

# Metadata Management for Integration and Analysis of Earth Observation Data

Akira Takahashi

Department of Social Informatics Graduate  
School of Informatics, Kyoto University  
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501,  
Japan

atakahashi@db.soc.i.kyoto-u.ac.jp

Hiroko Kinutani

Earth Observation Data Integration and Fusion  
Research Initiative (EDITORIA), The University  
of Tokyo  
Komaba 4-6-1, Meguro-ku, Tokyo 153-8505,  
Japan

kinutani@tkl.iis.u-tokyo.ac.jp

Masashi Tatedoko

Department of Social Informatics Graduate  
School of Informatics, Kyoto University  
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501,  
Japan

m.tatedoko@db.soc.i.kyoto-u.ac.jp

Masatoshi Yoshikawa

Department of Social Informatics Graduate  
School of Informatics, Kyoto University  
Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501,  
Japan

yoshikawa@i.kyoto-u.ac.jp

## ABSTRACT

Earth observation technologies have developed rapidly during the last decades. Substantial amounts of earth observation data have been acquired and stored among the literature and databases of various research fields such as climatology, oceanography, agriculture, and ecology. Analysis and integration of such data might produce valuable data products to promote understanding of the global environment, and to solve global environmental issues. However, most institutions store and manage their Earth observation data in their own manner, with little metadata. Scientists have to struggle hard to search for valuable data from data out of their research domain and seek their usage. In this paper we introduce a conceptual model for earth observation data. Utilizing a model to express earth observation item associated with ontologies, The model is an simple quintuple with information extracted from conventional data models, and is used to uniquely determine portions of earth observation data, which enables flexible annotation of earth observation data. We also introduce our systems for metadata management and user interfaces for encouraging user annotations of earth observation data that can help scientists discover and understand useful data that can support their research.

Keywords: Earth observation data, data annotation, data lineage, DIAS

## 1. INTRODUCTION

Earth observation data have increased both in volume and diversity in recent decades; and integrated use of earth ob-

ervation data has attracted much interest. Today, earth observation data are collected by many organizations and institutions from various fields of studies using methods such as in-situ observations, oceanographic observations, remote sensing, weather and climate models, and participatory citizen observation. Through integration of such data among different disciplines, we can achieve further understanding and provide comprehensive solutions to global environmental issues.

### 1.1 Data Integration and Analysis System

The Data Integration and Analysis System (DIAS) project is intended to facilitate multi-disciplined management of earth observation data. It is part of the Global Earth Observation System of System (GEOSS), a multinational project for management of earth observation data. Launched in 2006, DIAS is a part of the Earth Observation and Ocean Exploration System, which is one of five National Key technologies defined by the Third Basic Program for Science and Technology of Japan. The project is designed to coordinate cutting-edge information science and technology and various research fields examining the earth environment, to create knowledge enabling us to solve earth environment problems and to generate socioeconomic benefits. Several projects within the framework of DIAS have certain achievements in applications, such as integrated water resources management, agricultural production management, ocean circulation and fishery resources management, ecosystem conservation, and participatory monitoring programs.

### 1.2 Emerging problems

Within the DIAS framework, more than 100 terabytes of earth observation data were collected from organizations and stored in the core system of DIAS during 2007. Several hundreds of terabytes are to be stored in the next three years. The collection phase of valuable data has been successful. Nevertheless, scientists must confront several problems that hinder the use of the collected data.

First, most earth observation data have been acquired by organizations and institutions in their own manner; subse-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ICUIMC-09 January 15-16, 2009, Suwon, S. Korea

Copyright 2009 ACM 978-1-60558-405-8109101 ...\$5.00.

quently, the data have been managed in a domain-specific format, intended to be used by special application software, and requiring certain efforts on the part of scientists to use the data.

Next, few metadata are provided; scientists must struggle with great difficulty to discover and understand the data which meet their demands. However, managing metadata and providing high-quality data products might burden data providers with additional work.

To address the stated problems, we must establish a data model for earth observation data to support interoperability of the data products and to enable better management of metadata. We apply the model to metadata provided by the creator of the data products. Some geospatial metadata standards [1, 2] already exist to cope with interoperability of spatial data, but founding a data model for the practical use of earth observation data remains as an open problem. However, as described earlier, earth observation data are rapidly increasing in volume and diversity; unification of metadata is insufficient for discovery and understanding of valuable data. Using data annotation and lineage is necessary to support better methods of data discovery and encouraging research activities.

