

## **Digital Humanities and Research Methods in Religious Studies**

# **Introductions to Digital Humanities – Religion**

---

Edited by  
Claire Clivaz, Charles M. Ess, Gregory Price Grieve,  
Kristian Petersen and Sally Promey

## **Volume 2**

# **Digital Humanities and Research Methods in Religious Studies**



An Introduction

Edited by

Christopher D. Cantwell and Kristian Petersen

**DE GRUYTER**

ISBN 978-3-11-057160-8

e-ISBN (PDF) 978-3-11-057302-2

e-ISBN (EPUB) 978-3-11-057194-3

**Library of Congress Control Number: 2020944396**

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2021 Walter de Gruyter GmbH, Berlin/Boston

Cover image: Social network visualization. With friendly permission of Martin Grandjean.

Printing and binding: CPI books GmbH, Leck

[www.degruyter.com](http://www.degruyter.com)

# Table of Contents

Christopher D. Cantwell and Kristian Petersen

**Digital Humanities and Religious Studies: A “Why” To Guide — VII**

## Part I: Texts

Marcus Bingenheimer

**Digital Tools for Buddhist Studies — 3**

Lincoln A. Mullen

**The Making of *America’s Public Bible*: Computational Text Analysis for Religious History — 31**

Frederik Elwert

**Network Analysis of Religious Texts. Case Studies on Ancient Egyptian and Indian Religion — 53**

Rebecca Krawiec and Caroline T. Schroeder

**Digital Approaches to Studying Authorial Style and Monastic Subjectivity in Early Christian Egypt — 71**

## Part II: Images

Andrew Quintman and Kurtis R. Schaeffer

**Synthesizing Image and Text in the Life of the Buddha — 97**

James S. Bielo & Claire Vaughn

**Materializing the Bible: A Digital Scholarship Project from the Anthropology of Religion — 119**

Emily C. Floyd and Sally M. Promey

**Collaboration and Access in the Study of Material and Visual Cultures of Religion — 143**

Erin Walcek Averett and Derek B. Counts

**Scaling Religious Practice from Landscape to Artifact: Digital Approaches to Ancient Cyprus — 167**

### Part III: **Places**

Abhishek Amar

**Sacred Centers in India: Archiving Temples and Images of a Hindu City — 199**

J. E. E. Pettit, Fenggang Yang, Yuqian Huang

**Developing a Database of Religions in Contemporary China — 215**

Louis Kaplan and Melissa Shiff

**From Ararat to Kimberley: Activating Imaginary Jewish Homelands with Augmented and Virtual Reality — 227**

Caleb Elfenbein, Farah Bakaari, Julia Schafer

**Mapping Anti-Muslim Hostility and its Effects — 249**

### Part IV: **Issues**

Christopher R. Cotter and David G. Robertson

**Critique and Community: Podcasting Religious Studies — 273**

S. Brent Plate

**Public Pedagogy: MOOCs and Their Revolutionary Discontents — 291**

Wendi Bellar and Heidi A. Campbell

**Building Social Sites of Collaborative Research: A Case Study of the Network for New Media, Religion, and Digital Culture Studies — 303**

Russell T. McCutcheon

**Learning to Code: Digital Pebbles and Institutional Ripples — 319**

**Index — 339**

Christopher D. Cantwell and Kristian Petersen

## Digital Humanities and Religious Studies: A “Why” To Guide

In 2018, the American Academy of Religion hosted a special “wildcard” panel on digital research and teaching projects that encourage students to engage with their local communities. Titled “Teaching Local Religion with Digital Humanities: Objects, Methods, Pedagogies,” the session featured ten educators who contributed to eight different projects. Some of the projects were based at major research universities and worked with communities across the country. Others were housed at small liberal arts colleges and focused on one place of worship. But almost all of the projects connected with the scholars’ research agenda and employed the tools and methods of digital humanities (DH) in order to benefit from the work of student collaborators.<sup>1</sup>

Despite the panel’s explicit focus on teaching, the question and answer session that followed forayed into other matters. The very first question from the audience asked how—or even if—such projects counted in tenure and promotion portfolios. Another wondered how the projects’ directors received credit for their work. Finally, one audience members seemed to question the entire premise of the panel altogether, asking whether the time spent developing these projects would be better spent producing journal articles or monographs.

Such questions are important ones. They get at some of the tensions around integrating digital scholarship into the academic study of religion. But they also missed the point of the session. While the presenters were eager to talk about the pedagogical and intellectual benefits they had gained by employing digital methods, the audience was concerned with the professional issues that surround digital work.

---

1 “Teaching Local Religion with Digital Humanities: Objects, Methods, Pedagogies,” American Academy of Religion Annual Meeting, Denver, Colorado, 19 Nov. 2018. The projects featured at the roundtable were Emily Suzanne Clark, “Jesuit Missions on the Columbia Plateau”; Shana Sippy and Michael McNally, “Religious Diversity in Minnesota Initiative,” <https://religionsmn.carleton.edu/>; Gale Kenny, “Religion in the Archive”; Rachel McBride Lindsey, “Arch City Religion,” <https://www.archcityreligion.org/>; Christopher D. Cantwell, “Gathering Places: Religion and Community in Milwaukee,” <https://uwm.edu/gatheringplaces>; Rachel Kranson, “Pittsburgh Torah Scrolls Project,” <https://pittsburghtorahscrolls.wordpress.com/>; Amy DeRogatis and Isaac Weiner, “American Religious Sounds Project,” <https://religioussounds.osu.edu/>; Jennifer S. Leath, “Black Religious Denver”; Cara Lea Burndige, “American Religion and Refugees in the Heartland.”

It is a disconnect that also often defines conversations about digital humanities more broadly. From the moment scholars first took up the term “DH” in the early 2000s to describe work that either harnessed new technology for humanistic inquiry or studied new technologies humanistically, debates about DH tend to circulate around “how” questions. How is digital scholarship going to count? How is it going to be peer reviewed? How is it made?<sup>2</sup> Though, again, important, what can get lost in these more practical discussions is a sustained reflection on why one might want to take up digital methods in the first place. Why are an increasing number of scholars and educators finding digital humanities relevant to their work? What advantages have these methods presented to them? How have they benefited? And why might digital scholarship be vital for the study of religion’s future?

This collection aims to take up this later set of questions. Rather than add to the already robust literature on best practices in the production, evaluation, and promotion of digital scholarship, it aims to focus on the ways scholars have found that digital methods enhance their research, teaching, and service to the profession. Gathering together the lead investigators of more than a dozen projects, the collection covers a wide swath of the digital work now being done in the study of religion. It considers why podcasting and social media platforms might be the future in promoting the public understanding of religion. It showcases why digital projects might be an invaluable resource in helping our students become critical producers of knowledge about religion. And it dwells upon why twenty-first century technology might be the most productive avenue for answering some of the questions that have animated our fields for almost a century. Though by no means a comprehensive survey of the ever-growing body of digital work taking place, the collection does hope to serve as a snapshot of the inspirations and revelations that have propelled digital scholarship’s growth of late. In assembling the volume, we asked contributors to outline the journey of their inquiries and the logic behind the approaches they used. We hope that by

---

2 On the history of digital humanities and the perennial debates that have defined the field, see Susan Hockey, “The History of Humanities Computing,” in *A Companion to Digital Humanities*, Susan Schreibman, Ray Siemens, John Unsworth, eds. (Oxford: Blackwell, 2004), [http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213\\_brand](http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213_brand); Matthew Kirschenbaum, “What Is Digital Humanities and What’s It Doing in English Departments?” *ADE Bulletin* 150 (2010): 2–3; Matthew K. Gold, “The Digital Humanities Moment,” in *Debates in Digital Humanities*, Gold, ed. (Minneapolis: University of Minnesota Press, 2012), <http://dhdebates.gc.cuny.edu/debates>; Christopher D. Cantwell and Hussein Rashid, *Religion, Media, and the Digital Turn* (New York: Social Science Research Council, 2015).



mapping the terrain in such a way—showcasing why digital tools aided in producing, managing, or teaching religious studies—the reader might come away with an alternative perspective on researching their own subjects or be inspired to try new methods that resonate with their materials.

