What's the Big IDeA?: Considerations for Implementing an Institutional Repository

ABSTRACT
Continually increasing journal costs have pushed libraries and research institutions to consider alternative forms of scholarly publication. One such form is that of the institutional digital repository (IR). As an early implementer of DSpace, an open-source institutional digital repository software product, IUPUI offers those just beginning to think about IRs an overview of issues such as: choosing a repository platform, staffing and technology needs, metadata and controlled vocabulary concerns, promotion, and time challenges. While the article outlines the process IUPUI followed to create its own IR, the piece is universalized to address the concerns of any new IR implementer.

KEYWORDS: Institutional digital repositories, Institutional repositories, Digital archives, Open Access, Scholarly communication.

Institutional Repositories (IRs) have recently become major players in the Open Access movement. Libraries across the world are clamoring to get their own IRs launched, but when it comes to getting an IR off the ground, it is not as easy as simply downloading the software and waiting for it to populate itself. There are many manpower, technical, and policy issues that need to be ironed out in order for an IR to be successful. The following is based on the experiences of the authors and other library staff at Indiana University-Purdue University Indianapolis (IUPUI) in implementing their IR, IDeA. The intention of this article is to outline some issues to consider when thinking about starting an IR.

Open Access and New Forms of Scholarly Communication

The influences behind IUPUI's decision to implement an IR were not unlike many other libraries' reasons for taking the plunge. It became a sort of perfect storm: new challenges met with new possibilities and the IR was born. The new challenges IUPUI libraries faced included journal price increases paired with budget cuts; increasing awareness of the toll of copyright restrictions on authors; increasing dissatisfaction with the speed of the research cycle; and the advent of units of scholarship that were not successfully being disseminated by traditional publication methods. Open access models of scholarly communication were able to answer these challenges with their cost-effective, copyright-friendly, and technology-savvy means. Much of the culture of IUPUI libraries seems to be custom made for the support of open access to scholarship. The libraries have prided themselves on technological innovation and library administrators have encouraged trying new things.

Deciding to Take the Plunge

Once the library administration recognized the need for something different, it was time to investigate the possible avenues of open access scholarship. While open access journals were obviously an important means of furthering open access, it was

decided that the most practical way for the library to support open access publishing at that juncture was to explore the possibility of starting an institutional repository. The possible benefits of IRs identified included:

- Provided a functional framework to collect, preserve, and disseminate scholarly content
- Provided "branding" for an institution's scholarship
- Consolidated a campus' scholarly materials in one place, making scholarship available to a worldwide audience
- Preserved content

Helped to decrease the cost of accessing scholarship (long term benefit)

A task force was convened in 2002 to study the possibility of instituting an IR at IUPUI. The task force was made up of all possible library stakeholders: catalogers, archivists and special collections librarians, digital library team librarians, operations people, and professional school librarians. The task force set out to study the different products available for IRs, as well as what it would take (money, hardware, knowledge, staff time) to implement and maintain the IR. The task force also examined IUPUI's faculty culture in relation to how an IR might (or might not!) be used.

Choosing an IR Platform

When studying the IR platforms, the task force limited itself to open source products, essentially because no commercial products were available at the time and the library was not interested in building its own from scratch. Today there are more options. When selecting a platform it is important to keep in mind an institution's assets and how material in the IR will potentially be used. Each has its advantages and disadvantages when compared closely with the IR's intended use. Proprietary repository products have customer support and ready-made infrastructure, but may lack the customization capabilities needed and will certainly come at a price. Some disadvantages to using an open source repository platform include their lack of traditional customer support, built in storage space, and automatic metadata creation. On the upside, open source software is freely available and boasts great user bases that are willing to offer support via listservs and user groups. Developing your own platform offers the ultimate customization but will also require more expertise and manpower. Regardless of whether an institution uses a proprietary, custom made, or open source product, the promotion of the IR to the campus and beyond (which has proven to be a critical piece for many IRs) will be in the hands of the institution.

