FEATURED WEB RESOURCE: *Theological Commons*

by Gregory P. Murray

The Theological Commons (http://commons.ptsem.edu) is an online digital library supporting research in theology, religion, and related fields. This searchable system contains tens of thousands of resources and is continually growing. The majority of the content consists of public domain books, but the site also includes thousands of periodicals, audio recordings, and other formats. All of these resources are accessible publicly and at no cost.

Background

In early 2008, Princeton Theological Seminary entered into an agreement with Microsoft to digitize public domain print materials. Project funding would come from Microsoft, digitization would be performed by the Internet Archive at a scanning center located in Princeton Seminary's library, and Princeton Seminary would provide thousands of volumes on theology and religion for inclusion in Microsoft's Live Search Books service (Microsoft's answer to Google Books). Other institutions that had joined Microsoft's program included the British Library, Columbia University, Cornell University, the New York Public Library, the University of California, the University of Toronto, and Yale University.

A few months later, however, Microsoft ended its mass digitization initiative, leaving Princeton Seminary and others to consider how or whether to proceed without outside financial support. Remaining committed to its vision of contributing to the shape of a global digital library in which theological disciplines would be represented, Princeton Seminary decided to move forward by providing an annual budget for digitization and by retaining its relationship with the Internet Archive. The digitized volumes would be deposited in the Internet Archive's online digital library (http://archive.org).

Because the Internet Archive has been so successful in building its massive digital library in the service of its mission, "universal access to all knowledge," the sheer size of the library is as daunting as it is impressive. For the researcher in theology and related fields, navigating these seas of data to find relevant material can be a challenge. In late 2010, Dr. Iain Torrance, at that time the President of Princeton Theological Seminary, asked a small subset of library staff to consider how to improve discoverability and access to the thousands of volumes on theology and religion that Princeton Seminary and other institutions had digitized through the Internet Archive, to facilitate research by students, scholars, and pastors both locally and globally. Starting from this seed, and under the leadership of Dr. Clifford Anderson, at that time Curator of Special Collections at Princeton Seminary, the library's "digital team" began building a database and web-based user interface with this goal in mind.

The first step toward realizing this vision was content selection, which began with harvesting the metadata and full text of every item in the Internet Archive that had originated from Princeton Seminary's library. However, because the goal was to provide access to relevant resources, not to showcase Princeton's digital content, the digital library team subsequently took a detailed list of Library of Congress subject headings provided by Don Vorp, at that time Collection Development Librarian at Princeton Seminary, and performed searches in the Internet Archive system for digitized books with those subjects, irrespective of library of origin. Those items were then harvested in the same manner. This procedure soon amassed tens of thousands of digital texts, and in March 2012, the Theological Commons was publicly released as a free, web-accessible digital library.

Content from Other Libraries

The inclusion of digitized materials from other libraries merits further comment. The Theological Commons aspires to live up to its name as a "commons"— a central location or shared resource available to an entire community — in part by incorporating content from outside Princeton Seminary's own library collections. Less than half of the content in the Theological Commons originated from and was digitized by Princeton Seminary; the majority originated from other

Gregory Murray is Director of Academic Technology and Digital Scholarship Services at Princeton Theological Seminary, Princeton, NJ.

research libraries that also digitize their materials through the Internet Archive (Figure 1). (The content contributors are acknowledged in multiple places in the user interface, including the Frequently Asked Questions page and the Browse by Contributor feature.)

In addition, the site incorporates several "Featured Collections," some of which have been built through direct collaborations with other theological libraries. The Payne Theological Seminary and A.M.E. Church Archive is a discrete collection within the
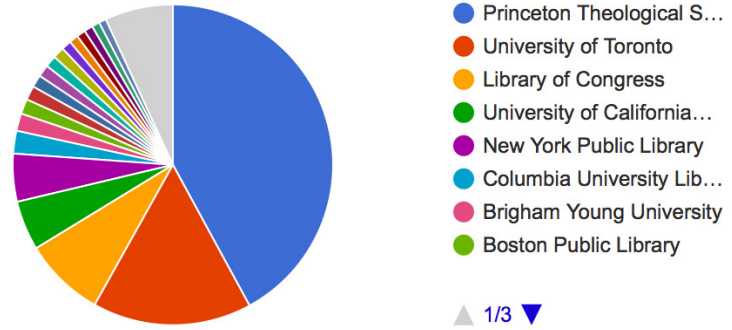

