

# Perspectives of Digital Humanities in the Field of Buddhist Studies

## Workshop & Symposium (Hybrid)

12 January & 13–14 January 2023

Buddhist Studies have been inspired by the development of Digital Humanities over several decades. While some tools reached maturity about twenty years ago, other tools have been developed very recently and the future offers even more exciting perspectives. This symposium brings together pioneers and newcomers of the world of Buddhist digital humanities in order to give insights into their projects and experiences. While the event is aimed at promoting a professional exchange, it also seeks to build a bridge between the developers and the users of these tools. For this purpose, an accompanying workshop will consist of introductory units on tools and methods like TACL, Social Network Analysis, and BuddhaNexus. The symposium-cum-workshop is a joint event by the Numata Center for Buddhist Studies and the Khyentse Center for Tibetan Buddhist Textual Scholarship at Hamburg University.

**Registration** for the Workshop Part I, II, III and for the Symposium: Please send an e-mail to Carsten Krause ([carsten.krause@uni-hamburg.de](mailto:carsten.krause@uni-hamburg.de)) before 5 January 2023.

More information at: [www.buddhismuskunde.uni-hamburg.de](http://www.buddhismuskunde.uni-hamburg.de)

### Venue

Hamburg University

### Funded by

Numata Center for Buddhist Studies, Hamburg University  
Khyentse Center for Tibetan Buddhist Textual Scholarship,  
Hamburg University

### Convenors

Carsten Krause, Sebastian Nehrdich

### Speakers

#### Workshop

Orna Almogi (Hamburg University)

Marcus Bingenheimer (Temple University)

Sebastian Nehrdich (Düsseldorf University, Hamburg University)

Michael Radich (Heidelberg University)

#### Symposium

Orna Almogi (Hamburg University)

Marcus Bingenheimer (Temple University)

Laurent Van Cutsem (Ghent University)

Christian Faggionato (University of Cambridge)

Oliver Hellwig (Düsseldorf University)

Leo Maximilian Koenig (Hamburg University)

Carsten Krause (Hamburg University)

Charles Li (French National Centre for Scientific Research, EHESS)

Patrick McAllister (Austrian Academy of Sciences)

Kiyoinori Nagasaki (International Institute for Digital Humanities, Tokyo)

Sebastian Nehrdich (Düsseldorf University, Hamburg University)

Alexander James O'Neill (SOAS, University of London)

Jörg Plassen (Bochum University)

Michael Radich (Heidelberg University)

Élie Roux (Buddhist Digital Resource Center)

Gregory Adam Scott (University of Manchester)

Sven Sellmer (University of Poznań, Düsseldorf University)

Leon Woltermann (Hamburg University)

# Perspectives of Digital Humanities in the Field of Buddhist Studies

## Workshop, Day 1 (Thursday, 12 January 2023)

Place: Akademischer Senatssaal, Edmund-Siemers-Allee 1, Main Building, 20146 Hamburg

Part I: 9:30–12:30 (including coffee breaks)

### “Using the Historical Social Network of Chinese Buddhism in Gephi”

Marcus Bingenheimer  
(Temple University)

12:30–14:00

*Lunch Break*

Part II: 14:00–17:00 (including coffee breaks)

### “The BuddhaNexus Platform”

Sebastian Nehrdich  
(Hamburg University, Düsseldorf University)  
Orna Almogi  
(Hamburg University)

17:00–17:30

*Coffee Break*

Part III: 17:30–19:00

### “TACL: A Tool for the Comparative Large-scale Analysis of Texts in the Chinese Buddhist Canon”

Michael Radich  
(Heidelberg University)

---

**The workshops are free of charge.** When registering via e-mail ([carsten.krause@uni-hamburg.de](mailto:carsten.krause@uni-hamburg.de)), before 5 January 2023, please let us know which parts you are going to attend – offline or online.

