

CHAPTER 4

Looking beyond the text: Opportunities and challenges in the digitisation of Sanskrit inscriptions

Francesco Bianchini

King's College, University of Cambridge

Abstract

This chapter provides an introduction to the main digital repositories of inscriptions from South and Southeast Asia. The digitisation of epigraphs in Sanskrit and other languages has found considerable impetus in recent years, being the focus of two ERC Synergy projects as well as a number of other scholarly initiatives. While producing reliable editions and comprehensive metadata remains a central concern, the digital environment brings unique opportunities to move beyond traditional printed editions. The repositories introduced below already offer some practical solutions as to how this may be achieved, especially through a more integrated approach to the epigraphic object as a whole. This includes, among other aspects, recording object data as well as the physical layout of the inscriptions, and the integration of GIS technologies. Despite recent progress, the chapter argues that more can be done in this direction. Another aspect that deserves further attention is the development of the database itself,

How to cite this book chapter:

Bianchini, F. 2023. Looking beyond the text: Opportunities and challenges in the digitisation of Sanskrit inscriptions. In: Palladino, C. and Bodard, G. (Eds.), *Can't Touch This: Digital Approaches to Materiality in Cultural Heritage*. Pp. 77–93. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bcv.e>. License: CC BY 4.0

especially in terms of advanced search and query capabilities, as well as cross-database communication. Lastly, the chapter raises the issue of the historical complexities involved in standardisation and efforts towards decolonisation. It suggests that a diversified approach involving independent teams that nonetheless communicate with one another may be a viable way forward.

Abstract (Italiano)

Il presente capitolo offre una panoramica delle principali collezioni epigrafiche digitali riguardanti l'Asia meridionale e sudorientale. La digitalizzazione di fonti epigrafiche in sanscrito e altre lingue ha di recente ricevuto una notevole attenzione, trovandosi al centro di ben due ERC Synergy projects nonché di altre iniziative accademiche. Certamente il miglioramento delle edizioni e delle banche dati attualmente disponibili rimane uno scopo importante. Tuttavia è imperativo sfruttare al massimo le opportunità offerte dal contesto digitale rispetto alle tradizionali edizioni cartacee. Le collezioni descritte qui sotto già offrono varie soluzioni pratiche in questo senso, in particolare tramite un approccio integrato all'oggetto epigrafico, che non consiste certo di solo testo. Tra i vari aspetti menzioniamo la registrazione di dati relativi all'oggetto stesso, nonché il formato del testo come appare sulla superficie iscritta, oppure la mappatura digitale che posiziona l'artefatto all'interno di un contesto archeologico e paesaggistico. Nonostante i risultati già raggiunti, il presente capitolo suggerisce che le possibilità non sono per nulla esaurite e guarda con trepidante attesa ai risultati di certi progetti attualmente in corso. Un altro aspetto da considerare è lo sviluppo delle banche dati e soprattutto degli strumenti di ricerca e interrogazione dei dati stessi. In ultima istanza, vanno anche considerate le varie complessità storiche che possono sorgere nel contesto post-coloniale, in cui la standardizzazione delle banche dati presenta sia opportunità che problemi. Si suggerisce qui sotto un approccio che valorizzi la diversità e allo stesso tempo la collaborazione tra team indipendenti di studiosi.

1. Introductory remarks

The study of pre-modern inscriptions in Sanskrit and other languages of Southern Asia has made great progress in recent years.¹ On the one hand, epigraphists have focused on what might be considered 'groundwork', i.e. producing reliable editions—in many cases for the very first time—and recording metadata systematically and comprehensively. On the other hand, scholars have also explored new ways and techniques for dealing with epigraphic texts and associated objects. This is particularly true of efforts towards the digitisation

¹ For an introduction to these sources see Salomon 1998; Francis 2018. On Sanskrit inscriptions from Southeast Asia see Griffiths & Lammerts 2015.

of Sanskrit corpora, with a number of epigraphic repositories now accessible online—each showcasing a different set of strategies, methods and techniques.

Improving the reliability of edited texts remains a key aim for epigraphists. In the digital environment, this process can be supported by the adaptation of EpiDoc XML encoding to the specific features and needs of Asian languages and scripts. However, the digital environment and EpiDoc itself also offer various opportunities to think beyond traditional printed editions.² Epigraphists have already started to explore a range of possibilities, for example in terms of documenting visual features, creating searchable databases of object metadata, and exploring digital mapping technologies (GIS).

Furthermore, the digitisation of Sanskrit inscriptions offers an opportunity to reflect upon a number of social and cultural issues. This involves reflecting on and transcending colonial and eurocentric ways of approaching Asian cultural heritage. Possible strategies include encouraging more inclusivity within epigraphic task-forces, welcoming the formation of digital repositories across multiple institutions, and proposing new ways of cataloguing and describing objects, not necessarily based on pre-existing colonial archival practices. On the other hand, there is also the question of how emerging Asian nationalisms can impact historical and cultural studies, and more specifically of how Asia-based digital repositories can come into a constructive dialogue with Western scholarship.³

The present chapter aims at highlighting and discussing these issues, without claiming to be comprehensive or conclusive. In introducing various online repositories, it seeks to highlight the positive, while also indicating where more progress could be made. In terms of methodology, the main suggestion made in this chapter can be summarised as follows: the digital environment can help revolutionise the way we think about inscriptions as well as the way we study them. In order to achieve this, I suggest focusing on what is not immediately visible, on what surrounds the ‘text’ and would otherwise remain implicit in most traditional printed editions. This includes recording the links between texts and textual layouts; texts and objects; objects and their monumentalized environment, and even monuments and (inter)regional cultural networks. Furthermore, there is an increasing need to build extensive databases that can be queried in advanced ways, not only in terms of locating textual strings, but also in terms of retrieving object features, as well as geographical and chronological parameters. Digital repositories that meet such criteria would significantly improve our ability to develop a more compelling—and in fact decolonised—form of historiography.

