

# Open-access platform of KU-ORCAS, Past, present and future

OKUMURA Kayoko

KU-ORCAS のオープンプラットフォーム—過去・現在・未来

奥村佳代子

KU-ORCAS は、泊園書院から東西学術研究所へと受け継がれた東アジア研究の流れを汲み、より開かれた研究拠点を目指して組織された。とりわけ5年にわたる研究活動の中心にあったのは、研究リソースのオープンであったと言えるだろう。関西大学総合図書館の個人文庫、東西学術研究所には東アジア関連の貴重な資料が豊富に所蔵されているが、構成員各自が収集してきた特徴的な資料も充実している。それらの画像公開に関しては、意図的な選別を避け幅広い分野とテーマを対象とした。総花的な印象を与えもするが、今後人文学の領域でもデジタル化が一層進むことを考えれば、可能な限り数多くの資料の画像を網羅することは正しい選択であったと言えるだろう。一方で、研究者はオープンにされた資料をどのように研究すべきだろうか。資料のあり方が変化するのに付随して、変化する研究のあり方に、人文学の分野に属する者としての関わり方があるはずである。

なお、本報告は、KU-ORCAS 全体の研究成果を踏まえて執筆したものである。特に、前センター長内田慶市関西大学名誉教授による概要説明や成果報告に依拠するところが大きい。

キーワード : openness, digital archives, East Asian cultural studies,  
digital humanities, Kansai University

## 1. Introduction

### 1.1 Background to KU-ORCAS

The Kansai University Open Research Center for Asian Studies, or KU-ORCAS for short, is an ongoing project for organizing the diverse and unique resources held at Kansai University for the purpose of creating digital archives for East Asian cultural studies and making them accessible, hoping to establish an internationally recognized research hub for this academic interest. This initiative was accepted to participate in the Private University Research Branding Project for the FY2017, funded by the Ministry of Education, Culture, Sports, Science and Technology of Japan, or MEXT.

East Asian studies at Kansai University has its roots in a private school, Hakuenshoin, that was established about 200 years ago during the Edo period. The school held a vast library on Chinese classics, which was donated to Kansai University in its entirety upon the school's closure in the mid-twentieth century. This private library gave the university the reason for establishing the Institute of Oriental and Occidental Studies. In this sense, the study of classical Chinese literature and philosophy is a fundamental part of the academic tradition of our university. The Institute of Oriental and Occidental Studies was established in 1951 and contributed to the development of East Asian studies at the university. Notable developments in this pursuit include the Center for the Study of Asian Cultures (CSAC), which operated between 2005 and 2009 as part of the MEXT Large-scale Academic Frontiers Project. The same initiative carried on between 2011 and 2015 (CSAC-II) through the MEXT Strategic Research Foundation Grant-aided Project for Private Universities. There was also the Global COE program by MEXT granted between 2007 and 2011. All these initiatives that went before became the foundation of KU-ORCAS, and it strives to be an “open” research institution.

## 1.2 An open institution versus siloed organization

We place a particular emphasis on the “open” research institution, and the reason for this—to develop an institution of this nature—was because we wanted to address the widespread issue of siloed organizations in academia.

Today, many institutions across the world pursue the development of digital archives, and yet there is a persistent problem of organizational barriers. The institutions themselves may create their own digital archives, but their resources are “closed” in their respective organizations due to the lack of lateral accessibility. No matter how rich and useful the information may be, the limited access to it means it serves far fewer people than it could otherwise, and the data that are rarely accessed will eventually be made obsolete and lost.

One of the underlying issues of this siloed information problem is the lack of international standards for data such as images.

To address this situation, an international framework for image data has recently been introduced. This is called the International Image Interoperability Framework, or IIIF (triple-I F). It facilitates the lateral accessibility to digitized image data across data-holding institutions.

This international web-content-sharing framework has been adopted by major academic and public institutions in many countries, including the Stanford University Library, the

British Library, the National Library of France, and the Bodleian Library of the University of Oxford. In Japan, also, it is adopted by a few institutions, such as the Center of Next-generation Humanities Development at the Graduate School of Humanities and Sociology, the University of Tokyo, the Kyoto University Library Network, the National Diet Library, and the National Institute of Japanese Literature. Following this trend, KU-ORCAS is digitizing the university's academic resources for China studies and Oriental studies in compliance with the IIF requirements and making the data accessible to the public.

This initiative has significant implications to the future of the studies in these areas considering that, in Japan, ICT has not yet been sufficiently deployed in these areas.

### 1.3 Issues concerning digitization

There are, however, issues in relation to pursuing this initiative, or at least some points to be considered.