# Perspectives of Digital Humanities in the Field of Buddhist Studies

## Symposium, Day 2 (Friday, 13 January 2023)

Place: Room 121, Edmund-Siemers-Allee 1, West Wing, 20146 Hamburg

9:15–9:30

*Introduction*

9:30–10:00

**“Multilingual Semantic Mining for Text Alignment and Parallel Corpus Building for Buddhist Languages”**

Sebastian Nehrdich

(Düsseldorf University, Hamburg University)

10:00–10:30

**“BuddhaNexus: Underlying Philosophy, Hitherto Achievements, Future Plans”**

Orna Almogi

(Hamburg University)

10:30–11:00

**“Digital Humanities at the Buddhist Digital Resource Center”**

Élie Roux

(Buddhist Digital Resource Center)

11:00–11:30

*Coffee Break*

11:30–12:00

**“Tracing the Impact of Buddhist Commentaries from Silla on Tang Dynasty Huayan: Towards Computer-assisted Analyses of Hidden Text Reuse”**

Jörg Plassen

(Bochum University)

12:00–12:30

**“Research Results Achieved by Applying the TACL Toolkit”**

Michael Radich

(Heidelberg University)

12:30–14:00

*Lunch Break*

# Perspectives of Digital Humanities in the Field of Buddhist Studies

## Symposium, Day 2 (Friday, 13 January 2023)

Place: Room 121, Edmund-Siemers-Allee 1, West Wing, 20146 Hamburg

14:00–14:30

### “On the Use of Historical Social Network Analysis in the Study of Chinese Buddhism”

Marcus Bingenheimer  
(Temple University)

14:30–15:00

### “Lineages as Network: An Inquiry into the Representation of *dharma* Lineages of Chan Buddhism in the *Zutang ji* 祖堂集”

Laurent Van Cutsem  
(Ghent University)

15:00–15:30

### “Cataloguing Chinese Buddhist Publications in the Chinese Religious Text Authority”

Gregory Adam Scott  
(University of Manchester)

15:30–16:00

*Coffee Break*

16:00–16:30

### “Using *n-akṣaras* to Model Sanskrit and Sanskrit-adjacent Texts”

Charles Li  
(French National Centre for Scientific Research, EHESS)

16:30–17:00

### “Utilization for Digital Images for Buddhist Studies”

Kiyoinori Nagasaki  
(International Institute for Digital Humanities, Tokyo)

17:00–17:30

### “What Can We Learn from DH Tools and Methods for Buddhist Studies in the Contemporary Field?”

Carsten Krause  
Leo Maximilian Koenig  
Leon Woltermann  
(Hamburg University)

18:30

*Dinner*

# Perspectives of Digital Humanities in the Field of Buddhist Studies

## Symposium, Day 3 (Saturday, 14 January 2023)

Place: Room 121, Edmund-Siemers-Allee 1, West Wing, 20146 Hamburg

9:15–9:45

### “ChronBMM – Dating Text Corpora Using Bayesian Mixture Models”

Oliver Hellwig

(Düsseldorf University)

Sven Sellmer

(University of Poznań, Düsseldorf University)

Sebastian Nehrdich

(Düsseldorf University, Hamburg University)

9:45–10:45

### “NLP Pipeline for Diachronic Tibetan and Newar Corpus Analysis”

Alexander James O’Neill

(University of London)

Christian Faggionato

(University of Cambridge)

10:45–11:15

*Coffee Break*

11:15–11:45

### “Retro-digital Editions of Sanskrit Texts: Many Problems, Some Solutions”

Patrick McAllister

(Austrian Academy of Sciences)

11:45–12:15

### “Data Design for a Navigation System in the Tibetan Buddhist Canon”

Bruno Lainé

(University of Vienna)

12:15–13:00

*Closing Discussion at Lunch Break*

## Abstracts: Workshop

Marcus Bingenheimer  
(Temple University)

### “Using the Historical Social Network of Chinese Buddhism in Gephi”

The first half of the workshop will introduce Gephi. Gephi ([gephi.org](http://gephi.org)) is a popular open source network analysis tool widely used in the Digital Humanities. (Please download before coming and test it as the workshop will not cover installation.)