### 1.3 Data annotation

Management of data annotation will bring further benefits to scientists through discovery of needful data and deeper understanding of the data. Today, data annotation is ubiquitous on the Internet. So-called Web 2.0 applications, such as Youtube [3], Wikipedia [4], Social network sites [5], and Social bookmark services [6] use annotation by users to enhance the value of their contents. We believe that this Web2.0 idea of user interaction is also applicable to the region of e-science. User annotation is expected to provide better understanding of data products; it might introduce a new scheme to evaluate earth observation data products.

Earth observation data come in various contents and formats. Remote sensing provide data in images covering a wide geographical area, whereas data from meteorological observations provide temporal sequential data at a certain geographical site. The actual data files might be provided by text formats or binary formats intended to be read by specific application such as NetCDF[7] or GrADS[8] To discuss how to manage and annotate such data, we must produce a conceptual data model that relies neither on data formats nor objects, and which can determine uniquely which data are annotated. It must be able to seek a URI for earth observation data management.

### 1.4 Related works

#### 1.4.1 Metadata modeling

Many works for geospatial metadata modeling exist. Some of the metadata are incorporated into the actual data format, hence are called self-describing formats. Regarding data modeling for software for grid data analysis, for example, NetCDF[7], GrADS[8], are some major works. Several standards are used for geospatial metadata, such as the Content Standard for Digital Geospatial Metadata (CSDGM)[2], which is used by the Federal Geographic Data Committee. Another standard is that of the International Organization for Standards[1].

#### 1.4.2 Metadata Management Systems

There are numbers of researches managing data annotations. Some works for managing annotations on HTML documents on the Internet includes Annotea[9] and [10, 11]. Social bookmark services such as delicious<sup>1</sup> and Hatena<sup>2</sup> and other services with user created contents, such as Youtube<sup>3</sup>, Flickr<sup>4</sup> shares user annotations and comments to evaluate and classify various contents. Recently, annotation management systems for genomic sequences have also been built [12, 13], as well as in the domain of data warehouses and scientific datasets [14, 15, 16, 17].

We designed and implemented an annotation management system for earth observation data. Conventional systems for managing annotation on earth observation data often attaches annotation on each data files, or a bigger granularity, such as the whole dataset. However, users may want to annotate data among several portions of each data files. For example, users may want to annotate data derived from a specific instrument throughout the dataset. We can achieve this by iteratively annotating each data. But the user semantic can easily be lost by adding another data file on the dataset. Thus we are concerned with preserving the user-semantic of annotations. To the best of our knowledge, this is the first implementation of an annotation management system for earth observation data that allows user-semantic preservation.

The rest of the paper is organized as follows. We introduce our conceptual model for managing annotations of earth observation data, in Section 2. Then we briefly introduce our system we have implemented, and present application interfaces to enhance collaboration between scientists in Section 3. Finally, Section 4 concludes the paper with some future works.

## 2. ANNOTATION MODEL OF EARTH OBSERVATION DATA

To maintain user semantic of annotations, the ability to manage annotation on various granularity is necessary. Our approach illustrated in Fig. 5 is to assume a virtual dataset with fine-grained data, and metadata are annotated against sets of small granules of data. We describe how to model each granules of earth observation data in the following subsection.

### 2.1 Conceptual model of earth observation data

To refer to a certain portion of earth observation data, we need a conceptual model of data that does not rely on data formats or objects. Consider an example of earth observation data shown in Table 1. The data are part of the WMO Resolution 40 dataset provided by NOAA [18]. Two tables exist within the data: a station list for denoting the geographical site of an observatory, and a data file, which shows the actual value of the data. To refer to the value in the second row and the fourth column (39.8), We must specify values of the STATION NAME, YEARMODA, the column name TEMP, as well as the dataset name WMO Resolution 40. These four pieces of information—the spatial attribute, the temporal attribute, the observational at-

