

## **Avoiding the Calf-Path: Digital Preservation Readiness for Growing Collections and Distributed Preservation Networks**

» Martin Halbert, Woodruff Library, Emory University

» Gail McMillan, Digital Library and Archives, Virginia Tech

ETD repositories often start with very idiosyncratic and ad-hoc beginning data storage structures, driven by exigencies associated with creating an effective electronic workflow for accepting and securely storing digital copies of theses and dissertations as either a replacement or supplement to parallel workflows for print copies. ETD repositories also tend to grow in an effectively unbounded manner over time. Much like the story of the wobbling Calf-Path in the poem by Sam Walter Foss, these early idiosyncrasies and unbounded growth can subsequently cause enormous problems in systematic efforts to digitally preserve content of growing collections. The most effective preservation strategies incorporate pre-coordinated replication of content in distributed and secure locations; such replication strategies become increasingly difficult when the content is stored using irregular practices in directory structures, metadata, and file naming conventions.

This paper will address “Calf-Path” problems by providing practical guidelines, suggestions, and recommendations for ETD repositories. These recommendations are informed by five years of experience in operating the MetaArchive Cooperative, a distributed digital preservation cooperative of cultural memory organizations which has grappled with standardizing transfer mechanisms and developed cost/effective strategies for distributed preservation of ETDs based on the LOCKSS open source software. In the course of the past six years the members of the MetaArchive Cooperative have identified a series of best practices for digital preservation readiness.

These best practices can benefit start-up programs which have not yet established regular procedures and standards for directory structures, metadata, and file naming conventions. This paper will document relatively simple principles and guidelines for such programs that can greatly improve the subsequent likelihood of implementing successful distributed digital preservation programs.

## **Framing Digital Assessts into Context: A Preservation Study in the Design of a Cryogenic Pressure Vessel Using “STEP Documentation”**

» Samrat Sarovar, Graduate Student, Mechanical Engineering, West Virginia University

» Victor Mucino, Professor, Mechanical and Aerospace Engineering, West Virginia University

The process of development of a product involves innumerable activities and transactions, spread across a wide array of fields, departments, people etc. Right from the requirements of the customer, conceptualization of the design to manufacturing, there is an exchange of information punctuating all the activities and transactions. One of the main problems the industry faces is the documentation of this information so that it can be used as and when required.

In the long run, most of the information is either not retrievable, or is worthless due to proprietary issues, unavailability of programs used to create it, and other common situations. Even in the scenario that all the information is available and re-usable, there is always the missing link, which might be due to the absence of a particular designer, or in most cases, the statement of purpose, and the context of the design.

This study addresses the issue of Long Term Retention of data, the documentation procedures used during the design process,

and the development of context for an engineering problem. The context of an engineering design is the crux of the design process, as, it is the context that defines the problem, the different sets of solutions to that particular problem, and all the decisions taken in the life cycle of the product.

Also the context unifies each decision and each detail documented in the design. Thus, Loss of context renders all the information about the product useless. The design of a Cryogenic Pressure Vessel is used as a case study to understand the process of design, the flow of information, and the role of context in the design of a product.

A cryogenic pressure vessel that is designed to be used as a case study is designed to store and supply liquid nitrogen for a superconductor application. This is complex engineering problem, as the process of design and manufacture is a plethora of activities and transactions between people from different walks of life.

Using this design process as a reference to create Engineering Scenarios, we identify the flow of information through the various activities and transactions involved in the development of the product. These Engineering Scenarios also give us an idea on how the information is documented at each stage.

In the broader aspect, this project will help us identify a methodology for archiving information about engineering techniques and experiments, and be fundamental in providing guidelines to better preservation of complex engineering data. This will facilitate an environment conducive to easier and faster research on any given topic relevant to the end user.

This study is also part of wider digital preservation and archiving efforts lead by the U.S. National Archives and Records Administration (NARA). The common mission is to ensure the long-term retention and usefulness of digital data as well as the complex relationships and contextual metadata among digital assets. West Virginia University (WVU) is among a growing number of partner institutions working with the National Archives to provide comprehensive and integrated archival standards, methodologies, systems and solutions to guarantee the preservation of the digital information that becomes part of the historic national record.

Further, the WVU Libraries have offered the use of the institutional repository, “wvuScholar”, as a collaborative test bed environment for future usability studies in metadata subject and contextual analysis as well as for preservation and archiving using the application of the proposed methodologies of this study.

## **BREAKOUT SESSION 2B**

### **COLLABORATIONS: INTER/INTRA DEPARTMENTAL COOPERATION**

» Moderator: Suzie Allard, University of Tennessee

» Thursday, June 11 11:15 a.m.–12:45 p.m.

## **Moving from a Decentralized to Centralized Model for Processing ETDs**

» Angela McCutcheon, Director of Thesis and Dissertation Services, Graduate College, Ohio University

Ohio University in Athens, Ohio, used a decentralized graduate college model for 25 years. Moving towards a centralized ETD processing model required coordinating with the library, graduate council, several graduate studies offices, and the new Graduate College. Ohio University spent two years centralizing its operations.

