**Institution Repository Research Project**

**11-Dec-2018**

**Client:** Dean of library Amy Stewart-Mailhiot.

**Team Members:**

Raj Bhetwal: Team Leader, Client POC, Data structure

Augustine Maddox: Software engineer

Theodore Ryan – Project manager, POC for vendor, Report/Presentation developer

Reginal Grant: Software engineer, AWS site coordinator

**Time line:**

12-Sept-2018: Raj contacts Amy Stewart-Mailhiot to coordinate project

20-Sept-2018: Team meeting with Amy Stewart-Mailhiot(client) for project set-up meeting

25-Oct-2018: Mid Term presentation to class

1–Nov-2018: Update Client meeting . Result of meeting client wanted team to just focus on Islandora

1. – Dec 2018: Final client meeting with up dated information on Islandora.

**Result of final brief: Client happy with what we had accomplished, and information provided**.

**Teams goals:**

1. Research Multiple freeware software systems that are used to develop and run Intellectual Repositories.
2. Reduce possible choices down to two systems we were conduct closer examination of systems
3. Get vendor possible general quotes while we downloaded a reduced version of the freeware systems to be able give a fair assessment of issues and challenges the client will face while creating the IR system

The project was to research and test software systems to recommend to the Saint Martin’s Library to us to create an Intellectual Repository (IR). This placed requirement to look not just at Isloandora and Dspace by our client, but we also identify other similar systems and compare them give the best overall recommendation that fits our clients goal. The main requirements we were looking at was system that could be up front easy for user to use to located saved data, file, or any other file typed saved in repository, looking at freeware to reduce costs, the requirement including programing languages to create and maintain, and any costs of having vendor to set-up and possibly run. This led to the goals of what are group wanted to achieve.

The original goals of the group helped to focus our groups efforts. First goal was research multiple freeware systems used to create Intellectual Repository for libraries. Second goal when down to two systems we were conduct closer examination of systems. This included downloading and running a sandbox version of each. This would lead to our need to update our client with a a briefing to the client. After this briefing the third goal was to get vendor possible general quotes while we downloaded a reduced version of the freeware systems to be able give a fair assessment of issues and challenges the client will face while creating the IR system. This lead us break down of the group to start our goals

The group was started by researching multiple IR freeware systems looking at their initial good qualities and issues. This is where we elimated E-Print as it has issues will large upload of data having to be reduced to smaller files to be uploaded. VuFind and Fendora was elimated as Islandora used all the same software languages but provided additional resources they did not. The other system looked at was Invenio that was on version v3. Closer look showed that they had just released this version few years ago requiring the older version having to be totally changed over as this was not very compatible with older versions of Invenio. This left with two original top runners of Islandora and Dspace. Both systems have similar capabilities with very helpful establish communities along with running sandboxes that we could use to test in further stages. The biggest downside during this stage was in ability for most software not to provide sandbox environment. VuFind dose provide a demonstration but his is limited in its uses. With agreement our group set to split to focus on Islandora and Dspace.

Raj Bhetwal and Ahmed Alharbi to focus on Dspace. DSPACE is a set of cooperating Java web applications and utility programs that maintain an asset store and an associated metadata store. The web applications provide interfaces for administration, deposit, ingest, search, and access. The asset store is maintained on a file system or similar storage system. The metadata, including access and configuration information, is stored in a relational database. Operating System **is w**ritten in Java, tested under Linux, Windows, and Mac OSX. DSPACE is a full stack web application, consisting of a database, storage manager and front-end web interface. Originally designed for libraries, the embedded DSPACE data model and approval workflows are familiar to librarians and archivists. Dpace had a lot of good qualities.

Some of the top qualities of Dpace were unlimited file types, uses a Qualified Dublin Core (QDC) based metadata schema, comes with various tools, and built in security. Unlimited files types for DSPACE can store any type of file. In addition, it auto-recognizes files of most common formats (e.g., DOC, PDF, XLS, PPT, JPEG, MPEG, TIFF). Dspace uses QDC allowing institutions the ability to extend that base schema or add custom QDC-like schemas. DSPACE can import or export metadata from other major metadata schemas such as MARC or MODS. Dspace comes with a suite of tools (batch ingest, batch export, batch metadata editing, etc.) and plugins for translating content into Dspace objects. Additionally, commercial plugins are available through service providers. Security forDspace provides its own built-in authentication / authorization system but can also integrate with existing authentication systems such as LDAP or Shibboleth.

Islandora was researched by Augustine Maddox and Theodore Ryan. Islandora is a digital repository system, meaning it is designed to both store and organize information. One of its strongest quality is the different languages used to create and operate it. It uses Fedora Commons is designed to have digital repositories such as ISLANDORA built upon it. Its purpose it to instill a modular design into any software based on it. It is designed to allow for a digital repository to easily communicate information from one module to another. DRUPAL is a content-management system, used as a web application framework for a lot of websites worldwide, and is written in PHP. APACHE SOLR is a search engine written in Java, based on the Apache Lucene project. The features it provided was presented to our client that had effect on the focus of the project.

