June 2012

A Brief Primer on Institutional Repositories

Bob Schatz

BioMed Central, robert.schatz@biomedcentral.com

A Brief Primer on Institutional Repositories

by **Bob Schatz** (North American Sales Manager, BioMed Central / Open Repository) <robert.schatz@biomedcentral.com>

My professional life resides in the world of open access. This is a fascinating place to be for a librarian who spent many years in the world of traditional publishing and bookselling, made all the more so by my now having sales responsibilities in North America for **BioMed Central's** *Open Repository* hosted repository service. I have to admit, this did not feel like a welcome assignment when I first got it. The world of repositories is not without technical and conceptual complexities which my print-oriented brain initially had a hard time adjusting to. I spent a year wrapping my head around all those complexities in order to be able to credibly discuss our service with prospective clients, and am learning more all the time. All this is good and helps keep **Bob** from becoming a dull boy.

When I talk to librarians who are asked to evaluate repositories for their institutions, it is clear they are challenged by the assignment. As with many technical products and services, a single line of inquiry frequently leads to peripheral issues making any analysis more involved than originally anticipated. In order to help my counterparts in libraries manage repository evaluations more effectively, I offer a basic primer of some of the many questions connected to selecting a repository. My intent is not to point people in a specific direction. Rather it is to help them identify the elements on which any analysis of repository services is likely to touch. By understanding the scope of the task, I hope the process of evaluation will run more smoothly and produce better outcomes.

Going into an evaluation, it will be helpful to understand and accept that repositories touch the interests of many stakeholders: librarians, faculty, students, administrators, recent PhDs, IT departments, archivists, and so on. Organizing your list of stakeholders in such a way that objectives can be set and decisions made can be daunting, but it is not impossible. Take heart, others have successfully navigated these waters and you can too. As you work through the issues identified below, think about whose interests or expertise will be touched by the questions asked and answers received. Identifying who needs to be brought into the process, what level of knowledge they have, and what their vested interests are may help you come to a faster, better decision about how your institution will provide repository services to the communities you serve.

Questions (and a few answers, but not many)

Why do you want a repository, and what will go into it? The motivation for, and intended use of, a repository is important to understand, though once in place your organization may well expand on those uses when everyone sees how the repository is working. What the repository will be used for may inform some of the decisions you make about the kind of repository you actually build or acquire. Typically your repository will be populated with works by and about your faculty and researchers, but how will you define that? Will it be just published works or will it include drafts, reports, and other unpublished or un-peer-reviewed items? If you intend to include published works, have you established any guidelines about how you will treat works that have copyright restrictions? Will you import metadata with links to external publisher Websites or only accept works for which there are no restrictions to access the full content? Will you allow authors to submit these works directly into the repository or only allow them to submit to a holding file where an authorized person can decide whether or not to accept the item into the live repository? If you can determine when a copyrighted work will be allowed to be openly accessed, do you want to import the content and hold it behind a embargo firewall until it is allowed to be accessed via the repository?

Will you include works only of your faculty or will you allow student works as well? What about dissertations? Do you want to restrict who can access the repository content or do you intend to make it open access? The answer may be both, but understanding which is more important will be useful to know. Are repository uses defined by any mandates? Are there specific objectives or timelines that have to be achieved? Will you accept just print-oriented items in the repository (i.e., documents and PDFs) or other digital object types, such as audio and video files?

The initial items to go into your repository: Once you have a feel for the potential use to which you'll put the repository, you'll need to have an understanding of what already exists and where it resides. This will help guide you in developing, either on your own or with a hosted service, how much server capacity you'll initially need and how you will pre-populate the repository. The demands of moving several hundred PDFs already residing on a university server are much different than finding and moving thousands of digital objects that may be on servers, laptops, flash drives, external Websites, and perhaps even old floppy disks stuck in the back of desk drawers.

