

# July 2017 Community Sprint Report

In Github

## Participants

- Trey Pendragon (Princeton University)
- Esmé Cowles (Princeton University)
- Justin Coyne (Stanford University)
- Noah Botimer (University of Michigan)
- Stuart Kenny (Trinity College)
- Michael Klein (Northwestern University)
- Brendan Quinn (Northwestern University)
- Chris Syversen (Northwestern University)
- Carrick Rogers (Northwestern University)

## Overview

Our first community sprint featured nine developers from five Samvera Partner Institutions coming together to spend a week working toward the creation of a minimum viable product. The current goal of the Valkyrie working group remains to demonstrate a MVP at SamveraConnect 2017.

As the initial sprint for the working group the goals for this sprint were quite broad due to multiple areas of the code base needing attention. The goals included:

- Improvement of documentation, specifically YARDoc, and testing
- Creation of an ActiveFedora Storage Adapter
- Proof of concept for file characterization
- Derivative creation support for multiple image types
- Standardized around use of UTC for all `DateTimes` within the codebase
- Fixity Checking within Valkyrie
- Creation of a File Characterization Proof of Concept

## Results

The working group was able to accomplish the majority of the tasks set out during the sprint planning. It was actually a pleasant surprise how few direct conflicts there were despite having so many devs operating on the same code base and across multiple time zones. Minimal time was spent dealing with merge conflicts and other issues compared to other Samvera Community sprints of similar size.

### Improvement of Documentation and Testing

Test coverage and Rubocop code linting reached 100% this sprint, keeping Valkyrie aligned with general community standards for TDD and code cleanliness. The majority of the codebase was also YARDoced during this sprint which will provide future benefit in regard to on boarding developers. The current README documentation regarding spinning up a development instance and running tests locally proved to be sufficient with no developers reporting issues regarding creating a local development instance and running the test suite.

### Creation of An ActiveFedora Storage Adapter

One of the key goals is demonstrate the ability to use a variety of options for metadata and file storage. This sprint, we added an adapter that uses Fedora for storing files, bringing the options supported to:

- \* Files:
- \* Disk
- \* Fedora
- \* Memory (for fast tests)
- \* Metadata:
- \* Fedora
- \* Memory (for fast tests)
- \* PostgreSQL
- \* Solr

### Support for Multiple Image Types

Previously Valkyrie had only supported the `image/tiff` mimetype, which while common for archival purposes is not sufficient of it itself. Work was done to expand support for the uploading of JPEGs, GIFs, PNGs, and Bitmaps. While this is not intended to be the final list of supported image types, the addition of four additional types (and support for generating derivatives based off those types) provides proof that Valkyrie is extensible with regard to file format.

## **Standardization around UTC for DateTime Storage**

One of the issues from using multiple storage locations (Fedora, Solr, Postgres, etc) is that different tools may format their DateTimes differently. The current Samvera standard is to store in ISO 8601 using the UTC timezone. This is originated within the community due to this being the standard for DateTime storage. Valkyrie now stores in ISO 8601 UTC Timezone across all storage locations and when queried for all DateTimes return as ISO 8601.

## **Fixity Checking**

Valkyrie now stores SHA256 hashes of all files uploaded to it upon ingest of those files. These hashes will provide the future foundation upon which to perform fixity checking for all content stored by Valkyrie. This allows fixity checking if it is not native to the specific tool being used to store the file.

## **File Characterization**

File Characterization remains an important task for all Samvera applications. The current established pattern is to use the FITS tool, which in turn wraps other tools. However FITS is no longer maintained and it may not meet the Samvera community needs going forward. This sprint saw Valkyrie implement a generic service pattern for using a file characterization tool and a specific proof of concept service using Apache Tika. In the future services using other tools such as MediaInfo, GraphicsMagick, etc can be implemented to meet the community needs. This pattern may also be of use to other Samvera community members who are looking for a replacement for FITS.

## **Areas for Improvement/Retrospective**

Developers of the working group did encounter some pain points with the new patterns. While the new development patterns used in Valkyrie were generally well received and did successfully address pain points found in older patterns, there was still a learning curve for developers taking part in the sprint.

To alleviate this the working group has resolved to produce additional documentation on the patterns prior to the next sprint. Additionally a subset of the working group will create tests for features to be written in the next sprint. These tests will provide a specification for developers participating in the next sprint to code against.

## **Moving Forward**

The working group will be conducting a further review of this sprint during its next weekly meeting and then begin grooming the backlog for another sprint. The next spring will likely focus heavily on interaction with Fedora, either directly via the Fedora API or via ActiveFedora. Additional research is required to determine which path the next sprint will focus on.