DIGITAL LIBRARIES LABORATORY

DEPARTMENT OF COMPUTER SCIENCE

MSc Proposal

STUDENT: .... signed on ....

Stefano Rivera <stefano@rivera.co.za>

SUPERVISOR: .... signed on ....

Hussein Suleman <hussein@cs.uct.ac.za>

UNIVERSITY OF CAPE TOWN

August 7, 2009

Digital Asset Management in a Content Management System

MSc Proposal

Stefano Rivera <`stefano.rivera@uct.ac.za`>

Supervisor: Hussein Suleman <`hussein@cs.uct.ac.za`>

August 7, 2009

1 Introduction

Groups with collections of digital objects of public interest, such as papers or photographs, often store them in Digital Library systems. Such systems take care of the common needs of these repositories: browsing and searching interfaces, and help with curation and preservation.

There are a few different popular digital library systems in common use, but they tend to share the same disadvantages of being monolithic, and difficult to use. More recently, there has been a trend towards more modular systems, where the front-end and store (Digital Asset Management System, or DAM) are separated via well-defined APIs. An example is FEDORA [10], which provides all the back-end functions for a digital library system but has no user interface. The API separtion has allowed multiple front ends to be developed for Fedora, such as FEZ [6].

The growth of the Web in recent years has seen the rise of the Web Content Management System (WCMS, or in this document, simply CMS). CMSs make it trivial for even non-technical users to deploy dynamic Web sites, without having to be experts in Web-design or re-inventing the wheel for every project. They provide a common framework for tasks such as user management, authentication, on-line editing, work-flow management and search. Many blogs and news websites are powered by CMSs and they are growing in popularity in almost every other part of the web.

As building websites in general is made easier by CMSs, most institutions and archives would have their own website powered by a CMS. It follows, then, that their digital library should be integrated into their CMS. This would provide a consistent user interface to visitors and give the archive maintainer the ability to design the digital library with the same tools as the website.

Currently, on-line digital libraries tend to be deployed as two separate systems:

- A CMS-backed (or static) website with institutional information and a link into the digital library system.

- The digital library system, completely independent of the website. With a different look and feel, and rarely customised beyond the basics, as most Web designers are not able to re-skin arbitrary web applications.

This proposal is for integration between CMSs and DLs. In particular, a plugin for a CMS that provides a fronted to a repository stored in Fedora.

The aim is to graft the usability and presentation customisability of a CMS onto an archive that provides well-managed, archival storage and standard digital library protocol interfaces. This will allow for easier management via a system that Web designers are already familiar with, and thus a better user experience. This is valuable, as digital repositories are currently difficult to deploy and customise, and even more difficult to integrate with a related website.

2 Related Work

Historically, Digital Library systems have tended to be large, monolithic systems. A good, modern, Open Source example of this is DSPACE [9]. DSpace implements submission, metadata management, storage, search, and a web interface. While internally, it is

1

modular, it presents itself as a monolithic whole, and is known to be difficult to extend.

Recently, a new front-end has been developed for DSpace, MANAKIN [7]. It allows for re-skinning of the archive (albeit in a Manakin-specific way, rather than a well-known CMS's template system), and provides template-components for specific content types.

There have been internal and research projects to build content management systems for digital libraries, a custom CMS has usually been built *around* the digital library, rather than taking a proven widely-applicable CMS and linking it fundamentally with a digital library. Specifically, front ends have been written for Fedora, such as FEZ [6] and ELATED [1].

Given Fedora's compartmentalised design, different use cases have been dealt with by different projects. The FEDORA LIBRARIAN INTERFACE [2] was created to facilitate easy curation and ingestion into the archive. It is a desktop, GUI application; applicable, and probably beneficial, to a full or part-time curator, but not the casual repository user.

More recently, the ISLANDORA project [8] has started to create a Fedora-integration plug-in for DRUPAL [4], a popular CMS. Islandora is still in early stages and has yet to publish any results, however examination of the code-base, and failed attempts to deploy it for evaluation, have led the author to believe that Islandora is not straightforward to implement. This may improve later, but for now means that Islandora is not a suitable base to build this research upon.

There has been some research on the subject of evaluating digital libraries in [5] and [3]. Blandford et al. [3] emphasise the importance of the user experience in digital library systems. Fuhr et al. [5] develop a comprehensive framework for evaluating complete digital library systems (as part of the DELOS initiative).

2.1 Content Management Systems

Content Management Systems have been slowly replacing both the traditional, static website and custom dynamic websites. They provide all the basic building blocks to build a website, and let the user get on with designing the look and feel and filling out the content.

While there is no definitive study of the advantages of using a CMS for a website, there

are related works, such as [11], showing the case for the CMS.

