

## Chapter 7

# The SAT Taishō Text Database

## A Brief History

A. Charles Muller, Masahiro Shimoda and  
Kiyonori Nagasaki

### INTRODUCTION

The *Taishō Shinshū Daizōkyō* 大正新修大藏經 (*Taishō canon*) is the Japanese edition of the East Asian Buddhist canon, compiled during the Japanese Taishō period (1912–1926). The Taishō compilation of Buddhist texts (mainly in classical Chinese, as well Japanese and some Indic scripts) began in 1924 (Taishō year 13) under the editorial direction of Takakusu Junjirō 高楠順次郎 (1866–1945) and Watanabe Kaigyoku 渡辺海旭 (1872–1933), and was completed in 1934, including a total of 100 volumes. This compilation is reported to have been largely based on a Korean Goryeo dynasty (918–1392) woodblock printed edition (*Goryeo Daejang Gyeong* 高麗大藏經) held in the collection of Zōjōji Temple 増上寺, further informed by Song (960–1279), Yuan (1271–1368), and Ming dynasty (1368–1644) texts, and also sources from the Zōjōji collection, materials in the Shōsōin 正倉院, and the imperial household agency collection, texts from Dunhuang, as well as materials preserved in temples, universities, and other private collections.

The twentieth century was a watershed period for Buddhist studies in Japan, in the sense that the field underwent a major transformation from being strictly the domain of religious organizations and institutions to its being secularized and restructured along Western scientific principles of history and philology and being taught at secular state institutions. The *Taishō Canon* was, in this environment, compiled and organized along philological and historical principles, and its compilation was a crowning achievement for Japanese Buddhology, as well as Japan as a nation. The comprehensiveness, clarity, and rigor with which the *Taishō Canon* was compiled, and then printed in modern typeset, made it the standard source for the study of the East Asian canon from the 1930s down to the present day.

One of the results of the secularization/scientization of Buddhist studies in Japan was that Buddhism would be taken up as a major object of research in Japan's secular and national universities, among these, its leading national university—the University of Tokyo (formerly, Tokyo Imperial University). The University of Tokyo became the de facto academic headquarters for Buddhist studies in Japan, and the university's graduate program in the Department of Sanskrit Literature (later, the Department of Indian Philosophy) became the premier program in Japan, staffed by scholars of the highest reputation. In the postwar period, this department became the location for the headquarters of the Japanese Association for Indian and Buddhist Studies (JAIBS), as well as the journal published by the same organization.

Coming up to the mid-1980s, the age of the personal computer emerged, and the need to think about the digitization of the core textual resources of Japanese Buddhism became an unavoidable reality. At least a few Japanese scholars working in the field of Buddhist studies (especially at the University of Tokyo) could not but gradually come to be aware of the new possibilities—and eventually, the new responsibilities—that were emerging. One of the first University of Tokyo faculty members to take the matter of digitization seriously was Prof. Hirakawa Akira 平川彰 (1915–2002), who took the initiative for creating the database of Japanese articles on Buddhist studies (the INBUDS database) in 1984. This database at first focused on articles published by the JAIBS, but later expanded its coverage to articles from a broad range of academic publications.<sup>1</sup> Hirakawa was joined in his efforts toward digitization in 1993 by Prof. Ejima Yasunori 江島恵教 (1939–1999), who became concerned about the need to digitize the *Taishō Canon*, and began to seek both funding and methods to begin this task. For funding, Prof. Ejima naturally applied for the standard Japanese JSPS Grants-in-aid, but in order to secure a sustainable source of monetary support, he also worked to form an organization of Japanese Buddhist temples that would contribute to support this monumental task.

The publisher of the *Taishō Canon*, Daizō Shuppan 大藏出版, had attempted to begin the digitization process, producing three *Taishō* volumes on three CDs. However, the process was slow, and the price of \$250.00 per CD was far too high to be realistically affordable for the average researcher. At length, Daizō Shuppan turned the project over to the leadership of Prof. Ejima, who had already taken over the responsibility for the digitization of the INBUDS database from Prof. Hirakawa. Working toward the digitization of both collections up through the early 1990s, Ejima came to the realization that the management of the digitization of *both* collections at the same time was beyond his capacity. So, in 1994, upon the entry into the University of Tokyo Department of Indian and Buddhist Philosophy faculty of junior professor Shimoda Masahiro (whom Ejima had taught as a graduate student),

Ejima “bequeathed” the project of the digitization of the *Taishō Canon* (named as SAT<sup>2</sup>) to Shimoda.

