International Journal of Digital Humanities (2023) 4:1–4
https://doi.org/10.1007/s42803-022-00064-5

INTRODUCTION

Special issue on digital humanities and East Asian studies

Alíz Horváth[1] · **Hilde De Weerdt**[2]

Published online: 8 February 2023

In 2020 we launched a broad call for contributions on East Asian digital humanities, covering topics from theoretical discussions to applied papers on digital methods and research. Based on the state of a diversifying field we envisioned potential contributions on OCR and digitization, text analysis of East Asian scripts, image and visual analysis, GIS, network analysis, topic modeling, East Asian digital infrastructure development, digital libraries and archives, software development, curricular developments in East Asian digital humanities, and collaborations in digital research. Excepting explicit treatment of curricular developments, the resulting special issue offers a window into the multifarious ways in which digital humanities in the field of East Asian Studies has developed in recent years.

Compared with the heavy emphasis on the development of digital text corpora in the last decade of the twentieth century, the achievements in digital East Asian Studies in recent years broadly fall within six areas.[1] First, the early investment in textual, biographical, and geographical databases has been maintained and extended. Second, researchers and research organizations across East Asia, Europe, and the Americas have developed a broad range of research-focused tools and platforms. Following up on the development of tools and platforms for text and image annotation, text analysis, and data visualization, there has also been significant collaboration in the conceptualization and design of interoperable and integrated platforms, with individual platforms borrowing from and linking to each other. Fourth, and perhaps most noticeably, there has been a rapid expansion of East Asian digital research. Across East Asia but also elsewhere there are now regular events, conferences, workshops, and summer schools; DH centers and programs have grown rapidly, especially in East Asia. This growing interest has also brought about a proliferation of East Asian

[1] For a more extensive discussion of these points, see De Weerdt (2021) For more extensive overviews, see Vierthaler (2020), Horvath (2022), Nagasaki (2019), Cha (2018).

✉ Alíz Horváth
horvath.aliz@btk.elte.hu

Hilde De Weerdt
hilde.deweerdt@kuleuven.be

[1] Eötvös Loránd University, Budapest, Hungary

[2] KU Leuven, Leuven, Belgium

language journals in Digital Humanities, English-language journals on digital Asia, special issues in various disciplines (history, literature, art history, religions, communication and media, etc.) and on methodologies (e.g., network analysis) in digital East Asian Studies. In addition, DH journal editors, including the editors of *IJDH*, conscious of the digital monolingualism that characterizes much of the work in DH, have made space for multilingual and non-European language research. These trends suggest that we are also seeing a diversification of East Asian digital research. Even though there is, especially in humanities research, still a focus on more computational, quantitative, and hypothesis-driven analyses in digital East Asian Studies, we are also seeing more explicit engagements with critical and theoretical work on the digital, methodological reflection, and experimentation with new formats and innovative forms of data analysis. Sixth, and finally, digital East Asian Studies is characterized by a strong sense of community. A plethora of social media groups have sprung up on Facebook, Wechat, H-Commons, and Twitter; and students and researchers have shared and discussed finished work and research in progress in podcasts, Wikis, Googledocs, and blogs.

Despite the steady growth and diversification in digital East Asian Studies, it also evident that significant challenges remain. Many of these challenges, organizational, technical, epistemological or governmental, are shared, as will also become evident from the papers in this volume. Others such as the distinctive features of East Asian languages and scripts, and different regimes of Internet governance, can also be best discussed in a comparative framework. The papers assembled here further highlight some of the fault lines within digital East Asian Studies, a term that may to a large extent reflect assumptions about the coherence of area studies to external observers, or, within East Asian or Asian Studies departments, tends to express aspirations rather than teaching or research practices. Digital Humanities in East Asian Studies has mostly been defined by the kind of methodological nationalism that is broadly characteristic of area studies. In East Asia too DH focuses on the national scene and the main standard languages: Chinese, Japanese, and Korean.

