Digitization of Buddhist cultural heritage

Marcus Bingenheimer

Throughout history, Buddhists have used all available means to encode and transmit the ever increasing volume of their textual heritage. After the death of the founder of Buddhism, the early community organized the transmission of a sizable corpus with the help of mnemonic recitation techniques. The earliest Indian epigraphy as well as the earliest manuscript fragments in Indian languages are connected with Buddhism, and the earliest extant printed book, dated 868 CE, is a Chinese translation of the Diamond Sutra.

Today, in the twilight of print, text is largely produced, transmitted, and stored digitally and, for better or worse, cultural heritage information is being digitized ever more comprehensively. In the field of Buddhist studies, texts were a natural starting point for digitization. Buddhist texts exist in a bewildering range of languages and genres, and there are several large canonical collections in Pāli, Chinese, Tibetan, Mongolian, and Manchu that overlap in complicated ways. Many texts have also survived in Sanskrit and prakritic languages, sometimes complete in the monasteries of Nepal and Tibet, sometimes fragmentary in the sands of South and Central Asia. Then there are modern translations into Japanese, Korean, Vietnamese, French, English, German, etc.

Since the late 1980s, various organizations have started to digitize these riches, scanning manuscripts and producing digital full text editions. Distributed online, vast amounts of Buddhist literature are now available, equally and freely, to the wider public. The effects on Buddhism of making all its texts available to all believers with an Internet connection are not yet fully understood, but the impact could be significant —comparable to that of the adoption of writing in Buddhism (which played a major role in the emergence of

Mahāyāna) or the discovery of printing in Europe (which was a condition for the Reformation).

Where to find Buddhist canonical texts online in reliable form? For Pāli the most widely used digital corpora are the *Chaṭṭha Saṅgāyana* corpus, the *Buddha Jāyanti* corpus, and the digitized version of the *Pāli Text Society* edition. For early Buddhist literature in general, *SuttaCentral* offers parallel full-text in ancient languages and the largest number of translations from Pāli texts into modern languages. It also makes all its data available in an exemplary fashion for download.

For the Chinese canon there is the *Taiwanese Chinese Buddhist Electronic Text Association* (CBETA) corpus, and the *Japanese SAT Daizōkyō Text* Database. Translations of Chinese Buddhist texts are less readily available online. An online bibliography of translations from the Chinese Buddhist canon shows that so far about 520 of c. 5500 pre-modern Chinese Buddhist texts have been translated into European languages, but not all of them are available digitally.[1] Other projects offer scans of manuscript collections that contain a large amount of Buddhist material. The *International Dunhuang Project*, for instance, offers scanned images of the manuscript witnesses for Chinese Buddhist texts, and the *Digital Library of Lao Manuscripts* preserves the rich heritage of Laotian manuscript culture.

Most of these datasets and initiatives are openly accessible, and many, but unfortunately not all, projects share their data freely via their websites or version controlled repositories such as Github. The digital data on offer now surpasses by far any single canonical print collection in terms of volume, acquisition cost, searchability, and portability.

While the digitization of texts has been quite successful, others aspects of Buddhist heritage digitization are less advanced. With a few notable exceptions, such as the *Huntington Archive,*[2] the high-end digitization of images, objects, and spaces has just begun. Many museums today make digital images of their holdings available, but an archive with faceted search across institutions and geared to Buddhist iconography still needs to be built. The 3D scanning and printing of Buddhist objects and sacred spaces are still at an early stage of development, but have strong potential for both teaching and research.

For scholars, one of the benefits of digitization is that we are now able to use computational methods to explore the language, the historical geography, the social networks and other facets of the Buddhist tradition in new ways. Individual researchers have taken steps into this direction using computational analysis, for instance,

to re-assess the attribution of translations, or to create data for historical social network analysis.[3] The challenge is now to integrate these new approaches into mainstream research and for graduate programs in Buddhist Studies to include training in digital methods and datasets.

Marcus Bingenheimer is an Associate Professor in the Department of Religion at Temple University
bingenheimer@temple.edu

Notes

1 http://mbingenheimer.net/tools/bibls/transbibl.html
2 www.huntingtonarchive.org and www.huntingtonarchive.osu.edu
3 See the attribution database by Michael Radich http://dazangthings.nz or emerging datasets for historical social network analysis http://mbingenheimer.net/tools/socnet

Language	Database	Website
Pāli	Chaṭṭha Saṅgāyana Pāli Text Society Corpus (at GRETIL) SuttaCentral	http://tipitaka.org http://gretil.sub.uni-goettingen.de https://suttacentral.net https://github.com/suttacentral
Chinese	Chinese Buddhist Electronic Text Association (CBETA) SAT Daizōkyō Text Database	http://cbetaonline.dila.edu.tw http://21dzk.l.u-tokyo.ac.jp/SAT/index_en.html
Tibetan	Asian Classics Input Project Buddhist Digital Resource Center Buddhist Canon Research Database Resources for Kanjur & Tanjur Studies Tibetan and Himalayan Library (THL)	http://www.asianclassics.org https://www.tbrc.org http://databases.aibs.columbia.edu https://www.istb.univie.ac.at/kanjur/rktsneu/sub/index.php http://www.thlib.org
Sanskrit	Göttingen Register of Electronic Texts in Indian Languages & related Indological materials from Central & Southeast Asia(GRETIL) Digital Sanskrit Buddhist Canon	http://gretil.sub.uni-goettingen.de www.dsbcproject.org
Multiple	International Dunhuang Project Digital Library of Lao Manuscripts	http://idp.bl.uk http://www.laomanuscripts.net

---

# Digitization projects at the Cultural Heritage Research Institute, Zhejiang University

Zhirong Li and Changyu Diao

The team at the Cultural Heritage Research Institute of Zhejiang University embarked on a series of major digitization projects in 2010. Headed by an archaeologist and a scholar in image processing, our members come from various disciplines including computer science, archaeology, art history, and digital humanities. Our mission is to establish a high-standard, comprehensive digital database of the cultural relics in China for the purpose of conservation, research and education.

At present, the team has conducted digitization work at more than a hundred archaeological sites, museums, and cultural institutions across twenty different provinces, cities, and autonomous regions in China. Our works encompass large-scale monuments such as historic architecture and Buddhist cave temples, and museum collections ranging from textiles to paintings, calligraphy, porcelains, and statues. We aim to maintain state-of-the-art technological standards in the process of scanning, archiving, preserving, and presenting cultural objects and sites.

Recently, with the collaboration of the Yungang Academy, we have successfully printed a to-scale 3D model of the rear chamber of Yungang Cave 3. This marks a significant advance in the digital conservation and reconstruction of cultural heritage in China.

Zhirong Li lzr5335@163.com and **Changyu Diao** diaochangyu@gmail.com are Assistant Directors of the Cultural Heritage Research Institute, Zhejiang University.
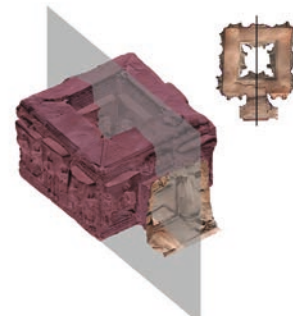
Left and below: Digital scans of the Xumishan Cave 45 in Ningxia.

Bottom: Scanned image of the east wall mural of the White Hall at Tholing Monastery, Ngari Prefecture, Tibet.