² On digital editions of inscriptions see also Filosa, Gad & Bodard (Chapter 3 in this volume).

³ See sections 4 and 5 below for further comments on these issues, and Baba (Chapter 2 in this volume) for further discussions on cultural heritage and digitisation.

2. Digitised palm-leaf manuscripts

Before diving into epigraphic sources, it is helpful to look at digital manuscript repositories, for these offer a range of digitisation strategies that are also relevant to inscriptions.⁴ For example, a number of Sanskrit manuscripts—mostly palm-leaf or paper—are available on the Cambridge Digital Library.⁵ One may access the specimen called *Suvarṇaprabhāsottamasūtra* (CUL MS Add.875), an 18th century paper manuscript copied in Kathmandu and preserving the text of a major Buddhist scripture.⁶ The interface presents high-quality zoomable images on the left, with detailed metadata on the right side of the screen (Figure 4.1). Both images and metadata (XML) can be downloaded by the user.⁷ The metadata offers a general description of the specimen and the text it contains, before listing various specifications, many of which address the physical object and its characteristics. Then, incipit, rubrics, and colophon are presented in roman transliteration, followed by bibliographic details. The user is thus presented with what appears to be meticulously collected and systematically arranged data.

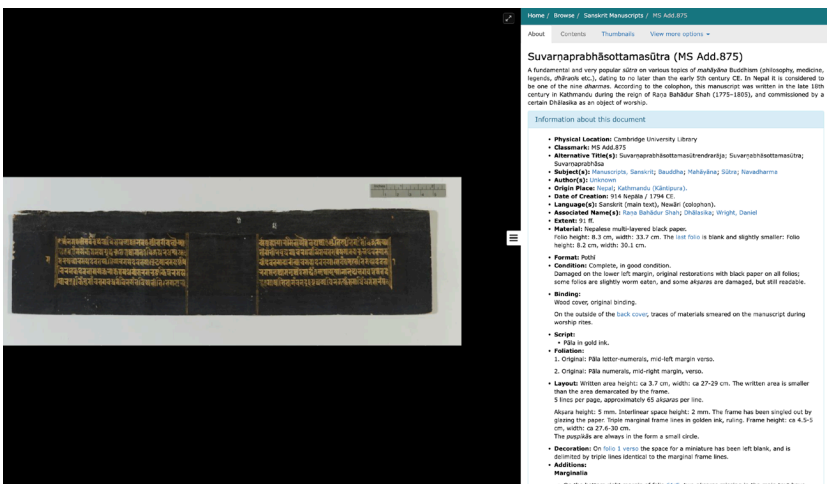


Figure 4.1: CUL, digitised Sanskrit MS (image and metadata).

⁴ On manuscript digitisation see also Woodward, Offner & Blackwell (Chapter 6 in this volume).

⁵ <https://cudl.lib.cam.ac.uk/collections/sanskrit/1> (Accessed January 2023).

⁶ <https://cudl.lib.cam.ac.uk/view/MS-ADD-00875/1> (Accessed January 2023).

⁷ In this chapter, I will not dwell on copyright issues, which represent of course a key aspect of digitisation.

The carefully edited texts can be easily checked on the basis of the images on the left side of the page. Editions also include punctuation marks, symbols, folio numbers, fillers, string holes, substrate defects and annotations, with conventions listed in a separate PDF.⁸

One may now ask which features make fuller use of the digital platform and the opportunities that come with it. For example, folio specifications within the edited texts link directly to the relevant image, which saves the user time; metadata often also includes hyperlinks. Indications about illuminations, bindings and marginalia link directly to the relevant folios. Authors and toponyms are part of the database. One interesting consequence of this is that by clicking on a place name one gets results from across the entire library, not just the Sanskrit section (clicking on ‘Nepal’ will bring up a number of Darwinian letters, for example). This might not always be useful to subject specialists but it can produce new insights and unexpected associations. The final portion of the metadata provides links to other specimens of the *Suvarṇaprabhāsottamasūtra* from two separate digital manuscript repositories (the International Dunhuang Project and the Nepalese German Manuscript Cataloguing Project). There is thus—to some extent—a synergy across different databases and platforms. This is in my opinion one of the key features to look for in digital collections, and a stepping stone towards more ambitious efforts, such as Open Linked Data.

Finally, each manuscript is arranged through ‘subjects’. This allows one to view the entire list of manuscripts associated in terms of genre, subject, language and so on. For example, opening the database link for *Vyākaraṇa* (a term indicating the emic grammatical tradition), the researcher has a glimpse of all relevant texts across the collection, arranged according to sub-topics and with indication of the number of available specimens. The drop-down menu ‘date’ gives the number of specimens available for each century. Taken with all due caution, this is information that can be interpreted and used in research, with potential implications that may transcend the history of the Cambridge collection itself.⁹

There is much about this database one would wish to find in repositories dedicated to epigraphic texts. However, epigraphic texts are in a way even more complex, for they are more intimately linked with landscapes and monuments, making GIS and in-situ photography quite indispensable.

⁸ <https://www.repository.cam.ac.uk/handle/1810/326907> (Accessed January 2023).

⁹ <https://cudl.lib.cam.ac.uk/search/advanced/results?subject=vyakarana> (Accessed January 2023).

3. Two pioneering repositories of Sanskrit inscriptions

The first digital epigraphic repository to adopt EpiDoc XML encoding was the *Corpus of Inscriptions of Campa* (2012), thanks to the efforts of Arlo Griffiths and various other scholars.¹⁰ This was then followed by a second repository called the *Early Inscriptions of Āndhradeśa* (2017).¹¹ In both cases, the EFEO (École française d'Extrême-Orient) collaborated with various partners, based in France or overseas.

