Free US ETDs (FUSE)

Promoting Open Access to American graduate research

Interview with Thomas Dowling, Founder and Architect of OATD.org, a new portal to open access theses and dissertations

Posted on **April 8, 2013**

Last week, a new Web service for finding open access theses and dissertations made its debut. OATD.org is the brainchild of long-time ETD advocate Thomas Dowling, 2009 recipient of the NDLTD Leadership Award and the former manager of OhioLINK's Electronic Theses and Dissertations Center. In this interview with *Free US ETDs* , Thomas shares his vision and plans for making OA theses and dissertations easier to discover and obtain.

GC: The FAQ document for the OATD.org site describes the service as a search tool for finding open access graduate theses and dissertations published around the world. Don't we already have this capability with other search tools such as the NDLTD Catalog , the Proquest PQDT database, OpenThesis.org, and WorldCat, not to mention Google? What makes OATD.org different?

TD: There are lots of ETD search sites out there but none of them specifically index Open Access ETDs. And the Internet search engines like Google find lots of stuff, not just ETDS, but it is nearly impossible to whittle large result sets just to theses. I am convinced there is a real need for a tool dedicated to OA theses so scholars can find them more easily. Libraries have been working hard to bring up other discovery services for our users but when it comes to search tools for theses and dissertations, we've fallen behind.

With OATD I'm taking extra steps to ensure the openness of the ETDs being indexed. I am not indexing sites that deliver commercially available documents, or open sites that contain records for printed theses are not online anywhere. Repositories often contain some OA theses and some that are not, so it is important to check the collection level metadata to see what the status of a given collection is.

GC: In this country, ETD professionals have been calling for a free US portal to ETDs so that we could have one place to search for our graduate works. Is that what OATD represents? If not, what would it take for OATD.org to be comprehensive in covering all US ETDs born digitally, say those from 2004 forward?

TD: OATD indexes ETDs from around the world but it shows a model of how an aggregated ETD service could work. It would not be hard to spin off a US-only index from OATD.

The hard part of the process is raising awareness of the issue – getting institutional willingness to make ETDs available for harvesting in the first place. All it takes for most repositories (DSpace, Digital Commons, Eprints, even ContentDM with the right plug-in) is to turn on the switch to allow OAI-PMH harvesting. I am even able to get some metadata by web crawling a university's ETD listing on their web site.

GC: Why is the OAI-PMH method of aggregating ETD metadata from myriad sites better than other methods? What are its limitations? Does it pick up metadata from all types of repositories?

TD: The PMH stands for "Protocol for Metadata Harvesting" so it was created specifically to collect and aggregate metadata from lots of unrelated repositories all over the Internet. Harvesting structured metadata with defined fields like author, title, subject provides the best data for searching. The alternative would be to use a Web crawler that can't make good distinctions between the various elements in the ETD record, like author or publication date.

And OAI-PMH is really pretty simple to implement. All that is needed is a base URL to the site where the ETDs live. Additionally, OAI-PMH allows the harvester to get everything from just one collection in the repository (say, all the ETDs that are available Open Access), or just everything that has been edited or revised since the last harvesting visit. So it supports a more continual, recurring approach to building a database for ETDs that grows over time.

GC: If an institution wants to push its ETD metadata to you instead of having OATD.org pull it via harvesting, are they able to do that? What metadata elements would need to be supplied? What format would the metadata need to be provided in?

TD: First, it is important that the site distinguish which of their ETDs are available via Open Access, so we can harvest just the OA ones for OATD. That can be done in different ways: by adding a metadata field in the record to flag the OA status or to flag items restricted by

audience of embargo period; or it can be done by setting the OA ETDs off from the others in a separate collection. Besides that, only a minimal set of metadata is absolutely critical for OATD, say "Author", "Title", and "URL" to the location of the ETD. And I am more than happy to work with a given institution that does not have its own repository to set up an Excel file to send the ETD metadata so it can be loaded into OATD.

GC: Your menu offers the option to limit search results to Creative Commons licensed works. Kudos, no other ETD search site offers this important feature! How does your system know that an ETD has a Creative Commons license attached?

TD: With some repository metadata, such as DSpace using Dublin Core, the record includes a DC.rights field that contains a url for the specific Creative Commons license attached to the ETD or to the phrase 'creativecommons'. So the harvester can detect the presence of the CC license that way. And there are comparable fields in other repositories using other metadata schemes.

But if the only evidence of a CC license is in the file itself, and OATD has to fish around the ETD document to find evidence of that license, things become much less reliable and consistent.

GC: What kind of support does OATD.org have to ensure it will scale and sustain over time?

TD: The Z.Smith Reynolds Library at Wake Forest University has committed to supporting the servers needed to run OATD. And I am in the process of forming a Steering Committee/Editorial Board to help provide guidance for site development.

GC: Many of us remember your work in developing the *WWW Index of Theses and Dissertations* at OhioLink, back when you worked for that consortium. Is OATD an extension of that project?

TD: I guess OATD is a reconceptualization and reengineering of the earlier ETD index. There have been some significant changes in how the index works. OATD does represent my ongoing professional interest and commitment to help make ETD scholarship more widely and freely accessible.

GC: Thanks for talking with FUSE about this valuable new service, and for all you do to support ETD discovery.

TD: You're welcome! If any institution would like help in making their OA ETD metadata available through OATD.org, please contact me at dowlintp@wfu.edu.

This entry was posted in **ETD metadata**, **OA portal to US ETDs**, **Open Access** by **gclement**. Bookmark the **permalink [http://sites.tdl.org/fuse/?p=582]** .