Shimoda was not, by any means, a technical maven who knew the best approaches to digitization of texts and, as an incoming junior professor, had but minimal power to make others work or to gain financial support for the project, so it was clearly a formidable task that had been laid at his feet.<sup>3</sup> But realizing the importance of the task, he began the long, difficult, and sometimes halting march toward the digitization of the *Taishō* in Japan.

The first thing needed was basic funding to support the work. The initial source of funding was a JSPS “Grant-in-aid for databases,” through which Shimoda was able to secure a basic level of ongoing support for the work of digitization. Additionally, vital ongoing support was obtained from a SAT support association comprising Buddhist temples organized by Takasaki Jikidō 高崎直道 (1926–2013) and Nara Yasuaki 奈良康明, called the Society for the Promotion of Buddhist Studies (Bukkyō Gakujutsu Shinkōkai 仏教学術振興会),<sup>4</sup> which paid for approximately half of the expenses. Dedicated collaborators from around Japan (approximately 250) offered their services as typists for the input of the text.

Technical challenges abounded, the foremost of these being the basic method of converting paper text into digital form. In those days, although OCR input was a viable approach for digitizing roman character-based texts with their fifty or so ASCII<sup>5</sup> characters, it was impossible to accomplish the digitization of a Chinese and Japanese character database containing some 12,000 types of old characters<sup>6</sup> in timeworn fonts to a viable level with OCR. The CJK Han Character section of the Unicode consortium was still only in the early stages of planning, and the Japanese computer character set at the time (JIS-x 0208) only included some 6,800 characters—only a little more than half of what would have been needed—even assuming that the hardware and software could read the fonts accurately. Thus, the SAT team had no alternative but to use the input method of double-typing,<sup>7</sup> also practiced by other major text database projects such as the Academia Sinica Text database<sup>8</sup> and the *Tripitaka Koreana*.<sup>9</sup> Yet there were 100 million characters to be typed in, many of which were not contained in the Japanese computer character set—even with the addition of 5,800 new ideographs in the newer JIS-x 0212 character set. The Unicode (ISO-IEC 10646) character set was becoming available in the mid-1990s,<sup>10</sup> but Japanese software engineers were slow in embracing and implementing it. Therefore, thorny technical decisions had to be made as to how to encode, and later display the large numbers of missing characters (*gaiji* 外字). Beyond the basic problem of ideographic coverage, the *Taishō Canon* contained charts, diagrams, and pictures, as well as Siddham and other non-Sinitic characters. Many problems needed to be

solved by Shimoda and his team of largely nontechnical Buddhist specialists, and progress was, in the early stages, slow and painful.

In some cases, the solution of technical problems could be resolved through the emerging standards developed in the realm of the computerization of Buddhist studies. There were a number of digitization projects in progress around the world, many of which were tackling similar problems.<sup>11</sup> There were also developments in general world standards for computers, including the gradual implementation of Unicode, and basic computer functions. But progress was nonetheless slow.

A shift in the management strategy of the project came during the year 2000, when, based on input received from experts from the IT sector, the project was able to improve significantly its strategy for digitization. During this period, the development of the SAT database was also much aided by the advice and efforts of Ishii Kōsei 石井公成 of Komazawa University and Moro Shigeki 師茂樹 (at the time a graduate student at Tōyō University, now a professor at Hanazono University). Shimoda also received much unexpected and invaluable assistance from many of his graduate students. During the latter stages, the input system was improved significantly based on the web collaboration system designed and implemented by Nagasaki Kiyonori (introduced in further detail below), who joined the project in 2005. In 2007, after more than a full decade of work, the basic textual digitization process was beginning to reach its completion.

As the task of digitization reached its completion, SAT members came to be convinced that the best course to follow in terms of publication would be a web-centric approach, rather than a CD-based or local-application-based approach. In this way, it distinguished itself from the other major East Asian canonical projects of CBETA<sup>12</sup> and *Tripitaka Koreana*. In fact, this approach had begun even in the late 1990s, when SAT was the first of the Buddhist canon digitization projects to make its data available for downloadable in plain text format with instructions. Thus, SAT was beginning to show its own distinctive impetus and orientations, in much part already inherent in its basic environment. With the project based within the JAIBS, located in the University of Tokyo, which together comprised one of the richest research repositories for detailed textual/historical studies of Buddhism, it was natural for SAT to take advantage of the situation. The SAT team was also coming to awareness that the future of humanities resource development lay not in CDs or hard drives, but in online web services. In this regard, the SAT project was fortunate to gain as a key member Nagasaki Kiyonori, at the time a faculty member at Yamaguchi Prefectural University.<sup>13</sup> Although originally a Buddhist studies specialist, Nagasaki had honed high-level web server and database programming skills and shared a vision for the development of Buddhist studies resources on a web platform. Not just “SAT online,” but “SAT

online sitting at the center of an array of other web-based Buddhist Studies research resources.”