<sup>1</sup><http://delicious.com/>

<sup>2</sup><http://b.hatena.ne.jp/>

<sup>3</sup><http://www.youtube.com/>

<sup>4</sup><http://www.flickr.com/>

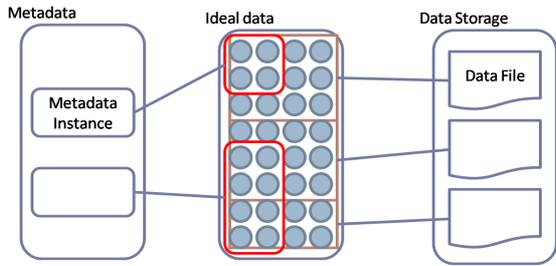


Figure 1: Annotating Data in flexible granularity

tribute and the dataset attribute—are the general information used to specify earth observation data of any kind.

Using these attributes, we model an earth observation datum, expressed as  $d$ , as the following quintuple.

$$d = (ds, s, t, i, v)$$

Each attribute of  $d$  describes an aspect of earth observation data. Actually,  $ds$  is a dataset identifier;  $s$  and  $t$  respectively specify spatial and temporal attributes. In addition,  $i$  is the observation item attribute;  $v$  denotes the actual value observed (or simulated). Using this quintuple, we can uniquely determine  $d$ , to which the earth observation datum is referring. In our model, attributes  $ds$  and  $i$  play key roles in specifying the instance of earth observation data. We present a further explanation of each attribute in the following.

### 2.1.1 Dataset attribute

This attribute denotes the dataset to which the data belong. The value of this attribute is an identifier of a dataset; it might indicate the source-satellite of the data, what climate model was used, which buoy was used, etc. Furthermore, a certain prefix might describe the data processing used.

### 2.1.2 Spatial and Temporal attributes

The spatial and temporal attributes denotes the spatial and temporal extent where the data is valid. The spatial attribute value  $s$  is a representation of a geospatial point or a region. The temporal attribute value  $t$  is a representation of a duration or a time.

### 2.1.3 Observational item attribute

Observation item might describe several characteristics of the data. For example, “max\_air\_temp” might denote that the observed value is temperature, and is the highest value in a certain period. We define characteristics separately; we define an observation item as a combination of such characteristics. The three characteristics used to define the observation item are listed below.

Target:

The target substance or phenomena of observation: air, rainfall, wind.

Property:

The observed property: e.g., temperature, mass, speed.

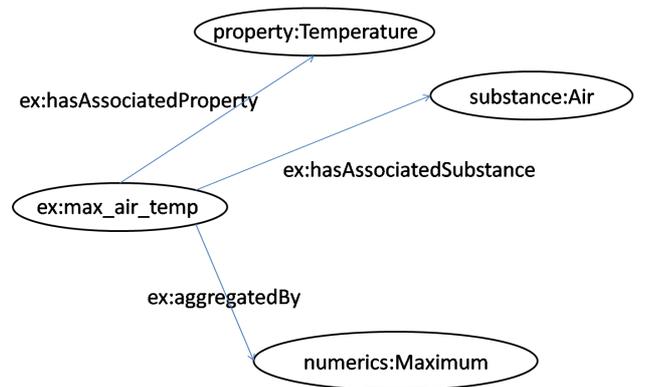


Figure 3: An Observation Item Instance: max\_air\_temp

Aggregation method:

The method the value was aggregated or calculated: e.g. maximum value, average.

The value of the observation item attribute  $i$  represents a combination of these characteristics. To determine the characteristics of  $i$  strictly, we use ontologies to describe these characteristics. Ontologies, such as the SWEET ontologies [19] maintained by the Jet Propulsion Laboratory [20], can provide classes that suits our needs. In figure 2, we show the correspondence between the characteristics and the ontology we can use from the SWEET ontologies.

Using ontologies, we can describe the observation item using RDF [21] Figure 3 depicts an example of an RDF graph representation of an observation item, max\_air\_temp: the highest value of air temperature measured.

The prefix  $ex$  used in Fig. 3 denotes the namespace of the ontology used in the DIAS project, which imports the SWEET ontologies and extends its vocabulary.

### 2.1.4 Value Attribute

The value attribute value  $v$  is a representation of the actual value observed, simulated or calculated data, and its unit of measurement, if it exists.  $v$  might represent values such as directions or weather, as well as scalars. In addition, null values might be used to indicate missing values.