Within the limits imposed by the submissions received, we have sought to reflect the diversification of digital research on East Asia, and the variety of argumentative styles and presentation formats. Because we strongly believe that data transparency is key in published research, and because promises to respond to data enquiries have been proven to be largely ineffective,[2] we have repeatedly asked all contributors to deposit all data used in their papers. We regret that not all authors had followed up on this by the time of publication.

We have grouped the articles of the special issue into subsections based on the ways they approach and use digital methods and conceptualize the role of the digital in their respective work. In this vein, readers will first encounter an essay on the thought processes and reversals in designing a digital art history project and research articles in Chinese and Korean social and cultural history. Stephen Whiteman's

[2] A recent survey in the biomedical sciences showed that "more than 90% of corresponding authors either declined or did not respond to requests for raw data." This matches our own experience. All data requests that De Weerdt has submitted to authors providing an email address for such requests have gone unanswered. Watson (2022)

article critically reflects on digital research methods and widely used digital corpora and discusses how such methods can bring together fragmentary, disparate, and unconventional sources in the history of late imperial Chinese landscape architecture. Subsequently, Shang Wenyi and Ted Underwood track shifts in the geographical distribution of civil service examination participants in late Tang China, arguing for a correlation between examination success in different regions and the sociopolitical status of their places of origin. On the other hand, Hu Jing introduces a novel approach to the study of early modern Sino-Korean relations by mapping traders and goods captured in diplomatic travelogues using MARKUS, a digital platform targeting scholars working on Chinese and Korean sources. Her article is followed by Wang Changsong et al.'s contribution which explores the environmental, literary, and aesthetic aspects of spatial patterns in the Yellow River region.

DH can also play a crucial role in the preservation of cultural and linguistic heritage. This aspect is of particular importance in the case of endangered, minority, and historical languages and non-Latin scripts which are, for the most part, severely under-represented and under-resourced in DH. Through this special issue, the *International Journal of Digital Humanities* aims to give more visibility to scholarly projects on the challenging reconstruction and preservation of historical or minority scripts and languages. Xu Duoduo's contribution constitutes a rare attempt of critically handling Dongba pictographs in a digital context. The fact that the Dongba characters in Xu's article had to be inserted as images by the publisher sheds light on the prevailing realities of many scholars working with non-Latin scripts. Jeffrey Tharsen, on the other hand, builds upon his self-created *Digital Etymological Dictionary of Old Chinese* and combines it with visualization tools and sound files to create a multimodal experience in order to (re-)create multilingual phonetic and phonorhetorical patterns. Tharsen's paper also demonstrates how old Chinese can potentially impact the digital study of other, often better-resourced languages.

The following set of papers in the special issue provides space for articles that either present the technical aspects of creating tools or critically analyze existing research infrastructures in East Asian studies. First, Matthias Arnold et al. report on the challenges of data reproducibility in text databases. They address the lessons learned from the experience of working with existing digitized Chinese newspaper collections by assessing their performance and degree of accessibility through the consistent application of FAIR principles.

Local gazetteers (*difangzhi* 地方志) constitute an important and multifaceted source base, particularly in Chinese historical studies. Keen scholarly interest in these materials fueled the creation of a versatile digital platform, LoGaRT (*Local Gazetteers Research Tools*), to facilitate the access and large-scale digital exploration of these works. The following two articles in the special issue reflect the importance of local gazetteers in the existing scholarship through the critical and more technical analysis of LoGaRT. The contribution of Chen Shih-Pei et al., the creators of LoGaRT, revisits the capabilities and challenges of working with local gazetteers digitally by combining their technical insights with concrete case studies. Their insights reflect not only on the realities and future horizons of LoGaRT, but also on the digital analysis of the local gazetteer genre itself. Subsequently, Liu Zhou et al. propose a novel technique in model training with BERT to increase tagging efficiency using data from the Chinese

Biographical Database (CBDB) and LoGaRT. Considering the existing challenges of integrating Natural Language Processing (NLP) into projects based on non-English texts and non-Latin scripts,[3] this paper offers a promising avenue towards improved results, particularly in Named Entity Recognition (NER) for Chinese.

