for a given institution and for the Fedora Commons software, and mapping contents of each object's "rights" elements in DC datastreams to equivalent PREMIS "rightsExtension" elements.
- Community Interest Groups: Islandora has an active Preservation Interest Group[6] that develops and recommends common approaches to preservation within the Islandora suite.

---

[1] SOAR (Scalable Online Archive & Repository) is a trademark of Truman Technologies, LLC and is registered in the US Patent and Trademark Office.

[2] Davis, D. (2011, August 13). Versioning- Fedora 3.4 Documentation.

[3] Access and view the BagIt module, documentation, and forum at https://github.com/islandora/islandora_bagit

[4] The checksum and checksum checker modules are community developed, and not yet supported by DGI, but future versions of Islandora will include these features. Access the checksum and checksum checker modules at https://github.com/Islandora/islandora_checksum and https://github.com/mjordan/islandora_checksum_checker

[5] Access the PREMIS module, documentation, and forum at https://github.com/islandora/islandora_premis

[6] Read the Preservation Interest Group's manifesto at https://docs.google.com/document/d/1o6GeNbFQsR5L_4BBe1UtwbORcRnW0ii0SXoGdEDfR98/edit#