

Establishing an Electronic Theses Repository Using Digital Commons

Yumi Ohira

Digital Initiatives Librarian/Assistant Professor

Dr. C.C. & Mable L. Criss Library

University of Nebraska - Omaha

Deborah L. White

Digital Resources & Initiatives Manager

Pittsburg State University

Abstract

Has your institution mandated an Institutional Repository for electronic theses? Do you feel intimidated with setting up an institutional repository? The authors will share our success and experience of working with the graduate school at two small universities in rural Kansas to establish an electronic theses program.

The repository serves as an Open Access solution for global dissemination. Both Pittsburg State University (PSU) and Fort Hays State University (FHSU) currently use CONTENTdm (CDM) as their primary digital repository. In 2015 both FHSU and PSU purchased and launched bepress Digital Commons (DC), a more robust repository. If you seek global discoverability, unlimited storage, efficient technical support, and the ability to share a wide range of file formats in one interface, then bepress Digital Commons (DC) is the most reliable platform.

The authors will share their experiences and challenges of adapting and implementing an IR at PSU and FHSU. Then the authors will compare and contrast the advantages and disadvantages of the Digital Commons and CONTENTdm. Finally, they will share the challenges associated with developing IR initiatives at their institutions which includes marketing, workflows, and collection development of ETD materials.

Introduction

Open Access institutional repositories (IRs) have dramatically changed the way that academic institutions around the world disseminate the intellectual research produced at their institutions. Ryan Crow describes the role of an IR as: "a digital archive of the intellectual product created by faculty, research staff, and students of an institution and accessible to end users both within and outside of the institution, with few if any barriers to access" (3). Both PSU and FHSU desired to share their intellectual product with a wider audience which is what led them to purchase bepress Digital Commons.

Institutional Repositories actively provide open access to a wide variety of scholarly materials that benefits the institution. Also, an IR functions as a vehicle to drive research communication across disciplines and around the world. While implementing IR initiatives,

academic institutions have faced many challenges or barriers. Debora Madsen and Jenny Oleen discuss the challenges which an IR faces as it matures in their 2013 article.

As an institutional repository (IR) matures it will face the challenge of how to scale up its operations to increase the amount and types of content archives. These challenges involve staffing, systems, workflows, and promotion. The desire to scale up the operation, expanding the number of faculty participants and content, was addressed as part of a library-wide reorganization that provided more staff working as a cross-departmental team. This staff expansion, in turn, created the need to redefine staff responsibilities, develop resources to manage workflows, and provide greater efficiencies. (Madsen & Oleen 1)

The authors identified several challenges at their institutions encouraging them to create new workflows and efficiencies in order to manage working in a small department, a cross-departmental team, or with redefined staff responsibilities.

Background Information

Located in the southeast corner of Kansas, Pittsburg State University (PSU) has a student population of more than 7,400. The PSU campus has only one library, Leonard H. Axe Library. Fort Hays State University (FHSU) is located in western Kansas, between Denver, CO, and Kansas City, KS. FHSU's enrollment hit 14,000 in 2015, including more than 6,000 online students in over 20 countries. Like the PSU campus, Forsyth Library is the only library at the FHSU campus. Both PSU and FHSU are a small universities in rural Kansas, but have had continual growth as one of the Kansas Board of Regents' universities.

PSU and FHSU currently use CONTENTdm (CDM) as their primary digital repository. In 2015, both institutions purchased and launched an open access digital repository from Berkeley Electronic Press (bepress) called Digital Commons (DC), which showcases a variety of scholarship produced by the university, such as theses and dissertations (ETDs), reports, conferences, journals, and peer-reviewed publications. Both institutions purchased DC to have a more robust repository for scholarship published by faculty and students. At the same time creating efficient ways to enhance the value and capture the global impact of the scholarship by making it globally discoverable. DC offers global discoverability, unlimited storage, efficient technical support, and the ability to share a wide range of file formats in one interface. PSU and FHSU have populated their IRs with digital content that includes theses while implementing the digital and scholarly communication initiatives across campus.

Literature Review

Making theses and dissertations available to the scholarly community is an integral part of the research process at the university. As the university aims to achieve access to theses and dissertations, an open access digital repository represents a key resource to realize that purpose. Theses and dissertations gain their visibility and discoverability through a digital repository. More than half of the institutions implementing an IR are making theses and dissertations available in their repositories where they can contribute to the impact of their institutions (Schöpfel). Accessing these electronic theses and dissertations (ETDs),

researchers are able to easily retrieve valuable knowledge that may not be in journal articles and other publications to expand their research activities.

The development of ETD implementation through an IR requires collaborative work from the university administration or the graduate school, academic departments, and the library, as well as participation from students. The strong support network in the university community is important to a successful implementation of an ETD project with a digital repository. Communication between the graduate school and the library is a crucial step to create an ETD workflow. Establishing a workflow for the ETD project is challenging due to the changing organizational culture at institutions (Reeves).

The case study at University of North Carolina Greensboro shows “How do these institutions handle the interdepartmental communication and collaboration needs of ETD programs?” They conclude that the strong communication and collaborative relationships between the university campus units and the library create the rich environment for providing opportunities to build and implement ETD programs efficiently through workflow evaluations and discussions of others’ ideas and thoughts (Early and Taber 13).

