As always, this blog is my opinion, and the stuff I write here is © Peter Sefton. I'm emphasising this because it is important to make the distinction here between what I personally think and my role at USQ, and the CAIRSS service we help provide for CAUL. I would like to note in particular **this** post does not represent the view of USQ or the Australian Digital Futures Institute and is not written on behalf of CAIRSS. This is post is my personal opinion.

I think that there is a gap in the market in the Australian Repository scene; I think a lot of sites would be interested in services and/or hosting around a completely open source software stack. In this post I will outline the kinds of services I think might be attractive, how they might be provisioned, which bits of software I think would be suitable.

I am writing this for three reasons.

1. I see frustration in the repository community that needs to be resolved. I've never used that phrase about the elephant in the room before, but I can smell one.
2. I think that if it is resolved then my group at the Australian Digital Futures Institute (ADFI) might be in a position to supply some of the services.
3. I don't think that it is appropriate for ADFI or USQ to try to enter the market as a service provider or I would not be posting this, I'd be talking to our office of commercialisation.

Regarding frustration: I have been at public meetings where people have stood up and said quite explicitly that a particular piece of commercial software is not working properly for them. I know that some of the sites that use that software are shopping around for other solutions. I know that at least one bought it but decided not to deploy it. I know that it in the past it has been hard to upgrade between versions. I know it has had performance and stability problems. And I have heard that it is very hard to get it configured to work for the ERA. You, the customers know who you are. My question is why don't you get together and (a) start making some noise and (b) encourage a bit of open competition? Together you represent an attractive market, with a set of similar repositories that are sitting in a very open easy-to-migrate system, with a dead-simple content model. If I was a customer I would be asking the vendor to make the whole thing open source, so the community could fix some of the bugs and/or seeing if there was a compatible open solution with someone offering support – open source means open market for services.

Which reminds me, where is the Australian Dorothea Salo, blogging from a repository rat-hole about what's really going on with repositories and the software that drives them? There's a gap in the blogospehere now that she's quit.

The gap: commercial repository service on an open source base

I know there is at least a small market for services because USQ has supplied such services to two small Higher-Education providers, helping them to set up their ePrints repositories and assisting them with sourcing a hosting provider. I know that there is a market for commercially supported repository software because a significant number of IRs in Australia are commercially supported. I don't know of anybody actively developing and promoting a service which offers commercial support for open software, something like Southampton offers for ePrints. (You can get support for Fez, apparently but as I write this the vendor is certainly not making it easy to find that out on their site).

So what I'd like to see is one or more providers set up shop offering the following:

1. Training, consulting, maintenance, help-desk support and software for running an IR where all the Software is open source.
2. Hosting for institutions who don't want to do it themselves.

That's it. It's simple. Generally speaking there is no more inherent risk in 'buying' a supported open source system than a closed one. In some domains buying proprietary may get you lots more features – but I don't see that in this market.

As a library you could treat it like any other software deal, except that at the end you not only get to keep the software, you can do what you like with it, subject to the terms of the open source license. I would imagine you would be looking at contact terms of something like 2 to 3 years to make this viable for the vendor.

How would it work?

There are a few models for this that I can think of, I'm sure you can think of more so use the comments below.

1. The Southampton/ePrints model, where the copyright holder for the software offers services. I believe their business is going well.
2. A third-party model, where someone just sets themselves up in business – the software is free – away you go!
3. A hybrid model where one group does the training, deployment and support work, while subcontracting development to the relevant software owners. For example if you were offering support for ePrints you might try to negotiate something with Southampton to help fund the development of the product and resource bug-fixing activities.

[Note that above I declared an interest in this idea – I **think** my employer might be interested in us doing some of the development and consulting in a model like 2 or 3]

Which repositories?

Based on a few years in this game, and running the technical team for the RUBRIC project, If I were going to set up a 'classic' Institutional Repository for the purposes of disseminating research via documents produced by your researchers, then I would probably pick ePrints. It works, it is stable and reliable, and it has all the workflow stuff you need for managing an IR, and every time I poll my technical team they agree. I think ePrints is less attractive for other stuff like photo collections and so on, but they keep working on it.

But of course the first phase for a potential vendor in this space would be to work out which packages **they** were going to support.

- If you were going after existing ARROW customers then you'd go for something that was compatible with their existing repositories, which happen to be set up and running in Fedora and mostly using the simple VALET system for ingest so you would probably look to keep the Fedora stuff and maybe migrate to the new ARROW-funded version of VALET, that my team helped write. I think that ePrints might be a good choice for some of these sites for their core IR functions as it vastly more configurable and usable than VALET, but that may be a bit hard to swallow. I am predicting a version of ePrints with a Fedora back-end, which might make people feel more comfortable about it.
- If you detected a need for DSpace support then that would be simple; support DSpace unless those users are all looking to get into Fedora in which case you can wait (probably a long time) to see if the new DSpace/Fedora foundation builds you a hybrid.
- Like I said before I already **know** there is a market for ePrints services.
- If you wanted to take on Digital Commons or DigiTool, I'm not sure what you'd do.

It has been a while since I have caught up with Fez and Muradora both of which are alive and well, so you

would want to consider those, at least.

But I don't think ePrints or any other IR is the whole story. I would prefer to see a much more flexible discovery system over the top of the IR. So if I were entering this market then I would also be pitching the thing that is popping up everywhere in library systems these days, Apache Solr.

For example, the NLA are testing a new Solr system across several data sources.

With lots of libraries using Solr for their catalogues via applications like VuFind I think Solr is a new standard bit of infrastructure which is actually much more important to the customers than what you have on the back-end. Who cares about DSpace vs Fedora vs ePritns when you can mash it all up together and make it easy for people to create their own view of the repository using a bit of Javascript embedded in a web page, or a little site written using a rapid development framework like Django or Rails.

Looking at our own situation, I think USQ should have a big fat Solr index of their course materials, their Intranet, their IR, their data (described using RIF-CS), feeds from their staff and student's workblogs, delicious tagged-links, digital assets from the photography department, etc. That doesn't mean it's all accessible to everyone, Solr can use limit queries for security so that certain groups of people only see certain things, as seen in The Fascinator which can act as a proxy in front of Solr to handle security.

If there's any interest in seeing the kind of service offerings I think would make sense spelled out in more detail let me know in the comments.