

# EARTH OBSERVATION DATA INTEROPERABILITY ARRANGEMENT WITH ONTOLOGY REGISTRY

Masahiko Nagai<sup>1</sup>, Masafumi Ono<sup>2</sup>, and Ryosuke Shibasaki<sup>2</sup>

<sup>1</sup>Geoinformatics Center, Asian Institute of Technology,  
km 42, Paholyothin Highway, P.O. Box 4, Klong Luang  
Pathumthani 12120, Thailand; Tel: + 66-2-524-5599;  
E-mail: nagaim@ait.ac.th

<sup>2</sup>Earth Observation Data Integration and Fusion Research Initiative, The University of Tokyo,  
Cw-503, IIS, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan; Tel: + 81- 3-5452-6417;  
E-mail: maono@iis.u-tokyo.ac.jp

**KEY WORDS:** Earth Observation Data, Data Interoperability, Ontology, Registry

**ABSTRACT:** Standardization organizations are working for syntactic-level interoperability. At the same time, semantic interoperability must be considered as a heterogeneous condition and also very diversified with a large-volume data. The ontology registry was developed and ontological information has been collected for data interoperability. This is a very challenging method for earth observation data interoperability because collaboration or cooperation with scientists of different disciplines is essential. Multiple semantic MediaWikis are applied to register and update ontological information as a part of the ontology registry, which promises to be a useful tool for users. In order to invite contributions from the user community, it is necessary to provide sophisticated and easy-to-use tools and systems, such as a table-like editor, reverse dictionary, and graph representation for sustainable development and usage of ontological information. Registered ontologies supply the reference information required for earth observation data retrieval. We proposed data/metadata search with ontology such as technical term and visualization of relations among dataset to very large scale various earth observation data.

## 1. INTRODUCTION

The global environment is lying on trans-disciplinary fields, such as meteorology, hydrology, geology, geography, agriculture, biology, and so on. It is essential to cross these trans-disciplinary fields for measures of the global environmental problems, such as climate change, global warming, various disasters, and so on. One of the key issues is data interoperability arrangement under the trans-disciplinary condition. There are two aspects of the data interoperability: syntactic interoperability and semantic interoperability. Improvement of both aspects of interoperability is needed for integrated use of heterogeneous data. To improve the syntax interoperability, many efforts have already been made such as standardization of data formats and development of XML-based data encoding rules, for example, ISO (International Organization for Standardization) standard and OGC (Open Geospatial Consortium) standard. Improvement of semantic interoperability requires common understanding among different ontologies, terminologies, taxonomies, including definitions and associations of various concepts/terms, name spaces, classification schemes and so on, which is collectively called an “ontology”. The word “ontology” was originally used in philosophy, to refer to the branch of metaphysics that deals with the nature of being. Currently, in context of knowledge sharing, the term means a specification of a conceptualization (Smith, 2003). In recent years, several institutions have initiated efforts to propose a standard ontology and/or terminology/taxonomy related with Earth Observation. SWEET (Semantic Web for Environment and Technology) by NASA (National Aeronautics and Space Administration) is one such ontology (Jet Propulsion Laboratory, 2010). FAO (Food and Agriculture Organization of the United Nations) is making similar kinds of efforts based on AGROVOC, that is a multilingual, structured and controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (FAO, 2010). Many other ontologies and terminologies/taxonomies are expected to be proposed by other expert/professional communities and institutions. For data interoperability, ontological information including terminology, taxonomy, glossary, etc., must be collected, managed, referred and compared; for example, data dictionaries, classification schemata, terminologies, thesauruses, and their relations are handled. Common understanding of heterogeneous semantic information is used for data sharing and data services such as supporting data retrieval, metadata design, information mining, and so on. In this study, the ontology registry is constructed for information sharing by using a Semantic MediaWiki, which helps to gather ontological information and associations for data interoperability among diversified and distributed data sources. Generally, ontology is applied to a strict and well-defined purpose, classes and instances such as a task ontology (Kitamura, et. al., 2001), but in this study, the scope of ontologies is not restricted and comprises any reference information based on terminology of technical terms for data interoperability. The ontology registry

creates a “knowledge writing tool” for experts, by extracting semantic relations from authoritative documents using natural language processing techniques, such as morphological analysis and semantic analysis for earth observation data interoperability of DIAS (Data Integration and Analysis System).