This paper examines the process of selecting IRs for ETD programs, including the advantages and disadvantages of the two platforms, CONTENTdm and Digital Commons which are used at PSU and FHSU. This paper will also consider the construction of an electronic theses collection and address in a fair amount of detail the workflows which were established to support integrating thesis materials into a new digital repository and the collaborative relationship with the graduate school in a small institutional setting.

Assessing Advantages and Disadvantages of bepress DC and CDM

Several institutional repository tools are available as an open source software (OSS) tool or proprietary software tool. Examples of OSS in use at academic institutions are DSpace, Eprints, and Fedora/Fez. These OSS tools are freely available and users can run and distribute the software, but systems and database administration, server maintenance and application support are required. On the contrary, the implementation of a proprietary type of institutional repository comes with a consultant and maintenance services. The most widely used proprietary IR software tools are CONTENTdm and Digital Commons (Amaral 1-3).

The process of selecting an IR employs a needs assessment to ensure the existing demand, content characteristics, technical and organizational capacity, and manpower needs and development. PSU and FHSU use CONTENTdm as a primary IR and they selected DC to implement a new IR. This selection is due to the manpower and technical issues that need to be ironed out in order for it to be successful.

Both CONTENTdm and DC are widely implemented in academic libraries to host an institutional repository and provide positive features, including presentation of the various types of digital materials. CONTENTdm is hosted on OCLC and is a stand-alone digital asset management system. CONTENTdm is best for image-based materials with a large

metadata structure which allows for granular metadata for those sorts of visual materials. Digital Commons is widely implemented as a flexible, robust and open-access institutional repository solution. Because of this both institutions thought Digital Commons was the best for showcasing scholarship produced by faculty and students.

To better evaluate the two platforms the authors looked to a document published in-house at FHSU in 2011 when the university first began investigating the establishment an IR, and also a report published by UNESCO comparing the current systems used for IRs (Bankier and Gleason 5 – 14; Weiss 44-49). The following table derives content from both documents by FHSU and UNESCO.

Table 1
Comparison of basic and major functionalities between CONTENTdm and bepress DC

CRITERIA	CONTENTdm (http://www.contentdm.org/)	bepress DC (http://www.bepress.com/ir/)
Open Source/Proprietary	Proprietary	Proprietary
Software or Hosted Service	Hosted Service	Hosted Service
Support Available	YES – via CDM	YES – via bepress (email, phone, resources, and community support)
Content File Formats	Some	All
Metadata Standard	Simple and Qualified Dublin Core and customizable metadata	Simple and Qualified Dublin Core and customizable metadata
Syndication (RSS, etc.)	NO	YES
Statistical Reporting	YES	YES
Design Rationale for IR Flexibility	NO	YES
Design Rationale for IR Accessibility	YES	YES
Design Rationale for IR Interoperability (OAI-PMH)	YES	YES
Implementation Technologies (Scripting language, Database, Operating System)	N/A: hosted (System requirements: Linux/Windows)	N/A: hosted
Storage	YES, but costs \$ over 196,000 items	NO (unlimited storage)
Batch processing	YES	YES
Journal Publishing	NO	YES

(Bankier, Bankier and Gleason; Weiss)

The table shows a comparison between CONTENTdm and bepress DC. Bepress DC clearly has the capability to bring scholarly materials, such as theses and dissertations, together in one searchable location, while CONTENTdm has the advantage when handling visual materials with flexible metadata.

ETD Workflow

Designing an effective workflow is one of the essential requirements for an ETD project to be successful. The development of an ETD program required the collaborative work of the graduate school and the library, as well as participation from the faculty and students.

PSU's Leonard H. Axe Library began the process of branding their DC platform in August 2015 which went live December 1, 2015. A working group between the PSU IT and Library Services was established to set-up the branding for our user interface for PSU Digital Commons (<http://digitalcommons.pittstate.edu>). Immediately after going live the Digital Resources and Initiatives Manager migrated only forty-five of the 119 theses from CONTENTdm (<http://axedigital.pittstate.edu>) to Digital Commons via the batch method using DropBox. DropBox requires an extra step of changing the URL for DC to acquire the file. PSU currently backs-up all of our digital materials on a QNAP which requires a login to access the files. However, working closely with DC support the process of using DropBox went rather smoothly. Library Services and the graduate school are currently developing workflows (presented below) for students to submit their theses and go through the review process in DC beginning in fall 2016. All retro scanning of theses is uploaded by the Digital Resources and Initiatives department.

Most of the collections in CONTENTdm will remain there for the simple fact that the Library likes the interface for our visual and audio collections. PSU is using DC primarily for keeping track of graduate and undergraduate scholarship and faculty scholarship when faculty wish to participate. Our first major collection consisted of photographs and video from the Student Research Colloquium in spring 2015. This is an annual event and statistics already show it to be one of the most accessed collections. Other collections recently added are Finding Aids, Annual Faculty Author Reception, material from the Gene DeGruson Memorial Lecture series, papers from two History classes, the Kanza yearbooks, and Open Pitt, the new home for OER (Open Educational Resource) produced at PSU.