The two repositories are similar enough in concept to justify treating them together. One can perhaps conceptualise the key aims of the editors as follows: achieving reliability, transparency, and comprehensiveness.¹² Reliability is particularly important here, for many of the previously available editions were simply faulty. In the field of Sanskrit epigraphy, investing time, efforts and resources into improving editions is still a valuable and legitimate enterprise, however unambitious this might seem from the perspective of more advanced historiographical research. As for transparency, the idea in both repositories is to provide the reader with an image against which they can readily check the edition. As for comprehensiveness, the repositories collect secondary sources that are scattered across old and rare journals, with information often found in archaeological notes, the retrieval of which can be very time consuming. Corpora that provide such valuable contributions are definitely to be welcomed.

The *Corpus of the Inscriptions of Campā* presented about 50 records, including inscriptions on steles, pedestals, door jams, dishes and vases. The website names at least three Asian specialists who acted as collaborators to the project.¹³ Although simple and somewhat rudimentary, it is a must-go for any scholar interested in the inscriptions of this ancient kingdom. The website has been discontinued but the materials continue to be expanded under the framework of an ongoing ERC Synergy project introduced below.

The key part of each record appears to be the edition, which is usually very detailed and accompanied by an apparatus with copious notes (Figure 4.2). Translations in both English and French are also offered. For our present purposes, it is important to note that the pictures usually cover not only estampages but also the objects themselves. Thus, one gets a clear idea of the substrates on which the texts have been inscribed and their architectonic, monumental, or utilitarian functions. The metadata is somewhat minimal but it includes

¹⁰ <https://isaw.nyu.edu/publications/inscriptions/campa/index.html> (Accessed January 2023).

¹¹ <http://hisoma.huma-num.fr/exist/apps/EIAD/index2.html> (Accessed January 2023).

¹² Read the author's statement here: <https://isaw.nyu.edu/publications/inscriptions/campa/about.html> (Accessed January 2023).

¹³ <https://isaw.nyu.edu/publications/inscriptions/campa/credits.html> (Accessed January 2023).

Corpus of the Inscriptions of Campā

HOME ABOUT CREDITS INSCRIPTIONS BIBLIOGRAPHY SEARCH

C. 216 Stela inscription of Hoà Lai

Please note: you are reviewing a preprint version of this publication. Contents here may change significantly in future versions. Scholars with specific interests are urged to consult all cited bibliography before using our texts and translations or drawing other significant conclusions.

Support Stela: sandstone; dimensions of the stela (excluding the tenon): h. 92 cm × w. 60 × d. 13 (the tenon adds 16.5 cm, bringing the total height to 108.5 cm); dimensions of the surviving part of the socle: h. 25.5 cm × w. 69 × d. 46.

Text The stela itself is engraved on three faces, and the socle is inscribed as well. Principal face A bears an incipit and 19 lines; lateral face b bears 11 lines; principal face C bears 15 lines; the socle bears 2 lines on two edges. All are written in *Sanskrit*. Together, these lines represent two chronologically distinct but textually intertwined inscriptions. The original inscription (C. 216.1) comprised 18 lines on the first, and 15 on the second principal face. The second inscription (C. 216.2) was subsequently inserted into the former, by adding a 19th line on face A, which was then continued onto the lateral face with three lines, after which the reader was to resume on the posterior face; presumably on the same occasion, extra text was added on the socle.

Date Text 1: 5 waxing Vaiśākha, 700 Śaka (6 April, 778 CE); Text 2: 7 waning Pusya, 760 Śaka (10 January, 839 CE).


Origin Temple of Hoà Lai (Ba Tháp, Bắc Phong, Thuận Bắc, Ninh Thuận, Vietnam).

The stela was discovered in 2006. We found the stela in the store room of the museum in 2009, as the one bearing inventory number 1444/D.25 (see *ECIC II*: 275-276).

Edition(s) First published in *ECIC II*, whence the present edition.

Facsimiles

- Estampage: EFEO n. 1900
- Estampage: EFEO n. 1004



Photograph, with scale, showing face C of inscription C. 216. Taken in the Phan Rang Museum by Arlo Griffiths on 12 October 2009.

Figure 4.2: Sample inscription from Vietnam (metadata and image).

a description of the support and used to comprise various hyperlinks which are now inactive. As per the original intention, by clicking on ‘gilded silver’ one could have viewed all the objects made of or covered in silver. The search engine itself seems rather limited and mostly oriented at finding text strings present within the corpus.

The EpiDoc XML file can be downloaded, but the website interface does not include the familiar toggle option between diplomatic and critical editions. It appears to me that the main reason—although certainly not the only one—for adopting EpiDoc encoding here is that it allows the editors to update the editions and keep improving the records as new information becomes available. Printed editions simply do not offer this kind of flexibility, whose implications are in my opinion mostly positive. For example, an editor may be reluctant to publish a corpus of inscriptions in print unless convinced that the readings are definitive. This may in some cases result in long delays to publications despite the fact that the editions have already reached a very high standard. The digital environment allows some of these valuable results to be shared with others (with provision of DOIs for example), leaving open the possibility of an update to a record where this becomes necessary. However, potential pitfalls of a more flexible environment include: the inability to bring the editorial process to a conclusion (returning again and again to the same record, switching between possible interpretations rather than actually improving the readings); the digital publication of records that are not sufficiently precise or mature (in the absence of a balanced supervision of the platform); a reluctance to pay more attention to other important aspects, such as metadata, GIS, or the development of search engines. These considerations do not relate in any way to this very useful digital collection and are meant to be general remarks based on my own experience of epigraphic digitisation.

