

Towards a digital research environment for Buddhist studies

Kiyonori Nagasaki and Toru Tomabechi
International Institute for Digital Humanities

Masahiro Shimoda
University of Tokyo

Abstract

This article will explain the structure and function of a digital research environment for Buddhist studies entitled Research Base for Indian and Buddhist Studies (RBIB). In the field of Buddhist studies (and especially in the area of Indian Buddhism), scholars face various obstacles in their efforts to share materials—in both paper and digital media. In establishing the RBIB, we have considered it to be of paramount importance to preserve the continuity of the prior tradition of studies within the paper medium, while at the same time developing a methodology that takes best advantage of the digital medium. After identifying six basic conditions that will satisfy the demands of both approaches, we have implemented a Web collaboration system that fulfills these conditions. Our system also makes functional parts of the RBIB that enable the presentation of the relationships between arbitrary fragments of related texts in various canonical languages through a user-friendly interface, which in turn allows users to learn how to edit and browse relationship data. Our system has thus far been highly evaluated by testers, and it has been further facilitated by recent developments in Information Communications Technology. Humanities digitization has a deep mutual relationship with this new technology.

Correspondence:

Kiyonori Nagasaki,
International Institute for
Digital Humanities,
5-26-4, Hongo, Bunkyo-ku,
Tokyo, Japan.

Email:
nagasaki@dhii.jp

With increasing number of electronic texts and other research materials available on the Web, the role played by digital resources has become more and more central in the field of the Humanities, and Buddhist Studies is no exception to this trend. At the same time, it must be admitted that this basic availability of digital resources has not necessarily led to a drastic improvement of the research environment. Although a considerable amount of Buddhist texts have been digitized and made available online, these resources are for the most part not optimally set up for advanced scholarly use. Most importantly, they lack standardized interface that would allow interoperation among them and work with equal efficiency for the philological data in

Sanskrit, Pāli, Tibetan, Chinese, and other languages in which Buddhism has been transmitted. In addition, the method of digital presentation of research results is yet to be established. Scholars of Buddhism are now in urgent need of a well-defined framework for the mutual integration of digital resources scattered across the Web. In this article, we shall first try to determine the requirements for the digital research environment of Buddhist Studies focusing primarily on its philological aspects and then present our tentative approach to the solution of the problem through the implementation of a collaborative Web framework called the Research Base for Indian and Buddhist Studies (RBIB).

Buddhism has created a huge corpus of scriptural, scholastic, philosophical, and liturgical works in the course of its long history. Scholars of Buddhism have to be able to handle them not only in different languages but also in different versions/recensions. Figure 1 illustrates in a simplified manner the historical relationship between different versions/recensions—from the archetype down to modern critical editions—of the *Mūlamadhyamakakārikā*, an Indian Buddhist philosophical treatise dated from the 2nd to 3rd centuries. As this example shows, one single text undergoes several changes in the process of transmission: the original text in Sanskrit was reproduced in the form of manuscripts (this process often introduces scribal errors), commented upon, translated into other languages (Tibetan and Chinese in this case), and so forth. In order to get closer to the original form of the text (though always hypothetical), any philological study should take into account the text's transmissional history and the mutual relationship between versions/recensions (see Steinkellner,

1988). This requires comparative analysis of the textual information provided by extant witnesses (in Sanskrit, Tibetan, and Chinese/in the form of manuscripts, xylographs, and printed editions in the case above). In addition, exegetical literature and other related works, if available, are also to be taken into consideration as secondary witnesses. The unit of textual comparison can vary: witnesses may be collated on the level of whole text, fascicle, chapter, paragraph, sentence, word, syllable, or character.

As a basic tool for the comparative analysis as described above, a collaborative digital research environment for Buddhist Studies should meet several requirements. First of all, the textual data provided by each resource should be punctuation-neutral. Because it is not explicitly indicated but sometimes strongly affects interpretation of the text, the word or sentence separation, which can be a matter of argument especially in Sanskrit and Chinese texts, must not be imposed by the data provider but should be left to the choice of the scholar who