Nagasaki took over the technical management of the database in 2007, started off by taking the currently available SAT texts and setting them up in a fast, searchable database format. In 2008, the database went online. He then began to add interactive functions, starting with the INBUDS article database, creating an environment where users could select terms in a Taishō text and directly see if they were included in the titles or keywords of the articles contained in the INBUDS article database. Since this INBUDS data was already the property of the JAIBS, the matter of getting access to it was relatively straightforward. At the same time, Nagasaki took technical responsibility for the administration of the INBUDS database, and he worked to expand the search functions and coverage of that bibliographical resource significantly.

## WEB COLLABORATION

A major step in interoperative function development was achieved in 2008 when, using Web API techniques, Nagasaki set up an interoperative function between the SAT database and the *Digital Dictionary of Buddhism (DDB)*,<sup>14</sup> wherein, based on a type of XML resource description file,<sup>15</sup> users of the SAT database who selected a portion of text would be presented with a list of terms within that text contained in the DDB, along with basic meanings, pronunciations, and links to the full entries in the DDB. At this point in time on the web, this kind of interoperative arrangement between two separate resources was relatively new, but quite significant, as, instead of loading the entire DDB dataset into the SAT database, both SAT and the DDB could maintain their independence as separate entities, yet at the same time add significant value to each other.<sup>16</sup> Using these DDB XML data extracts, Nagasaki was also able to add other distinctive search functions to SAT, including searching capabilities based on English or Hangeul input.

From here began the development of a SAT-based “ecosystem” that included both API-based interoperation and other sorts of datasets that were being developed at various ranges of proximity to SAT. The point was not to only create an online database of the Buddhist canon, but to take the lead in starting an entirely new approach to humanistic studies. Thus, the notion of a “knowledge base” of the humanities, which would be sufficiently persuasive to receive acknowledgment by those who were unfamiliar to these new approaches—approaches that would eventually become known by the label of “Digital Humanities.” At the most basic level, this included web-based inputting and proofreading of the base texts. Complementing the translation-tool array that started with the DDB linking, first came the inclusion of parallel

text data from the Bukkyō Dendō Kyōkai 仏教伝道協会 (BDK) English Tripiṭaka translation project,<sup>17</sup> linked in such a way that lines of Taishō text that matched translated lines in the BDK texts would now appear in a separate pop-up window. Further bibliographical information was provided by linking to the SARDS database.<sup>18</sup> In 2009, the function of being able to search for PDFs contained in CiNii<sup>19</sup> by title and keyword through INBUDS was established. These and several other new applications and updates were included in the SAT 2012 edition. Thus, in 2012, in addition to the revamping of the search engine and extensive basic correction of the textual content,<sup>20</sup> readers had access to the *Digital Dictionary of Buddhism* along with the article databases of INBUDS and SARDS. On a micro level, users were provided with access to the ideographic information resources of CHISE<sup>21</sup> and the Unihan database.<sup>22</sup>

Even with the full digitization of the text of the *Taishō*, plain text presentation by itself was found to be inadequate in many cases. For example, in the case of notes, or other additions to the original source text, such as deliberate font-size changes, and so forth. Thus, the pages of the *Taishō* were also scanned into image format and these images were made directly available, aligned through links to searched pages, and these were set up to allow zooming in and out. Further developments in the implementation of images would follow later.

In terms of the basic dataset, although the main task of digitization had been completed in 2008, a number of problems remained to be resolved. Foremost among these was the *gaiji* (missing character) problem. The *Taishō Canon* contained thousands of characters that were not yet available in the international character standard. Of course, developments in Unicode, especially with the continued expansion of CJK Unified Ideographs coverage in Unicode 2.0 and 3.0, were gradually reducing the number of ideographs that needed to be represented with GIF images. But still, even with Unicode covering some 40,000 ideographs, the SAT database contained more than 6,000 ideographs that still needed to be displayed with GIF images. Without in some way getting these ideographs registered into Unicode, such a situation could have continued indefinitely.

