Introducing Vireo: an ETD Submittal and Management System for DSpace

Adam Mikeal, Scott Phillips, John Leggett and Mark McFarland

Texas A&M University Libraries
{adam, scott, leggett}@library.tamu.edu

The University of Texas Libraries
m.mcfarland@austin.utexas.edu

Introduction

The Texas Digital Library (TDL) is a consortium of public and private institutions from across the state of Texas. Founded in 2005, TDL exists to provide a common digital infrastructure for the state, and to promote the scholarly activities of its member schools. TDL currently offers a suite of services to its members, each of which plays a role in creating an online scholarly community for the state. These services—scholarly blog hosting, wiki infrastructure for collaboration between research groups, online peer-reviewed journals and workflow management, and digital repositories—provide increased visibility for the member schools and their scholarly output, and seek to leverage the economies of scale inherent in collaborative partnerships.

One of TDL's earliest services was a federated collection of electronic theses and dissertations (ETDs) from several of its member schools. Beginning with contributions from Texas A&M University and the University of Texas at Austin, the collection has been steadily growing in volume and participants. The university systems of four of the contributing schools in the ETD Repository project comprise more than 40 campuses, nearly 400,000 students, and 130,000 faculty and staff; The University of Texas alone processes nearly 1,400 ETDs every semester. The scale of this collection and the associated challenges of ingestion and management quickly became evident.

Meeting these challenges led to the formation of a state-wide ETD repository for managing the entire life-cycle of ETDs. The Texas ETD Repository is a large effort that spans multiple independent initiatives, all of which interact to support the overall task of managing ETDs in Texas. Metadata, identity management, repository interfaces, submission workflows, and data preservation each constitute significant projects in their own right, and these responsibilities are delegated to various groups within TDL.

This presentation will describe *Vireo*, the customized submission and workflow management application that TDL developed for DSpace, and it's role within the Texas ETD Repository. We will describe its current implementation as a Manakin aspect and theme, and discuss the future plans for the application, including its release to the repository community under an open source license.

Implementation

Vireo is built on *Manakin*, the new XML-based interface framework that first shipped with DSpace 1.5 in April 2008. Built on the Apache Cocoon platform, Manakin provides the ability to modify the look-and-feel of the repository at the community, collection, or item level, as well as a clean, modular way to introduce new functionality into the repository (Phillips 2007). Manakin uses two primary mechanisms to enable customization of the repository: *themes* that customize the look-and-feel of the repository interface; and *aspects* that act as a plug-in architecture for introducing new functionality or behavior to the repository.

Constructed as a paired set of Manakin themes and aspects, Vireo is essentially a customized window into a set of DSpace collections inside the repository. The aspects add the customized submission process used by the students, and the functionality needed to implement the iterative review workflow used by the university staff members. The themes apply a highly customized look-and-feel that provides a rich set of "Web 2.0"-style interface features. In this way, Vireo extends the default feature set of DSpace to create a complete solution for ETD management: point of submission, approval workflow, and publication.

Vireo replies on several other technologies or standards developed as part of the larger ETD Repository project:

- **TDL MODS Profile**. TDL developed a metadata standard, expressed as a MODS-XML profile to ensure that metadata for ETDs was captured and stored in a consistent manner across its participants (Surratt 2006, Rushing 2008). Vireo uses this document as the authoritative standard for its internal data storage, and follows its recommendations for conversion into standard Dublin Core or ETD-MS when necessary. When a Vireo-managed item is moved into a **Published** state, a MODS XML file is written into the DSpace item as a new bitstream.

- **Identity Management**. TDL selected Shibboleth to handle distributed identity management, and Vireo leverages Shibboleth's ability to transmit authorized user attributes between systems. This provides university-validated information (such as name or department), while reducing the potential for error caused by manual
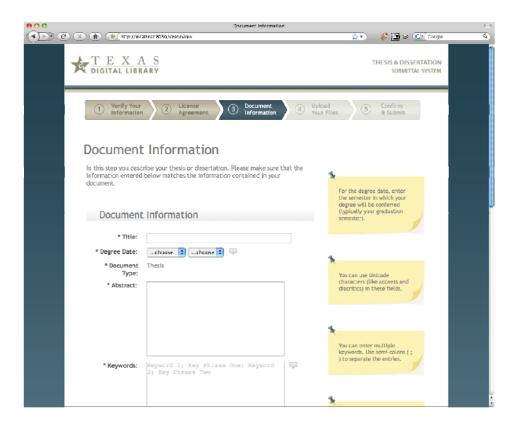

Figure 1. Student submittal interface.

duplication of data.

