What You're Up Against

Dorothea Salo
Digital Repository Librarian
University of Wisconsin
3 December 2007

Let's start a repository!

Faculty have all their work just lying around... they'll love having a place to store it!

The software is free!

It's essential infrastructure!

The metadata's easy, just Dublin Core!

We're fixing scholarly communication!

All we have to do is set it up!
Maybe a little marketing...
Faculty will do the rest!

All we need is a mandate!

blue-sky thinking. There's a lot of this. It's usually what informs the planning process -- and lest we forget, planning processes don't usually include librarians or consultants who have actually *done the work* to run an IR. When you hear these things, look out!

Reality check

The software is useless.

Nobody needs these things.

Faculty have all this work / publish just lying around, they... they'll take more work please / and don't care!

The word 'metadata' scares faculty!

Faculty have no ideas about it! / Maybe understand marketing... / Faculty won't do anything with it.

Publishers have us outgunned with our own faculty.

We have the faculty and data we need / We have the faculty and what to do.

Because this is the reality. Faculty draw a careful line between what's important and what isn't, and they don't leave important stuff just lying around -- or they don't THINK they do, anyway. It's really no different than the print world. When a faculty member retires, does he come by with tidy files? No, with big ugly dusty unorganized boxes! His computer is no different! Essential infrastructure -- I'd love to see someone try to get Cliff Lynch to repeat that. I suspect he's already embarrassed by it.

There is no good repository software. There is no good repository vendor service. By "good" I mean "something that faculty are going to find useful and flock to on their own initiative." It's all broken.

Fundamental ideology

IF YOU BUILD IT...

... they will come!

Why is it broken? It's broken because the fundamental ideology on which we opened these services was deeply mistaken. What did we believe? If you build it, they will come. This was a nice idea! I love this idea!

Well, we built it.

They aren't coming.

But it was wrong. Well, we built it. They aren't coming. I'm sure all of you here have read as much of the literature as I have. This probably isn't news! But it still *is* news to our library administrators, to the top thinkers in open access, even to the very programmers who make our software! The thing is, none of *them* gets blamed when uptake is low. That's on *us*.

If I sound frustrated, well, I am. I won't lie to you, this is a really hard business to be in right now. It will be until academic libraries reframe some expectations and think a little bit harder, and that's not happening yet. Part of the reason I'm here is that I want it to.

Faculty: the few

- Peter **Suber**
- Stevan **Harnad**
- **Physics** and **computer science**
- "Open Access **Anthropology**"
- Editorial **revolts**
- Faculty-senate **initiatives**

So let's talk about our faculty, and why they aren't flocking to our services. A few of them are genuinely on our side, and we should hug them and love them and call them George and otherwise appreciate them, because they are the real true believers. Peter Suber and Stevan Harnad are doing their level best to evangelize and counteract the fear, uncertainty and doubt emanating from many quarters.

Physics and computer science are light-years ahead of the rest of us. The trick there is, they're not engaged with *us*. They're engaged with their *disciplines* and their repositories. This is only natural; it's what we should expect. But the build-it-and-they-will-come ideology didn't allow for it, and so we don't have a plan to work with it.

Anthropology, you have to love these people, their professional society sells out to Wiley/Blackwell, but do they go gently into that good toll-access night? They do not. They raise hell, they start new journals, they make themselves heard, and bless them for it. Faculty senates are ratifying calls for faculty to retain some of their IP rights, and that's great; if your institution hasn't, get your library administration lobbying for it. Even this, though -- none of these as I've read them has been even symbolically linked with an IR. There's no answer to the question "okay, I have my rights, now what?" Nobody is tracking what happens when the addenda get used. Nobody is systematically collecting materials covered by the addenda. So ideologically speaking this is great, a real step forward -- but in pragmatic terms, it doesn't help repositories at all.

Faculty: the many

- Worried about their reputations

- Worried about their societies

- Don't care about our problems

- Unaware how much they're giving up

- Unwilling or unable to change their workflows
  (and we're not helping much)

They are, unfortunately, also the *few*. Most faculty are hanging out on benches asleep waiting for the bus to pass by.

Faculty don't have the same concerns we do. So faculty do *not* think about pricing when they submit articles hither and yon. They care about their reputation in their field. They care VERY MUCH about the health of their scholarly societies, which is another reason to be very, very careful when talking about OA to them. A lot of societies put all their cashflow eggs in the journal-subscription basket, and they're in a very poor position to move to OA. Those are problems that are very salient to faculty. Our problems? Not so much. To some extent, that's our fault; we've done a little too well at insulating them from the cost of journals.