## Why Digital Humanities?

As the case studies presented here hopefully will demonstrate, there are as many reasons for drawing upon digital methods as there are digital projects. Scholars often find themselves drawn to digital methods because they help grapple with a question peculiar to their field, or because they open up new avenues for addressing perennial issues related to teaching, research, or service. But taken as a whole, the essays collected here do point to a series of overlapping inspirations that have propelled the field. For instance, several of the authors collected here turned to digital tools because analyzing a body of source material by traditional methods proved impractical. Lincoln Mullen’s interest in documenting the most widely quoted passages of scripture throughout American history would have required him to read hundreds of millions of newspaper articles. Similarly, Frederick Elwert’s research on the frequency with which certain characters or ideas appear in sacred texts across time, space, and tradition would have necessitated a lifetime of close reading and cataloging. These are issues that confront scholars across the academy today as the sheer volume of material or data related to our work grows exponentially. And by using textual and network analysis software, both Mullen and Elwert were able to process colossal quantities of data over a reasonable amount of time in order to contribute to debates in their field in new and vital ways.

The ability of digital research methods to facilitate the interpretation of data goes beyond just computational analysis, however. A number of the projects featured here have turned toward digital platforms in order to foster collaboration or consolidate resources around a particular kind of source material. Rebecca Krawiec and Caroline T. Schroeder’s work on the Coptic *SCRIPTORIUM*, for example, not only uses text analysis software to analyze Coptic texts. It also takes advantage of open-source text encoding methods to create a shared database of Coptic items from multiple institutions across the globe. James Bielo and Claire Vaughan’s piece on the Materializing the Bible project, meanwhile, similarly curates a collection of texts, images, and multimedia related to those instances when Christian communities across North America attempted to make Bible stories “real” through art, sculptures, plays, and other public performances. What’s more, the expanse of Bielo and Vaughn’s work is also made possible

by the ability of many digital asset management systems, which allow them to host a variety of kinds of media in one place. Through their intentional curation, juxtaposition, and organization, Bielo and Vaughn, like Krawiec and Schroeder, open new avenues of inquiry and analysis.

In addition to creating new collections of source material, moreover, digital methods also allow scholars to reach and constitute new publics as well. Wendi Bellar and Heidi A. Campbell, for instance, argue in their contribution that social network platforms can actually facilitate the kind of interdisciplinarity the study of religion claims to foster. On the Network for New Media, Religion, and Digital Culture Studies they run, scholars from any discipline can create a profile in order to find a potential research partner or learn about new work that might have been published in another field. Christopher R. Cotter and David G. Robertson take this argument even further, claiming that new media outlets like podcasting not only connect scholars with each other, but can also connect scholars with the wider public. As founding contributors of the Religious Studies Project, Cotter and Robertson hope that their work will impact the public understanding of religion. The episodes and essays they host on their site are aimed at both scholars across the academy and lay readers who take an interest in the field.

As Cotter and Robertson's work suggests, one of the central promises of digital research methods are their ability to blur boundaries between previously separate categories or spheres of work. For the Religious Studies Project, this means collapsing the distance that exists between the academy and the wider public. With Coptic SCRIPTORIUM, the line where Krawiec and Schroeder's research begins and their service to the profession ends is blurred. Digital scholarship's multimodal nature is often what makes it incredibly difficult to fit into existing themes of promotion, peer review, and tenure. But for many of the contributors here, digital scholarship's genre-busting potential is also its greatest promise. Caleb Elfenbein's project documenting instances of anti-Muslim hostility in the U.S., for instance, is a class assignment, research project, and public resource all at once. Students scour the media for reports of Islamophobic or anti-Muslim hate crimes that official outlets fail to collect. This material then gets posted on an interactive map that journalists, activists, and policy makers can refer to. Abhishek Amar's work in developing virtual recreations of lost or inaccessible Hindu temples in India similarly builds upon the digital inclinations of twenty-first century students in order to make them partners in the research process. In a different vein, however, Louis Kaplan and Melissa Schiff's project blurs the boundaries between scholarship and art, fact and speculation. Their work developing speculative recreations of failed efforts to create a Jewish homeland through virtual reality and other map-based software both documents these overlooked endeavors and interrogates the impulses that tie them together.

Each of these projects, in short, drew upon digital tools, platforms, or methods for a multitude of reasons. It is important to note, however, that despite this diversity all of the efforts documented here share a common understanding about the nature of DH work. Every project in its own way turned to new technologies because they allowed them to advance the study of religion’s longstanding commitment to research, teaching, and public service. Emily Floyd and Sally Promey’s essay on the Material and Visual Cultures of Religion (MAVCOR) perhaps documents this most clearly. MAVCOR began in part as an archive of sources related to religion’s material and visual manifestations. But those connected with MAVCOR intended the archive to serve as a resource that could promote greater collaboration between art historians, visual studies scholars, and the study of religion. As part of its efforts, MAVCOR’s team published an edited volume and established a new journal. Though it is published online in order to take advantage of the web’s ability to publish images and other material at a lower cost than print publications, MAVCOR’s journal functions, at its core, as an exemplar of best practices in scholarly publication.

The same is true for the other projects showcased here. The turn to the digital is in many ways a means to augment and enhance the work we do in analog. Quality digital work should never exist solely of its own accord. It should contribute to the study of religion’s ongoing efforts in print, in the classroom, and in the public. The works gathered here exemplify this commitment to the field. Our hope is that their work can inspire others in their efforts as well.

## What to Consider

While the reasons why scholars turn to digital work are diverse, the process by which projects came together, developed, and then launched is more uniform. The growth in digital scholarship in the last decades has provided a roadmap that many projects follow, and the contributions collected here are no different. From the essays readers may glean several themes that arise when using digital methods or producing digital scholarship. Taken together they draw attention to several things to consider when embarking on a digital humanities project. They include:

- *Do your homework.* Before starting a project do some exploratory research on the “how to” of DH. There is a growing body of resources, many of them open source and freely accessible, on how to employ certain methods, do certain

types of analysis, and use specific tools.<sup>3</sup> As you learn about new methods and tools some of what you'll encounter may seem beyond your abilities or patience. Don't get discouraged. Keep moving forward in your DH discovery and see where it brings you. Often you will not need to master lots of technical tools or languages in order to produce results that can advance your research. But it is important that you select the right tool for your project, and consulting the plethora of DH "how to" guides will help in that selection.

- *Collaboration is key.* Many projects will require the labor of several contributors, often those with very different skill sets. As with traditional research, collaboration may benefit from including various disciplinary specializations (i.e. different regional foci, expertise in varied traditions, or methodological approach, etc.). Digital projects, however, also often require collaboration with individuals from other fields altogether, including programmers, archivists, librarians, and students. Finding the right collaborators is key, and there are many outlets to do so. If you lack access to digital scholarship networks at the institution where you work, there are a number of DH blogs and professional associations you can connect with. Of particular interest will be the The Humanities and Technologies Camp (THAT Camp), which the American Academy of Religion and the Society for Biblical Literature sponsor at the start of their joint annual meetings.<sup>4</sup>

---

**3** The literature and resources here are immense. For resources in the field see, The Digital Humanities Literacy Guidebook, (Cleveland: Carnegie Mellon University, 2018), <https://cmu-lib.github.io/dhlg/>; Doing Digital Scholarship (New York: Social Science Research Council, 2018), <https://labs.ssrc.org/ddsh/>; Digital Humanities Now (Fairfax, VA: George Mason University, 2009), <http://digitalhumanitiesnow.org/>. Commonly cited introductory texts in the digital humanities also include Gold, ed., *Debates in Digital Humanities*; Daniel J. Cohen and Roy Rosenzweig, *Doing Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web* (Philadelphia: University of Pennsylvania Press and Fairfax, VA: Center for History and New Media, 2005), <http://chnm.gmu.edu/digitalhistory/>; Anne Burdick, et al., *Digital Humanities* (Cambridge, MA: MIT Press, 2016); Elieen Gardiner, *The Digital Humanities* (New York: Cambridge University Press, 2015); Claire Battershill and Shawna Ross, *Using Digital Humanities in the Classroom: A Practical Introduction for Teachers, Lecturers, and Students* (New York: Bloomsbury Academic, 2017); D. Berry, ed., *Understanding Digital Humanities* (New York: Palgrave Macmillan, 2012); Geoffrey Rockwell and Stefan Sinclair, *Hermeneutica: Computer-Assisted Interpretation in the Humanities* (Cambridge, MA: MIT Press, 2016); Jentery Sayers, *Making Things and Drawing Boundaries: Experiments in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2018); Cantwell and Rashid, *Religion, Media, and the Digital Turn*.

