

Institutional Digital Repository Phase I Final Report

This is a final report describing the activities surrounding Phase I of a one-year pilot project called the University of Notre Dame Institutional Digital Repository. After outlining the goals and methods of the project, the report enumerates ways the project could be continued. The seventy-some people who participated in the project are now looking to administrators across the University to become familiar with the contents of the report and set its future course.

An abridged, HTML version of this report can be found online at:

<http://www.library.nd.edu/idr/documents/idr-final-report.shtml>

Eric Lease Morgan & Team IDR
December 18, 2006

TABLE OF CONTENTS

EXECUTIVE SUMMARY 5

INTRODUCTION 7

ENVIRONMENTAL SCAN..... 11

TYPES OF CONTENT IN THE IDR 15

 EXCELLENT UNDERGRADUATE RESEARCH 15

 PUBLICATIONS FROM THE INSTITUTE OF LATINO STUDIES 16

 KELLOGG INSTITUTE WORKING PAPERS 17

 ENGINEERING CITATIONS 17

 COMPUTER SCIENCE TECHNICAL REPORTS 18

 ELECTRONIC THESES & DISSERTATIONS 18

 ART IMAGE LIBRARY 19

TECHNOLOGY BEHIND THE IDR..... 21

RAISING AWARENESS THROUGH MARKETING..... 25

 AWARENESS OF SCHOLARLY COMMUNICATION ISSUES 25

 STRATEGIC ALIGNMENT 26

PRESERVATION..... 29

COPYRIGHT 31

METADATA 33

A CHINESE MENU OF CHOICES..... 35

RECOMMENDATIONS 41

APPENDIX A: TEAM IDR AND THE PEOPLE THEY WORKED WITH 45

APPENDIX B: COPYRIGHT ENVIRONMENTAL SCAN 49

APPENDIX C: COPYRIGHT AND DUE DILIGENCE FOR THE IDR 53

**APPENDIX D: UNIVERSITY OF NOTRE DAME ART IMAGE LIBRARY
COPYRIGHT REPORT II**..... 55

APPENDIX E: PERMISSION TO USE WORK 59

APPENDIX F: METADATA REPORT 61

APPENDIX G: IDR MARKETING PLAN - DRAFT 71

Executive summary

HERE IS THE BRIEFEST OF SUMMARIES REGARDING WHAT WE DID, WHAT WE LEARNED, AND WHERE WE THINK FUTURE DIRECTIONS SHOULD GO:

-
- 1) **What we did** - In a nutshell we established relationships with a number of content groups across campus: the Kellogg Institute, the Institute for Latino Studies, Art History, Electrical Engineering, Computer Science, Life Science, the Nanovic Institute, the Kaneb Center, the School of Architecture, FTT (Film, Television, and Theater), the Gigot Center for Entrepreneurial Studies, the Institute for Scholarship in the Liberal Arts, the Graduate School, the University Intellectual Property Committee, the Provost's Office, and General Counsel. Next, we collected content from many of these groups, "cataloged" it, and saved it into three different computer systems: DigiTool, ETD-db, and DSpace. Finally, we aggregated this content into a centralized cache to provide enhanced browsing, searching, and syndication services against the content.

 - 2) **What we learned** - We essentially learned four things: 1) metadata matters, 2) preservation now, not later, 3) the IDR requires dedicated people with specific skills, 4) copyright raises the largest number of questions regarding the fulfillment of the goals of the IDR.

 - 3) **Where we are leaning in regards to recommendations** - The recommendations take the form of a "Chinese menu" of options, and the options are be grouped into "meals". We recommend the IDR continue and include: 1) continuing to do the Electronic Theses & Dissertations, 2) writing and implementing metadata and preservation policies and procedures, 3) taking the Excellent Undergraduate Research to the next level, and 4) continuing to implement DigiTool. There are quite a number of other options, but they may be deemed too expensive to implement.

Introduction

THE PROBLEM WE ARE TRYING TO SOLVE

The three-fold purpose of the Institutional Digital Repository (IDR) is closely aligned with the goals of the University. The IDR's three goals are:

1. to make it easier for students to learn
2. to make it easier for instructors to teach
3. to supplement the scholarly communications process

With these goals in mind the IDR is defined as a set of digital objects combined with sets of services applied against those objects – think “digital library”.

**GREAT CONTENT IS
CREATED HERE AT
NOTRE DAME,
RANGING FROM
THESES &
DISSERTATIONS
WRITTEN BY
GRADUATE
STUDENTS TO
ESSAYS AND
POSTERS BY
EXCELLENT
UNDERGRADUATES.**

Great content is created here at Notre Dame, ranging from theses & dissertations written by graduate students to essays and posters by excellent undergraduates. It includes the output of institutes that explore and elaborate upon issues pertinent to their associated fields of study. It includes articles written by faculty as the result of their research and scholarship can be found as well as scanned images used to teach art history and other subjects.

By making greater amounts of University scholarly output accessible on the Web, more people with similar interests will be able to find, read, and comment on the good work happening at Notre Dame. Instead of hiding the University's “candle under a bushel basket”, the IDR can make content

more accessible to the outside world and exploit the globally networked environment in which we live and work.

INTRODUCTION

The more than 1,300 items comprising the current content of the IDR are diverse in subject matter and format. For example, the subjects represented by the IDR include but are not limited to: Anthropology, Asian Studies, Biology & Life Science, Business, Economics, Engineering, European Studies, Geology, History, Latin American Studies, Latino Studies, Political Science, Science and Technology, Social Science, Sociology, and Theology & Religion. This content is manifested by more than:

- 650 images
- 500 articles
- 350 theses & dissertations
- 250 technical reports
- 100 working papers

The IDR is searchable through a number of simple or advanced interfaces. It can be browsed by author, department, subject, and format. The IDR is not necessarily intended to be a destination, but rather a platform for syndicating content. Consequently, information contained in the IDR can easily and seamlessly be incorporated into the Web pages of the campus-wide portal, a department's website, or an individual's home page.

The IDR's digital objects can be just about any computer file created at the University for teaching, learning, or research. Types of content that could be included are: working papers, data sets, pre-prints, pictures, movies, sounds, technical reports, conference presentations, etc. Computer formats that could be included are: Word documents, PDF files, JPEG images, tab-delimited text files, MPEG files, LaTeX files, Postscript files, etc.

The services for the IDR would be the same sorts of services expected from any repository that includes search and browse, but also could include other value-added services. Some examples of value-added services are: What's New?, create my vita, show me my Google Page Rank, syndicate my content to the campus-wide portal, syndicate my content to departmental Web pages, create content- or subject-specific search and browse interfaces, show me who has looked at my page, show me who links to my page. In short, the IDR is a digital library designed to address a number of the University's priorities and the articulated needs of University students, instructors, and scholars for teaching, learning, and research.

The primary goal of the IDR is to explore and supplement ways to enhance learning, teaching, and research. It attempts to solve that problem to by collecting, organizing, archiving, and providing access to value-added services against content "born digital" here at Notre Dame.

I N T R O D U C T I O N

The balance of this report outlines the current environment of institutional repositories and scholarly communication; what Team IDR did to learn about providing institutional repository services; a number of issues that have come to our attention because of these efforts; and finally, a number of options available to University administrations for taking the IDR to the next step.

Environmental scan

PEOPLE'S EXPECTATIONS REGARDING ACCESS TO INFORMATION HAVE SIGNIFICANTLY CHANGED WITH THE ADVENT OF GLOBALLY NETWORKED COMPUTERS.

In combination with sets of nebulously resolved intellectual property rights issues, these expectations have created an environment where some traditional best practices are increasingly seen as dysfunctional and new sets of best practices are yet to be established. The Academe, specifically the scholarly communications process, has not gone untouched in this regard. This section paints a picture of the current environment and how it came to be.

By the early to mid-1990's almost every faculty and staff member in institutions of higher education had computers on their desktops that were connected to the Internet. The typewriter had all but disappeared to be replaced with the word processor. Through the use of internal networks, files were easily duplicated from computer to

**IN 1994 STEVAN
HARNAD... IDENTIFIED
THIS TREND AND
WROTE THE
"SUBVERSIVE
PROPOSAL" THAT
SUGGESTED
AUTHORS PUBLISH
THEIR OWN
SCHOLARLY WORKS
THROUGH THE USE
OF "PUBLIC FTP"
... MAKING THEM
FREELY AVAILABLE
ON THE INTERNET.¹**

computer. Scholarly works such as journal articles and monograph manuscripts were sent to publishers via email. Computers allowed people to create exact copies of their "born digital" creations.

At the same time, the scholarly publishing industry was undergoing great changes. Smaller publishing houses were being purchased and subsumed by larger publishers. Academia saw an increasing number of specialized titles. Fewer and fewer journals were being published by institutions of higher education and instead they were being sold to and published by for-profit institutions.

In 1994 Stevan Harnad, then a cognitive scientist at Princeton University, identified this trend and wrote the "Subversive Proposal" that suggested authors publish their own scholarly works through the use of "public FTP (file transfer protocol)", thereby making them freely available on the

Internet.¹ This was an influential proposal, and essentially the first articulation for something that would later be called “open access”.

By the mid-1990’s the scholarly publishing industry had leveraged increasing monopolies and, combined with a number of other factors, significantly raised prices, especially journal prices. These other factors include: junior faculty mandated to write and to publish in particular journals in order to achieve tenure (“publish or perish”); legal agreements that transfer intellectual property rights from author to publisher, and librarians pressured to maintain broad as well as deep collections. In such an environment for-profit publishers have been able to successfully raise the prices of their products and services by at least 7 percent / year for the past fifteen years. Consequently, libraries commonly pay \$2,000/year for a subscription to a scientific journal and \$150/year for a humanities journal. As of 2004 scientific books cost around \$100 each and humanities books cost about \$50 each.

In the late 1990’s and beginnings of the 2000’s, we saw the birth of the World Wide Web and the “dot-com boom”. While the technology behind the Web was created in the very early 1990’s, it did not become popular until the mass distribution of the “graphical Web browser” by Netscape, Inc. Companies fueled by venture capitalists sprang up all over the place. Using HTML, people rushed to put their content on the Web. With easy availability of all sorts of content, people spent hours “Web surfing.”

Libraries responded in two ways. First, they began creating digital surrogates of some of their more rare and special collections. Second, they created lists of useful Internet resources akin to records in the venerable library catalog. Universities, university departments, and faculty reacted similarly by creating sets of “home pages” describing what they were doing and why it was important. University computing centers offered space for individuals to host their own sites, where online vitas, research agendas, and publication lists were created. Publishers responded by making more of their content available electronically and in full-text forms. Publishers also changed their subscription model. Often now libraries license the ability to view content, but not necessarily to own it.

Probably the most visible by-products of the dot-com boom are Yahoo and Google. Yahoo began its life doing what many libraries were doing, namely creating lists of Internet resources, organizing them into groups, and allowing people to browse the collection. Yahoo is an acronym for “Yet Another Hierarchical Official Oracle”. Google focused its attention on search instead of browse. By crawling the Web, counting the number of times people linked to various websites, indexing the content, and providing a very simple interface for searching the index, Google became the de-facto standard for searching the Web. If there is one thing that changed people’s expectation regarding access to information, that one thing is Google. Enter a few

¹ <http://www.arl.org/scomm/subversive/sub01.html>

words. Get back a list of hits. Click on an item. Get the content. Fast. Easy. Usually very satisfying.

Academia, especially in Europe, was taking notice of these changes, and in 2002 the phrase “open access” was coined in the Budapest Open Access Initiative. The Initiative, a public statement regarding the dissemination of peer-reviewed scholarly materials, defined “open access” as a kind of free and unrestricted online availability of scholarly materials. From the Initiative:

An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge.²

Since the Initiative was articulated, numerous other statements have been written to advocate the importance of open access publishing including: the Bethesda Statement on Open Access Publishing, the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, and the Declaration On Access To Research Data From Public Funding.