Although similar, the *Early Inscriptions of Āndhradeśa* corpus presents further features of interest. It is a larger repository consisting of 173 records. The careful editing and the provision of object images as well as estampages provide both reliability and transparency. Here we see the toggle options, which allow one to switch between ‘logical’ and ‘physical’ versions of the text, as well as the XML markup itself.¹⁴ Some features related to textual layout are also recorded systematically, which is particularly welcome. For example, a diamond shape (◊) indicates blank space used as a punctuation mark (Figure 4.3).¹⁵ Unfortunately the metadata does not present hyperlinks, except for bibliographical entries. Perhaps the most exciting feature is an exploratory GIS page which presents both a drawn map and a link to a QGIS cloud. Lastly, a number

Metadata

Support	Āyaka pillar; h. 275 × w. 47 × d. 36 cm.
Text	Middle Indo-Aryan, Southern Brāhmī script. h. 66 × w. 51 cm .
Date	Sixth regnal year of Siri-Virapurisadatta, approximately between 225 and 275 CE.
Origin	Erected originally at findspot.
Provenance	Discovered between March 1926 and October 1927 at Nagarjunakonda, Site 1, on the South side of the Mahācaitya. Mistakenly, Raghunath 2001 : 197 indicates Site-2 instead of Site-1. Identified at Nagarjunakonda Museum (acc. no. 285) in February 2016.
Visual Documentation	Photo(s): photos AL 2017 Photo(s) of estampage(s): estampage Leiden N8 Vogel Raghunath
Editors	Stefan Baums, Arlo Griffiths, Ingo Strauch and Vincent Tournier.
Publication history	First described and edited by Vogel 1929–30a : 13, 18-9 (B4). Re-edited here from the Leiden estampage and after autopsy of the stone.

Edition

Logical Physical XML

(1) sidhaṁ namo bhagavato devarājasakatasa supabudhabo◊dhino◊ savamñuno sava[sā] (tā) (2)nukaṁpakasa jitarāgadosamohavipamutasa mahāgaṇivasabhagaṁdhaha[thi] (sa) (3) saṁmasaṁbugasa dhātuvaparigahitasa mahācetiye kulahakānaṁ bālīkā (4) mahāsenāpatisa mahātalavarasa vāsīṭhiputasa hiraṇmākanāṁ khaṁḍacalikikaṁmaṇakasa (5) bhayā mahāsenāpatini culacātisiriṇikā °apano °ubhayalokahitasukhani(6)vāṇathanāya °imaṁ selakhaṁbhaṁ paṭīṭhapitaṁ ti raṁṇo sirivirapurisadatasa (7) ◊ sava 6 vāpa 6 diva 10

Figure 4.3: Metadata and ‘logical’ edition of inscription from Southern India. Note use of diamond-shaped symbol to indicate blank space.

¹⁴ <http://hisoma.huma-num.fr/exist/apps/EIAD/works/EIAD0008.xml?&odd=teipublisher.odd> (Accessed January 2023).
¹⁵ <http://hisoma.huma-num.fr/exist/apps/EIAD/conventions.html> (Accessed January 2023).

of Indian scholars and institutions are mentioned for their role in facilitating access to various resources and sites, although no South Asian scholar is mentioned among the main editors (Credits page).

Generally speaking, both corpora seem to have been designed primarily by philologists whose focus is to edit and interpret the inscriptions. This, to some extent, is to be expected. And yet one of the main opportunities provided by a digital environment is that the digitisation process can lead to something much more comprehensive. It is very encouraging in this regard to see that GIS was added to one of the corpora (albeit in an exploratory way), and that great care was taken into providing pictures of the objects. Some art historians were also involved in various capacities, which is a further positive sign.

Arguably, however, these very valuable repositories are not yet fundamentally different from printed editions, at least not in a way that revolutionises how these fascinating sources are approached and studied.

4. Digital epigraphy and two recent ERC Synergy projects

The digitisation of Sanskrit epigraphy has been a central component of two ERC Synergy projects: Asia-Beyond Boundaries (2014–2020) and DHARMA (2019–2025). For both projects, it is still too early to offer a comprehensive evaluation of how deeply and substantially they have changed the way we look at Sanskrit epigraphy, or the extent to which ‘synergy’ has been achieved. This term in my opinion implies both bringing together scholars who work on different corpora or regions, as well as scholars who work in different ways, transcending the disciplinary boundaries between philology, archaeology, or art history. For our present purposes, it will suffice to offer a few introductory remarks on the breadth of possible digitisation strategies, mostly based on my experience as a student collaborator of the former of these two projects.

When it comes to the digital repository of Asia Beyond Boundaries, I would like to focus on the current *Siddham.network*¹⁶ online repository, not its earlier incarnation as *Siddham*.¹⁷ The reason is that the current version offers some tentative alternative approaches to some of the issues outlined above. The earlier *Siddham* was the almost singlehanded achievement of Dániel Balogh, who—along with Arlo Griffiths—should be mentioned among the currently leading (digital) Sanskrit epigraphists (Balogh 2019). In many ways, *Siddham* continued along the lines of the corpora outlined in the preceding section. One additional feature was that a separate entry was provided for each object (assigned with a unique object identifier). Thus, text and object were given almost equal weight in the *Siddham* system. The main way of recording meta-data was through detailed spreadsheets, which could to some extent be queried

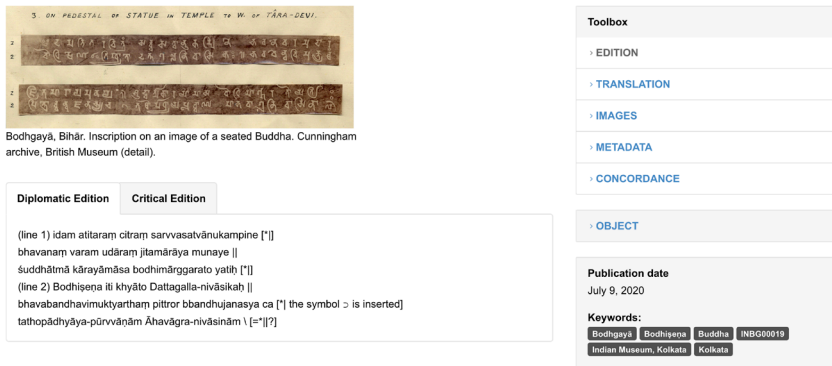
¹⁶ <https://siddham.network/> (Accessed January 2023).

