INTERDISCIPLINARY UTILIZATION SYSTEM OF EARTH OBSERVATION DATA WITH MULTI DOMAIN GLOSSARY

Masafumi ONO[1], Masahiko NAGAI[2], Ryosuke SHIBASAKI[3]
[1]Earth Observation Data Integration and Fusion Research Initiative, the University of Tokyo Cw503 stitute of Industrial Science 4-6-1 Komaba, Meguro-ku,Tokyo 153-8505, Japan; Tel: +81-3-5452-6417;
Email: maono@iis.u-tokyo.ac.jp

[2]Geoinformatics Center (GIC), Asian Institute of Technology (AIT) , km 42, Paholyothin Highway, P.O. Box 4, Klong Luang, Pathumthani 12120, THAILAND; Tel: + 66-2-524-5599;
E-mail: nagaim@ait.ac.th

[3]Center for Spatial Information Science, the University of Tokyo, Cw503 Institute of Industrial Science 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan; Tel: + 81-3-5452-6417;
E-mail: shiba@csis.u-tokyo.ac.jp

KEY WORDS: Earth Observation Data, Interoperability, Interdisciplinary, Metadata, Glossary

ABSTRACT: Earth observation data is useful for various research domains such as agriculture, hydrology, meteorology, biodiversity, and oceanography. These research contributions lead to the better understanding of earth cycle or climate change.

DIAS (Data Integration and Analysis) is a Japanese project providing a system infrastructure to integrate and analyze earth observation data. Currently, the DIAS core system contains various kinds of data such as satellite image, ground observation sensor data, civil survey data and simulation model output collected in cooperation with many experts in Asia or Africa regions.

However, most of general search functions are not enough to handle various data in many associated domains. Therefore, we propose our developing system to encourage earth observation data utilization. This system is constructed by using multi domain glossaries related to earth observation science. Also, because this system provides a user interface of keyword graph, it can navigate users from their well-knowing keywords to data entry.

1. Introduction

Earth observation data is the indispensable resources for enhancing earth science researches, and also dealing with global issues such as climate change, typhoon, flood, drought, tsunami, poor harvest, pest damage, and ecosystem destruction. From the effort of earth science developments and analysis, we can currently find many observation systems over the world.

GEO (Group on Earth Observations), which was a global voluntary network organized in 2005, leads the coordination with such observation systems. The main mission of GEO is to build Global Earth Observation System of Systems (GEOSS) continuously. On the official Web site, GEO releases the 2009-2011-work plan for accelerating GEOSS (GEO 2010). The work plan contains several tasks and sub-tasks about architecture developments, data managements, capacity buildings and application implementations.

On the other hand, there is a Japanese project named as DIAS (Data Integration and Analysis System) committing with this global movement (EDITORIA 2006). DIAS has the strong motivation for advancing earth environmental science researches not only in Japanese region but also global region. So, the DIAS member consists of many researchers on various fields such as agriculture, hydrography, oceanology, meteorology. Through the collaborations with several work teams including Asian experts, DIAS has been collecting huge volume of data set such as satellite image, climate model simulation data, and so on.

March 11 in this year, a huge disaster hit Japan. That is an unforgettable day for most Japanese people, because the victims are counted over twenty thousands people. Currently, many activities are under processing for the purpose of supporting sufferers and rebuilding many destroyed cities. To support their activities, some work teams use or develop information systems for monitoring, analyzing and predicting environmental parameters, for example climate, wind speed or tsunami. As outputs of each running processes, many original data sets are being generated day by day. However in most cases, such data formats are bespoke and application specific. So, harmonization with different types of data is a very practical and serious issue in Japan.

In DIAS architecture, data harmonization is realized by metadata technology. And the semantic data association and data retrieval are realized by ontology technology including several glossaries in some domains related with earth environmental science. According to both technologies, DIAS manages about several hundreds of earth environmental data set. A part of archived data set is released last October not only to private members but also to public, if the usage is non-commercial. At the same time, our developed systems are available through DIAS portal site. After DIAS ends the activity phase 1 during five years last year, DIAS starts the activity phase 2 this year. So, DIAS goes the next stage. Also certainly, the current provision for DIAS resources is extended.

This paper is organizes as follows. Section 2 describes the DIAS architecture about storage infrastructure, data policy and quality assessment. Section 3 introduces our developing interoperability arrangement system. In the section 4 we concludes the summary.