I have just mentioned that ICT is not fully leveraged in the Chinese and Oriental studies in Japan, but this is not because researchers of these academic pursuits are old-fashioned in their orientations. It is because, for one thing, it requires so much effort to carry out digitization and maintain publicized data on a continued basis, which is a basic requirement for leveraging ICT. The difficulties extend to securing funding for the undertaking, and this in turn is a time-and-energy-consuming endeavor. We KU-ORCAS face this problem as well, and we are doing our best in this struggle.

While it is widely accepted that digital data about research materials and methodologies are useful, and therefore digitization should be encouraged, there seems to be a sentimental resistance, among some of us, based on an idea that one must fight for digitized data. This attitude unintentionally accepts the idea that data should be kept secluded and heavily guarded. It is therefore true that there are physical obstacles to the digital data accessibility, which limit what such data about research materials and methodologies could be used for. Consequently, this situation could be detrimental to the progress of the Chinese and Oriental studies in Japan.

There is another point to be considered. The advantages of digital technology are not only for research purposes. The technology makes access to knowledge far easier than before. I think many of us enjoyed the benefits of digital technology during the confinement under the coronavirus pandemic, like watching live concerts over the Internet. It is fun to watch live images and hear live sounds even though the experience is virtual. However, we need the Internet access to have this experience, and furthermore, the experience

itself is an entirely different affair from the thrills and excitement that can be had from a real experience.

Now, this perspective also applies to research activities. Imagine that you are studying paintings. Digital data of paintings are sufficient and useful to know what are illustrated in a painting. However, you cannot perceive its overall or general presence and ambiance from such data. The same can be said about text-based documents, as well. You could only understand the significance of their existence by having the actual articles in front of you. Therefore, we should explore ways in which digitization of data is complemented by enhanced preservation of, and open access to, actual resource materials.

## 2. KU-ORCAS today

KU-ORCAS is in its final year of the project, and I now describe where we are at.

When it started, KU-ORCAS had three research units. To be an “open” organization, we maintained the research units to be “flexible” by allowing the research groups and their members to be changeable, and it still is in this way today. It currently comprises four research units and a pilot unit which our members created at their own initiatives by archiving information of their own choices.

### 2.1 Research units and their themes

[Unit 1: Text-based cultural contact between the Orient and Occident]

This unit has developed a digital archive of the documents within the university archive in relation to the subject of language contact (dictionaries, language textbooks, missionaries’ reports, etc.), and the basic idea is to form an integrative archive by linking its entries with records and documents housed in institutions overseas, such as the British Library, the National Library of France, the Vatican Library, and the Harvard University Library.

[Unit 2: Academic traditions and networks in Osaka seen in the East Asian context]

This unit aimed to develop a comprehensive archive of documents primarily from the Hakuenshoin, which has the roots of Kansai University’s academic tradition. Their archive also includes visual data of the paintings by artists from pre-modern Osaka, mainly composed of the works included in the university collection, and others that are found elsewhere in Japan and abroad.

[Unit 3: Spaciotemporal perspective on antiquity Osaka/Kyoto: cities and monuments]

This unit undertook the digital archiving of the excavation records, archaeological

artifacts, and literature that were accumulated by the university's research initiatives on the ancient port-town of Nanba and political center Asuka. The unit also carried out investigative research on the tumuli from the Asuka period, and the results were presented in an exhibition. They intended to include these results in their digital archive. Another objective of the unit was to study ancient texts and maps, as well as drawings of temples, in relation to the suburban areas of Kyoto and the areas along the Yodo river in Osaka, to add to their digital archive.

[Unit 4: Turning Japanese classics into digital formats]

This unit aimed to popularize new studies of the Japanese classical literature Man'yōshū and research methods by creating an open resource about how to digitize classical texts. It is expected that the information will be applicable to other historical documents. The unit's objective is to support open science and contribute to the development of public humanities.

[Pilot unit: Digitization of East Asian cultural studies]

The pilot unit was formed by the members of the four units. The objective was to create digital archives other than those developed by the four units. These archives were organized by themes each researcher decided for themselves. There are several archives already published, and one of them is for documents related to Asian films.

The research units and groups aimed to develop open research infrastructure, and their activities have borne fruit in the following areas.