### 2.1.5 Example

We portray an example of earth observation data in Fig. 4, as derived from a value in Table 1. We used ISO standards to describe the spatial and temporal attributes. However this is merely an example. We have no intention of specifying how to implement the description of the attributes. The value of the observation item attribute mean\_air\_temp represents the average air temperature of the day.

In our data model, each data bears spatial attributes, temporal attributes and observation item attributes. In most cases, however, earth observation data are provided as a dataset. In this subsection, we discuss how to treat a set of earth observation data, namely an earth observation dataset. We describe an earth observation dataset  $D$

Station list						
USAF	WBAN	STATION NAME	CTRY	LAT	LON	ELEV
:	:	:	:	:	:	:
477550	99999	HAMADA	JP	JA	34.9	132.067
477560	99999	TSUYAMA	JP	JA	35.067	134.017
477590	99999	KYOTO	JP	JA	35.017	135.733
477610	99999	HIKONE	JP	JA	35.283	136.25
477620	99999	SHIMONOSEKI	JP	JA	33.95	130.933
:	:	:	:	:	:	:

Data file										
STN	WBAN	YEARMODA	TEMP	DEWP	SLP	STP	VISIB	WDSP	MXSPD	GUST
477590	99999	20080101	37.4	25.6	1012.5	1006.7	15.5	3.3	5.1	999.9
477590	99999	20080102	39.8	28.9	1019	1013.2	17.1	3.5	8	999.9
477590	99999	20080103	42.9	28	1020.2	1014.5	21.7	2.7	6	999.9
:	:	:	:	:	:	:	:	:	:	:

Table 1: Example of earth observation data: WMO Resolution 40

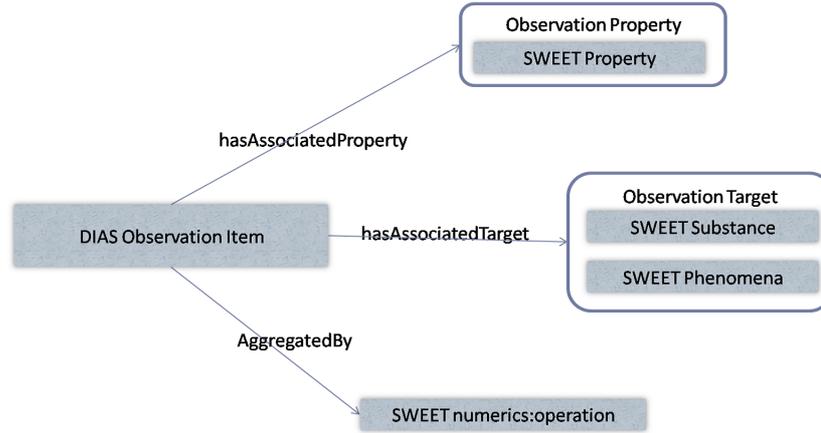


Figure 2: Usage of SWEET ontologies

```

(
wmoresolution40,
+35.017+135.733.
2008-01-01T00:00:00+09:00/
2008-01-02T00:00:00+09:00,
mean_air_temp,
37.4 F
)

```

Figure 4: Examples of data

as follows

$$\begin{aligned}
D &= \{d_1, \dots, d_n\} \\
&= (DS, S, T, I, V)
\end{aligned}$$

$DS, S, T, I, V$  represents a set of values of each attributes, and is defined as follows. .

$$\begin{aligned}
DS &= \{ds_1, \dots, ds_n\}, S = \{s_1, \dots, s_n\}, \\
T &= \{t_1, \dots, t_n\}, I = \{i_1, \dots, i_n\}, V = \{v_1, \dots, v_n\}
\end{aligned}$$

In the followings, we define some properties an earth observation dataset has.

Definition 1. Let  $D$  be an earth observation dataset, and

$i$  be a value of an observation item attribute.

The spatial extent of  $i$  in  $D$  is the minimum bounded rectangular region which includes every values of the spatial attribute of data included in  $D$  and has  $i$  as the observation item, and is denoted as  $|S_D(i)|$ .