¹⁷ <http://siddham.uk/> (Accessed January 2023).

(or sorted). The internal EpiDoc encoding guidelines prepared by Balogh were reasonably robust and detailed and yet also manageable and could be mastered in a relatively short period by dedicated students.¹⁸ The encoder was still able to navigate and follow the Sanskrit text of a fully marked-up edition, due to the relatively limited amount of code.

The current *Siddham.network* repository represents primarily the vision and ideas of one of the ERC project's PIs, Michael Willis (Figure 4.4).¹⁹ To put it simply, the repository hands over control to the user, who is free to create their own 'community' (a curated group of epigraphs which are mutually related, e.g. by dynasty or place) and adopt whatever editorial policies seem appropriate. Very few of the communities being developed currently use EpiDoc at all. The user might also be surprised to see a 'Google search' box on the opening page. This indicates an intention to gradually move away from expensive custom-search engines that attempt to predict what a user will input, in favour of something more dynamic, that may slowly and gradually 'learn on its own.' For example, there are a number of ways a Sanskrit term can be romanised. Therefore, expensive search engines could be designed to automatically include the most common alternative spellings. Alternatively, one could put trust in a search engine that learns over time as more user data makes itself available. This of course remains more of an aspiration than a reality at present, for epigraphists are likely to ask very specific phonological questions that require an equally precise answer.

Inscription (INBG00019) on an image of the Buddha from Bodhgaya



3. ON PEDESTAL OF STATUE IN TEMPLE TO W. OF TĀRA-DEVĪ.

2. ॐ नमो भगवते वासुदेवाय ॥ ॐ नमो भगवते वासुदेवाय ॥ ॐ नमो भगवते वासुदेवाय ॥

2. ॐ नमो भगवते वासुदेवाय ॥ ॐ नमो भगवते वासुदेवाय ॥ ॐ नमो भगवते वासुदेवाय ॥

Bodhgaya, Bihar. Inscription on an image of a seated Buddha. Cunningham archive, British Museum (detail).

Diplomatic Edition **Critical Edition**

(line 1) idam atitaram citram sarvasatvānukampine [*]
bhavanam varam udāram jītamārāya munaye ||
suddhātmā kārayāmāsa bodhimārggarato yatīḥ [*]
(line 2) Bodhiṣeṇa iti khyāto Dattagalla-nivāsikah ||
bhavabandhavimuktyartham pitror bbandhujanasya ca [*] the symbol > is inserted|
tathopādhyāya-pūrvvānām āhavāgra-nivāsinaṁ \[*]||[*]

Toolbox

- EDITION
- TRANSLATION
- IMAGES
- METADATA
- CONCORDANCE
- OBJECT

Publication date
July 9, 2020

Keywords:
Bodhgaya Bodhiṣeṇa Buddha INBG00019
Indian Museum, Kolkata Kolkata

Figure 4.4: Image, 'toolbox' and edition (not encoded in EpiDoc) of sample inscription on *Siddham.network*.

¹⁸ For an overview of markup strategies developed by epigraphists working on Sanskrit and other South and Southeast Asian languages, please refer to the DHARMA EpiDoc guidelines quoted below.

¹⁹ Willis has authored a monograph on the cultural landscape of the Gupta period, covering many key epigraphic sources (2009).

These policies mean that *Siddham.network* is particularly useful as a note-taking tool to be used while researching or even while conducting fieldwork. In short, quality will vary depending on the editor(s) of the individual communities and their standards and methods of work. Furthermore, many texts and records can still be very hard to find and query. Currently, there is no way for the external user to view all of the communities present within the repository (there are plans to achieve this through the introduction of a drop-down menu).

Be that as it may, *Siddham.network* does encourage us to ask a number of deeper questions. This can be a very useful exercise for scholars and project managers alike. Can we ever and should we ever attempt to control or predict the way people work? Should we let people take control, as long as there is a way of bringing each other's work into dialogue (for example via sets of keywords)? In the longer term, is it not more appropriate to expect AI to learn by itself, rather than marking up every single element of a sentence manually? Perhaps such approaches need not result in decreased quality and reliability, provided there are ways for scholars to interact and build on each other's efforts.

Incidentally, the approach just outlined could prove particularly useful when it comes to the heritage of countries which have experienced colonisation. There seems to be an intrinsic problem with scholars and institutions imposing a top-down approach in terms of mark-up guidelines, unique identifiers, and database structures. *Siddham.network* does not solve any of these issues but its flexible, user-based approach can at least inspire a certain form of inclusivity.

Be that as it may, how does *Siddham.network* attempt to build a database? This is mostly achieved via hyperlinks and sets of keywords, which should in time generate an underlying network.²⁰ In practice, keywords are often either too generic or too specific (e.g. the name of a remote Indian village associated with a single copper-plate inscription). Another interesting aspect is that the 'Concordance' window provides links to the PDFs of open access articles, which makes it a very valuable instrument for research. Links to other repositories can also be provided and although no proper GIS is in place, recording map coordinates is possible and encouraged.

One example of a community that has reached a good state of development is the one labelled 'Bodhgaya epigraphy', which collects inscriptions from the site associated with the Buddha's awakening.²¹ This has been prepared not by an epigraphist but by an archaeologist, Daniela De Simone.²² It is particularly encouraging to see that other professional figures can find the repository useful enough to dedicate considerable time to it. De Simone has collected not only the editions, but plenty of photographic evidence of objects (with dedicated

²⁰ <https://siddham.network/inscription/inbg00014/> (Accessed January 2023).

²¹ <https://siddham.network/community/bodhgaya-epigraphy/> (Accessed January 2023).

²² De Simone is the PI of the project *Excavations at Bodhgaya*, sponsored by the Shelby White and Leon Levy Program for Archaeological Publications (2021–2023).

pages and IDs), as well as estampages. Most of the secondary literature can now be accessed directly through the links in the concordance. Plenty of other communities have achieved significant results and can already be of great use to students and researchers alike.

