

## Digital Resources



### East Asia Resource Library in Slovenia: Open a matchbox and access global knowledge of the region

Zlatko Šabič and Mirjam Kotar

In September 2017 a delegation from a Chinese institution came to visit the location of the East Asia Resource Library (EARL), which is hosted by the University of Ljubljana's Faculty of Social Sciences, in Slovenia. This is a modern, spacious building, home to about 3,000 students, researchers and faculty and a venue for countless events at all levels. The delegation could hardly hide their surprise when they realised that the EARL was made up of three rooms which barely exceed 200 square meters. They thought that the entire Faculty of Social Sciences was the 'Library'.

We, too, were taken a bit by surprise by such an observation, but we also took a lesson from that impression. Since then, we never fail to explain to our guests that the word 'space' has different meanings in different contexts. In Slovenia, which given its geographical determinants belongs to a group of smaller countries on this planet, physical size does not matter much. There are many determinants that relativise it. What follows is a story that proves this.

In November 2015, a group of four scholars –synologists, japonologists and experts in

international relations from the University of Ljubljana –began to meet regularly to reflect on the academic landscape in Slovenia and the neighbouring countries with regard to studying the East Asian region. They agreed that there were several study programmes devoted to East Asia in the region, however, they have comparatively speaking relatively modest support in literature and in primary sources. To access these (re)sources, students and experts need to travel to East Asia. So the group started wondering: if all these resources are difficult to reach, why not bring them closer to us?

With technological advances, the idea how to make 'smallness' irrelevant seemed doable. After all, digital databases already cover most of secondary sources, and digitalisation of primary sources is progressing rapidly. Hence, with the support of experts on digital databases, a concept of a regional hub for resources devoted to East Asia started to take shape. By the end of 2015, work on the EARL began. Its mission has been embraced and supported by the University of Ljubljana's Faculty of Arts and Faculty of Social Sciences. Both faculties are large teaching and research institutions (<https://www.fdv.uni-lj.si/en/home>; and <http://www.ff.uni-lj.si/an>). On 17 May 2016, the Deans of the two Faculties signed an agreement to establish the East Asia Resource Library – EARL (<https://www.fdv.uni-lj.si/en/library/earl>).

The EARL was structured as follows. Symbolically, the EARL signified a region. It provided designated spaces for each of the participating institutions, which are also called 'corners'. Officially, the 'corners' are named as follows (in alphabetical order): China Corner Reading Beijing, Japan Corner, Korea Corner, and Taiwan Resource Centre for Chinese Studies. Each section is organised differently. Embassies from Korea and Japan in Slovenia serve as facilitators of information and contacts with the Japan Foundation and the Korea Foundation that provide the financial support for electronic databases, books and other resources. The EARL also collaborates extensively with the Capital Library of China, with the Taiwan Resource Centre for Chinese Studies and various nonprofit institutions from East Asia. All these institutions contribute to the unique concentration of knowledge about the region. EARL offers access to several East Asian

databases, to over 3,000 titles of printed books and other materials and is also a lively social place, where students, scholars, practitioners and interested public from the region meet.

The EARL is yet another embodiment of the conviction that Humanities and Social Sciences must go hand in hand to understand what is going on around us and how we got there. Interdisciplinarity matters. One might have the most detailed knowledge about politics or social fabric of any corner in this world; yet, if one does not understand the culture and speak the language people speak in one part of the (East Asian) region or the other, one shall never have a full understanding of it. In modern social sciences and humanities, it is safe to say that this kind of thinking has not yet been internalised. The EARL is built on the premise that the 'wall' between social sciences and humanities is not really high if we use the right ladder, the one which is built from the wealth of interdisciplinary knowledge and experience of scholars coming out from the two branches of science.

In conclusion, let us go back to the reflection on size. We see the EARL not in terms of the small, landlocked room, but, in terms what it *really* offers. It is a large port, with access to the open sea of resources, going into millions of books, articles and historical documents. In international relations, Open Sea signifies a space with no borders. In our world of science and teaching, knowledge is our ocean. This is why EARL's motto is: *knowledge knows no borders*. It is in this vast space that people with knowledge meet, compare notes, and discuss. With this in mind the EARL hosts presentations, lectures, gallery, cultural events, and also academic, professional and unofficial meetings. A very special value of the EARL is the unreserved commitment of our partners to provide the EARL with the very best the region as a whole can offer. In trying times like the ones we live in today, access to resources and a possibility to compare them is the only way forward to bridge irrational political differences and discuss the region with only one vision in mind: to assure and sustain lasting peace.

Zlatko Šabič Director General of the East Asia Resource Library (EARL).  
Mirjam Kotar Head of the Central Social Library, Chief Coordinator of the EARL.