The features that mad Islandora so appealing was its original design, support for file types, extensive documentation on implementation, sandboxes availability, multi-language, and support for XML metadata. , It was designed initially by a college library system, the Robertson Library of the University of Prince Canada. Support for any file type with use of Fedora and multiple lange functioialit with use of Drupal. It has a modular framework for defining specific data types, including standard audio, PDF, images, paged content, videos, and web archives. There is documentation for learning how to implement ISLANDORA available from the ISLANDORA homepage. With the goal of creating a system that is both easy to implement and simple to learn, this fits our needs very well with readily available sandbox for testing and learning. Support for any XML metadata standard, including unique schemas. The database has data-entry/editing forms for any XML schema, support for the creation of relationships between objects, A flexible faceted search (Solr), and Automation of the transformation of assets.

Theodore Ryan reached out to multiple users that were already using both systems and received reply back from Pacific Lutheran University that uses Dpace and Oregon State library that uses Islandora. Using this information along with what the group had so far on 1-Nov-2018 the team gave update presentation to the Saint Martin’s University Dean of the Library Amy Stewart-Mailhiot(Client). Showing what we have learned and access ability to each sandbox along with reviews from other users and her own personal experience she optioned for the team to focus on Islandora. This require the team to refocus and set new goals for itself. Goal one was to get working version of Islandora on AWS account to gather first hand experience setting up from scratch a Islandora system. Research data systems that could be presented that is usable for the IR. Then last was contact vendors that provide set-up or services for Islandora to get general quotes without committing the University.

The reaching to finding vendors was rocky especially when trying to general quote of cost of set-up and running system. Three vendors focused on were Agile Humanites, Discovery Garden, and Born Digital that were orignally located on Islandora service companies page. Agile Humanites responded for set-up and installation on a client server with base cost for setup and installation of Islandora 7 along with standard, community supported solution packs on a client-maintained server is $3600 + HST. They followed with running Islandora for university general amount of annual services agreement for Islandora. This Includes the equivalent of 40 developer hours (at an agency rate of $150/h), with 24 allocated to general maintenance, security updates, and backups, and 16 to minor feature development at cost about 40h @ $150/h = $6000 + HST per year. The next is Discovery Garden that responded with list questions that was filled out to the best our ability without getting into specifics. There was no response from Discovery Garden after that. The last was Born Digital where they attempted to set up a phone meeting. This fell thru due to mis communication and conflicting work scheadules.

The Raj Bhetwal researched data capability and structure use for Islandora. Everything in Islandora's repository is three types of objects. These different objects are a data object, collection object, and content model object. These objects are made up of datastreams. These datastreams are the parts of an Object, e.g.  a binary image file, a thumbnail, a metadata stream (such as Dublin Core metadata). These Objects are set-up to relational database with having relationships with one another. This information is stored in a special datastream (RELS-EXT) that describes relationships. This is created by a primary key set-up with Persistent Identifiers (PIDs) that are unique in the repository for each object. This is done by implementation of a graph data structure which allows for flexible creation of rich digital resources of disparate types. The system that has been set up to help maintain this and operation is Islandora are the solution packets.

Solution packets that bundle tools supporting specific types of content and knowledge domains. This elegant solution packages the correct standards support, content models, and related tools (such as viewers) for a given content type into modules that can be associated with a collection. This elegant solution packages the correct standards support, content models, and related tools (such as viewers) for a given content type into modules that can be associated with a collection. There are multiple solution packets like Standard Image Solution Pack that enables support for JPEG, PNG and GIF images, Book Solution Pack that Enables creation of book collections and the ingesting of pages, and Audio Solution Pack. That enables support for WAV and MP3 files. Other solution packets are Large Image Solution Pack, Newspaper Solution Pack, Paged Content Solution Pack, PDF Solution Pack, Video Solution Pack, Binary/File Solution Pack, Entity Solution Pack, Digital Humanities (XML) Solution Pack, and Scholar Solution Pack.

There are three deployment options for Islandora that can be utilized. The 1st is service company provides you with a hosted instance of Islandora scaled to your requirements deployed on either your own or a hosted infrastructure. The second is Islandora OnDemand cloud instance (SaaS), which is quick but has some limitations on how much customization can be performed. The last is using the university IT resources with supporting infrastructure and skills they can obtain Islandora from [github](https://github.com/islandora) and configure an instance for the university. The first and second options come with maintenance provisions so your instance is always up-to-date with the latest releases of Islandora (usu. with new functionality).The last option requires your IT resources to maintain your instance. The third option deployment you can choose to handle as much of the setup as the organization has in resources. Service providers like mentioned earlier are available to consult, help with design, configuration, migration, and training.