The concept of open access publishing has manifested in two different ways: peer-review journals and repositories. To date the Directory of Open Access Journals lists about 2,400 peer-reviewed, open access journal titles.³ These journals all are “free”, scholarly, and cover a wide range of disciplines and subjects. The Public Library of Science and BioOne have become major publishers of open access journals that disseminate a large body of scientific literature.

Open access publishing also is manifested in repositories of two types: institution and discipline. Institutional repositories are collections such as pre-prints, working papers, post-prints, technical reports that originate from a specific university or college. MIT’s repository is a good example. Collections of electronic theses & dissertations are probably the most popular form of repository. Their content is aggregated into the Networked Digital Library of Theses and Dissertations. Other repositories are subject-based. The most notable subject-based repository is arXiv, hosted at Cornell University. It contains a huge collection of scientific literature with heavy emphasis on

² <http://www.soros.org/openaccess/read.shtml>

³ <http://www.doaj.org/>

ENVIRONMENTAL SCAN

physics. A project called OpenDOAR is attempting to list as many repositories as possible.

People's expectations have changed since the advent of Google. They expect to do some sort of Internet search, be presented with a list of possible items of interest, click on an item, and read the result. Usability studies show that users are frustrated with libraries because there is too much choice, things are too difficult to search, and electronic full text may not be available. Full text unavailability could mean that the item does not exist digitally. More probably it exists digitally, but the item is copyrighted by a publisher who secured the rights from the original author and now provides access to it through a proprietary, non-standard interface. Journal article publication, unlike book publication, is not about paying the author but rather about the author's work getting cited and recognized. If Academia were to make a greater amount of its "born digital" content freely available online via open access journals and/or repositories, then indexers (like Google) would find it, and libraries would collect and preserve it. Most likely the sphere of knowledge would grow faster and more people would be aware of the Academy's accomplishments.

Types of content in the IDR

SOME STORIES

Phase one of the IDR was a one-year pilot project. Its scope was limited to the content of four groups of people and a larger number of associated individuals:

- images from the Art Image Library – some 10,000 pictures used to support classes in art history
- electronic theses & dissertations – roughly 275 items submitted and accepted by the Graduate School in the past three years
- excellent undergraduate research – a couple of dozen essays, reports, and presentations highlighting exemplary work done by undergraduate students
- intellectual output generated by faculty and researchers – about 1,000 working papers, formal publications, and technical reports from the Kellogg Institute, the Latino Studies Institute, and the Computer Science Department

Each of the associated individuals and their co-workers have been working with one or more of the 27-member Team IDR from the Libraries (see Appendix A) to collect, organize, preserve, and disseminate the digital content. The following sections outline the work done to accommodate these collections.

Excellent undergraduate research

In order to spotlight meritorious scholarship done by undergraduate students, a number of excellent undergraduate research materials have been added to the IDR.⁴

Late in 2005 a Vice President and Associate Provost brought together a people from a number of University departments in an effort to begin the process of highlighting excellent undergraduate research. These departments included: Anthropology, Architecture, Biological Sciences, Electrical Engineering, FIT (Film, Television, and Theater), the Gigot Center for Entrepreneurial Studies, the Institute for Scholarship in the Liberal Arts, and the Nanovic Institute. At that time it was decided each

⁴ <http://dewey.library.nd.edu/morgan/idr/undergrad/>

TYPES OF CONTENT IN THE IDR

department would investigate ways to submit examples of “excellent” undergraduate research to the IDR.

Site visits were made, relationships were built, and a laissez-faire policy was adopted regarding what would get submitted and when; each department was to use its own standards for defining “excellence”. In addition it was decided each submission should include a short biography and a picture of the student, as well as a copy of his/her work. All of these things were expected to be displayed as a part of the project, and special software was written and procedures put into place in order to make this come to fruition.

Since the University in this instance is essentially operating as a publisher of content it did not own, a copyright permissions form (see Appendix E) was drafted by University General Counsel essentially stating four things:

1. I, the undersigned, own the copyright to this material.
2. I grant the University the non-exclusive right to distribute this material in the IDR.
3. I retain my copyrights to this material.
4. I retain the right to have the material removed from the IDR as long as the request is submitted in writing.

As materials were submitted to the IDR, a lot of time was spent getting the copyright permission forms signed. In the end, a total of 14 works were submitted from the Anthropology Department, Biological Sciences Department, Business School, and Nanovic Institute.

Publications from the Institute of Latino Studies

Thirty publications from the Institute of Latino Studies have been incorporated into the IDR.⁵

The Institute of Latino Studies publishes research reports, policy briefs, monographs, an annual news magazine, and more. The majority of their content is not in a digital format but thirty of their publications exist as PDF files. Working with the Institute we acquired the digital content, created records describing them, and added them to the IDR. As with the content of the Computer Science Department, the descriptions of the records are not as complete as they could be because of a lack of subject expertise by the people doing the data-entry. This emphasizes the need to work more closely with content creators.

⁵ <http://dewey.library.nd.edu/morgan/idr/latino/>

TYPES OF CONTENT IN THE IDR

Kellogg Institute working papers

Using the previously existing metadata from the University Libraries' catalog, we were able to create and include 161 records describing Kellogg Institute working papers into the IDR.⁶

The Kellogg Institute has a collection of at least twenty year's worth of working papers, that is, essays and articles written by Faculty Fellows, Visiting Fellows and Guest Scholars. Approximately half of their 300 papers are available digitally with the majority already cataloged by the Libraries.

Working with the Institute, the Team IDR acquired the digitized working papers, used the library catalog as a base, and created database records for each of the papers. Then the metadata was harvested into the centralized cache where browsable interfaces were provided similar to the one on Kellogg's website but with an additional searchable interface to the collection, something the Kellogg website does not support.

This particular IDR experiment was especially fruitful because of the full descriptive cataloging that had already been done.

Engineering citations

Through an almost entirely automated process more than 300 engineering citations were added to the IDR.⁷

One of the primary purposes of the IDR is to accomplish scholarly communication by demonstrating and making available the good work done at the University. This work

**USING THIS
AUTOMATED
MEANS WE
WERE ABLE TO
INCORPORATE
A LARGE
QUANTITY OF
DATA QUICKLY,
THOROUGHLY,
AND
ACCURATELY.**

is partially exemplified by articles published in scholarly journal literature. In an effort to demonstrate how such content can be retrospectively included in the IDR, the Libraries' bibliographic databases were searched for articles authored by Notre Dame faculty in the various engineering departments. The resulting sets of bibliographic citations were saved in a database application called EndNote and subsequently exported to an XML file. The XML was converted into a format usable by the DSpace importing routines and associated each citation with a dummy PDF document. The dummy PDF documents were used because the actual articles are quite likely copyright protected. After the data were ingested, they were harvested into the centralized cache where they are searchable and browsable in the same manner as the IDR's other content.

⁶ <http://dewey.library.nd.edu/morgan/idr/kellogg/>

⁷ <http://dewey.library.nd.edu/morgan/idr/>

TYPES OF CONTENT IN THE IDR

Using this automated means we were able to incorporate a large quantity of data quickly, thoroughly, and accurately.

Computer science technical reports

Using a very manual process, about 250 computer science technical reports were incorporated into the IDR.⁸

For more than a decade the Computer Science Department has been making various technical reports available on its website. Earlier documents are saved as postscript files. Later documents are saved as PDF documents. As a demonstration, the older postscript files were converted to PDF and names, titles, dates, and abstracts were entered into the database. These items were then harvested and cached into the central IDR for searching and browsing.

Because the people doing the data entry did not have the necessary subject expertise, some things such as keywords were not added to the system unless they had been provided by the original document. This process taught us about the necessity for working more closely with content creators.

Electronic theses & dissertations

The entire corpus of the University's electronic theses & dissertations collection has been incorporated into the IDR.⁹

A few years ago the University Libraries, in conjunction with the Graduate School, implemented a system for collecting, indexing, searching, and disseminating electronic versions of theses & dissertations. The process centers around a piece of open source software called ETD-db and supported by Virginia Tech. As graduate students (both Masters and Ph.D.) approach graduation they are encouraged by the Graduate School to submit their theses or dissertations electronically as PDF documents. Upon submission, students are asked for their name, degree, title of work, abstract of work, a few keywords describing their work, and whether or not they would like their work to be freely distributed. The Graduate School makes sure the work is correctly formatted and signed by all parties involved. Upon approval from the Graduate School, the work gets passed on to the Libraries where it gets more formally described and added to the library catalog. Finally, the work is made available to the general public, if so specified. To date there are about 360 theses & dissertations in the system.

Using a metadata harvesting protocol called OAI-PMH, information about the theses & dissertations was copied to a centralized cache and sets of value-added services were provided against it. Just like the out-of-the-box ETD-db interface, browsable interfaces were provided based on names and titles. Similarly, searchable interfaces were created against names, titles, abstracts, and keywords. Unlike the standard ETD-

⁸ <http://dewey.library.nd.edu/morgan/idr/?cmd=term&id=41154>

⁹ <http://dewey.library.nd.edu/morgan/idr/etd/>

TYPES OF CONTENT IN THE IDR

db application, the IDR is able to provide lists of theses & dissertations based on departments. This enables academic units to display lists of recently completed works. Since Google has been able to crawl the ETD-db over the past few years, other people have been able to link to its content. Such links increase the works' relevance, and the "PageRank" is able to be displayed by querying Google for this information.

This project demonstrated additional value-added services the IDR can provide against content "born digital" at the University.

Art Image Library

About 650 images from the Art Image Library have been made a part of the IDR.¹⁰

The Art Image Library is a branch of the University Libraries. It is staffed by one full-time curator and a number of part-time students. Located in O'Shaughnessy Hall it includes about 230,000 slides depicting art and architecture from all over the world with an emphasis on Western culture. Most of the slides were created by taking photographs from books. To a lesser extent the slides were purchased by the University or given to the Library by faculty. Each slide is classified, given an accession number, and stored in sets of large metal cabinets.

The primary audience of the Art Image Library is the art historians who use it to teach art history. To make use of the collection, instructors search and browse the library collection and/or identify images from their own collections for classroom use. The library staff then digitizes the necessary content and makes the resulting files available on a shared file system (hard disk). Instructors incorporate the images into PowerPoint files for in-class presentations.

About 10,000 of the slides have been digitized over the past few years. A number of differing automated systems have been used to organize and access these images. Each system had its own strengths and weaknesses. None of the systems has been comprehensive regarding the total number of slides in the collection and none of the systems has been used to thoroughly and consistently describe each slide.

To automate the work of the Art Slide Library to a greater degree, an application called DigiTool was purchased. As a digital asset manager designed for libraries, it promises methods for easily describing images using Dublin Core metadata elements, access control at the item level, and customizable browser-based as well as computer-based interfaces.

Because of a number of challenges, migrating the digital images and their associated metadata into DigiTool proceeded in fits and starts. First, Dublin Core was not seen as an optimal metadata language for art history content. Instead VRA Core was desired, and consequently, time was spent creating a mapping between VRA Core to Dublin

¹⁰ <http://dewey.library.nd.edu/morgan/idr/?cmd=term&id=21392>

TYPES OF CONTENT IN THE IDR

Core. A relational database schema was also designed to manage VRA Core metadata. A previously existing VRA Core database was then identified but the acquisition of the database was delayed many times and proved too time-consuming to implement. Second, the metadata records in the database system most recently used by the library were determined to be incomplete and inconsistent. This made any migration from the existing database to a VRA Core-based system difficult, if not impossible, to automate. It also made it very difficult to ingest the metadata into DigiTool. Third, the OAI-PMH interface implemented by DigiTool and used to create the IDR's centralized cache was not standards-compliant. The problem was exacerbated by an upgrade to DigiTool that broke the interface all together.

