Gandhari.org

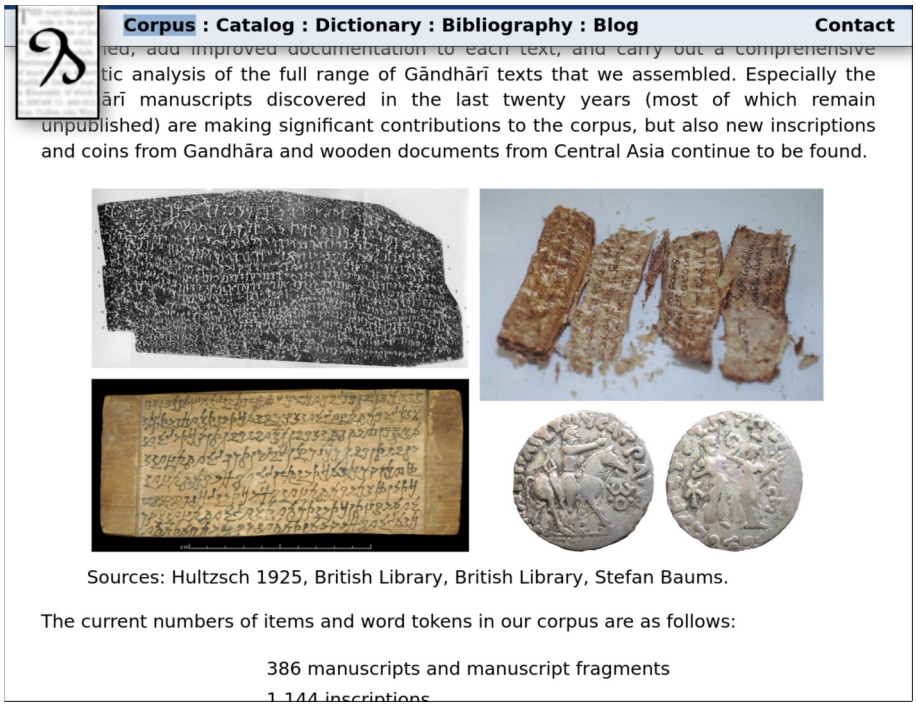
A Research Environment for Gāndhārī Buddhist Texts and Beyond

Stefan Baums
University of Munich
baums@lmu.de

Andrew Glass
Microsoft Corporation
asg@uw.edu

Overview

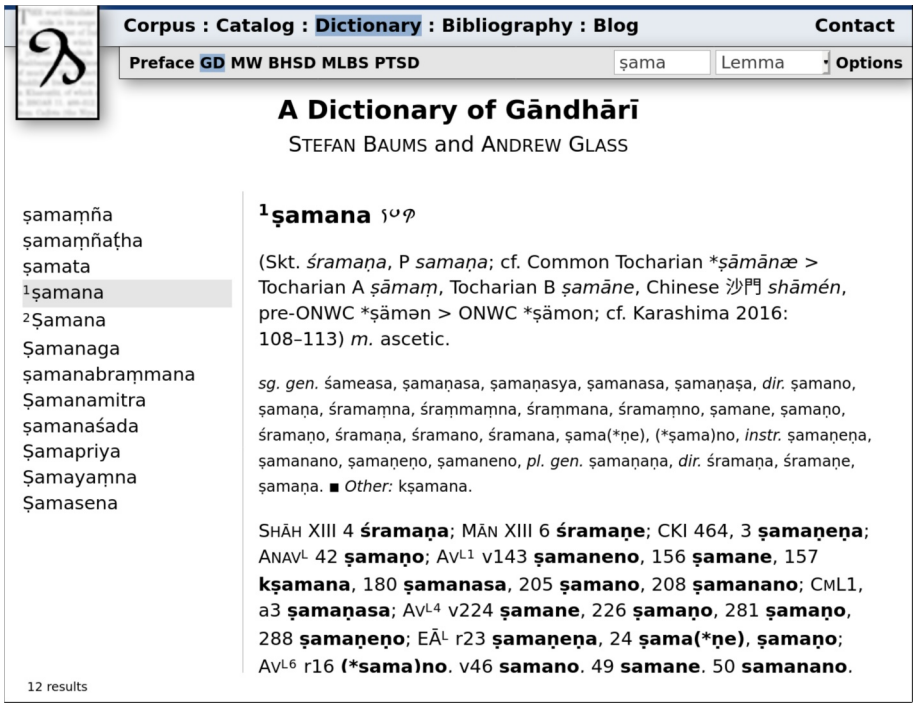
The earliest Buddhist manuscripts are preserved to us in the form of birch-bark scrolls dating from as early as the first century BCE to the third century CE and written in the Gāndhārī language and Kharoṣṭhī script. These manuscripts go back to the very beginning of the use of writing by Buddhist communities and as such cast valuable light on processes such as the transition from oral to written transmission and the formation of Buddhist canons. Gāndhārī manuscripts also formed the basis for the earliest translations of Buddhist texts into the Chinese language, and thus laid the groundwork for the export of Buddhism from India on its way to becoming a true world religion.