Comparatively, FHSU's Forsyth Library launched the DC platform, branding their implementation, FHSU Scholars Repository (<http://scholars.fhsu.edu/>), in December 2015, and officially made an announcement to the campus departments in January 2016. FHSU Scholars Repository has currently a few collections built: faculty papers, OERs (Open Educational Resources), e-Journals, and Archives & Special Collections materials. With FHSU IT support, the Library has set-up the campus proxy server as a publicly accessible server for a batch upload processes. The DC system allows FHSU to upload multiple items at a time by implementing this batch feature. The FHSU Forsyth Library uses FileZilla (<https://filezilla-project.org/>) as a FTP client to connect the server. Some digital collections, such as e-journal publications, have been migrated from CONTENTdm to the DC platform

by implementing the batch process. The theses collection on the CONTENTdm platform (<http://contentcat.fhsu.edu/cdm/>) are targeted as a next migration.

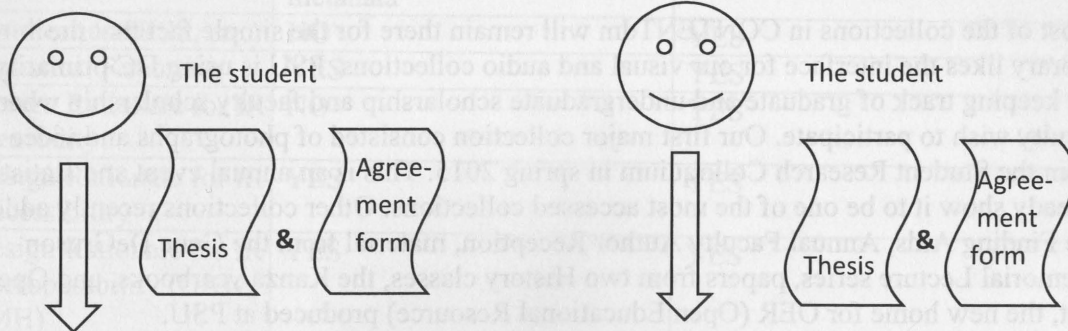
The students are required to submit their theses to the graduate school office and those theses would be presented on the CONTENTdm platform. The theses collection on this platform contains nearly 3000 items from 1930 through the present. Comparatively, many institutions use *ProQuest* to disseminate and archive their theses and dissertations via *ProQuest*, however, both PSU and FSHU do not participate in *ProQuest* theses program.

Because PSU is in transition, their students are just beginning to set up their accounts in Digital Commons and submitting their own thesis work through the system. The “Past ETD Workflow at PSU” in fig. 1 is very similar to FHSU’s and applied to submissions through spring 2016. PSU has approximately ten to twelve submissions for spring term and two to five for fall term. In contrast, PSU has approximately twenty to twenty-five for spring and five to eight in the fall.

The current ETD workflow at FHSU is simple because the total number of theses submitted per semester is small and the submission to *ProQuest* is not required (see fig. 1). The average number of theses submitted by per semester is less than ten papers. The library receives thesis materials with signed repository publishing agreement forms electronically via the Graduate School Office, then presents those materials on the CONTENTdm platform.