Figure 1. Contributing Institutions

broader Theological Commons containing materials originating from Payne's archives and digitized through the Internet Archive, funded by a grant to Payne Seminary from The Arthur Vining Davis Foundations. The content is organized into subcollections with accompanying descriptions, along with scholarly historical essays that provide further context and insight into the content. Another featured collection within Theological Commons is Missionary Research Library Pamphlets, containing selected pamphlets from the Missionary Research Library Archives, which are held by The Burke Library at Union Theological Seminary. Columbia University Libraries, of which The Burke Library is a part, is funding the ongoing digitization of the pamphlets collection, and new content is regularly incorporated into the Theological Commons.

Features and Usage

The digitized print materials (books, periodicals, and theses) are fully searchable; each digital object includes not only descriptive metadata but also the full text of the original work, as generated by OCR

Figure 2. Search Box

software. Audio, video, manuscripts, and photographs are searchable only by metadata, although many A/V items will include full text in the near future (more on this below). The search box, which appears at the top of every page in the site, is deliberately simple (Figure 2). It provides keyword searching across all textual data and metadata, and accepts commonplace techniques such as using quotation marks for exact phrases. The user can also opt to search only titles or author names. A Search Help page describes advanced search techniques that allow the user to perform arbitrarily complex searches.

The interface also provides facets (Figure 3), by which search results can be filtered according to predefined categories, namely date of publication, format of the physical item, language, and subject. The subject facet is designed to represent a balance between complexity and simplicity. The terms for this facet are derived not from the vast array of Library of Congress subject headings, but rather from the Library of Congress classification scheme. In this approach, each digital resource is assigned to a broad taxonomy of knowledge.

For convenience, the user can search first and then apply facets to filter the results, or select facets first and then perform a search; the end result is the same. The search terms and facets currently in effect are clearly indicated at the top of each page of results, and the user can remove any individual search constraint at will. Importantly, the interface allows the user to select more than
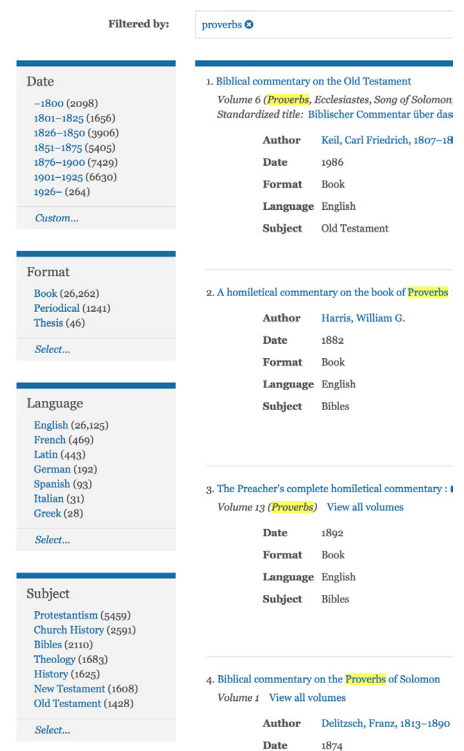

Figure 3. Faceted Search Results

[Read online] [Get PDF] [Copy to Kindle]

Biblical commentary on the Old Testament

Volume 6 (Proverbs, Ecclesiastes, Song of Solomon)   View all volumes
Standardized title: Biblischer Commentar über das Alte Testament.

Author	Keil, Carl Friedrich, 1807–1888
Date	1986
Format	Book (1208 pages)
Language	English
Subject	Old Testament
Contributor	Brigham Young University-Idaho, David O. McKay Library

Figure 4. Item Page

4. Biblical commentary on the Proverbs of Solomon

Volume 1 (View all volumes)

Author   Delitzsch, Franz, 1813–1890

Figure 5. "View all volumes"

one value from any given facet simultaneously, allowing the user to apply a complex combination of multiple formats, languages, and subjects. Such combinations can include either a predefined or custom range for the date of publication. These features exponentially multiply the paths by which a researcher can navigate this digital library to pinpoint content of interest.

Because facets can be selected independently of any keyword search, the facets also serve as a browsing feature. Additionally, the user can browse by author, collection, contributor (library of origin), or title.

