Oxford Bibliographies
Your Best Research Starts Here

Digitization of Buddhism (Digital Humanities and Buddhist Studies)

Marcus Bingenheimer

LAST MODIFIED: 24 SEPTEMBER 2020
DOI: 10.1093/OBO/9780195393521-0265

Introduction

The period between 1990 and 2010 saw a momentous change in the way humans store information. The transition from a society that encodes its information mainly in analogue ways to one that relies mainly on digital media has far-reaching consequences for each of its subsystems, including religion and academia. The well-understood materiality of analogue media, which encode information in unique, persistent, easily addressable items, which are embedded in economic and legal arrangements, has been replaced by a regime where most information is encoded digitally. Computationally mediated, digital information can be quickly produced, changed, multiplied, and transmitted, but is always reliant on a many-layered infrastructure of network, hardware, and software standards. How is Buddhist heritage digitized and how does that impact Buddhist studies? Buddhists, from the very beginning of their tradition, have often been "early adopters" and eager to use whatever new media were available to store, manage, and transmit their cultural heritage. With the advent of writing in India, Buddhism is mentioned in the earliest examples of Indian epigraphy (the Aśokan edicts, 3rd century BCE), and the oldest surviving Indian manuscripts (*c.* 1st century CE) are of Buddhist texts. In China, Buddhism became the first religion to make use of printing to copy their sacred scriptures. Famously, the earliest dated printed book (868 CE) is a Chinese version of the Diamond Sutra. In Buddhist studies, like in other fields of academic inquiry, researchers had to learn within a generation to digitally access and manage *primary sources* (see Digitization of Primary Sources) and *research tools* (see Digitization of Research Tools). *Cyberspace* has become a new frontier for research into contemporary Buddhism (see Buddhism in Cyberspace). Similar to other fields in the Humanities, the application of *research methods specific to digital data* (see Application of Computational Methods in Buddhist Studies), however, is still in its infancy. This article is neither a link list, nor a bibliography in the traditional sense, but an attempt to survey the landscape of initiatives and approaches toward the use of computational methods in Buddhist studies. To prevent link rot, I cite URLs only where projects are not easily findable via a simple online search for their name. Most of the resources in this article are the product of teamwork; very few are created by a single person alone. Because of this, I generally forgo mentioning individuals, focusing instead on the institutions that maintain a resource. Acronyms are only given where they are widely used.

Digitization of Primary Sources

The primary sources for research in the Humanities are texts. Digitization in the Humanities means first and foremost to model analogue representations of text in digital form. Beyond "text" in the narrow sense, this ideally includes the textures of music, art, ritual practice, architecture, and other aspects of human creativity. One problem in reporting on digital scholarship is the fluid nature of digital information. Online resources are ephemeral and URLs are not reliable over time. Although an ecosystem of standard frameworks (e.g., open archival information system, or OAIS), stable identifiers (e.g., digital object identifier, or DOI; Purl), and trusted digital repositories (e.g., Zenodo, Dryad) for data has evolved, few of the major initiatives in Buddhist

studies partake in it so far.

Digital Canonical Collections

Although what counts as a "canon" varies considerably between traditions, canonicity is a common denominator. As such the "canon" is a natural starting point for full-text digitization. Between *c*. 1990 and 2010, canonical editions in Pali, Chinese, and Tibetan as well as a large number of Buddhist Sanskrit texts were digitized from various print editions. Digital canonical collections were driven by the desire to disseminate the texts and to make them searchable. Buddhist digital texts are thus generally freely accessible, but often difficult to assess and cite. Even the most advanced digital editions, such as Chinese Buddhist Electronic Text Association (cited under Chinese), are still generally cited to their print originals. We have not yet managed to create digital editions that are recognized and cited as independent critical editions in their own right.

Pali

Digital Pali Buddhist texts are readily available online, but often suffer from a lack of technical documentation and metadata. Digital Pali texts were created as part of independent digitization initiatives in India, Sri Lanka, and Thailand, but later often copied and aggregated elsewhere on the web. In the 1990s, the Thai BUDSIR (Buddhist Scriptures Information Retrieval) initiative produced one of the earliest digital editions, but the texts seemingly never made it beyond the CD stage.

Buddha Jayanti Tripitaka Project.

The Sri Lankan Buddha Jayanti Tripitaka Project digitized the Pali canon based on the government-sponsored Sinhalese Buddha Jayanti print edition (1956–1990). Anecdotal evidence suggests that the digital version of the Buddha Jayanti edition is less well proofread than the VRI corpus. It can be found on various websites. By far the most stable way of accessing the Buddha Jayanti corpus is via the Göttingen Register of Electronic Texts in Indian Languages (cited under Sanskrit).

Digital Chaṭṭha Saṅgāyana Edition.

Organized by the Vipassana Research Institute (VRI; Dhammagiri, India) based on the Sixth Council Burmese edition of the Pali canon (compiled 1954–1956). Widely copied online, it is the only digital version of the Pali canon that includes the commentaries (*aṭṭhakathā*) and sub-commentaries (*ṭika*). The current official online version (2020) on the VRI website is *Chaṭṭha Saṅgāyana Tripitaka*, Ver. 4.0. The best interface for this corpus is the Digital Pāli Reader by Yuttadhammo, available online.

Pāli Text Society corpus.