Current ETD Workflow at FHSU

Past ETD Workflow at PSU



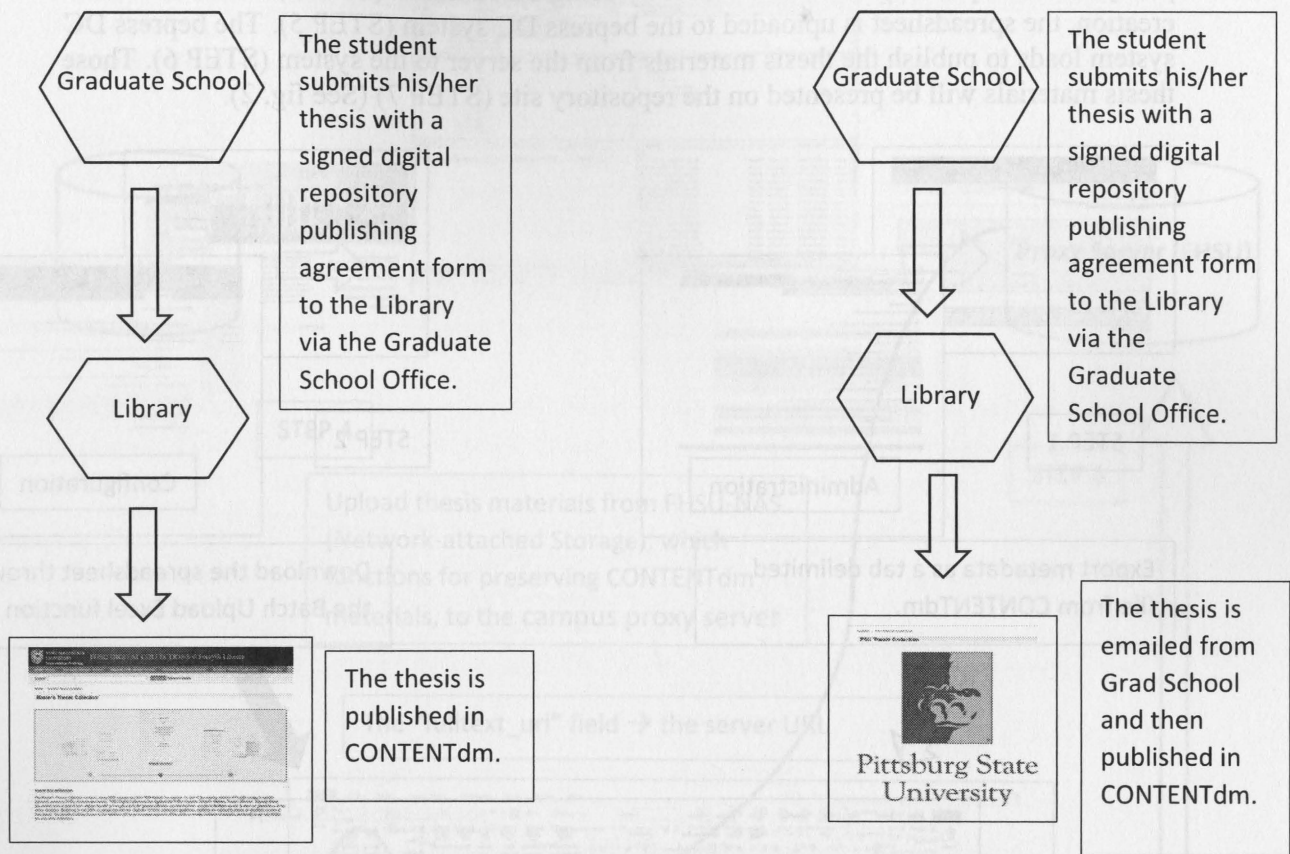


Fig. 1. Past at PSU and current workflow at FHSU.

As the FHSU Forsyth Library has not implemented the theses migration from CONTENTdm to DC, the following Figures 2 and 3 are outlining potential future plans for their theses program development.

Before implementing the theses migration from CONTENTdm to Digital Commons, an institutional repository publishing agreement form would need to be updated. The agreement form would state that theses will be deposited into the FHSU's institutional repository, FHSU Scholars Repository. With the Graduate School Office's agreement, the migration process would be implemented.

The first step of the migration process would be a batch creation which allows uploading a number of thesis materials, which are currently presented on the CONTENTdm. At the time, CONTENTdm provided a metadata export function. FHSU staff can export metadata describing the Thesis materials on CONTENTdm (STEP 1). Then, they can download the spreadsheet through the bepress DC batch tools (STEP 2). The spreadsheet consists of metadata fields, such as a title, abstract, author name (s), and full text URL. The "full text URL" field indicates the URL of the item which is on a publicly accessible server which the bepress system will access to and copy the file at the URL provided and store it (STEP 3). Targeted thesis materials to the batch process are uploaded from the FHSU-NAS (Network-attached Storage), which functions for preserving CONTENTdm materials, to the campus

proxy server or publicly accessible server by using a FTP client (STEP4). After the batch creation, the spreadsheet is uploaded to the bepress DC system (STEP 5). The bepress DC system loads to publish the thesis materials from the server to the system (STEP 6). Those thesis materials will be presented on the repository site (STEP 7) (See fig. 2).

