

The Tibet Oral History Archive Project and Digital Preservation

Linda Cantara (linda.cantara@case.edu)
Case Western Reserve University

The *Tibet Oral History Archive Project*¹ (TOHAP) is part of the research and education program of the Center for Research on Tibet in the Department of Anthropology at Case Western Reserve University.² The Center was created in 1987 by Melvyn Goldstein, John Reynold Harkness Professor of Anthropology, and Cynthia Beall, Sarah Idell Pyle Professor of Anthropology, to generate and disseminate new knowledge about Tibetan culture, society, and history, and was the academic pioneer in opening Tibet to in-depth anthropological and historical research. The TOHAP builds on a series of fieldwork-based studies that have examined the adaptation of Tibetans to high altitude, and the changes that have occurred since Tibet's incorporation into the People's Republic of China in 1951.

The *Tibet Oral History Archive* includes three primary collections:

- *The Common Folk Oral History Collection*: nearly 2,000 hours of interviews with hundreds of ordinary rural and urban Tibetans about their life experiences. Since the number of individuals in Tibet who were adults in 1959 -- the end of the traditional era -- is rapidly dwindling, there is particular urgency to document the voices of ordinary Tibetans in order understand the diversity of life as it was lived in Tibet as well as the way the salient historical events played out among the different strata of society.
- *The Political History Collection*: approximately 400 hours of historical interviews with former Tibetan government officials who played important roles in modern Tibetan history, including His Holiness the Dalai Lama. These interviews cover the traditional period before Tibet was incorporated into the People's Republic of China (1913-1951) and the subsequent period up to the end of the Cultural Revolution in 1976.
- *The Drepung Monastery Collection*: approximately 350 hours of interviews with about one hundred monks who were members of Drupung Monastery, Tibet's largest monastery, at the end of the traditional era. These interviews are unique in that they provide the only in-depth window into large-scale monasticism in traditional Tibetan society.

Conducted primarily in the Tibetan language, the interviews were taped on audio cassettes which have subsequently been digitized in three formats: archival WAVE files, medium format QuickTime files, and compressed delivery MP3 (MPEG) files. The interviews have been transcribed and translated into English and were initially saved as Microsoft Word documents. Professor Goldstein, Editor of the Archive, has partnered with Kelvin Smith Library to prepare the audio files and transcripts for online dissemination and long-term preservation. For online dissemination via the World Wide Web, we are converting the Word documents to plain text and encoding them in XML using the Text Encoding Initiative (TEI) Document Type Definition (DTD) for Transcriptions of Speech.³ To facilitate understanding, the Archive will also include a glossary of terms, encoded in XML using the TEI-DTD for Printed Dictionaries.⁴ A programmer has been hired to create a Web-based tool for creating the glossary and an application for automatically encoding extended pointer notation to link terms in the transcripts to their definitions in the glossary. Work is also underway to design an end user interface which will include browse and search functions. In the meantime, we are temporarily transforming the XML files to XHTML and using the *Greenstone Digital Library Software* to facilitate local access.⁵

A larger concern, however, is how to ensure long-term preservation of and access to the Archive. In 1996, the Commission on Preservation and Access (CPA) and Research Library Group (RLG) Task Force on Archiving of Digital Information published a seminal report on the long-term preservation of digital resources.⁶ Since then, virtually every significant publication about digital preservation has indicated that primary responsibility for initiation and management of the metadata necessary to ensure long-term access to digital resources begins with the creator of the resource. Traditionally, it has been the role of librarians and archivists to ensure long-term viability of and access to cultural heritage materials, but this is not within the realm of expertise of the majority of scholars in the humanities and social sciences. Thus, if the creators of digital resources are responsible for initiating lifecycle documentation of the descriptive, administrative, and structural metadata necessary to migrate, emulate, or otherwise translate existing resources to future hardware and software configurations -- a task foreign to most discipline-based scholars -- close collaboration with information technology professionals early in a project is imperative.