A digital corpus of the widely used Pāli Text Society (PTS) edition was produced in collaboration with the Dhammakaya Foundation, Thailand, between 1989 and 1996. The texts are now distributed on Göttingen Register of Electronic Texts in Indian Languages (cited under Sanskrit) under a Creative Commons Attribution-ShareAlike 4.0 license. The digital PTS corpus so far consists mainly of the basic *mūla* texts of Vinaya, sutta, and abhidhamma; only a few of the commentarial and paracanonical works from the PTS print series are currently available digitally.

Chinese

Like with Pali and Tibetan, there are several independent digitization projects for the Chinese canon. These were organized by groups in Hong Kong, Taiwan, Japan, and Korea. Due to Chinese, Japanese, Korean (CJK) character encoding issues, the creation of digital editions of Chinese texts has been especially challenging. Currently, the two main collections are the Chinese Buddhist Electronic Text Association (CBETA) corpus and the SAT Daizōkyō Text Database (SAT, or *Saṃgaṇkīkṛtaṃ Taiśotripiṭakaṃ*, "Society for the Creation of the Taishō Tripiṭaka"). Both started out collaboratively with the aim to digitize the Taishō canon (Tokyo, 1924–1934), the most widely used edition of the Chinese Buddhist canon. However, their text base has diverged in the early 21st century and the actual overlap of searchable text between them consists only of about 2270 texts (Taishō Volumes 1–55 and Volume 85), that is, the Indian scriptures translated into Chinese, and the works composed in Chinese until *c*. 8th century by Chinese, Korean, and Japanese Buddhists. In addition to these, the SAT interface searches other *c*. 1220 Buddhist texts from the Taishō canon (Volumes 56–84 and Volumes 86–97), which are not contained in the CBETA corpus. These comprise mostly texts by Japanese authors written after the 8th century. The CBETA corpus, on the other hand, contains approximately 2350 additional digital texts from various sources, which are not accessible through the SAT website. These are mostly texts by Chinese authors written after the 8th century. As a rule of thumb, when studying Japanese Buddhism (or the Japanese commentarial tradition on Indian and Chinese texts), one ought to work with the SAT website. For research on Chinese Buddhism, one should make use of the latest version of the CBETA corpus.

Chinese Buddhist Electronic Text Association.

Maintained by Dharma Drum in Taiwan, the CBETA corpus can be accessed online and, more conveniently for searching, online here as well. It is the only Buddhist Chinese corpus that is available freely for download in multiple formats (EPUB, XML, pdf, etc.). Crucial for researchers in digital humanities (DH) is the CBETA GitHub repository, online. The CBETA corpus offers revised punctuation for many Taishō texts and has expanded the Taishō apparatus by increasing the number of witnesses and by adding emendations. The 2018 version of the corpus contains approximately 4,620 texts.

SAT Daizōkyō Text Database.

The SAT corpus, maintained by at Tokyo University, is reliably accessible via its website. The interface allows for searches of the complete Taishō canon (3,283 texts) as well as the *Jōdoshū zensho* 浄土宗全書, the "Collected works of the Pure Land School." Helpful for translators is the linking of highlighted texts with the Digital Dictionary of Buddhism (cited under Dictionaries, Authority Databases, and Network Data), and to the English translations published by Bukkyō dendō kyōkai (aka Numata Foundation).

The Tripitaka Koreana Knowledgebase Project.

Organized by the Research Institute of the Tripitaka Koreana (Seoul). Like SAT, which is a digital edition of the Taishō, the Tripitaka Koreana Knowledgebase Project is an attempt to carefully model one particular edition, in this case, the first and second printing of the "Korean edition" of the Chinese canon (14th century). The project has been intermittently online since at least 2000, but currently seems to be dormant.

Tibetan

There are two main independent digital collections of Tibetan Buddhist texts. The US-based Asian Classics Input Project (ACIP), and the Buddhist Digital Resource Center (BDRC) (formerly the Tibetan Buddhist Resource Center, or TBRC). Next to these, the Austrian Resources for Kanjur & Tanjur Studies is an important clearing house for Tibetan online editions.

Asian Classics Input Project.

Since 1987 ACIP, an independent nonprofit organization, has produced digital plain-text versions of the Kangyur (*bka' 'gyur*), the Tangyur (*bstan 'gyur*), and various Sungbum collections. The digital texts come without metadata about their provenance, which limits their use for research. As a whole, however, the corpus might be used for corpus linguistic analysis and related approaches.

Buddhist Digital Resource Center.

BDRC, a nonprofit founded by E. Gene Smith (b. 1936–d. 2010), has digitized, catalogued, and archived a large number of Tibetan texts. Most works are distributed under a Creative Commons license. A few texts are restricted based on cultural commitments to stakeholders. Most texts contained in the BDRC corpus are distributed as scans (in pdf), others are available as full text. Both come with copious metadata.

Resources for Kanjur & Tanjur Studies.

The Resources for Kanjur & Tanjur Studies are part of the Tibetan Manuscripts Project Vienna at the University of Vienna. It provides both transcriptions and scanned sample images of a range of printed and handwritten editions, as well as links to digital editions from other organizations (e.g., ACIP, Adarsha).

