dissertations, making them highly accessible, useful and share-able — at absolutely no cost.

Patents Online is the company behind www.freepatentsonline.com, the #1 worldwide Website in terms of audience in the patent space. With its worldwide patent database, freepatentsonline.com averages 10 million unique monthly users, and 1 million users have registered (registration is free and gives additional functionality).

Having started in the patent space, Patents Online's mission has expanded to include offering free access to all types of technical and academic documents. Theses and dissertations form an important body of documents in this space. OpenThesis.org will allow authors to upload their theses/dissertations, and the documents will then be vaulted for posterity and made searchable anywhere in the world. If an author (or an author's family) has only a paper-copy of a thesis/dissertation, the site will give guidance on how to get it scanned into computer-readable form, for subsequent uploading. OpenThesis.org will also work directly with Universities to ensure as comprehensive a database as possible. The existence of a comprehensive ETD database will benefit Universities in licensing efforts, and authors via prospective job opportunities.

As is currently true for www.freepatentsonline.com, visitors to OpenThesis will be able to register in about 1 minute, for free, and gain access to special features such as the ability to organize documents into folders, annotate documents, share documents with other users, and set up alerts for automatic notification of new documents of interest.

A few other important aspects of OpenThesis.org:

- Authors will create accounts that will be e-mail-verified as part of uploading their document, and can choose on their very own author page (www.OpenThesis.org/their-name) to show their contact information and/or LinkedIn profile page as desired.

- Every College and University will have a free microsite, e.g., www.OpenThesis.org/Massachusetts-Institute-of-Technology, such that they can link to a repository of documents from their own authors, co-branded with their logo.

- The site will allow users to search the entire document collection, or to limit the search to a particular institution. The documents from a given institution also can be browsed with that institution's microsite.

- OpenThesis.org will offer, via partner companies, the ability to print and bind theses and dissertations faster and cheaper than through traditional sources.

Theses and Dissertations represent an important compendium of vetted research. OpenThesis.org will ensure that these works, which authors created with formidable investments of inspiration and time, and which universities have ratified as part of awarding graduate degrees, are instantaneously available anywhere in the world via the Web at no cost.

NDLTD Union Catalog / VTLS Visualizer

» Panelist: Vinod Chachra, VTLS

This presentation will discuss in depth the structure and use of the NDLTD Union Catalog. Information will be provided on how the data is harvested, indexed and made available for access. Statistical information will be provided on the size of the database (in excess of 750,000 ETD's) and its source broken by continent, language and country. Useage statistics will be provided showing the source, frequency and pages viewed. Finally pointers will be provided on what to do with the metadata to make dissertations from you institution more accessible to the world at large.

Topical Categorization of Large Collections of Electronic Theses and Dissertations

» Panelist: Edward Fox, NDLTD Director, Department of Computer Science, Virginia Tech

» Panelist: Venkat Srinivasan, Department of Computer Science, Virginia Tech

Objectives: The NDLTD Union Catalog has metadata for over 600,000 Electronic Theses and Dissertations (ETDs) in diverse languages from universities around the world. The users can access these ETDs through various search and browse Web interfaces reachable through the NDLTD Website (example, from Scirus and VTLS). We aim to improve those services in two ways. First, we develop approaches to build larger collections of ETDs, which consist of ETDs not only collected via NDLTD's Union Catalog, but also those collected through focused crawling of many universities' Webpages. Second, we develop approaches in order to make these large collections more amenable to being used by students and researchers.

Methods and Results: We have identified repositories for some universities that host ETDs but that are not yet part of NDLTD. We have developed custom crawlers in order to crawl some of these repositories as well as the NDLTD Union Catalog in order to harvest ETDs and their metadata (where permissible). Our current collection has about 40,000 ETDs from Union Catalog for our initial experimentation, and we actively continue to collect more ETDs.

We also have developed a categorization system, based on the Library of Congress categorization system and Wikipedia, that is more suitable for categorizing ETDs, and have categorized ETDs into the resulting category tree. Users can first browse this category tree based on their needs and then can either browse a particular node, or search it for items of interest.

Conclusions: Through focused crawling, we have been able to increase content available to users, and made it available at a single place. Categorization of ETDs has helped organize the ETDs semantically in order to make it easier to find relevant information. As part of future work, we will improve our methods to collect as many ETDs as possible from the NDLTD Union Catalog and from various universities around the world, categorize them, and provide a Web interface facilitating access.

NEW TRENDS PLENARY PANEL SESSION

» Moderator: Vinod Chachra
» Thursday, June 12 4:15 p.m.–5:45 p.m.

The Semantic Electronic Scientific Thesis

» Panelist: Peter Murray-Rust, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Lezan Hawizy, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Jim Downing, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Joe Townsend, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Nick Day, Unilever Centre, Department of Chemistry, University of Cambridge, UK

» Panelist: Peter Sefton, University of Southern Queensland

We have developed a range of tools and protocols that allow the creation, validation, and re-use of ""born digital"" theses in scientific domains, especially disciplines reporting chemical

information and results. The primary authoring tools in science are Word and LaTeX, both of which create documents with structures (chapters, sections, etc.) and semantics (annotated paragraphs, tables, graphs, etc.). In many cases the theses also contain raw and processed scientific data which can be at least as important as the natural language text.