This included: one Web site location for all forms and information regarding how to file a thesis or dissertation; one set of TAD format/layout guidelines; one location for documents to be submitted; one Word template document that can be used by all students to write a thesis or dissertation; Word training for formatting manuscripts in any writing style (APA, MLA, etc.); Thesis and Dissertation (TAD) Process Workshop for all thesis and dissertation students; a more streamlined method of Library cataloging; and future opportunity to standardize all thesis and dissertation forms for all colleges.

In this presentation Angela will discuss the challenges, processes used to resolve conflict, and successes experienced in moving towards a centralization models for formatting manuscripts, streamlining ETD operations and procedures, and working closely with the Library to get document cataloged efficiently.

### **Going Digital: The Implementation of an Electronic Honors Thesis (eHT) at WVU and its impact**

- » Keith Garbutt, The Honors College, West Virginia University
- » Molly Simis, The Honors College, West Virginia University

In 2006 the Honors Program at West Virginia University (WVU) became The Honors College, as part of that change the formally optional Honors Thesis became a requirement for all students who wished to graduate as a University Honors Scholars. To facilitate the collection and storage of these undergraduate thesis WVU Libraries collaborated with the Honors College to provide an Electronic Honors Thesis Repository (eHT).

This presentation will discuss the issues and challenges associated with this implementation, particularly issues associated with undergraduates as users of the system and the effect of a moderator (The Honors Dean) who is not a professional librarian. In addition we will discuss the unintended, but positive, consequence of the move to a digital thesis, the creation by students in the Honors College of an externally funded undergraduate academic journal.

### **From Paper to Pixels: Rough Spots and Roadblocks on the Way to ETDs**

- » Max Read, Faculty of Graduate Studies, The University of British Columbia
- » Bronwen Sprout, UBC Library, The University of British Columbia

This paper will describe the process of developing an ETD program from pilot project to mandatory ETD submission, and will discuss issues and solutions from the perspective of both the Library and the Faculty of Graduate Studies.

In December 2005 the UBC Library and Faculty of Graduate Studies initiated a pilot project for ETD submission to the Library's new electronic repository. We will discuss what we did to prepare for the pilot project, issues that arose from it, and subsequent actions.

When ETD submission opened in November 2007 we immediately got 50% of theses submitted electronically, and that percentage has increased. Theses submitted on paper are scanned to PDF and handled the same way as ETDs. The most challenging issue was the question of archiving ETDs. We decided to discontinue the older technology of microfilm/microfiche/paper backups, and to put resources into developing a reliable system for electronic archiving. We will discuss the process that led to this decision.

Some unexpected issues arose; creative arts faculty and students did not want their theses online, and we had an increase in requests to withhold theses on the grounds that journals would not publish previously-online material. We will discuss how we resolved these issues, albeit temporarily, and what is being done to address them further.

Technical support is now available to students through a Library computer lab, and we will be proposing that electronic submission become mandatory. We will discuss the outcome of this proposal, and any procedures we develop to streamline the submission process.

## **BREAKOUT SESSION 2C**

### **OPEN ACCESS: THE FUTURE OF OPEN ACCESS**

- » Moderator: Julia Blixrud, Association of Research Libraries (ARL)
- » Thursday, June 11 11:15 a.m.–12:45 p.m.

### **Repository usage statistics – Can you count on them?**

- » Simon Bevan, Cranfield University, UK
- » Paul Needham, Cranfield University, UK

**Objective:** The objective of this paper is to describe work undertaken as part of the Publisher and Institutional Repository Usage Statistics (PIRUS) project to develop COUNTER compliant statistics for material contained in IRs. The initial aim of the project was just to consider journal articles but it became clear that this format limitation was an unnecessary constraint. ETDs were an obvious candidate as the need for reliable statistics pertaining to ETDs has never been greater.

**Methods:** The project was undertaken via a combination of desk research, survey and practical development testing. The study had some synergy with a parallel funded project entitled Embed (Embedding IR into the institutional research process) for which academic staff were already requesting data on usage statistics.

**Results:** A methodology for providing COUNTER compliant usage statistics to measure full-text usage in repositories has been developed. PIRUS/Embed synergy resulted in the development of a dashboard style interface to statistics. Usage statistics, stripped of robot usage have resulted in significantly lower figures but are now credible, compatible and consistent.

**Conclusions:** A bid for funding for the PIRUS2 has been submitted. This will expand on the work undertaken in PIRUS and will include a case study on the recently launched EThOS service.

It is hoped that comparisons with other services can be developed along with a consideration of functionality to aggregate statistics for different copies of the same items within different systems. Currently the COUNTER standard relates to full-text downloads.

Given that item usage may be a significant metric in terms of, for example, conversion rates (item views to full text views), one future consideration should be to consider the measurement of metadata views. Item views may also be important for researchers, institutions and funders.