Sanskrit

No complete "canon" of Buddhist texts in Sanskrit has survived, and indeed many extant Sanskrit Buddhist texts might never have been considered "canonical" in India. Modern print editions of Buddhist Sanskrit texts often rely on manuscripts found in Nepal, Tibet, and India. There are two main efforts to digitize Buddhist Sanskrit, usually based on modern print editions of individual texts. The first is the Göttingen Register of Electronic Texts in Indian Languages (GRETIL) and related Indological materials from Central and Southeast Asia. Though covering Indology in general, GRETIL contains a sizable collection of digital Buddhist texts in Sanskrit (and Pali and other Prakrits). The second is the Digital Sanskrit Buddhist Canon (DSBC).

Digital Sanskrit Buddhist Canon.

Since 2003 maintained by the University of the West, in collaboration with partners in Nepal, the DSBC is a dedicated collection of Buddhist Sanskrit texts. The (undated) introduction online says "369 scriptures are freely offered" on the DSBC website. Metadata for the digital texts is minimal and the interface lacks faceted search. The texts are not packaged for download. Many, but not all of the texts, are also shared via GRETIL.

Göttingen Register of Electronic Texts in Indian Languages (GRETIL).

GRETIL is among the oldest and best maintained collections in its field. It has avoided the usual mistake of focusing on the interface and instead consistently offered well-digitized texts with basic metadata for download individually or aggregated in zip files. This includes approximately 270 Buddhist Sanskrit texts as of 2020. Thus, the GRETIL corpus enables sophisticated corpus linguistic approaches as well as searches across user-defined sections of the corpus.

Manuscript Collections

Next to collections of digital plain text, there are a number of digital collections focusing on manuscripts. While digital plain text is often normalized in one way or other, epigraphy or manuscript studies rely on unique witnesses of a text. Many of the Buddhist scriptures that are part of the Dunhuang corpus have been made digitally available via the International Dunhuang Project (IDP), which has provided crucial infrastructure to the field of Dunhuang studies by making the material more widely accessible. There also are several projects that aim to preserve the rich manuscript heritage of Southeast Asia.

Database of Khmer Manuscripts.

Hosted and maintained by the École Française d'Extrême Orient, the database consists of digitized negatives of photographs taken between 1990 and 2005. There seems to be no metadata associated with the digital facsimile. As of March 2020, the online database search interface does not seem to work.

The Digital Library of Lao Manuscripts.

Provides scans of Laotian Buddhist manuscripts (15th to 20th century). Images of approximately twelve thousand texts are findable by text title, ancillary term, language, script, category, material, location, and date via an exemplary faceted search function. Made available both via an online interface as well as packaged for download with professional metadata.

Digital Library of Northern Thai Manuscripts.

The Digital Library of Northern Thai Manuscripts is currently the main clearinghouse for several digital manuscript collections from Northern Thai (Lanna). Together approximately six thousand manuscripts in various scripts (Thai, Burmese, Shan, Lanna, Lue, etc.) are made available as scans with basic metadata.

International Dunhuang Project.

Dunhuang manuscripts are relevant for the study of late Indian, medieval Chinese, and early Tibetan Buddhism. Hosted by the British Library, the IDP has brought together collections from the United Kingdom, France, Germany, Russia, China, and Japan. Although not all collections have been fully digitized so far (about 30 percent of the Stein collection remains unscanned), and not all that is digitized is released, much has been made available and is distributed via the IDP website and its mirrors (many in dire need of maintenance). IDP currently offers access to over half a million images of over one hundred thousand manuscripts, paintings, artifacts, and photographs.

Myanmar Manuscript Digital Library.

Hosted at the University of Toronto, this is a database of Burmese palm-leaf manuscripts, scanned in situ in temples and libraries in Myanmar.

Aligned Online Editions

Besides large-scale canonical editions and digital facsimile collections, another option for digital text is to produce high-end

digital editions of individual texts. Such editions offer user-defined views of the text and can present complex, multilingual text clusters through hyperlinks and alignment. Many Buddhist texts are ideally read in clusters of parallel texts in the same or other languages. Thus aligned editions are useful, but producing and maintaining stable, digital aligned editions is labor intensive, and so far cannot be automated. The Open Philology project at Leiden is currently working on issues of automated alignment specifically for Buddhist literature.

Digital Comparative Edition of the Shorter Chinese Saṃyukta Āgama (T.100).

This detailed, aligned Text Encoding Initiative (TEI) edition with all Chinese, Pali, Sanskrit, and Tibetan parallels of the Bieyi za ahan jing (T.100) has been prepared for a project at Dharma Drum, Taiwan. The TEI source files are available.

Thesaurus Literaturae Buddhicae.

Developed as part of the Bibliotheca Polyglotta at Oslo University. The website presents approximately ninety multilingual Buddhist texts aligned in Sanskrit, Tibetan, Chinese, Mongolian, and English. The texts are chunked in (loosely defined) sentence or paragraph units for parallel reading. The data is limited to online use, as no downloadables of the linked texts are made available.

Yogācārabhūmi Database.

Chinese, Sanskrit, and Tibetan versions of the Yogācārabhūmi are available in an (online-only) interface optimized for parallel reading. Maintained at Dharma Drum.

Art, Architecture, Ritual

Whereas, in spite of often insufficient metadata and standardization, the digitization of Buddhist scriptures has been rather successful; very few initiatives have attempted to tackle the digitization of Buddhist art, architecture, and ritual and oral transmission. In principle, we should be able to enter a virtual Buddhist 8th-century heritage site as easily as pulling up an 8th-century Buddhist text. However, legal and economic factors apart, there is a time lag in the technological tool chain between the management of digital text and the management of 2D and 3D images as well as audio, video, and interactive virtual environments. There have been, for instance, several attempts to create digital representations of Dunhuang caves, but none so far has proved sustainable.