Interfaces

Because Vireo interacts with two different audiences—students who submit manuscripts, and administrative staff that review and approve those manuscripts—it required the construction of two unique interfaces. In many ways, Vireo is two applications that share the same underlying data, each modifying that data in different ways, according to differing rules (Mikeal 2008).

In almost all cases, the student submittal interface will only be used by a particular individual once. The student submittal application is a *novice interface*, as each user approaches the interface fresh, and remains so throughout the short time he interacts with it, not staying in the interface long enough to develop expertise (Shneiderman 1997). This interface design was approached using the multiple step "wizard" paradigm familiar to users of modern graphical operating systems such as Windows or OS X (van Weile 2000). A progress bar dominates the top of the screen, signifying the current stage of the five-part submittal process (see Figure 1).

The administrative workflow interface epitomizes the *expert interface*. Used by only a handful of staff at a given institution, many of these users will interact with the interface for several hours each day, as it will be integral to their job. It performs a complex set of tasks, and seeks to optimize for staff time and efficiency over user-friendliness. This interface was approached using the "job queue" paradigm. The initial screen shows an unfiltered list of all submissions in
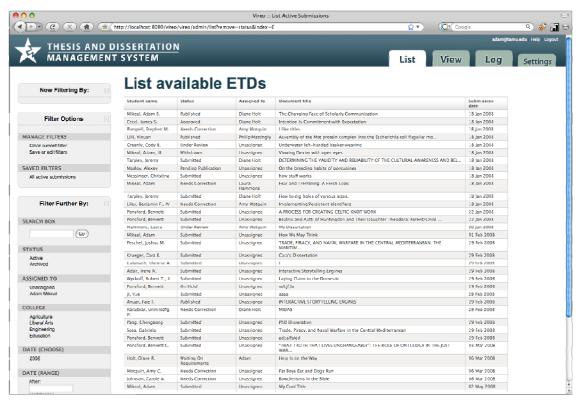

Figure 2. Administrative staff interface.

the system, sorted by submission date (see Figure 2). A set of *filters* is available to the left of the list, providing a faceted browsing experience, allowing the staff member to create customized queries. Staff users are granted significant control over their environment, with customization options available to individual users and to managers on a school-by-school basis.

Deployment

Since the ETD submission and approval process is a mission-critical service for university graduate schools, the introduction of a new information system into existing processes must be handled carefully. TDL has implemented a phased rollout to its member schools, starting with a small demonstrator at each school, and gradually increasing its use over a period of several semesters. Texas A&M University became the first school to complete this testing phase and move Vireo into full production mode. The University of Texas at Austin is currently within its phased testing period, and Texas Tech University is scheduled to begin in spring 2009. New schools will continue to be added in this staggered manner, ensuring that any emergent scalability concerns can be managed as they occur.

Conclusion

The Texas ETD Repository is a large project with many interconnected pieces. Vireo is a critical component in this project, allowing the data ingested into the repository to use a standard, consistent set of metadata and providing sophisticated workflow tools for the administrative staff in the graduate schools. Its development directly into the DSpace application stack allows for a complete solution to the ETD management problem, from submission to publication.

The Texas Digital Library has been an active participant in the open source community since its inception and remains committed to the philosophical perspective of the free software movement. In that spirit, once documentation and testing is completed for the initial release cycle, Vireo will be released under an open source license, including all generic documentation and training materials produced. Under the current timeline, this places the open source release in summer of 2009.

Acknowledgements

References

Mikeal, Adam, Tim Brace, Scott Phillips, John Leggett and Mark McFarland. "Developing a Common Submission System for ETDs in the Texas Digital Library". In *Proceedings of the 10th International Symposium on Electronic Theses and Dissertations*, Uppsala, Sweden. June 13-16, 2007.

Phillips, Scott, Cody Green, Alexey Maslov, Adam Mikeal and John Leggett. "Manakin: A New Face for DSpace". *D-Lib Magazine*, Vol. 13 No. 11, November 2007.

Rushing, Amy. "Texas Digital Library Descriptive Metadata Guidelines for Electronic Theses and Dissertations, Version 1.0". Prepared for and published by the Texas Digital Library, May 2008.

Shneiderman, Ben. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishing Company, 1997.

Surratt, Brian. "MODS Meets Manakin: Innovations in the Texas Digital Library's Thesis and Dissertation Collection". In *Proceedings of the 9th International Symposium on Electronic Theses and Dissertations*. June 7—10, 2006. Quebec City, Canada.

van Welie, Martijn, H. Trætteberg. "Interaction Patterns in User Interfaces". In *Seventh Pattern Languages of Programs Conference*, 13-16 August 2000, Illinois, USA.