The ERC Synergy project DHARMA has recently taken off and can only be mentioned in passing here. Within a comprehensive and nuanced approach to epigraphy and its archaeological contexts, it seems to place particular importance on the reliability of edited texts and the consistency of a highly developed system of EpiDoc mark-up. A large number of encoded epigraphs are already available for consultation.²³ The mark-up guidelines, co-authored by Balogh and Griffiths, have already been published and encompass more than 150 pages.²⁴ These guidelines are now the go-to resource for anyone interested to know more about the Indological contributions to EpiDoc mark-up.

One of the interesting aspects of the DHARMA project is that it employs a large number of scholars from South and Southeast Asia. The project offers opportunities for young scholars to gain professional training and academic qualifications within the EU, for example at universities in Paris. This form of partnership based on the provision of 'training' can be seen as part of the wider Western move towards trading in consultancies and knowledge-based products, given that Asian nations now increasingly have their own experts, infrastructures and cultural institutions in place.

In conclusion, the amount of resources provided for the study of Sanskrit epigraphy by these two ERC projects represents nothing less than a once-in-a-generation opportunity to change the way we think and study epigraphic texts from Southern Asia.

5. A Thai epigraphic project

In practical terms, the best results in the digitisation of inscriptions—both academically and in terms of the decolonisation of heritage—are probably achieved by fostering a variety of approaches. The repositories mentioned above are very different from one another, and that is arguably a good thing. But for the range of opportunities to be fully explored, there need to be repositories developed independently from each other, by different teams and institutions. This is particularly true of repositories designed and run in Asian countries. For example, Indian scholars have taken major steps towards the digitisation of an immense

²³ The DHARMA project team has recently made available a significant number of digital editions, meticulously edited and often accompanied by English and/or French translations: <https://erc-dharma.github.io/#tfc-collection> (Accessed July 2023).

²⁴ <https://hal.archives-ouvertes.fr/DHARMA/halshs-02888186v1> (Accessed January 2023).

cultural heritage. The focus has been on manuscripts and art-historical artefacts, although some epigraphic sources can already be consulted too. The Museums of India repository allows one to view object images as well as basic metadata.²⁵

A fascinating Thailand-based epigraphic database is the “Inscriptions in Thailand” project by Princess Maha Chakri Sirindhorn Anthropology Center, built in collaboration with Silpakorn University, Bangkok.²⁶ The website calculates the number of available records at 2456. Opening sample records, one finds basic metadata, alongside estampages and photos of the objects (Figure 4.5). There is some interlinked data, for example according to scripts, places and languages. Editions and translations are provided—although sometimes only in Thai script and Thai language—and can be downloaded as PDF (EpiDoc encoding is generally not employed by this particular project). While the website interface is available both in Thai and English, the reliance on Thai for deciphering the records sends a strong signal in terms of the repository being primarily meant for a local readership. One fascinating element, which one may take as decolonising, is the systematic use of the Buddhist Era in the metadata, instead of the Christian/Common one. Also,

The screenshot displays the website for the Princess Maha Chakri Sirindhorn Anthropology Centre. The header includes the center's name in Thai and English. The main content area is titled 'จารึก' (Inscription) and 'จารึกหมายเลข 2 บนสุรูปศิลา' (Inscription No. 2 on a stone stupa). It features a grid of images showing a stone stupa and close-ups of inscriptions in Thai script. Below the images, there is a section for 'คำอ่าน-แปล' (Transcription and Translation) with a PDF download icon. To the right, a detailed metadata section provides information about the inscription, including its location, date, and the center's contact details.

จารึก
THE PRINCESS MAHA CHAKRI SIRINDHORN ANTHROPOLOGY CENTRE

จารึกหมายเลข 2 บนสุรูปศิลา < ย้อนกลับ

จารึก

จารึกหมายเลข 2 บนสุรูปศิลา

โพสท์เมื่อวันที่ 13 ก.พ. 2550 13:59:58 (อัปเดตเมื่อวันที่ 22 เม.ย. 2566 12:59:24)

ชื่อจารึก จารึกหมายเลข 2 บนสุรูปศิลา

ชื่อจารึกแบบอื่นๆ หลักที่ 32 จารึกหมายเลข 2 บนสุรูปศิลา, นว. 3

อักษรที่เขียนจารึก ปัลลวะ

ศิลา หินทราย

ภาษา บาลี

จำนวนด้าน 1 ด้าน มี 1 บรรทัด

วัตถุจารึก ศิลาจารึก

ลักษณะวัตถุ สรูปบนฐานศิลา

ขนาดวัตถุ สูง 390 ซม. เส้นผ่าศูนย์กลาง 130 ซม.

บัญชี/ทะเบียนวัตถุ 1) กองหอสมุดแห่งชาติ กำหนดเป็น "นว. 3"

2) ในหนังสือ ปะฐุมพินิจจารึก ภาคที่ 3 กำหนดเป็น "หลักที่ 32 จารึกหมายเลข 2 บนสุรูปศิลา"

3) ในหนังสือ จารึกในประเทศไทย เล่ม 1 กำหนดเป็น "จารึกบนสุรูปหมายเลข 2 บนสุรูปศิลา"

ปีที่พบจารึก ประมาณ พ.ศ. 2470

สถานที่พบ บริเวณพระปฐมเจดีย์ วัดพระปฐมเจดีย์ ตำบลพระปฐมเจดีย์ อำเภอเมือง จังหวัดนครปฐม

ผู้พบ ไม่ปรากฏหลักฐาน

ปัจจุบันอยู่ที่ พิพิธภัณฑสถานแห่งชาติ พระนคร กรุงเทพมหานคร ถนนหน้าพระธาตุ แขวงพระบรมมหาราชวัง เขตพระนคร กรุงเทพมหานคร (สำรวจเมื่อ 31 ตุลาคม 2564)

คำอ่าน-แปล

ด้านที่ 1

Figure 4.5: Metadata and photographic samples of inscription on the Inscriptions in Thailand database.