Huntington Photographic Archive of Buddhist and Asian Art.

This largest independent archive of Buddhist art represents the field documentation efforts by Susan and John Huntington from 1969 to the present, and contains scans of more than one hundred thousand photographs. The data is so far not published under an open license, but made available to researchers online without charge. The site is currently hosted at the University of Chicago.

Photo Dharma.

As of 2019, this independently maintained site contains more than 14500 photographs of Buddhist sites and art objects

published under a Creative Commons Attribution-Share Alike License. Especially notable, the panoramic, 360-degree photographs of Buddhist sites from South and Southeast Asia.

SAT Taishōzō Image DB.

The Saṃgaṇkīkṛtaṃ Taiśotripiṭakaṃ (SAT) team at the University of Tokyo has created a search interface for the heavily illustrated Volumes 86–96 of the Taishō edition of the Buddhist canon. The images, mostly from the realm of East Asian esoteric Buddhism, can be searched by keywords, magnified, and tagged, and are published according to the International Image Interoperability Framework (IIIF) standard.

Tibetan and Himalayan Library.

According to its website, the Tibetan and Himalayan Library (THL) is designed as "a publisher of websites, information services, and networking facilities relating to the Tibetan plateau and southern Himalayan regions." Its interface provides access to a large collection of approximately thirty thousand photographs, audio and visual material, a map collection, and Tibetan language tools. THL is maintained at the University of Virginia.

The UMA Library of Recorded Voice.

Thousands of hours of digitized tape recordings of the first generation of Tibetan lamas to arrive in America, made by Jeffrey Hopkins. The library features the Fourteenth Dalai Lama, as well as over a dozen other prominent Lamas teaching various canonical texts. The library is currently hosted by the UMA Institute for Tibetan Studies at UMA Tibet. All materials are made available under a Creative Commons license.

Digitization of Research Tools

Next to the digitization of primary sources, Buddhist studies has profited greatly from the digitization of print and the development of born-digital research tools. The most important categories here are *dictionaries* capturing the vast multilingual terminology of Buddhism, *catalogues* of primary sources, and *bibliographies* of secondary scholarship.

Dictionaries, Authority Databases, and Network Data

Lexicography is the backbone of Buddhist studies in the sense that no student or researcher working in the field can do without them. The late 20th century saw the digitization of print dictionaries as well as the appearance of born digital dictionary and terminological databases. Next to online dictionaries like the Digital Dictionary of Buddhism (DDB), there are offline dictionary platforms (GoldenDict, Babylon, Pleco, etc.) that allow for local storage and custom searches over a large range of glossaries. Whereas online lookup tools serve a wider audience, freestanding glossaries are more relevant for digital humanities (DH) researchers, as they can be used for natural language processing tasks.

All_Index.

The all_index is an important tool for computational linguistics on East Asian Buddhist texts. The single file "all_index.xml" indexes a large number of East Asian Buddhist dictionaries resulting in approximately 290,000 terms. Started by Urs App and

Christian Wittern, it is now maintained by Charles Muller, who makes the 2010 version available online.

Dharma Drum Buddhist Person Name Authority.

Contains structured, densely referenced biographic information on more than forty-three thousand persons from Buddhist historical sources. Online lookup tool and downloadable data set. Emphasis is on Chinese Buddhist monastics and laypeople, with some coverage of Japanese, Korean, and Indian Buddhists. Includes information about lineage, and links to authored works and biographies.

Dharma Drum Buddhist Place Name Authority.

Contains structured, diachronic information regarding places appearing in Chinese Buddhist sources. More than fifty-eight thousand geo-referenced entries. Due to licensing issues, the downloadable data set contains only approximately twenty thousand entries, but these include most Buddhist sites (temples etc.).

Dharma Drum Glossaries for Buddhist Studies.

Fifteen different glossaries for Buddhist studies from the ancient Mahāvyutpatti (Sanskrit-Tibetan-Chinese) to the modern Karashima glossaries on individual texts produced at Dharma Drum, Taiwan. Together with the general digital dictionaries for Chinese, Sanskrit, Pali, Tibetan, etc., that are available elsewhere on the web, the glossaries are the basis for a robust off-line dictionary system for Buddhist studies.

Dictionary of Gāndhārī.

Based on the corpus of Buddhist manuscripts written in the Gāndhārī Prakrit, this online only dictionary contains approximately seven thousand entries (as of 2017).

Digital Dictionary of Buddhism.

Next to a large number of entries by individual researchers, the DDB has incorporated the Soothill-Hodous *Dictionary of Chinese Buddhist Terms* and Lewis Lancaster's *The Korean Buddhist Canon: A Descriptive Catalogue*. As of 2020, it contains approximately seventy-four thousand entries. The DDB is a subscription service, but visitors can query up to ten terms per day by using a generic log-in ("guest") without password.

Social Network Data for Chinese Buddhist History.

A large SNA data set comprising referenced links between more than seventeen thousand Chinese Buddhists ranging from the Han dynasty to the Republic. Synthesized from the Dharma Drum Buddhist Person Name Authority, marked-up biographies of monks and nuns, and other sources. Independently maintained at GitHub.