Every item in the system has its own item-level page (Figure 4), listing all relevant metadata and providing multiple ways to access the content. For digital resources derived from print materials, the user can read the page images online in the Internet Archive's BookReader, download a PDF (with searchable text included), or, if the quality of the OCR-generated text is sufficiently high, copy the text to a Kindle book reader. For audio and video, the user can listen/watch directly within the Theological Commons using the integrated streaming media player. Similarly, photographs are displayed in an integrated image viewer, allowing close inspection. In the case of archival files, content is scanned at the folder level, and any given folder can be viewed in the BookReader.

From the item-level page of any given text-based resource, the user can search within that particular text. Each search result includes a snippet showing the keyword(s) in context, with a link to the corresponding page image where the keyword occurs. To help researchers find related items, each item-level page also includes a "More like this…" feature showing items with similar Library of Congress subject terms.

Because the Theological Commons includes numerous serials and multi-part monographs, it is worth noting that these sets of related items have been reconstituted in the user interface, even though each item was digitized individually. Every item that is a member of such a set is displayed with a link labeled "View all volumes" which in turn displays all items in the set (Figure 5). Uniform titles (and in some cases, multiple editions of a given work) are similarly grouped, providing instant access to all members of that logical grouping.

In all of these ways, the Theological Commons represents a blending of the aims and methods of mass digitization, in the manner of Internet Archive or Google Books, with the informed curation and accurate resource description that are the hallmarks of librarianship. The goal is to strike a balance between quantity and quality in a way that maximizes the reach and usefulness of these digital resources for research and ministry.

Current and Future Directions

In mid-2013, the Henry Luce Foundation awarded Princeton Theological Seminary a $1.5 million grant for the expansion of the Theological Commons in two important directions: to move beyond text by digitizing and providing access to audio and visual materials, thereby promoting digital media as an integral part of the study of theology; and to move beyond the West, by digitizing and providing access to theological material published in or focused on Africa, Asia, and Latin America, thereby providing resources of relevance to diverse communities of scholarship and faith around the globe.

The first direction has led to the Princeton Theological Seminary Media Archive, one of the "Featured Collections" within the Theological Commons. Since the early 1950s, Princeton Seminary has maintained a consistent practice

of recording the many public lecture series and institutes held at the Seminary, along with major sermons, addresses, and interviews. The result is an archive of several thousand unique, previously unpublished audio recordings that until now existed only on aging reel-to-reel tapes. Thanks to the Luce Foundation grant, these audio recordings (and a small number of video recordings) have been digitized. The "digital team" at Princeton Seminary library is actively verifying the audio quality and metadata accuracy of these recordings. As of this writing, over 1,600 recordings are publicly accessible in the Theological Commons, and the number is growing continually. In addition, the team is obtaining professional transcriptions of many of the recordings. Because digital audio and video recordings contain no searchable text (other than their accompanying descriptive metadata), transcriptions will allow the verbal content of lectures and sermons to be fully searchable.

The second direction of the Luce Foundation grant — shifting the intellectual center of gravity in the Theological Commons beyond North America and Europe — has led to the Latin America Collection, which contains materials from Princeton Seminary's deep collection of Latin American books and periodicals. In addition to public domain content, this collection includes several current or recent Latin American journals and book series (thanks to legal agreements between Princeton Seminary and the publishers). While the Latin America Collection continues to grow, Princeton Seminary library is also digitizing content pertaining to Christianity in Africa and Asia. One major effort currently underway is digitization and metadata preparation for the Moffett Korea Collection held by Special Collections at Princeton Seminary library. This collection contains the papers of scholar and Presbyterian missionary Samuel Hugh Moffett (1916-2015), including a sizable photograph collection, all of which pertains to the history of Korean missions, the Korean churches, and religion in Korea. The digital representation of this collection is expected to be available in the Theological Commons by the end of 2016.

Current Status

As of this writing, the Theological Commons contains over 88,000 resources ranging from the year 975 to 2015, in eight formats, ninety-six languages, forty subjects, and six featured collections. New content is added continually.

The Theological Commons is developed and maintained by the Digital Collections unit within the Office of Academic Technology and Digital Scholarship Services at Princeton Theological Seminary Library: Maggie Hasegawa, Digital Library Application Developer; Donna Quick, Digital Production and Quality Assurance Specialist; and Christine Schwartz, Metadata Librarian and XML Database Administrator. Gregory Murray, Director of Academic Technology and Digital Scholarship Services, is team leader and project manager. The team welcomes comments and questions at digital.library@ptsem.edu.