**4** Other resources include the Alliance of Digital Humanities Organizations, <http://adho.org/>; MLA Commons, <https://mla.hcommons.org/core/>; Network Infrastructure for Nineteenth-Century

- *Think iteratively.* Traditional models of scholarly production based around print typically presume a project is finished upon publication. The appearance of a monograph or journal article represents years of labor. Digital scholarships, however, can—and often should—work in reverse. The launch of a digital project marks the beginning of the work to be done rather than its end. The projects featured here are growing and changing, not finished. What this means is that project directors have the ability to think about their work in phases or stages, allowing for both the project and the project team to grow and learn over time. If a project seems too daunting, consider breaking it into a series of phases that provide a set of benchmarks that can be set. The publics you envision or engage with can then provide multiple moments of connection and benefit. This iterative thinking also relates to scale, allowing a project to grow in scope over time.
- *Audience(s) is key.* A given project may address a number of constituencies. Who your intended audience is will shape the types of choices you make regarding the tools you employ, the user interface, privacy, accessibility, and phases. Projects might simultaneously be used by academic and general audiences, with their findings being used for scholarly research, activist outreach, and educational purposes all at the same time. These various components might develop over time but it is key to think about how your project will take shape as it moves beyond your computer and is taken up by others.
- *Is this ethical?* While digital tools and analysis may make new approaches available to researchers we should always think about the implications and social consequences of our research. Some projects may put religious actors, especially those from vulnerable communities, at risk. When designing projects one should consider: Am I endangering anyone by making this information available publicly? How can I protect my subjects? How might my data be used in the future? The ethical issues related to data and privacy may force you to imagine your project’s objectives or methods in necessary ways.
- *Does this “count?”* For many scholars the question of institutional expectations may shape the types of scholarship they can pursue, especially in relation to tenure and promotion requirements. If one is not hired specifically to do digital humanities work, then a DH project may not fulfill narrow definitions of “scholarship.” In these cases, academics often take an approach that places digital scholarship as an extra outcome in addition to more tra-

ditional publishing methods, such as peer-reviewed journals and books. Part of the reason institutions don't recognize digital projects as fully as they might is because they may not be legible to reviewers using longstanding departmental standards. The American Academy of Religion (along with organizations such as the American Historical Association, Modern Language Association, and others) has produced "Guidelines for Evaluating Digital Scholarship," which are intended to aid evaluating committees in assessing digital scholarship. Part of the effort can be done by recognizing more traditional criterion for scholarship, such as grants, academic reviews, citations, or classroom use.

- *Think about sustainability.* While the ability to grow and evolve digital projects over time is one of the medium's benefits, this feature also raises the spectre of a lifetime of upkeep and maintenance. Therefore, it is incumbent to think about the lifetime of your project early on in order to identify both the kind and amount of resources you will need to maintain it. Not every project needs to involve a decade-long commitment to build a colossal archive that can call forth a new field of inquiry. Smaller scale projects can be completed in either a year or a semester's time and then left up with little change. Keeping projects online, however, often requires regular hosting and registration fees. So, accounting for when a project will sunset, and considering where the data from a project will live after the interface might break, are topics worthy of consideration early on.

While there are many issues to keep in mind when developing a digital project, the most important, perhaps, is managing expectations and keeping open the opportunity for surprise. Results derived from a DH project may not align with the initial expectations one may have going into a project. Alignment between the two shouldn't be the sole standard for measuring success. There are many unexpected elements that will arise when doing digital work but often these surprises will result in new opportunities. Unanticipated issues may be raised through the accumulation of valuable data, which may require learning how to use a new digital tool in order to analyze this data. Even grave challenges or out-right failures can lead to the next productive stage of a particular project. Luck (good or bad) can often be a factor in how a project takes shape or in the results it garners. What matters is that you thoughtfully and intentionally engage with these tools as you would any research method. And like other forms of inquiry and analysis, the best way to engage with them thoughtfully is by considering other examples. This is what we hope these essays provide.

## A Guide to the Volume

As there are an infinite number of reasons scholars turn to DH and numerous ways these projects take shape, there are many ways to structure this volume as well. One version could have structured the collection by research method, dividing the essays by the kind of software or tool that they use. Another could have arranged the volume by activity type, grouping together those essays focused on research, teaching, or public engagement. We chose, however, to arrange the volume by source type, with sections on texts, images, places, and issues. Such an arrangement is in service of the fact that the kinds of sources scholars work with often dictate the tools or techniques used. Part I on “Texts,” for instance, includes pieces on four projects whose data is primarily textual in nature. Marcus Bingenheimer’s contribution focuses on the ways digitization and text encoding are changing the nature of research within Buddhist studies. The advent of computational methods have made the creation of encoded corpora valuable scholarly contributions to the field. Lincoln Mullen’s essay, meanwhile, showcases the kind of public scholarship that can emerge from employing textual analysis. His America’s Public Bible Project mined a massive collection of digitized American newspapers to pull out the frequencies of specific Bible quotes. In addition to advancing Mullen’s own scholarship on the topic, the application of digital methods also resulted in the creation of a public-facing resource that allows anyone to chart the use of a particular passage of scripture over time. Similarly, Frederik Elwert’s contribution documents how the digitization of textual material allows for the application of methods commonly associated with the study of institutions, organizations, or relationships. By tagging keywords, character, and parts of speech, Elwert’s SeNeReKo project was able to chart the history of religious contact by the appearance of certain themes in ancient Egyptian and Buddhist texts. Like Bingenheimer, Mullen, and Elwert, Rebecca Krawiec and Caroline T. Schroeder’s essay on the Coptic SCRIPTORIUM project underscores how the digitization and encoding of texts are vital scholarly contributions.

Building upon Part I’s focus on texts, Part II on “Images” considers the digitization and interpretation of sources beyond the written word. Andrew Quintman and Kurtis R. Schaeffer’s essay, for example, illustrates how digital methods can help bring textual and visual material in conversation with each other. The Life of the Buddha project they discuss annotates large Tibetan murals with both scholarly commentary as well as passages from the texts that inspired the work of art. Similarly, Erin Walcek Averett and Derek B. Counts’ contribution highlight how a suite of digital tools have brought fresh insights to archaeological remains



from Cyprus. Beyond rich, 3D modeling of artifacts that allow for levels of scrutiny beyond the human eye, digital methods also allow archaeologists to link this data to provenance, geographical information, and scholarly commentary. James Bielo and Claire Vaughn's contribution, meanwhile, demonstrates how the web's ability to keep multiple kinds of data from myriad locations in a single site allows for the creation of specialized research archives. The Materializing the Bible project they discuss has, to date, collected material related to nearly five hundred attractions that made Bible stories real and manifest in some way. Finally, Emily Floyd and Sally Promey's piece on the Material and Visual Cultures of Religion project demonstrates the kind of scholarly outcomes that can come from this kind of critical curation. As previously mentioned, the MAVCOR project includes both a digital archive as well as a peer-reviewed journal that draws upon that archive as a source of reflection and analysis.

In contrast to Parts I and II, Part III on "Places" focuses less on a particular kind of source and more on a mode of thought encouraged by the digital turn. Focusing on questions of space, or using geographic information systems as a means of analyzing material has become such a prominent part of DH work that some have argued that it constitutes its own field.<sup>5</sup> Abhishek S. Amar's essay on the 3D modeling of historic Hindu and Buddhist sites of pilgrimage, for instance, explicitly grapples with the kind of transformative work scholars must do to their data in order to make it machine readable. But the final renderings of these sites, Amar claims, would allow for the study of these sites across time, combining textual, archaeological, and art-historical analysis. Spatial thinking also allowed J. E. E. Pettit, Fenggang Yang, and Yuqian Huang to shed light on territories devoid of other kinds of source material. In creating a database of 72,000 temples, mosques, and churches in China, they were able to approximate the relative size of these religious communities in a country that refuses to release such data. This ability of space to make real that which may seem absent also defines Louis Kaplan and Melissa Shiff's essay on Mapping Ararat. They use augmented and virtual reality software to project failed attempts at establishing a Jewish homeland onto the landscape. In this way, they call forth the religious imaginaries that often animate communities or individuals, instantiating the visions that have motivated people. Finally, where Kaplan and Shiff build speculative topographies, Caleb Elfenbein, Farah Omer, and Julia

---

5 On the importance of "spatial humanities" see Richard White, "What is Spatial History?" *Stanford Spatial History Lab: Working Paper*, February, 2010, <http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>; David J. Bodenhamer, John Corrigan, and Trevor M. Harris, *The Spatial Humanities: GIS and the Future of Humanities Scholarship* (Bloomington: Indiana University Press, 2010).