DIAS is a Japanese national key project having missions to archive such earth environmental data and then to analyze global phenomenon through combining and processing these data such as observation data, numerical model outputs, and socio-economic data provided from the fields of climate, water cycle, ecosystem, ocean, biodiversity and agriculture. The aim of DIAS is to share earth observation data and knowledge among different disciplines. Currently, many researchers in the science and engineering fields are participating in DIAS. DIAS is one of GEOSS (Global Earth Observation System of Systems) activities in Japan.

## 2. ONTOLOGY REGISTRY

In order to collect ontological information with above requirements, a registration system is developed based on Semantic MediaWiki (version SMW1.2). Semantic MediaWiki is a feature-rich wiki implementation. Semantic MediaWiki handles hyperlinks and has simple text syntax for creating new pages and cross-links between terms (Semantic-mediawiki.org, 2010). Entry words, definitions, sources, and authors are handled as nodes with tags, and relations to other terms are handled as links. Those terms are surrounded by other relational terms. Here, each ontology or terminology is managed by separate Wiki, for example, SWEET Wiki is created for SWEET and AGROVOC Wiki is created for AGROVOC.

### 2.1 Registration

At first, ontological information is added to Semantic MediaWiki by the developed tool automatically converting form text, RDF, and AWL to XML and importing to Wiki. Sometimes, ontological information is manually registered from book and Web pages. These existing dictionaries or glossary are already considered as ontological information. OCR (Optical character reader) is sometimes used to digitize the sources. Secondly, symbols and abbreviations, such as related words and synonyms are extracted from the dictionary and converted from semantic structure to syntactic structure by natural language processing. Finally, imported ontological information is modified by authorized users with editing function of the Wiki as shown in Figure 1.

In Semantic MediaWiki, a visual depiction of content is expressed by tags. It is not easy to add or select appropriate relations by tags without knowledge of computer science and miss spelling, so in this study, we developed a table like editor as a wiki plug-in. The table editor links to editing page of the Wiki by pop-up window and suggest appropriate tags to control community authoring. The table editor is implemented by dhtmlGrid, v1.2 standard. XML is prepared for Web server. Figure 2 shows Semantic MediaWiki table editor, by which the user can browse and edit explanations of a term without ambiguity of tags. Semantic Media Wiki displays not only definitions, but also relations of terms among multiple Wikis. The table editor is applied in order to modify relations of terms by using a table without tags.