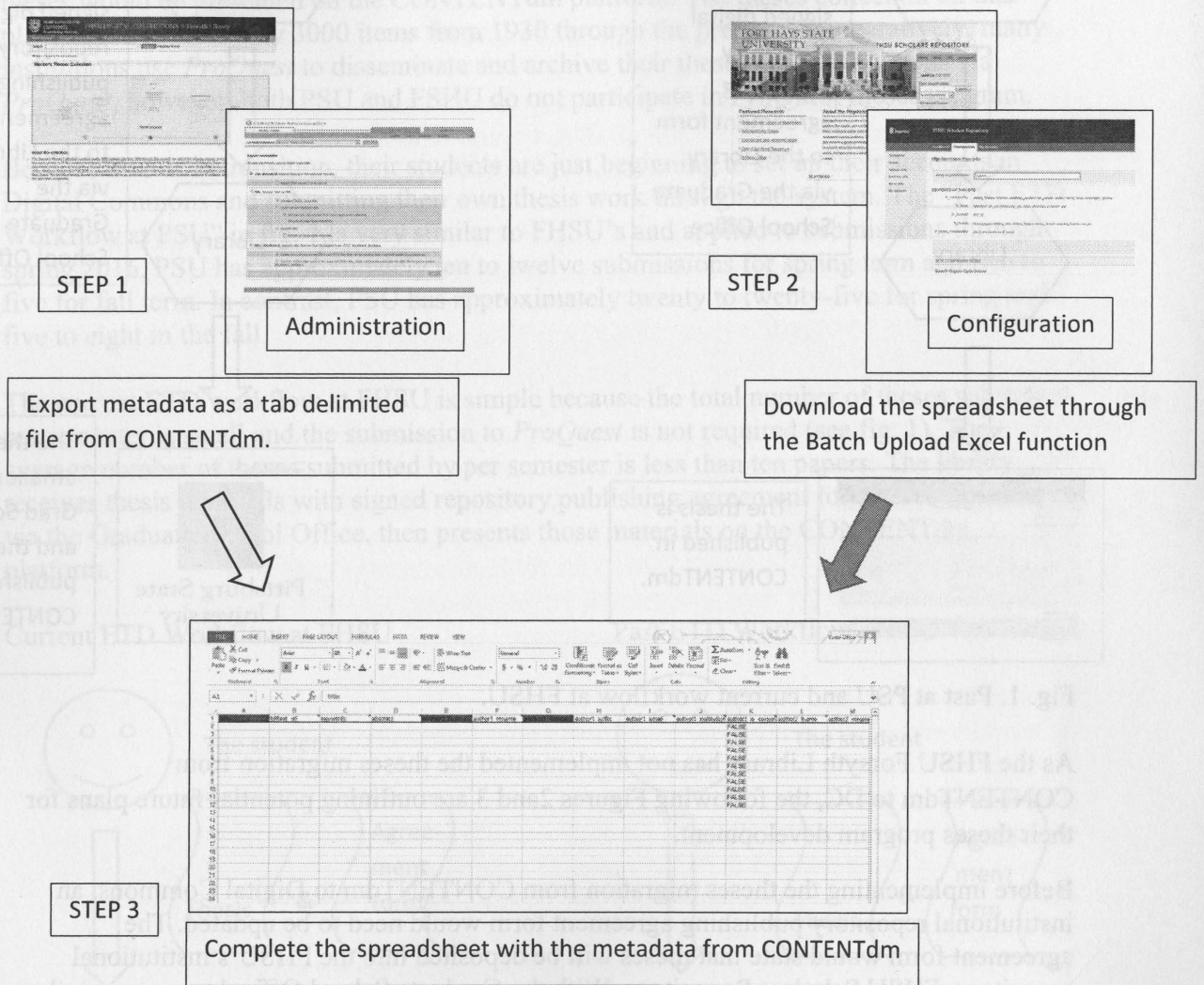


Fig. 2.1 Theses migration: CONTENTdm – bepress DC at FHSU.

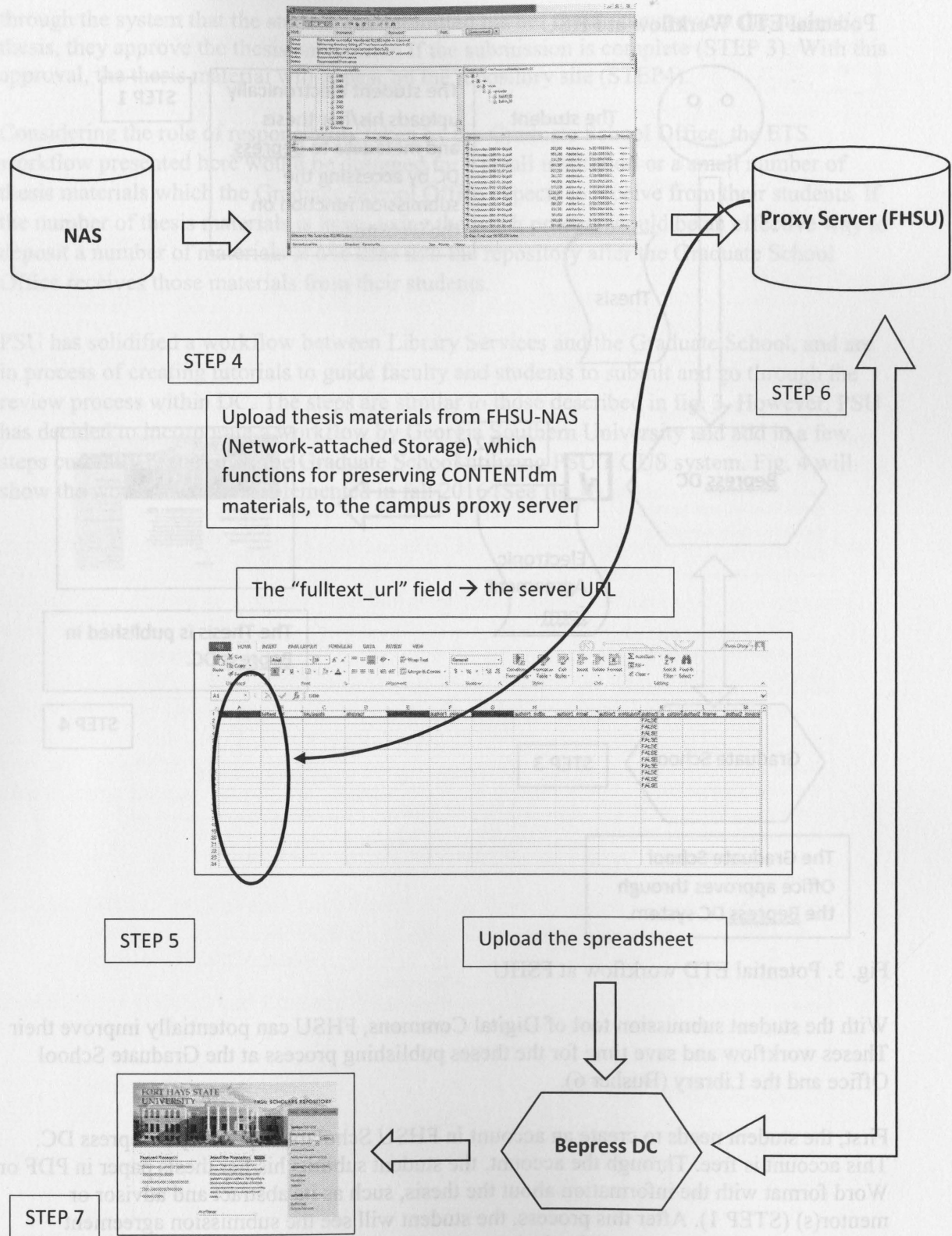


Fig. 2.2 Theses migration: CONTENTdm – bepress DC at FHSU.