### Naval Kishore Press – digital: From hidden treasure to open access

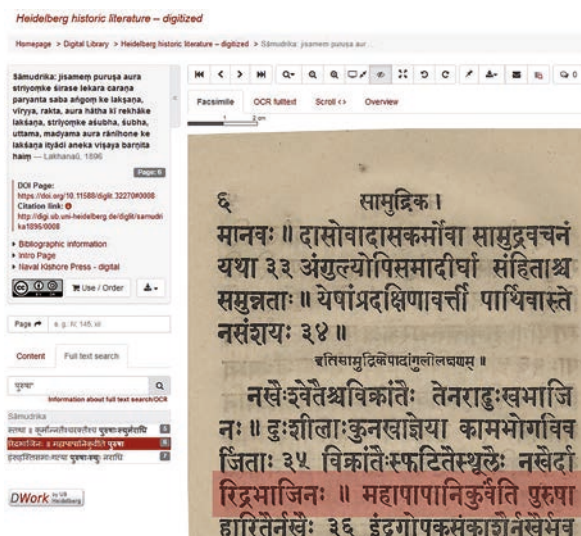
Nicole Merkel-Hilf

The Naval Kishore Press was established in the north Indian city of Lakhnau in 1858 by Munshi Naval Kishore (1836-1895). In the following decades it grew to one of India's most important publishing houses. During Naval Kishore's lifetime the press published around 5,000 titles covering literature in Hindi, Urdu, Arabic, Persian and Sanskrit on subjects as diverse as religion, education, medicine, school-books, popular editions of Sanskrit literature, and much more. The library of the South Asia Institute (SAI) at Heidelberg University holds a representative cross section of the Naval Kishore Press' publications with 1,400 titles in print and around 700 titles on microfilm.

In order to make this treasure more visible for scholars the Naval Kishore Press Bibliography has been set up by using the open source software VuFind. The bibliography is intended as a provenance database and aims to provide access to bibliographic records as well as digitized online editions of works issued by the Naval Kishore Press that are distributed in libraries worldwide –and not only to the SAI library collection. Currently we are enriching the bibliography with 1,200 title records from the Bodleian Library in Oxford. The bibliography will then contain more than 3,500 entries from eight different libraries.

From the mid-19th century onwards wood pulp paper was used for printing which tends to be acidic and therefore paper deterioration is a problem for the printed part of the collection. For reasons of preservation the Naval Kishore Press –digital project was initiated by the SAI library and Heidelberg University Library.<sup>1</sup> Within this project, selected Hindi and Sanskrit titles in Devanagari script from the Naval Kishore Press collection are digitized, but the primary aim of Naval Kishore Press –digital is to offer scholars more than a digitized image facsimile. The goal is to produce machine-readable texts that can be further edited online by using digital editing techniques.

Suitable OCR software especially for South Asian scripts has long been unavailable due to the complexity of the writing systems and has turned out often to be unsuitable for mass digitization projects. For the Naval Kishore Press –digital project two text recognition methods have been used –the OCR software for Hindi and Sanskrit developed by ind.senz and, more recently, a data model trained by Transkribus.<sup>2</sup> For the training of the model 200 pages of a so-called 'ground truth' transcription was produced, i.e. an accurate representation of the text on the image facsimile. The ground truth transcription and the images are then used



to train a recurrent neural network to get a data model to automatically transcribe more texts from the Naval Kishore Press collection. With an error rate of 5,59% on a random test set the results are very promising and we are using the model now on the digitized Hindi and Sanskrit texts of the Naval Kishore Press collection.

For the web presentation of the digitized images and the OCRed full-texts created with Transkribus the software 'DWork – Heidelberg Digitization Workflow' is used, an in-house development by Heidelberg University Library. It provides a variety of functions for the use of digital copies, such as thumbnail overview, zooming in and out, full text search, and various navigation features as well as components for annotations.

Words or phrases from the Hindi and Sanskrit texts can be searched in Devanagari script or in Latin transliteration and the results are highlighted in the image facsimile as well

as the recognized text. Furthermore, users can download a high quality OCR-PDF of the facsimile from the project website where the text is also fully searchable in both scripts.

The annotation tool implemented in DWork allows scholars worldwide to work collaboratively on a text or text corpus independent of place and time. Each annotation can be entered comfortably via a web form, is provided with the name of its author and can be reliably referenced and quoted by being assigned a DOI. Revisions are saved as new versions, while earlier versions remain still visible and can be accessed through the revision history.

Both resources can be accessed on CrossAsia:  
<https://themen.crossasia.org>

Nicole Merkel-Hilf Chief coordinator "South Asia" FID Asien, SAI Library.

#### Notes

- 1 Naval Kishore Press –digital is part of a larger, three-year project "Fachinformationsdienst Asien" (FID Asien), funded by the Deutsche Forschungsgemeinschaft (DFG) until the end of 2018. The FID Asien project is cooperatively carried out by the State Library in Berlin, Heidelberg University Library and the South Asia Institute. The web portal CrossAsia is used as the central access point to the project results and for scientific information in Asian studies (<https://crossasia.org/en>).
- 2 <https://transkribus.eu/Transkribus>