Catalogues

The primary literature of Buddhism spans more than two thousand years and more than twenty languages. Buddhist scholars through the ages had to produce catalogues and finding aids to manage their collections. This is especially true for China and Tibet where canonical editions were in principle open and could accrue additional texts over time. To find and disambiguate titles, and assess their textual history in the thicket of canonical editions can be a formidable task even with the help of modern databases.

Buddhist Canon Research Database.

Designed as online look-up tool, the Buddhist canon Research Database is hosted online at Columbia University. It specializes in Tibetan and Sanskrit texts, but includes mapping to the Taishō edition. At present, it contains approximately ten thousand bibliographic records for primary texts, with another twelve thousand bibliographic records for the associated secondary literature, while the full-text interface offers search features for a large corpus of Tibetan canonical texts.

Chinese Buddhist Canonical Attributions Database.

Maintained at Heidelberg University, this cutting-edge database collects information regarding author- and translatorship attributions for the Chinese canon. As many texts are dated on the basis of these (often erroneous) attributions, this clearinghouse enables digital researchers to assemble better quality corpora for digital humanities (DH) tasks.

Digital Database of Buddhist Tripitaka Catalogs.

Initiated by Aming Tu (b. 1953–d. 2016) and maintained at CBETA, this is still the best way to quickly check online which of the many Chinese canonical editions contain a given sutra.

SuttaCentral.

SuttaCentral has aggregated a comprehensive database of early Buddhist texts that links text clusters of Pali, Chinese, Sanskrit, and Tibetan text. Next to the identification, SuttaCentral offers links to full texts and modern translations in more than twenty languages. All data is freely available on Github, packaged in an exemplary fashion. SuttaCentral encourages reuse and copying of its data.

Bibliographies

Bibliographies are indispensable. Since 1844, when Eugene Burnouf wrote the first modern history of Buddhism, thousands of monographs, editions, articles, translations, and edited volumes regarding Buddhism have appeared in Japanese, French, English, Mandarin, German, and a host of other languages. Buddhist studies as a field lacks a comprehensive multilingual bibliography since the *Bibliographie Bouddhique* founded by J. Przyluski in 1928 was discontinued after 1967. The citations in this section are digital attempts to mitigate the problem.

Bibliographical Sources for Buddhist Studies from the Viewpoint of Buddhist Philology.

By far the most professionally annotated current bibliography for Buddhist studies. Maintained online by the Buddhist Bibliography Project at the International College Postgraduate Buddhist Studies, Tokyo. Distributed as "e-document," i.e. a

searchable pdf, which is occasionally updated. The current version 2.6 (March 2020) has 1288 pages containing an estimated seven thousand entries.

Bibliography of Buddhist Studies Bibliographies.

This simple online list contains bibliographic references for 148 Buddhist studies bibliographies (mostly printed, some born digital).

Bibliography of Translations from the Chinese Buddhist Canon.

Maintained since 2001, this slow growing bibliography specializes in translations from the Buddhist Chinese canonical editions into "Western" languages. As of 2020, listed are approximately twelve hundred translations of approximately 550 texts. It can be downloaded as a simple HTML list or accessed as a SPARQL endpoint.

Dan Martin's TibSkrit.

The 2014 version of this idiosyncratic, but very useful collection of references for person and text names with an emphasis on Indian Tibetan Buddhism is still available online. A 2008 version is available as part of the Dharma Drum Glossaries for Buddhist Studies (cited under Dictionaries, Authority Databases, and Network Data).

Digital Library and Museum of Buddhist Studies.

Hosted at National Taiwan University, this is currently the largest online bibliography for Buddhist studies. Incorporates the Indian and Buddhist Studies Treatise Database data set until 1999. As of March 2020, the database contains approximately 411,000 entries and seventy-three thousand full-text articles. Coverage of Chinese secondary literature is especially comprehensive. The interface is adequate, but a download of the data is not possible. Query results can, however, be received by email.

"Genchi" Bibliography of Jodo-Shinshu Scriptures in Western Languages.

Independently maintained and covering primary as well as secondary literature, this large and well organized bibliography focuses on the *Jōdō Shinshu* School of Japanese Buddhism.

H-Buddhism Bibliography Project.

Since 2012 this is part of the H-Buddhism mailing list. The project is a Zotero-based attempt that aims to pool the many private bibliographies by users of the H-Buddhism mailing list. As of 2020, there are approximately ten thousand items on file. This is less than either the National Taiwan University's Digital Library and Museum of Buddhist Studies or the Indian and Buddhist Studies Treatise Database, but the records are more easily available for computation (e.g., for research in trend analysis).

Indian and Buddhist Studies Treatise Database.

Going online in 1998, Indian and Buddhist Studies Treatise Database (INBUDS) is maintained by the Japanese Association of

Indian and Buddhist Studies 日本印度学仏教学会. This is the largest bibliography of Japanese secondary literature on Buddhist studies. Western and Chinese literature are less well represented. Very laudably, INBUDS has made a snapshot of its data set (as of 2015) available for download, which contains sixty-nine thousand entries.

Buddhism in Cyberspace

Cyberspace here means all forms of online communication: portals, bulletin boards, blogs, apps, online groups, video games, virtual worlds, etc. As fora for contemporary Buddhism, these have attracted the attention of researchers. Most of the studies and overviews about these new venues, however, still were published in print.