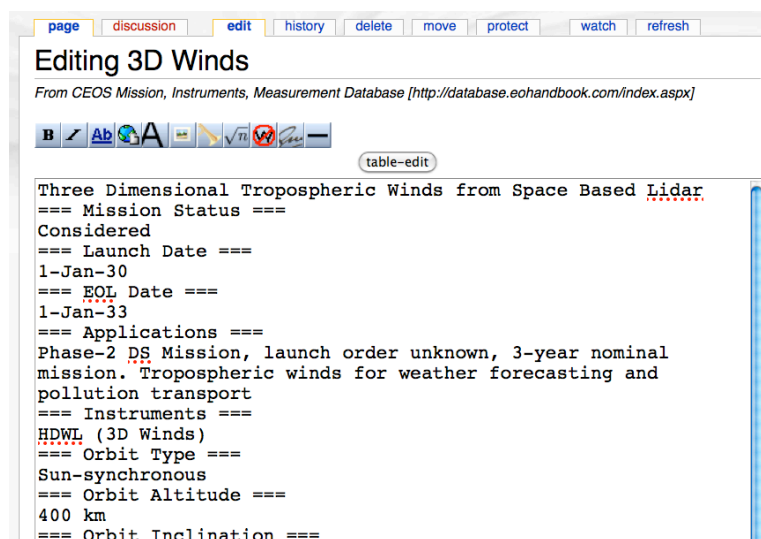


Figure 1 Semantic MediaWiki editing interface

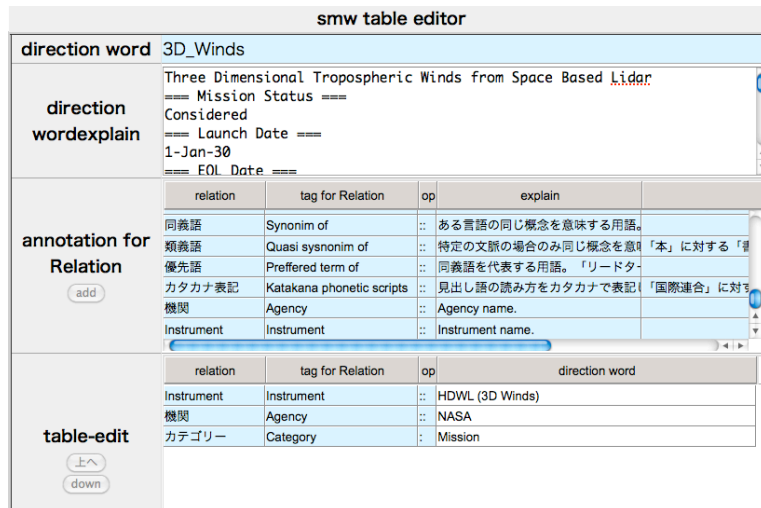


Figure 2 Table editor which is Semantic MediaWiki plug-in

## 2.2 Information Retrieval

Registered ontological information with Semantic MediaWikis is retrieved by a reverse dictionary. A reverse dictionary describes a concept of a term from definitions and associations of terms. The reverse dictionary is developed based on GETA (Generic Engine for Transposable Association), which was developed by the National Institute of Informatics, Japan (Takano et. al., 2000). It comprises tools for manipulating large-dimensional sparse matrices for text retrieval through more than one Wiki in all together. GETA is an engine for the calculation of associations such as similarity measurement of multiple Wikis. In order to create matrices to find similarity, morphological analysis is conducted for word segmentation and listing of ignored words for calculation. The query is “earth environment observation by satellite or air-craft”. The result is “remote sensing”. As an example of information retrieval, suppose a user wants to know about a “satellite for sea surface temperature”. The reverse dictionary returns the answer as a list of terms with similarity scores as shown in Figure 3., such as “sea surface temperature” and “MSMR” in CEOS terminology, “Thermal sea power” in GEMET terminology, and so on. The reverse dictionary relates data by calculation of similarity by using a definition. The user without basic knowledge can discover that a “MSMR” instrument is good for monitoring sea surface temperature and that sea surface temperature is related to “Thermal sea power”.