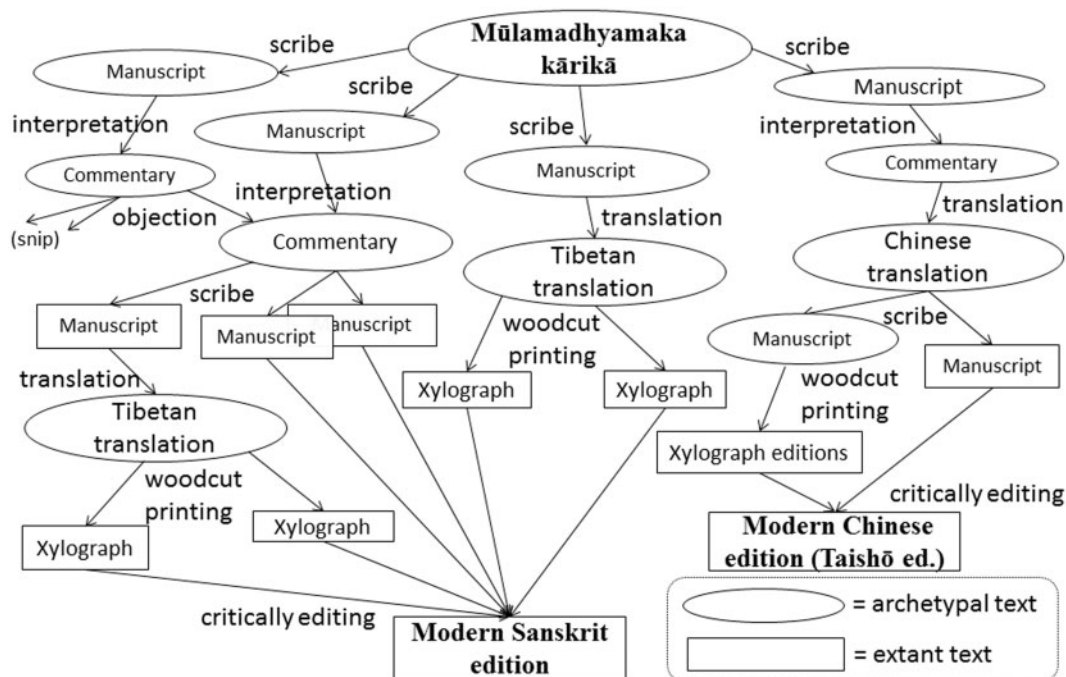


Fig. 1 An example: 'मूलमध्यमककारिका' (*Madhyamakakārikā*) and some related texts

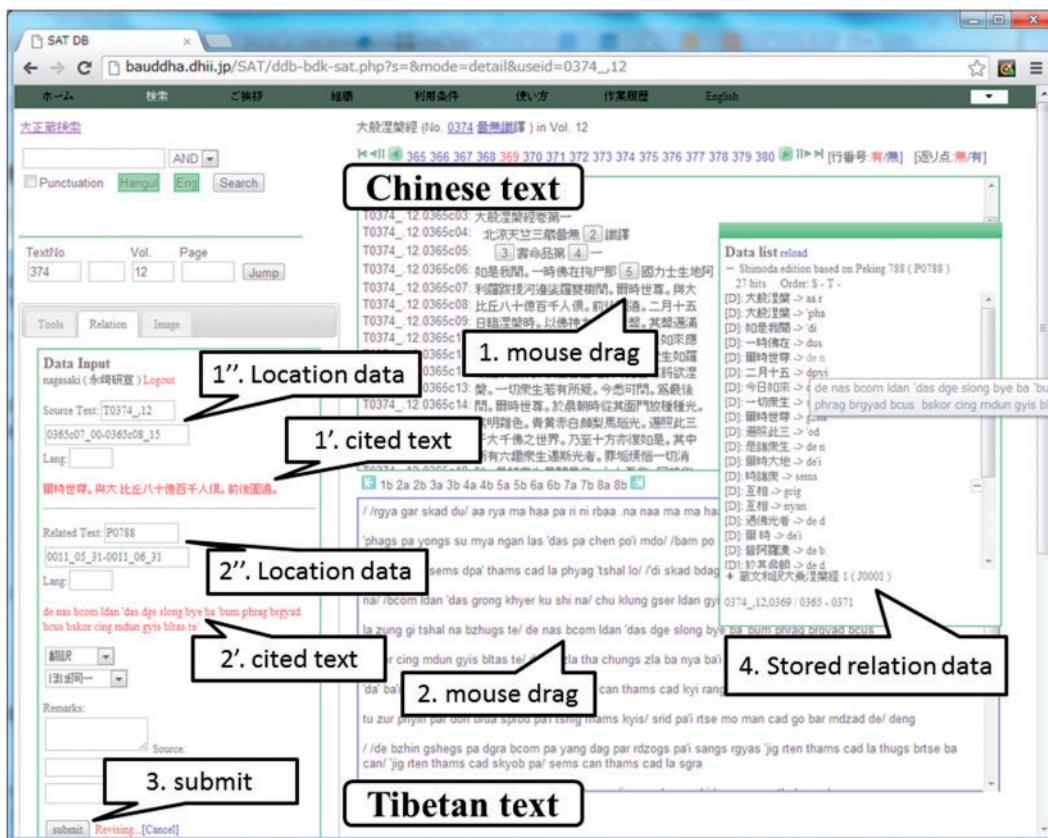


Fig. 2 The procedures of inputting relationship data on an environment

uses the resource. This also ensures the flexible use of the data and the user can freely select the textual units for comparison. The second requirement is an efficient system for storing annotations. Information on variant readings, links to other witnesses, to modern translations, to research papers, and other philological notes are to be appended to the text data through the annotation system consisting of a separate database shared by multiple text databases. To enable the use of complex annotations, the system should allow one annotation to overlap with another, following the stand-off approach described in Caton (2007). While based on these two principles, i.e. punctuation-neutral text data and separate annotation system, the actual implementation of the digital research environment should also be able to preserve the continuity from the past research results published in legacy

media. The latter requirement can be realized by using the traditional referencing method converted to URI (e.g. T0001,01,0001a01 for *Taishō Shinshū Daizōkyō*, Volume 1, Page 1, row a (of a, b and c), Line 1).

Following these basic technical principles, the RBIB aims to provide a collaborative research platform on the Web, where different text databases are connected through open Web API and a scholar can act as both contributor and beneficiary using an easy-to-learn, user-friendly Web interface. We have so far been trying to implement the concept of the RBIB using the SAT (Saṃganīkīkṛtaṃ Taiśotripitakam, a text database of the Chinese Buddhist canon with 150,000 accesses per month, <http://21dzk.l.u-tokyo.ac.jp/SAT/>; see Nagasaki et al., 2009) as a pilot case. Although the RBIB implementation on the SAT is still under development

大般涅槃經 (No. 0374 曇無讖譯) in Vol. 12

◀◀|| 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 ||▶▶ [行番号:有/無] [返り点:無/有]