2. DIAS architecture

2.1 Data management

The DIAS core system is organized as a traditional centralized database that has huge storage ability. The management of the core system, such as administration or performance, is handled by an information engineering team of the University of Tokyo. The storage contains various kinds of and huge volume of data set. These data set are collected as a result of some negotiations and collaborations with data providers through several past projects related to DIAS. Even now, such archived data variety is increasing every period. So, the managers are reinforcing the disk space in accordance with increasing variety.

Data harmonization for such data variety is realized by comprehensive DIAS metadata. In most case, this metadata is created and attached to DIAS dataset by data provider.

Data policy and usage limit for data users are different in every data set. Basically, original data provider decides the data policy on his own data when they archived. Some data principles require a registration or certification, and another principles require just only citation when you use it. You can confirm the details by checking each metadata,

Data quality is checked by developed application tools when registering and uploading data (Tamagawa 2008). However, this tool is basically customized as application and data specific. So the covering format of the tool is limited.

2.2 Metadata

Each DIAS data set is generated through its own lineage from the created time. This leads the fact that each data model, structure, header information, metadata are independent and not totally unified in the initial state. Such heterogeneous data set should be harmonized for data integration and fusion researches. This is the reason that DIAS developed comprehensive metadata.

DIAS metadata is designed as a subset of ISO 19115 published by ISO/TC211 (ISO 2003). ISO 19115 is a specification of geographic information metadata standards. When we checked archived DIAS data, we can understand most earth environmental data has some kinds of spatial and temporal attributes. This is why DIAS adopts ISO metadata.

The elements of ISO standard are totally about 400 items and can be classified into 14 packages such as entity set, data identification, contact, extent, spatial representation and data quality information and so on. More specifically, the spatial representation package in ISO 19115 contains elements for representing grid-based space and vector-based space. The extent information package can have ability of describing spatial extents for GIS based geometry such as bounding box or polygon and temporal extents for time period.

However, there are some problems on naïve usage of ISO standard. In one case, the original equipments for spatial representation are not enough. Basically, original ISO metadata is considered on the assumption of geographic space based usage. However, some climate model data in DIAS has unique grid expression instead of general geographic model. For example, the vertical axis of ISO model grid is considered as metric unit. But the vertical axis of specific climate model grid is defined as hectopascal or non-regular interval level. In another case, the ISO original elements for temporal expression are not enough, because some simulation data has unique expression such as virtual simulation time that cannot be described as Coordinated Universal Time (UTC).

To solve these unfits cases, DIAS metadata have some extra extensions. The extension is mainly adopted for code list set that is selection elements of XSD. By using DIAS code list, unique grid model or vertical temporal expression can be represented.

In addition, some metadata element is defined as string data type that is not any selection-based, formalized or restricted data types. In one aspect, the string data type has some merits and advantage, because it is very flexible and allows any free text descriptions. But too much flexibility leads the difficulties for extended service developments such as data retrieval on metadata, because text-matching quality becomes down.

To solve this problem, we introduce vocabulary management mechanism to write metadata information is necessary for service developments. We will explain this mechanism in next chapter.

2.3 Vocabulary management

Good vocabulary management is very important for improving heterogeneous data management. The current majority of data access services are keyword-based. In the data access service, the process needs the appropriate matching the user requested query with the index generated from metadata values. If the gap between the vocabularies of query set and index set, data access process will fail.

Our approach on vocabulary management is to reuse the existing resources. Firstly, we collect the authoritative vocabulary resources such as ontology, terminology, thesaurus and glossary. Each resource is created in different domain, discipline and format. So secondly, we pick up the set of title, description and association from all resources. And then we rearrange them totally as something like a multi domain glossary set. Index set for data access service is created from this. In this process of picking up, a part of original strict properties and complex structures (for example, the original ontology has) are lost, because of light-weighting. But, too complex structure does not match with comprehensive management. So, this approach does fit the current status of DIAS rather than depending on the strict rules of conceptual ontology.

By using two technologies, metadata and vocabulary management, we developed a utilization system including a data access service. We will explain it in the next section.

3. Interoperability portal system

We have an interoperability utilization system by using vocabulary management system. This system is available through an interface on Web portal.

DIAS interoperability portal provides data/matadata search, technical term search and visualization of relations among dataset to very large-scale and wide variety of earth observation data registered in the DIAS core system.

This portal system is released not only to private members but also to public last October. So, you can access this site at the following URL address < http://dias.csis.u-tokyo.ac.jp/op/en/>.