## 2.2 Outcomes of the five-year project

The major achievements attained through this five-year project are as follows:

- (1) To make research resources openly available, we have developed and published multiple digital archives, which amounted to more than 6,000 items during 2020. This is a 300% increase on the number of items in 2017. The number of accesses to these archives also increased by 130%.
- (2) To open up research institutions engaging in digital archiving, we linked up with 16 pioneering museums, libraries, and archival organizations across the world and made our intellectual resources freely available.
- (3) To create an open environment to share and discuss the know-how concerning the development of digital archives and their uses as well as the challenges involved in creating digital archives, we held more than 63 symposiums and colloquiums as platforms for this purpose. More than 4,000 people participated. We also translated

and published the guidelines on digital humanities that were available outside Japan.

- (4) As for the provision of open resources on research outcomes, we published various research projects, numbering 199 scholarly papers in 2019 alone. Some of these were presented at domestic and international conferences, and others were published in academic journals. The number of presentations was in hundreds, which was far more than the target we initially set to ourselves.

Overall, the project has been a great success, with a number of figures surpassing their initial expectations.

## 2.3 The archives

Now, I present some of the digital archives we have published.

### 2.3.1 CSAC digital archive continued and further developed

CSAC and CSAC-II, which operated before KU-ORCAS, produced a digital archive of various research projects. We leveraged this as one of our main archives and added more content.

Firstly, there is the Modern Chinese Literature database, launched in 2006 as part of a Grants-in-Aid project led by Professor Keiichi Uchida, former director of the Asia Open Research Center.

Approximately 400 items, mostly produced by Europeans during the 19th century, were digitized and organized to allow full-text search, cross-document concordance search, search using either simplified or traditional Chinese characters, as well as the aligned display of texts and images. We are confident that this has already made significant contributions to studies in humanities, but we intend to bring it to higher levels in the future. More will be described about it later.

### 2.3.2 Database of private-collection archives

Kansai University holds many private archives in relation to East Asia that were developed by individuals, and they have been digitized in parts for the first time through our project. These are unique materials, perhaps the only examples in the world, let alone in Japan. We hope they will remain useful resources for those who may need them as they can be accessed online. The following archives are only a few examples:

Naito archive (33500 books)

—Konan Naito (antique classical Chinese library)

Nagasawa archive (30497 books)

—Kikuya Nagasawa (Japanese and Chinese literature)

Nakamura archive (33491 books)

—Yukihiko Nakamura (Japanese texts)

Masuda archive (16184 books)

—Wataru Masuda (Lu Xun, “the eastward dissemination of Western learning”)

Yoshida archive (2479 books)

—Isaburo Yoshida (Asian diplomacy)

Kito archive (10309 books)

—Shingyo Deguchi (Japanese texts)

Hakuen archive (16954 books)

—the Fujisawas: Togai, Nangaku, Kokoku and Koha

So far, approximately 6,000 copies of published books have been digitized, of which about 3,000 copies are made accessible online.

Meanwhile, the university recently acquired the Masuzawa archive, which is an invaluable source of knowledge in terms of Chinese studies and language learning in modern Japan. We are currently converting its content into a digital format as quickly as possible.

These private-collection archives include many non-published texts, such as letters and calligraphic works, as well as non-text items. We also try to include these in our digital archives. Some of them have already been made accessible online.

These include the following:

Chinese sealing lutes (about 20 items)

Letter written by Konan Naito (about 8,000 letters: all digitally scanned)

Seals contained in the Hakuen archive (about 300 items) and

scrolls from the Naito archive as well as from the university’s rare article collection (about 700 items: ready for online publishing)

We have thus converted many documents and items concerning East Asia into the electronic format, covering a wide area of interest. Most of these resources were attained through various Chinese and Oriental studies at our university. Therefore, they can be

said to represent the history of the Chinese and Oriental studies at Kansai University.

## 2.4 Digitization in Asia today

Meanwhile, digitalization is advancing rapidly in Asia, as we have learned today from the presentations given by our colleagues from China and Hong Kong. We could say, the most remarkable progress is the large-scale the Scripta Sinica database of the Institute of History and Philology, Academia Sinica. Their efforts to develop digital resources started about 35 years ago, and the database has grown today to cover approximately 700 million characters, and the database is fully searchable.

In China, also, the China Academic Digital Associative Library (CADAL) started in 2001 as a national-level joint project of Zhejiang University and Chinese Academy of Engineering. Their archive contains 7 million copies of digitized literary resources. The Japanese National Diet Library and Kansai University attempted to join them and affiliate with their library, but this has not been realized. Today, the University of Tokyo is the only participant from Japan.

Apart from these, there is an international institution, the Pacific Rim Research Library Alliance, or PRRLA. This initiative is joined by thirty-three universities from Hong Kong, China, Taiwan, Macau, Australia, Canada, and the United States. Unfortunately, no active participation is made from Japan (except by Kansai University, which subscribed to the alliance when the former director of KU-ORCAS, Professor Uchida, was the chief librarian [of the university library]. However, our membership is only nominal today).

