Rethinking Repositories

Time to (re-)focus on the researchers?

Paolo Mangiafico
Duke University Libraries
Office of Copyright & Scholarly Communication
@paoloman

CNI Spring Meeting
April 2017

[Photo by Mark Zupan, Duke University Libraries: https://www.flickr.com/photos/dukeunivlibraries/31519764812/]
I thought it would be useful to start with one of the early definitions of IRs – from Cliff Lynch’s 2003 essay: “Institutional Repositories, Infrastructure for Scholarship”
“A university-based institutional repository is a set of services that a university offers to members of its community for the management and dissemination of digital materials created by the institution and its community members.

It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution.”

There’s a pretty big scope defined here:

[quote on screen]

And I think this is what we’ve all be striving for.
“an institutional repository
it not simply a fixed set of
software and hardware”

He also says: [quote on screen]

Yet here we are 14 years later, and in many cases we’ve treated it that way,
trying to develop a single core set of infrastructure that can accommodate the variety of materials we think should be preserved and made available,
which ends up feeling like we’re trying to squeeze many differently-shaped pegs into the same round hole.
Building repositories is hard, and especially when we aim for this kind of one-size-fits-all type of system, it’s slow going.

So it makes sense that we often end up relying on tools and processes that are customized for particular needs and particular stakeholders, and that repository programs have sometimes bifurcated in the ways described by my colleagues, rather than being singular all-encompassing efforts.

At Duke, and at Carolina and elsewhere, I think we’ve successfully accomplished the relatively easy parts of institutional repository goals, the low hanging fruit – or at least we’re well on track to accomplish them. Things like:
digitized library collections, where we have a small number of content creators in a relatively controlled environment and
traditional publications, where we’re talking about things that are relatively small and bounded, like PDFs, and that have well-defined ways to describe, cite, discover, and consume them

and where the biggest challenges are getting the rights to make them available (which is what open access policies do) and convincing authors to take the time to do the deposit (which is more difficult)
As we start to work more on making repositories serve more complex needs and a greater variety of content, this is where the road really starts to go uphill, so I think it’s natural that we’re struggling a bit.

[Image by Baker County Tourism](https://www.flickr.com/photos/basecampbaker/28125552966 used under CC BY-ND license]
“At the most basic and fundamental level, an institutional repository is a recognition that the intellectual life and scholarship of our universities will increasingly be represented, documented, and shared in digital form, and that a primary responsibility of our universities is to exercise stewardship over these riches: both to make them available and to preserve them.”

Returning to Lynch’s 2003 essay, he writes: [quote on screen]

Which to me says our focus should be on the wide variety of scholarly materials our communities are creating in digital forms, and on the emerging workflows that are becoming an increasingly important part of scholarship in the 21st century, not just on the things we ourselves have digitized or that have already been published elsewhere, and we’re just hosting an open access copy.
And yet…

our focus often tends to stick closely to the library's collecting goals, and when researchers who have their own goals come to us with their needs, we aren’t well prepared to help them.

We are on track to meet our own goals and processes, and the requests coming to us from faculty seem extraneous.
While for practical reasons repositories might have their own goals and staff and infrastructure supporting them, they aren’t really a program in themselves, any more than the library building or library stacks are a program in themselves.
So we have to ask ourselves:

How can we make infrastructure provisioning and support efficient, but at the same time not allow it to constrain the variety of programming and service needs of the broader organization, or attempt to drive them?

[ Image by Eugenio Mazzone https://unsplash.com/photos/fyywo2o2a7e used under CC 0 license ]
I don’t have answers to this, but I do strongly believe that the way forward is not IT-driven, and not repository-driven, but service-driven, and stakeholder-driven.

We need to keep a sharp focus on the key stakeholders and their goals and desires, and always work to make sure what we’re doing is meeting those targets.

We need to make sure that we keep a focus on the repository as a means to an end, not an end in itself.
And I think that circles us back to the question of who these services ultimately are for?

- are the libraries themselves the key stakeholder? For the preservation aspects, probably so, since most everyone else isn’t thinking that long term, or at least isn’t willing to invest in the long term like libraries are.

- but for almost everything else, especially for materials created by members of our communities – the scholarly outputs of the institution – we have to meet their immediate needs, their short term incentives, their collaborative workflows (and usually pretty messy processes), and their timeframes.

[ Image by UK Chevening Scholarships https://www.flickr.com/photos/66669315@N07/9362061241 used under CC By license. ]
To be successful at this we probably need to think more like startups, who focus first on building a community of engaged users, whose needs are being met by the program, and then build the things that we care about behind the scenes.

We need to show immediate value, give people a reason to care, and a reason to come back, and a reason to tell their colleagues about it and invite them to use it too.
• Tools and services must be in tune with researchers’ workflows, which are often discipline-specific (and sometimes even project-specific).
• Researchers resist top-down and/or mandatory schemes.
• Researchers favour a “cafeteria” model in which they can pick and choose from a set of services.
• Tools and services must be easy to use.
• Researchers must be in control of what happens to their data, who has access to it, and under what conditions. Consequently, they want to be sure that whoever is dealing with their data (data centre, library, etc.) will respect their interests.
• Researchers expect tools and services to support their day-to-day work within the research project; long-term/public requirements must be subordinate to that interest.
• The benefits of the support must clearly visible - not in three years’ time, but now.
• Support must be local, hands-on, and available when needed.