Shafer discuss how they also make invisible landscapes visible. By geo-referencing hate crimes against Muslims and other instances of anti-Muslim hostility, Elfenbein, Omer, Shafer, and others demonstrate both the pervasiveness and the regional specificities of American Islamophobia.

Where the collection's first three sections focus on particular types of data or sources, the final section, Part IV, focuses on how scholars are using DH tools and methods to address issues that impact the study of religion more broadly. Christopher R. Cotter and David G. Robertson's essay documents their ambitious plan to create a social media strategy for the academic study of religion. The Religious Studies Project they discuss hosts rigorous debates about theory, method, and interpretation. But it presents this material in genres accessible to wider audience such as blogs and podcasts. S. Brent Plate, meanwhile, reflects upon his experience teaching a Massive Open Online Course, or MOOC. While debates in DH might have moved beyond MOOCs to discuss more practical concerns related to online education, Plate's essay offers a great deal of insight related to how scholars can, and cannot, reach broader audiences online. Wendi Bellar and Heidie A. Campbell's essay, meanwhile, explores how digital platforms can reach new audiences within the academy. Like Floyd and Promey's essay on MAVCOR, Bellar and Campbell's discussion of the Network for New Media, Religion, and Digital Culture Studies site documents how the web has become a place for new subfields to form. In the absence of support from or cooperation between different professional organizations, Bellar and Campbell have created a space online where interlocutors can converse and collaborate. Finally, the collection concludes with a vital, but surprisingly overlooked discussion: the impact digital tools or platforms are having on religious studies departments. An essay by Russell T. McCutcheon discusses one of the oldest attempts to harness new technology to promote the study of religion at the university level. Through blogging, podcasting, and other new media outlets, the University of Alabama's Department of Religious Studies has made the work of generating majors a site of vibrant disciplinary discourse in the field as well.

While the above outline structures the volume, it is, as we mentioned, not the only way to approach the text. We encourage you to engage with the collection in ways that will inform and inspire your own work, diving into those essays that are relevant to you. Those who are interested in the use of digital tools for what might be called “pure” academic research, we recommend you look at the essays by Bingenheimer; Mullen; Elwert; Krawiec and Schroeder; Averett and Counts; Amaral; and Pettit, Yange, and Huang. For those who are interested in how digital platforms can help academic research reach broader audiences can consult Mullen; Floyd and Promey; Averett and Counts; Kaplan and Shiff; Elfenbein, Omer, and Shafer; Cotter and Robertson; Plate; and Bellar and Camp-

bell. Those interested in DH pedagogy, meanwhile, should see Bielo and Vaugh; Amar; Elfenbein, Omer, and Shafer; and McCutcheon.

But we hope you will engage with all of the essays, and draw from them the inspiration necessary to develop your own digital work. Because as each of the essays collected here demonstrates, the digital work happening now in the study of religion has built upon the work done in other disciplines across the academy. If the study of religion is to continue to grow, it will need to take up this kind of work and make it its own.

## Selected References

- Battershill, Claire and Shawna Ross. *Using Digital Humanities in the Classroom: A Practical Introduction for Teachers, Lecturers, and Students*. New York: Bloomsbury Academic, 2017.
- Berry, D., editor. *Understanding Digital Humanities*. New York: Palgrave Macmillan, 2012.
- Bodenhamer, David J., John Corrigan, and Trevor M. Harris. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press, 2010.
- Burdick, Anne, et al. *Digital Humanities*. Cambridge, MA: MIT Press, 2016.
- Cantwell, Christopher D. and Hussein Rashid. *Religion, Media, and the Digital Turn*. New York: Social Science Research Council, 2015.
- Carnegie Mellon University. *The Digital Humanities Literacy Guidebook*. <https://cmu-lib.gitlab.io/dhlg>.
- Cohen, Daniel J. and Roy Rosenzweig. *Doing Digital History: A Guide to Gathering, Preserving, and Presenting the Past on the Web*. Philadelphia: University of Pennsylvania Press, 2005.
- Gardiner, Eileen. *The Digital Humanities*. New York: Cambridge University Press, 2015.
- Gold, Matthew K. "The Digital Humanities Moment." In *Debates in Digital Humanities*, edited by Matthew K. Gold, <http://dhdebates.gc.cuny.edu/debates>. Minneapolis: University of Minnesota Press, 2012.
- Hockey, Susan. "The History of Humanities Computing." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, John Unsworth, [http://digitalhumanities.org/3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213\\_brand](http://digitalhumanities.org/3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-2-1&toc.depth=1&toc.id=ss1-2-1&brand=9781405103213_brand). Oxford: Blackwell, 2004.
- Kirschenbaum, Matthew. "What Is Digital Humanities and What's It Doing in English Departments?" *ADE Bulletin* 150 (2010): 2–3.
- Rockwell, Geoffrey, and Stefan Sinclair. *Hermeneutica: Computer-Assisted Interpretation in the Humanities*. Cambridge, MA: MIT Press, 2016.
- Roy Rosenzweig Center for History and New Media, George Mason University. *Digital Humanities Now*. <http://digitalhumanitiesnow.org/>.
- Sayers, Jentery. *Making Things and Drawing Boundaries: Experiments in the Digital Humanities*. Minneapolis: University of Minnesota Press, 2018.
- Social Science Research Council. *Doing Digital Scholarship*. <https://labs.ssrc.org/dds/>.

White, Richard. “*What is Spatial History?*” *Stanford Spatial History Lab: Working Paper*,  
<http://www.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=29>.



---

## Part I: **Texts**



Marcus Bingenheimer

# Digital Tools for Buddhist Studies

## Introduction

Buddhists have rarely hesitated to embrace new forms of communication. The earliest Indian epigraphy (3<sup>rd</sup> century BCE), and the earliest manuscript fragments in Indian languages (1st century BCE/CE) are connected to Buddhism. The earliest extant printed book, dated 868 CE, is a Chinese translation of the Diamond Sutra.<sup>1</sup> Throughout its history Buddhism has used whatever means available to encode, disseminate, and maintain its growing corpus. Buddhist texts were first composed and transmitted in India in a cultural environment that valued the mnemonic techniques of oral transmission. Later Buddhists became eager “early adopters” of two other emerging information technologies—writing and printing. Below I address the shift of Buddhist heritage information into the digital under three main headings: the digitization of Buddhist texts and images, the digitization of scholarly tools (dictionaries, bibliographies etc.), and the application of computational methods on those data.

## Data: Digitization of Primary Sources

### Digital Editions of Canonical Texts

Through the centuries Buddhists have managed their shifting corpora via the changing media of oral, handwritten, printed, and now digital text. Part of the conceptual apparatus for these endeavors is the notion of canonicity, a central

---

**Note:** I am grateful for responses by: Venerable Ānandajoti, Rupert Gethin, Paul Hackett, Susan Huntington, Bryan Levman, Charles Muller, Michael Radich, Miroj Shakya, Sam Van Schaik, Venerable Sujato, Venerable Upatissa, Jeff Wallman, and Christian Wittern. Any mistakes in the information about the resources and all opinions expressed about them are my own. The discussion below is meant to be comprehensive, but it is not complete, and I apologize to all deserving projects that I fail to mention. I have chosen to assess only projects of which at least some data is made freely available.

<sup>1</sup> For the connection of Buddhism and the invention of printing in China see T. H. Barrett, *The Woman Who Discovered Printing* (New Haven: Yale University Press, 2008).

and early concern for Buddhists.<sup>2</sup> Although there is no single, stable Buddhist canon that is used in all traditions, the concept of canonicity, both fluid and robust, has played an important role in shaping how Buddhists perceive of their textual heritage. It is thus not surprising that first digitization efforts were aimed at producing digital editions of the “canon.” Most of the digital canonical editions were created independently from each other, and as a result we have several overlapping versions of the Pāli, Chinese, and Tibetan canon. These are often modeled on twentieth century print editions.

## Pāli

The Pāli Canon exists in three major independent digital versions, most of which were created in the late 1980s and 1990s. These have been copied across the net, often with minor changes along the way.<sup>3</sup> As a result, digital Pāli texts are easy to find online, but their provenance and editorial standards are often undefined. This makes them difficult to cite and to rely on for philological research.

