

# The Tokyo Trial at Richmond:

## Digitizing the Sutton Collection of Documents from the International Military Tribunal for the Far East

*by Chris Kemp and Suzanne Corriell*

**M**any librarians and readers are by now familiar with Laura Hillenbrand's book *Unbroken*, which has spent over a year on the *New York Times* best seller list for hardcover nonfiction. In *Unbroken*, Hillenbrand chronicles the story of Louis Zamperini, a one-time Olympian who was taken prisoner and tortured by the Japanese

while serving in the U.S. Army Air Forces during World War II. In her research for later chapters in the book, Hillenbrand routinely cites documents from the war crimes trials of those who perpetrated crimes against prisoners of war. The University of Richmond (UR) is fortunate enough to hold many of these documents in the special collections of the William T. Muse

---

**Chris Kemp** is the head of digital initiatives for Boatwright Memorial Library at the University of Richmond. He can be reached at [ekemp@richmond.edu](mailto:ekemp@richmond.edu).

**Suzanne Corriell** is the head of reference and research services for Muse Law Library at the University of Richmond. She can be contacted at [scorriell@richmond.edu](mailto:scorriell@richmond.edu).



A view of the IMTFE courtroom at the Japanese War Ministry Building in Tokyo, May 3, 1946.



*Accused war criminals in the prisoners' box, May 14, 1946.*

Law Library: the papers of David Nelson Sutton, a prosecutor during the Tokyo War Crimes Trial. A new project at UR hopes to bring many of these primary documents to the digital world.

David Nelson Sutton, a 1915 graduate of the University of Richmond and a 1920 graduate of the University of Virginia's law school, practiced law in West Point, Virginia, where he served as the commonwealth's attorney from 1928 to 1946. Invited to participate in the trial of General Tojo and other Japanese leaders after the end of World War II, he served as associate counsel to the prosecution before the International Military Tribunal for the Far East (IMTFE), better known as the Tokyo War Crimes Trial. The prosecution assigned Sutton to duties corresponding to his fluency in Chinese, as exemplified by his two trips to China to investigate

charges associated with the Rape of Nanking and Japan's illegal narcotics trade. He presented evidence and questioned witnesses during the course of the trial pertaining to these topics. At the end of the trial in 1948, Sutton returned to Virginia, resumed his law practice, and served as president of the Virginia State Bar Association from 1948–1949.

The Tokyo trial itself was just one of many judicial proceedings undertaken by the Allied powers in order to prosecute Japanese military and civilian leaders after the end of World War II. It was, however, the highest-profile trial. Justices from eleven countries (Australia, Canada, China, France, India, the Netherlands, New Zealand, the Philippines, the Soviet Union, the United Kingdom, and the United States) composed the tribunal, which received its au-

thority and charge from the supreme commander for the Allied powers, General Douglas MacArthur. Twenty-eight men, representative of the upper echelon of the Japanese military and government, were charged with waging aggressive war (otherwise known as crimes against peace), war crimes, and crimes against humanity. In many respects, the Tokyo trial followed the lead of the Nuremberg trials, which ran from November 20, 1945, to October 1, 1946.

Contrasted with the Tokyo tribunal's nearly two-and-a-half-year proceedings (April 29, 1946, to November 12, 1948), the Nuremberg trials seem to have proceeded in a relatively timely fashion. There are many reasons for the length of the proceedings in Tokyo, but a major cause was the need for language translation both during the proceedings and for the documenta-



*David Sutton presenting his summation of the charges against one of the accused to the IMTFE, February 20, 1948.*

*Below, portrait of David Sutton by Kōichi Okada.*



tion of the trial. Real-time translation took place to accommodate the numerous nationalities involved in the court's proceedings. Many long hours of discussion resulted from the phrases in English without equivalent meanings in Japanese, and vice versa. Further, although English was the official language of the trial, court-produced documentation was also translated into Japanese. Documentary exhibits also came to the court in numerous, sometimes vastly different, languages, which were then translated into Japanese and English for the convenience of defense attorneys,

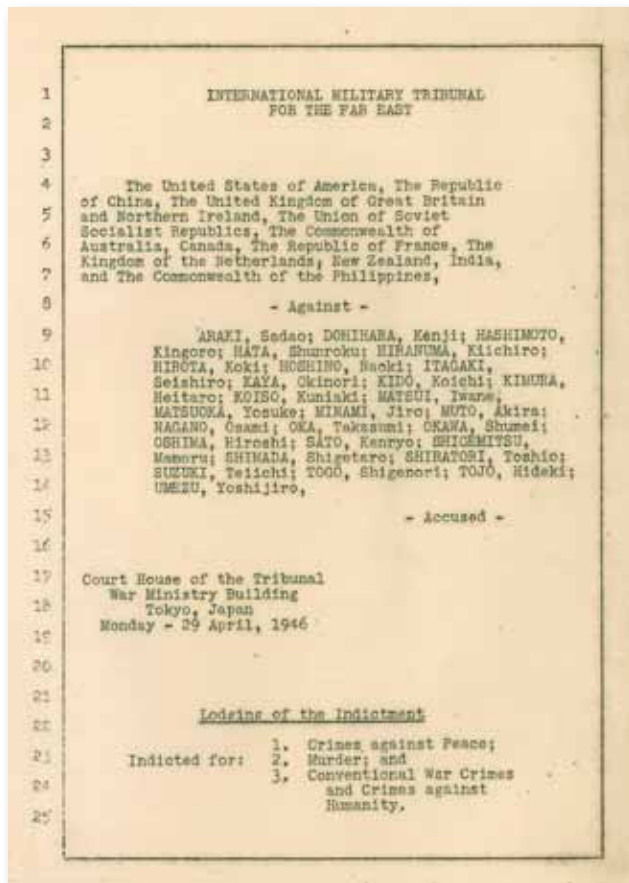
prosecutors, justices, and other officers of the court.