To a large extent these challenges have been overcome through the combined use of brute force manual data entry and automation. A subset of the digitized content was identified. Each of the descriptions of the items in the subset was supplemented with subject terms garnered from ARTstor. These descriptions, along with the associated images, were then ingested into DigiTool. Finally, the DigiTool OAI-PMH interface was configured and harnessed to allow the IDR to provide searchable/browsable services against the 630 digital images in the collection.

Technology behind the IDR

MAKING THE SYSTEM TICK

This section describes the technological infrastructure behind the IDR.

**THE
COMPUTER
TECHNOLOGY
USED BY THE
IDR IS A
COMBINATION
OF LEGACY
SOFTWARE,
COMMERCIAL
SOFTWARE,
AND OPEN
SOURCE
SOFTWARE.**

The computer technology used by the IDR is a combination of legacy software, commercial software, and open source software. For example, the University's Electronic Theses & Dissertations Project was implemented a few years ago using Virginia Tech's ETD-db application. While not suited for all types of content such as images, the ETD-db is very well suited for theses & dissertations. Since the Graduate School was using this application with a great deal of success, it was deemed silly to migrate its functionality to another system. It was determined that an application called DigiTool would be used to automate the collection, description, and dissemination of content from the Art Image Library. This commercial software is supported by Ex Libris, the same vendor of the Libraries' integrated library system. DSpace, an open source software application specifically designed for institutional repository use, is used to collect, describe, and disseminate things like working papers, technical reports, and the excellent undergraduate research.

Each of these applications has its own strengths. ETD-db does very well what it was designed to do. DigiTool provides finely grained access control. DSpace is a "free" application widely supported by the academic community and probably ranks as the most popular in its class. At the same time, each of these applications has its own weaknesses. For example, each application operates in its own silo so that it is difficult to integrate the searching process. More importantly, each of these systems operates as a turn-key application so that it is very difficult to change anything but the most rudimentary cosmetic customizations.

Luckily each of these three systems supports a harvesting protocol called OAI-PMH. This protocol allows the metadata to be collected from each of the systems, to cache it locally, and then to provide searching/browsing services against the cache. MyLibrary,

the same software used to drive much of the Libraries' website, is used to provide this functionality. Because MyLibrary is more like a toolbox and less like a turn-key application, there is much more control over the appearance of the implementation as well as more ability to provide additional functionality that none of the turn-key applications supports. Below is an enumeration of some of these features:

1. **Enhanced Dublin Core** - sets of facet/term combinations were created and inserted into DSpace fields. When the content is harvested, its physical structure is noted, parsing and caching it accordingly. This process allows the materials to be described using fields beyond Dublin Core. It also paves the way for greater searching/browsing later on.
2. **Name authority** - For a subset of the implementation, specifically the Excellent Undergraduate Research, photographs and short biographies of students needed to be included in browsable displays. To accomplish this, personal name-based facet/term combinations were created in our database. These authorities defined a key which pointed to a directory containing a JPEG image and more detailed information regarding the author. By combining these facet/terms, the files in the directory, and the records in DSpace a browsable list of authors was able to be created complete with pictures and bios.
3. **Standards-based search** - Each application provides search, but not in a "standard" way, nor against each other. By indexing the cache and providing an SRU (Search/Retrieve via URL) interface to the index, these problems can be overcome. Moreover, by using such an approach our underlying software can be replaced or updated without changing the overall functionality of the system.
4. **Syndicating content with "widgets"** - The name of the Web game is, "Put the content where the users are. Don't expect the users to come to the site." To this end a number of one-line Javascript "widgets" were created to allow Web Masters to insert content from the institutional repository into their pages. For example, the Aerospace Engineering department might want to list the recently completed theses/dissertations, or a faculty member might want to list his/ her publications. These "widgets" allow the content to be syndicated, even to the campus-wide portal.
5. **Pseudo peer-review** - Faculty want to be recognized (and cited) by their peers. Currently the standard for this practice is to publish in peer-reviewed journals and gather citation counts from ISI. As more and more content is made available on the Web, things like Google PageRank and links from remote documents may supplement peer-review and citation counts. Using Google's programming interface it is possible to retrieve a PageRank and lists of linking

documents. As a demonstration, this has been implemented against some of the browsable lists in the ETD collection.

6. **Thumbnail browsing** - DigiTool is used to store images for art history classes. Not only is this content fraught with copyright issues, but also the content is based on pictures, not words. As content is harvested from DigiTool we are able to calculate the location of thumbnail images on the remote server. Consequently, standard search interfaces can be provided to the content to display pictures of the hits as well as descriptions.

Raising awareness through marketing

WHILE THE GOALS OF THE IDR ARE LAUDABLE, THE IDR DOES NOT SELL ITSELF.

People are often too busy to consider change. In this developmental phase, Team IDR learned from several different sources about the necessity of raising awareness through marketing and advocacy.

Awareness of Scholarly Communication Issues

Two members of Team IDR and the Libraries' Associate Director for User Services attended an Association of Research Libraries Institute on Scholarly Communication in July 2006. Institute speakers recommended creating a systematic plan for increasing scholarly communication effectiveness. Plan objectives would include:

1. Raising awareness
2. Enabling people to comprehend the issues
3. Taking ownership
4. Promoting activism
5. Transforming the environment

We learned a number of things about promoting awareness, marketing, and advocacy on issues surrounding scholarly communication:

1. Know your audience
2. Focus on the change-makers such as people with authority
3. Know the audience's turf and use their issues

The following comments shared by teaching/research faculty in the group helped us begin to understand some of the marketing challenges we will face:

RAISING AWARENESS THROUGH MARKETING

- “The bosses—people making decisions—are paper-based. Junior faculty might want to participate to a greater degree but they have conflicting priorities.”
- “Faculty value models of excellence. Peer-review is very valuable even if it is not perfect.”
- “Not everybody will participate. Faculty are not computer literate. People are time challenged. Consider tenured and untenured faculty and compare them to teachers and researchers. Tailor your message to each of the four groups (tenured teachers and tenured researchers, untenured teachers and untenured researchers).”

Strategic Alignment

To become and remain successful, the IDR must have a continuous flow of new content and it must also be able to substantiate that the content is being accessed, viewed and applied by other scholars. The symbiotic relationship of these two requirements implies that marketing the IDR will concentrate both on solicitation of content and promoting the IDR as an information resource.

A review of the literature confirms that institutions have faced an uphill battle in marketing their repositories to faculty. (See Appendix G for detailed information). At least one institution responded to this obstacle by focusing its institutional repository marketing efforts on promoting and archiving student work. This approach, “students – and eventually faculty - ...develop some conception of the issues surrounding copyright, fair use, licensing, and alternative publishing models” (Nolan, Costanza, 2006), exposes future academics to the potential benefits of open access in scholarly communication.

The success potential for any marketing effort increases when the need is clearly defined and when the innovation is aligned with institutional goals. The University’s recent selection to participate in the Carnegie Academy for the Scholarship of Teaching and Learning, with emphasis on enhancing undergraduate research, creates a pointed need for a central repository for archiving and promoting exemplary student work. When considering the various paths a marketing plan for ND’s IDR might take, focus on Excellent Undergraduate Research consistently rises to the top. This direction, at least in the early phase of IDR promotion, provides a unique opportunity to present the IDR as an innovation that successfully meets its stated goals:

**THE SUCCESS
POTENTIAL FOR
ANY MARKETING
EFFORT
INCREASES
WHEN THE NEED
IS CLEARLY
DEFINED AND
WHEN THE
INNOVATION IS
ALIGNED WITH
INSTITUTIONAL
GOALS.**

RAISING AWARENESS THROUGH MARKETING

1. Make it easier for students to learn
2. Make it easier for faculty to teach
3. Supplement scholarly communications

Initial focus on Excellent Undergraduate Research does not preclude a long-term goal of marketing the benefits of the IDR to faculty. If the effort to spotlight student-produced content in the IDR is successful, it is likely that the IDR's appeal to faculty would increase.

Preservation

THE IDR IS A COLLECTION OF DIGITAL OBJECTS, AND AS SUCH IT PRESENTS A NEW CHALLENGE FOR THE LIBRARIES.

Preserving digital objects requires proactive, ongoing intervention. Analog content is inherently more stable and, even if it sits unused, can often be read 50-100+ years after it is produced. If digital objects go unused for that length of time they would become inaccessible. This is because the technology used to create and to read the digital content is ever-changing. For long-term accessibility (preservation), active human intervention is required. One widely accepted method for preserving digital objects is migration. Migration involves transforming digital materials from one format to another as software and technology change. GIF to JPEG. JPEG to JPEG2000. WordStar to WordPerfect. WordPerfect to Word. Word to PDF. Etc. In this environment preservation is a self-selecting process. It often happens as a matter of course for digital objects that are heavily used. The documents that are accessed most get migrated from format to format as needed. It is also an example of the principle alluded to by Thomas Jefferson, "...let us save what remains: not by vaults and locks which fence them from the public eye and use in consigning them to the waste of time, but by such a multiplication of copies, as shall place them beyond the reach of accident." A lesser used and more costly method for the preservation of digital objects is emulation. Emulation is a process whereby older computer technologies are mimicked to enable the display of obsolescent digital formats. This process requires access to outdated software and hardware, along with the accompanying documentation. Emulation allows for the display of content as it looked. Migration can lead to the loss of data and information since characteristics of the original, e.g. formatting, may not translate to the newer version.

When it comes to the IDR, probably the most important items for preservation are the theses & dissertations. Three years ago the Graduate School, in conjunction with the Libraries, implemented the Electronic Theses & Dissertations project. In this system graduate students are encouraged to submit their theses or dissertation electronically in a PDF format. Items sent through this process are neither being printed nor archived. They are sent to a company called ProQuest that has contracted to preserve the materials, and, if Proquest goes out of business the Library of Congress has committed to their continued preservation.

PRESERVATION

The theses & dissertations are primary literature. By definition they are unique pieces of information to the University of Notre Dame community. These materials are being regularly backed up to tape archives which are stored in an adequate environment and routinely checked to ensure that the data is still readable. In addition, the IDR employs a number of best practices for the preservation of digital materials: 1) widely used, open standards, 2) widely used document formats, and 3) non-proprietary software. These measures enable the Libraries to retrieve the content from the underlying system and more easily migrate from one form to another.

The other content of the IDR is being preserved by creating duplicate copies of the content. Kellogg working papers exist at the Kellogg Institute. Much of the Institute for Latino Studies content is in paper form. The images from the Art Image Library exist as photographic slides. Any articles, preprints, etc in the IDR may have become published works and therefore duplicated through the publication process. A more proactive approach to preserving the unique content of the Electronic Theses & Dissertations may be something to consider.

Copyright

THE ISSUES SURROUNDING COPYRIGHT ARE THE MOST CHALLENGING ISSUES FOR THE IDR.

Most people believe intellectual expressions are inherently owned by the individuals or organizations creating them. These beliefs have manifested themselves in patent, trademark, and copyright laws. Under the current copyright laws of the United States things are copyrighted as soon as they are recorded on a fixed medium, and unless the recorded works are considered works for hire, the copyrights are owned by the creators. Rights can be transferred only through written legal agreements from creators to other parties.

Much but not all of this thinking is embodied in a University policy, “University Ownership of Intellectual Property Arising from University Research”. For example, the policy explicitly states “The University owns all rights to all patentable inventions arising from University research”. This sounds very much like a work for hire. When it comes to more traditional scholarly outputs, though specific exceptions are made:

However, consistent with long-standing academic tradition, the University does not normally claim ownership of works such as textbooks, articles, papers, scholarly monographs, or artistic works. Creators, therefore, retain copyright in their works, unless they are created under a grant or sponsored program that specifies ownership rights in some entity other than the creator, they are subject of a contract modifying ownership rights, or they are otherwise addressed in this policy.