²⁵ http://museumsofindia.gov.in/repository/record/nat_del-92-53-29127 (Accessed January 2023).

²⁶ <https://db.sac.or.th/inscriptions/> (Accessed January 2023).

it appears that only occasional use is made of the French system of cataloguing Southeast Asian and specifically Khmer inscriptions. While many Western scholars would be quite willing to adopt new cataloguing guidelines set in place in collaboration with Asian colleagues and their institutions, at the same time there is a need to avoid unnecessary confusion. How a new system might work in practice remains obscure, although it would certainly require considerable international cooperation (Khmer/Angkorian inscriptions alone have been found across a number of modern nations, including Cambodia, Thailand, and Laos). As a regular user of the database, I found dealing with Thai scripts, the Buddhist Era and the lack of familiar inscriptions IDs rather challenging.

Another valuable aspect of the Anthropology Center's repository is that it provides a map interface which is searchable.²⁷ For example, one can search for the Sanskrit inscriptions in the Northeastern Thai province of Buriram, after which one currently gets three hits which are pin-pointed on the map and can be then browsed on separate windows. The editions themselves are currently being improved and published in a separate series edited by Thai epigraphist Ajahn Sombat Mangmeesuksiri at Silpakorn University.

6. Database, query, analysis

So far I have primarily examined website interfaces, encoding, and inclusivity within digital projects. However, one aspect which none of the above repositories has fully mastered so far is in a way the most basic one: the database itself. If we are to move beyond printed editions in a true 'epistemic turn', the database itself would need to be at the centre of attention, rather than the edited text. Keywords and (hyper)links are certainly useful, and so are edition-image interactions and powerful GIS. However, one should also focus on the range of search options. For example, a particularly useful digital tool available to Sanskritists is Harry Falk's Indoskript (Figure 4.6).²⁸ This is essentially a palaeographic database that allows one to search for a certain letter (or, more precisely, a syllable graph) and view how it appears on a variety of manuscripts and epigraphs over various centuries. The search function allows one to narrow it down to certain centuries and to select certain areas of the map. For example, one can search for the syllable 'ka' from 300 BCE to 100 BCE, focusing on witnesses from Western India. This is a simple and yet powerful way of making data accessible. Ideally, the records digitised through the repositories mentioned above, should also be searchable palaeographically or at least integrated with Indoskript. It may not be enough to offer an image so that the edition can be checked against it. Various parts of that image could be actively integrated within the database itself.²⁹

²⁷ <https://db.sac.or.th/inscriptions/map> (Accessed January 2023).

²⁸ <http://www.indoskript.org/letters> (Accessed January 2023).

²⁹ An attempt can be found in Bianchini 2023, based on a seventh century Maitraka inscription from Gujarat. On the vast epigraphical corpus of the Maitrakas see Schmiedchen 2018.

Letters

Transliteration (% = wildcard) Lower date: Upper date: Collections: Historical persons: Historical places:

ka -300 2000

Manuscript ID: Longitude (left): Longitude (right): Latitude (top): Latitude (bottom):

Search

1 2 3 4 5 6 7 8 9 ... 23

Figure 4.6: Partial overview of letter ‘ka’ in Western India, date range 300 BCE–200 CE.

SEALang :: SEAClassics

Indic entry a e i o r t d h m n

Corpus search Go! Clear all

Search V+ C+ restricted clear

context +/- 5 order by date

Collocates Neighbors sum don't sum

Summary #distribution

Search V+ C+ CV+

Corpus restriction / display Clear all

Display Text(s) Kwc(s) Both selected below

Extract lexicon order by alphabet

K.num	Reign(s) / era(s) / years (contiguous)	Site(s)	Region(s)
1	Unknown		
6	pre-Angkor (to 876)		
7	Angkor (877-1380)		
8	post-Angkor (1381 on)		
9			
11			
18			

Language(s)	Script(s)
Cambodian	Khmer
Thailand	Sanskrit
Laos	Pali
Vietnam	Thai

Highlight items Unique in corpus

Corpus of Khmer Inscriptions

Note that this search has been **restricted** to selected texts (click **clear all** to removed these restrictions). Restrictions are: Inscription(s): 1.

1 item

Inscription 1 (site: Vat THLEN, area: CHAUDOC, Saka year: 6-7C)

(1) vā ta śivadeva saraṁ ta kurāk kandiṣ cap vā kandos 1 ku tai dau
(2) jvan ta mratāñ klon jeshapura kalla klon bhavapura 'atmay
(3) 'ācāryya isānadatta dau ka dhāra ge ta mratāñ ge tel poñ
(4) kumārāsānti ta kloñ ge ptā ta 'nak poñ pre 'ācāryya isāna-
(5) -datta loh ge man supratīṣṭha vrah karmatāñ 'añ śrīśānka-
(6) -rañārāyana poñ oy ge ta vrah vñāk nuḥ upakalpa mratāñ
(7) kloñ bhavapura pre tān spun tān bho 'ācāryya śilābhadrā 'mac
(8) vñāk kañlon kurāk kandiṣ pras ni poñ śivadeva kamṁoy kamṁon
(9) poñ kumārāsānti pantiñ ket mratāñ kloñ bhavapura pre mok
(10) oy ge ta vrah karmatāñ 'añ śrīśānkanārāyana ai cmōñ
(11) vā kandos 1 ku tai 1 kon ku vā jloy 1 ku jūñ thāñ 1 vā so 1 ku 'antis 1 ku kam'el
(12) vā trok 1 vā jiva 1 ku dhuli 1 vā vinita 1
(13) sre ralon ruñ 'amvi thai luc pu yoi loh tem cikan 'amvi
(14) rāñ travan tān dallan loh thai luc travan isāvakumāra ti le kyel
(15) gui ralon ver thai ket travan tān dal man loh ta poñ pravara....
(16) 'amvi tem saññāyana loh ta sre vrah kloñ vā candravira 'amvi ...
(17) ...vā tān dal rāñ plu loh travan ver rāñ plu pi....
(18) ... kyel guha loh travan tān
(19) travan kvah 'añ sin loh travan poñ vidyābhadrā d....
(20) ai karom kyel gui pāk ramlon vera ple pi katje
(21) sre ai tel thai ket tem kadam ralon piyy le kyel
(22) sre 'amvi tem ransi... pi sre man kloñ jlen lak ta vrah loh
(23) kloñ gann loh travan poñ śiladeva pramoṣ mās 'amc....
(24) 'amvi travan vā tvellan dan gui ralon kandiñ 'añ ...thail sam lo...
(25) rāñ gui sre kloñ ci mratāñ thai ket gui kyoy tvākk damṁon loh gui
(26)ṭ sare 1....

