

Articles by NDL staff

## Welcome to the NDL Digital Library Cafe

Research and Development for Next-Generation Systems Office, Digital Information Distribution Division  
Digital Information Department

*This article is a translation of the article in Japanese of the same title in NDL Monthly Bulletin No. 671 (March 2017).*

On November 24 and 25, 2016, the National Diet Library (NDL) sponsored the first in a series of lectures called the "NDL Digital Library Cafe." Held at the Tokyo Main Library, this was the NDL's first use of the "science cafe" format to provide the general public with a clear and easy-to-understand presentation of recent research and the latest trends in digital libraries.

What is a "science cafe?" The name "science cafe" refers to events where the general public can discuss scientific topics with the researchers who study them. Science cafes are increasingly popular at universities and research facilities as means of facilitating communication between the general public and science professionals. The NDL Digital Library Cafe is a place to discuss a wide range of topics related to digital libraries.

The NDL provides the general public with access via the Internet to a variety of digital libraries, including the NDL Digital Collections. The utilization of these digital libraries—not only via the Internet but also on the library premises—for research, application development, and other objectives is expanding. The NDL Digital Library Cafe featured presentations by invited guests, who described two new ways that the NDL Digital Collections are being utilized and then discussed these innovations with an audience of roughly 20 participants.

### November 24 Open Data, Civic Technology, and the NDL Digital Collections: Developing websites for researching regional history using the NDL Digital Collections

The use of civic technology—which is primarily the use of information technology to enhance public participation in national and regional development—is now spreading throughout Japan in response to nationwide promotion of open data. Mr. Shusaku Higashi, director general of the Open Knowledge Foundation Japan, provided a basic explanation of civic technology and presented examples of how open data can be used by the general public. Mr. Takashi Koike, head of Midori IT Office, LLC, demonstrated how the NDL Digital Collections could be used to overlay the names of villages in the provinces of Musashi and Sagami during the late Edo period onto present-day maps and linking these names to documents on local history. He also described interesting discoveries he made and the difficulties he encountered while creating this map.



A map of villages in the provinces of Musashi and Sagami during the late Edo period

### Q&A with the guests

Q: What is the connection between open data and your research interests?

A: Data is just a means to an end. A passionate commitment to research interests is what is important! What is needed to enhance interaction is a liaison who can connect people with research interests to those who can provide relevant data.

### November 25 The NDL Digital Collections: The road to automated text digitization

The NDL Digital Collections contains approximately 350,000 modern Japanese books that were published during the time from the Meiji period to the middle Showa period. These books are available free of charge to the general public, but it is impossible to search these works by text, because the digitized materials are provided only in image file formats. Optical Character Recognition (OCR) software is designed to read text characters from digital images and convert those images to digital text files. At present, however, OCR technology is not yet capable of accurately reading digital images made from modern Japanese books. Mr. Kazuki Joe, professor of the Faculty of Science at Nara Women's University, is working to enhance OCR technology by collecting and analyzing image data of Japanese characters from the NDL Digital Collections. Mr. Joe described the background to his

research, the challenges he faces moving forward, and research products that incorporate artificial intelligence in the development of OCR.

#### Q&A with the guests

Q: How difficult is it for the software to recognize one character from another in modern printed matter?

A: As long as the document is in good condition, OCR is not difficult at all. The difficulties arise when the pages are creased or soiled, but these are issues for future study.

Q: I can well imagine that you need to collect a great many varieties of printed characters for this development. Are there any characters that appear only rarely in these historical documents?

A: The use of "deep learning," which is able to learn to anticipate the features of a particular character based on those of similar characters is one alternative method.

The NDL looks forward to continuing to introduce the future potential of digital libraries through the NDL Digital Library Cafe. We hope you will participate in the next one.



(Translated by Rie Watanabe)