A lot of them simply do not realize what their publishing agreements are making them sign over. They think they own their work when they don't. This is a hard problem to solve, because -- look, go up to a faculty member and tell him he hasn't made the best possible deal for himself. Then get on the bus to the hospital when he pushes your teeth in. I exaggerate -- but not much; these discussions get really contentious.

The other problem is workflow. Faculty are generally receptive to the idea of self-archiving, but they're *not* willing to put much -- or any! -- effort into it. Even the less-than-ten-minutes per article it takes to stick something in an institutional repository is like pulling teeth!

We have not as yet done enough to help with that. Most folks with IRs have now realized that the tech is the easy part, it's the *services* that make or break the project. And the state of automation is also poor; we need to connect a great big pipe between faculty desktops and IRs, and we just haven't done it yet.

Keystrokes?!
(you expect them to type actual keystrokes?)

- Faculty have *never* preserved their work before. Why should they have to do it *now*?

- Do they know *where* their work is?

- Do they *own rights* to it? Do they *care*?

- Our *interfaces* are horrible. All of them.

- Do they *trust us* to do this? Do they *know* we do it?

- *What's in it for them*?

So we go to them saying, deposit your work, here, it's just a few web forms, no big deal. And they're like, keystrokes? I have to type actual keystrokes? If you look underneath, though, there's some other stuff going on there that you have to contend with. We're asking them to do something they've genuinely never worried about before!

...

So, yeah, value proposition? They do this work, what's in it for them? Will it help them get tenure? Will it help them get their next grant? Will it help them keep their in-progress research organized? Will it help them collaborate across institutional boundaries? No? No. Well, what good is it, then?

The basic problem is...

We don't have anything faculty want...

... and what faculty want, we don't have.

THE IDEOLOGY HAS FAILED.

It's not as though there aren't repository-like functions that researchers need, from version control to access control to high-quality data storage to name it. But IRs can't do any of that. And what IRs do -- collection, metadata, and preservation -- faculty are pretty sure they don't need, and they're not wrong.

So the "build it and they will come" ideology is bankrupt. It's failed. Even though we predicated our services on it.

Pernicious effects

"... *poke* *poke*..."

"Resources? Staff? But faculty are doing all the work, right?"

"My publisher will hate me and I'll never get tenure and no one will respect me again and I'll throw myself off a bridge..."

"Training? ...uragement? ...deposit? Bah, a ...posit only takes ten ...inutes!"

"Customization? It works out of the box!"

"How dare you take in student projects *(learning objects, images, digitized special collections, etc.)*? Repositories are about the PEER-REVIEWED LITERATURE!"

"Of course I own the rights! I wrote it, didn't I?"

"What's that thing got to do with me?"

"Versioning in the software? All we want is the final product!"

"We did all the planning. So everything will be just fine."

"Why is it so empty? Where did we go wrong? Where did I go wrong?"

The persistence of this bankrupt ideology has a number of pernicious effects that you need to watch out for and be ready to counteract. It starts at the beginning, right during the planning process. There's a lot out there about the planning process for an institutional repository, and frankly, 90% of it is complete garbage. Useless happytalk. Technical platforms, collection policy -- that stuff is useless. The only collection policy is "you take what you can get." So they go through the planning process, and they think to themselves, we don't need to throw a whole lot of staff at this! The faculty are typing the keystrokes!

Well, I've explained why that one doesn't fly. You need to be planning for what I call "mediated deposit," which is where you do the keystrokes on behalf of faculty. I'll talk more about that in a minute.

And then you have to deal with faculty and the whole rights thing, and for that I'll point you to Peter Murray-Rust's blog at Cambridge. Peter is a chemist and explains beautifully why this is such a problem.

And then there's repository software. Remember that planning process? It probably assumed that out-of-the-box software was good enough for the purpose, because goodness knows that's the line Harnad pushes. It's garbage. You need to plan for developer and designer hours, and you need to be prepared to push open-source developers for features they don't understand that you and your faculty need, such as document versioning.

Let me tell you, I am not a popular person among DSpace developers. I got cut down to size by MacKenzie Smith the other week, in fact. But DSpace isn't gonna move unless we move it!

And then there are your wonderful colleagues at the library you work in. I work with some wonderful, wonderful people! Unfortunately, they have no idea what I do, no idea that I need their help, and no incentive to give it! And the library literature, and the library professional organizations, *are not helping*. IRs are a hot research topic right now, so I for one get poked a lot, like a lab rat. More surveys than you could shake a stick at; I could hire a whole full=timer just to deal with surveys and interviews! The thing is, all this poking never manages to tell me anything I don't already know. It doesn't give me strategies to deal with all the stuff I'm talking about today. And it doesn't give me hope. Really pretty useless.

In fact, one of the commonest reactions I get from fellow authority-obsessed librarians is, "Why do you have *that stuff*?"

