



Documentation Center on Contemporary Japan (Reischauer Institute, Sponsor): Online Electronic Texts

Kazuko Sakaguchi
Documentation Center Director

Until recently, the dominant form of access to electronic texts (e-texts) has been offline media such as CD-ROMs. Nowadays, however, with the dramatic growth and development of the Internet and the increased power and capacity of personal computers, the popularity of online e-texts is growing rapidly. While re-writable CDs have become popular for personal data storage, read-only CD-ROMs as a form of access to e-texts have become gradually obsolete. The main reason is that this type of media cannot be updated, whereas online e-texts can easily be updated due to the vast storage capacity of large computer servers. Also, online e-texts do not require shipping and can supply as many users as needed instantaneously, and so are suitable for on-demand publication. Reprinting titles that are in short supply or out-of-print is easier as well. In addition, needless to say, online e-texts allow one to jump directly to references through high-powered links and to convert format into Braille or to voice with multi-media for the vision-impaired. Depending on the software, online e-texts allow one to enjoy a variety of readings and usage. Accordingly, I would like to take this opportunity to explore several databases for online e-texts, as this publication form is likely to increase in the near future.

Project Gutenberg (<http://prom.net/pg>)

The disclosure of information and the intellectual traditions held in the public domain are the foundations for democracy. Project Gutenberg, a pioneer of e-texts, was initiated by Michael Hart, then affiliated at the University of Illinois in 1971, based on his philosophy that such cultural heritage is common property for all people. Although the Internet was de-

veloped initially for military purposes, Project Gutenberg is an example of how the Internet has expanded for academic use and peace work. It reflects an underlying spirit of the foundation of the United States. The Project covers a wide range of genres including the U.S. Declaration of Independence, the Bible, U.S. Census, the Presidential Inauguration Address, and creative literature, which currently includes over 3,000 documents and works. While the majority of documents in Project Gutenberg is in English, it also includes other languages. In Japanese, the project includes Akutagawa Ryunosuke's *Rashōmon* and a translation of *The Small Catechism of Martin Luther*. The most distinguished feature is its sophisticated cataloging search function that enables one to search six different fields, consisting of author, title, subject, notes, language, and class in the Library of Congress (LC) classification system. For those familiar with the LC classification system, it is easy to use the Project's search function. For instance, by clicking PL in the LC classification, you might find *Rashōmon*, Okakura Kakuzo's *The Book of Tea*, and Lafcadio Hearn's *Kwaidan*. Furthermore, when clicking subject headings for each title, groups of works in each subject are gathered. Downloading, however, is limited to a standard text file format.

The On-Line Books Page (<http://digital.library.upenn.edu/books>)

In addition to the above mentioned Project Gutenberg, a number of other digital development projects are in progress individually across the world, such as The Oxford Text Archive (<http://ota.ahds.ac.uk/index.html>). The On-Line Books Page (OLBP) is another of those independent on-going projects offering a consolidated seamless search for about thirty databases. The OLBP has been edited and maintained by John Mark Ockerbloom. It moved its main server from Carnegie Mellon University to the University of Pennsylvania in 1999. Since its establishment in 1993, the OLBP has digitized and aggregated over 12,000 e-texts. The majority of its collection is from English speaking countries; however, it also includes others. Among Japanese works, for instance, it includes Natsume Sōseki's *Kokoro*. In addition to author and title search,

the OLBP has the richness of the LC classification subject headings. When clicking a subject heading of the classification, titles related to the subject appear. With usage of a high-powered link, it links to related reference materials or to pages with more detailed information. The OLBP also has lists of both banned books and prizewinners. Moreover, under "A Celebration of Women Writers," which lists titles by author and by country, one might find it worthwhile to examine the translation and summary of works by Murasaki Shikibu, Sei Shonagon, Ono-no Komachi, and Abutsuni.

Aozora Bunko (Blue-sky Archive) (<http://www.aozora.gr.jp/>)

Because the publishing industry is dictated by profits, unprofitable books are not given extended opportunities, and remainders are simply cut off. Even out-of-stock items are often not reprinted unless demand assures a sufficient number of copies to warrant reprinting. It is not unusual that out-of-stock books cease publication within a year. Aozora Bunko, or Blue-sky Archive, is a private digital library that responds to a writer's desire for one's publications be read while filling a reader's desire to read items which are no longer available. It was launched by Tomita Michio, a free-lance journalist, in 1997. This Japanese version of Project Gutenberg offers free access to digitized literature in the public domain, not only items which are no longer copyrighted after 50 years from the author's death but also items still under copyright restrictions (with the author's permission). Selection of titles for inclusion, input, proofreading, and making files, all these tasks depend on volunteers. Thus far the project has digitized more than 1,200 titles as of the end of 2000. The list includes not only works by Japanese authors but also translations of foreign writers as well as English translations of Japanese authors. In addition to literary works, it also includes law.