Definition 2. Let  $D$  be an earth observation dataset, and  $i$  be a value of an observation item attribute, the temporall extent of  $i$  in  $D$  is the shortest duration which includes every values of the temporal attribute of data included in  $D$  and has  $i$  as the observation item, and is denoted as  $|T_D(i)|$ .

The next two values are defined only in specific cases.

Definition 3. Let  $D$  be an earth observation dataset, and  $i$  be a value of an observation item attribute. If all the values of the spatial attribute of data included in  $D$  and has  $i$  as the observation item are in the same shape and area, we call the shape and size the spatial resolution of  $i$  in  $D$  and denote it as  $\lambda_D^s(i)$ .

Definition 4. Let  $D$  is an earth observation dataset, and  $i$  is a value of an observation item attribute. If all the values of the temporal attribute of data included in  $D$  and has  $i$  as the observation item have the same length, we call the length the time cycle of  $i$  in  $D$  and denote it as  $\lambda_D^t(i)$ .

As you can see in the models NetCDF or HDF-EOS supports, there are three principal types of geographic distribution of earth observation data, namely point, grid and swath. When data in  $D$  which has  $i$  as the observation value are distributed in the form of a grid,  $\lambda_D^s(i)$  is defined as follows:

$$\lambda_D^s(i) = (|lat_i|, |lon_i|)$$

Where  $|lat_i|, |lon_i|$  are the length of zonal and meridional edges of the spatial resolution respectively. In this case, we can define an order among the spatial resolution as follows:

$$\lambda_D^s(i_0) \geq \lambda_D^s(i_1) \iff$$

$$|lat_{i_0}| \geq |lat_{i_1}| \wedge |lon_{i_0}| \geq |lon_{i_1}|$$

$$\lambda_s(i_0) = \lambda_s(i_1) \iff$$

$$|lat_{i_0}| = |lat_{i_1}| \wedge |lon_{i_0}| = |lon_{i_1}|$$

If the spatial resolution or the time cycle is common in every  $i$  of  $D$ , we simply denote them as  $\lambda_s, \lambda_t$ .

From a discussion at W3C, annotation can be defined loosely as: [22].

Any object that is associated with another object by some relation.

Today, annotations on the Web arise in various forms. They can be RDFs, or simple notes or comments, or a number of stars to express users' preferences. That definition might be acceptable considering most contents of the Web, but we require a slight alteration in the definition to discuss annotation for earth observation data. Within our model, annotation is defined as a relation between earth observation data and annotation data.

We model an annotation datum, denoted as  $a$ , as the following triple.

$$a = (u, t, c)$$

Therein,  $u$  describes a user who practiced the annotation,  $t$  is the valid time of the annotation, and  $c$  is the content of the annotation. We denote an annotation  $A$  of a set of earth observation datum  $D$  with annotation datum  $a$  as follows.

$$A = (a, D)$$

There is a demand among data users to aggregate earth observation data and annotate them at once when annotating earth observation data. Figure 5 portrays models of such annotation. In that figure,  $D$  and  $A$  respectively denote conceptual vector spaces where earth observation data and annotation data are denoted as points: annotations  $a, b$ , and  $c$  represent the three general types of annotation.

Annotation  $a$  represents a single data annotation by which the subject of the annotation is an earth observation data instance. We denote such annotation as follows.

$$A = (a, (DS, S, T, I, V))$$

Annotation  $b$  and  $c$  are annotations whose subject is a region in vector space  $D$ . We allow two methods to denote these kinds of annotations. The first method is to ignore some dimensions of the earth observation dataset. For example, we might want to annotate every datum with the same dataset attribute, temporal attribute, and observation

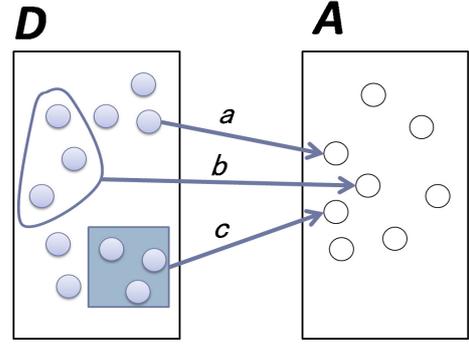


Figure 5: Models of annotations

attribute. To meet this requirement, we use an asterisk to represent “do not care”. Such annotation with the do not care attribute is written as follows.

$$A = (a, (ds, *, t, i, *))$$

The second method is to use comparison expressions to determine a subset of a dataset. We allow the use of selection conditional expressions, which is defined as follows, in the annotation.