Because of the laws and policies, copyrights for things like articles, working papers, technical reports, etc. are owned by the faculty and students who wrote them until they are transferred in writing.

To legally post excellent undergraduate research to the IDR a copyright agreement had to be written and signed by undergraduates. The agreement essentially states four things:

1. I, the undersigned, am the copyright owner of this work.
2. I grant the University the non-exclusive right to freely distributed this work.

COPYRIGHT

3. I retain my ownership rights to this work.
4. I may have my work removed from the repository upon written request.

Trying to catch up with the undergraduates to physically sign the agreement was very time consuming. Time schedules did not mesh. Students left the University and email addresses did not work. The postal service delayed correspondence. With only a single exception, Team IDR secured signed copyright forms for all the examples of undergraduate research. The one exception was an example where the undergraduate had a co-author (a faculty member), who was not willing (or able) to sign the copyright form.

Working with traditional scholarly output proved to be even more difficult. Since scholarly publishers require authors to sign copyright agreements, usually transferring copyright from author to publisher, getting permission to redistribute published articles means getting signed agreements from publishers. While an increasing number of publishers allow authors to make their works available on the Web after an embargo period, this is the exception and not the rule. Additionally it is usually only applicable to personal websites, not institutional or subject-based repositories. This is why there are no articles, only dummy PDF documents, in the Engineering collections of the IDR.

This being the case, working with the Kellogg Institute and the Institute for Latino Studies proved much easier. The institutes have a mandate to disseminate their locally generated information as freely as possible. Visiting scholars are expected to write working papers and reports intended to be made accessible to the general public, and not necessarily through traditional publishing venues. These two institutes already have policies in place stating that work done there is owned by the University as well as the authors.

There is much misunderstanding regarding copyright across campus. In general creators of copyrightable materials do not know what their rights are. People do not understand that they are able to negotiate copyright transfer agreements. Copyright laws are nebulous in our current digital environment where information is increasingly bought and sold. On one hand, the Academy is about “expanding the sphere of knowledge”, in turn making it easier for people to benefit from the fruits of its labor. On the other hand, copyright is freely signed away to publishers whose primary market is the Academy.

Metadata

THE DEFINITION OF METADATA IS “DATA ABOUT DATA”.

This section outlines the role metadata and metadata creation played in the IDR.

Much of this section’s content is based on the observations and experience of a subset of Team IDR—the Working Group on Organizing and Describing the Digital Content of the University Libraries’ IDR whose report is presented in Appendix F.

Metadata can be divided into three types: 1) descriptive, 2) structural, and 3) administrative. Descriptive metadata is used to pull like things together, create homogenous sets of information, and make explicit various pieces of implicit information about content. This section focuses primarily on descriptive metadata. Structural metadata describes the physical make-up of content and will be elaborated upon in the Preservation section of this report. Administrative metadata, used to accommodate a workflow within a system, is not relevant to this particular discussion.

Team IDR spent much of its time creating descriptive metadata to describe and provide access to digital content. Some of this content was in the form of images. Some of the content was in the form of working papers and technical reports. Based on experience, different forms of content require different forms of descriptive metadata. For example, working papers and technical reports often have authors and titles. On the other hand, images of specific parts of the Vatican are not so easily described. For this reason various descriptive metadata schema have been created. Dublin Core, consisting of fifteen commonly used fields/elements is quite popular. It forms the basis of describing things in DSpace and the ETD-db. It is also an inherent component of our harvesting protocol (OAI-PMH). On the other hand, VRA Core has been created in an effort to better describe images of art and architecture. It includes a variety of additional descriptive fields/elements such as material, technique and style Period in order to meet the specific needs of users of art images. One size does not fit all, and different metadata schema were used to describe the materials in the IDR.

Through the process of creating metadata, the need for policies, guidelines, and authority control was reinforced. A guiding principle of librarianship is “Save the time of the user.” One way to implement this is through the use of name authority lists.

Authors are often cited with different forms of their names, such as: first initial and last name; full first name and full last name, etc. When searching for people's names it is important to bring together all the documents by the same person, even if their names were cited differently in different documents. Authority control identifies a single authorized form of a given personal name, corporate body, series, subject term, etc. and provides cross-references from variant forms for the user. The applications used in the IDR do not include authority control functionality. In lieu of significant software enhancements, policy and procedural manuals can be created to partially compensate for the lack of authority control features. Without very sophisticated software, computers will fail to retrieve all of the documents on a specific topic or by a specific person. This is especially true in the humanities. Creating metadata for IDR content reinforced the need for authority control, or, at the very least, the need for data-entry guidelines and policies.

The thorough application of descriptive metadata to IDR content requires specialized skills and collaboration. Based on the experience of other universities who have implemented institutional repositories, "If you build it, they won't come" was learned. Put another way, it takes more than bringing up DSpace for it to get populated with content. This is true primarily because authors do not want to spend the time doing data-entry and applying metadata. In an attempt to overcome this issue all the data-entry into the IDR was done by members of Team IDR. In fact, the data-entry people constituted the largest proportion of people working on the IDR. The skills, knowledge, and experience of these people varied significantly. Some had little or no formal metadata experience. Others had the experience of thirty years of library cataloging. Some had a complete understanding over the use of their computer. Others needed a bit of training on computer fundamentals. Most importantly, doing data-entry on behalf of authors emphasized the need for subject-specific knowledge. In order to adequately describe the content of a technical report or a working paper, the metadata specialist needs to know or have direct access to another person who knows about the subject area. "Is the name of this species of frog significant to the description of this paper, and if so, then what is that name?" "To what degree is it important to know that this paper describes the differences between a compiled computer language and an interpreted language?" Without this sort of knowledge or without easy access to a person with this sort of knowledge, the application of descriptive metadata will be incomplete at best. In short, we learned that people applying descriptive metadata need to have the following skills:

- attention to detail
- basic computer skills
- knowledge of computer file formats
- familiarity with a variety of metadata formats and their applications
- subject/language expertise

A Chinese menu of choices

THIS IS LIST OF POSSIBLE CHOICES – A MENU – REGARDING FUTURE DEVELOPMENTS OF THE IDR.

In general, things are listed a return on investment basis. Items with lower costs but higher returns are given priority. Put another way, items at the top of the list are easier to implement than items at the end of the list.

1. **Do nothing** - It is not necessary for the Libraries to do anything differently from the way things are being done now. Beyond the work done regarding the ETD's, there is no reason why we must take on additional responsibilities. This is the least expensive option, but instead of time such an option might cost the Libraries in terms of opportunities. (Costs: 0 additional people)
2. **Continue the ETD's** - The ETD process works and is increasingly being used. Continuing the course with the ETD's will cost no additional time, except if the University wants to increase awareness of its existence and if the Libraries were to implement preservation policies against the ETD content. (Costs: 0 additional people)
3. **Dedicate specific people with specific skills to any future IDR efforts** - The laissez-faire policy taken by the IDR Pilot Project generated a lot of energy. It made many people feel included and it offered new opportunities to many people across the Libraries. At the same time, many people were able to volunteer because they knew the Project was short-lived. Any additional developments regarding the IDR will require dedicated staff with specific skills. The number of staff and the type of skills are dependent on the IDR activity. At the very least a few policies and procedures about metadata and preservation ought to be written. Beyond that there will need to be communicators who liaison between the Libraries and the University, computer types to maintain hardware and software, metadata experts who can implement metadata guidelines, and possibly subject-specialists who can assign controlled vocabulary terms. (Costs: at least 1 ongoing computer programmer and/or systems administrator; at least 1 ongoing metadata specialist to implement guidelines; at least 1 ongoing liaison to facilitate communication between the Libraries and its audiences.)

A CHINESE MENU OF CHOICES

4. **Write metadata guidelines and implement them** - If the IDR continues, then a set of policies and procedures needs to be written and implemented regarding metadata. The guidelines should cover the aspects of content in the IDR, not all metadata aspects. (Costs: 1 - 3 metadata specialists for a limited period of time.)
5. **Write preservation guidelines and implement them** - Preservation issues need to be addressed sooner rather than later. If the IDR is continued, then a set of preservation guidelines ought to be written and implemented for the items in the IDR. (Costs: 1 - 3 preservation specialists for a limited period of time.)
6. **Take the Excellent Undergraduate Research to another level** - Just about everybody believes continuing the IDR in regards to the Excellent Undergraduate Research is a good idea. An implementation scenario would be the following. Build interest across campus. Have departments determine what is excellent. Have students complete an online form including name, biography, abstract, photo, and content. Save the content to a queue. Wait for signed copyright form to be received. Update item in queue with controlled vocabulary terms. Publish content. (Costs: 1 computer programmer for an ongoing basis with high startup costs but low ongoing costs; 1 metadata specialist for an ongoing basis; at least 1 liaison between the Libraries and departments with excellent undergraduate research.)
7. **Fill DSpace with Libraries content** - Maybe the Libraries should practice a bit of what it preaches. By creating a DSpace instance, customizing/enhancing its graphic design, and populating it with Libraries-created content this option will give the Libraries more experience and demonstrate to the University community what can be done regarding open access publishing. (Costs: 1 graphic designer for a short period of time; 1 computer systems administrator for an ongoing period of time; at least 1 liaison between DSpace and Libraries staff/faculty.)
8. **Work more closely with the Open Courseware Project** - The Open Courseware Project is collecting, organizing, preserving, and disseminating selected undergraduate classes taught here at the University. One of their goals is to highlight the undergraduate education available here at the University, a goal very similar to Excellent Undergraduate Research aspect of the IDR. By working more closely with Open Courseware we, the University, will be able to build on each other's strengths. For example, the Libraries can offer its metadata expertise and preservation practices. Open Courseware can facilitate communication between the Libraries and instructors. (Costs: 1 computer programmer to on an on-going basis help maintain the Dspace; at least 1 metadata specialist for a limited period of time; at least 1 liaison to communicate between the Libraries and Open Courseware.)

A CHINESE MENU OF CHOICES

9. **Build it and hope they come** - One option is to better customize/enhance our implementation of DSpace, tell people about its existence, and allow people from the University community to self-populate it. While this is a relatively inexpensive option, experience has taught us that few people will contribute. “If you build it, they don’t come.” (Costs: 1 graphic designer for a short period of time; 1 computer systems administrator for an ongoing period of time; at least 1 liaison for an ongoing period of time.)
10. **Programmatically collect faculty publications** - Importing metadata from bibliographic databases into DSpace is relatively easy. Search databases for Notre Dame faculty content. Save records to EndNote. Delete everything but “green” publications. Export EndNote data and import it into DSpace. The only additional task would be to acquire the actual article/content from the authors and import that too into DSpace. Such a process is thorough, eliminates costly data entry, and reduces metadata overhead. At the same time it would require a liaison to track down the content from authors. This could be very time consuming. (Costs: 1 librarian to search bibliographic databases on a regular basis; 1 computer programmer to import data into DSpace on a regular basis; at least 1 liaison to gather content from authors on an ongoing basis.)
11. **Explore the use of images that do not have copyright issues** - Copyright raises the biggest number of questions regarding institutional repositories. The images of the Art Image Library are no exception. By incorporating images in the public domain the Libraries and the University can investigate the various metadata and computer technological issues surrounding an image-based repository to a greater degree. Something like the Lantern Slides in the Architecture Library may be good candidates. (Costs: at least 1 person to scan slides; at least 1 person to investigate metadata issues; at least 1 computer person to manage hardware/software combinations such as DigiTool; at least 1 person to oversee the creation of user interfaces.)
12. **Establish relationships with Notre Dame Report to capture faculty output** - Much of every issue of Notre Dame Report includes citations listing the scholarly output of Notre Dame faculty and staff. This content could be manually entered into a bibliographic database program like EndNote or programmatically entered via the current Notre Dame Report form. Once acquired, the citations could be added to DSpace. Someone would then have to acquire the actual content. (Costs: at least 1 liaison to communicate between the Libraries, authors of content, and Notre Dame Report; 1 computer programmer; maybe 1 person to do data-entry.)
13. **Raise copyright awareness across campus and market IDR services** - People’s awareness of copyright issues are all over the map. Some people’s conceptions are out-dated. Some people’s conceptions are ignorant of the

facts. Few people have a holistic view of what copyright can and does do. By making more people aware of copyright issues, and by making people more aware of the possibilities for institutional repositories, the goals of institutional repositories can be more easily obtained. Unfortunately, such a process requires a great deal of time. Educating and bringing up to speed the entire campus on such a complex issue cannot be done in short order. (Costs: at least 1 full-time communicator person who is extremely knowledgeable about copyright issues and spends at least a few years doing it.)