Perhaps the most influential digital edition of Pāli Buddhist texts is the final CD version (Ver. 3) of the *Chaṭṭha Saṅgāyana Edition* that was published by the Vipassana Research Institute (VRI) in late 1999.<sup>4</sup> As the name *Chaṭṭha Saṅgāyana* implies, the VRI corpus is a digitization of the printed canon as redacted by the sixth council that was held in Yangon from 1954 to 1956. The strengths of the VRI corpus are that the texts have been proofread, and that it alone among digital editions of the Pāli canon includes the commentaries (*aṭṭhakathā*) and sub-commentaries (*ṭīka*). Markup links connect the commentaries to the *mūla* text, making it possible to build interfaces that present the *mūla* together with two layers of commentaries.

---

<sup>2</sup> For a discussion of Buddhist notions of “canonicity” across traditions see Jonathan A. Silk, “Canonicity” In *Brill’s Encyclopedia of Buddhism*, ed., Jonathan A Silk, Vol. 1, Languages and Literatures. (Leiden: Brill, 2015), 5–37.

<sup>3</sup> E.g. the World Tipiṭaka Edition, produced in Thailand, that aims to improve Burmese Chaṭṭha Saṅgāyana edition. Another large, on-going edition project to watch is the Dhammachai Tipiṭaka Project (2010–) in Thailand. This also aims at a digital edition (Alexander Wynne, “A Preliminary Report on the Critical Edition of the Pāli Canon being prepared at Wat Phra Dhammakāya,” *Thai International Journal of Buddhist Studies* 4 (2013): 135–170, and Bryan Levman, “Towards a Critical Edition of the Tipiṭaka,” *Journal of Nāṇasaṃvara Centre for Buddhist Studies (JNCBS)* 87 (2018): 87; but as of today, digital text does not yet seem available.

<sup>4</sup> The Pali Tipitaka, <https://tipitaka.org>. All URLs given here and below were accessed Jun 2019 where not otherwise indicated. The VRI edition was also published in a Devanagari print edition.



It is unclear whether or in how far the online texts currently available on the VRI website, called Chaṭṭha Saṅgāyana Tripitaka Ver. 4.0, were edited beyond the last Chaṭṭha Saṅgāyana CD (Ver. 3) version. As with all digital editions of the Pāli canon there is lack of technical documentation, or indeed any documentation or meaningful metadata. Digital editions need, like their print counterparts, information as to who created the resource, when and where, and what editorial decisions were made (and why) in converting the printed into a digital text. Development on the VRI corpus seems to have stopped some years ago, though a search engine for the corpus (Windows only) and an iPhone app has been made available. These days the best way to use the VRI corpus is via the *Digital Pāli Reader* that is developed and maintained by Yuttadhammo Bhikkhu.<sup>5</sup>

A second digital edition of the Pāli canon is the Sri Lankan *Buddha Jayanti Tripitaka Project*, which has digitized the Pāli Canon from the government sponsored Sinhalese Buddha Jayanti edition (1956–1990), an edition that was created partly in response to the Burmese Chaṭṭha Saṅgāyana. The digital version of the Buddha Jayanti corpus, seems less well proofread than the VRI corpus, but it too has been available since the 1990s, and can be found on various websites. Next to the core texts of the Pāli Tripiṭaka, the Buddha Jayanti corpus comprises a small number of paracanonical and commentarial works, as well as texts on history, grammar and rhetoric. One stable way of accessing the Buddha Jayanti corpus is via the Göttingen Register of Electronic Texts in Indian Languages (GRE-TIL) (see below).

The third digital Pāli corpus, still hardly noticed by the scholarly community, is the release of the *Pāli Text Society* edition online under a CC License via GRE-TIL. The digitization is the result of a collaboration between the PTS and the Dhammakaya Foundation in Thailand between 1989 and 1996. The original aim was, as so often in the 1990s, to produce a CD. After two CD versions, this line of distribution was discontinued, and in 2014 the texts were released on GRE-TIL. The digital PTS corpus so far consists only of the Pāli Vinaya, Sutta and Abhidharma, none of the commentarial and paracanonical works from the PTS print series seem currently available digitally.

To date (June 2019), the files on GRE-TIL contain the copyright notice: “This file is (C) Copyright the Pali Text Society and the Dhammakaya Foundation, 2015. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 In-

---

<sup>5</sup> “Digital Pali Reader,” *Pali.Sirimangalo.Org*, <https://pali.sirimangalo.org> and Digital Pali Reader repository “Yuttadhammo/Digitalpalireader,” GitHub, <https://github.com/yuttadhammo/digitalpalireader> (the repository also contains XML versions of the Thai and Burmese editions of the Pāli Canon). <https://www.digitalpalireader.online/>.

ternational License.”<sup>6</sup> Moreover, the PTS files in GRETEL contain the following disclaimer: “These files are provided by courtesy of the Pali Text Society for scholarly purposes only. In principle they represent a digital edition (without revision or correction) of the printed editions of the complete set of Pali canonical texts published by the PTS. While they have been subject to a process of checking, it should not be assumed that there is no divergence from the printed editions and it is strongly recommended that they are checked against the printed editions before quoting.”

Working with digital Pāli text, at this stage the recommendation is to search the VRI Pāli canon via Yuttodhamma’s reader in order to have full access to the commentarial strata, then use the PTS editions in print or pdf to corroborate difficult or doubtful passages.

For simple searches with convenient access to translations and parallels it is best to use the SuttaCentral website (see Sec. 2.2), which hosts emended versions of the VRI corpus, and might at one point add commentarial literature.

## Chinese

The history of Buddhist canonical collections in Chinese has long been studied, especially in Japan.<sup>7</sup> The production of digital Chinese Buddhist texts was and is slightly more challenging than for texts in Indian languages or Tibetan, because Chinese cannot be presented meaningfully in alphabetic transcription. For a long time the rendering of Chinese character variants was an endemic problem for Chinese digital text. Only the advent of Unicode in 1993, and especially the addition of Extension B in 2001, put an end to the confusion of encoding systems and normalization strategies that had hindered the digitization of Chinese.

Like with the Pāli and Tibetan corpora, different organizations have produced independent digital editions of Chinese Buddhist texts, often framed around a specific canonical edition. The two main collections to date are the *Chinese Buddhist Electronic Text Association* (CBETA) corpus, and the *SAT Daizōkyō Text Database* (SAT for Saṃgaṅkikṛtaṃ Taiśotripitākam “Society for the Creation of the Taishō Tripiṭaka”). Both started out collaboratively as projects to digitize the Taishō Canon,<sup>8</sup> the canonical edition of the Chinese Buddhist Canon that,

---

<sup>6</sup> Many of the earlier editions are already in the public domain.

<sup>7</sup> Nozawa lists more than 1200 works related to the topic “canon” (*Daizōkyō* 大藏經) published in Japan between 1879 and 2003.

<sup>8</sup> *Taishō shinshū daizōkyō* 大正新脩大藏經. Tokyo: Taishō issaikyō kankōkai 大正一切経刊行会, 1924–1934.

published between 1924 and 1934, became authoritative for East Asian Buddhist Studies.

CBETA was founded in Taipei in 1998 in emulation of the Pāli Text Society, with the aim of providing reliable digital versions of Buddhist texts to a user community that comprised Buddhist believers as well as researchers in Buddhist Studies. CBETA always considered its main task to provide accurate Chinese Buddhist texts in many different formats. The texts can be read online at [www.cbeta.org](http://www.cbeta.org), but are also available for download in multiple formats (epub, mobi, pdf etc.) for different devices and applications. Before the advent of the mobile versions most users availed themselves to the (Windows only) CD version (current: Version 2018) with the dedicated CB Reader search engine. A sophisticated online interface for researchers, with some analytic functions, has been released in 2014.<sup>9</sup>

CBETA published its first CD version containing Vols. 1–55 and 85 of the Taishō Canon in 1999. Since then it has successively added texts from numerous other canonical editions, most importantly the *Manji shinsan zoku zōkyō* 卍新纂續藏經 (Tokyo, 1905–1912), which contains 1230 Chinese Buddhist texts that are not part of the Taishō canon. In addition, the CBETA corpus comprises 285 texts from the Jiaxing Canon 嘉興大藏經, some 250 Buddhist temple gazetteers, and other texts not included elsewhere from various sources and editions. The 2018 version of the corpus contained 4,621 texts.<sup>10</sup>

Crucial for researchers in the digital humanities (DH) is that CBETA provides its corpus in XML/TEI format on github.<sup>11</sup> This is currently the most comprehensive and usable collection of open access Chinese Buddhist texts for the application of corpus linguistics and related forms of analysis.

