Thoughts on the Fusion of Informatics and Literature

Hisashi Yasunaga, Emeritus Professor
National Institute of Japanese Literature (NIJL)

Looking back on my experiences with Grants-in-Aid for Scientific Research (Kakenhi), the first thing that comes to mind is a project funded by the Ministry of Education, Science, Sports and Culture (MESSC) under Grant-in-Aid for Special Project Research that was implemented from 1976 to 1979 on the theme, "The Formation process of information systems and the organization of scientific information." Led by the late Professor Hiroshi Inose, this was a massive project that cost over 1 billion yen altogether and involved the participation of more than 500 researchers. I was placed on the general research team and was involved in the development of inter-university computer networks. Our research took place prior to the advent of the internet and was aimed at implementing the precursor to the Inter-University Computer Network. As a priority initiative of the MESSC, it received tremendous support in the form of grant funding. It sought to build on the research accomplishments of a project under Grant-in-Aid for Special Project Research implemented from 1973 to 1975 on "Advanced information processing of large scale data over a broad area" (Chief Investigator: the late Professor Takehiko Shimanouchi) and made major pioneering contributions to Japanese research in the informatics field.

The first stage of my career as a researcher chiefly comprised information and communications engineering-related research at the university level. In the latter stage, I shifted my research work to the National Institute of Japanese Literature (NIJL), an Inter-University Research Institute Corporation. Research at the university level is primarily conducted by individual researchers on an independent basis in the course of their educational activities. The activities at Inter-University Research Institute Corporations comprise goal-oriented research undertakings that involve no educational duties. Recently, however, Inter-University Research Institute Corporations have become affiliated with the Graduate University for Advanced Studies (Sokendai), playing a role in the arena of higher education. I concur with the observation made by Professor Akihiro Watanabe (Nara National Research Institute for Cultural Properties) on the Japan Society for the Promotion of Science (JSPS) website: namely, that research institutions in the humanities field are engaged entirely in commissioned research, with nothing performed by independent researchers.

Computer utilization was the focus of the research work performed at NIJL. A broad diversity of

research themes derived from that focus, and furthermore, within the context of informatics, applied research in the humanities was still uncharted territory. For example, not even a single, effective method had yet been developed for the entry of Japanese text, including the classics of Japanese literature. (At that point in time, the JIS Japanese character standard had only recently been approved.) Funding for commissioned research is one component of an institution's operating budget. However, that funding cannot be utilized for any other purpose and the allocations are typically in small denominations. Outside funding was essential for research on computer utilization. Our application for Kakenhi constituted a request outside the scope of commissioned research. Kakenhi would be vital to the pursuit of our research and we worked in earnest to acquire it. Fortunately, thanks to the extensive grant assistance we did receive, we succeeded in setting the stage for the utilization of computers in research on Japanese literature. Words alone cannot fully express my gratitude. However, the amount of research grants received stood out within the institute and caused pressure for other grant applicants. The computer field has drawn heavy criticism as a money pit. Although the humanities have long been thought of as fields that do not require heavy spending on research, the task of organizing and digitizing information is something that demands huge and sustained injections of funding.

Database compilation is a prime example. In the early days, Japan had not made much progress in building its own databases for academic use and accordingly coped by adapting databases from overseas sources for use by data centers running large mainframe computer systems.
In 1985, I was appointed to serve as a Senior Scientific Research Specialist, a role that involved me in initiatives for the organization of academic information, the inauguration of the National Center for Science Information Systems (now known as the National Institute of Informatics, or NII), and the expansion of funding for Grant-in-Aid for Publication of Scientific Research Results (Databases). We needed to create databases of the specialized information resources that were independently held by inter-university research institutions. The dissemination of information resources on Japanese literature in particular had become a focus of strong demand.