Potential ETD Workflow at FHSU

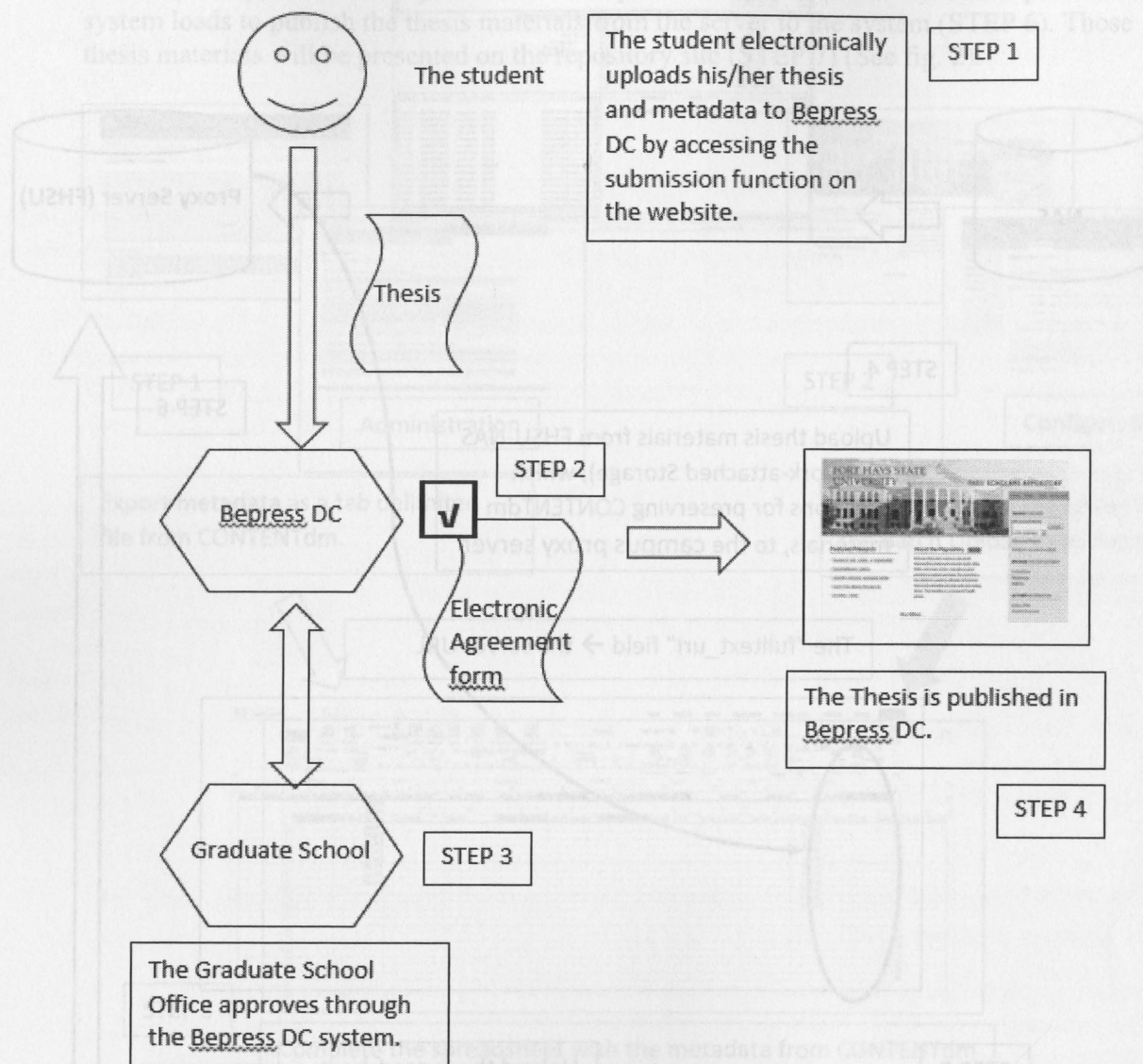


Fig. 3. Potential ETD workflow at FSHU

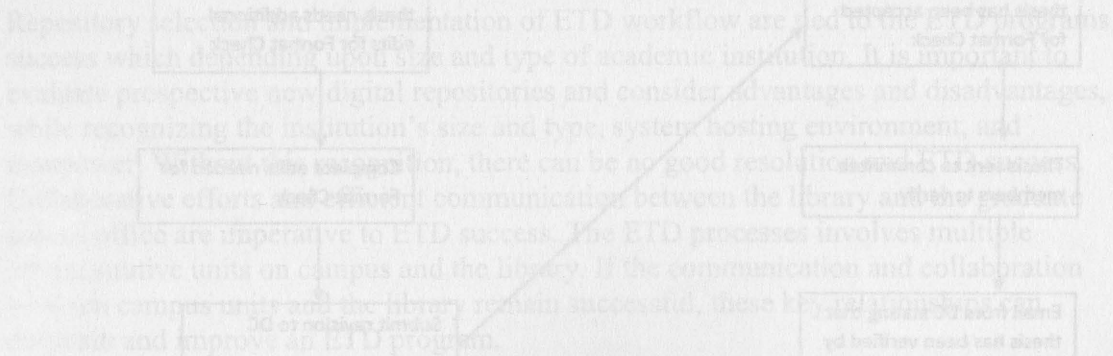
With the student submission tool of Digital Commons, FHSU can potentially improve their Theses workflow and save time for the theses publishing process at the Graduate School Office and the Library (Busher 6).