In the second half of the workshop we will explore the Historical Network of Chinese Buddhism with the help of Gephi. The data for the Historical Network of Chinese Buddhism is at: [https://github.com/mbingenheimer/ChineseBuddhism\\_SNA](https://github.com/mbingenheimer/ChineseBuddhism_SNA).

The data contains more than 18K people that appear in Chinese Buddhist sources and their relationships, spanning the time from c.150 to c.1900 CE. With it it is now possible to visualize and analyze social relationships in Chinese Buddhism from the ego-network level of an individual, to larger dynamics across generations. After the workshop attendees will know how to find their way around Gephi and use the Historical Network of Chinese Buddhism for their research. They will be able to identify persons in the network, understand how to trace connections back to primary sources, visualize ego-networks, and enrich the dataset with their own data.

Sebastian Nehrdich  
(Düsseldorf University, Universität Hamburg)  
Orna Almogi  
(Hamburg University)

### “The BuddhaNexus Platform”

BuddhaNexus is a database devoted to the study of Buddhist texts and literary corpora in Pāli, Sanskrit, Tibetan, and Chinese, with particular emphasis on evolution of scriptures, formation of canons, and intellectual networks. It offers new ways to study the various stages in the evolution of a specific scripture, intertextuality between various scriptures, the linkage between treatises and scriptures, and possible intellectual networks, by locating in the various texts and textual corpora (approximate) textual matches, be them acknowledged citations, “borrowed” textual passages, or parallel passages. We will introduce the database and demonstrate its current abilities by presenting examples for its application on specific research questions.

## Abstracts: Workshop

Michael Radich  
(Heidelberg University)

### **“TACL: A Tool for the Comparative Large-scale Analysis of Texts in the Chinese Buddhist Canon”**

This session gives a hands-on introduction to a suite of digital tools for the study of the Chinese Buddhist canon: TACL, which helps users analyze texts to discover various types of intertextual relations, including sources, borrowing, and distinctive style. Participation in this session will be made smoother if participants go to <https://dazangthings.nz/tacl-gui-one-stop-shop/> before the session, and download the “TACL GUI starter kit”.

## Abstracts: Symposium

Sebastian Nehrdich  
(Düsseldorf University, Hamburg University)

### “Multilingual Semantic Mining for Text Alignment and Parallel Corpus Building for Buddhist Languages”

Multilingual contextual word embedding models have taken the world of Natural Language Processing (NLP) in the past years by storm. Usually trained on terabytes of data, they achieve impressive performance on downstream NLP tasks such as bi-text retrieval while at the same time drastically reducing the amount of necessary aligned training data. Since considerable amounts of Buddhist texts preserved in Chinese and Tibetan are translations from other languages, as well as the fact that an increasing number of ancient Buddhist texts is being translated into modern languages such as English, Japanese and Modern Chinese, the construction of comprehensive parallel corpora of these languages is an important step to facilitate their research as well as the development of NLP applications for tasks such as machine translation. I will discuss the adaptation and training of large multilingual languages models for bitext mining on Buddhist corpora, how the resulting data is presented for philological research in the BuddhaNexus database and how machine translation systems perform when being trained on this data.

Orna Almogi  
(Hamburg University)

### “BuddhaNexus: Underlying Philosophy, Hitherto Achievements, Future Plans”

BuddhaNexus aims at facilitating the study of Buddhist texts in Pāli, Sanskrit, Classical Tibetan, and Classical (Buddhist) Chinese, particularly in view of the history of their composition and their impact on later texts. Towards this end, it identifies approximate textual matches between texts written in the same language. Altogether 220,000,000 such matches have been located, revealing a very high level of intertextuality between the texts. In my talk, I shall outline the underlying philosophy of the project, the new research dimensions it opens up for Buddhist textual scholars and the challenges it brings along. Finally, I shall also delineate the plans for future developments.

## Abstracts: Symposium

Élie Roux

(Buddhist Digital Resource Center)

### “Digital Humanities at the Buddhist Digital Resource Center”

BUDA was built from the ground up without reference to any earlier software. In developing BUDA we focused on creating new ways to access, discover and reuse the Buddhist Digital Resource Center's (BDRC) large collection through state of the art data technology.

