Behind the Scenes at the Digital Archives
Clifford B. Anderson

When you think of the Department of Special Collections, do you imagine rows of antiquarian folios? Or do you picture box after box of papers and manuscripts? Or do you think about cabinets of photographs and maps?

What do all these traditional images of Special Collections have in common? In a word, locality. To see them, you have to travel to Princeton and make sure to arrive at Special Collections during our public hours. We do hope that you will have a chance to visit us soon!

However, there are many scholars who cannot afford to make the trip to Princeton. Of course, we’re glad to help out with reference requests. (Keep them coming to special.collections@ptsem.edu!) But we have not traditionally been able to provide a full copy of Calvin’s 1536 *Institutes*—the “wear and tear” costs have just been too high.

Things are changing. This issue of our newsletter highlights several ways in which we are providing full access to portions of our rich collections to scholars around the block and around the world. Expect more to come.

In the interim, we hope that you enjoy learning about what it takes behind the scenes to move from an analog archive to a digital archive.

Suggestions for what you would like to see us work on next? Let us know! And thanks for visiting us on the web: http://libweb.ptsem.edu/collections.

John Calvin at the Internet Archive
Clifford B. Anderson and Kenneth Woodrow Henke

The Princeton Seminary Library is working in partnership with the Internet Archive to digitize a substantial number of public domain books. The Internet Archive is a non-profit organization “founded to build an Internet library, with the purpose of offering permanent access for researchers, historians, and scholars to historical collections that exist in digital format.” The Internet Archive has digitized more than 1,285,000 books to date. Access to these materials is open and free from commercial restrictions.

As of this writing, the Internet Archive has digitized more than seventeen thousand books from the Seminary Library’s collections and made them available for anyone to review, redistribute, remix, and (we hope) read. You can access these books at www.archive.org/details/Princeton.

In celebration of the 500th anniversary of John Calvin’s birth (July 10, 1509), this issue of the *Luce Library Bulletin* highlights a number of antiquarian editions of Calvin’s works digitized from our Special Collections. You can view these books in a variety of formats. We recommend starting with the “Read Online” format, by which you can leaf through books using your web browser. If you would like a higher fidelity version, download the PDF. Finally, you will also find a link to the “full text” of books. Please note, however, that the text is uncorrected and will likely contain transcription errors.

(continued on page 2)
Estienne printed four editions of the Greek New Testament, two editions of the Hebrew Scriptures, and numerous Latin and Greek classics before immigrating to Geneva in 1550, where he published a French Bible, a Latin Bible, and the first edition of Beza’s Latin New Testament as well as the writings of John Calvin. The 1559 edition of the Institutes was his final work, completed just days before he died. If you follow the link, you can see his printer’s device (based on Romans 11:17–24) on the first page with its warning in Latin against becoming proud and conceited in one’s faith. (see www.archive.org/details/ institutiochrist1559calv)

We have also scanned other Latin editions, including François Perrin’s 1569 Geneva edition (see www.archive.org/details/ institutiochrist00calv) and the 1576 London edition of the learned French Huguenot refugee Thomas Vautrollier, which contains a sonnet in French on the death of John Calvin opposite the title page of Book One. (see www.archive.org/details/ institutiochristlond00calv)

The earliest English translation of the Institutes in print is by Thomas Norton, a lawyer, member of Parliament, and accomplished poet who also contributed some twenty-eight psalm translations to the famous Sternhold and Hopkins metrical psalter. The first edition was printed in 1561. (see www.archive.org/details/ institutionofchr1561calv)

Demand proved high and the volume went through numerous editions. In later editions, Thomas Norton allowed his name to appear on the title page as translator; he also added a fine preface in which he explains his approach to translating the work and commends it for study. (see www.archive.org/details/institutionofchr1578calv)

Norton’s English translation remained the standard until John Allen’s 1813 translation.