The SAT corpus, maintained and hosted by a team at Tokyo University, is currently accessible via its websites, the most recent version of which is dated 2018.<sup>12</sup> The interface allows for searches of all or selected sections of the Taishō canon (3,283 texts) as well as the *Jōdoshū zensho* 浄土宗全書, the “Collected Works of the Pure Land School.”<sup>13</sup> Especially helpful for translators is

<sup>9</sup> At: CBETA Online Reader, [cbetaonline.dila.edu.tw](http://cbetaonline.dila.edu.tw).

<sup>10</sup> The number of files in the archives available at: CBETA Chinese Electronic Buddhist Association, <http://cbeta.org/download/ebook.php>.

<sup>11</sup> “Cbeta-Git/Xml-p5a,” Github, <https://github.com/cbeta-git/xml-p5a>. For textual analysis of the Taishō texts only this version is preferable: “Cltk/chinese\_text\_cbeta\_taf\_xml,” Github, [https://github.com/cltk/chinese\\_text\\_cbeta\\_taf\\_xml](https://github.com/cltk/chinese_text_cbeta_taf_xml).

<sup>12</sup> The SAT Daizōkyō Text Database, [http://21dzk.l.u-tokyo.ac.jp/SAT/index\\_en.html](http://21dzk.l.u-tokyo.ac.jp/SAT/index_en.html).

<sup>13</sup> This collection contains 493 texts (Main Collection 306 texts + Supplement 187 texts) from the Japanese Pure Land school. The online texts are hosted at: “浄土宗全書,” テキストデータ

the linking of highlighted texts with the *Digital Dictionary of Buddhism* (see Section 2.1.), and to the corpus of English translations in the *Bukkyō dendō kyōkai* (BDK) corpus of translations. The interface also allows to view scans of the print edition, which is often helpful, especially when dealing with Siddham script, illustrations or variant characters.

Another recent contribution is a search interface for the heavily illustrated volumes 86–96 of the Taishō.<sup>14</sup> The images can be searched by keywords, magnified, and tagged, and are published according to the IIIF standard. Considering the dearth of open data regarding Buddhist Art, this is a welcome addition.

CBETA and SAT are often seen as giving access to the same data, because both started out (and collaborated) on digitizing the Taishō. However, their text base has diverged over the last fifteen years and the actual overlap of searchable text consists only of about 2,270 texts (Taishō Vols. 1–55 & Vol. 85), which is most of the Indian scriptures translated into Chinese, and the works composed in Chinese until c. the 8<sup>th</sup> century by Chinese, Korean and Japanese Buddhists. Those texts are nearly identical in both corpora, as both rely on the Taishō, however, the CBETA corpus offers revised punctuation for many texts, and has expanded the apparatus. It also provides (transparently marked) emendations where the CBETA editors judge the Taishō text to be erroneous and the mistake does not become apparent in an apparatus entry.<sup>15</sup>

In addition to the 2,270 texts shared between the SAT and the CBETA corpus, the SAT interface searches another c. 1,220 Buddhist texts from the Taishō canon (Vols. 56–84 and Vols. 86–97), which are not contained in the CBETA corpus. These are mostly texts by Japanese authors written after the 8<sup>th</sup> century, the majority composed in Buddhist Chinese. In contrast, the CBETA corpus contains another c. 2,351 texts from various sources, which are not accessible through the SAT website. These are mostly texts by Chinese authors and written after c. the 8<sup>th</sup> century.

As a rule of thumb, whoever studies Japanese Buddhism (or the Japanese commentarial tradition on Indian and Chinese texts) ought to work with the

---

ベース, <http://jodoshuzensho.jp/jozensearch>, where the collection is maintained. Parts of this collection overlaps with the Taishō.

<sup>14</sup> At: SAT Taishōzō Image DB, <https://dzkings.l.u-tokyo.ac.jp/SATi/images.php>.

<sup>15</sup> The latter practice carries the danger that users may be unaware that they are looking at an emendation. This is mainly an interface issue, as the emendations are transparently marked in the digital master text (XML/TEI) and the original wording is always preserved.

SAT website. For research on Chinese Buddhism one should make use of the latest version of the CBETA corpus.<sup>16</sup>

To date, the main difference between the two projects from a DH perspective is that CBETA aims to produce a wide ranging corpus of Chinese Buddhist texts, and distributes these texts in various formats under a CC license, whereas SAT aims to provide an online research platform for the Taishō edition of the Chinese canon.

A third digital project regarding the Chinese canon is the *The Tripitaka Koreana Knowledgebase Project* by the Research Institute of the Tripitaka Koreana (Seoul). Like SAT it is an attempt to carefully model one particular edition, in this case the first and second printing of the Korean edition of the Chinese canon. The project, however, seems to have been dormant for some years now. The website has been offline at times in the past, and used to be in dire need of maintenance and internationalization.<sup>17</sup> The Tripitaka Koreana Knowledgebase at one point offered scans of the surviving portions of the first printing of the canon, and could be an important resource for research into the printing history of the Buddhist canon, but in its current stage is difficult to use, at least via the English interface.

## Tibetan

Various groups have worked on electronic editions of Tibetan canonical collections and the “Collected Works” (*gsung ‘bum*) by later authors. The two most widely used digital collections are the *Asian Classics Input Project* (ACIP), and the *Buddhist Digital Resource Center* (BDRC) (formerly Tibetan Buddhist Resource Center).

Since 1987 the ACIP has produced distributables in plain-text format of the Kangyur (bka’ gyur), the Tangyur (bstan gyur) and various Sungbum collections.<sup>18</sup> The text entry of ACIP was accomplished by involving Tibetan communities in India, Tibet and Mongolia. The produced texts lack metadata about the editions from which they were created, which limits their usability for research on the level of individual texts. As a whole, however, the plain-text corpus might be used for corpus linguistic analysis and related forms of research.

---

<sup>16</sup> For searches of the CBETA corpus the recommendation today is: CBETA Online Reader, <http://cbetaonline.dila.edu.tw>.

<sup>17</sup> I was not able to search the database with current versions of Firefox, Opera, or Chrome.

<sup>18</sup> Downloadable at: Asian Classics Input Project, <http://www.asianclassics.org>.

BDRC was founded as the Tibetan Buddhist Resource Center in 1999 by E. Gene Smith (1936–2010), who dedicated his life to the preservation and dissemination of Tibetan texts. BDRC has digitized, cataloged, and archived a large number of culturally significant works, securing the once critically endangered Tibetan literary corpus and making it widely accessible. Smith's effort counts among the great success stories in cultural heritage preservation. Most works are distributed under a Creative Commons license. A few texts are restricted based on cultural concerns by stakeholders (e.g. regarding esoteric texts).

Some texts contained in the BDRC corpus are distributed as scans (in PDF) from original documents, others are available as full text. Both come with metadata, which helps to trace provenance, and makes them usable for research on the level of individual texts.<sup>19</sup> Currently, the website allows download only of single texts and even for this a user account is needed. However, the BDRC team is ready to consider requests in case full-text access is needed for DH related analysis.

In 2015, the Board of Directors voted to broaden the Center's preservation mandate to include texts in languages beyond Tibetan, including, among others, Sanskrit, Chinese, and Pāli. To reflect its expanded mission, the Center's name was changed to Buddhist Digital Resource Center.

Many researchers in Tibetan Buddhism make use of both the ACIP and the BDRC collection in one form or another. They often search for material via Paul Hackett's *Buddhist Canons Research Database* (see Sec. 2.2).

Still another Tibetan canon project is less well known, but deserves more attention. The "Resources for Kanjur & Tanjur Studies" (rKTs) provides both transcriptions and scanned images of a range of printed and handwritten editions via a minimalist website.<sup>20</sup> It is part of the Tibetan Manuscripts Project Vienna (TMPV) directed by Helmut Tauscher and a good place for online research into the edition history of the Tibetan canon.

## Sanskrit

Though one or more canonical collections in Sanskrit might have existed in India at one point, none have survived *in toto*, and no Sanskrit "canon" as such has been compiled or translated as a whole. Nevertheless, a great wealth of Buddhist

---

<sup>19</sup> At: Buddhist Digital Resource Center, <https://www.tbrc.org>.

<sup>20</sup> Resources for Kanjur & Tanjur Studies, <https://www.istb.univie.ac.at/kanjur/rktsneu/sub/index.php>.

Sanskrit texts have survived, often complete in the monasteries of Nepal and Tibet, sometimes fragmentary in the sands of South and Central Asia.