Up to this point in time, the Unicode Consortium had only been accepting new character applications from national standards bodies through its International Rapporteur Groups (IRG). In other words, ideographs found in the writings of local countries could only be added to Unicode via a proposal submitted by that country's officially designated IRG. Fortunately, through the kind help and intervention of the members of the Japanese IRG, most importantly Kobayashi Tatsuo 小林龍生, Suzuki Toshiya 鈴木俊哉, and Kawabata Taichi 川幡太一, a proposal was put forth to allow the submission of candidate ideographs from major international research organizations and

projects, in this case, the SAT project. The Han Ideograph IRG committee accepted this proposal, and this opened the way for the inclusion of the 6,000 ideographs identified by SAT.<sup>23</sup>

During the years after the 2012 update, other functions have been added. Collaboration has been arranged between SAT and the *Tripiṭaka Koreana* (TK) project, wherein textual selections within the SAT database can be viewed in their TK parallel. Efforts are being made toward the handling of Siddham characters and images within the *Taishō*. Links to manuscripts and prints of canonical texts held at other institutions and locations are also expanding rapidly. Starting from the RITK linkage mentioned just above, SAT texts are now linked to images in libraries at Waseda University, the University of Tokyo, the British Library, National Diet Library, and others. Thus, SAT is steadily expanding its functions and links in various directions, in various media and technologies.

The most important recent developments lie in the area of the enhanced ability to manipulate and present images. Taking advantage of the new possibilities offered by the International Image Interoperability Framework (IIIF),<sup>24</sup> SAT is using IIIF to present the images of monks, mandalas, rituals, and so forth contained in a twelve-volume section of the *Taishō* (called the Zuzōbu 図像部), as well as the images of the text of the *Jiaying Canon* (J. Kakōzō 嘉興藏) from the University of Tokyo Library.<sup>25</sup>

## SAT AND DIGITAL HUMANITIES

There is another aspect of the SAT project that extends beyond the creation of a state-of-the-art database replete with tools, applications, and interoperation with other resources. That is its role as the starting point for the creation of the study of Digital Humanities (DH) at the University of Tokyo, and its further influence for all of Japan. This is not to say that all of DH in Japan sprung from SAT—Japanese researchers have been using computers for the study of literature, art, history, and other humanistic fields for more than three decades. But these studies had been led, and carried out for the most part by computer scientists from a relatively technical-oriented perspective. They had also been limited to Japanese-language publications carried out exclusively by Japanese-language scholars.

On the other hand, the SAT project, in order to continue to develop as an academic and scientific enterprise, needed to maintain and strengthen its academic and scientific orientation within the University of Tokyo and in Japan in order to receive due financial support from the JSPS granting agency. In other words, once the database had been completed, ongoing funding could not be obtained through database-creation grants. Rather, the project needed

to be defined in terms of advanced academic research. Shimoda and other SAT project members saw that the emerging movement of DH (already long flourishing in Europe and North America), led and defined by scholars with humanities backgrounds, provided exactly the kind of language and framework which define the future trajectory of the project. SAT members began to attend international DH conferences, soon becoming deeply involved in the worldwide DH community. As a result, members of SAT in Tokyo began to link up with other researchers around Japan at institutions such as Kyoto University, Osaka University, Dōshisha University, Ritsumeikan University, Tokyo Institute of Technology, National Institute of Informatics, and so forth, who were also interested in the DH approach. A community began to develop, and in 2011 the Japanese Association of Digital Humanities (JADH) was formed with Shimoda as chair. In 2013, the JADH was accepted by the Alliance of Digital Humanities Organizations (ADHO)<sup>26</sup> as a constituent member. Thus, in a sense, the existence of the SAT database was a pivotal factor in the development of DH in Japan.

Armed with the language and framework of DH, Shimoda and the SAT team in 2010 obtained a grant-in-aid at level (A) for the creation and expansion of SAT as a “knowledge base” for Buddhist studies, which lasted until 2014. In 2015, Shimoda obtained the larger JSPS grant-in-aid level (S)<sup>27</sup> for the further development of SAT as a research base, including the investigation and application of various DH methodologies, including, for example, the application of TEI/XML principles to SAT and other Buddhist textual materials. The ripples from SAT’s DH involvement have also had effects locally at the University of Tokyo, where a fledgling DH curriculum is being established.

Thus, the SAT project expects to continue to grow in many ways along with developments in the broader realm of DH, ever trying to provide a more efficient, broader-reaching, and increasingly functional “research base” around the teachings of the Buddha spoken more than two millennia ago.