14. **Systematically capture conference presentations happening on campus -** Every year a number of conferences are held here at the University. As a part of the scholarly communications process, these conferences produce a number of tangible items lending themselves to inclusion in an institutional repository: audio and video recordings, various manifestations of presentations (PowerPoint slides, Word documents, PDF file, etc.), posters, etc. In an effort to collect, organize, archive/preserve, and then re-distribute these items for the benefit of the conference attendees as well as the wider scholarly community, the Libraries could systematically deposit these materials into a repository. Again, the hardest part of this task would probably surround the copyright issues. (Costs: at least 1 communicator/facilitator to liaison between the Libraries and the conference leaders, at least 1 metadata person to do data-entry or lead presenters to do their own data-entry.)
15. **Systematically discover and incorporate faculty home page content/links -** Many faculty create Web pages for themselves, and on these pages they make available versions of their publications, even though this practice may be a breach of contract. In an effort to more systematically highlight the scholarly output of the University, academic departments and authors, the Libraries could identify faculty websites, incorporate their data into a DSpace instance, and provide enhanced services against the content. Creating the technical infrastructure would not be too difficult. Doing the data-entry would be a large and tedious task. More importantly, without copyright clearance, the Libraries (and therefore the University) would be increasing its liability and exposure to risk. Getting the copyrights from publishers and/or getting authoritative content from authors may be more expensive than the computer system maintenance and data-entry combined. (Costs: 1 part-time computer person to create and maintain the technological infrastructure; 1 part-time to full-time metadata person to identify and do data-entry against content; 1 part-time to full-time person to gather published content and acquire copyright permissions.)
16. **Take the Art Image Library & DigiTool project to the next level -** While many of the Art History faculty already have created sets of images and incorporated them into PowerPoint presentations for their classes, there will only be an increasing need for an art image database. This database should

enable instructors to supplement their existing teaching tools as well as empower students to enhance their learning opportunities. To create such a database in the Art Image Library using DigiTool, a number of things will need to happen: 1) articulate and write down a collection development policy, 2) exhaustively and systematically describe the images, 3) thoroughly understand how to use DigiTool, 4) save the images and the metadata to DigiTool, 5) create user interfaces against the DigiTool application, and 6) go to Step #1. The most difficult tasks in this process are the metadata creation and the full exploitation of DigiTool. (Costs: at least 1 part-time person to articulate a collection policy, at least 1 additional part-time to full-time person with art history subject expertise to create metadata, at least 1 part-time to full-time computer person to thoroughly understand DigiTool, at least 1 part-time graphic designer to create user interfaces to DigiTool.)

17. **Work more closely with existing “publishers” across campus** - A number of groups across the University formally publish content. The University Press. The Institute for Latino Studies. A few faculty who edit journals. These groups have similar goals, one of which is to disseminate the intellectual capital created here and across Academia. Through a process of building stronger relationships with these groups, focusing on our commonalities, and learning to exploit the Internet as a distribution medium, the Libraries could play a larger role in the scholarly communications process. It would be entirely possible to create a synergy using the strengths of each group to create a thing whose sum is greater than its parts. The downside of this process will be each group’s ability to focus on similarities and not differences. A challenge in any environment. (Costs: at least 1 communicator person spending at least one year building relationships, at least 1 computer person to create and maintain systems meeting everybody’s needs.)

Recommendations

THE MEALS

The previous “Chinese menu” of choices was a list of possible future directions, but implementing everything on the list is impractical. The University, and the Libraries in particular, can simply not afford to do it all. Assuming the Libraries has three to four people with the necessary skills who are willing and able to spend at least some of their time on future IDR efforts, the Chinese menu has been grouped into the following “meals”, and Team IDR recommends the meal called “The Safe Bet”.

Do Nothing – This option includes only menu option #1 where we, the Libraries, do nothing differently from what we are doing now.

The Necessary Appetizers – Each of the following “meals” requires resources and sets of guidelines. Without these items any future efforts will be spent inefficiently. This “meal” is a prerequisite for every other meal on the list. They are the key ingredients and include:

- Dedicate specific people with specific skills to any future IDR efforts
- Write metadata guidelines and implement them
- Write preservation guidelines and implement them

The Safe Bet – This is the recommended meal. It represents what Team IDR believes is most implementable with the resources at the Library’s disposal. Along with The Necessary Appetizers, it includes:

- Continue the ETD’s
- Take the Excellent Undergraduate Research to another level

RECOMMENDATIONS

The Safe Bet, Plus – With a bit of extra effort the Libraries may be able to add on to the Safe Bet and build synergies with an additional partner:

- Continue the ETD's
- Take the Excellent Undergraduate Research to another level
- Work more closely with the Open Courseware Project

DigiTool – The Libraries is dedicated to implementing DigiTool, and here is a “meal” for making that happen:

- Explore the use of images that do not have copyright issues
- Take the Art Image Library & DigiTool project to the next level

The Low Road – This “meal” represents the path many North American libraries have taken in regards to open access publishing. While popular, it does not seem to be very effective:

- Fill DSpace with Libraries content
- Build it and hope they come

Aggressively Capture Content – This “meal” will definitely highlight the academic and scholarly work produced by the University. This meal is also avant garde. By choosing this set of options the Libraries will be definitely demonstrating ways it can use its traditional skills to exploit the current technological environment. It is also very expensive:

- Programmatically collect faculty publications
- Establish relationships with Notre Dame Report to capture faculty output
- Systematically capture conference presentations happening on campus
- Systematically discover and incorporate faculty home page content/links

The High Road – This final “meal” is heavy on advocacy and raising awareness. It focuses on educating people regarding the related issues and banking on changing their behavior accordingly. It includes:

R E C O M M E N D A T I O N S

- Raise copyright awareness across campus and market IDR services
- Work more closely with existing “publishers” across campus

Appendix A: Team IDR and the people they worked with

MANY PEOPLE FROM INSIDE AND OUTSIDE THE UNIVERSITY LIBRARIES WERE INVOLVED WITH THIS FIRST PHASE OF THE REPOSITORY.

Within the Libraries Team Institutional Digital Repository consisted in the following people:

Julie Arnott (Preservation)

Laura Bayard (Documents Access & Database Management)

Andy Boze (Desktop Computing and Network Services)

Pascal Calarco (Library Systems)

Clara Enriquez (Documents Access & Database Management)

Robert Fox (Digital Access and Information Architecture)

Beata Frelas (Business Information Center)

Anastasia Guimaraes (Catalog and Database Maintenance)

Tom Hanstra (Library Systems)

Mandy Havert (Library Systems)

Christine Johnson (Catalog and Database Maintenance)

Diane Kennedy (Reference)

Mary Lehman (Electronic Resources and Serials Access)

Pat Loghry (Cataloging)

APPENDIX A

Natalia Lyandres (Serials Cataloging)

Denise Massa (Art Slide Library)

Laurie McGowan (Reference)

Mary Claire McKeown (Cataloging)

Eric Lease Morgan (Digital Access and Information Architecture, Leader)

Trudie Mullins (Collection Development)

BobbieLou Philotoff (Serials Cataloging)

Christian Poehlmann (Business Information Center)

Elaine Savely (Digital Access and Information Architecture)

William Sill (Desktop Computing and Network Services)

Felicia Smith (Documents Access & Database Management)

Marsha Stevenson (Department of Arts & Architecture)

Lisa Stienbarger (Serials Cataloging)

Ladonna Weeks (Government Documents Technical Services)

Nelson Weindling (Serials Cataloging)

At one point there was a call for volunteers to help with some data-entry and a few more people from within the Libraries came forward:

Judith Ann Conner (Acquisitions)

Alice Frost (Acquisitions)

Christeena Listenberger (Acquisitions)

On various visits to campus academic departments a number of library liaisons participated:

Carol Brach (Engineering Library)

Hector Escobar (Reference)

APPENDIX A

Laura Fuderer (Collection Development)
Scott Van Jacob (Collection Development)
Rob Kusmer (Cataloging)
Dan Marmion (Information Systems and Digital Access)
Margaret Porter (Reference)
Kathryn Ryan-Zeugner (Reference)
Linda Sharp (Reference)

The institutional digital repository effort is a joint effort between people who work in the Libraries and people who work outside the Libraries. This is a list of people, sans the authors, who have contributed to the Repository effort:

Charles Barber (Art, Art History)
Terri Bays (Kaneb Center)
Kevin Bowyer (Computer Science)
Gary Bernstein (Electrical Engineering)
Randy Coleman (Art, Art History)
Jim Collins (Film, Television, and Theater)
Caroline Domingo (Institute for Latino Studies)
Michael Edwards (Intellectual Property Committee)
Tim Flanagan (General Counsel)
Tracy Grimm (Institute for Latino Studies)
Martin Haengi (Electrical Engineering)
Doug Hall (Electrical Engineering)
Susan Hartel (School of Architecture)
Shari Hill (Graduate School)
Kara Kelly (School of Architecture)

APPENDIX A

Peter Holland (Film, Television, and Theater)

Dennis Jacobs (Office of the Provost)

Jeff Kantor (Intellectual Property Committee)

Kathee Kiesselbach (Nanovic Institute)

Nick Laneman (Electrical Engineering)

Greg Madey (Computer Science)

Jessica McManus Warnell (Gigot Center for Entrepreneurial Studies)

Sara Method (Art Slide Library)

Steve Moriarty (Art, Art History)

Melissa Paulsen (Gigot Center for Entrepreneurial Studies)

Kathleen Pyne (Art Slide Library)

Tim Ready (Institute for Latino Studies)

Gretchen Reydams-Schils (Institute for Scholarship in the Liberal Arts)

Robin Rhodes (Art, Art History)

Kelly Roberts (Kellogg Institute)

Charles Rosenberg (Art, Art History)

Mark Schurr (Anthropology)

Sharon Schierling (Kellogg Institute)

Yaakov Sloman (Electrical Engineering)

Elizabeth Spencer (Intellectual Property Committee)

Barb Turpin (Graduate School)

Michelle Whaley (Biological Sciences)

Appendix B: Copyright environmental scan

COPYRIGHT AND THE NOTRE DAME IDR: AN ENVIRONMENTAL SCAN AND RECOMMENDATION FOR FURTHER ACTION.

Pat Loghry and I recently investigated the state of copyright as it affects universities, faculty, and students in general as well as at the University of Notre Dame. In the course of our investigation we examined the copyright policies at numerous State and Private institutions of higher education, met with a copyright specialist at Indiana University, and attended a conference on copyright at Ball State University.