General Surveys and Edited Volumes

In recent years a number of edited volumes have probed how digital data and methods are affecting Buddhist studies (Grieve and Veidlinger 2015, cited under Buddhism Online; Veidlinger 2019). These contain article-length overviews of individual projects (Muller 2019), or certain aspects of the field (Hackett 2019). Veidlinger 2018 is a reflection on digital media in the context of the history of Buddhist transmission.

Bingenheimer, Marcus, Christian Wittern, and Jinhua Chen, eds. *Special Issue: Buddhism and Technology*. *Journal of the Japanese Association of Digital Humanities* **5 (2020).**

A special issue on digital humanities and Buddhism, containing a number of articles ranging from technical topics such as word-segmentation strategies for Buddhist Chinese to ethnographic research of Buddhism online.

Hackett, Paul. "Digital Encoding, Preservation, Translation, and Research for Tibetan Buddhist Texts." In *Digital Humanities and Buddhism: An Introduction*. **Edited by Daniel Veidlinger, 91–110. Berlin: DeGruyter, 2019.**

Currently the most up-to-date, concise overview of the development of digital Tibetan text. Hackett is also the creator of the Buddhist Canon Research Database (cited under Catalogues).

Muller, Charles. "The Digital Dictionary of Buddhism and CJKV-English Dictionary: A Brief History." In *Digital Humanities and Buddhism: An Introduction*. **Edited by Daniel Veidlinger, 143–158. Berlin: DeGruyter, 2019.**

A firsthand account of the development of the oldest online dictionary of Buddhism, the Digital Dictionary of Buddhism (see under Dictionaries, Authority Databases, and Network Data).

Veidlinger, Daniel. *From Indra's Net to Internet: Communication, Technology, and the Evolution of Buddhist Ideas*. **Honolulu: University of Hawai'i Press, 2018.**

Currently the most sustained media theoretical discussion of the changes experienced by Buddhism in relationship to the media of transmission. Discusses the digital transformation in the context of previous historical media changes such as from oral to written, and from manuscript to print.

Veidlinger, Daniel, ed. *Digital Humanities and Buddhism: An Introduction*. **Berlin: DeGruyter, 2019.**

Two introductions and ten chapters on various topics focusing on digital humanities approaches and tools for Buddhist studies. Veidlinger's introduction to the volume is a useful and concise overview of the topic.

Buddhism Online

The study of contemporary Buddhism online involves observations regarding Buddhist discussion fora (Sharapan 2018), apps (Grieve 2017b), or blogs (Foulks McGuire 2015). Ethnographic fieldwork in virtual worlds (especially Second Life) is becoming a distinct form of research in its own right (Connelly 2010, Grieve 2017a, Falcone 2019). Grieve and Veidlinger 2015 contains twelve essays that discuss various facets of Buddhism online. The research in English so far is still rather US centric. Although globally less than 10 percent of Internet users are in the United States as of 2020, there are only few studies (such as Baffelli, et al. 2011) of Buddhism in different online cultures.

Baffelli, Erica, Ian Reader, and Birgit Staemmler. *Japanese Religions on the Internet: Innovation, Representation and Authority*. **New York: Routledge, 2011.**

Most chapters in this informative collection touch on how different aspects of Japanese Buddhism are represented online. Topics include the Shikoku pilgrimage, online challenges to traditional centers of authority in Jōdo Shinshu and Sōka Gakkai, and celebrity representation.

Connelly, Louise. "Virtual Buddhism: An Analysis of Aesthetics in Relation to Religious Practice within Second Life." *Heidelberg Journal of Religions on the Internet* **4.1 (2010): 12–34.**

An analysis of the aesthetics of two virtual Buddhist places in Second Life. Points out that in spite of the emphasis on Buddhist visual culture, other sensory experiences (imitation-touch) too are significant.

Falcone, Jessica. 2019. "Sacred Realms in Virtual Worlds: The Making of Buddhist Spaces in Second Life." *Critical Research on Religion* **7.2 (2019): 147–167.**

Introduces a helpful distinction between virtual spaces in Second Life that are community-practice-oriented and those that are individual-practice-oriented. Argues that "authentic" Buddhist practice is possible in virtual worlds.

Foulks McGuire, Beverley. "The Way of the Blogisattva: Buddhist Blogs on the Web." In *Buddhism, the Internet, and Digital Media: The Pixel in the Lotus*. **Edited by Gregory P. Grieve and Daniel Veidlinger, 204–220. New York: Routledge, 2015.**

A study of the English-language Buddhist blogosphere around 2015. Notes the existence of nearly three hundred Buddhist blogs, and the yearly "Blogisattva Awards" for the best Buddhist blog.

Grieve, Gregory P. *Cyber Zen: Imagining Authentic Buddhist Identity, Community, and Practices in the Virtual World of Second Life*. **New York: Routledge, 2017a.**

The most sustained discussion of ethnography in cyberspace so far. Focuses on Grieve's fieldwork among the Buddhist communities in Second Life.

Grieve, Gregory P. "Meditation on the Go: Buddhist Smartphone Apps as Video Game Play." In *Religion and Popular Culture in America*. **Edited by Bruce D. Forbes and Jeffrey H. Mahan, 195–213. Oakland: University of California Press, 2017b.**

Discusses, in a largely positive assessment, how the smartphone mindfulness app "buddhify" is used to practice meditation, and how its design draws on lessons learned from video gaming, in effect melding spiritual pursuit with play in line with other "not religious but spiritual" identities.