John Calvin’s Bible commentaries are represented by his early commentary on Romans (1540) (see www.archive.org/details/commentariijinepi00calv) and his Harmony of the Synoptic Gospels (1584) (see www.archive.org/details/harmoniaexevange00calv). Even if your Greek is a little rusty, you should look at our copy of the Greek translation of Calvin’s catechism. The scanned volume comes from the library of Samuel Miller, and inside the front cover is a note in his handwriting stating that this volume once belonged to the seventeenth century Genevan theologian Benedict Pictet (a nephew of Francis Turretin, whose theology text was the one primarily used in the early days of Princeton Seminary) and that it had been procured for Miller by none other than the famous Genevan historian of the Protestant Reformation, J.H. Merle d’Aubigné. (see www.archive.org/details/stoicheiosistesc00calv)

Finally, we call attention to a volume of Calvin’s sermons translated into English from Latin in 1561 that has a wonderful engraving of John Calvin at age fifty-four inserted opposite the title page. The preface says in introducing Calvin’s preaching that “…all his sermons seme nothing els but the swete licour of the scriptures and liuely word of god set furthe before our eyes in Christalline vessels to entice vs to beholde them and to prouoke vs to tast and to smel of these lycours of lyfe…” (see http://www.archive.org/details/fouregodlyesermo00calv)

---

Special Collections Staff

Curator of Special Collections: Clifford B. Anderson
Reference Archivist: Kenneth Woodrow Henke
Manuscript Librarian: Sarah Seraphin
Project Archivist: Robert Golon

Stephen D. Crocco, James Lenox Librarian

The Luce Library Bulletin is published twice a year (spring and fall) in cooperation with Speer Library. Current and back issues are available online.
Special Collections houses more than 400 manuscript collections. Every day we work to process and describe these materials for research use, placing them in archival enclosures and creating inventories of their contents. We write biographies, checking books and records for accurate information. We decipher handwritten notes, identify photographs, and try, in short, to make the unorganized orderly. The fruits of our labor can be found in the finding aids to our collections.

Our finding aids guide researchers to the cache of resources in our collections. They list the contents with varying levels of description. And though we strive to inventory every box that we acquire, the task of processing and describing every collection is enormous. Thus, a number of our collections are unprocessed, and a number of our described collections need to be brought up to the current standards in archival description. As the expectation for online access to information grows, we face the additional challenge of publishing our finding aids online.

Until recently our finding aids were only available offline as print resources. The Seminary Libraries’ online catalog (or OPAC) contains record-level summaries for 267 collections. Now, finding aids for more than 200 collections are available through their online cataloging records.

We began the process of converting our finding aids from Microsoft Word documents to machine-readable form in the spring of 2008. In order to ensure that we were only putting accurate information online, the first step was to determine which of our collections were completely processed and fully described. We conducted a box-level inventory of all of our collections, and determined which had finding aids. After verifying that a finding aid listed contents for each box, we began the process of revising their structure and enriching their content, based on archival standards.

Encoded Archival Description (EAD) is a standard for encoding full-text finding aids in machine-readable form, similar to the function of the MARC format for books. The structure of EAD allows for the multi-level, hierarchical description that is present in finding aids. Descriptive information is encoded in pre-defined tags, making the document machine-readable based on standardized content. To fit our finding aids into the structure of EAD, we had to impose both structure and content.

For each finding aid that we were preparing to encode we created a collection summary that contains the name of the creator, dates included, total number of boxes (as well as measurement in linear feet), and languages included. Most of the existing finding aids contained biographical information or chronological lists of major life events, but where this information was lacking we added it. We also added scope and content notes, which summarize the contents of the collection, and administrative information, including processing information, access restrictions, and instructions for citation. We also checked for spelling errors and edited content for clarity and quality. The resulting documents, we feel, are easily navigated and well described. And, although the project of restructuring began with the goal of displaying the finding aids online, we continue to provide access to our most up-to-date copies of these documents in print form in the Seminary Libraries’ reference areas.

Once we had imposed the new structure and content, the process of encoding began. At a two-day workshop on EAD, hosted by the Society of American Archivists, I learned its basic structure, practiced encoding a finding aid, and discovered many resources online for librarians and archivists that are starting EAD implementation projects. The Encoded Archival Description Tag Library, Version 2002, is a resource created and maintained by the Society of American Archivists Encoded Archival Description Working Group and the Network Development and MARC Standards Office of the Library of Congress. It is available in print and online.

EAD is an XML-based standard. Each tag in EAD is designed to hold the various descriptive elements of a finding aid. In determining our best practices for the use of EAD tags, we referred to the RLG Best Practice Guidelines for Archival Description (2002) and the Virginia Heritage Project Best-Practice Guidelines for Archival Description (2003). To begin, we designed a basic XML template that would match the structure of our finding aids, and then we began copying and pasting information from the Word documents into the XML document. We referred to the EAD Cookbook 2002 to find an XSLT stylesheet that would transform the encoded finding aids into HTML in a consistent way. Once we found a stylesheet that was similar in format and color to our web site style, we worked hard to customize its structure so that it would display the information according to Princeton Seminary standards.