As a matter of course, attorneys from the prosecution and defense, and the members of the tribunal themselves, all received copies of the court-produced documentation. These documents comprise the vast majority of the portion of Sutton's personal papers that were given to UR's Muse Law Library following his death in 1974. The collection consists of approximately 85,000 pages of official trial transcripts, exhibits, depositions, opinions, and a set of Sutton's personal papers and artifacts from his time in Japan and China. Most of the Sutton Collection is not unique—many of the materials are duplicated in other archives around the world, and are as geographically close to the University of Richmond as the Virginia Historical Society and University of Virginia. Some of the documents have been gathered together and published previously in a condensed format, and some of the Japanese versions have been digitized and reproduced online by the National Archives of Japan as images linked to catalog records (<http://www.archives.go.jp/english/>).

At the university, an effort to organize the collection was led

by John Barden, currently the director at the Maine State Law and Legislative Reference Library. Following that initiative, the collection was identified as a candidate for digitization following the completion of the IMLS-funded *Richmond Daily Dispatch* project (<http://dlxs.richmond.edu/d/ddr/>) by the university's Digital Task Force, a working group consisting of staff members from Boatwright Library, Muse Law Library, and the Center for Teaching, Learning, and Technology.

Even though it was an incredibly important event in the formation of modern Japan, the Tokyo trial is, relative to other seminal war crimes trials, a somewhat unstudied event in history. Western scholarship pertaining to the event has been renewed during the last ten years, with several books being published by university presses on the subject. Almost in response to the unasked question, the Boatwright Memorial and Muse Law libraries at the University of Richmond began forming an ambitious plan to digitize the Sutton Collection in 2008. The overarching goal was (and remains) to build a standards-based and openly accessible representation of the collection with academic and legal research in mind.



Left, page one of the court-produced trial transcripts.

Below, a map from Exhibit 1255, "Sketch Map of Pearl Harbor after the Attack of 7 December 1941."



Based on the volume of documents in the collection and the limited staff resources at UR, we decided to outsource the production of digital images and text. After the initial planning period, production began in the fall of 2009. The ongoing workflow calls for the careful packaging of each folder of archival materials in acid-free and watertight shipping materials and sending them to the vendor's imaging facility. There they are imaged according to the Library of Congress's *Technical Standards for Digital Conversion of Text and Graphic Materials* (<http://memory.loc.gov/ammem/about/techStandards.pdf>), and derivative TIFF and JPEG files are created. Optical character recognition (OCR) software extracts text from the images, allowing basic markup to be performed on the corrected OCR output. The vendor returns

the documents and all data files to the university on external hard drives, where each physical page is reviewed to make sure that every page that was sent gets returned. The internal process of quality control begins at that point, with staff members and student employees ensuring the integrity and accuracy of all data returned by the vendor.

This project, like many of Boatwright's other text-based digital projects, uses XML (Extensible Markup Language) in the TEI (Text Encoding Initiative) format to encode bibliographic metadata, metadata internal to the documents, and the structure of the documents themselves. TEI markup allows wonderful benefits—enhanced full-text searching, named entity identification and normalization, increased utility for online applications—but it is expensive

in terms of labor. Choosing a vendor that performs data operations overseas helped mitigate the cost of basic, structural TEI markup, but we opted to encode the documents more in-depth locally. This allows more precise control over the intricate TEI structures that are required to perform complex data manipulations and present the texts in different ways. For example, specific XML tags are required within the files to normalize personal and organizational names throughout the collection, to link entire documents or portions of documents to others, and to georeference place names. Proper TEI tagging, combined with the use of predefined thesauri and a lot of hard work, will enable interesting and potentially revealing visualizations of the resulting data.

As of the end of 2011, the project has progressed substantially.