## NOTES

1. <http://21dzk.l.u-tokyo.ac.jp/INBUDS/search.php>. Prof. Hirakawa helped to arrange the rental of a house for this purpose.
2. SAT is an acronym for the Sanskrit neologism “*Samganikīkṛtaṃ Taiśotripitakam*.”
3. Shimoda reports that he made his best attempt at turning down Prof. Ejima’s request, but to no avail.
4. This organization still actively supports digital Buddhist studies projects. See <http://butsugakushin.org>.

5. ASCII stands for American Standard Code for Information Interchange. Computers can only understand numbers, so an ASCII code is the numerical representation of a character such as ‘a’ or ‘@’ or an action of some sort. The ASCII character table is available at <http://www.asciitable.com>.

6. In fact, 12,000 is a relatively small number, when one considers that the Korean Tripitaka, a major source for the *Taishō Canon*, included more than 30,000 different characters. The number of characters available to the typesetters for the *Taishō Canon* was less than 12,000, which means that many characters were arbitrarily normalized in the process of printing.

7. That is, two typists input the same text, and the two versions are later checked for differences by computer. Even with modern advances in OCR, this method is still regarded by many as being more effective for a large project. The longer the work proceeds, the greater is the speed and accuracy of the typing.

8. <http://hanji.sinica.edu.tw>.

9. The *Tripitaka Koreana* project initiated their efforts in 1993 under the leadership of Ven. Chongnim of Haeinsa. Based on the generous support of Samsung, they completed the digitization of the Korean canon in 1999. A timeline of the project can be found at [http://kb.sutra.re.kr/ritk\\_eng/intro/introProject03.do](http://kb.sutra.re.kr/ritk_eng/intro/introProject03.do).

10. The history of Unicode release dates is provided at <http://unicode.org/history/publicationdates.html>. One must keep in mind that the “release” of a version of Unicode does not mean that it becomes immediately usable on computers. Especially in the earlier days, some regions were slow to adopt Unicode for cultural and political, as well as technical reasons.

11. Early pioneers of the digitization of Buddhist studies began to connect with each other, and began to gather for academic meetings. The main venue for these early meetings was the Electronic Buddhist Text Initiative [EBTI] (<http://www.buddhism-dict.net/ehti>.) The rapidity of the digitization of all of the Buddhist canons, along with the development of Buddhist studies research tools, can be in part attributed to the regular meetings of the EBTI.

12. The history of CBETA is provided in the prior volume in this series in the article by Aming Tu, “The Creation of the CBETA Electronic Tripitaka Collection in Taiwan.” See Jiang Wu and Lucille Chia (ed.), *Spreading Buddha’s Word in East Asia* 321–335. In addition to the information provided in Tu’s article, it should be understood that it was SAT that made it possible for CBETA to publish their version of the *Taishō*, as it was the SAT representatives who persuaded the president of the Daizō Shuppan to relax its stringent copyright restrictions toward Taiwanese publishers. Daizō Shuppan, which had suffered severe damage by the pirated editions published in Taiwan during the postwar decades, had originally flatly refused to accept the proposals submitted by CBETA to digitize the Tripitaka. In response to the request put forth to SAT by CBETA, SAT, after careful discussion, came to the conclusion that CBETA’s work, even though in direct competition with that of SAT, should be supported by SAT as much as possible, since the database of the Buddhist canon, regardless of the provider, should be equally shared by contemporary and future generations as common property to all the people. With this decision in mind, SAT members, escorting delegates from CBETA, visited the president of Daizō Shuppan to ask for

the permission to be given to CBETA. The president, based on this recommendation from SAT, eventually agreed. This was in 1998, one year after SAT's first volume of the Prajñāpāramitā sutras had been published on the web.

13. Nagasaki was introduced to the project by Suzuki Takayasu 鈴木隆泰, former student of Shimoda and the former manager of INBUDS.

14. A dictionary/encyclopedia of Buddhist terms edited by A. Charles Muller, based on Sinitic headwords, which went online in 1995, presently including more than 66,000 entries. <http://www.buddhism-dict.net>.

15. A small data extraction of each entry in the dictionary in XML format, providing a pointer to more detailed information.

16. It is important to clarify that the availability of the DDB extracts file is by no means, and has never been, a special limited arrangement between SAT and the DDB. The file is available for any other online web resource project that wishes to use it in a proper way. For example, it is presently being used by the Smarthanzi web application (<http://www.smarthanzi.net>), the DDBAccess program (<http://download.smarthanzi.net/ddbaccess>), and the Chinese Āgama pages of Sutta Central (<http://suttacentral.net>).