*GRETIL* (the “Göttingen Register of Electronic Texts in Indian Languages and related Indological materials from Central and Southeast Asia”) is among the oldest and best managed repositories for digital Indian texts.<sup>21</sup> Safely hosted by the Niedersächsische Staats- und Universitätsbibliothek Göttingen, the formidable team has early on avoided the principal mistake of digital resources (over-reliance on a particular interface) and distributed well-curated text files with basic metadata. Conceived as a repository for Indology in general, *GRETIL* has become a valuable resource for digital textual studies of Indian and Pāli Buddhism. The main file format is HTML with some metadata information at the beginning of each file. Longer texts are often split in several files. To date it contains some 250 Buddhist Sanskrit texts that can be downloaded in one single zip-archive. In another major contribution *GRETIL* recently (2017) has added digitized versions of fourteen Sanskrit dictionaries in CSX format, allowing for aggregated search in platforms such as GoldenDict.

The other Buddhist Sanskrit project of note is the *Digital Sanskrit Buddhist Canon*, since 2003 maintained by the University of the West.<sup>22</sup> The digital texts are produced from available print editions at the Nāgārjuna Institute of Buddhist Studies in Kathmandu. Over the years, the Nāgārjuna Institute has assembled an important collection of texts. The texts have been proofread, have basic metadata associated, and are made available through the project website. The interface, however, is lacking in faceted search and, as so often, does not offer the data archived for download. Fortunately, many of the texts are shared with permission via *GRETIL*, from which users can assemble their own collections.

Though there is a considerable overlap between *GRETIL* and the Digital Sanskrit Buddhist Canon there are texts which are unique to either of the repositories. At this stage, the recommendation is to get the latest version of the Buddhist Sanskrit texts from *GRETIL* and search the files via a text editor or a grep-like command line tool. In addition, one should search the Digital Sanskrit Buddhist Canon.

---

<sup>21</sup> *GRETIL*, <http://gretil.sub.uni-goettingen.de/>.

<sup>22</sup> Digital Sanskrit Buddhist Canon Project, [www.dsbcproject.org](http://www.dsbcproject.org).

## Thematic Collections

Next to collections of “canonical” texts—however defined—there are a number of important digital collections that do not work with the canon as a primary category for organizing text, but which are centered in various ways on geographic regions, single material collections, topics or genres.

The *Huntington Photographic Archive of Buddhist and Asian Art* is the largest independent archive of Buddhist Art. It represents the field documentation efforts by John and Susan Huntington from 1969 to the present.<sup>23</sup> The archive was first established at Ohio State University in 1986. The collection is currently being accessioned by the University of Chicago Libraries, where it will be permanently housed and maintained.

The material Huntington Archive contains more than 250,000 original slides and photographs documenting the artistic traditions of Asia from ancient to modern times. The collection emphasizes Buddhist material, but also includes significant holdings on Asian art in general. Currently, some 60,000 photographs are available online,<sup>24</sup> with the goal to have all remaining images online by 2020. In addition to the image database, the Archive’s website includes online exhibitions and educational materials on Asia and Asian art, including useful maps. The database is currently transitioning its metadata to the VRA Core (Ver. 4) standard that will help to unify the Archive’s terminology and classification system.

The data is so far not published under an open license, but made available to researchers online without charge. To reproduce the images in publications, researchers still need permission from the archive and fees might apply.

The *Tibetan and Himalayan Library* (THL) was begun in 2000 under the leadership of David Germano and was one of the first large online collections of scholarly information about Tibet.<sup>25</sup> Today the THL is designed, according to its website, as “a publisher of websites, information services, and networking facilities relating to the Tibetan plateau and southern Himalayan regions.” Its interface provides access to a large collection of c. 70,000 photographs, audio and visual material, a map collection, and Tibetan language tools. Among these the “THL Tibetan to English Translation Tool” is especially noteworthy.<sup>26</sup>

---

<sup>23</sup> The Huntington Archive of Buddhist and Asian Art, <http://www.huntingtonarchive.org/>.

<sup>24</sup> The Huntington Archive of Buddhist and Asian Art, <http://huntingtonarchive.org/database.php>.

<sup>25</sup> The Tibetan & Himalayan Library, <http://www.thlib.org/>.

<sup>26</sup> “THL Tibetan to English Translation Tool,” The Tibetan & Himalayan Library, <http://www.thlib.org/reference/dictionaries/tibetan-dictionary/translate.php>.



Tied to the canonical catalogs are helpful bibliographies of secondary literature for many texts in the canon. The THL is designed as online library and offers no distributable data. The emphasis is on linking historical and textual information to be accessible online in the THL interface.

Another project that provides image data is the *International Dunhuang Project* (IDP) that aims at making the richness of Central Asian manuscripts available. These manuscripts are indispensable for the study of late Indian, medieval Chinese, and early Tibetan Buddhism. IDP was established in 1994 to coordinate international teams of conservators, catalogers, researchers and digitization professionals to ensure the preservation of the Eastern Silk Road collections and to make them freely available online. Hosted by the British Library, IDP has brought together collections and stakeholders from the UK, France, Germany, Russia, China and Japan. Although not all collections have been fully digitized so far (about 30 % of the Stein collection remains unscanned), and not all that is digitized is released, much has been made available and is distributed via the IDP website.<sup>27</sup> IDP currently offers access to over half a million images of over 100,000 manuscripts, paintings, artifacts, and photographs.

Also focused on manuscripts is the *Digital Library of Lao Manuscripts*, which aims to preserve the rich heritage of Laotian Buddhist manuscripts (15<sup>th</sup> to 20<sup>th</sup> century).<sup>28</sup> It contains images of c. 12,000 texts, which are findable by title, ancillary term, language, script, category, material, location, and date via an exemplary faceted search function. The data is the happy result of the Preservation of Lao Manuscripts Programme of the Lao Ministry of Information & Culture, which was supported by the German Ministry of Foreign Affairs from 1992 until 2004. According to the website the criteria for the selection for microfilming were “historico-cultural importance, cultural diversity or regional representation, age (all manuscripts over 150 years old) and quality of the manuscript. Within these general guidelines, priority for microfilming was given to extra-canonical literature, all manuscripts which were thought to represent indigenous literary traditions, and all texts of a non-religious nature.”<sup>29</sup> The texts were originally preserved in microfilm format, but have been digitized and are now made available both

---

<sup>27</sup> International Dunhuang Project, <http://idp.bl.uk>. Single collections can sometimes be accessed through other portals as well. E.g. the Pelliot collection has been made available through Gallica, [gallica.org](http://gallica.org). The Berlin Turfan Collection can be found at: Digitales Turfan Archiv, <http://turfan.bbaw.de/dta/>.

<sup>28</sup> Digital Library of Lao Manuscripts, <http://www.laomanuscripts.net>.

<sup>29</sup> “About DLLM: Background,” Digital Library of Lao Manuscripts, <http://www.laomanuscripts.net/en/about/background> (Sept. 2020).

via an online interface and packaged with professional metadata for download in pdf format.

In Taiwan, there are the projects conducted at the *Library and Information Center of the Dharma Drum Institute of Liberal Arts*. Over the last fifteen years a steady series of some twenty projects have produced open data on various aspects of Buddhist culture.<sup>30</sup> To mention only five:

- Among the larger projects at the Library and Information Center was a visualization platform for *Gaoseng zhuan* collections, i.e. Buddhist biographical literature, that allowed users to explore the information in different views, e.g. on a map or as social network. The social network information that was produced during this project is the largest of its kind in Buddhist Studies.<sup>31</sup>
- The *Digital Archive of Buddhist Temple Gazetteers* resulted in the full-text digitization of some 250 local histories of Buddhist temples, which are important sources for the study of Buddhism in late imperial China.<sup>32</sup> Data & metadata is published in the form of METS archives, the full text is encoded in XML-TEI.
- The *Catalog Database of Republican Era Buddhist Journals* is an detailed online catalog of two large print collections of Buddhist periodicals published between 1912 and 1950. It allows for searches by topic and genre.
- *Buddhist Temples in Taiwan* is a geo-referenced dataset of c. 5,500 temple in Taiwan, which includes historical and religious information about most sites. Data & metadata is published in XML. An online interface visualizes the distribution of temples on a time line. The data is joined with the largest image database of temples on Taiwan.
- The *Buddhist Authority Databases* were designed to provide authority data for various projects at Dharma Drum that needed to disambiguate the names of persons and places, text titles, and East Asian calendar dates.<sup>33</sup>

---

**30** Digital Archive Project, [http://lic.dila.edu.tw/digital\\_archives\\_projects](http://lic.dila.edu.tw/digital_archives_projects).