Now, what "that stuff" is varies. It may be student projects, learning objects, images, special collections stuff, whatever. But IRs were sold, again, on the "access to the peer-reviewed literature" ideology, so nobody has any real idea how completely dependent we are on the kindness of faculty who don't have any particular reason to be kind to us!

It's not just your colleagues, either. Anybody read the report from Ithaka on university presses? Did you see the remark calling IRs dusty attics? Ooooh, diss! And yes, all of this hurts. It is frustrating and demoralizing and it can really make you wonder why you're doing this. Watch out for that. It's not you. It's not your fault. It's the system you're embedded in.

The simple fact is...

... nobody's coping well with the ideology failure.

Not in the United States, anyway.

Now, there's hope. It's coming from England and Australia, mostly. You want to watch what the Monash University guys are doing with ARROW and DART and ARCHER and all that stuff. You want to watch the SWORD initiative, helmed in England, and all the nice code that Les Carr and the EPrints people are writing. That's where hope is. It's not here. Not yet.

Two choices

- You can make your service valuable to faculty on FACULTY TERMS...

  - ...which means a lot of design and development work, because repository software is so hopelessly bad...

  - ...or hooking the repository to a sturdier bandwagon...

- Or you can do the deposit work on their behalf...

  - ...which means a lot of software and workflow development, as well as ongoing costs.

- NOTHING ELSE WORKS.

  - If you can't commit to one or the other, get out of this business (or don't get into it in the first place).

After all this doom and gloom, I'm sure you're wondering what I can possibly tell you that will help you get content into your IR. Well, I'm here to tell you that you have essentially two choices. One, you can build a service that is useful to faculty on faculty terms. How do you find out what faculty need? Well, I'd watch the higher-ed literature for mention of what's being called "cyberinfrastructure" or "research computing," because honestly, that's where the action is. Failing that, focus groups can be educational, and you may be able to pair with campus IT to offer them.

The thing is, what faculty want is not what you got. You got an IR. They don't want an IR. Your job, and the job of your developers, is going to be turning an IR into something they want. Usage statistics, versioning, a pipeline from existing campus storage, whatever -- you have to figure out how that works. Alternately, if you *have* a cyberinfrastructure initiative happening on campus, you need to jump on that bandwagon. Trust me, you can't help Big Science, because they take care of themselves. But you can help the people who aren't big enough to be Big Science, and IT will be happy to offload them to you.

Your other option is to do it for them. Be a collection developer. Type the keystrokes. Canvass departmental websites and disciplinary repositories. Now, I warn you, library admin is not going to like this option, because it takes staff and ongoing commitment and both of those cost money. (CLICK) Still. One or the other, because the alternative is an empty repository. Nothing else works.

Oh, they'll tell you other stuff works. Marketing? Marketing does not work. I will show you all the marketing that's actually useful: a trifold brochure that you can hand people. Get a cheap color inkjet printer, 28-pound paper from the office-supply store, and go to town.

Mandates? Yeah, right. In your dreams. I mean, if you see a chance to attach the repository to somebody's review process, do it! but don't waste your breath chasing mandates. That takes senior management, which you aren't.

The software

- If you have the choice, go for EPrints.

- Arrange a customization workflow with IT up-front.

  - You *will* have to customize it. Repeatedly.

  - If IT won't let you or help you, forget it.

- Think *now* about how to integrate with other campus content-creation services.

  - Pay attention to SWORD. Close attention.

  - Got webspace? Build in an "Archive It!" button.

So if you're still in this business after that, let's talk software. If you have the choice, my suggestion is EPrints (and yes, I run DSpace). Its developers are on the ball, offering features that faculty will actually find useful. It's got statistics that don't suck. It's got an author-contact button. It's got embargoes built-in. Just really basic things that DSpace is too busy arguing about Java frameworks to code in.

No matter what you choose, though, you need to bargain for some developer and designer time for customization. Because of the way these software packages work, you'll need customization help on *every upgrade of the software*, and upgrades happen pretty often, because the software is young still. If you can't get this help, pick up your toys and go home. Seriously.

Next, you need to be thinking about how to hook up the IR with other campus services. This is why I mention SWORD, which is a new web-services deposit gizmo based on the ATOM publishing protocol. It's got plugins for DSpace and EPrints, and if you're running one of those two packages, you want it.

What you want to *do* with it, and again you'll need your friendly IT folks for this, is figure out how to hook up with other campus services. If you've got a local webspace thing, that's what I mean -- you want an Archive It! button in there. It won't be easy. I don't have these things yet, though I'm certainly pushing for them. But you can't get what you don't ask for.

The service model

- The repository by itself is useless.