First, the student needs to create an account in FHSU Scholars Repository – bepress DC. This account is free. Through the account, the student submits his/her thesis paper in PDF or Word format with the information about the thesis, such as its abstract and advisor or mentor(s) (STEP 1). After this process, the student will see the submission agreement (STEP 2). Once the student clicks the box to indicate his/her agreement with the terms, the thesis material will be submitted. When the Graduate School Office receives notification

through the system that the student has submitted his/her thesis, they review the student's thesis, they approve the thesis submission if the submission is complete (STEP 3). With this approval, the thesis material will appear on the repository site (STEP4).

Considering the role of responsibility taken by the Graduate School Office, the ETS workflow presented here would be designed for a small institution or a small number of thesis materials which the Graduate School Office expects to receive from their students. If the number of thesis materials is large, using the batch process would be an effective way to deposit a number of materials at one time into the repository after the Graduate School Office receives those materials from their students.

PSU has solidified a workflow between Library Services and the Graduate School, and are in process of creating tutorials to guide faculty and students to submit and go through the review process within DC. The steps are similar to those described in fig. 3. However, PSU has decided to incorporate a workflow by Georgia Southern University and add in a few steps currently required by the Graduate School utilizing PSU's GUS system. Fig. 4 will show the workflow to be implemented in fall 2016 (See fig. 4).



Works Cited

Arnsper, Megan. "Institutional Repositories, Open Source Options, and Libraries." *Core.ac.uk CORE*, 2008. Web. 13 June 2016. <<https://core.ac.uk/download/files/418/11684820.pdf>>

Bankier, Jean Gabriel, and Kenneth G. Gorman. "Digital Repositories: A Comparative Study." *Cultural Organization*, 2014. Web. 13 June 2016. <<http://www.bnl.gov/bnlweb/academic/MULTIMEDIA/DC/DCI/pdf/www.institutional-repository-comparative-study.pdf>>

Busher, Casey. "Saving Trees, Time, and Money: Best Practices & Procedures for ETDs." *digitalcommons.bepress.com*. *BePress Digital Commons*, 2014. Web. 13 June 2016. <<http://digitalcommons.bepress.com/etds/>>

Crow, Ryan. "The Case for Institutional Repositories." *A Monthly Report on Research, Library Issues, and Action from ARL, CNL and*