Grieve, Gregory P., and Daniel Veidlinger. *Buddhism, the Internet, and Digital Media: The Pixel in the Lotus*. **New York: Routledge, 2015.**

Eleven chapters and an introduction on a wide variety of topics. Next to contributions on historical and methodological aspects, the volume presents case studies on Buddhist blogs and apps, virtual places, and communities. Many chapters exemplify how the study of online Buddhism can now be considered a subfield in the study of contemporary Buddhism.

Prebish, Charles. "The Cybersangha: Buddhism on the Internet." In *Religion Online: Finding Faith on the Internet*. **Edited by Lorne L. Dawson and Douglas E. Cowan, 135–147. New York: Routledge, 2004.**

An early overview of Buddhism online, roughly from the coinage of the word "cybersangha" (1991) to the rise of blogs (after 2004). Covering early Buddhist Internet resources such as discussion forums, online journals, and how Buddhist online communities could be understood through the idea of a "*sangha* of the four quarters" (p. 135).

Sharapan, Maria. "Discovering Buddhism Online: A Translocative Analysis of Tibetan Buddhist Forum Discussions." *Heidelberg Journal of Religions on the Internet* **13 (2018).**

A detailed, theoretically founded discourse analysis of a Tibetan Buddhist online forum. Discusses how Tibetan Buddhism is adopted by non-Buddhists as reflected in online forum discussions.

Application of Computational Methods in Buddhist Studies

First attempts have been made within Buddhist studies to use digital methods borrowed from computational linguistics, bioinformatics, geography, and network science.

Computational Linguistic and Stylometric Analysis of Buddhist Texts

Phylogenetic analysis has been used to built stemmatic relationships between different witnesses of a text (Maas 2008, Apple 2014). Computer-assisted frequency analysis (Radich 2014, Radich 2017a, Radich 2017b) and Principal Component Analysis (Hung, et al. 2010; Bingenheimer, et al. 2017) was applied to attribute or de-attribute translatorship of text. Researchers have also started to look into automated term extraction (Wang, et al. 2014; Bingenheimer 2015) and n-gram extraction (Handy 2019).

Apple, James B. "Fragments and Phylogeny of the Tibetan Versions of the Mañjuśrīvihārasūtra: A Case Study in the Genealogy of Tibetan Kanjurs." *Annual Report of the International Research Institute for Advanced Buddhology* **17 (2014): 293–335.**

Based on cladistic, or phylogenetic, analysis of textual variations between five Dunhuang manuscripts and seventeen Kanjur editions, Apple is able to show how these different versions of the *Mañjuśrīvihārasūtra* are connected in a single stemma. Rooted in Dunhuang witnesses and ranging to 18th-century print editions, the phylogenetic analysis (like Maas 2008, using PAUP) was confirmed via a "Bootstrap analysis" with satisfactory results.

Bingenheimer, Marcus. "The Digital Archive of Buddhist Temple Gazetteers and Named Entity Recognition (NER) in Classical Chinese." *Lingua Sinica* **1.8 (2015): 1–19.**

The Digital Archive of Buddhist Temple Gazetteers was manually marked-up for Person and Place names. The article describes previous attempts at named entity recognition (NER) for classical Chinese texts. It gives some examples for synchronic and diachronic research with NER data, demonstrates the impossibility of brute force, dictionary-based NER, and suggests the corpus as training data set for more advanced solutions.

Bingenheimer, Marcus, Jen-Jou Hung, and Cheng-en Hsieh. "Stylometric Analysis of Chinese Buddhist Texts—Do Different Chinese Translations of the Gaṇḍavyūha Reflect Stylistic Features That Are Typical for Their Age?" *Journal of the Japanese Association for Digital Humanities* **2 (2017): 1–30.**

A stylometric analysis focusing on grammatical particles in Buddhist Chinese. Using Principal Component Analysis the authors are able to show that translations from the same period can be correlated by particle use. Compares T.278, T.279, T.293, and, as control, T.294.

Handy, Christopher. "A Context-Free Method for the Computational Analysis of Buddhist Texts." In *Digital Humanities and Buddhism: An Introduction*. **Edited by Daniel Veidlinger, 183–208. Berlin: De Gruyter, 2019.**

Describes an algorithm for parsing n-grams of variable length out of Sanskrit, Tibetan, and Chinese text. This is a basic, but nontrivial task and a prerequisite for other types of linguistic analysis. The program written in C is made available by the author online.

Hung, Jen-jou, Marcus Bingenheimer, and Simon Wiles. "Quantitative Evidence for a Hypothesis regarding the Attribution of Early Buddhist Translations." *Literary and Linguistic Computing* **25.1 (2010): 119–134.**

Tests a hypothesis by Mizuno Kōgen, who suggested that a group of single *Madhyama-āgama* sutras, which were traditionally attributed to different translators, should in fact be considered as translated by Zhu Fonian and Dharmanandin. Using Principal Component Analysis, the authors were able to confirm that the sutras in question should indeed be considered to have been translated by the same translators and that the traditional attributions are therefore incorrect.