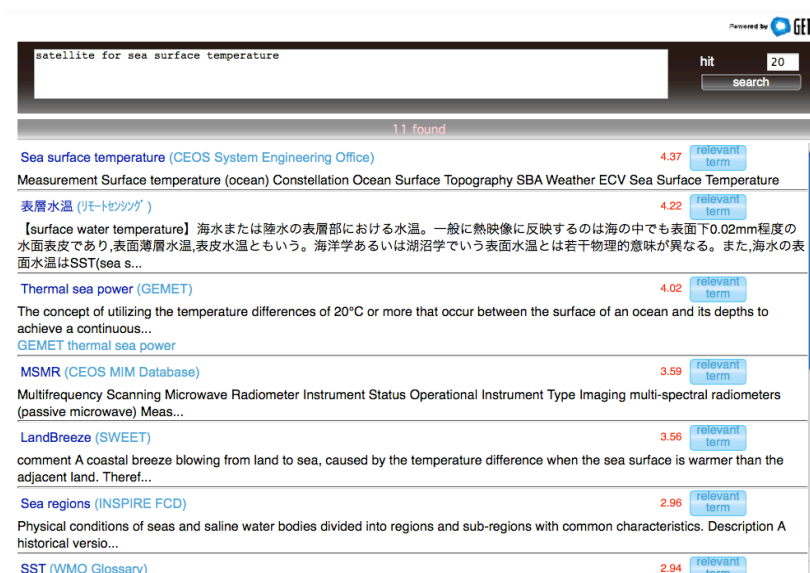


Figure 3 Reverse dictionary

## 2.3 Graphical Representation

In order to compare associations among the different key words from various ontology and terminology which is managed by each Wiki, graph representation as shown in Figure 4. is useful. The graph representation is developed

by KeyGraph that is open source of Java library. XML data that is constructed in the Wiki is visualized with the result of information retrieval by the reverse dictionary. All the related terms from various ontologies and terminologies are represented at once.

One of the examples is graph representation is a term from landuse classification schema in Thailand and Indonesia. The term “water body” landuse class can be found in both countries. Apparently, both landuse classes are the same, but the level of hierarchy is a bit different in each classification schema. In the case of Indonesian landuse, “water body” does not include watercourses, but “water body” in Thailand includes all water-related geographical features. Consequently, graph representation proves a clear distinction between the two terms. Then, the new information such as the relations of “water body” in both countries can be created that “water body” class in Thailand is the same as “water” class in Indonesia. This kind of information is treated as newly-created ontological information, and is added through the Semantic MediaWiki. The ontological information can grow autonomously by adding relations, becoming more and more useful.

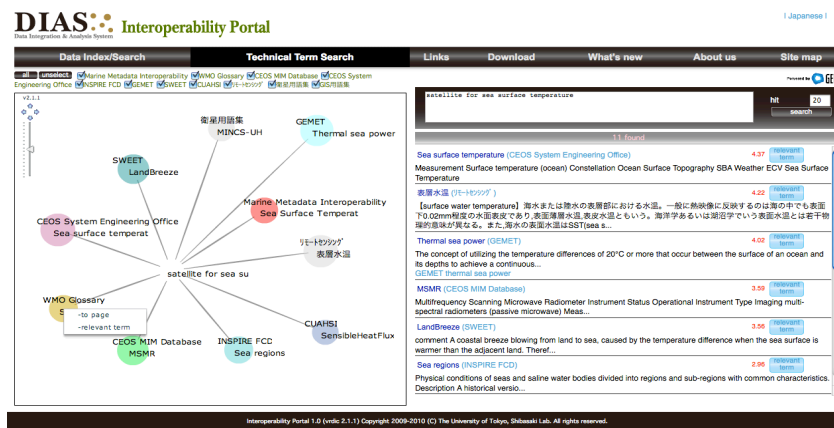


Figure 4 Graph representation for reverse dictionary with Wiki

#### 4. DATA RETRIEVAL AND IMPLEMENTATION

DIAS is tackling a large increase in volume of the earth observation data. DIAS has been developing a core system for data integration and analysis that includes the supporting functions of life cycle data management, data search, information exploration, scientific analysis, and partial data downloading. DIAS is also tackling a large increase in diversity of the earth observation data. For improving data interoperability, DIAS is developing a system for identifying the relationship between data by using ontology on technical terms and ideas, and geography. DIAS also is acquiring data base information from various sources by developing a cross sectoral search engine for various databases. Interoperability portal for DIAS has been developed. This portal provides data and metadata search, technical search and visualization of relations among dataset to very large scale and various earth observation data registered in the DIAS core system.

Most of earth observation data commonly in DIAS have spatial and temporal attributes such as the geographic coverage and the time stamp of data creation with scientific keywords. The metadata standard is published by the geographic information technical committee (TC211) in ISO 19115 and 19139 series. Accordingly, DIAS metadata is developed based on ISO/TC211 metadata standards. From the viewpoint of data users, metadata is useful not only for data retrieval and analysis but also for interoperability and information sharing among experts. In DIAS, document centric metadata registration tool has been developed for reducing time for creating metadata. Since various kinds of datasets stored in DIAS core system increase, it is necessary to support searching datasets based on keywords, spatial conditions, and temporal conditions with created metadata. Datasets are classified into some categories based on such criteria as GCMD Science Keywords or GEOSS social benefit areas.