Protocols and standards for digital preservation are now under vigorous development, yet there are still many unknowns. For the short-term, multiple copies of the audio and XML files will be maintained in multiple locations at Case Western Reserve University, both at the *Center for Research on Tibet* as well as in *Digital Case*, Kelvin Smith Library's *Fedor* repository.⁷ For the long-term, the Asian Division of the Library of Congress

has expressed interest in hosting the completed Archive. To prepare the *Tibet Oral History Archive* for deposit with the Library of Congress, we are creating a Submission Information Package (SIP) in compliance with the *Reference Model for an Open Archival Information System (OAIS)*,⁸ using the Metadata Encoding and Transmission Standard (*METS*), a metadata standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library.⁹ This paper will present a prototype for scholar-librarian collaboration in the digital preservation of multimedia resources, including a discussion of the practical aspects of constructing a *METS* document for the *Tibet Oral History Archive*, with particular attention to the multiple metadata standards that must be bundled with the digital files to create a robust Submission Information Package.

-
1. This project is sponsored by the Henry Luce Foundation with additional support from the National Endowment for the Humanities (grant no. RZ-20585-00) and the National Geographic Society.
 2. The Center for Research on Tibet Web Site is <http://www.case.edu/affil/tibet/index.htm>.
 3. Chapter 11 of the *TEI Guidelines (P4)*; see <http://www.tei-c.org/P4X/TS.html>.
 4. Chapter 12 of the *TEI Guidelines (P4)*; see <http://www.tei-c.org/P4X/DI.html>.
 5. *Greenstone* is open source software for building and distributing digital library collections, produced by the *New Zealand Digital Library Project* at the University of Waikato, and developed and distributed in cooperation with *UNESCO* and the *Human Info NGO*. See <http://www.greenstone.org>.
 6. Commission on Preservation and Access (CPA) and Research Library Group (RLG). *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. May 1996. Online at <http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>.
 7. *Fedora™ Flexible and Extensible Digital Object Repository Architecture* -- is an open source digital repository management system, developed by Cornell University and the University of Virginia, available at <http://www.fedora.info>.
 8. A SIP is "an information package that is delivered by the producer [of a digital object] to the *OAIS* for use in the construction of one or more AIPs [Archival Information Packages]." See "OAIS Terms". *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Cornell University Library. 2003. Online at <http://www.library.cornell.edu/iris/dpworkshop/working/terminology/oais.html>. See also, *Consultative Committee for Space Data Systems (CCSDS). Reference Model for an Open Archival Information System OAIS*. CCSDS 650.0-B-1. ISO 14721:2003. January 2002. Online at <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>.

9. *METS* is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the *Digital Library Federation*. See <http://www.loc.gov/standards/mets>.

Bibliography

The Center for Research on Tibet's Web Site. Accessed 2005-03-29. <http://www.case.edu/affil/tibet/index.htm>

"Chapter 11: Transcriptions of Speech." *TEI Guidelines (P4)*. Text-Encoding Initiative. Accessed 2005-03-29. <http://www.tei-c.org/P4X/TS.html>

"Chapter 12: Print Dictionaries." *TEI Guidelines (P4)*. Text-Encoding Initiative. Accessed 2005-03-29. <http://www.tei-c.org/P4X/DI.html>

Fedora. Cornell University and the University of Virginia. Accessed 2005-03-29. <http://www.fedora.info>

Greenstone. University of Waikato. Accessed 2004-07-16. <http://www.greenstone.org>

METS. Digital Library Federation. Accessed 2005-01-25. <http://www.loc.gov/standards/mets>

"OAIS Terms." *Digital Preservation Management: Implementing Short-term Strategies for Long-term Problems*. Cornell University Library. Accessed 2005-03-29. <http://www.library.cornell.edu/iris/dpworkshop/working/terminology/oais.html>

Reference Model for an Open Archival Information System (OAIS). CCSDS Secretariat. Accessed 2002-01. <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

Waters, Donald, and John Garrett. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Accessed 2005-03-29. <http://www.rlg.org/legacy/ftpd/pub/archtf/final-report.pdf>