A key feature is the new ability to access and reuse BDRC's 20+ million images through IIIF. Our bibliographical and prosopographical database is now openly accessible through a standard data format, RDF. We will explain the new model of the database and how to access and query it. Etexts have also been completely reworked and standardized through a new collaborative project, OpenPecha. OpenPecha provides a stand-off annotation format with a suite of tools, with the ambition of becoming a new standard in digital humanities.

Finally, we will show how BUDA federates catalogs and images from multiple sources and indicates provenance and attribution. We hope that through a set of common standards and a focus on open access with attribution, BDRC can become a digital Buddhist commons where data and images related to Buddhist studies can be distributed and accessed.

Jörg Plassen

(Bochum University)

### “Tracing the Impact of Buddhist Commentaries from Silla on Tang Dynasty Huayan: Towards Computer-assisted Analyses of Hidden Text Reuse”

Korean and Japanese scholars have produced extensive collections of quotations from early Korean Buddhist texts in Chinese and Japanese sources solely based on open citations. However, conventional research on textual influences supports the expectation that systematically uncovering hidden text reuse will yield valuable additional material that will testify to an even more pervasive influence of Silla texts on “Chinese Buddhism”.

Drawing on generous support by Volkswagen Stiftung and Bochum University for our Hanmun Lab, over the next four years we will carry out an in-depth study on text reuse of Hwaōm Buddhist texts from Silla within Tang period (617–906) commentaries, identifying parameters of text reuse through philological analysis, molding these insights into extensions to a C++/CUDA program as well as SQL and R scripts, and synthesizing the results into a preliminary survey on the impact of the Silla texts on Tang Huayan.

Taking the influence of Wōnhyo 元曉 (617–686) on Fazang 法藏 (643–712) and Chengguan 澄觀 (738–839) as an example, the talk will provide a glimpse into what is possible already at this stage.

## Abstracts: Symposium

Michael Radich  
(Heidelberg University)

### “Research Results Acheived by Applying the TACL Toolkit”

In this talk, I will present in brief a number of research findings that have been arrived at (in part) by the application of the TACL suite of software tools to problems in textual analysis and intertextuality. I will focus on a selection of case studies that showcases a range of applications and research questions.

Marcus Bingenheimer  
(Temple University)

### “On the Use of Historical Social Network Analysis in the Study of Chinese Buddhism”

Can historical network analysis contribute to our understanding of Chinese Buddhism? The talk will describe some of the methods and data currently available for the application of formal network analysis. "Formal network analysis" in this context denotes the analysis and visualization of clearly defined, computable network data. After an introduction to network analysis as a method, the presentation will describe its application with the help of one study. A closer look at the network region modeling the late 16th early 17th century reveals two distinct stages in the late Ming Buddhist revival. Following the Wanli revival brought about by a group of well-studied, famous monks, the network perspective asserts the centrality of Miyun Yuanwu 密雲圓悟 (1567–1642) and his students who came to dominate 17th century East Asian Buddhism. It can be argued that the Buddhism of Miyun's lineage differed considerably from the syncretic, inclusive Buddhist discourse of the "great monks" of the Wanli era. This important change in the history of Buddhism is visible in the network.

## Abstracts: Symposium

Laurent Van Cutsem  
(Ghent University)

### “Lineages as Network: An Inquiry into the Representation of *dharma* Lineages of Chan Buddhism in the *Zutang ji* 祖堂集”

This paper is a first attempt at examining the genealogical framework of “lamp records” (*denglu* 燈錄) of the Chan Buddhist tradition as networks, using analytical tools and methods of Historical Social Network Analysis (HSNA) and graph theory. As an exploratory study, the primary objectives are to investigate the possibilities offered by HSNA for research on the socio-religious milieu of production of Chan records, explore the advantages of this approach over classic lineage charts, and reflect on its limitations in contrast to more philologically oriented studies. The paper focuses on the Chan community portrayed in the *Zutang ji* 祖堂集 (Collection of the Hall of Patriarchs). It shows that a relatively marked predominance is given to the branch stemming from Xingsi 行思 and Shitou 石頭, both from the perspective of the genealogical network and the space allotted in the text. The study also highlights possible irregularities in lineage claims by contrasting metrics of degree and betweenness centrality with complementary data (e.g., length of the entries).