A quick scan of the legal literature as well as conversations with copyright experts quickly led us to conclude that, counter intuitively, copyright is not a well-settled area of the law. University policies are sometimes vague, and vary from institution to institution. We also discovered that copyright as applies to faculty creative work is a particularly ill-defined area of law. There is no real consensus on whether the Copyright Act's work-for-hire provision applies to faculty or graduate students for that matter. Packard and Dreyfuss note that many scholars assert that faculty output is work-for-hire and thus universities may assert ownership of the rights assigned in copyright. Conversely, Reichman seems to indicate that in practice, little has changed in the 1976 act from prior practice. There was some concern among the speakers at the Ball State conference that this was an ill-defined area and the work-for-hire provision may apply to graduate students in some circumstances.

We also investigated policies at other institutions. We did this primarily by reviewing college and university web sites for public information. We discovered that most institutions of higher education spelled out their policies on intellectual property most clearly when addressing patents. Some had exceptionally good clear policies and guidelines on copyright, but this was not always the case. MIT is a particularly good example. <http://web.mit.edu/ist/topics/security/copyright/> In addition to providing clear and easy access to MIT's copyright policies, this website also contains links to other useful sites addressing copyright.

APPENDIX B

The University of Notre Dame has posted its intellectual property policy on the web. <http://www.nd.edu/copyright/> This policy is reasonably clear:

The University owns all rights to all copyrightable materials (including computer programs, software, or multi-media productions) that are works made for hire under copyright law or that are required to be assigned to the University by the contract terms of a grant or sponsored program. However, consistent with long-standing academic tradition, the University does not normally claim ownership of works such as textbooks, articles, papers, scholarly monographs, or artistic works. Creators, therefore, retain rights in copyright in their works, unless they are created under a grant or sponsored program that specifies ownership rights in some entity other than the creator, they are the subject of a contract modifying ownership rights, or they are otherwise addressed in this Policy.

The University will ordinarily waive its rights to intellectual property created by student creators where the use of University facilities, equipment, or other resources has been properly authorized, except when:

- faculty or staff involvement is substantial,*
- the work is part of a larger University work or specifically commissioned by the University,*
- the use of facilities, equipment, or other resources is substantially in excess of the norm for educational purposes, or*
- the intellectual property resulted from the student's employment with the University.*

Informal discussions with Notre Dame faculty, staff, and students quickly led us to believe that these members of the university community have little understanding, on the whole, of how copyright affects their work. Often times there was the perception among the creators of the work, particularly exceptional undergraduate research, that they retained copyright to their work when in fact it was unlikely that this was truly the case. The most frequent scenario for losing full rights seemed to occur when a student (or even faculty member) submitted a paper for publication and signed away rights to the article. Excellent undergraduate research is often submitted for publication, and business plans (a variation of undergraduate research for Business students) are often submitted to contests.

Our recommendations are as follows:

- There should be links to copyright information on the IDR web page, but these links should be limited to the US Copyright Office <http://www.copyright.gov/> and the University of Notre Dame policy <http://www.nd.edu/copyright/>

APPENDIX B

- A Creative Commons like license, with a checkbox, should be developed for those wishing to deposit items in the IDR. There may be someone on the IDR team who can do this in conjunction with the University Counsel's office.
- Some method should be developed to maintain an archival copy of these agreements when a users submits materials to the IDR
- The University Libraries should facilitate or support, rather than offer, informational or instructional sessions for the Notre Dame community on copyright issues
- Team IDR should develop list of useful copyright sites

Referenced Works

Copyright or Copy Wrong: An Analysis of University Claims to Faculty Work. Ashley Packard. 7 Comm. L. & Pol'y 275

The notion of scholarly authorship is complicated, however, by the Copyright Act's work-for-hire provision. The Copyright Act defines a work made for hire as "a work prepared by an employee within the scope of his or her employment." It further states that "In the case of a work for hire, the employer or other persons for whom the work was prepared is considered the author for purposes of this title, and unless the parties have expressly agreed otherwise in a written instrument signed by them, owns all of the rights comprised in the copyright."

Many scholars have interpreted this to mean that faculty, as university employees, are creating works made for hire when they produce academic materials and scholarly articles, and that the university, as employer, is actually the owner of those works.

Computer Programs As Applied Scientific Know-How: Implications of Copyright Protection for Commercialized University Research. J. H. Reichman 42 Vand. L. Rev. 639

Yet this result seems rather surprising under a Copyright Act that aimed to strengthen authors' rights, especially when the legislative history conveys no intention to disturb the "teachers' exception" to the works-for-hire doctrine that was firmly established under the 1909 Act. To equate a general duty to write with a duty to produce specific works for a university distorts the nature of academic employment and downgrades the professorial rank to that of an ordinary staff member. Professors are expected to advance their universities' reputations for high-quality scholarship by publishing suitable research of their own, largely because scientific publications enhance the ability of these same universities to attract the kinds of funds and personnel that ensure fulfillment of their educational missions. That professors are attracted to teaching because research support is provided or that they write to obtain tenure and retain its full benefits hardly entitles a university to regard itself as the author of a scholarly product over which it has exercised no direct supervisory control whatsoever. In ascertaining the scope of university employment for purposes of copyright ownership, moreover, one finds no evidence to suggest that trade usage had altered the well-established teacher exception at the time the 1976 Act was adopted.

The Creative Employee and the Copyright Act of 1976. Rochelle Cooper Dreyfuss 54 U. Chi. L. Rev. 590

The copyright statute enacted in 1976 has modified prior law. Under the Copyright Act of 1909, courts and commentators regarded the work for hire doctrine, which deems an employer the owner of work prepared within the scope of employment, as largely inapplicable to teachers. Commentators have, however, argued that Congress eliminated this exception to the work for hire doctrine when it passed the Copyright Act of 1976. According to these commentators, the 1976 Act permits universities to claim copyright to, and even "authorship" of, their faculty's output.

Appendix C: Copyright and Due Diligence for the IDR

AFTER MUCH DISCUSSION WITH OUR LEGAL COUNSEL, I WOULD RECOMMEND THAT THE LIBRARY REPOSITORY INSTITUTE THE FOLLOWING POLICIES FOR COPYRIGHT COMPLIANCE AND DUE DILIGENCE:

-
1. As many copyright permissions as possible need to be obtained from the person submitting the content, prior to acceptance of that content. If there are materials other than original text, such as photographs, audio/video clips, data sets etc. there must be an acknowledgment that permission to use these has been granted and if possible a copy of the permission should be included with the submission. If there is not a copy of the permission, then the repository should confirm that it has permission to mount the item.
 2. For materials that have already been submitted to the repository with out the prerequisite permissions, obtain latest address from the alumni office and attempt to obtain permission for the use of the item. If we are unable to obtain permission for use from the copyright holder, then the item should **NOT** be included in the repository. Should there be an item that by its nature is valuable to the content of the repository and we are unable to obtain the permission for its use, a case can be made to counsel for the inclusion of the item. Council will then assess the possible risks for the use of each particular item and then the repository will need to determine if the risk is an acceptable one before the content is included in the repository.
 3. Since permissions are required before the content is submitted to the repository, it is suggested that the repository develop an online compliance form that may be printed for the permissions file and included with the content. Using the copyright clearance format, the form should allow the author to check the type of submission as well as the type of permission that she/he is granting the library.
 4. A compliance file should be maintained for the repository that contains one signed original form that would be housed in the compliance office. This file

APPENDIX C

would also contain copies of any quoted/included content permissions for materials that the author included but is not his/her original work.

Appendix D: University of Notre Dame Art Image Library Copyright Report II

Prepared by Denise J. Massa, May 19, 2006

OBJECTIVES: COPYRIGHT POLICES, GUIDELINES OR STATEMENTS REGARDING THE IMPLEMENTATION OF AN ONLINE IMAGE LIBRARY.

DigiTool, an online image library enables the University of Notre Dame, through the Hesburgh Library, to manage and preserve the locally administered digital collection of the Art Image Library. It is expected to potentially house over 200,000 images as the digital collection continues to grow over the years.

The Art Image Library plays a central role within the development of digital images for classroom instruction. It has served the education interest of professors in the Art, Art History and Design Department and the Notre Dame community. The principle objective for the Art Image Library is to continue to create, use and manage copyrighted works of art through an online image library. In corporation with fair use in the electronic age, the Art Image Library recognizes the importance and purpose of copyright in conjunction with appropriate use of copyrighted works. In practice, there is a requirement that we supply full source information, to the extent that we can, to our Copyright Office for obtaining permissions.

Statement

The use of the Art Image Library DigiTool Database and its contents, in good faith adhere to copyright restrictions. All images and text are to be used for educational and teaching purposes only. The contents of the database may only be used for comment, criticism, review, analysis, discussion and other similar purposes associated with instruction. Images and image text from DigiTool may not be shared with other educational institutions. At no time is it permissible to use images or image text for commercial/business purposes or for them to be modified, adapted, redistributed or republished. The user has the choice of clicking on the "I Agree" or "I Do Not Agree" buttons to enter the website. If that user clicks on "Don't Agree," he or she is then sent to the copyright code website or legal counsel of the University of Notre Dame

APPENDIX D

Section A. Acquisition and Use of Digitized Images for Art, Art History and Design Department Courses

(Based on Image Collection Guidelines: The Acquisition and Use of Images in Non-Profit Educational Visual Resources Collections prepared by the Visual Resource Association Committee on Intellectual Property Rights)

The Art Image Library will collect digitized images for use in specific art, art history and design courses using the following approach.

Acquisition -

Pre-existing analog image collection. The Art Image Library will identify images in its collection specifically needed for digitizing, and make a reasonable effort to identify rights holders and seek permission for use via the University Copyright Office.

Attribution -

- A. Department will make reasonable inquiry to identify rights holders and seek permission or determine whether the image is in the public domain.
- B. To credit sources for digitized images, the Art Image Library will maintain an attribution file with Digitool to include creator, title, date, publisher and other pertinent information.

Display -

- A. Digitized images will be maintained as an Art Image Library collection on a local server in a file accessible only by faculty, staff and students of the Art, Art History and Design Department and then within the University. The attribution file will be maintained with this collection file.
- B. Separate files will be set up for each individual class using digitized images, including an attribution file for the images used. These separate class files will be accessible only to students registered for the class (password and/or user log on agreement) and only for the semester during which the class is taught. Faculty and staff will have access to the images for use in class or for other professional presentations. Students may have access to the images for presentations for class assignments.

Section B: Campus Copyright Rights and Responsibilities

What are the copyright restrictions?

Rights pertain to the original sources and the digital images. The Art Image Library will apply fair use principles to the use of copyright materials. The images may be used for

APPENDIX D

educational purposes only: study, teaching, scholarly research. They may not be downloaded, copied, retained, printed, shared, modified or otherwise distributed except as provided for by permitted educational uses. They may not be re-published in any way, including mounted on an unrestricted web site.

Copyright Access Statements for Viewing Images

- Outside the University of Notre Dame Domain –
- You are not authorized nor permitted to view this image.
- Restricted for Notre Dame IP address only.

Section C: Current Practice

Brown, Dartmouth, Duke, Emory, Georgetown and Johns Hopkins (identified as library peer institutions in a report to the Provost in 1994) have access to digital collections controlled by username and password for students enrolled in a class. A copyright information page can be linked stating what a user can and cannot do with the materials without obtaining permission.

The VRA Committee on Intellectual Property Rights developed the statement ‘Image Collection Guidelines: The Acquisition and Use of Images in Non-Profit Educational Visual Resources Collections.’ (<http://www.vraweb.org/copyright/guidelines.html>). This section of the guidelines (under A. Acquisition, Section 3) is particularly relevant to our discussion:

3. Images created by copy stand photography and scanning from published materials for inclusion in the permanent archive are subject to the following considerations:
 - a) images of suitable quality are not readily available at a reasonable cost and in a reasonable time from any of the options listed above
 - b) images will not be shared between or among other educational institutions if such use is prohibited by the terms of their acquisition.
 - c) images will be used for comment, criticism, review, analysis, discussion, or other similar purpose associated with instruction or scholarship
 - d) images will be used for purposes that are both nonprofit and educational. If these conditions can be met, it is likely that making images and digital files from published materials will be within “fair use” as outlined in the Copyright Act of 1976.