DSpace

IUPUI University Library's task force ultimately suggested using MIT and Hewlett-Packard's open source software, DSpace. It is certainly not for every IR, but it fit IUPUI's needs nicely. Some of the reasons DSpace was selected:

- Open source—no cost for the IR software (though other software was needed to make it run well—most of which IUPUI already had in place)
- Pedigrees of the institutions involved in its development coupled with the impressive list of early adopters
- Good functionality and features
- "Guaranteed" preservation of many common file formats

- Use of Qualified Dublin Core metadata
- Provisions for controlled vocabulary
- Use of URIs—persistent network identifiers that eliminate online citation "decay"
- Workflow largely controlled by the groups of submitters—freeing up librarian time and giving submitters a feeling of ownership

People and Leadership

After the IR platform was decided upon, the next step involved gathering the people who would be responsible for launching and maintaining the repository. One of the most important characteristics of people who will be involved with the IR is enthusiasm—specifically in scholarly communication issues, the application of technology, or the promotion of the intellectual assets of an institution. Without interest, some people will not have the patience to stay involved in the project for the long haul. That being said, interest alone will not get an IR running. If the interested people do not have the necessary skill sets to start and maintain an IR, those skills must be found in other people. If no one in the library owns all of the skills highlighted in the next section, they will need to be found either elsewhere on campus or perhaps in new library hires.

From this collection of people, it is important to identify a leader. This ensures the project moves along and that tasks are delegated appropriately. It is sometimes difficult when the people assigned to the IR are not given release time from their current assignments to pursue the repository project. It is, however, common in the start-up of many IRs (especially in such tight budget times) and something to consider when thinking about leadership and staffing.

Skills

Some of the skill sets that will need to be considered include:

- Subject expertise
- Database design expertise
- Metadata expertise
- Information organization expertise
- Campus-wide communication expertise

Subject expertise is helpful when it comes to knowing the publishing culture of specific disciplines, as well as in identifying what campus assets reside in different departments. Database design is necessary even when not building the IR from scratch. Information organization, interface design, and database design skills all help to ensure that the IR's relational database is set up in an intuitive fashion which allows users to easily find what they need. Communication skills will be necessary to market the IR to both faculty and campus administration.

Hardware and Software

After assembling the lists of people and skills that are needed, it is necessary to look at the hardware and peripheral software that are required to effectively run the repository. This, like many other aspects of building an IR, will depend on what platform has been chosen and the intended function of the IR. At IUPUI, it was decided to use a dedicated server for IDeA. This allows staff to easily restart the IR server for upgrades and testing without disrupting other services. It also affords the preservation and

dissemination of very large amounts of data and offers the ability to host a test bed for trying out new upgrades and extensions to the IR. Another important technical consideration is how data will be backed up. It may be as simple as designating that it will be backed up using the same method as the rest of your library data, but it is important to consider nonetheless. Stability of the service and back up are important because the IR will be billed as a secure and reliable storage method. It is also important to consider the processing speed of staff and end users' computers, as well as the network speed on campus. This will affect how efficiently items will be uploaded into the IR and accessed by users.

Install the Software

As mentioned before, it is not as easy as plug and play. Most IRs are not stand alone products. For example, there are various supporting software packages that enable DSpace to function. During IUPUI's initial implementation the instructions that DSpace supplied concerning these various software applications (including Redhat Linux, Java,Tomcat, Apache, Ant, PostgreSQL) were version specific. For the sake of cost, many institutions will choose to use the version of these applications which their institution is already running, rather than upgrade to the specific version the IR product employs. Therefore the IR installation instructions may have to be modified to fit whatever particular cocktail of versions the library is currently using.

Test the IR

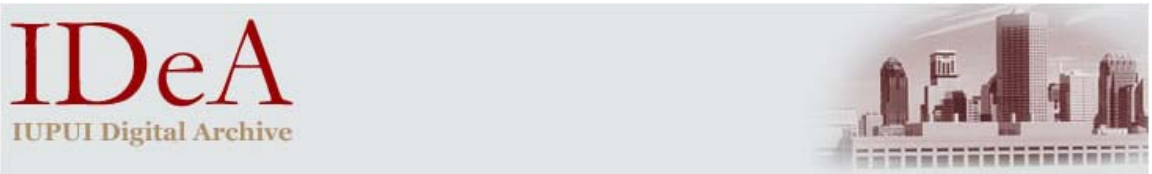
IUPUI has two instances of IDeA running at all times; the publicly accessible IR and the test instance. It is suggested that this second instance not only be used for the initial testing of the IR but for every subsequent upgrade of the software. As with any new database or service product it is important to test every aspect of the product's proposed capability. Test from perspectives of user, administrator, and any intermittent role that may be included (for example DSpace allows for general users to have some Collection level administrative duties). Consider:

