**NAGASAKI — Significance**

<2 images>

**Significance of Linking between Past and Present, East and West, and Various Databases**

Nagasaki, K., Hackett, P., Muller, A. C., Tomabechi, T. and Shimoda, M.

It is the case with Buddhist studies—as with most other fields in the humanities—that various types of digital resources have been developed. So far, there have been several core digital resources based on traditional series of Buddhist scriptures, indexes, and dictionaries, which had been published within the long era of paper media. It is, in fact, efficient to use such traditional research tools in order to connect the digital resources with existent secondary sources. In this context, digitization makes us aware of what we have been depending upon; it also helps to reveal issues that have been inherent in our work, even during the era of paper media. We will describe the significance of digitization for Buddhist studies from this viewpoint through a recent case in the SAT Database (SAT DB) (Nagasaki et al., 2013).

The SAT DB is a web service that delivers an integrated research environment for Buddhist studies based on a corpus of digital classical Chinese Buddhist scriptures. The SAT DB has connected with several academic and cultural resources such as the Digital Dictionary of Buddhism, a parallel corpus of the Buddhist scriptures translated into English, and several bibliographical databases and character databases—all linked by implementation of convenient interfaces. The web services are accessed over 300,000 to 500,000 times per month. The SAT DB has recently been linked text by text with several other digital resources, such as the Buddhist Canons Research Database (BCRD)1 and some other digital repositories. We will briefly describe the BCRD below.

The BCRD is the only comprehensive index to Tibetan Buddhist canonical materials, providing cross-references between primary materials and detailed publication information to secondary literature, while incorporating hyper-text links to online resources. Standard library cataloging systems provide bibliographic information at the ‘item’ level (a monograph, a serial, etc.) only. This level of cataloging information is inadequate for classical collections such as the various Buddhist canons. To illustrate by example, the Tibetan Buddhist canon consists of approximately 5,000 individual works of varying authorship; an individual recension of the Buddhist canon is typically represented in standard library catalogs by either one or two bibliographic records. All other relevant details for accessing individual works are relegated to secondary reference literature. The same holds true for the vast majority of subsequent commentarial literature that has often been published in large collections grouped thematically or by author.

The BCRD has compiled complete documentation for the Tibetan Buddhist canon, as well as another cross-linked 3,500 post-canonical works. It also provides the raw bibliographic resources to enable advanced research by fully documenting previous research and to provide constantly updated bibliographic references to print and digital resources in multiple languages. Such a project benefits a broad range of scholars in the humanities by providing detailed and accurate reference information otherwise not readily available, and in a quick and concise manner.

In addition, the BCRD provides full-text searching of the entire Tibetan Buddhist canon (15 million syllables). At the present time, it has more than 700 active users and deploys cutting-edge techniques of Natural Language Processing to enable intelligent searching and precise identification of search results. As a result, user feedback has been enormously positive, and users have described the resource as ‘revolutionary’ and ‘a huge advance . . . [and] enormously beneficial to scholars in the field’.

The linkage between the BCRD and the SAT DB has been made possible based mainly on a corresponding catalogue (Ui et al., 1934) that was a milestone in the paper media—that is, the two present resources are connected with the past resource. It has streamlined the workflow of browsing corresponding scriptures across the Chinese and the Tibetan, which had been important and inevitable but needed complex procedures in order to explore transition of various aspects of Buddhism. Moreover, each separate database project is now relieved from having to do the various works of building mutual databases. It will also be useful to explore further issues in both corresponding scriptures and to improve earlier catalogues.

The SAT DB also has begun to gather URLs of digital images of primary resources of Buddhist scriptures, such as manuscripts and woodcut printings, which are published in various cultural data repositories. We have tentatively obtained approximately 600 URLs and have uncovered several problems. In the case of secondary resources, the SAT DB has already provided a stable linking service by semi-automatic search by use of Web API in several search engines of journal articles. However, in this case, we were not able to use some popular convenient methods such as Web API due to the fact that the metadata in each repository are not unified and the contents cannot be retrieved by texts. Our collection must be done by manually checking each image. In Japanese cultural web resources, at least two integrated search engines are enabled, but neither provides enough search function for our usage. HathiTrust2 and gallica,3 from which we gathered URLs, are also not sufficient. In the case of HathiTrust, the names of scriptures are not yet regularized with our databases. In the latter, as it includes various fragments of manuscripts of the scriptures, we need to identify the location of them one by one in order to provide convenient usage. So far, we have targeted the repositories of HathiTrust, gallica, the National Diet Library of Japan (NDL),4 the University of Waseda, the University of Ryukoku, Ritsumeikan University, the National Institute of Japanese Literature, and so on (see Figures 1 and 2). As the targets will increase further, we should check URLs continuously. Meanwhile, we should analyze each repository and design appropriate metadata for our database and disseminate these for efficient interoperability according to several standards, because some repositories lack publication data, some lack a method of writing, and some don’t include format of metadata. Moreover, we released a Web API to provide our link data so that other providers can avoid reinventing the wheel. Thus, it will also be useful to provide a model case for interoperability of Eastern cultural resources.



Figure 1. The dialog to link with other resources.

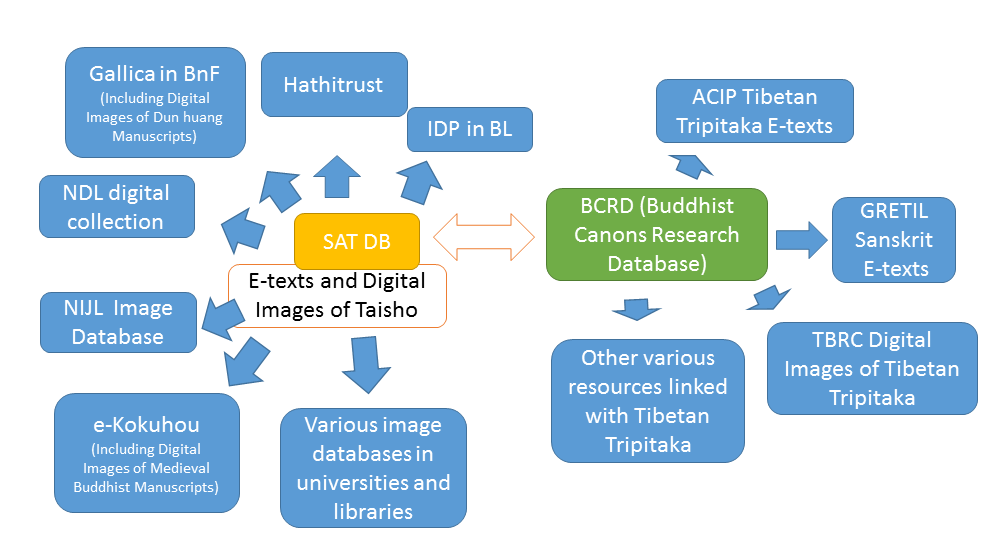


Figure 2. The relationship of digital repositories via SAT DB and BCRD.

**Notes**

1. http://www.aibs.columbia.edu/databases/New/index.php.

2. http://hathitrust.org/.

3. http://gallica.bnf.fr/.

4. National Diet Library Digital Collection, http://dl.ndl.go.jp/?\_\_lang=en.

**References**

**Nagasaki, K., Tomabechi, T. and Shimoda, M.** (2013). Towards a Digital Research Environment for Buddhist Studies. *Literary and Linguistic Computing,* **28**(2): 296–300.

**Ui, H., et al.** (eds). (1934). *A Complete Catalogue of the Tibetan Buddhist Canons.* Sendai, Japan.