Gregory Adam Scott  
(University of Manchester)

### “Cataloguing Chinese Buddhist Publications in the Chinese Religious Text Authority”

The Chinese Religious Text Authority (CRTA) was founded in late 2019 as an online, open-content repository for bibliographic and scholarly data on religious texts produced in China prior to 1949. It runs on Mediawiki with a database plug-in to allow the data to be queried in the style of a library catalogue, and exported for use by other researchers and projects.

Among the more than 3000 current entries, approximately 2500 are Buddhist publications from the late nineteenth and early twentieth centuries, including scriptural texts, monographs, and periodicals. These were produced by scriptural presses, small Buddhist publishers, and larger commercial presses. Many of the familiar monastic and lay figures from this era of Chinese Buddhism, including Taixu, Yinguang, Wang Yiting, and Ding Fubao, are represented in the catalogue as authors and editors.

My presentation will explore how I am gathering, editing, formatting, and presenting bibliographic data on Chinese Buddhist publications from the modern era, and how such a dataset may contribute to research in Buddhist Studies. It will examine how best to incorporate such data production into a research project that will result in articles and a monograph. My aim is to provide some guidance for others who might want to incorporate such an approach into their own research.

## Abstracts: Symposium

Charles Li

(French National Centre for Scientific Research, EHESS)

### “Using n-akṣaras to model Sanskrit and Sanskrit-adjacent texts”

Despite – or perhaps because of – their simplicity, n-grams, or contiguous sequences of tokens, have been used with great success in computational linguistics since their introduction in the late 20th century. Recast as k-mers, or contiguous sequences of monomers, they have also found applications in computational biology. When applied to the analysis of texts, n-grams usually take the form of sequences of words. But if we try to apply this model to the analysis of Sanskrit texts, we are faced with the arduous task of, firstly, resolving sandhi to split a phrase into words, and, secondly, splitting long compounds into their components. This paper will present a simpler method of modelling a Sanskrit text as n-grams, by using n-akṣaras, or contiguous sequences of akṣaras. This model reduces the need for sandhi resolution, making it much easier to use on raw text. We will present a number of applications of this model, from text mining and sequence alignment (collation) to inferring a phylogenetic network of interrelated Sanskrit and Sanskrit-adjacent (Hindi, Newari, Tamil, etc.) texts.

Kiyoinori Nagasaki

(International Institute for Digital Humanities, Tokyo)

### “Utilization for Digital Images for Buddhist Studies”

Nowadays, digital images are available on the Web from many cultural institutions around the world, which include many Buddhist scriptures and paintings. How to utilize these images in Buddhist studies has become an important issue. Recently, the number of IIF-compliant images and TEI-compliant bibliography and texts has been increasing, and they can be used in various intensive ways through Web APIs. This presentation will introduce some methods for aggregating digital images from around the world in an easily accessible form and utilizing them for Buddhist studies with the IIF specification and TEI Guidelines.

## Abstracts: Symposium

Carsten Krause  
Leo Maximilian Koenig  
Leon Woltermann  
(Hamburg University)

### “What Can We Learn from DH Tools and Methods for Buddhist Studies in the Contemporary Field?”

In Buddhist studies, many new DH tools and methods deal with the question of how to process canonical texts that have been collected over several centuries. They reveal questions of authorship and other personal data, text changes, adaptations and omissions of the past. But what can we learn from this for Buddhist studies in the contemporary field where canonization processes are still ongoing and internet-based textual material is subject to permanent and dynamic change? This presentation discusses an ongoing work to develop a personal database of prominent Buddhists in China today. While the project aims at a creative combination of openly available data from the Internet (Baidu, Wikipedia, etc.), it is accompanied by questions worthy of discussion, some of which can potentially be solved using existing tools and experiences, others involve unforeseen challenges and strive to promote interdisciplinary exchange.