Figure 4.7: The search engine of SEALang, Corpus of Khmer Inscriptions.

Another example are the records on SEALang Classics, particularly the *Corpus of Khmer Inscriptions*, based on the extensive work on Old Khmer by Philip Jenner.³⁰ Much effort seems to have gone into designing the search engine here, although primarily for philological purposes. Apart from a very precise way of searching for text strings, the site offers the ability to narrow it down by region, site, time period, reigning monarch, language, and script (Figure 4.7). It should be feasible for repositories of Asian inscriptions to offer at least this much in terms of querying possibilities as well as further options.

³⁰ <http://sealang.net/classic/khmer/> (Accessed January 2023).

7. Conclusion

To sum up, the digitisation of inscriptions from Southern Asia is an ongoing and complex process, although some very valuable results have already been achieved. The availability of readily accessible, reliable, and open-access digital records represents an advantage to both scholars and students of South and Southeast Asia. While one should keep in mind the contingencies and requirements of funding bodies, as well as realistically attainable goals, a fascinating question is to what extent digitisation practices actually reach beyond traditional printed editions, opening the door to new ways of thinking. In this regard, I think it is crucial to gauge how the multidimensionality of artefacts is represented—i.e. if text, objects, and archaeological landscapes are integrated—and thus the extent to which collaborative interdisciplinarity is achieved. Powerful search engines and GIS tools that explore the whole scale of epigraph multidimensionality are further desirable components of an ideal database that could potentially revolutionise research.

Last but not least, a plurality of voices, methods and approaches can help ensure that we do not fall back on unequal ways of engaging with Asian cultural heritage. Multiple databases, no matter how different, can still link to each other and need not be isolated. Ultimately, the digitisation of Asian inscriptions should be a way to celebrate our shared passion for history and historiography, enhancing it with the valuable tools offered by modern-day technology.

8. References

- Balogh, D.** (2019). *Inscriptions of the Aulikaras and their associates*. Berlin: De Gruyter.
- Balogh, D., & Griffiths, A.** (2020). DHARMA Encoding Guide for Diplomatic Editions. [Technical Report] EFEO; Humboldt-Universität (Berlin); CEAIS – Centre d'Études de l'Inde et de l'Asie du Sud. Retrieved from: <https://hal.archives-ouvertes.fr/DHARMA/halshs-02888186v1>.
- Bianchini, F.** (2023). Legally binding: the textual layout of a copper-plate grant from South Asia. *Manuscript and Text Cultures* 2(1). <https://doi.org/10.56004/v2.1fb>.
- Cambridge Digital Library – Sanskrit Manuscripts.** (2015). Retrieved from: <https://cudl.lib.cam.ac.uk/collections/sanskrit/1>.
- Corpus of the Inscriptions of Campā.** (2012). Retrieved from: <https://isaw.nyu.edu/publications/inscriptions/campa/index.html>.
- ERC DHARMA.** (2020). Github Repository. Retrieved from: <https://github.com/erc-dharma>.
- Early Inscriptions of Āndhradeśa.** (2017). Retrieved from: <http://hisoma.huma-num.fr/exist/apps/EIAD/index2.html>.

- Francis, E.** (2018). Indian Copper-Plate Grants: Inscriptions or Documents? In: A Bausi, C Brockmann, M Friedrich and S Kienitz (Eds.), *Manuscripts and Archives: Comparative Views on Record-Keeping*. (Studies in Manuscript Cultures). Berlin: De Gruyter, 387–417.
- Griffiths, A., & Lammerts, Ch.** (2015). Epigraphy: Southeast Asia. In J. Silk, O. von Hinüber & V. Eltschinger (Eds.), *Brill's Encyclopaedia of Buddhism: Volume One, Literature and Languages* (pp. 988–1009). Leiden: Brill.
- Indoskript 2.0.** (2017). URL: <http://www.indoskript.org/>.
- Museums of India. National Portal and Digital Repository.** (2020). Retrieved from: <https://museumsofindia.gov.in/repository/home>.
- Salomon, R.** (1998). *Indian Epigraphy: A Guide to the Study of Inscriptions in Sanskrit, Prakrit, and the Other Indo-Aryan Languages*. New York; Oxford: Oxford University Press.
- Schmiedchen, A.** (2018). Kings, authors and messengers: The composition of the Maitraka copper-plate charters. In B. Shelat and T. Parmar (eds), *New horizons in Indology. Prof. Dr. H.G. Shastri commemoration volume*. Ahmedabad: Shri Nandan H. Shastri, 35–41.
- SEAlang.** SEAcassics: Corpus of Khmer Inscriptions. Retrieved from: <http://sealang.net/classic/khmer/>.
- Siddham.** The Asia Inscriptions Database. Retrieved from: <https://siddham.network/>.
- The Inscriptions in Thailand Database.** (2019). The Princess Maha Chakri Sirindhorn Anthropology Centre. Retrieved from: <https://db.sac.or.th/inscriptions/>.
- Willis, M.** (2009). *The Archaeology of Hindu Ritual : Temples and the Establishment of the Gods*. New York; Cambridge: Cambridge University Press.