About one hundred and fifty substantial Gāndhārī manuscripts are now known, all but one of them discovered in the last twenty-five years, and research on them and their publication are still in their beginning. The fragmentary state of many of the manuscripts as well as the nature of the language and script present great difficulties to the interpretation of this uniquely important material. It is essential to study the manuscripts in the context of the complete corpus of contemporary Gāndhārī inscriptions, administrative documents and coin legends to arrive at an

understanding of their content, and to consider physical and geographical aspects of the carriers of all these texts to gain a well-rounded picture of Buddhism at the crossroads of ancient Gandhāra.

In 1997, one of the nominees – Andrew Glass – began the development of digital tools to help in this formidable task, starting from the preparation of digital texts for the only Gāndhārī manuscript then published (the Khotan Dharmapada) and the Senavarma inscription. He was joined by the other nominee – Stefan Baums – in 2002 at the University of Washington, and by 2003 they had developed a first public database system containing several hundred Gāndhārī inscriptions as well as the Khotan Dharmapada. This first system was replaced in 2007 by the system currently still in use (and continuously updated) on the website operated by the nominees, Gandhari.org (https://gandhari.org). On this website, the nominees assembled a complete corpus of all published Buddhist (and some non-Buddhist) texts in the Gāndhārī language, consisting of currently 2,761 items with a total of 114,377 word tokens. Uniquely among Buddhist digital corpora, the Gandhari.org corpus combines manuscript and epigraphical sources and thus casts light on all aspects of the Buddhism of Gandhāra reflected in any type of written source. It presents images of the inscribed objects in parallel with their textual content and with a rich set of metadata situating and organizing the material in time and space. On the basis of this database, Baums and Glass are writing the first ever *Dictionary of Gāndhārī* (supported by a *Catalog* and *Bibliography* of source texts), which has reached 7,686 articles with a total of 38,657 references and continues to grow at a brisk pace as new material becomes available. One feature of Gandhari.org appreciated far beyond the field of Gāndhārī studies is the convenient parallel availability of Sanskrit, Buddhist Sanskrit and Pali dictionaries that allow researchers to trace the development of words through all the early languages of South Asian Buddhism.