3 Research Questions

3.1 Can Content Management Systems and Digital Libraries be seamlessly integrated?

Content management systems are designed to provide easy to use and user-administrable Web sites. Technically, content management systems are not specifically designed for administering a digital library, but rather textual content, closer to Web pages. Similarly, DAMs are designed to provide archival-quality structured storage for digital objects, first and foremost, with the user interfaces and customisation neglected. In the case of Fedora, there is no user-interface whatsoever, so the question is whether Drupal can fulfil that role or not.

The CMS is likely to have a far shorter lifespan then the content being archived (and its accompanying DAM), so the archive should be laid out sensibly, according to the DAM's best practices.

The objective of combining the two systems is to confer the advantages of both systems onto the whole.

However, while all the major CMSs have plug-in mechanisms for adding functionality, plug-ins rarely replace the data store. It is not expected to be trivial to implement such a fundamental change, as it impacts on a lot of the CMS's infrastructure.

Such a deep intrusion into the core of the CMS could conflict with other plug-ins and negate the benefit of using a CMS in the first place. Bringing the worst of both systems together would not be a positive outcome.

This question is clearly limited in scope to the systems used and a successful outcome is partially a sign of favourable design in the CMS used.

It is understood that Fedora has a very broad design and this system will not be able to expose all its functionality. While this will limit the applicability of this system, it is expected that the simpler, reduced scope system will be more usable within its scope. "Easily usable" and "generic, widely applicable" are likely to be opposing objectives.

3.2 Does this cater to the needs of End Users?

The CMS would need to provide front ends to much of the functionality of the digital library, at the very least:

- browsing

- retrieval

- ingestion

- search

These must fit in with the design of the CMS from both technical and usability perspectives. From the user's point of view, the CMS should present a consistent whole, not something that feels like a system within a system.

Can the CMS confer its usability upon the digital repository? The CMS is designed for usability — does handling digital repository objects within the CMS feel natural to a CMS user?

3.3 Would this be easily deployable by System Administrators?

Such a system is of little use if it cannot be easily deployed. One of the main advantages of a content management system is the reduced setup time and complication, compared to a custom system.

Maintenance of the CMS can often involve re-installing it or its plug-ins, so configuration is not a once-off consideration. Administering the digital repository should not be substantially different to administering anything else within the CMS.

4 Methodology

4.1 Implementation

The first task will be to implement a Fedora plug-in for Drupal, similar to Islandora. The design should be as simple as possible, but allowing for, at the very least, browsing, searching and ingesting documents such as would be found within a research group's institutional repository. See Figure 1 for an architecture overview.

The choice of Drupal is mostly based on the author's prior experience in developing plug-ins for Drupal, but there are other suitable
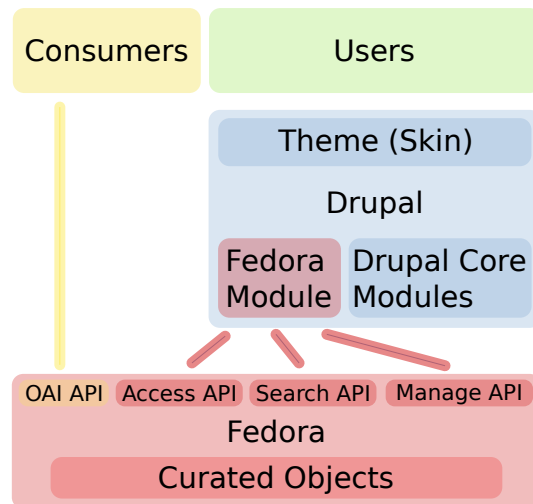

Figure 1: Proposed Architecture

CMSs that could be used in its place, such as Plone.

There should not be anything specific to this front-end within the asset management system. If such data is *necessary*, it should be kept to a minimum, and be removable (i.e. it should not form the core of the repository).

4.2 Evaluation

User experiments and evaluation will be required.

4.2.1 System Deployment

A study, in which Drupal users, and administrators familiar with Drupal, install the module should give an indication of ease of deployment. They should also be able to judge how well the system fits into the CMS and how well the system fits the digital library role.

It is unlikely that we will be able to evaluate more than a couple of Drupal administrators, and it is unlikely that they will have any experience of Digital Repositories. However, they will have experience in dealing with Drupal modules.

Similarly, Digital Library software users, who have deployed other Digital Library solutions, can give an indication of whether adding a CMS to the mix provides any ease in deployment, compared to the other Fedora front-ends or competing digital library systems.

4.2.2 User usability

A user study, based on tasks such as browsing, ingestion and search will give an indication of the quality of the front end. An important question is how integrated the repository components are with the rest of the site.

Most university post-graduate students (the likely pool for the survey) have experience in using on-line digital repositories and should know what to expect in such a system.

4.2.3 Case Study

A live deployment of the system, with useful data from one of the digital repositories managed by the department could provide the basis for a case study.

Given the special skills required by the subjects in an evaluation with expert users, it is unlikely that many subjects will be studied. A case study should provide data to cover areas that the user evaluations are not capable of.