Oliver Hellwig  
(Düsseldorf University)

Sven Sellmer  
(University of Poznań, Düsseldorf University)

Sebastian Nehrdich  
(Hamburg University, Düsseldorf University)

### “ChronBMM – Dating Text Corpora Using Bayesian Mixture Models”

Corpus linguistic studies are often based on the assumption that the temporal structure of the corpus under consideration is known. While such an assumption holds true for important texts of many European and East Asian corpora, the dates of pre-modern Indian texts that were (and are) proposed in scholarly research often vary by several centuries. The project ChronBMM develops probabilistic Bayesian models that use linguistic features in order to narrow down the temporal range of important Vedic texts and thus provide future research with a more reliable temporal structure of the Vedic corpus. Although ChronBMM concentrates on Vedic texts, the models developed are generic and can therefore be applied to comparable cases such as Classical Sanskrit and Tibetan.

## Abstracts: Symposium

Alexander James O'Neill  
(University of London)

Christian Faggionato  
(University of Cambridge)

### “NLP Pipeline for Diachronic Tibetan and Newar Corpus Analysis”

Our AHRC funded project is the first comprehensive diachronic investigation of egophoric marking, the marking of the conscious self. Through state-of-the-art diachronic corpus methods, we will trace the lexical origins of egophoric markers and compare Tibetan and Newar developments side by side.

In the present talk, we will present a fully implemented NLP pipeline to create deeply-annotated corpora of several historical and contemporary Tibetan and Newar varieties with five different layers of annotation, ranging from morphosyntactic to information-structural annotation. We will describe in detail our annotation process – both for Tibetan and Newar – and the three phases of our NLP annotation pipeline consisting of: (1) pre-processing – normalisation and standardisation, (2) word segmentation, POS tagging and sentence segmentation and (3) parsing and information structure annotation. Phases (2) and (3) have an additional step of manual annotation to maintain a gold standard. In the end, we will discuss how to adapt our NLP-tools to other Tibetan and Newar varieties that are both low-resourced and under-researched from an NLP perspective.

Patrick McAllister  
(Austrian Academy of Sciences)

### “Retro-digital Editions of Sanskrit Texts: Many Problems, Some Solutions”

The press of the Austrian Academy of Sciences publishes the series “Sanskrit Texts from the Tibetan Autonomous Region” (TAR). This series, currently standing at around twenty volumes, is focused on editions of Sanskrit works that have survived in manuscripts in the TAR since at least the thirteenth century CE. The volumes published so far present only a small subset of the works available in these manuscripts.

I will present an ongoing effort to switch to a realistic workflow for producing both a printed publication and a machine-readable version of the editions. I will introduce the main considerations shaping this workflow – technical, practical, and scholarly, and then give special attention to the relation between the technical design goals and the scholarly considerations: it is their relation that motivates scholars to adopt the workflow in their own editorial projects, and also allows them to ask questions about their works that traditional editions cannot (or cannot easily) answer.

## Abstracts: Symposium

Bruno Lainé  
(University of Vienna)

### “Data Design for a Navigation System in the Tibetan Buddhist Canon”

In this presentation, I will address the theme of data design for a navigation tool in the Tibetan Buddhist canon. For the rkts.org site, three types of data are produced: xml, json and sql; xml for catalogs, json for indexes and sql for concordances. The organization of the data in a tree schema also allows to structure the whole corpus and makes it possible to keep an overview of the existing data pool and of the parts still to be developed.

The data in xml format allow direct visualization of the structures specific to the cataloged collections; they are a practical tool for entering catalogs. Data in json format are compact files used to generate index lists. Data in sql format allows quick access to a large number of concordances between canonical collections.

The rKTs site combines these three types of data to present results to a user's search (title, quotation search, etc.) and to provide the possibilities of navigating in the whole corpus.