Designing metadata catalog system for ecological information

Geunchul Park*, Soonwook Hwang*, Taesang Huh*
*KISTI(Korea Institute of Science and Technology Information), Korea
E-mail : gcpark@kisti.re.kr, hwang@kisti.re.kr, tshuh@kisti.re.kr

1. Introduction

The ecological environment in today's society is changing rapidly as a result of external stimulus including development and pollution caused by human among others. In order to prevent destruction of the environment resulting from development by human and to prepare for possible natural disaster, a system that can collect and share ecological information is in vital need. In particular, ecological data collects various types of measurement data, photos, video footages and others in various forms, and as the process is carried out for an extended period of time within a broad scope, the amount collected is quite significant. In order to effectively manage, utilize, and share these data, a metadata catalog system that can extract the summary information from the collected data and manage the extracted information, is needed. This thesis designed metadata catalog system for the purposes of effective management, utilization and international sharing of ecological information.

2. Ecological information metadata

Metadata refers to data regarding data, in that it is information that follows a specific data in order to analyze, classify, as well as add supplementary information to a certain raw data, namely structured information. Metadata is a data assigned to data in accordance with the fixed rule, mainly to efficiently search for the desired information from mass information or from a pool of information created in forms from which are difficult to search. Currently, metadata is used in an extremely wide scope, ranging from information regarding images taken with a digital camera to result data from particle accelerator experiments.

Metadata is designed with the characteristics and field of utilization of the raw data taken into consideration. The raw data for ecological information metadata is information collected in various formats including different types of measurement data, images and video footages. In order to design ecological information metadata, one must study and analyze ISO19115, UNEP and other remote-sensing related metadata standard which is already defined and being used, as well as EML, DarwinCore and other ecology related metadata standard and incorporate the information from the study and analysis in the designing process. In addition, metadata from UK's ECN and NEON from the United States which are countries that already collect and manage ecological information with the future metadata sharing in mind, must also be studied and analyzed.

Ecological information metadata must be designed based on classifications according to its characteristics including basic data, specific data, and extended data. Basic data is metadata that every data possesses and the information contained includes file name, date created, date updated, size and others. Specific data is a metadata possessed by data in specific format and in the case of picture files, the information includes resolution, date of photograph, photograph equipment and others. Extended data is metadata which the data registrant inputs directly and is information that is used to explain the applicable data, utilization related information and etc.

3. Metadata Server

Metadata server performs the role of storing metadata via direct communication with the database and providing metadata information at the client's request. The functions of the server can largely be divided into basic functions, security management functions, data sharing functions, and API for service. Basic functions refer to create, update, read, delete functions used in managing metadata. As metadata can most effectively be managed using hierarchical method, creating function in order to support this, is needed. Furthermore, a function which allows for read using various conditions, as well as modifying and deleting of metadata is needed. Hierarchical management function must be materialized in a similar form to that of a command language used in managing the directory and file in Linux and for conditional read of the data, it must support SQL Query Statement. In security management function, the necessary functions include certification function which certifies the user access as well as authorization function that controls data access authority. In the case of metadata, per group authority management of data is vital and it must allow for the user to select the access authority per domain of specific data.

For effective sharing of data, grid/cloud base data sharing function should be designed. It is a vital function for the purposes of international data sharing and it must support various grid certification method and standard encryption protocol. For encryption necessary for data communication, it must support standard SSL. Replication and Federation functions for effective sharing of metadata were also designed. Replication is a function that automatically produces copies and manages mass data for the purposes of metadata sharing. Federation is a function that allows for accessing of high volumes of metadata which are stored in dispersed manner in various sites, as though it is all stored in a single site. Metadata must provide service in connection with various external interfaces. Accordingly, design interface such as TCP Streaming, WS-DAIR and etc. for the purposes of communication with the outside. Java, Python and other API library must be provided in order to utilize external application.

4. External access interface

Design two forms of interface in order to access the metadata from outside. One should be an interactive command based client and the other should be a web-application. Interactive command interface will be managed mostly by the administrator and communicate via TCP Streaming with the server. Materialize it in a client program form which can be executed in the Linux system. Web-application should be created to make it possible to access various data read and manage functions suitable for ecological information with a single click. In addition, create Excel File upload and download functions in order to create and read mass data.

5. Metadata extension function for ecological information

Design basic and specific metadata automatic extraction function in order to compose metadata more easily from the collected ecological information or the information being collected. Regarding previously collected data, use batch job process in order to create uniform automatic extraction of metadata and registration function. For newly collected data being registered, created a web-application plug-in that automatically extracts and fills into registration tool, the basic and specific metadata from the registration process. The user who registers ecological information must check to verify the extracted basic and specific metadata. Additionally, the extension metadata must be manually entered in.

For international sharing of ecological information metadata, automatic conversion support function must be created, regarding metadata with different format and storage system. In order to do so, a function that can extract the stored metadata in various format and a function that can automatically convert and store external metadata of various format, must be created.

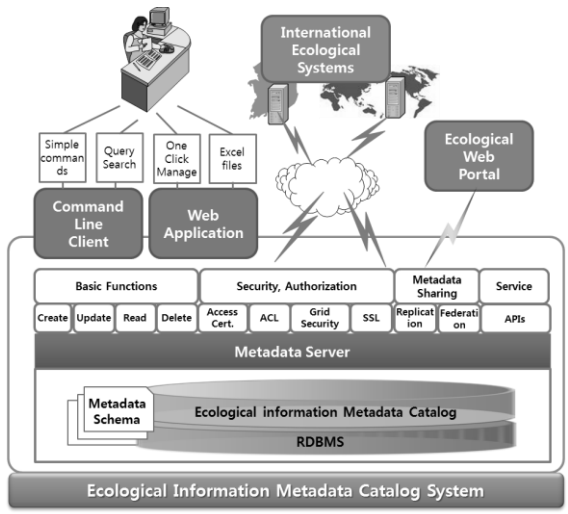

Figure 1. Ecological Information Metadata Catalog System

6. Conclusion

This thesis strived to design metadata catalog system for ecological information. This included designing metadata schema for ecological information and designing a function required in metadata catalog system. Metadata extension function was also designed in order to effectively collect and share ecological information. In the future, when the actual system construction completes, we expect that it would be used effectively in collecting, managing and utilizing of the ecological information.

7. References

[1] Sunil Ahn, Namgyu Kim, “Performance analysis and optimization of AMGA for the large-scale virtual screening”, Software: Practice and Experience, John Wiley & Sons, Inc., vol. 39, no. 12, pp.1055-1072, 2009.
[2] Santos N, Koblitz B, “Metadata services on the grid”, Proceedings of the Advanced Computing and Analysis Techniques, DESY, Zeuthen, Germany, 22–27 May 2005.
[3] NEON official web site, retrieved from http://www.neoninc.org/
[4] ECN official web site, retrieved from http://www.ecn.ac.uk/