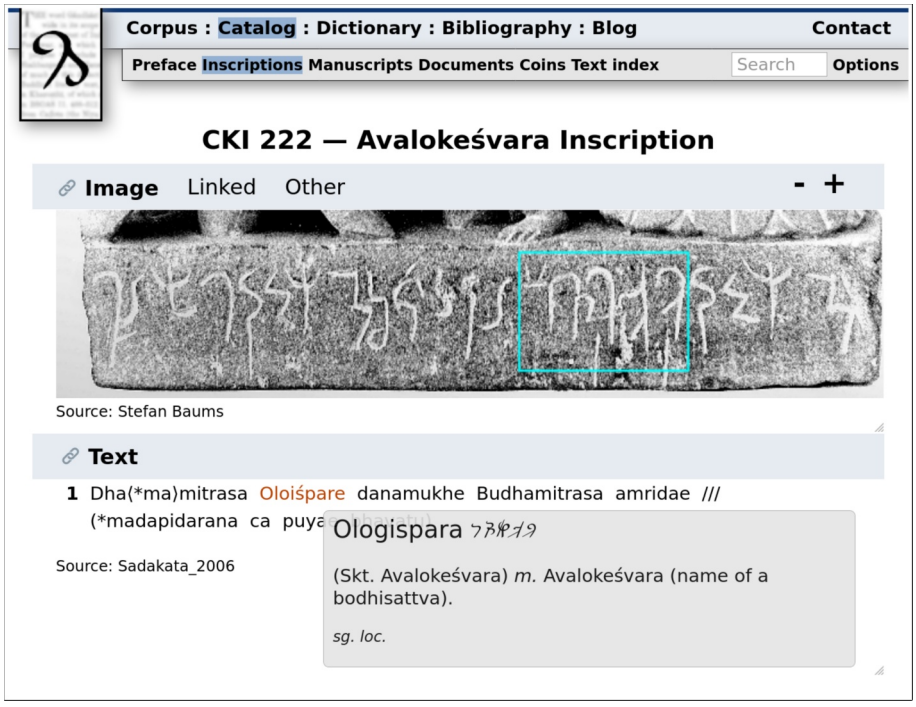
Unicode and Font Development

In order to further support the computer-aided study and publication of Kharoṣṭhī-script material, Baums and Glass developed a proposal for the encoding of the Kharoṣṭhī script, which be-

came part of the Unicode standard in version 4.1 (March 2005, updated in version 11.0, June 2018). It was followed by another proposal from Baums and Glass for the ancient Brāhmī script, accepted into the Unicode standard in version 6.0 (October 2010, updated in version 7.0, June 2014). After designing a comprehensive and widely used transliteration font (Gandhari Unicode), Glass then developed a free and open-source Kharoṣṭhī font (Kharoshthi Unicode) that in combination with the character coding gives full support for the use of Gāndhārī texts in their original script in digital environments. Another Kharoṣṭhī OpenType font developed by Glass as part of the Segoe font set is now included in the Windows operating system. The Gandhari.org website offers the option of displaying *Dictionary* lemmata in the original script using this technology, and an option to display all Gāndhārī content on the site (including full source texts) in the Kharoṣṭhī script is under development. Last not least, the experiences of Baums and Glass with the computer encoding and rendering of the Kharoṣṭhī and Brāhmī scripts led to the development, by Glass, of a Universal Shaping Engine for the Windows operating system, which has since added support for many more of the world's complex scripts (including the Siddham script as used in the Taisho canon) across operating systems.

New Features and READ

Ever since they first deployed the current Gandhari.org software (written by Glass) in 2007, the nominees made plans for a second-generation system with advanced features such as fine-grained text-image linking, the semi-automatic generation of glossaries for individual texts, a dependency-grammatical syntax module, the generation of paleographic reports, the parallel management of more than one historical or current edition of a given source text, and roundtripping between the internal data format of the Gandhari.org database and a set of TEI schemata. Apart from improving the usefulness of Gandhari.org as a research resource, the nominees felt that these features would also benefit other scholars and empower them to compile similar databases for other Buddhist textual corpora, with the possibility of eventually building a Linked Data network of databases in the field of textual Buddhist studies.