In Japan, Waseda University and the National Diet Library have been proactive in digitizing their libraries for a long time. Recently, notable progress has been made by others, such as the National Institute of Japanese Literature, with their project to develop a database of pre-modern Japanese works, and Kyoto University, Shimane University, and so on.

Meanwhile, Europe and America are quite advanced in that many institutions have established digital archives of the historical documents of East Asia in their collections, as we all know. There is also a dedicated open-access online platform for Chinese philology and linguistics. As we have seen, digitization is advancing not only in the western part of the world, but also in Asia.

Our KU-ORCAS is now coming to a close as a project, but we are hopeful because we can learn from many good practices in the world to consider how we can make the most of what we have developed and what kind of system we should design in the future.



### 3. “New” KU-ORCAS in the future—digital archives and studies in humanities

KU-ORCAS will come to an end in three months. We must now consider how to make use of this body of digital resources.

In terms of research, we know for certain that humanities research cannot proceed without involving digital technology. We need therefore to explore how to leverage the digital archives in our research activities. We need to look for new research methodologies and apply them in our future studies.

As I mentioned earlier, digital resources are becoming a major trend in humanities research today. This trend does not stop at serving for scholarly work, but it also has implications to the preservation of knowledge passed on through history. The move toward the digital conversion of resources is perfectly in line with our motto, “私蔵は死蔵”, “A private archive nobody has access to is the same as the absence.” We believe that historical documents must be accessible to all, thereby they can serve their true purposes.

This is not to say that we should be content with developing digital archives. From researchers’ perspective, those archives must be relevant to new research methodologies.

Japanese and Chinese are particularly difficult languages in terms of handling digitized text documents, because of the lack of word dividers, which makes word separation particularly a challenging task. However, text mining technology is expected to make a significant improvement in this problem. Furthermore, in China, the rapid advancement in the studies of natural language processing is leading to solutions of many difficulties that could not be solved before.

For example, the progress of Chinese lexical studies used to amount to the number of hand-written lexicon cards, which correlated with the volume of reading done for the studies. This style has changed significantly today thanks to new tools, such as corpus, concordance and full-text search as well as comprehensive vocabulary indexes.

These tools have also enabled collaborations with other fields of study, such as conceptual history and history of ideas. Enriching data with additional information makes diverse approaches possible.

For example, Google N-gram Viewer is useful for investigating keywords particular to certain historical periods.

Text mining allows us to extract valuable information from text databases to analyze, for example, word frequency, correlations of concurrences, occurrence tendency and chronological analysis by words and phrases.

Other than these, there is the Text Encoding Initiative, which helps to give

universality and permanence to digital text documents through the use of the XML format. This will allow, for example, to visually collate words.

These new research methods and progressive digitization do not mean to dismiss more traditional methods. However, it would be impossible to generate new knowledge without adopting them in the areas where traditional methods were unable to tackle problems, such as natural language processing and large volume text analysis. This being said, we have to remind ourselves of one issue when introducing those technology-aided methods, that is, to ensure that the digital text data must be accurate.

In the Chinese language, there are several writing systems, and the same word (or character) is written distinctly differently. The difference of written characters is less important when studying the language semantically, and so the texts may be adapted to one system or another as appropriate. However, differences in writing systems form a significant component in philological and linguistic analyses, let alone Chinese studies. In the Chinese context, the differences of some specific characters may indicate the periods in which the texts were written. If these differences were erased, the documents would permanently lose their value as study materials.

Although there are several different types of digital resources according to the purposes, we must ensure that digital texts are accurate and highly universal for handling. This is our task in the future. There are many things we must work on in order to develop an appropriate research environment, and the generation of digital data is certainly one of them. And this is a task for everyone, including myself and other colleagues who may not be a digital expert.

## Conclusion

KU-ORCAS has been a successful initiative to “liberate” digital resources and “address the problem of siloed information” as it set out to do.

We would like to continue the efforts to improve open research resources, open research groups, and open research know-how. We will strive to provide a research platform that is useful not only for those who are engaged in East Asian studies, but also the general public, students, and other stakeholders. It is our hope that, through our efforts, we will make contributions to the advancement of digital humanities in East Asian cultural studies. We will keep going to develop such digital archives and research platform that have a wide appeal, not only to some researchers, but to any people who are interested in the subject.

There is only so much that one university can do alone. It is important that we extend our network of collaboration with various institutions in the world and work together to build further knowledge in humanities that supports a knowledge society, which is becoming a part of our reality.