When we were satisfied with the design of our templates, and had encoded about a dozen finding aids from scratch, we contacted codeMantra, a data conversion company, to develop a strategy to outsource the encoding process. Since January, we’ve worked with codeMantra to encode nearly 200 finding aids. As the project continues and we learn more about EAD and methods of presenting encoded finding aids online, we are exploring new ways to search and display them in our digital library.

As noted above, we have already created links to the finding aids through their cataloging records. But the EAD mark-up allows for more a sophisticated manipulation of data, which we are just beginning to envision. For example, dates and events from each chronology are encoded, and can be easily searched and retrieved. Therefore, as we continue writing chronologies and encoding more finding aids, we hope to manipulate or collate that data to create an online timeline of the many important individuals and events throughout the Seminary’s history. As the process of editing, encoding, and displaying our finding aids continues, we hope that you’ll enjoy exploring them!
The Future of Libraries and the Role of Metadata

Christine Schwartz

With the Internet’s rise in importance for information discovery, the role of a library is changing. Libraries need to accommodate digital resources and new ways of finding information. In fact, over the last fifteen years the web has become so ubiquitous it no longer feels new anymore. One thing is still uncertain: what this change means for the future of books, reading, scholarship, publishing, and libraries. Librarians are seeking to provide the best stewardship for library resources, both print and digital. We are trying to respond to users’ needs in an ever more complex information environment.

I have been considering these changes about one specific aspect of library work: cataloging. Just two years ago, I started writing a blog called Cataloging Futures whose focus is the future of cataloging and metadata in libraries (http://www.catalogingfutures.com). At the time, I was working as the head cataloger in Princeton Seminary’s Speer Library. For those unfamiliar with this aspect of librarianship, cataloging is the systematic process of organizing and providing access to all types of library materials: books, periodicals, DVDs, CDs, etc.

The work catalogers do to organize and provide access to digital resources is becoming an increasingly important part of the library. In fact, we have a new job title: metadata librarian. Just a year ago I was given that job and I now work exclusively with digital resources. So, as I explore these changes in the profession on the Cataloging Futures blog, I’m also experiencing them in my day-to-day work at Princeton Seminary. I’m developing a new skill set while still relying on my background in traditional library cataloging.

“Metadata,” or data about data, is the word used for cataloging digital resources. In fact, metadata can be found everywhere online. Wherever information about a web site or photograph is attached to the item itself, you have metadata.

The work we do to provide access to Princeton Seminary’s digital library (http://digital.library.ptsem.edu/) is quite different is several ways from traditional cataloging. One of the big differences is the workflow. With cataloging, we describe the physical items one at a time and add new items into the existing classified library collection. With digital resources, we usually work on a project that encompasses batches of items all at once. A digital project often includes several people collaborating together, with the metadata librarian as one member of the team.

Another difference is that traditional cataloging is a highly standardized, rule-driven system. The results of our work, bibliographic records, are globally shared among libraries in the international WorldCat database (http://www.worldcat.org/). We don’t recreate the wheel but instead use each other’s records in a process called “copy cataloging.” Metadata creation, on the other hand, is an area still new and under development. There are a variety of metadata schemas to choose from and the work is less standardized. Also, we are often describing unique special collections and archival materials and so the focus is on local needs, not sharing records. This means that individual libraries often choose metadata schemas and standards that suit their local needs for specific digital projects.

Metadata work also requires librarians to acquire better information technology skills. This past year I learned XQuery, a computer programming language for querying and manipulating XML data, the format of our metadata. In the future, much metadata work will be automatically generated, and so will be less labor intensive then traditional cataloging. The human role in metadata creation will be limited to adding value where computers cannot. Two examples where human review is necessary are complex subject analysis and name disambiguation, distinguishing one author’s name from another.

Metadata is also more complex than cataloging in the type of data we record. While cataloging deals with description and providing classification and subject access, metadata captures both of these as well as administrative, technical, preservation, rights, and structural information about digital resources. There is also added complexity working with the digital items themselves. Digital files require more work in the area of preservation. They are more difficult for libraries and archives to preserve over time.

So while the work libraries have always done remains, we face many challenges adapting to the web. The changes in cataloging illustrate just one aspect of the exciting changes on the way for Princeton Seminary’s historic library collections!