Initially, resources were lacking to develop such an ambitious new system, but with the establishment in 2012 of the Buddhist Manuscripts from Gandhāra project at the Bavarian Academy of Sciences and Humanities (designed by Baums with strong support for digital tool development in mind), the realization of Baums and Glass's vision became possible. In order to emphasize the general usability of the system in Buddhist and other ancient textual studies, the new software system was named "Research Environment for Ancient Documents." The bulk of the programming work on READ is carried out by Stephen White (Venice), and while the Bavarian Academy continues to provide the large majority of its funding, a growing number of projects and institutions have made contributions and adapted READ to many languages ranging from Gāndhārī, Sanskrit, Pali and Tibetan to Egyptian, Latin and Mayan. Within the field of Gāndhārī studies, the new software system supports the production of future volumes in the publication series Gandhāran Buddhist Texts (University of Washington Press, to be released under Creative Commons BY licensing), for which Baums also serves as co-editor. The source code of READ is released as free and open-source software under the terms of the GNU General Public License. In parallel with this development effort, Glass completely rewrote the user-facing Gandhari.org website software to make use of READ as a new backend for Baums and Glass's philological work. The new system can already be seen in action under the address https://gandhari.org/beta, with the immediate benefit of contextual popups providing lexical information for all texts covered in the *Dictionary*. Baums and Glass are now beginning to make use of the many other features and standards support that the new software system brings to their site for the benefit of its users and improved interaction with other digital resources in Buddhist Studies and beyond.

References

Baums, Glass. 2006. "Towards a Computer Encoding for Brāhmī." In: Adalbert J. Gail, Gerd J. R. Mevissen and Richard Salomon, eds., *Script and Image: Papers on Art and Epigraphy,* Delhi: Motilal Banarsidass Publishers (Papers of the 12th World Sanskrit Conference, Vol. 11.1), pp. 111–143. Online

Baums, Stefan, and Andrew Glass. 2002– a. *A Dictionary of Gāndhārī*. https://gandhari.org/dictionary

———. 2002– b. *Bibliography of Gāndhārī Studies*. https://gandhari.org/bibliography

———. 2002– c. *Catalog of Gāndhārī Texts*. https://gandhari.org/catalog

———. 2007. "Proposal for the Encoding of Brāhmī in Plane 1 of ISO/IEC 10646." https://www.unicode.org/wg2/docs/n3491.pdf

———. 2013. "Die Lexikographie der Gāndhārī-Sprache." *Akademie aktuell: Zeitschrift der Bayerischen Akademie der Wissenschaften* 44: 44–47. Online

Baums, Stefan, and Andrea Schlosser. 2018. "Birkenrinde und Computer." *Akademie aktuell: Zeitschrift der Bayerischen Akademie der Wissenschaften* 64: 38–39. Online

Glass, Andrew, and Stefan Baums. 2017. "Additional Characters for Kharoṣṭhī Script." https://www.unicode.org/wg2/docs/n4812-kharosthi-add.pdf

Glass, Andrew, Stefan Baums and Richard Salomon. 2003. "Proposal to Encode Kharoṣṭhī in Plane 1 of ISO/IEC 10646." https://www.unicode.org/wg2/docs/n2524.pdf

Glass, Andrew, and Shriramana Sharma. "Proposal to Encode 1107F Brahmi Number Joiner (Revised)." https://www.unicode.org/wg2/docs/n4166.pdf

Pollock, Sheldon. 2012. “Sanskrit Studies in the United States.” In: Radhavallabh Tripathi, ed., *Sixty Years of Sanskrit Studies (1950–2010), Vol. 2: Countries Other than India*, pp. 259–310. New Delhi: Rashtriya Sanskrit Sansthan. [P. 284.]

Veidlinger, Daniel. 2019. *Digital Humanities and Buddhism: An Introduction*. Berlin: De Gruyter. (Introductions to Digital Humanities – Religion, Volume 1.) [Pp. 12–13, 31, 229.]

Gandhari.org website (production version): https://gandhari.org

Gandhari.org website (beta version): https://gandhari.org/beta

Gandhari Unicode font: https://github.com/gandhariunicode/gandhari_unicode_font

Kharoshthi Unicode font: https://github.com/kharoshthiunicode/kharoshthi_unicode_font

READ software repository and information: https://github.com/readsoftware/read

Universal Shaping Engine: https://docs.microsoft.com/en-us/typography/script-development/use