

Digital Library Platform Working Group Final Report

20 Feb 2012

Charge

The newly-appointed Digital Platform Working Group of the Digital Initiatives and Services Committee will:

1. Continue the comparative study of Islandora and SobekCM begun by the earlier subcommittee.
2. Evaluate the relative strengths and weaknesses of the two systems for implementation by FCLA and its successor organization. The Subcommittee is encouraged to draw upon individuals throughout the SUS, at FCLA and beyond, as needed, for expertise and insight.
3. Prepare a report to be delivered to CSUL members by February 15, 2012. The report will discuss the pros and cons of the two systems under consideration, and recommend how either or both might be used to create an effective, efficient, and sustainable system.
4. Lead a discussion of the report at the March CSUL meeting.

Members: Emily Gore, FSU, Chair
Lee Dotson, UCF
Lois Widmer, UF

Digital Library Landscape

In the initial report submitted by the DISC, a list of digital library systems contrasted and compared before the systems were narrowed to Sobek and Islandora. The functional requirements of those systems were evaluated as thoroughly as possible in the time allowed, but the evaluation was never completed. This Digital Library Platform Working Group found no problems with the report of the prior work of the DISC subgroup, but wanted to take a broader look at the items we feel that should be evaluation points when selecting a system, a landscape survey, in addition to looking at the functional requirements addressed in the initial report.

System functionality can always be enhanced, and in comparison at any given point of time, one system may fit the needs of an organization better than others. However, broader issues like developer support, user community, and even the underlying architecture are fundamental and not as easily changed. In the table below, we have outlined some of these important issues and the current status for both Sobek and Islandora. For this information we relied largely on responses from the developers of Sobek and Islandora, not unlike you would in an RFP.

	Sobek	Islandora
Support	Currently developed/supported	Open Source - user driven

	<p>by UF Programmer Mark Sullivan and the Digital Library Center at UF. Released as Open Source. Github repository and released in SourceForge. Assistance from UF provided for free to SUS's. FCLA has begun exploring the code and is actively working on a port of the database to PostgreSQL and associated code changes.</p>	<p>and supported; Drupal and Fedora; multiple listservs; Google groups for users and developers; Github repository and vendor available for fee for assistance with Discovery Garden. DGI currently has 20 full-time staff and 6 regular contract developers (defined as someone who works at 30% or greater of a FT position). Our current hiring plan has us growing to 26 full-time and 12 contractors by the end of this year . The average size of a Digital Asset Management company, according to the industry literature, is 20-40.</p> <p>In addition the core Islandora team at UPEI has 7 full-time staff.</p>
<p>User Community Installations vs. Contributors</p>	<p>Installations = 5 (UFDC/dLOC /NewspaperCat, Wolfsonian, FCLA testing, One in Vietnam being worked on by a grad student and professor, One in Austin, Texas collection parameters unknown)</p> <p>Contributors = 100+ users through the University of Florida Digital Collections, collaborative Digital Library of the Caribbean, Wolfsonian, and other smaller projects.</p>	<p>100+ installations including CARL consortium in Colorado and UCLA, collections of considerable scale that would be comparable to our collections; over 100 more users slated to come online in the near future.</p> <p>Active Islandora, Fedora and Drupal listservs/community of users. Research articles/case studies/presentations on this stack solution.</p>
<p>Upgrade/Feature Requests</p>	<p>direct request to Mark Sullivan; priority ranking yet to be determined.</p>	<p>can contribute code to Github and have code/features reviewed then possibly added to release. Work contracted with Discovery Garden for hire is usually folded into the next</p>

		release of the open source product.
Implementation and Migration	Florida work would be led by FCLA and Mark Sullivan. SUS day 1 implementation list already developed by DISC and FCLA. Windows server/C# knowledge is required. Open source and all contributions will initially be reviewed by UF/FCLA until non-profit organization is created to assist with steering of development. Could migrate from Windows directory structure to Fedora Repository backend with sufficient interest. Runs on windows .NET.	Open Source products from thier inception - Linux server based, and FCLA has extensive Linux knowledge; Drupal knowledge also exists at FCLA; Fedora written in Java; Priorities and data model will need to be established. Data migration from FCLA's Digitool instance may be possible with Discovery Garden outsourcing in "one-click" fashion or could be performed in house. Migrating University of Florida Digital Collections (UFDC) content from SobekCM to Islandora to be determined, if there is interest.
Training and Documentation	Documentation written by Mark Sullivan, developer, and available online at http://ufdc.ufl.edu/sobekcm/ . Patron Help and Technical Help for SobekCM are available online. The Contact Us form goes to a small group of DLC staff, who address the technical questions. Training: selected UF DLC staff provide training, to date at no fee.	Extensive documentation provided on the Islandora site updated with each release available online at https://wiki.duraspace.org/display/ISLANDORA/Islandora Previous documentation remains archived on the site. Islandora Camp and Red Island Repository Institute held by UPEI each year. Discovery Garden training services available for a fee if needed.
Funding and Corporate Partnerships	DLOC partnership. UF supported.	In addition to full time staff listed above, Islandora/UPEI has current grant funding - ongoing \$2.3 million funding through 2014 from the Atlantic Innovation Fund (AIF) from the Atlantic Canada Opportunities Agency (ACOA). Partnerships with DuraSpace, Arrow Electronics, Oracle, Truman Technologies

System Features

The committee members visited UF's Digital Library Center to learn more about the development and use of Sobek, an open source system, at UF. Programmer Mark Sullivan provided the group with an overview of the system, and UF DL staff members discussed how Sobek is used in their digital library workflow. Sobek is written in C#, approved as a standard by ECMA (ECMA-334) and ISO (ISO/IEC 23270:2006), and runs in a Windows server environment. Sobek uses SOLR to power its search and operates under a GNU general public use license.