Registered metadata and ontological information is managed in the interoperability portal. Figure 5 shows the index of data search. Dataset is accessed from four categories, persons, places, keywords, and organization. Persons are the responsible person name for dataset. Places are the location of dataset such as country and city name. Keywords are the related keywords that are controlled by ontology registry. And Organization is the information of data provider. The interoperability portal helps keyword retrieval. The result shows not only related dataset but also related researchers name, organization, and available location of dataset.

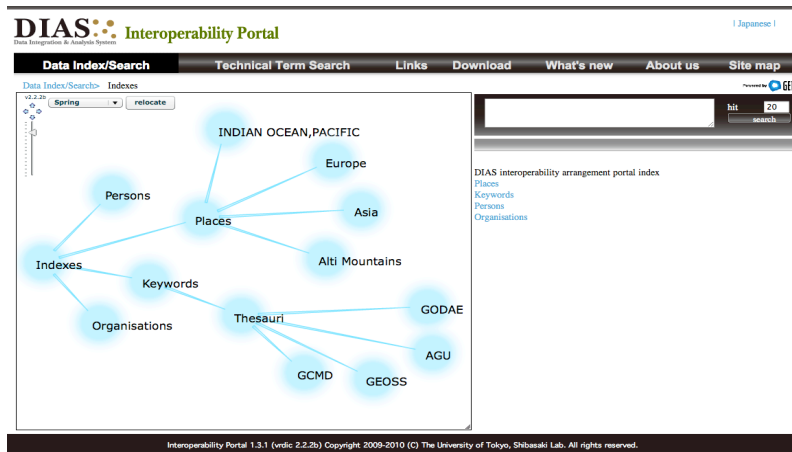


Figure 5 Index of data search with ontology

## 5. CONCLUSION

According to the improvement of observation technologies and earth science studies, a large amount and various kinds of earth observation data including remote sensing data, satellite images and model simulation data are globally being produced by many experts and researchers. At the same time, many kinds of ontology, taxonomies, thesauruses, and gazetteers are being produced in various fields. The ontology registry needs to be developed as a showcase and as a basis for the comparative analysis for better semantic interoperability among diversified earth observation data. The ontology registry is carrying out a component of DIAS and GEOSS interoperability infrastructure.

We have developed the ontology registry and collected the authoritative glossaries, dictionaries, terminologies and ontologies about the earth observation domain and also developed the multi-referential reverse dictionary including them. The reverse dictionary supports getting unknown meanings from our own vocabularies. Our proposed approach is beneficial for earth observation data interoperability management with ontological information.

## REFERENCES

B. Smith, Preprint version of chapter “Ontology”, in L. Floridi (ed.), Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, pp.155–166, 2003.

Jet Propulsion Laboratory, California Insutitute of Technology, Semantic Web fir Earth and Environmental Technology. Available: <http://sweet.jpl.nasa.gov/index.html>

FAO, AGROMOC Thesaurus.  
Available: <http://aims.fao.org/website/AGROVOC-Thesaurus/>

Y. Kitamura, M. Kashiwase, M. Fuse and R. Mizoguchi, “Deployment of an ontological framework of functional design knowledge,” Advanced Engineering Informatics, Volume 18, Issue 2, pp. 115-127, April 2004.

Semantic-mediawiki.org, Semantic Mediawiki, Available:  
[http://semantic-mediawiki.org/wiki/Semantic\\_MediaWiki](http://semantic-mediawiki.org/wiki/Semantic_MediaWiki)

A. Takano, Y. Niwa, S. Nishioka, M. Iwayama, Toru Hisamitsu, O. Imaichi, H. Sakurai, Information, “Access based on Associative Calculation”, In Lecture Notes in Computer Science LNCS:1963, Springer, 2000.