

FY2021 Annual Report for International Joint Research
International Joint Digital Archiving Center for Japanese Art and Culture (ARC-iJAC),
Art Research Center, Ritsumeikan University

[International Joint Research to Utilize the Center's Facilities and Equipment]

Date (2023/01/04):

1. Title of the research project	
“Natural Language Processing for a Geospatial Exploration of Japanese <i>Ukiyo-e</i> Prints”	
2. Research leader	
Name	Organization and title
Ewa Machotka	Stockholm University, Dept. of Asian and Middle Eastern Studies, Associate Professor
3. Co-researcher (Total : 5 persons)	
Name	Organization and title
Ryo Akama	Ritsumeikan University, Professor
John Pavlopoulos	Stockholm University,, Affiliated Researcher
Panagiotis Papapetrou	Stockholm University, Professor
Konstantina Liagkou	Athens Technology Center, Research Assistant
Marita Chatzipanagiotou (member until December 2021)	Athens University of Economics and Business, Research Assistant

4. Overview of the research project (about 150 words). Note: If you have changed your project since the time of application submission, please write clearly where you made changes.

The proposed project aims at a large-scale digital geospatial exploration of places depicted in Japanese early modern *ukiyo-e* landscape prints, and focuses on a dataset of prints accessible through the Ukiyo-e Portal Database hosted by the ARC. The first stage of the project, which is still ongoing, focuses on the extraction of information on the depicted locations by applying the Natural Language Processing (NLP) technology (e.g. Named Entity Recognition) on their meta-data; and their subsequent geolocation on the map of Japan in order to identify, map and contextualize the frequency and distribution of places in the cultural history of the first half of the 19th c. The second stage of the project, which is planned in the near future, will concern visual analysis of the images performed with the use of neural encoders (pre-trained to classify/caption generic images) to explore feasibility of iconographical identification of these places and involved challenges related to artistic mediation linked to the process of their visual representation. Overall, these analyses will generate a new information on geo-temporal distribution of the culturally significant sites and their visual depiction contributing to a better understanding of Japanese landscape prints in general.

5. Overview of the research results. Note: We may use this section for the Center's PR.

The first stage of our exploratory mixed-method analysis of the places depicted in Japanese early modern prints featuring *meisho* (lit. famous places) has so far delivered promising results as well as opened up new methodological and conceptual pathways to realize our overall research goals aiming at a large-scale geospatial study of landscape images in Japan. First, by employing the NLP approaches such as transfer-learning and Named Entity Recognition (NER) and applying our fine-tuned recognition model on a large dataset of prints, we provided a use-case of how a macroanalysis (or so-called 'distant viewing') of a visual dataset can be successfully undertaken in the context of art historical research of Japanese early modern visual culture. We also identified a number of methodological challenges related to this mixed-method analysis rooted in the complexities of the Japanese early-modern writing system, problems with adequate identification of historical place names, and specificities of pre-modern printed medium. Second, we developed "Ukiyo-e Distant Viewer", an online application for geolocation of the automatically recognised places on the map of Japan, that enables exploration of frequency and spatial distribution of these places across the map and identification of culturally significant places at a given time period. Third, we used the visualised outcome to perform an error analysis and compared our ML-based NER with gazetteer-based approaches. Our analysis also opened up new research pathways via the application of historical gazetteers. Furthermore, we also experimented with Optical Character Recognition (OCR), in order to investigate the following hypothesis: could the automatically recognised (not human-transcribed) text be used to extract place named-entities in order to increase the scale of our study? Answering this research question has a wide significance (beyond our study) especially considering a large amount of Japanese early-modern textual materials (also pervading visual culture) that still remain un-transcribed, and their subsequent limited accessibility for a general public. Hence, with this study we outlined the path for future research that will bring us closer to the understanding of early-modern landscape images and Japanese culture at large.

6. Research activities Note: We may use this section for the Center's PR.

(1) Books

(2) Articles:

- "Distant Viewing of Ukiyo-e Prints." 2022. *The Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*: 5879–5888. Machotka E., Pavlopoulos J., and Liagkou K. Peer-reviewed.
- "Automated Recognition of Geographical Named Entities in Titles of Ukiyo-e prints." *Association for Computing Machinery* (2021): 70–77. Machotka, E. Pavlopoulos, J., and Chatzipanagiotou, M. Peer-reviewed.

(3) Presentations:

- "Mapping *meisho*: NLP for Japanese Art History", December 2022, International Conference "The Digital Turn in Early Modern Japanese Studies", University of Cambridge. Presenters: Machotka E., Pavlopoulos J., Liagkou K.
- "Mapping Japanese Landscape Prints with NLP: Challenges and Solutions", November 2022, International Workshop "Error Correcting HTR for Historical Manuscripts", Ca'Foscari University of Venice. Presenters: Machotka E., Pavlopoulos J., Liagkou K.
- "HTR Error Correction: An Overview of the Recent HTREC Challenge", November 2022. International Workshop "Error Correcting HTR for Historical Manuscripts", Ca'Foscari University of Venice. Presenter: Pavlopoulos J.
- "Improving Named-Entity Recognition on titles of *Ukiyo-e* prints: Towards a 'Distant Viewing' in Art History", July 2022, Digital Humanities Conference 2022 (ADHO) *Responding to Asian Diversity*, Tokyo. Presenters: Machotka, E. Pavlopoulos, J., Liagkou K. Peer-reviewed.
- "Distant Viewing of *Ukiyo-e* Prints." June 2022, The 13th Language Resources and Evaluation Conference (LREC), Marseilles. Presenters: Machotka E., Pavlopoulos J. Liagkou K. Peer-reviewed.
- "Automated Harvesting of Geospatial Data from *Ukiyo-e* Prints with NLP", February 2022, Annual Report Meeting, The International Joint Digital Archiving Center for Japanese Art and Culture, Ritsumeikan University, Kyoto, Presenters: Machotka E., Pavlopoulos J.

- "Automated Recognition of Geographical Named Entities in Titles of *Ukiyo-e* prints", December 2021, Digital Humanities Workshop (DHW 2021), Kiev, Presenters: Machotka, E. Pavlopoulos, J., Chatzipanagiotou, M. Peer-reviewed.

(4) Symposiums and/or research meeting you organized:

- "International Workshop "Error Correcting HTR for Historical Manuscripts", Ca'Foscari University of Venice, November 2022, Participants no.: 16 speakers. Organizers and chairs: Machotka E. and Pavlopoulos J., Organizing Institutions: Center for Digital and Public Humanities at the Ca'Foscari University of Venice, and Stockholm University.

(5) Other research activities (Lectures to the general public, and appearances in/contributions to mass media)

(6) Academic awards

(7) Grants-in-Aid for Scientific Research -KAKENHI

(8) Competitive grants other than KAKENHI

- "Strategic Research Funding" of the Board of Human Sciences, Stockholm University

(9) Other achievements