Although the Aozora Bunko deals only with digital books, its approaches are quite humane and user-friendly. By clicking letters by author and title indexes in Japanese fifty syllables, one will get a library card which contains not only bibliographic information but also an introduction to the au-



thor as well as a summary of the work. For downloading documents, users can choose preferred files in three different types: zip text, HTML, and Expanded Book (software for bookmaking invented by Voyager). Text files are of two kinds: one with kana reading along with Chinese characters, and the other without kana. The list has less restrictive “fair use,” which deserves special merit. With the exception of certain copyrighted titles, one can even use most for commercial purposes. Furthermore, one can reproduce or redistribute downloaded files, either free or non-free. In general, those titles with copyright protection are placed in the author’s designated server space and can be linked from a library card at the Aozora Bunko. Corrections and updates to the list are frequently publicized through its bulletin board named *soramoyo* (weather) and *mizutamari* (puddles). Thus Aozora Bunko functions not only as a listing of digitized books but also as communication space on the Internet, and this should not be overlooked.

Japanese E-texts List

(<http://jcmac5.jc.meisei-u.ac.jp/etext-i.htm>)

The Japanese E-texts List has been developed by Shibata Masao at Meisei University for his own personal academic use and has included items both free of charge and commercial, including CD-ROMs and comprehensive listings of works in Japanese literature. Here one may find items not found in the above mentioned Aozora Bunko. Although it lacks a search engine, it has various indexes that follow the periodic division of Japanese history; up to the Edo period, titles are listed in chronological order of their release, while the titles after the Meiji period are listed by the authors’ year of birth. Furthermore, it has an author index, which is arranged by Japanese syllables as well as by the dates of the authors’ birth and death. The ingenuity

of these indexes is helpful for viewing the entire picture of the history of Japanese literature. One can also obtain information about authors by entering the author index directly. As the list is updated frequently, one can obtain the most current information on new listings and corrections. For similar types of listings, one might use the Japanese Literature Text File by Okashima Akihiro at Fukui University (<http://kuzan.f-edu.fukui-u.ac.jp/bungaku.htm>), which has useful web information on Japanese literature with a brief review. In addition, the Japanese Literature Text File (<http://www.konan-wu.ac.jp/~kikuchi/linkd.html>) by Kikuchi Shin’ichi at Konan Women’s University offers URL addresses, which is a useful feature.

Japanese Text Initiative

(<http://etext.lib.virginia.edu/japanese/>)

A joint project developed by the University of Virginia and the University of Pittsburgh, the Japanese Text Initiative is an archive of digitized titles in Japanese literature ranging from *Man’yōshū* (Ten Thousands of Leaves) to Yokomitsu Riichi’s *Kikai* (Machine) with both text and image marked using SGML (Standard Generalized Markup Language). Though limited in the number of inclusions (40 titles as of the end of 2000), the Initiative deserves special merit for its bilingual search function in both Japanese and English, since it aims to include non-native speakers of Japanese. Because the founding project at the University of Virginia Library Electronic Text Center was launched to support academic purposes as a counterpart to commercial electronic texts, the Initiative digitizes only titles to be offered free of charge on the Internet. It is known for its accuracy of bibliographic information and completeness of data input. It also includes many illustrations and other visual references drawn from special collections and manuscripts. Furthermore, the uniqueness of the

Initiative is a parallel text of *Ogura Hyakunin Isshu* (One Hundred Poems by One Hundred Poets), which is worthwhile for teaching. The Initiative recently received the Digital Archives Award by Digital Frontier Kyoto for its contribution in preserving cultural tradition in a digital archive, with cutting-edge technology useful for users both academic and non-academic.

It is a boon to readers to be able to obtain books wherever and whenever one desires as far as one has access to the Internet. However, this must be a threat for those in the publishing industry and for parties dependent on copyright; hence copyright issues exist, made complicated by different copyright laws in a global age. As this deeply affects our lives, it will be challenging to resolve these complex copyright issues in this digital century. The future of online e-texts that emerges from the gap between the principles of democracy—that intellectual heritage is public domain—and the idea of copyright is still uncertain, and yet deserves our special attention.