The last stage is setting up a working Islandora to show to the client. The group coordinated with Professor Nelson and the Engineering school department to secure an AWS site to set up servers for downloading and setting up a working Islandora working system. First Augustine set-up a sandbox/slim version of Islandora on his personal laptop. He first set-up Oracle VM virtual box where then downloaded Ubuntu Linux operation system. This is where he downloaded a sandbox version of Islandora.

During this time Reginal “Reggie” Grant helped to coordinate and set-up a site on Amazon Web Services (AWS). AWS provides Amazon Elastic Compute Cloud (EC2) that offers wide range of operating systems. This what lead to basic ideal to us Ubuntu a Linux operating system. This gives us the flexibility to users to design their operating systems to best suit their computing needs. Other options are available for modification, including the number of virtual CPUs, as well as Instance Storage capacity. This also provided us with AWS instructing the user to configure the firewall called the security group. This allows to set of firewalls rules that controls the traffic for you instance. This gave the group the ability to access ports 22, 1194,943, and 443 so that Augustine could establish connection t outside network. Augustine coordinated with Reggie to get access to the groups AWS site.

Here Augustine began to download from Github the containers and documentation to set up an actual Islandora system. Here we ran into issues where software update and lost of AWS site due to miscommunications. Augustin started setting up Islandora where he ran in considerable challenges as a lot of the systems were showing failures. Half way thru trouble shooting Islandora updated downloads in Github. Augustine used the new downloads was able to get some of the systems starting t work. This is where he ran into secondary problem near the end before our final presentation. The Amazon Web Services site we had secured and been locked and deactivated. When research into this issue by Reggie , he discover that credit card used to secure site was about to be charged. This was caused when the servers running on AWS site were not put into sleep mode causing them to run over on allowed utilization time a month.

A final meeting was set up with Dean of library Amy Stewart-Mailhiot(client) to give her final brief of what he had achieved and be able to present all our information. During this brief she communicated with Reggie on desire what she would like to do moving forward with him if he chooses to continue this project. Here we demonstrated the slim down version of Islandora on Augustine laptop and covered our short comings in getting multiple feedbacks from vendors and issues when attempting to download full version of Islandora on AWS EC2. The group provided the client with a copy of final brief with links to sources we used. Asked the client if she was happy with our results and she said yes.

The next steps for this project to continue would find out what issues to fix to finish the download of Islandora in AWS. Then come up with to plans to continue with one being locate a server for use with 1.4 TB of memory space to set-up intuitional supported Islandora for the IR. Second source would be to price cost of cloud out source compare to using on university sources to house system and gather more in-depth cost estimate from vendors to set-up and run Islandora system for university. Collect the costs for both and provide t client to develop budget request to pass up to the university to see which it will fund.

Resources:

<http://www.eprints.org/uk/>

<https://vufind.org/vufind/>

<https://invenio-software.org/>

<https://www.slideshare.net/anjesh/digital-library-repository-invenio-vs-dspace>

<http://wseas.us/e-library/conferences/2010/Faro/DNCOCO/DNCOCO-16.pdf>

<https://wiki.duraspace.org/display/ISLANDORA/Getting+Started+with+Islandora>

<https://wiki.duraspace.org/display/ISLANDORA/Installation>

<https://www.digitalocean.com/community/tutorials/how-to-create-a-self-signed-ssl-certificate-for-apache-in-ubuntu-18-04>

<https://webcache.googleusercontent.com/search?q=cache:-BqbAv45giQJ:https://islandora.mnpals.net/pals/islandora/object/PALSrepository%253A345/datastream/OBJ/download/2013_Islandora_Basic_Training_Manual.pdf+&cd=9&hl=en&ct=clnk&gl=us>

[https://aws.amazon.com/ec2/?sc\_channel=PS&sc\_campaign=acquisition\_US&sc\_publisher=bing&sc\_medium=ACQ-P%7CPS-BI%7CBrand%7CDesktop%7CSU%7CCompute%7CEC2%7CUS%7CEN%7CText&sc\_content=ec2\_p&sc\_detail=ec2&sc\_category=Compute&sc\_segment={creative}&sc\_matchtype=p&sc\_country=US&s\_kwcid=AL!4422!10!71949349564215!71949660862469&ef\_id=W4BI7QAAAI3QQW3M:20181209075713:s](https://aws.amazon.com/ec2/?sc_channel=PS&sc_campaign=acquisition_US&sc_publisher=bing&sc_medium=ACQ-P%7CPS-BI%7CBrand%7CDesktop%7CSU%7CCompute%7CEC2%7CUS%7CEN%7CText&sc_content=ec2_p&sc_detail=ec2&sc_category=Compute&sc_segment=%7bcreative%7d&sc_matchtype=p&sc_country=US&s_kwcid=AL!4422!10!71949349564215!71949660862469&ef_id=W4BI7QAAAI3QQW3M:20181209075713:s)

<https://aws.amazon.com/>

<https://duraspace.org/dspace/>