We have developed vocabularies and ontologies to describe such theses and, for example, are devloping an authoring tool for creating semantic chemistry in a Word environmant. We urge that instituitons encourage semantic theses and have been developing a proof-of-concept (ICE-TheOREm). Here a student can assemble a thesis from components which can me managed locally on on a server and create either Word2007 or ODT-compatible documents. Such theses preserve all the semantics and data.

In practice, current theses are deposited as PDFs and to re-use their contents we must resort to natural language processing (OSCAR3) or semi-structured data tools (OSCAR-DATA). PDF has no semantics and reconstruction is seriously lossy but we can often extract meta/data by machine. We urge institutions always to deposit the native Word2007 or LaTeX documents as well as any PDF; in this way they will capture far more of their science.

We also demonstrate lightweight semantic repositories which provide an embargo mechanism for all or part of the thesis (TheOREm, using ORE). The metadata for the documents (including structuring) are converted to RDF which can be queried with SPARQL providing great flexibility.

Our toolkit and examples are based on the premise that all software, protocols and content should be Open.

We thank JISC, and Microsoft Research for support.

ProQuest Dissertations and Theses Database User Survey:
The first large-scale survey of dissertation information seeking behavior

» Panelist: Amanda Ross, ProQuest LLC

OBJECTIVE: To develop a clearer profile of researchers who use dissertations and a better understanding of how dissertations are used in the research cycle in order to present dissertations in more effective ways to support research.

METHODS: "Users" are those who accessed the ProQuest platform and/or ProQuest Dissertations and Theses during the course of the survey (April 20 – May 15, 2008), and responded to an invitation link to the survey instrument. An incentive drawing was included. 3,034 individuals took part in the survey.

The survey instrument – mostly of structured single response questions – included a two multi-element questions using a 1-10 rating scale, and two open-ended questions on the object of the search and on usage of competitive databases like PQDT.

RESULTS: Graduate students account for nearly half of database users; undergraduates about a third. Nearly half of all those who searched the dissertation database are either studying for doctorates or working on master's theses. The corollary is that a majority of those accessing PQDT are not doing so. While librarians in this survey were not themselves frequent users of PQDT, the college or graduate school library Website is an "extremely important" influence on accessing the PQDT or ProQuest platforms.

PQDT is likely to be accessed quite specifically with the intention of reviewing dissertations or theses – and often. About one in five users accessed PQDT at least five times in the month immediately prior to the survey. Social sciences, business and education are the three leading disciplines associated with ac-

cessing PQDT. Other important areas are the arts and humanities generally, and medical sciences.

CONCLUSIONS: This survey suggests that there is more to learn about non-student researchers and a need to find ways to provide them access to dissertation research. It also supports the importance of dissertations as primary source material in current research.

Students as Advocates

» Panelist: Julia Blixrud, Scholarly Publishing and Academic Resources Coalition (SPARC)

There are many challenges in talking with students about ETDs. Some of the important topics involve those surrounding open access and author rights. Reaching out to students to engage them in these issues can be a complicated process. There are, however, an increasing number of student groups that have become interested in these issues and finding ways to connect with them can lead to increased understanding about the value of ETDs.

The Scholarly Publishing and Academic Resources Coalition (SPARC) has been working with a variety of student groups on open access issues. The techniques in working with students are to find common areas of concern and leverage collaborations. In addition, taking advantage of technologies that appeal to the students increases the reach of the programs.

This paper will provide information on several SPARC student activities. The Right to Research campaign responded to a growing demand from the college student community for tools and resources to express their support for Open Access to research. Students are also interested in Open Educational Resources and SPARC is working with student groups to raise awareness about them as well. The annual SPARKY contest provides awards to innovative videos expressing student perspectives on sharing information. A student open access blog provides a means for students to communicate among themselves.

SPARC's work with students and student groups suggests many students have embraced the concept of open access and are receptive to raising issues they see as impeding their ability to access information. These activities suggest ways others might consider in their work with students and ETDs.

GLOBAL OUTREACH PLENARY PANEL SESSION

BRIDGING THE KNOWLEDGE DIVIDE: EXPANDING GLOBAL OPEN ACCESS

» Moderator: Frances O'Brien, Dean, WVU Libraries
» Friday, June 13 9:30 a.m.–10:45 a.m.

OceanDocs and Open Science Directory:
Two facets of the information policy of UNESCO/IOC-IODE

» Panelist: Marc Goovaerts, Information Technology, Hasselt University Library

The International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission of UNESCO (IOC) has developed, in the past five years, the tools for a modern information policy focused on increasing accessibility of scientific literature in the field of marine science