- But it can form part of useful services:
  - Research assessment and faculty bibliographies.
  - Retirement archiving.
  - Publishing and copyright consulting or assistance.
  - Collaboration and other cyberinfrastructure spaces.
  - Conference capture and archival. Small-journal archival.
  - ETDs.

- Figure out the services you can offer that fit your faculty's needs and your goals.

Finally, you need a service model. A value proposition. Something other than the tired old stuff about citation impact and preservation capacity that faculty just don't listen to. There are some possibilities, and no, you don't have to do it all, just pick some things that you know you can do and do well.

Research assessment and faculty bibliographies. At Wisconsin we've been working on a little gizmo called the BibApp, in collaboration with UIUC, and we've found out that although faculty think it's really important to keep their online bibliographies up-to-date, they're too lazy to do the work. In a lot of disciplines, we can automate that kind of work through RSS search feeds from the bigger databases and indexes. Build a SHERPA checker, and throw a few students at downloading suitable articles for the repository, and you're really onto something. Or we think we are.

Retirement archiving... it's not big now, but it's going to become a growth industry. Talk to your campus archivists and records managers. See what they've got that you can have.

Publishing and copyright are confusing as all get-out now and only going to become more so. All you need to get into *this* business is to be more up-to-date than faculty, and frankly, that ain't hard.

I've talked about cyberinfrastructure -- you can't build it on your own, but you can certainly try to be part of it. You can also reach out to your campus's needs for preservation of small-conference proceedings and local publications. Figure out what you can offer that's useful.

You and your colleagues

- You need to win Most Collegial awards.

  - Your colleagues won't help the IR, but they might help you.

- You need your library administrators' ear. And their help.

  - You can't get a mandate.

  - You can't require ETDs.

  - You can't allocate library staff time or funding.

  - You can't even talk to provosts and deans and IT brass.

  - Start beating the drum now, and don't stop.

Dealing with your colleagues, who don't know what open access is, don't care, and don't even know how to introduce you to faculty. The first law is, you must make yourself a valued colleague. They won't lift a finger for the IR, but you've won half the battle if they'll help you just because you're you.

Two, you need to make a pest of yourself with your library's administrators. This is because they have power and connections that you do not...

You and me

- We do not have a community of practice.
  - No journals, just "special issues."
  - No regular conferences (in the US), just random workshops.
  - No professional associations, or even SIGs.
  - Not even a lousy mailing list!
- We are not sharing...
  - ... policies.
  - ... code.
  - ... ideas.
  - ... the burden.

If anything, we're being talked at. Not talked to, and certainly not listened to.

- I can't fix this one alone. I've tried.

Finally, I want to point out one last barrier. And that's the profession, which is not serving us. Cataloguers have AUTOCAT, web librarians have web4lib... we've got nothing. We don't have journals. We don't have consistent conferences; DASER was great, but it's dead, and Open Repositories isn't in the country most years. I'm not going to the next one; I can't afford to. We don't have a professional organization, or even a special-interest group. Not even a mailing list!

This is seriously holding us back. We aren't sharing code; for this and other reasons, innovation in repository-space is frankly languishing except at universities with in-house developers. We aren't sharing policies and solutions to policy problems -- I get email from people who are facing something novel and are ashamed of asking questions! This is absurd. We're not sharing ideas, not on any large scale. And we're not sharing the burden. This is a hell of a job to be in. Support is helpful. It's not out there.

We're not being talked to. We're being talked *at*. This is the first full-on presentation I've given about the experience of running an IR, and I'm one of the louder IR managers out there, honestly. If you want more people like me, and I'm not saying you do, but if you do, you need to speak up, on evaluation forms and when calls for papers come out.

I've tried. A journal I pitched in on folded before its first issue came out. I started a bulletin board and mailing list. They folded for lack of interest. I'm fresh out of ideas. I'd love to see some people here pick up the ball.

Credits

- http://flickr.com/photos/jacobrask (Cover)

- http://flickr.com/photos/purpleslog (Let's start)

- http://flickr.com/photos/bossco (Reality check)

- http://flickr.com/photos/laffy4k (Fundamental ideology)

- http://flickr.com/photos/oneservant2go (We built it, The basic problem, The simple fact)

- http://flickr.com/photos/rickabbott/ (Faculty: the many)

- http://flickr.com/photos/DeclanTM (Keystrokes)

- http://flickr.com/photos/anonfx (Two choices)

- http://flickr.com/photos/alex-photos (The software)

- http://flickr.com/photos/myklroventine (Service models)

- http://flickr.com/photos/richardsummers (You and your colleagues)

- http://flickr.com/photos/kazk (You and me)

Thank you!

... and good luck!

dsalo@library.wisc.edu
AIM: mindsatuw