- Metadata
- Ingestion process
- Searching capability
- Navigation
- Various file formats (acceptability of)

IR Interface/Design

When a library creates its own IR from the ground up the web design of the site will be an integral part of the initial planning process, but for those using out-of-the-box products it will be important to determine where a customized public interface fits into the list of IR implementation priorities. For IUPUI, customization came second to populating the repository, meaning for the first year the focus was on soliciting material to be added to the IR. It was not until year two that the interface was considerably customized.

Figure 1
IDeA Homepage header, an example of customizations that did not come until year two of implementation.



Organizational Method

The software package may decide the overall method of organization the IR, but certainly some theme of organization will still lie in the implementer's hands. Institutions will need to consider the main purpose of its IR and users. It is also important to think about access points. This is no new issue—think access at the journal level vs. access at the article level.

A few possible organizational methods:

- Hierarchical: a tree with branches—as seen in Figure 1
- Subject based: for example, a section of scholarship on women's history
- Entity based: for example, a section of scholarship from the institution's Department of History
- Format based: for example, a collection of reusable learning objects

A combination of the above methods is highly likely. IDeA is bound by the structure DSpace employs (an overall hierarchical structure) but has aspects of entity, subject, and format based organization as well. Within the hierarchy collections are primarily entity based, as are the majority of scholarship producing groups on campus. IDeA implementers have also considered format based organization for collections of reusable learning objects, electronic theses, and dissertations; and subject based organization for faculty members wishing to submit scholarship as an individual rather than as part of a campus entity.

Figure 2
Hierarchical organization example, DSpace's default organization scheme

- Community level: **School of Liberal Arts**
  - SubCommunity level: **Department of History**
    - Collection level: **Dr. Robert L. Smith Lecture Series**
      - Item level: the metadata or **catalog record**-description of the lecture
        - File(s) level: a **pdf** of the lecture, a **tiff** of an image displayed during the lecture, a **wav** file of the audio of the lecture, and a **PowerPoint®** presentation used during the lecture

In Figure 2 it is possible to recognize how the method of organization will greatly affect access points. Here the metadata record describes the lecture as a whole. If the author wanted to describe each file (pdf, tiff, wav, ppt) in-depth, then he would have to move up one level in the hierarchy. Placing each file at the Item level allows more

minute description but also loses the mechanism that allows these various files to be immediately associated with one another.

Metadata

The schema and amount of metadata chosen will be one of the main factors effecting item discovery. As metadata is intrinsically connected to the method of organization it too is oft determined by the software product chosen. DSpace employs Qualified Dublin Core, which is one of the reasons IUPUI chose DSpace. Dublin Core is Open Access Initiative (OAI) compliant and allows for easy, standard description of a broad spectrum of scholarship. Metadata considerations when selecting or creating your IR:

- Is Descriptive, Administrative, Preservation metadata needed, if so how much?
- Specific purpose of the particular institution's IR?
    - To target a particular form of scholarship such as Learning Objects? If so consider SCORM (Shareable Content Object Reference Model)
    - Be broad enough to include a variety of scholarships types and be OAI compliant? If so consider Dublin Core

Metadata Considerations once an IR is Established:

- Determine minimum required fields, if any
- How will controlled vocabulary be handled? Some options:
    - Set a specific vocabulary to be used by all
    - Require a controlled vocabulary but allow the participating groups of scholars to select a subject or form specific controlled vocabulary
    - Do not require a controlled vocabulary

Controlled vocabulary is tricky. When presenting IDeA to campus communities the importance of controlled vocabularies and how they will ultimately allow better and greater access to their scholarly research was emphasized. But controlled vocabulary creation is time consuming and often goes uncompleted if faculty are submitting content to the IR themselves.