In this section, we explain the detail of the system with the captured print screen image.

3.1 Data associated information search

DIAS interoperability portal system equips with data associated information search service. This service provides the basic keyword search function to find archived and released DIAS data set.

Figure 1 shows this portal image. There are several tabs on the top of the figure. You can switch another function to select this tab. The “Data Index/Search” tab means data associated search service.

The visualization is implemented by Flash Action Script. The algorism of visualization is based on force directed graph. More specifically, there are three types of force directed model implemented, spring, boid and complex model. Spring model is a model regarding the edge between nodes as spring. In this model, gravitation and repulsive force by the spring decides the balance of nodes positions. Boid model is a computer model to implement coordinated animal motion, for example bird flocks. The rule focuses on individual unit rather than the whole flock, Complex model calculate with both.

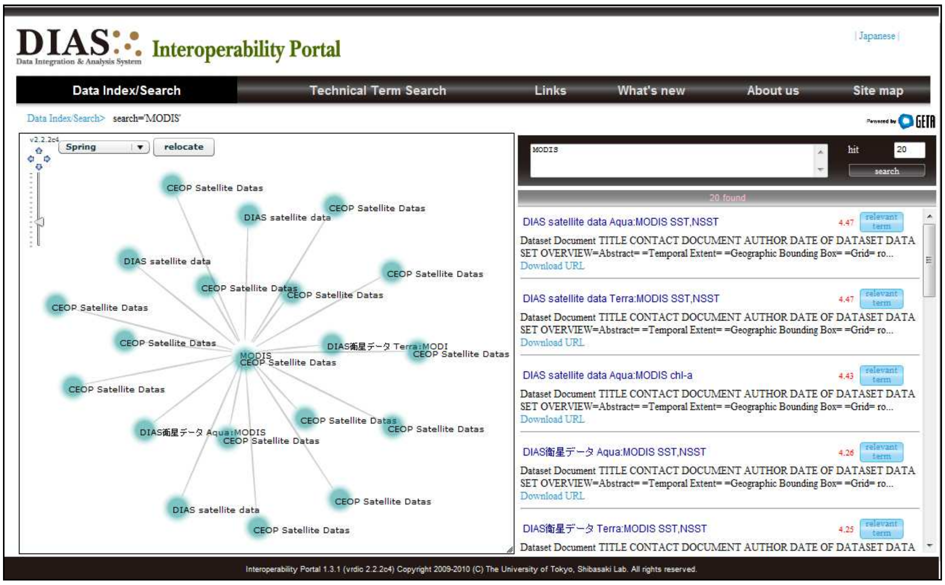

Figure 1: Data association information search on DIAS interoperability portal

3.2 Linked to the metadata management and download server

This service provides metadata information and data download link because this system is linked to metadata management server and data download server through the network. If you click each node in Figure 1, context menu is open. From the context menu or search results, you can connect to metadata information and data download link.

The left side of Figure 2 shows the metadata description, and right side shows data download pages.

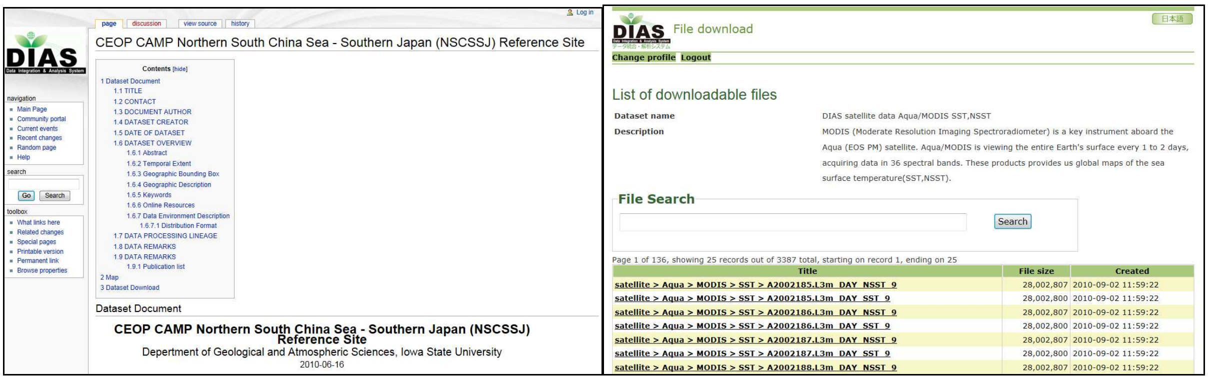

Figure 2: Metadata and data download page

3.3 Term and its definition search

Sharing knowledge among different interdisciplinary academic communities requires the appropriate vocabulary management (Ono 2010). The portal system provides not only data association information search function but also term and its definition search function from the multi domain glossary.