Finally, the issue concludes with a review article by Alíz Horváth on recent developments in digital Japanese studies. Her contribution complements the predominantly China- and Korea-related articles presented here by focusing on various forms and interpretations of what it means to work digitally in the context of Japanese humanities.

References

Cha, J. (2018). Digital Korean studies: Recent advances and new Frontiers. *Digital Library Perspectives, 34*(3), 227–244. https://doi.org/10.1108/DLP-04-2018-0013

De Weerdt, H. (2021). Key Challenges in Digital Chinese Studies. In *Digital East Asian Studies, Digital Modern Languages Seminar*, King's College London. Web. https://www.youtube.com/watch?v=51D3C-3XHZQ Accessed 10 Nov 2022.

Dombrowski, Q. (2020). Preparing non-English texts for computational analysis. *Modern Languages Open, 1*(45), 1–19. https://doi.org/10.3828/mlo.v0i0.294

Horvath, A. (2022). Digital brush talk: Challenges and potential connections in east Asian digital research. In D. Fiormonte, S. Chaudhuri, & P. Ricaurte (Eds.), *Global debates in the digital humanities* (pp. 127–140). University of Minnesota Press.

Kung, S. (2020). *Chinese Natural Language (Pre)processing: An Introduction*. Web. https://towardsdatascience.com/chinese-natural-language-pre-processing-an-introduction-995d16c2705f Accessed 15 Nov 2022.

Nagasaki, K. (2019). Contexts of digital humanities in Japan. In S.-H. R. Wong, H.-P. Li, & M. Chou (Eds.), *Digital humanities and scholarly research trends in the Asia-Pacific* (pp. 71–90). IGI Global, Information Science Reference.

Princeton University, DARIAH, Haverford College. (2021). *New Languages for NLP: Building Linguistic Diversity in Digital Humanities*. Web. https://newnlp.princeton.edu/ Accessed 5 Nov 2022.

Shalaan, K., Sidiqqui, S., Alkhatib, M., & Monem, A. A. (2018). Challenges in Arabic natural languages processing. In N. El Gayar & C. Y. Suen (Eds.), *Computational linguistics, speech and image processing for Arabic language* (pp. 59–83). World Scientific.

Vierthaler, P. (2020). Digital humanities and east Asian studies in 2020. *History Compass, 18*(11), 1–16. https://doi.org/10.1111/hic3.12628

Watson, C. (2022). Many researchers say they'll share data — But don't. *Nature, 606*, 853. https://doi.org/10.1038/d41586-022-01692-1

Yasuoka, K., Wittern C., Morioka, T., Ikeda, T., Yamazaki, N., Nikaido, Y., Suzuki, S., Moro, S., & Fujita, K. (2022). 古典中国語(漢文)Universal Dependenciesとその応用. [Koten Chūgokugo (Kanbun) Universal Dependences to sono ōyō]. 情報処理学会論文誌, *63*(2), 355–363. https://doi.org/10.20729/00216242.

[3] Many of these challenges are related to the different stages of digital text analysis, such as problems with OCR accuracy, proper character display, annotation and tokenization issues, linguistic differences between English and other languages etc. These could be considered preliminary steps but their accuracy (or the lack of it) can significantly affect the outcome of further analysis. These problems manifest themselves even more seriously in the case of non-modern and right-to-left scripts. For further details on these challenges, see for example Dombrowski (2020), Shalaan et al. (2018), Kung (2020), Yasuoka et al. (2022). In addition, the NEH-funded New Languages for NLP: Building Linguistic Diversity in Digital Humanities project, organized by Princeton University, DARIAH, and Haverford College (2021–2022) also aimed at increasing linguistic diversity in NLP (since such resources currently only support 85 languages) through the collaborative work of ten teams with a focus on classical Arabic, Kanbun, old Chinese, Quechua, Yoruba-Efik, Yiddish, Kannada, Ottoman Turkish, literary Russian, and Tigrinya. See https://newnlp.princeton.edu/.