(1) The Database Organization of Resources on Japanese Literature
Following my appointment to NIJL in 1983, I engaged in a structural analysis of diverse resources and information relating to Japanese literature and began forming basic concepts for their organization from an informatics perspective. To that end, I received Kakenhi for projects in Databases, General Scientific Research, Developmental Scientific Research, and International Scientific Research. These efforts paved the way for the creation of Japanese-language databases (including information on the classics of Japanese literature). A diverse array of databases resulted from this research and development undertaking, including a catalog of early Japanese

books in library collections, a unified catalog, and a catalog of published research, as well as full-text databases (e.g., *The Compendium of Japanese Classical Literature* published by Iwanami Shoten and *Hanashibon taikei* [The Compendium of Japanese humorous stories] published by Tokyodo Shuppan), and images of the originals. A microfiche-based catalog launched in 1987 was the first example of a Japanese-language database on Japanese literary resources. That database was also produced in a CD-ROM version and won strong acclaim from users inside and outside Japan.

A five-year Kakenhi project was launched in 1988 for research on a database that would contain the full text of selected Japanese literary classics. Comprising subprojects in the Databases, Developmental Scientific Research, and Scientific Research on Priority Areas categories, this undertaking yielded a full-text database of the approx. 100 volumes and 600 literary works in *The Compendium of Japanese Classical Literature* published by Iwanami Shoten. Following its trial publication in 1998, this database continued to earn high marks from the international community. The system of data description rules (KOKIN rules) for that database was a product of efforts in applied research at the earliest stage, and has been shown to possess descriptive functions comparable to rules in the Standard Generalized Markup Language (SGML) and Text Encoding Initiative (TEI) that it predated.

The resources and information available for research in the humanities field are typically quite extensive and diverse and incorporate complicated structures for data expression that mix text and numbers with images and audio. The same may be said for the academic resources and information that are generated during research. Undertakings in education and research utilize these resources and information in an integrated way. Through projects under Grant-in-Aid for Scientific Research and Scientific Research for Priority Areas, I studied appropriate formats for these resources and information, identified their data structures, and explored approaches to organization based on my findings. I also pursued research to determine whether databases could be harnessed to obtain new knowledge and insights through research on Japanese literature. The e-book of *Soseki and London* is one example of a model simulation with that purpose in mind.

(2) International Collaboration
In FY 2001 I embarked on a five-year project in Scientific Research (S) on the theme of "International collaboration for Japanese literary studies." The interim and final reviews of this project accorded it high marks (A+). A research colleague, Assistant Professor Shoichiro Hara (currently a professor at Kyoto University), made huge contributions to this project. Working in collaboration with foreign universities, research institutions, and academic societies, we compiled

collections of academic resources and information on Japanese studies and made these databases publicly available over the internet. In particular, the databases organize practically all of the academic resources and information available on Japanese studies in Italy and France and are extensively utilized worldwide. The sharing of such resources is now a must for education and research at Sokendai.

I was also involved in a project under Grant-in-Aid for Special Project Research and joint research on systems for the sharing of cultural information resources. Launched in FY 2004, this project had the cooperation of over 30 researchers from several institutions including Professor Mamoru Shibayama (Kyoto University) and Professor Hara, and was focused on the study and implementation of frameworks for the integration, meta-searching, and utilization of multiple databases by researchers in the humanities. For this undertaking, we actually linked around 30 individual databases run by eight of the humanities research institutions that participated in that project, devised a system for meta-searches, and demonstrated its viability through field tests. Currently, new research is being carried forward on these findings as part of an integrated backbone project aimed at sharing the resources of the National Institutes for the Humanities.

Back in Focus: Professor Inose and Haiku
Upon transferring to NIJL, I immediately faced difficulty and was puzzled and bewildered for several years thereafter by a series of experiences that made me feel out-of-place within this new world I had entered. As a concept, the fusion of literature with informatics sounds great in theory but is not always that easy to achieve in practice. I found myself consulting with Professor Inose frequently. On one occasion, although he said he was just joking, the Professor told me to create a haiku using a computer. Of course, it's becoming commonplace now for computers to beat their human opponents in games of shogi or chess but they are still not able to create a haiku. I see this as a difficult issue and I've given up on the idea of pursuing it.