I remember about 5 years ago coming across this report titled “What researchers want” that summarized what they learned in this way.

This was specifically related to research data services, but I think the lessons here apply to any of the other things we’re trying to do with repositories that are aiming to archive and make available materials created by our researchers.

[quote on screen]
How well are we doing with most of these?
What would we need to change to better meet these goals?
The researcher focus has to be reflected in the language we use too.
For example, let’s stop calling it an “institutional repository” when we talk to faculty, since that resonates with hardly anybody.
Should we call it an archive? Or digital publishing? Or something else that conveys what it does for the stakeholders?
At Duke, when talking with faculty about our open access program and the deposit mechanisms for it, even though the materials ultimately end up in and are served from a repository, we seldom use that term.

Instead, we say something like “make your work available to anyone who wants to read it, via your own profile and a Google or Google Scholar search” and “make sure that your work remains accessible and citable long into the future”.

We’ve integrated publication deposit into the workflow faculty use to update the profiles that appear on the university’s web site and their department and school web sites, so the deposit isn’t an end in itself but part of a larger effort where they have other incentives to do it.

I don’t have time to demo that in this session, but can refer you to a webinar I did on this in January where I get into it in a lot of detail. To download the slides or watch the video, google Duke OCLC RIM webinar.
We also need to do better at showing value by reflecting back indicators of use and impact.

- things like download stats are helpful, especially if we can customize them to the individual or departments or other org units, to show them stats specifically related to things they've created or are interested in

- being able to show how many other works cite your work, or who is saying what about it in different contexts – people are really interested in that, and it creates a positive feedback loop

- this is why you see things like Altmetric badges in Figshare or tools like that.
And I really like what Harvard has done with their DASH Stats package layered on their DSpace repository.

And in the session yesterday about Michigan’s Deep Blue repository Jim Ottaviani showed some impressive stats collecting they’re doing, which they reflect back to their stakeholders.
None of these ideas are really new – those of us who have been following the development of institutional repositories for some time probably remember these papers.

[papers on screen]

They highlighted the idea that we need to better understand and meet researcher incentives in order to encourage them to engage.
And even before that, there was what’s called the Del.icio.us Lesson, which Joshua Porter wrote about in 2005, reflecting on the success of sites like Delicious and Flickr:
The lesson was:

Personal value precedes network value.

What this means is that if we want to build something that many people will actively engage with, and where the aggregation of their engagement is what builds the real value, then each contributor needs to get individual value from participating.

Otherwise they have little incentive to spend their time contributing to it.

It needs to do something useful for them.
What value are we providing to our individual stakeholders within the university?

Returning to institutional repositories, here too

We need to provide a service that’s compelling to our users, that solves their problems, saves them time, meets their needs,

and then build the network value, the institutional value, as a beneficial side effect of meeting the needs of many diverse individual actors
If we don’t do this, we risk building elegant but empty cathedrals of bits,
systems perhaps full of carefully curated content
but void of people who feel like they own the space, and feel comfortable in it, and that it does something for them.

[ Image by Ed Triplett, Duke University: https://today.duke.edu/2016/05/3dchapel ]
These are all things that most of us have thought about and probably planned for, but still, I've seen very few institutional repositories that really do this well, that really make it clear who they serve and what they’re there for in a way that’s compelling to the key stakeholders beyond those of us who build and maintain them.

The examples that I can mostly think of are not at universities, but are commercial services like Figshare or ResearchGate or academia.edu, which have managed to create compelling experiences for academics, but also have an entirely different set of incentives that don’t necessarily align with the goals of our universities.

A question for discussion: should we be partnering with these kinds of services? Let them do the user-facing stuff, which they seem to have figured out, and then we focus on the preservation or other things we can’t trust them to do? Harvesting out of these kinds of systems, along the lines of what Herbert van de Sompel and his colleagues described in one of yesterday’s sessions?

[ Image from Wikimedia https://commons.wikimedia.org/wiki/File:Documents_stacks_in_a_repository_at_The_National_Archives.jpg used under CC BY license. ]
So to conclude, I think as we go forward with university repositories, we should refocus our efforts on the researchers and their needs. If we do it well it will help build a bigger community of people who care about our repositories because the repository services do something for them. It will nurture a stakeholder community outside just the library, and that will help the repository programs succeed and be sustainable over a longer period of time, and better meet the goals they were initially designed for.

[Image by TenTenUK https://www.flickr.com/photos/tentenuk/18672186072 used under CC BY license.]
I’ll stop here, and hope we’ve all provided some ideas that will help stimulate some discussion for the rest of the time.

These slides are made available under the Creative Commons Attribution–ShareAlike 4.0 International (CC BY-SA 4.0) license.  
https://creativecommons.org/licenses/by-sa/4.0/