The group also discussed Islandora with Mark Leggot, UPEI. Mark demonstrated the new features of Islandora release 11.3 and demonstrated sites the group had not seen, including the CARL consortium in Colorado. Their transition to Islandora is just coming online. They have 15 members, ranging in size and collection scale, and all but one who will be transitioning to their new shared Islandora repository. (The one noted exception is choosing to stay with DigiTool.) The content in this repository is varied and includes scholarly works typically found in an IR as well as other image and text-based collections.

Islandora is an open source framework developed by the University of Prince Edward Island's Robertson Library. Islandora combines the Drupal and Fedora open software applications to create a digital asset management system that is capable of powering traditional digital collections for libraries and museums as well as atypical collections such as the DuraSpace website and the Scalable Online Archive and Repository (SOAR). The Fedora and Drupal user communities are well established with the [Fedora Commons Registry](#) listing 300 known installations and a [July 2010 website analysis](#) estimating the number of live Drupal sites at 7.19 million. Fedora is RDF-based, which prepares our content for the semantic web/linked data future. Additional open source applications can be added to this core stack to create "Solution Packs" or "Sprouts" which can address institutional or content specific needs. An example of a solution pack is the institutional repository feature. Islandora operates under a GNU general public use license.

Since the initial review of the systems last summer, both SobekCM and Islandora have continued development and have added new features to their systems. While the functional requirements document proved useful for narrowing the options from a field of many to the top two, further devotion to minor details is not relevant as a decision making tool for choosing between two open source solutions with continual improvements.

Recommendations

The majority of the members of this task force believe that the most versatile and forward-looking solution is the one with the most robust architecture, supported by the largest number of developers and the largest user community, running on the most widely available open source platform. For these reasons, we feel that there is no question that Islandora is the better option for the SUS.

In terms of the implementation, the group is presenting two options for CSUL to consider:

- 1) Adopt Islandora/Fedora immediately and begin developing an implementation plan.

Advantages: We go live without anticipated transition to another system in the foreseeable future. Modification of code should be minimal since this system is designed to work for consortium. Native Linux and Drupal environments, which FCLA and other SUS members have expertise in. Open source communities of Fedora, Islandora and Drupal can provide feedback and assistance and provides an environment of multiple developers/code contributors. Fedora is RDF-based, which is the language of the semantic web and the basis of “library linked data.” From UCLA Libraries on their recent selection of Islandora: “One of the principal attractions of the Islandora framework is that it builds on the strengths of robust open source solutions with large user bases and tested track records.” Fedora is no doubt a scalable solution. Many tests have been performed regarding scalability, and there is no doubt it scales to handle a consortium of our size. Multiple Fedora instances can be searched across as well if necessary.

Considerations: This transition will take at least 6 months to be completed according to estimates supplied by Mark Leggott, UPEI/Islandora. The transition, if sought to be completed in 6 months, will require that we hire Discovery Garden to help with implementation at an estimated 30 days at \$1200/day = \$36,000. (This is roughly the amount of the current DigiTool yearly maintenance.) Additionally, Fedora will require group decisions on data models to be made before we get too far into the implementation. This will require consensus among the potential users.

2) Continue with the Sobek implementation. If this is the decision made, the majority of the work group recommends that this approach be taken only with an eye towards Fedora/Islandora/Sobek integration over the next several years, in the belief that the RDF-based Fedora repository is the common underlying platform we would like to see drive our common digital library platform. The third member of the work group diverges somewhat from this view, recommending instead a stable four to five year commitment to the SobekCM platform, allowing time for completion of the grant-funded development of Islandora as well as a likely merger or shakeout among the other systems. In this scenario, SobekCM development should continue, preferably with an eye toward convergence with a leading system, likely Fedora-based.

Advantages: Implementation time may be short since this work has already begun. Mark Sullivan is able to dedicate time to help FCLA with the work that needs to be done. Since FCLA has begun implementing Sobek, a list of “Day 1” must haves has been developed based on features and material display options currently available to DigiTool users that need to be applied in the Sobek environment. There are tools in place like the Sobek METS editor that are in wide use and allow for creating packages that can then be FTP’d to the Florida Digital Archive.

Considerations: Overall time and effort will be greater to implement Sobek first and then to implement a Fedora-based system within the next few years. This may be considered a waste of FCLA’s and SUS’ resources and there may be resistance among the users to migrate their collections multiple times. Sobek is written in C# and hosted on Windows servers - there is no expertise at FCLA, so they are having to learn as they go. They are most familiar with Linux environments and open source tools. This system has a very small user community and essentially one developer which troubles the majority of this platform working group.