Maas, Philipp A. "A Phylogenetic Approach to the Transmission of the Tibetan Kanjur—the Akṣayamatinirdeśa Revisited." In *Bauddhasāhityastabakāvalī: Essays and Studies on Buddhist Sanskrit Literature Dedicated to Claus Vogel*. **Edited by Dragomir Dimitrov, Michael Hahn, and Roland Steiner, 229–243. Indica et Tibetica 36. Marburg, Germany: Indica et Tibetica Verlag, 2008.**

The, to my knowledge, first application of cladistics, or phylogenetic systematics, to the study of Buddhist texts. Cladistic algorithms in textual studies are used, in conjunction with text critical scholarship, to create a stemma of a text's historical development. This is a proof of concept article that tests the PAUP software to a section of the Akṣayamatinirdeśa, for which the stemma had already been researched.

Radich, Michael. "On the Sources, Style and Authorship of Chapters of the Synoptic *Suvarṇaprabhāsottama-sūtra* T664 Ascribed to Paramārtha (Part 1)." *Annual Report of the International Research Institute for Advanced Buddhology* 17 (2014): 207–244.

Radich here pioneers the use of a suite of Python programs called TACL (Textual Analysis for Corpus Linguistics; see also TACL's Documentation) to collect data on stylistic markers of translators and their teams. Based on statistical frequency analysis of n-grams combined with traditional scholarship, this method sheds light on the complex translation history of the *Suvarṇaprabhāsottama-sūtra* (T.663, T.664, and T.665).

Radich, Michael. "On the *Ekottarikāgama* 增壹阿含經 T 125 as a Work of Zhu Fonian 竺佛念." *Journal of Chinese Buddhist Studies* 30 (2017a): 1–31.

Using data from TACL combined with a thorough study of traditional information on T.125 from catalogues, etc., to build a benchmark corpus for Zhu Fonian. This is then used to argue that the translation of T.125 too should be attributed to Zhu Fonian, based on its correlation with the benchmark corpus.

Radich, Michael. "Problems of Attribution, Style, and Dating Relating to the 'Great Cloud Sūtras' in the Chinese Buddhist Canon (T 387, T 388/S.6916)." In *Buddhist Transformations and Interactions: Essays in Honor of Antonino Forte*. Edited by Victor Mair, 235–289. Amherst, NY: Cambria, 2017b.

Again using TACL in combination with traditional scholarship, this stylometric study affirms the attribution of T.387 to Dharmakṣema, while ruling out the hypothesis that T.388 was translated by Zhu Fonian. Stylometric analysis shows that T.388 instead has a close affinity with the idiom employed by Dharmakṣema and his team.

Wang, Yu-Chun, Karol Chia-Tien Chang, Richard Tzong-Han Tsai, and Jieh Hsiang. "Transliteration Extraction from Classical Chinese Buddhist Literature Using Conditional Random Fields with Language Models." *Computational Linguistics and Chinese Language Processing* 19.3 (September 2014): 25–38.

A paper on how to extract transliterated Indian loanwords from Chinese Buddhist texts. Using the *Saṃyukta-āgama* and the Lotus Sūtra as examples, the authors employ Conditional Random Fields, a statistical modeling method often used in natural language processing, to identify transliterated Indian terms.

Other Approaches (GIS, SNA, etc.)

Beyond textual analysis, researchers are starting to use information mined from digital text to build arguments from geographic (Protass 2019, Jensen 2019) and social network data (Bingenheimer, et al. 2011; Bingenheimer 2018).

Bingenheimer, Marcus. "Who Was 'Central' for Chinese Buddhist History? A Social Network Approach." *International Journal of Buddhist Thought and Culture* **28.2 (December 2018): 45–67.**

Based on a large social network data set for Chinese Buddhist history (see Dictionaries, Authority Databases, and Network Data), the author uses two formal metrics from network analysis—degree and betweenness centrality—to show how such measures can identify significant features within Chinese Buddhist history. While degree centrality identifies translators and patrons, betweenness centrality highlights Chan and Vinaya masters.

Bingenheimer, Marcus, Jen-Jou Hung, and Simon Wiles. "Social Network Visualization from TEI Data." *Literary and Linguistic Computing* **26.3 (2011): 271–278.**

Explains how the Dharma Drum corpus of biographies of eminent Buddhists was encoded, and how network data can be extracted from it.

Jensen, Christopher. "Mapping Religious Practice in the Eminent Monks: Theoretical and Methodological Reflections." In *Digital Humanities and Buddhism: An Introduction*. **Edited by Daniel Veidlinger, 160–182. Berlin: De Gruyter, 2019.**

Uses geo-referenced data derived from Dharma Drum Authority Databases and the marked-up corpus of Biographies of Eminent Monks in an attempt to illustrate the different distribution patterns of places mentioned in the *Gaoseng zhuan* T. 2059 and the *Xu gaseng zhuan* T.2060.

Protass, Jason. "A Geographic History of Song-Dynasty Chan Buddhism: The Decline of the Yunmen Lineage." *Asia Major* **3d ser. 32.1 (2019): 113–160.**

Uses lineage data and geographic information science (GIS) to show why and how the Yunmen Chan lineage suddenly disappeared in the 12th century. Instead of doctrinal developments, Protass traces the demise of the Yunmen school to its geographic exposure during the anti-Buddhist campaign of Huizong and the destruction caused by the Song-Jin wars.

back to top