Figure 3 shows one case of searching with the keyword of “temprature”. The graph of left side visualizes the vocabulary-associated information about the keyword. The right side is the result list of the keyword-based query. If you mouse over on one term index, semantic relation of the term (such as “Hypernym of”) will be presented.

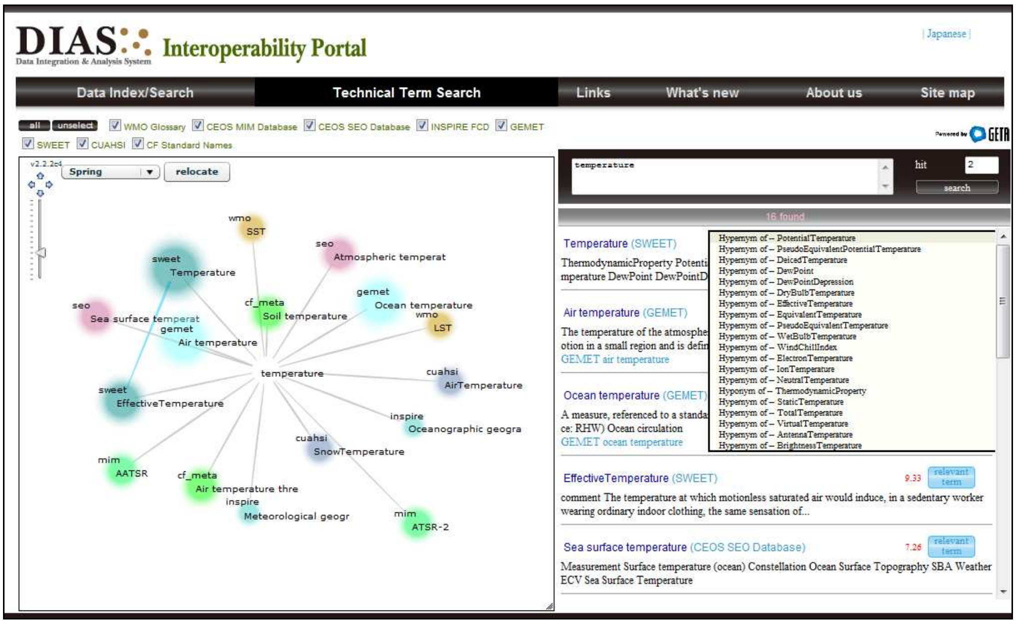

Figure 3: Term and its definition search on DIAS interoperability portal

4. Conclusion

This paper introduces a Japanese project for earth observation data integration and our developing interoperability system. The interoperability system can navigate data associated information and term associated information in the wide variety of data and disciplines about earth environmental data. We believe these approaches lead innovative interdisciplinary researches with data integration, fusion and analysis.

ACKNOWLEDGMENT

This study is supported by DIAS (Data Integration and Analysis System). DIAS is part of the Earth Observation and Ocean Exploration System, which is one of National Key Technologies in Japan.

REFERENCE

EDITORIA, 2006, creating global knowledge and awareness to support better management of one precious Earth, Available online at: http://www.editoria.u-tokyo.ac.jp/dias/english/index.html (accessed 25 August 2011)

GEO, 2010. GEO 2009-2011 WORK PLAN, Available online at: http://www.earthobservations.org/documents/work%20plan/geo_wp0911_rev3_101208.pdf (accessed 25 August 2011)

ISO, 2003, ISO19115:2003 Geographic information – Metadata, ISO/TC211,

Tamagawa K., Ikoma E., Ohta T., Williams S., Kitsuregawa M., Koike T. Ikoma E., Ohta T., Williams S., Kitsuregawa M., Koike T., 2008, Advanced Quality Control System for the CEOP/CAMP In-Situ Data Management, IEEE Systems Journal - GEOSS-Special Issue, 2, 3, 406-413

Ono M, Nagai M, Shibasaki R(2010), Development of the multi-referential reverse dictionary for interoperating earth observation data, 31nd Asian Conference on Remote Sensing (ACRS), Hanoi, Vietnam,1-5 November