How much server capacity will be needed will be affected by how many and what kinds of items will be going into the repository. A hundred gigabytes of server capacity may hold years' worth of documents and PDFs, but be totally inadequate to accept a collection of hundreds of large video files. Understanding the likely initial and long-term makeup of your repository's content will help you design or acquire a repository that is technically capable of meeting long-term needs. If you intend to use a hosted service, what server capacity approaches are offered? Can you grow the repository incrementally or are you required to commit to more capacity than you are likely to need? Understanding these parameters will help you plan realistically, and prepare for the associated budget implications.

What will the repository look like? Do you want to integrate the user interface (UI) with the rest of your organization's Web presence or do you want to create a totally separate look? Do you have standard logos, fonts, or colors that you are required to adhere to? Do you want a framed, columnar look or something more open? What kinds of graphics do you want to incorporate and how do you want them displayed? Does it matter where on the UI you place the search screen? Do you want any links to external sites on the UI? Do you want to display any welcome or introductory messaging? Do you need the capability to display that messaging in more than one language? What kind of help screens or support documents will your users have access to?

What kind of metadata do you want to collect? Will metadata records conform to accepted standards like Dublin Core? Will records be harvestable according to OAI-PMH protocols? Can you choose which metadata fields will be required with item submissions? Will the repository allow you to modify that for different parts of the repository? Can you change metadata tags and the order in which fields appear?

What kind of service and architecture will you acquire? Hosted or local? Open source or proprietary architecture? What version will you get and is that the latest version available? Does the architecture affect how you'll be able to move items or collections into the repository? Does it affect your options if you decide later to migrate to a different repository? Will you get user documentation and administrator training as part of your initial set-up? What ongoing documentation will you receive? Has your IT department or service provider added any additional features to the repository? What are they? How do they work? Are they included in the pricing you've been quoted or are they extra? What added features are desirable? Would you make use of faculty/researcher highlight pages if available? What kind of site/collection/item statistics can you expect? Will you get assistance in moving existing items and collections into the repository when

it is brought up? How, where, and how often will the repository be backed-up? Does that include disaster recovery back-ups?

Have you developed functional and/or technical specifications for your provider? Does it spell out and accurately reflect the objectives of the repository? Does it differentiate between desired functions and required? If you are insisting on technical functionality, like the integration of your LDAP SSO system in with the repository, have you provided specifications that will help your IT department or hosted supplier determine what it will take to comply? Are deadlines and measures of success spelled out? Have you asked stakeholders why they are insisting on certain features? Do you really need what you say you need? How do your requirements affect startup deadlines and costs? Are those acceptable? Is your provider willing to negotiate in order to get the project off the ground? Are you?

What other things do you want to consider? Can you get a trial of the repository before you make a commitment? Is it live? For how long will you have access? What information will you be required to provide to set-up a trial?

What legal agreements are associated with the start-up process? Who in your organization needs to vet and approve those documents? Are there clauses in the agreement that conflict with organizational policies? Can they be changed? Once documents are signed, how long does your provider, whether local or an outside company, need to bring up the repository you've decided on? Are there any external events that dictate when the repository has to go live?

How will you promote the repository to constituent user groups once it goes live? What will you do with all your free time when you get to hand off the repository project to someone else? (Just checking to see if you are still awake.)

What do you need to know about your provider? If you are hosting your own repository, does your IT Department have any experience doing that? Do they have expertise in the kind of repository architecture you'll be using? What other demands do they have on their time? Do they have, and will they adhere to, a schedule of back-ups (both local and third-party)? Will they be able to commit to evaluating new versions of the repository software on a timely basis? Will they commit to a regular schedule of upgrades?

If using a hosted service, what do you know about the company? How long has it existed? What expertise does it have in providing repository services? How well run is it as a business (i.e., will it be around for the long haul)? Does it have clients similar in needs and demands as your organization? Does it have other kinds of clients who may push the company to develop new and innovative enhancements you might not think of? Do the people you interact with seem knowledgeable?

While all of this may seem daunting, it need not be. By using these questions, or other ones that make sense to your particular setting, you can create a repository evaluation plan that help you assess your options, weigh alternatives and, with any luck, produce repository services which your internal and external users will find of value and will provide a showcase for the works produced on your campus or in your organization.

Good luck, and happy hunting. 🦐