**31** A follow-up project added figures from the Song, Yuan, Ming and Qing dynasties. The current size of the network is at c. 17,500 actors (data available at: Socnet Resources, <http://mbingenheimer.net/tools/socnet/>.)

**32** Digital Archive of Chinese Buddhist Temple Gazetteers, <http://buddhistinformatics.dila.edu.tw/fosizhi/>. Potential DH uses of the corpus are outlined in Marcus Bingenheimer, "The Digital Archive of Buddhist Temple Gazetteers and Named Entity Recognition (NER) in Classical Chinese," *Lingua Sinica* 1, no. 8 (November 2015): 1–19.

**33** Buddhist Studies Authority Database Project, <http://authority.dila.edu.tw/>. The calendar authority is described in Marcus Bingenheimer et al., "Modeling East Asian Calendars in an Open Source Authority Database," *International Journal of Humanities and Arts Computing* 10, no. 2 (September 2016): 127–144.

The data is made available in XML in packaged archives which are updated on a monthly basis.<sup>34</sup> There is also an open API for outside projects that wish to work with linked data. As of 2020, the Dharma Drum Buddhist Person Authority, which is continually developed and expanded, is the largest digital onomasticon for Buddhist Studies.

Next to long term projects that enjoy strong institutional support, sometimes outstanding data collections are provided by individual researchers. Special mention should be made of the well-curated *Ancient Buddhist Texts* of Venerable Ānandajoti, who has prepared a number of annotated digital editions of Pāli and Sanskrit texts.<sup>35</sup> All the material, which includes rare works from the grammatical tradition, is available via his website and downloadable in pdf, epub and mobi format. Value is added to many of the better known sūtra texts by the annotation. The material is well packaged, and, one would hope, will one day find its way into a long-term archive such as Zenodo.

Whether individual or institutional projects, long term sustainability is always an issue for digital resources; as is the creation of an environment for collaborative research.<sup>36</sup> Thankfully, data producers are increasingly aware of this and Dharma Drum, *Ancient Buddhist Texts*, *Suttacentral* and many others make archival packages of their data available. More and more open Buddhist data can be found archived on Github and other version controlled platforms. Such repositories open the possibility of new forms of collaborative research, e.g. the joint development of digital editions.

---

<sup>34</sup> The Place Name Authority contains a large number of entries originally created by the GIS team at Academia Sinica. This part of the data is not included in the downloadable archives.

<sup>35</sup> At: Ancient Buddhist Texts, <https://ancient-buddhist-texts.net>. Ānandajoti has also assembled a rich collection of over 13,000 photos of Buddhist art and sites at: Photo Dharma, <https://www.photodharma.net>.

<sup>36</sup> In the field of Buddhist Studies, we were e.g. lucky that the translations from the Pāli Canon collected on the *Access to Insight* webspace have been incorporated into the *Suttacentral* corpus (s.b.). *Access to Insight*, <https://accesstoinight.org>, was started 1993 by John Bullitt, but, although still online, the website is not longer maintained and since 2013 users are advised to download the Legacy Edition: “Download the Whole Website ATI Offline Edition,” *Access to Insight*, <https://accesstoinight.org/tech/download/bulk.html>. *Suttacentral* has also incorporated the comprehensive site of German translations of Pāli texts, which was created by Wolfgang Greger: *Tipitaka, der Palikanon die Lehre des Theravada*, <http://palikanon.de> and first went online in 1998.

One repository for premodern Chinese texts that is designed along those lines is *Kanripo*, developed since 2013 by Christian Wittern.<sup>37</sup> The repository combines texts produced in-house with texts collected from other projects and sources on the internet. It combines the large *Siku quanshu* 四庫全書 collections, which contain the output of 2,500 years of Confucian literati culture, the Daoist canon and the CBETA corpus, thus allowing to research terminology across genres and traditions.<sup>38</sup> The total number of texts in *Kanripo* is currently close to 10,000 individual items. Texts that are available multiple times in these collections are consolidated into one entry and, where available, digital facsimiles of the text are juxtaposed with the full text. The texts are released under a Creative Commons (BY-SA) license. The collection is working with users who want to add texts to the repository and works towards a “sinological common” that can provide reliable sources for research.<sup>39</sup>

Most of the resources mentioned so far concentrate on collecting and presenting texts in only one Buddhist language. However, in Buddhist Studies almost every text is a cluster of texts and research often involves comparing different translations. To prepare aligned text that assists with such comparisons is time consuming, and only a few multi-lingual digital editions have been attempted so far. The largest project is the Thesaurus Literaturae Buddhicae (TLB), which was developed by Jens Braarvig as part of the Bibliotheca Polyglotta.<sup>40</sup> The website presents dozens of multilingual Buddhist text clusters in Sanskrit, Tibetan, Chinese and English. The texts are chunked in (loosely defined) sentence or paragraph units. So far the data is limited to online use, one would hope that the dataset of linked text could one day be made available for download. The alignment of sentence-size chunks would be very helpful for e.g. the computational analysis of translation vocabulary.

For a Chinese Āgama text (T.100) a detailed, aligned TEI edition with all Chinese, Pāli, Sanskrit and Tibetan parallels has been prepared for a project at

---

37 *Kanripo* is short for “Kanseki 漢籍 Repository”. For information about the provenance of the digital corpora aggregated in *Kanripo*.

38 The project aims to collect all texts from the *Siku quanshu* 四庫全書, *Sibu congkan* 四部叢刊, *Zhengtong Daozang* 正統道藏, *Daozang jiyao* 道藏輯要 and the CBETA corpus 電子佛典集成. A few texts are still missing as of Sep. 2017.

39 “Kanseki Repository 漢籍リポジトリ,” GitHub, <https://www.kanripo.org>. <https://github.com/kanripo>. The texts can be queried and read at: Tipitaka, der Palikanon, die Lehre des Theravada, <https://www.kanripo.org>.

40 Bibliotheca Polyglotta, <https://www2.hf.uio.no/polyglotta/index.php?page=library&bid=2>.

Dharma Drum.<sup>41</sup> Similarly Chinese, Sanskrit and Tibetan versions of the Yogācārabhūmi are available in a dedicated interface.<sup>42</sup>

## Digital Tools for Scholarship

After the digitization of primary sources, the other obvious target of digitization were research tools such as dictionaries, catalogs, and bibliographies, which could be modeled relatively easily as databases.

### Lexicography

The multilingual character of the Buddhist tradition has resulted in a large number of dictionaries, glossaries and encyclopedias. Lexicography is not merely an concern for modern Buddhist Studies, but has been part of Buddhist scholasticism for centuries.<sup>43</sup>

While, on the one hand, dictionaries originally designed for print have been digitized,<sup>44</sup> the most widely used online dictionary of Eastern Buddhism is the “digital-native” *Digital Dictionary of Buddhism* (DDB), an original and innovative creation by Charles Muller.<sup>45</sup> Started as a private dictionary lookup tool in the late 1980s, Muller took the DDB online in the 1990s making it one of the earliest surviving online tools for Buddhist Studies. Conceived of as a collaborative project, contributions of various sizes and types have since been made by over 300 scholars. For each Chinese lemma, the DDB provides an array of definitions, which are individually credited to their contributors. It offers pronunciations in Mandarin, Korean, Japanese and Vietnamese, as well as pointers to print dic-

---

41 T.100 別譯雜阿含 project, <http://buddhisticinformatics.dila.edu.tw/BZA/>.

42 Yogācārabhūmi Database, <http://ybh.chibs.edu.tw/>.

43 Norman dates the earliest extant work of Pāli lexicography, the *Abhidhānappadīpika*, to the late 12<sup>th</sup> century. For Buddhist Chinese, where the prolific translation and transliteration of Indian terms made glossaries indispensable, the earliest glossary is the *Fanfangyu* 翻梵語 (T.2130), at least parts of which can be dated to the 6<sup>th</sup> century. For Tibetan—and later Mongolian and Manchu—Buddhist texts the most influential glossary that unified translation practice was the *Mahāvīyutpatti*, compiled in the 8<sup>th</sup> to 9<sup>th</sup> century.

44 Notable for its wide availability on different platforms is e.g. the *Foguang dacidian* 佛光大辭典 (first printed 1988).

45 Digital Dictionary of Buddhism, <http://www.buddhism-dict.net/ddb>.