Section D: Educational Materials About Copyright Law (Notre Dame’s Copyright Office)

For more on the fair use principles and copyright see:

Fair Use of Copyrighted Materials, University of Texas
<http://www.utssystem.edu/ogc/intellectualproperty/copypol2.htm>

Fair-Use Issues, Copyright Management Center, IUPUI
<http://www.copyright.iupui.edu/fairuse.htm>

APPENDIX D

Resources - Intellectual Property Rights/Copyright, VRA (Visual Resources Association)
<http://www.vraweb.org/resources.html>

Copyright & Art Issues, Christine L. Sundt, University of Oregon
<http://www.uoregon.edu/~csundt/copyweb/>

Appendix E: Permission to Use Work

I _____ having produced an original work entitled _____
_____ (the “Work”) hereby grant the University of Notre Dame du Lac, for inclusion in a publicly accessible digital repository, permission to reproduce, publish, have published, display, perform publicly, or distribute in any medium and any location this Work, or any portion thereof. I represent that the Work is my own original product, that I own all rights in the Work, and that the University’s use of the Work hereunder does not infringe anyone else’s rights. I represent that I retain all rights in the Work and that I have verified my authority to grant the permission hereunder, including in the case that the Work has been published or otherwise included in any other venue in any medium. I acknowledge that the University has agreed to include the Work in its digital repository in reliance on my representations. I hereby fully and forever discharge and release the University from any claim for damages of any kind arising out of the use or publication of the Work by the University, and agree not to sue or otherwise initiate legal proceedings against the University for such use. My permission to use the Work as set forth herein is non-exclusive to the University of Notre Dame and is royalty-free, and it may be revoked only on written notice to the University, in which case the Work will be removed from the repository within a reasonable amount of time.

Signature

Date

Printed Name

Appendix F: Metadata report

REPORT OF THE WORKING GROUP ON ORGANIZING AND DESCRIBING THE DIGITAL CONTENT OF THE UNIVERSITY LIBRARIES' IDR

PRELIMINARY DRAFT
JUNE 26, 2006

Prepared by
Anastasia Guimaraes
Mary Lehman
Natasha Lyandres
Mary McKeown
Lisa Stienbarger

University Libraries of Notre Dame

Scope

The purpose of this document is to provide a preliminary assessment of the issues and challenges identified thus far in describing, organizing and providing intellectual access to the information and digital objects within the Libraries’ Institutional Digital Repository (hereafter, IDR). While there are many different activities which are critical and necessary for the creation and maintenance of the IDR, this document will focus on identifying what must be in place in order to provide the kind of metadata which will make the IDR a truly valuable and usable information resource for the University and the academic community at large.

Role of Metadata

Metadata is most simply defined as “data about data.” In order for users to identify and access what the IDR contains which will be of interest to them, the individual objects within the IDR or collections of related objects must have records associated with them which contain various details about the objects (i.e. the data about the data). The type of metadata which supports the indexing, discovery and identification of digital resources is known as descriptive metadata and is the focus of this report. Other types include administrative metadata (used in managing and administering information resources) and structural metadata (used to document relationships among digital objects).

Metadata Standards

There are a number of metadata standards in existence, each with their own particular strengths (and weaknesses). It is clear from the working group’s experience with the IDR that one size does not fit all. It is critical that each collection within the IDR be described using the metadata standard most appropriate to that collection. This sentiment has been expressed by many other institutions in documents similar to this one. ⁽¹⁾ Issues to be considered when choosing a metadata standard for a particular digital collection include:

- the subject matter of the collection
- the format of the collection (e.g. text, images)
- skills needed and available to apply the standard consistently
- how the collection will be used and by whom
- software requirements/limitations

APPENDIX F

Metadata standards in use in the IDR

1. Dublin Core

The Dublin Core metadata standard is a set of fifteen elements used to describe digital resources and facilitate their discovery. DSpace, ETDdb and DigiTool, three software platforms currently used in the IDR, all support Dublin Core. Dublin Core was created with simplicity in mind in order to allow non-specialists to easily create metadata that would facilitate the retrieval of digital content. Dublin Core uses commonly understood terminology applicable to a variety of disciplines. The fifteen core elements are:

- Contributor
- coverage
- creator
- date
- description
- format
- identifier
- language
- publisher
- relation
- rights
- source
- subject
- title
- type

2. VRA Core

The VRA Core is a data standard for the cultural heritage community consisting of a metadata element set as well as “an initial blueprint for how those elements can be hierarchically structured. The element set provides a categorical organization for the description of works of visual culture as well as the images that document them.” (2) The value of VRA Core lies in its recognition of those special attributes of artworks and visual materials that cannot easily be provided for in a simple, generic scheme such as Dublin Core. While VRA Core is more complex, it was easier to identify the appropriate VRA Core data elements to use for the content from the Art Image Library than to “artificially” construct Dublin Core elements with locally created qualifiers. The VRA Core elements are:

- work, collection, or image
- agent
- date
- description

APPENDIX F

inscription
location
material
measurements
relation
rights
source
state Edition
style Period
subject
technique
textref
title
work
type

What We Have Done So Far

The content in the IDR currently consists of 14 preliminary collections, comprising faculty publications and excellent undergraduate research residing in DSpace; Electronic Theses and Dissertations (ETDs); and the Art Image Library collection of digitized slides (this collection is still in its initial stage and will be managed using DigiTool). Several strategies for providing metadata for the latter collection have been explored. Existing metadata for the slide collection in a FileMaker Pro database proved unusable due to a lack of standardized “cataloging” techniques, no authority control over names or subjects and past practices in place to accommodate the needs of printing the data for labels adhered to the slides themselves. Attempts to map the data to Dublin Core for ingestion into DigiTool were exceedingly difficult, due both to the nature of the existing metadata and the inappropriateness of Dublin Core for the nature of the collection. The current methodology is to catalog the images anew, directly into VireoCat, a recently available free product which uses FileMaker Pro and VRA Core standards.

Appendix A includes a detailed description of each collection in addition to the standards, software applications and skills that were necessary to create the metadata for each. [Removed for the sake of brevity. –ELM]

What We Have Learned

What worked well?

Cataloging background of many data entry participants

Participants’ willingness to volunteer for data entry assignments

Participants’ willingness to expand their knowledge and learn new skills (e.g. new metadata scheme – Dublin Core; new tool – DSpace, etc.)

APPENDIX F

Participants' willingness to collaborate

Participation of computer support specialists

Ability to enrich the IDR metadata from already existing content (e.g. controlled vocabulary from records in the Libraries' online catalog; full text databases)

What did not work well?

Varying level or lack of cataloging experience, subject and/or language knowledge, and computer skills among projects' participants

Absence of documented IDR cataloging policies and guidelines, including content and data standards, for each collection

Absence of overall coordination among various metadata creation projects

Inconsistencies in data retrieval and presentation due to absence of authority control for names and subjects

Inability to migrate data from one format to another (e.g. FileMaker Pro database to Dublin Core) due to past practices

Need to create and supply data in the form of facet and term combinations due to shortcomings in the DSpace software related to data harvesting; this data often duplicates data appropriately assigned to existing Dublin Core elements Subject, Coverage, Format, or Type

Data entry in DSpace supported by dropdown menus only during initial record creation (editing of an existing record, such as to add facets and terms, becomes more time consuming and prone to error and inconsistency)

Required Skills and Knowledge for Data Entry in the IDR

Skills and knowledge required for data entry in the IDR range from basic, to intermediate, to skilled depending on the collection. Data entry for some collections may require advanced subject knowledge and/or expert cataloging skills due to the nature of the digital objects.

Staff participating in data entry should have basic computer skills; some knowledge of cataloging, metadata standards and authority control; and knowledge of various software applications in which the digital objects are created. Some files will be batch loaded into the IDR: this requires advanced computer skills, including programming.

APPENDIX F

Collections within the IDR are subject specific, making it necessary to have subject specialists available to assist in creating, or applying existing, controlled vocabularies to facilitate a robust harvest of data to create browsable lists across IDR collections.

Below is a list of the minimum required skills/knowledge for IDR data entry participants.

Attention to detail	
Basic computer skills (keyboarding, copy and paste, use of macros)	
Knowledge of computer file formats, including:	
Adobe	PDF
Postscript	
Microsoft	PowerPoint
Microsoft	Word
WordPerfect	
ZIP	
Cataloging background, familiarity with MARC format for bibliographic records	
Familiarity with Dublin Core and VRA Core metadata standards, and others that may be introduced to accommodate future collections	
Knowledge of Authority and Subject control, including use of authoritative sources	
Understanding of use of Facet/Term combinations	
Familiarity with various software applications, including:	
ALEPH	
D-Space	
ETDdb	
DigiTool	
EndNote	
FileMaker	
My	Pro
Photoshop	Library
WebCT	

APPENDIX F

For a comprehensive list of required skills for each collection entered to date, see Appendix A, IDR Collection Descriptions and Requirements. [Removed for the sake of brevity. –ELM]

Recommendations

1. Recommended elements/Content Standards

The working group recommends devising an authoritative list of IDR content standards. This document should be made available to all prospective IDR contributors regardless of their level of metadata expertise to ensure accuracy, consistency, and interoperability of IDR records. The guidelines should be developed in accordance with NISO's A Framework of Guidance for Building Good Digital Collections (2nd edition, 2004) <http://www.niso.org/framework/Framework2.html>

The guidelines should include required and optional administrative, descriptive and structural metadata elements for two types of IDR records: collection level and object/item level.

The guidelines should identify core metadata elements for all IDR collections; and core and recommended elements for each IDR collection.

Guidelines for each metadata element should include the following information:

Label (if not DC element, it should also include crosswalk to DC)

Definition and scope

Core/Recommended

Controlled vocabularies — The working group believes that certain attributes of described digital objects must be expressed using standard controlled vocabularies whenever possible. These attributes may include, but are not limited to, personal names, corporate names, genre terms, subjects. The working group recommends devising an authoritative list of IDR controlled vocabularies for each metadata element. Using controlled vocabularies can ensure semantic consistency, reduce spelling errors, and make it easier for users to search for items. The working group understands that the choice of controlled vocabularies will depend on the content of each digital collection, available resources, subject expertise, user expectations, and system capabilities.

Inputting guidelines – This section should include instructions on how to enter dates, personal and corporate names, how to choose keywords for subject access, how to provide subject access to foreign language materials, etc.

2. Controlled Vocabularies

There is currently no mechanism for linking and querying external authority files from IDR storage (DSpace and DigiTool) and search (MyLibrary) systems. Further work on developing this area is recommended.

Faceted classification as “controlled vocabulary”

The use of facet/term classification (already in use in our DSpace environment) offers an opportunity to incorporate controlled vocabulary, providing a range of subsequent benefits.

Facet/term classification is implemented in DSpace metadata records as a locally defined qualified Dublin Core element “description.ft” via dropdowns in the data entry template. (See Appendix B for facet/term combinations currently available. [Removed for the sake of brevity. –ELM]) Any expansion of the facet/term classification needs to be made in coordination with the Libraries’ vocabulary group, TERMinators, which is responsible for the oversight of the vocabulary used by the database driven website (DDW) with consideration for how it is used by other tools.

The controlled vocabulary presented in facet/term classification enables systematic and consistent harvesting for presentation as browsable groups and listings within and across IDR “collections.” (See Appendix C for an example. [Removed for the sake of brevity. –ELM])

3. Training opportunities for metadata creators

Formal training must be provided for staff doing data entry (in metadata standards, authority control, application of controlled vocabularies, software, etc.) especially if continued use is made of staff drawn from all areas of the Libraries.

