Invited Paper

CSIS Clearinghouse for Academic Research Communities in Japan

Masatoshi Arikawa and Takeshi Sagara
Center for Spatial Information Science (CSIS)
The University of Tokyo
4-6-1 Komaba, Meguro-Ku, Tokyo 153-8904, Japan
{arikawa,sagara}@csis.u-tokyo.ac.jp
http://www.csis.u-tokyo.ac.jp

Abstract

Center for Spatial Information Science (CSIS) has developed and run our own Clearinghouse[1] for various academic researchers to share spatial data since 1999. This paper presents activities of the academic Clearinghouse and the backgrounds of it.
Keywords: *GIS, Clearinghouse, Metadata.*

1 Backgrounds and motivations

Center for Spatial Information Science (CSIS) was founded as a research institute of the University of Tokyo in April 1998. The center has been run by Ministry of Education, Culture, Sport, Science and Technology of Japan. The center has a mission of leading and supporting spatial information scientists, especially researchers in universities and public research institutes. The academic discipline *spatial information science* may sound new but it can be considered as the same discipline as *geographic information science* or *geographic information system*, both of which can be abbreviated as GIS.

We briefly explain what the spatial information science is. In the real world, we find many kinds of phenomena and social problems that are closely related to spatial factors. When we attempt to analyze these phenomena or solve these problems, it is necessary to establish a common method. Spatial information science is such a method. The goal of spatial information science is to develop a general systematic way to construct, manage, analyze, integrate and communicate spatial data (e.g. natural, social, economic and cultural data associated with location data), and to study its applications to natural, human and social sciences and engineering.

The aims of CSIS are the followings:

1. to develop and promote spatial information science
2. to develop the spatial data infrastructure to support spatial information science
3. to carry out joint studies with universities, governments and private companies

The main theme of this paper is to present a framework of our Clearinghouse for academic researchers. The activities of our Clearinghouse are conducted to achieve the above aim (2). CSIS has tried to construct the spatial data infrastructure which allows academic researchers to use various kinds of spatial data which are generally too expensive or difficult to get. At the beginning of the development of our academic Clearinghouse in 1998, we designed it to find what kinds of spatial data CSIS has collected by retrieving the metadata of the spatial data through the Web. At that time, there was no Clearinghouse in Japan. The FGDC[2] provided free software for starting Clearinghouse, but it did not deal with Japanese characters. Thus we had to develop our own Clearinghouse from scratch. Activities of international standardization of geographic information (ISO/TC211)[3] had been conducted at that time. While giving respects to their standardization activities, we developed our own ISO/TC211 compliant format for metadata of spatial data which were written in Japanese. Also, we adopted XML for describing the meta data and the format was defined by DTD. Our format was mainly based on the level 1 of *ISO/TC211 Geographic information - Metadata*. Small parts of the level 2 of it were also adopted in our format.

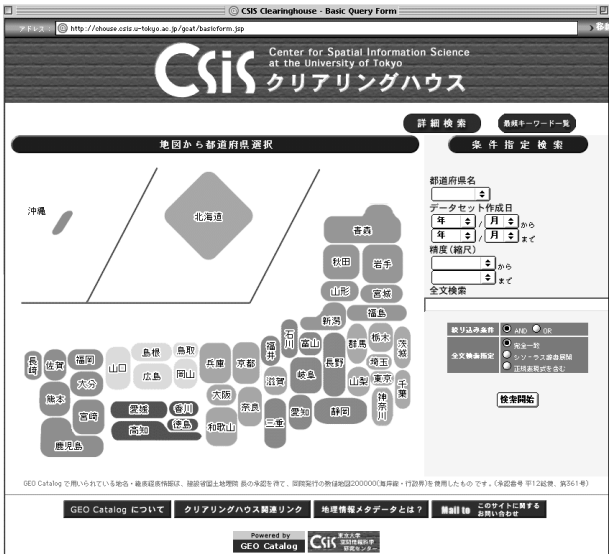

Figure 1. Top page of CSIS Clearing-house.

Figure 3. An example of the content of a metadata in a table style display format.

Figure 2. A result of selecting an area of interest .

Figure 4. Description for items in meta-data.