The vendor has imaged nearly 90 percent (about 76,400 pages) of the collection and delivered basic TEI files for just over 80 percent. In what may have been an overabundance of caution, quality control of the data files returned to UR started off a little more slowly than anticipated. However, staff and volunteers from the Boatwright and Muse libraries were quick to learn and generous with their time, thus swiftly closing the gap. All basic TEI files corresponding to the trial proceedings—almost 50,000 pages of text—have been checked and enhanced by departmental staff members or volunteers and verified by the project coordinator. The project will be making its first appearance in the classroom during the spring semester of 2012, when

Dr. Jan French from UR's Department of Sociology will use digital versions of selected trial transcripts and exhibits in a course focused on human rights.

The choice of repositories for the online version of the archive seemed at first to be a simple one. For the last five years, UR has been using DLXS, a digital access and repository system developed by the University of Michigan's Digital Library Production Service, to display collections online. DLXS has done the job fairly well, but several issues have required UR to move away from that system. Among these concerns are streamlining workflows, focusing on data design and project management rather than programming and web design, and integrating

digital library projects more fully into the systems architecture of the university. As a direct result of the focus on deeply encoded texts, the library selected the eXist native XML database as its data management tool for collections. The university's Web Services Department, which is responsible for the content management system and all public-facing websites for the university, had been working to incorporate eXist into the public site in numerous ways, most notably as the application serving data to the online directory. The library's data-driven digital projects were a natural fit for the system. In collaboration with the departments of Web Services, University Communications, and the Digital Scholarship Lab, Boatwright demonstrated



*The justices of the International Military Tribunal for the Far East, July 22, 1946.*



it could quickly and successfully mount an eXist-based collection in the fall of 2010 with the *Secession: Virginia and the Crisis of Union* project (<http://collections.richmond.edu/secession/>). The benefits of using eXist have been numerous within the library, but the advantages of developing relationships among administrative, technical, and academically focused departments on a small campus cannot be overstated.

Tangentially related to the use of TEI and the eXist database, an ongoing side project aims to develop a standards-based TEI annotation application with the help of partners and consultants. Coding detailed TEI is an exceptionally labor-intensive process, and our libraries have limited personnel resources. The primary purpose of the TEI annotator is to reduce the number of hurdles required to annotate and enhance TEI XML files through the use of a graphical, web-based interface. Running within a web browser, the application would equip subject experts or students who have little or no experience looking at TEI markup or XML code in general with a mechanism to make annotations within transcriptions of primary source documents.

We hope that the TEI annotator will allow us to integrate the development of semantic markup into the curriculum, allowing students to critically interact with primary source materials and providing a valuable service to the encoding process at the same time. The system runs on an eXist database and is written primarily in JavaScript, XSLT, and XQuery, with XML files that contain project-specific configuration data. While the application is still in development, substantive progress has been made, and library staff members are working to code the controlled vocabularies and thesauri necessary to enable named entity and

geographic location normalization with the system. The University of Richmond is excited by this work for its potential application to other local digital projects, and we hope that it will enable other institutions with similar challenges to begin development of semantically rich, text-based digital collections.

Outreach efforts have been ongoing since the planning stages in

---

### **The University of Richmond is excited by this work for its potential application to other local digital projects ....**

---

2009. Dr. Tze Loo of the university's History Department has a deep research interest in the Tokyo trial and has served as an advisor and enthusiastic supporter of the project, pointing the team in a number of valuable directions. Loo's suggestion that the project team make contact with leading researchers on the subject bore immediate fruit, and ultimately led to a site visit from Dr. Morten Bergsmo, director of the International Criminal Court's Legal Tools Database (LTD) project. The aim of the LTD is grand: in an effort to equip legal practitioners in developing nations with the information and tools they need to prosecute war crimes, crimes against humanity, waging aggressive war, and genocide, the project is attempting to compile all primary source documentation dealing with international criminal law. The LTD is a completely open resource, yet one of its major gaps has been the Tokyo trial material. In early 2011, the Muse Law Library, as the archive of record, signed a partnership agreement with the ICC to deposit its docu-

ments and metadata into the LTD. Partnering with this important and freely available resource will provide an additional access point to the Tokyo trial materials and allow UR to contribute internationally to the greater good.

Additional contacts have been made with other organizations that are planning similar projects that have not yet begun. These include the University of Canterbury in New Zealand and the Netherlands Institute for War Documentation in Amsterdam. We are maintaining informal contact with these organizations in hopes that the data produced among all the projects will be interoperable, creating the potential for virtual access to a complete set of all court-produced documents.

As an ongoing project, the effort to digitize and present the Sutton Collection is far from complete. Our effort has the potential to become a leading resource for materials relating to the Tokyo trial and, with the help of our faculty partners, to demonstrate relevancy of the trial to current issues in international criminal law and to the development of Japan's role in modern East Asia. As the project team learns more about the collection, consults with similar projects, and continues to implement innovative applications, processes are constantly updated. The coming year should bring further progress, and we look forward to getting a portion of the materials online, making contact with other similar projects, and becoming more deeply involved with the Legal Tools Database project. It is UR's hope that its endeavor to digitize and present the entire Sutton Collection online will spark new scholarship and understanding of the trial while highlighting the contributions and experiences of a Virginian during this significant episode in twentieth-century history. **VI**