4. Collaboration with contributors

For some collections, the cataloging expertise of the Libraries’ needs to be meshed with the subject expertise of contributing authors to achieve the most effective content description.

APPENDIX F

References

(1) Documents consulted detailing guidelines and “best practices” for creating metadata for digital collections:

California State Library Metadata Standards / prepared by Liz Bischoff. – September 29, 1999. – 23 p.

<http://www.library.ca.gov/assets/acrobat/metadocfinalrev.PDF>

Dublin Core Metadata best practices / CDP [Collaborative Digitization Program] Metadata Working Group. – Version 2.1, September 2005. – 66 p.

<http://www.cdpheritage.org/cdp/documents/CDPDCMBP.pdf>

A framework of guidance for building good digital collections / NISO Framework Advisory Group.

2nd ed. (2004). – 33 p. (p. 20-28 identify six metadata principles)

<http://www.niso.org/framework/framework2.pdf>

CIC metadata guidelines / CIC Metadata Portal. – version 1.0, 06/18/2004.

<http://cicharvest.grainger.uiuc.edu/dcguidelines.asp>

Indiana Digital Library metadata best practices for use of qualified Dublin Core. – August 6, 2004. – 19 p.

A chart of the DC elements and qualifiers with “comment,” “cataloging notes,” and “examples”

<http://www.statelib.lib.in.us/www/isl/diglibin/idlbest.pdf>

Name and subject guidelines for digital collections / Metadata and Digital Library Services, University of Oregon Libraries. – Last revision: Dec. 30, 2005.

Site includes link to various data dictionaries for several DSpace projects.

<http://libweb.uoregon.edu/catdept/meta/digsubj.html>

Repository metadata guidelines / [New Jersey Digital Highway] ; compiled by Rhonda J. Marker. – January 31, 2006. – 79 p.

<http://www.njdigitalhighway.org/documents/metadata-guidelines.pdf>

APPENDIX F

The importance of content standards in digital libraries / Leslie Johnston, University of Virginia Library. – 2006.

Constitutes Chapter 5 in: MyLibrary manual / [Eric Lease Morgan et al.]

<http://dewey.library.nd.edu/mylibrary/manual/ch/ch05.html>

Metadata guidelines for the Indiana Digital Library. – Aug. 2, 2004. – 6 p.

<http://www.dlib.indiana.edu/workshops/indl04/handout7.pdf>

Indexing & metadata for digital objects / Washington State Library, Digital Best Practices.

<http://digitalwa.statelib.wa.gov/newsite/collection/indexing.htm>

(2) VRA Core 4.0: Introduction. – Beta draft, Dec. 14, 2005. – 12 p.

http://www.vraweb.org/datastandards/VRA_Core4_Intro.pdf

Appendix G: IDR Marketing Plan - Draft

Prepared by Laurie McGowan
August 15, 2006

OVERVIEW

Marketing the IDR is not a separate phase of the project. Rather, marketing and implementation plans are continuously incorporated into the project as it progresses from prototype through final product. This has already occurred throughout the development process, as evidenced in the solicitation and participation of various stakeholder groups. A broad spectrum of library staff is involved in data entry, devising cataloging protocols, interacting with other campus units to solicit content, etc. This reflects best practice methodology in involving stakeholders early in the design and development of an innovation in order to maximize support.

This marketing plan is being written at the end of Phase Two. Some events/activities included here are planned before the review and assessment phase. Future marketing would require more focus and resource allocation than was contemplated earlier in this project. To ensure credibility, future marketing activities are contingent upon library and university administration endorsement of further development of the IDR. Future marketing initiatives should only proceed if there is a firm administrative commitment to sustain the IDR.

Marketing Goals

Marketing of the IDR is an enabling objective, designed to position the IDR so that it can meet its stated purpose:

1. Make it easier for students to learn
2. Make it easier for faculty to teach
3. Supplement scholarly communications

The IDR marketing objectives are:

1. Influence faculty, students, and administrators to contribute content to the IDR so that it remains current and relevant.

APPENDIX G

2. Inform a broad audience about the existence of the IDR; educate various market segments about IDR benefits to their particular group, influence faculty and students to use the IDR as an information resource;

These objectives are interdependent, i.e., the outcome from each marketing domain profoundly influences success potential in the other marketing domain. The ability to continuously recruit new content will depend to a large degree on evidence that the content is being read and cited by a broader audience. Likewise, sustainable user interest in the IDR will depend on the breadth, quality, relevance and currency of content.

IDR Marketing Leaders

Subject Liaisons

Because of their relationships with academic departments and institutes, Library Subject Liaisons are uniquely positioned to promote the benefits of a robust IDR to their T&R colleagues. Established communication channels are ideal means for educating faculty about open source scholarly publishing and IDR services, soliciting content and influencing faculty and students to make use of the IDR as an information resource. Examples of Subject Liaison marketing activities are:

- Introducing agenda items discussing the IDR in academic department/institute meetings
- Informal departmental round tables
- E-mail to department/institute members
- Sponsor IDR workshop for department/institute members
- Discussion of IDR in librarian consultations for new faculty and grad students
- Demonstrate author web page to a group of departmental faculty and staff
- Share statistical results (citations, downloads) with department faculty and staff to reinforce effectiveness of IDR.
- Add a section about the IDR to the What Can Faculty Do? page (http://www.library.nd.edu/colldev/library_budget_cuts/what_can_faculty_do.shtml)
- Include a section about the IDR in subject-specific library instruction sessions

APPENDIX G

Reference/Instruction Librarians

“The single most important criterion of an IR’s value to our faculty members is that other people find, use, and cite the work that they put into it. Even the most enthusiastic supporters of IRs will soon lose interest if this criterion is not met.” (Foster, Gibbons, 2005) The public service orientation of Reference Librarians qualifies this group to take a leadership role in promoting the IDR as a research resource. Reference librarians are also influential in the library classroom and can promote and encourage the use of the IDR as an information resource. Reference librarians interact regularly with a broad spectrum of patrons from the ND community and can also act as solicitors of content for the IDR. (Phillips, et al, 2005)

Examples of Reference Librarian marketing activities are:

- Recommend the IDR as an information resource at the point of need
- Include a segment about the IDR in library instruction sessions
- Add a section about the IDR to library promotional literature
- Design and develop IDR educational materials in print and web formats
- Sponsor general faculty/grad workshops on IDR issues
- Act as a distribution channel for IDR information from other sources
- IDR Display on bulletin boards and/or display case

Subject and Reference Librarians who agree to assume the role of marketing leader should be prepared to participate in training sessions so that they can confidently present the IDR as a vital and effective tool for scholarly communication.

IDR Task Force

If the IDR moves forward as a library-led campus project, it will require commitment of human and physical resources, and a spirit of collaboration among various library units and also between the library and the broader campus community. Marketing leaders must be conversant about functional and technical aspects of the IDR in order to effectively present the concept to T&R faculty and students. Team leaders of staff who enter content must be aware of anticipated demand levels in order to plan effective workflows. Those who create promotional and educational items about the IDR must be informed about user needs, successful applications and limitations. These examples illustrate the importance of cross-functional communication and collaboration regarding the establishment and maintenance of the IDR. The IDR has

APPENDIX G

the potential for significant organizational impact on the library. Consideration of this aspect should be part of the Task Force charge.

An IDR Task Force would be an overarching, cross-functional body for policy creation and review authority, dissemination of information, sharing of ideas, allocation of resources, technical oversight and development, and evaluation of project effectiveness. Logically this group would comprise several specialty sub-groups:

- Training (Internal and External)
- Metadata standards
- IDR services (author web pages, citation statistics, etc.)
- Marketing and Promotions
- Technical Issues
- More as needed

The Task Force would be a vehicle for timely communication among IDR leaders about issues of importance to multiple interest groups.

Team IDR members who have already established relationships with other campus departments for solicitation of content should be included in the initial membership of the IDR Task Force. Their experiences will be an asset to newer participants. If possible, several early adopters from T&R faculty should also be included in the membership. These should serve in an advisory capacity to minimize the time commitment and increase the likelihood of participation.

Challenges

A review of the literature indicates that even MIT, a leader in the development of the IR and developer of the open source software, D-Space, has wrestled with the challenge of obtaining sufficient content from faculty and students to justify the ongoing expense of maintaining a relevant IR. (Foster, Gibbons, 2005)

...what faculty members and university researchers want is to do their research, read and write about it, share it with others, and keep up in their fields But even those who are most committed to the role of professor complain of overwork, resist clerical responsibility, and resent any additional activity that cuts into their research and writing time. (Foster, Gibbons, 2005)

For this reason, it is especially important to identify and market to the so-called gatekeepers; i.e., those who control access to and may influence faculty and students

APPENDIX G

about participation in the IR. At Notre Dame, this group would include administrative staff and research assistants, who are likely to be involved in physically uploading or transferring content to designated library staff. It is imperative that this group be included in promotional efforts to overcome any perception that the IR content submission process is burdensome.

Conclusion

There is evidence that institutional repositories are gaining acceptance within the worldwide scholarly community. "In terms of the product life cycle the technology of IRs is gathering critical mass and should move to maturity within the next 5 years." (Swanepoel, 2005).

The marketing effort will take place in phases and each phase will focus on a target group. Motivating factors for each target group will be different. Marketing efforts need to be tailored to the perceived needs of each target group.

1. The first phase has already occurred with the call for participation from a broad spectrum of library staff. Excellent response indicates that there is a great deal of interest in the project and that there is ample goodwill among library staff to advance and refine the product. This is a benefit to the further development of the project, as it lends credibility and demonstrates success in recruiting people. Further marketing efforts will be facilitated by the energy from this phase.
2. The second marketing phase will be led by Eric Morgan and volunteer members of Team IDR. It will focus on recruitment of Subject and Reference Librarians to further the promotion of the IDR. Librarians must feel that participating in promotion of the IDR is likely to enhance relationships with their constituents. They must also feel comfortable that the IDR will truly be an asset for the ND community, that the library brings unique skills to the project in developing and maintaining the IDR, and that there will be continuing administrative support for the project. The development of peripheral services - such as author pages, citation records will also play a large role in making participation in the IDR an attractive venture. Both Subject librarians and Reference librarians should be trained to be able to demonstrate these features when discussing the IDR with their own target groups.
3. The third phase will involve Subject and Reference Librarians marketing the IDR to the broader academic community. This phase will require the continued support of IDR specialists, web and print materials, and a budget for sponsorship of promotional events as needed. Such events are most likely to be effective in informal departmental meetings (such as Cafecito in the Latino Studies Institute). Subject Librarians should have at least nominal budget authority to sponsor simple catering for academic department meetings. Reference Librarians should be able to integrate discussion of the

APPENDIX G

IDR into library instruction sessions and support materials without incurring additional costs. Library Instruction typically hosts at least one faculty/grad workshop during the academic year and this would be a suitable forum for presenting the IDR and appropriate applications.

More detailed marketing plans will evolve through the IDR Task Force and the efforts of marketing leaders as the IDR expands its role at Notre Dame.

References

- Foster, Nancy F., Gibbons, Susan (2005). Understanding faculty to improve content recruitment for institutional repositories. *D-Lib Magazine*. 11(1). Retrieved 6/22/2006 from <http://www.dlib.org/dlib/january05/foster/01foster.htm>.
- Phillips, H., Carr, R., Teal, J. (2005). Leading roles for reference librarians in institutional repositories: One library's experience. *Reference Services Review*. 33(3), 301-311.
- Swanepoel, Marinus. (2005) Digital repositories: All hype and no substance?. *New Review of Information Networking*, 11, (1), 13-25. Retrieved 6/22/2006 from Academic Search Premier Database site.