2 User-friendly interfaces and other characteristics

User-friendly interfaces have been considered very important for researchers in many fields to use CSIS Clearinghouse. Thus we designed the user interfaces (Figure 1) of CSIS Clearinghouse to be different from the one based on the FGDC Clearinghouse. If the researchers are experts of spatial data, it is appropriate for them to specify specific values for items of their interests to retrieve metadata for spatial data. However, most of the researchers who visit our Clearinghouse may not be the experts of spatial data. We designed the user interfaces based on this assumption. There are mainly two types of spatial information science researchers who use the Clearinghouse. The first type of researchers are mainly interested in local areas. It will be convenient for them to easily specify their intended areas on the user interfaces. The top page of CSIS Clearinghouse provides the Japan map for this purpose. The other type of researchers use data covering wide areas such as all parts of Japan. They are interested in themes or layers rather than areas. The FGDC style user interfaces are not appropriate for these researchers because they have to learn the data schema to retrieve spatial data. Our user interface provides a keyword search function which is based on text information retrieval. Researchers only input some keywords of their interests in the field to find metadata of their interests. Sometimes, it is not easy for them to find good keywords, so our Clearinghouse provides a list of frequency of keywords which are inputted by past visitors. Researchers can browse the list of keywords. The keywords can help the researchers inspire what keywords are appropriate for finding their intended spatial data. Also, in order to help researchers make queries to CSIS Clearinghouse, we have been preparing an experimental advanced function of making use of a thesaurus in order to intelligently interrupt requirements from researchers. Because thesauruses we currently used are designed for general uses, not for GIS, the function of using them has not worked well so far. However, if we can obtain or create a better thesaurus of terms concerning GIS or spatial data in the future, it might greatly help researchers. The durations of spatial data are useful items for all researchers to specify values of the durations for selecting their intended metadata from CSIS Clearinghouse. The top page has the fields of the durations to be specified directly by researchers.

Figure 2 shows a result of a query to CSIS Clearinghouse. It shows a part of the list of metadata as the result with a map which presents the extents of currently listed-up metadata. The map with the extents of metadata helps researchers know easily the contents of the metadata. The map was produced by IncrementP Co. and is served through the Internet by Internet GIS Co., an application service provider (ASP). The use of ASP for base maps in Clearinghouse may be the first case in Japan as well as in the world. If a researcher selects one metadata from the list, the researcher can browse the content of the metadata (Figure 3). As the definitions of items in the metadata are generally difficult for general researchers to understand, our user interfaces provide additional pages for explanations about items by clicking an item on the pages showing the content of the metadata (Figure 4). For experts of spatial data, the FGDC style user interfaces are also available in our Clearinghouse. In addition to browsing the metadata in the table form, the researcher can download the metadata themselves as the XML formed files. Then, they can process the metadata as their intended formats only for non-profit uses. Our Clearinghouse also provides researchers with a function of converting CSIS style metadata into JMP (Japan Metadata Profile) style metadata. JMP style is recommended by Geographic Survey Institute of Japan [4]. JMP style is defined by XML and is subject to ISO/TC211 Geographic information - metadata. Geographic Survey Institute also provides Isite-J as free software. Isite-J was modified from Isite which was originally provided by the FGDC in order to deal with Japanese characters. Any organizations can use Isite-J and launch their own Clearinghouse as a node of the FGDC Clearinghouse networks which have gateways for managing pointers to nodes. First, general researchers visit some gateways of the FGDC Clearinghouse, then make queries to find some information about the metadata. Isite uses ISO23950 protocol that is different from HTTP protocol, the most popular protocol for the Web. CSIS Clearinghouse does not use ISO23950, but uses HTTP protocol because it is much easier to maintain and develop real systems of Clearinghouse. CSIS Clearinghouse system is also offered free of charge so that any organization can start their own Clearinghouse using its free software. We have prepared some manuals to help researchers start their own services using CSIS Clearinghouse smoothly. The system of CSIS Clearinghouse is composed of all free software. The system is all programmed in Java and uses a Web Server and a Servlet. We recommend Apache as the Web server and Tomcat as the Servlet for the current system of CSIS Clearinghouse.