Suggestions to Aid with Controlled Vocabulary Creation:

- Create a list of links to subject specific controlled vocabulary thesauri, making it easier for campus groups to select and hopefully use a particular subject specific vocabulary
- Create or locate a limited list of broad terms that could apply to all subject areas and include them as a pull down list at the point of subject category metadata creation

Lead by Example

Begin populating the IR by adding library related and created scholarship. This allows IR administrators to establish and solidify their procedures and to encounter the real time issues that will arise as campus members begin submitting items to the repository. For example, IUPUI established a University Library Community within its IR. This forced the IR's administrators/librarians to think about and attempt to answer questions that other campus members would also be asking, such as:

- What does 'scholarly material' mean to us?
- What kind of controlled vocabulary do we want to apply?
- Who will be in charge of reviewing the content of submitted items?
- Who will check that metadata has been completed fully and accurately?
- What copyright concerns do we have?
- How will 'publishing' in IDeA effect our promotion and tenure process?

Promotion

IUPUI did not begin its IR venture with a promotion plan in place. Having such a plan would have greatly benefited promotion efforts in the long run. "[Margret] Branschofsky has found that as many as five to seven tailored and personalized impressions about DSpace are required to catch the interest of new potential IR participants. . . Creating tailored and personalized impressions requires a thorough knowledge of how the IR could meet actual needs of people. . .Often the most persuasive arguments for submitting content to an IR come not from administrators or librarians but from close colleagues at other institutions." (Gibbons 2004, p.57).

Promotional Ideas

- Identify and contact already-published faculty members
- Ask retiring faculty to submit items
- Seek out well-known campus authors
- Research-of-the-Day, highlight a new IR submission on the homepage everyday
- Provide seminars and workshops in the library for the campus
- Present at departmental meetings
- Develop scripts for library staff to engage with campus groups
- Know "green" journals—journals that allow author self-archiving of some sort, see SHERPA at http://www.sherpa.ac.uk/ for more information
- DSpace Federation Guide http://www.dspace.org/implement/index.html has reproducible promotional material and additional promotional ideas

Innovate

Undoubtedly many libraries will use existing faculty and staff to run their IR. A major challenge IDeA administrators still face is carving out time in their already compact schedules to address the challenges of IR administration. One way in which IUPUI has addressed this hurdle is to take advantage of the wealth of knowledgeable and willing students in IUPUI's Library Science Program. In the 2005 Spring Semester IDeA administrators partnered with Dr. Mary Alice Ball's Systems Analysis and Design course. The students in this course formed groups based on the needs IDeA administrators proposed as most crucial (Improved ingestion process, Metadata: specifically for reusable learning objects, and Promotion). Not only were students able to participate in a professional, cutting edge library project, but the IDeA administrators gained well-researched potential solutions to difficult IR questions and tangible products (such as training guides to be distributed to IR submitters).

Conclusion

A common thread throughout this IR discussion has been the need to focus on the specific purpose for creating an IR at a given institution. The purpose will be based on the needs of a given campus and its scholarship producing membership. This purpose will drive the organizational method, the metadata schema, and the manner in which the IR is promoted. Indeed it is important to recognize that an IR is not right for every library or institution. This piece outlines the skills, people, philosophies, and resources required. However it also reminds the reader that success is not dependent upon an abundance of all of these components, rather a thoughtful consideration of what goals, and therefore what resources, are most important.

Visit IDeA at: https://idea.iupui.edu

SOURCE CITED

Gibbons, S. (2004), "Establishing and Institutional Repository", *Library Technology Reports,* vol. 40 no. 4.

ABOUT THE AUTHORS

Kristi L. Palmer (klpalmer@iupui.edu) is a metadata/cataloging librarian at IUPUI University Library. In addition to her work with technical services and digital collections she also provides collection development and research assistance in the area of History and Women's Studies.

Emily Dill (eadill@iupuc.edu) is Public Services Librarian at Indiana University-Purdue University Columbus. Emily is a member of the IDeA repository working group and also serves as reviewer for the University Library Community of the repository.