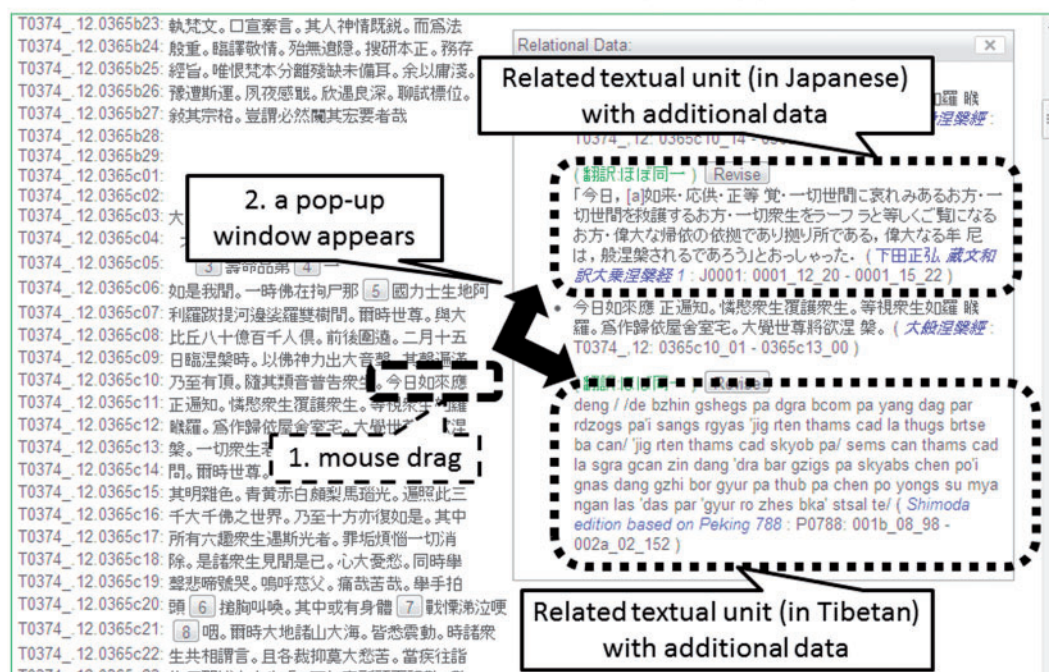


Fig. 3 The procedures of browsing the relational data on an environment

but partly open for the use of the general public, it is currently being tested intensively by our collaborators who have so far reported that the system is efficient and its use is easy to learn. Thanks to the development of Ajax, of which our system makes full utilization, the RBIB function on the SAT has been greatly improved compared with the older system described in Nagasaki (2008).

The RBIB implementation on the SAT currently consists of two basic components: (1) a database for the relationship information between text data and (2) a Web user interface. The relationship database stores the locations of the textual units in the text databases, and, optionally, annotations, type of relationship, contributor's name, and other information. The location information is stored as the start and end positions of the selected textual unit in the text database. The RBIB Web interface allows an easy selection of the text range by click-and-drag mouse operation. Figure 2 is a SAT screenshot

showing an example of relationship data creation between a Chinese text (upper central window) and a related Tibetan text. By just selecting a text region with mouse click-and-drag, the system creates the text unit's location information as URI and stores it along with the selected text in the database (right-hand window). To view the stored relationship data, the user has to only select a text region with click-and-drag: if the selected text unit has a data related to it, the system displays it in a separate window as shown in Fig. 3. The relationship database can be used by other digital resources outside the SAT through an open Web API. On the latest version of the SAT, this function is also applied to search and browse an English–Chinese parallel corpus of Buddhist scriptures. Additionally, this implementation will be extended so that image data can be handled.

As Lavagnino (2009) described, scholarly digital resources should be prepared for several distinct

levels of users. Even in a scholarly collaborative space, it is important for the quality control of the data to set up such distinction. The RBIB on the SAT takes the following approach to this question. The contributors of the RBIB are registered on three different permission levels, namely, observer, editor, and administrative editor. An observer has no permission to create the relationship data and can only check the progress of the work. An editor, main contributor of the data, inputs his/her own data using the Web interface of the RBIB. An administrative editor, besides being a data contributor, is responsible for the data consistency and quality as well as for the decision of data release. Non-contributing users are only permitted to browse the released data through the public Web interface. This separation of tasks according to the permission levels helps the better management of the project as a whole.

In closing, through the implementation of the RBIB, we have been continuously made aware of the importance of good user interface for this kind of project. While seemingly a superficial matter, the quality of the user interface can greatly affect the course taken by the project. As noted by Cummings (2007), Humanities scholars often stay reluctant to undertake or participate in a digital project exactly because of the steep learning curve of software usage. A collaborative project must overcome this problem. With a better interface, it is possible to expand the circle of collaborating scholars and this in turn contributes to the improvement of the research environment. We sincerely hope that the RBIB, as a test case, will help the future

development of the Web collaboration not only in the field of Buddhist Studies but also in the Humanities in general.

Acknowledgements

This work was supported by MEXT/JSPS KAKENHI grants No. 22242002 and No. 22700255.

References

- Caton, P. (2007). *Distributed Multivalent Encoding, Digital Humanities 2007 Conference Abstracts*. University of Illinois, Urbana-Champaign, pp. 33–4.
- Cummings, J. (2007). The text encoding initiative and the study of literature. *A Companion To Digital Literary Studies*. Oxford: Blackwell Publishing, pp. 451–76.
- Lavagnino, J. (2009). Access. *Literary and Linguistic Computing*, 24(1): 63–76.
- Nagasaki, K. (2008). A collaboration system for the philology of the Buddhist study. *Digital Humanities 2008 Book of Abstracts*, University of Oulu, Finland, pp. 262–3.
- Nagasaki, K., Charles Muller, A., and Shimoda, M. (2009). Aspects of the interoperability in the digital humanities. *Digital Humanities 2009 Conference Abstracts*, University of Maryland, MD, pp. 375–7.
- Steinkellner, E. (1988). Methodological remarks on the constitution of Sanskrit texts from the Buddhist pramāṇa-Tradition. *Wiener Zeitschrift für die Kunde Südasiens*, Band XXXII: 103–29.