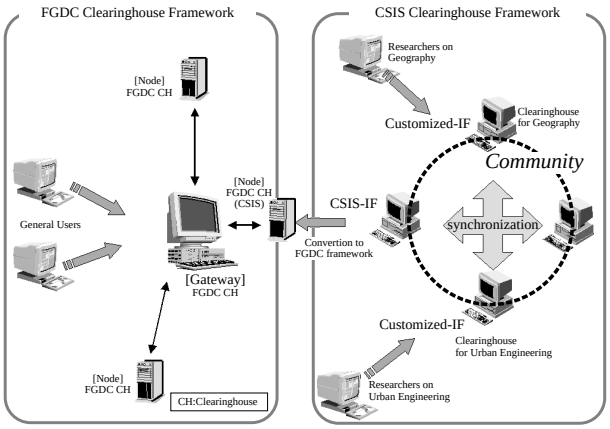

Figure 5. Configuration image of communication between multiple Clearinghouse nodes based on communities.

3 Multiple Clearinghouse nodes based on communities

The current version of CSIS Clearinghouse is a centralized system. We plan to extend it to a distributed system which will be different from the FGDC Clearinghouse framework. We assume that the update of the metadata is generally not frequent and the network speed is fast. Based on this assumption, it is natural to synchronize all sites of CSIS style Clearinghouse at the same time through the Internet. It means that all CSIS style Clearinghouses run by various organizations can have one global view of CSIS style Clearinghouse networks. It also makes the retrieval speed very fast because all data are stored in local disks. As we explained, the variety of the metadata for academic research purposes is wide depending on application domains. It is not natural to have only one metadata format such as JMP as a standard in Japan, particularly for academic uses. For this reason, our framework allows each application domain to have their own metadata formats which are extended from CSIS metadata format. Then, all member sites in the same community can be synchronized globally. Also, there will be gateways between different communities such as the FGDC Clearinghouse (Figure 5). We hope that the gateways make inter-Clearinghouse networks in Japan.

4 Conclusion

This paper has presented some backgrounds and approaches of CSIS Clearinghouse. The node of CSIS Clearinghouse was only one in September 2001, but three nodes are planned to start at Kyushu University, Osaka City University and Hokkaido University this year. We are now developing a distributed multiple Clearinghouse system. Next year, several nodes of CSIS Clearinghouse will be synchronized and shared in multiple communities. CSIS also provides other services: a spatial data sharing service and an academic portal site called *GISSchool*[5] for GIS people. The spatial data sharing service is related to CSIS Clearinghouse. If visitors to the CSIS Clearinghouse have already begun to collaborate with some staff of CSIS, they can download spatial data themselves from the spatial data sharing system in CSIS after retrieving CSIS Clearinghouse. The academic portal site GISSchool gives pieces of URL information of useful Web contents to support GIS people from an academic viewpoint. Examples of the useful contents are educational materials, experiments and so on. One characteristic of GISSchool is that any GIS people can contribute their recommended contents about GIS to GISSchool. GISSchool also has a powerful function of ranking for submitted contents and of automatically structuring of contents on the Web pages.

Acknowledgements

We would like to thank Professor Atsuyuki Okabe, Associate Professor Yukio Sadahiro and Professor Yasushi Asami for their efforts on realizing CSIS spatial data infrastructure for academic researchers of which CSIS Clearinghouse is a part. We appreciate the cooperation of all people related to CSIS Clearinghouse in developing the ideas and making real systems.

References

[1] Clearinghouse of Center for Spatial Information Science, Univ. of Tokyo (In Japanese): <http://chouse.csis.u-tokyo.ac.jp>.

[2] Federal Geographic Data Committee, USA: <http://www.fgdc.gov>.

[3] ISO/TC 211 - Geographic information: <http://www.statkart.no/isotc211>.

[4] Clearinghouse of Geographical Survey Institute, Ministry of Land, Infrastructure and Transport, Japan (In Japanese): <http://zgate.gsi.go.jp>.

[5] *GISSchool*, Center for Spatial Information Science, Univ. of Tokyo (In Japanese): <http://gisschool.csis.u-tokyo.ac.jp>.