17. <http://www.bdk.or.jp/english>. Bukkyō Dendō Kyōkai (BDK) America was founded in 1986 by Rev. Numata Yehan, having as one of its primary goals the translation of texts from the *Taishō Canon* into English. To date, more than eighty works have been translated.

18. South Asia Research Documentation Services (bibliographic database) <http://www.sards.uni-halle.de>.

19. CiNii (Scholarly and Academic Information Navigator, pronounced "sigh-knee") is a database service which can be searched with academic information of articles, books, journals, and dissertations. It is an exemplary academic resource, providing access to full-text PDFs and so forth, which, in North America, are supplied mostly by expensive for-pay services such as JSTOR, Wiley, and so forth. <https://support.nii.ac.jp/en>.

20. For a list of persons involved in the proofing and correction of the SAT 2012 database, scroll to the bottom of [http://21dzk.l.u-tokyo.ac.jp/SAT/members\\_en.html](http://21dzk.l.u-tokyo.ac.jp/SAT/members_en.html).

21. The Character Information Service Environment (CHISE) project attempts to collect and organize into a knowledge-base information about characters in the scripts of the world. <http://chise.zinbun.kyoto-u.ac.jp>.

22. <http://www.unicode.org/charts/unihan.html>.

23. The inclusion of these ideographs in Unicode is still in progress. The IRG does not simply accept batches of ideographs and uncritically add them. Each character has to be checked with its source references, and it has to be established as being either fully unique or part of a variant plane including other ideographs. At the time of this writing, approximately 3,000 SAT ideographs have been approved.

24. From the IIF website (<http://iif.io>): "A growing community of the world's leading research libraries and image repositories have embarked on an effort to collaboratively produce an interoperable technology and community framework for image delivery. IIF (International Image Interoperability Framework) has the following goals: (1) To give scholars an unprecedented level of uniform and rich access

to image-based resources hosted around the world. (2) To define a set of common application programming interfaces that support interoperability between image repositories. (3) To develop, cultivate and document shared technologies, such as image servers and web clients, that provide a world-class user experience in viewing, comparing, manipulating and annotating images."

25. <http://dzkimgs.l.u-tokyo.ac.jp/SATi/images.php>.

26. <https://adho.org>.

27. The only (S) level project selected from candidates among the field of the humanities in Japan in 2015.

## WORKS CITED

- Academia Sinica Text Database. <http://hanji.sinica.edu.tw/>  
 Alliance of Digital Humanities Organizations (ADHO). <https://adho.org>  
 Bukkyō Dendō Kyōkai (BDK). <http://www.bdk.or.jp/english>  
 Butsugakujutsu shinko kai 仏學術振興会. <http://butsugakushin.org>  
 CHISE (Character Information Service Environment). <http://chise.zinbun.kyoto-u.ac.jp>  
 CiNii (Scholarly and Academic Information Navigator). <https://support.nii.ac.jp/en>  
 Digital Dictionary of Buddhism. <http://www.buddhism-dict.net>  
 Electronic Buddhist Text Initiative (EBTI). <http://www.buddhism-dict.net/ebti>  
 Indian and Buddhist Studies Treatise Database. <http://21dzk.l.u-tokyo.ac.jp/INBUDS/search.php>  
 International Image Interoperability Framework (IIF). <http://iif.io>  
 SAT Taishō Database. <http://21dzk.l.u-tokyo.ac.jp/SAT/ddb-bdk-sat2.php>  
 SAT Taishōzō Image DB. <http://dzkimgs.l.u-tokyo.ac.jp/SATi/images.php>  
 Smarthanzi. <http://www.smarthanzi.net>  
 South Asia Research Documentation Services (SARDS). <http://www.sards.uni-halle.de>  
 Sutta Central. <http://suttacentral.net>  
 Tripiṭaka Koreana 高麗大藏經. <http://kb.sutra.re.kr>  
 Unicode Consortium. <http://unicode.org/>  
 UniHan Database. <http://www.unicode.org/charts/unihan.html>  
 Wu, Jiang and Lucille Chia (ed.). *Spreading Buddha's Word in East Asia: The Formation and Transformation of the Chinese Buddhist Canon*. New York: Columbia University Press, 2016.

Author's note: The research for this article was supported by a grant-in-aid from the Japanese Society for the Promotion of Science under the title of "Construction of a New Knowledge Base For Buddhist Studies: Presentation of An Advanced Model for the Next Generation of Humanities Research." Project Number: 15H05725 (Shimoda Masahiro)