I am now feeling worried about my talk at #ILI2009 – I can't think what to say without sounding dismissive of current academic libraries and I know that will upset people. But I came away from #ETD09 with feelings that I was 20 years in the past – the library is a priesthood and normal mortals are expected to visit its temple. We hear from time to time that the library is trying to engage with faculty and the faculty don't listen. Sorry, that's like saying to the mice that you have a better mousetrap. Unless you provide something they want they have no reason to come. And for scientists that's essentially true already.

Over the last few days I have posted about my talk at #ILI2009 asking for suggestions, especially about repositories. I've not got any. I read the FriendFeed comments on my posts (http://friendfeed.com/petermr) – I won't quote but I think you can read them. I had asked whether I could search the whole world repository collection by content, and I assumed I would get some help. Ok, whatever – maybe I will sometime.

Meanwhile I revisited OpenDoar (The Directory of Open Access Repositories) which was set up a few years ago by JISC/OSI/SPARC and others. It's a worthy effort and it collects information about repositories and publishes it. There are 1409 repositories and the list is updated daily. It states:

*Open*DOAR is an authoritative directory of academic open access repositories. Each *Open*DOAR repository has been visited by project staff to check the information that is recorded here. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories.

> *As well as providing a simple repository list, OpenDOAR lets you search for repositories or search repository contents. Additionally, we provide tools and support to both repository administrators and service providers in sharing best practice and improving the quality of the repository infrastructure.*

I'd heard of the content search because someone (perhaps Peter Suber's blog) posted an account that by using Google Custom search they results were at least as good as searching by human metadata. This seemed believable to me then and I believe it even more strongly now. Human metadata does no scale either with volume or diversity – software does. So the search site shows:

> *Search Repository Contents*

> *OpenDOAR is pleased to present a trial search service for the full-text of material held in open access repositories listed in the Directory. This has been made possible through the recent launch by Google of its Custom Search Engine, which allows OpenDOAR to define a search service based on the Directory holdings.*

> *Users of this service can search through the world's repositories of freely available research information, with the assurance that each of these repositories has been assessed by OpenDOAR staff. This quality controlled approach will minimise (but not eliminate!) spurious or junk results, and lead more directly to useful and relevant information.*

Clear

As this is a trial service, please send us [feedback](#) on your experiences.

To search for open access repositories rather than their content, please use the [Find](#) page.

This service does not use the OAI-PMH protocol, or the metadata held within repositories. Instead, it relies on Google's indexes, which in turn rely on repositories being suitably structured and configured for the Googlebot web crawler. If you are an administrator and your material is not being retrieved, first check that your repository is listed in [OpenDOAR](#). If it is listed, you may need to review your set-up against Google's [Guidelines for Webmasters](#) and see the related pages in the [Webmaster Help Center](#), especially the FAQ on [how Google crawls sites](#). There is also excellent advice on [How to Facilitate Google Crawling](#) prepared by [Peter Suber](#).

This sounds like what I want so I tried it with my query. You will remember that when I searched Google Scholar for theses with the term "aminobenzoic" I got only 2 hits, but I knew there were many more because I'd seen over 20 in one repository alone. So I searched...

*Results **1 - 10** for **aminobenzoic**. **(0.25** seconds)*
[*QUT ePrints*](#)

*23 Jan 2009 **...** Smith, Graham and Botta, Raymond C. and Lynch, Daniel E. (2000) The 1:1 adduct of 4-aminobenzoic acid with 4-aminobenzonitrile.*
eprints.qut.edu.au/13136/
by G Smith - 2000 - [Cited by 2](#) - [Related articles](#) - [All 11 versions](#)
[*Radiation chemical studies of P-aminobenzoic acid derivatives **...**](#)

*by Karl Ford Nakken Published in 1966, Universitetsforlaget (Oslo). Radiation chemical studies of P-aminobenzoic acid derivatives. Karl Ford Nakken **...***
openlibrary.org/b/OL5575399M
[*SEPARATION OF p-AMINOBENZOIC ACID BY REACTIVE EXTRACTION. 1 **...**](#)

*21 Apr 2009 **...** The comparative study on the reactive extraction of p-aminobenzoic acid with Amberlite LA-2 and D2EHPA in two solvents with different **...***
eprints.kfupm.edu.sa/104280/
by AI GALACTION - 2008 - [Cited by 1](#) - [Related articles](#)
[*Docking of oxalyl aryl amino benzoic acid derivatives into PTP1B*](#)

*PTP1B inhibitors such as Formylchromone derivatives, 1, 2-Naphthoquinone derivatives and Oxalyl aryl amino benzoic derivatives may eventually find an **...***
www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2637956
by N Verma - 2008 - [Related articles](#) - [All 4 versions](#)
[*Decarboxylation of substituted 4-aminobenzoic acids in acidic **...**](#)

*Decarboxylation of substituted 4-aminobenzoic acids in acidic aqueous solution. **...***
*Dewey: 547/.637. LC: QD341.A7 T6. Subject: Aminobenzoic acids. **...***
openlibrary.org/b/OL5767515M

Google™
Custom Search

These are real documents and real results containing aminobenzoic in the text and should be

really useful. But they really aren't for several reasons:

- There is no indication how many hits I have got. I know there are 10 pages, but how many more after that. Google gives a number, OpenDoar/GoogleCSE does not.

- There is no indication of priority. OK, we have got used to Google's page rank through the eigenvectors of hyperlinks but we need some guidance. Perhaps the number of accesses? Maybe the number of lexical occurrences of the search term? There are other clever things that could be done with Lucene.

- There is no indication of the type of document. A thesis, a data set, a Green-Access publication?

- The search cannot be restricted to document types (e.g. theses , which is what I want).

What I want to be able to do is retrieve ALL documents which are relevant to my search. Here I can only do this by manually clicking on each one.

To be slightly fair to OpenDoar, this was an experimental service installed 3 years ago and I suspect with no maintenance since. But on the other hand this is the single reason why a scientist might actually want to use repositories – to search for a term (yes, I know I really want chemical searches but I can inly do it if the priesthood allows it – see next post).

The whole site and project is clearly not aimed at the general scientist but at repository managers and experts – it says so...

> *The aim is to provide a comprehensive and authoritative list of such repositories for end-users who wish to find particular archives or who wish to break down repositories by locale, content or other measures. OpenDOAR will also provide listings to third-party "service providers" - typically search services who wish to use the categorised lists within their service. This will increase the accessibility and use of the content of these repositories, which will benefit the authors of the research material and the researchers who wish to find it.*

So OpenDOAR provides a platform that service providers can build services on – fair enough. It's been going three years or more, how many services are there based on on it? And are they useful to scientists like me – can I get things there that I can't get easier at PubMedCentral or PubChem?

Anyway I (or rather Nick Day) is a service provider (CrystalEye) and we know how to build crawlers and we know how to index text and chemistry so this looks great. The repository is Open - I assume – so why don't we just add theses to CrystalEye's offering?

I'll let you know in the next post how we get on...