5 Anticipated Outcomes

The system should be of interest to any group with a CMS based website that wishes to deploy a content management system for its digital repository.

It should be able to form a reasonable document repository (research papers and technical reports, for example), but will not necessarily support many other use cases out of the box.

The experiments should help prove the case for such integration, or show that Drupal is not currently a good choice for presenting a digital repository.

The system may not be appropriate for large, heavily used digital libraries, as it is limited by the ability of Drupal and Fedora to scale to such load.

6 Work Detail

6.1 Risks

6.1.1 Unsuitability of Drupal for the task

There is a risk that Fedora store simply cannot be integrated into Drupal in any meaningful way. Drupal's architecture expects to have all objects within the CMS be stored within its database, as NODES. That is not the planned approach for this integration, as the objects in the repository should be completely within the DAM, not the CMS. While there are plug-ins that does not go the node-per-object route, they tend to integrate poorly with the rest of a Drupal site.

The success of the integration can, unfortunately, only be determined some way into the development.

6.2 Timeline

There is a deliberate window for 3 months in this schedule, to allow for unexpected events. Additionally, at some point, any notable results will be written up into a conference paper.

6.2.1 Preparation

Duration: 4 months

Background reading in the field of Digital Libraries and investigating research possibilities.

6.2.2 Developmental

Duration: 6 months

The development of the project is largely about getting a demonstrable system up and running. The order of development should roughly follow this guide:

1. Prototype: Provide a way to view, say, PDF files stored in Fedora within Drupal.

2. Ingestion: Allow objects to be submitted via Drupal.

3. Browsing: Present the content of the repository to Drupal as a browsable system.

4. Search: Integrate a search engine in the Fedora repository with Drupal's search interface.

6.2.3 Experimental

Duration: 2 months

6.2.4 Write-Up

Duration: 3 months

6.3 Resources Required

This is not a resource intensive project, and only requires development time.

The personal computer equipment within the research lab should be all that is required to develop the software component of this research. All the software used is Open Source, and any products of this will likewise be released under an Open Source licence.

6.3.1 Experiments

For the evaluations: subjects, a couple of computers, and a suitable workspace will be required. The physical needs can be met by the experiment room within the department. The computers available within the research lab should also be sufficient (a desktop for the subject, and possibly a server or two).

Research subjects may be more difficult to obtain, as a few experienced administrators and Drupal users will be needed. It is hoped that the author can persuade some of the Drupal users acquaintances to participate in the experiments.

6.4 Deliverables

One deliverable of this project is a plug-in for a CMS to store objects within a DAM.

More generally, the aim is to produce some experimental data to prove that this is a good approach for the deployment of digital repositories.

References

[1] ACHANANUPARP, P., AND ALLEN, R. Developing a student-friendly repository for teaching principles of repository management. In *DigCCurr: International Symposium in Digital Curation, Chapel Hill, NC, US* (2007).

[2] BAINBRIDGE, D., AND WITTEN, I. H. A fedora librarian interface. In *JCDL '08: Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (New York, NY, USA, 2008), ACM, pp. 407–416.

[3] BLANDFORD, A., BUCHANAN, G., AND JONES, M. Usability of digital libraries. *International Journal on Digital Libraries 4*, 2 (October 2004), 69–70.

[4] BUYTAERT, D. Drupal, an open source content management platform. http://drupal.org/, May 2009.

[5] FUHR, N., TSAKONAS, G., AALBERG, T., AGOSTI, M., HANSEN, P., KAPIDAKIS, S., KLAS, C.-P., KOVÁCS, L., LANDONI, M., MICSI, A., PAPATHEODOROU, C., PETERS, C., AND SØLVBERG, I. Evaluation of digital libraries. *International Journal on Digital Libraries 8*, 1 (November 2007), 21–38.

[6] KORTEKAAS, C. Don't keep it under your hat: A Fez design description and case study. In *Fedora Users Conference* (June 2006), pp. 1–21.

[7] PHILLIPS, S., GREEN, C., MASLOV, A., MIKEAL, A., AND LEGGETT, J. Manakin: A new face for DSpace. *D-Lib Magazine 13*, 11/12 (December 2007).

[8] POUND, P. Islandora Fedora-Drupal module. http://www.fedora-commons.org/confluence/display/ISLANDORA/Islandra, May 2009.

[9] SMITH, M., BARTON, M., BASS, M., BRANSCHOFSKY, M., MCCLELLAN, G., STUVE, D., TANSLEY, R., AND WALKER, J. H. DSpace: An open source dynamic digital repository. *D-Lib Magazine 9*, 1 (January 2003).

[10] STAPLES, T., WAYLAND, R., AND PAYETTE, S. The Fedora project: An open-source digital object repository management system. *D-Lib Magazine 9*, 4 (April 2003).

[11] TALLOJU, M. Usability of web content management systems. Tech. rep., Kansas State University, 2007.