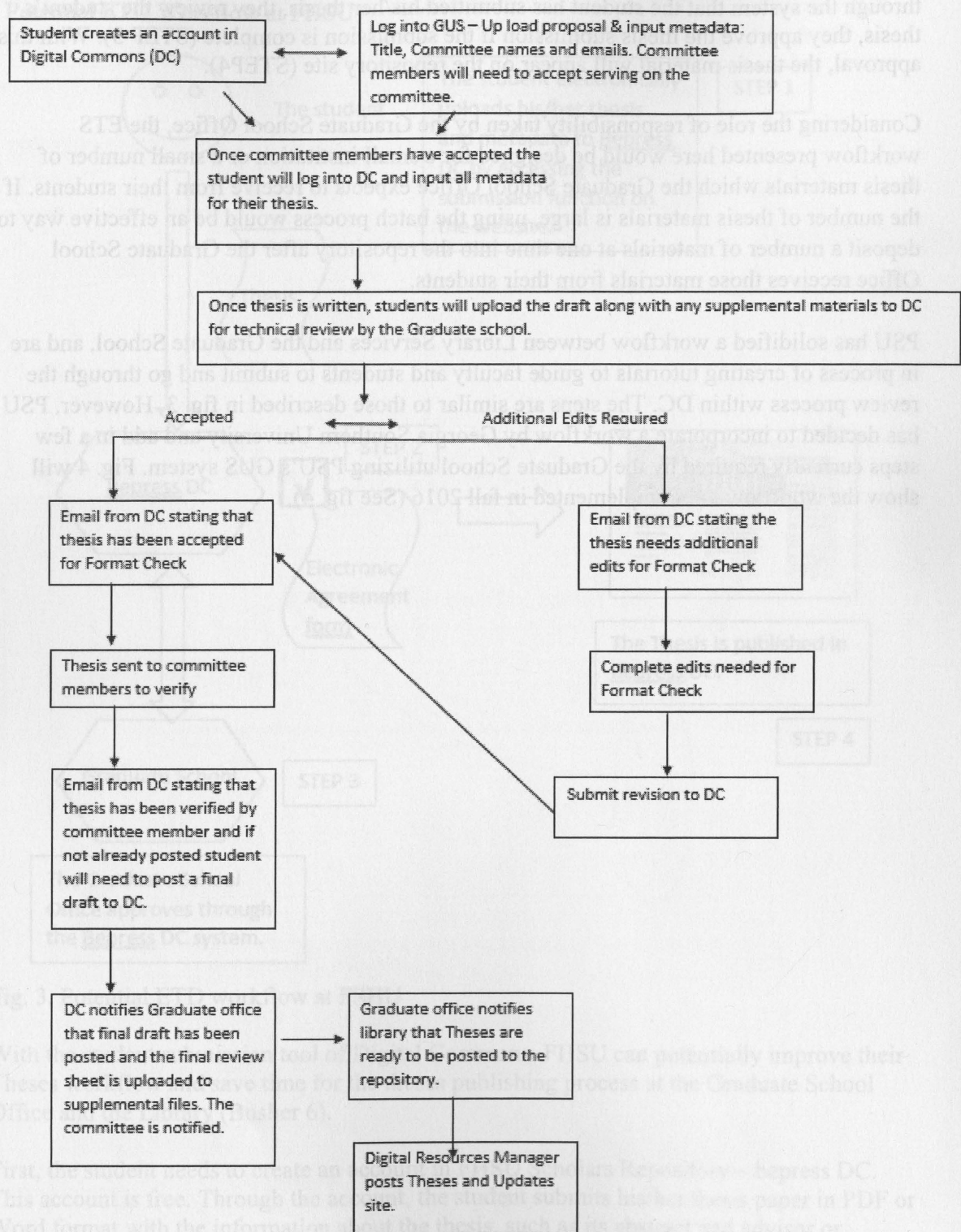


Fig. 4. PSU fall 2016 electronic thesis submission using DC

PSU has a few extra checks and balances in their workflow to incorporate GUS. As all persons involved actually work through the process some of these steps may change or disappear. PSU is no longer requiring students to submit a physical copy to Special Collections unless the student refuses to allow their thesis to be available through the repository, and the decision has been approved by the department and Graduate School.

Conclusion

A growing number of scholarly works are deposited into institutional repositories and openly available. Students are benefitting from the institutional repository by depositing their theses and dissertations. The efforts to develop and establish strategies guiding deposition of theses and dissertations into IRs will promote greater appreciation of the impact of scholarly output. The aim of developing workflows to promote the ETD programs and collaboration across campus will enhance the distribution of digital content. Cross-campus collaboration and inter-library collaboration are essential to build a robust ETD program and encourage further dialog about the need of academic institutions in the future. The aim of ETD programs include providing greater recognition and exposure to the wealth of information and scholarship that theses and dissertations represent.

Repository selection and implementation of ETD workflow are tied to the ETD programs success which depending upon size and type of academic institution. It is important to evaluate prospective new digital repositories and consider advantages and disadvantages, while recognizing the institution's size and type, system hosting environment, and manpower. Without this recognition, there can be no good resolution and ETD success. Collaborative efforts and efficient communication between the library and the graduate school office are imperative to ETD success. The ETD processes involves multiple administrative units on campus and the library. If the communication and collaboration between campus units and the library remain successful, these key relationships can maintain and improve an ETD program.

Works Cited

- Amaral, Megan. "Institutional Repositories, Open Source Options, and Libraries." *Core.ac.uk*. CORE, 2008. Web. 13 June 2016. <<https://core.ac.uk/download/files/418/11884820.pdf>>.
- Bankier, Jean Gabriel, and Kenneth Gleason. "Institutional Repository Software Comparison." *www.unesco.org*. United Nations Educational, Scientific and Cultural Organization, 2014. Web. 13 June 2016. <http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CI/CI/pdf/news/institutional_repository_software.pdf>.
- Busher, Casey. "Saving Trees, Time and Money: Best Practices & Procedures for ETDs." *digitalcommons.bepress.com*. bepress Digital Commons, 2014: Web. 13 June 2016. <<http://digitalcommons.bepress.com/etds/7>>.
- Crow, Ryan. "The Case for Institutional Repositories: A SPARC Position Paper." *ARL: A Bimonthly Report on Research, Library Issues, and Action from ARL, CNI, and*

- SPARC 223 (August 2002): 1-4. Web. 13 June 2016.
<<http://www.arl.org/storage/documents/publications/arl-br-223.pdf>>.
- Early, Mary G., and Anne Marie Taber. "Evolving in Collaboration: Electronic Thesis and Dissertation Workflows in North Carolina." *Collaborative Librarianship* 2.1 (2010): 4-18. Web. 13 June 2016. <https://libres.uncg.edu/ir/uncg/f/A_Taber_Evolving_2010.pdf>
- Lowery, Ashley D. "Submitting Your Thesis/Dissertation into Digital Commons@GeorgiSouthern." *cogs.georgiasouthern.edu*. Southern Georgia University, 2016. Web. 30 Dec. 2015<<http://cogs.georgiasouthern.edu/etd-submission-links>>.
- Madsen, Debora L. and Jenny K. Oleen. "Staffing and Workflow of a Maturing Institutional Repository." *Journal of Librarianship and Scholarly Communication* 1.3 (2013). Web. 13 June 2016. <<http://doi.org/10.7710/2162-3309.1063>>.
- Reeves, Sharon. "How to set up an ETD program." *www.ndltd.org*. Networked Digital Library of Theses and Dissertations, n.d. Web. 13 June 2016.
<<http://www.ndltd.org/resources/manage-etds/set-up-an-etd-program>>.
- Schöpfel, Joachim. "Adding Value to Electronic Theses and Dissertations in Institutional Repositories." *The Magazine of Digital Library Research* 19.3/4 (2013). Web. 13 June 2016. <[doi:10.1045/march2013-schoepfel](https://doi.org/10.1045/march2013-schoepfel)>.
- Weiss, Andrew. "Creating an FHSU Campus-Wide Digital Institutional Repository." *www.fhsu.edu*, Fort Hayes State University, 2011. Web. 13 June 2016.
<<http://www.fhsu.edu/uploadedFiles/academic/library/digital/IR-FHSU-Proposal-Final%281%29.pdf>>.