**Definition 5.** When  $X(\in \{DS, S, T, I, V\})$  is an attribute of dataset  $D$ ,  $Y$  is a set of value of  $X$ , and  $\theta \in \{\in, \notin\}$  denotes a membership operator,  $X\theta Y$  is a conditional clause of  $D$ . In addition, when  $X(\in \{S, T, V\})$  is an attribute of dataset  $D$ ,  $y$  is a value constant  $\theta$  is a binary operation in the set  $\{<, >, \geq, \leq, =, \neq\}$   $X\theta y$  is also a conditional clause of  $D$ . The selection conditional expression is defined as shown below.

1. a conditional clause of  $D$  is a conditional expression of  $D$ .
2.  $\neg l$  is a conditional expression of  $D$  when  $l$  is a conditional expression of  $D$ .
3.  $l_1 \wedge l_2$  is a conditional expression of  $D$  when  $l_1, l_2$  are conditional expressions of  $D$ .
4.  $l_1 \vee l_2$  is a conditional expression of  $D$  when  $l_1, l_2$  are conditional expressions of  $D$ .

We give an example of an annotation in which information is annotated to data with  $ds, s, i$  as dataset, spatial, observation item attributes respectively, and temporal attributes that represent that represent durations after date  $X$ .

$$A = (a, (ds, s, t \geq X, i, *))$$

We give a practical example of an annotation for additional understanding.

$$A = ((Akira, 2008-08-31T15:00:00+09:00, systematic\ error), (ds, *, t < 1990-01-01, i \in \{air\_temperature, precipitation\}, *))$$

This annotation denotes that a systematic error exists in data where their dataset attribute is  $ds$ , the observational

```

A = ( ( Akira Takahashi,
2007-08-31T15:00:00+09:00,
<iso:CI_ResponsibleParty>
  <iso:individualName>
    Akira Takahashi
  </iso:individualName>
  ...
</iso:CI_ResponsibleParty>
),
(ds, *, *, *, *, *)

```

Figure 6: Annotation with XML documents

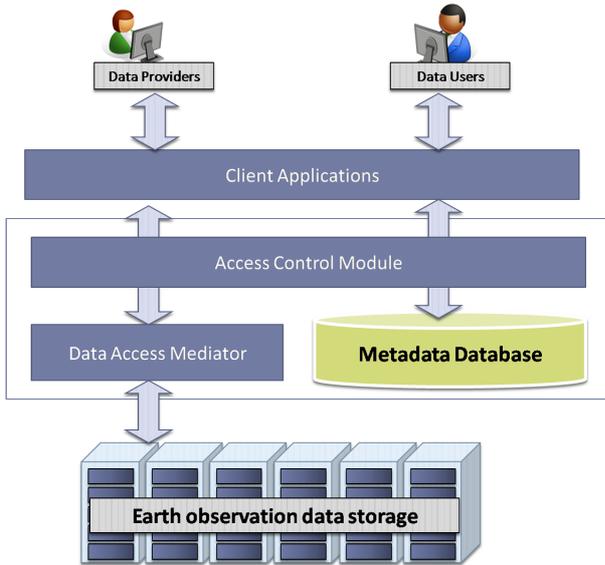


Figure 7: System Overview

items are air temperature and precipitation measured at any place before January 1, 1990.

In our data model, we specify no syntax that an annotation content might take; generally, no restriction defines what users can annotate. However, it might be useful if the annotation were available in a machine-readable format. Additionally, we might want to specify the semantics of the annotations to distinguish them and avoid mutually exclusive annotations. Therefore, we use XML documents for annotation contents. Using well-known schemas for marking up the annotation might increase the interoperability of the annotation. We present an example of an annotation using markups with classes defined using iso19115 metadata standards [1] in Fig. 6

### 3. PLAN OF SYSTEM IMPLEMENTATION

In this section, we explain about our system we have implemented, and introduce some application interfaces for metadata annotation.

#### 3.1 Overview of Our System

Figure 7 represents the overview of our systems

The data access mediator gives the mapping between our conceptual earth observation datum model and the actual

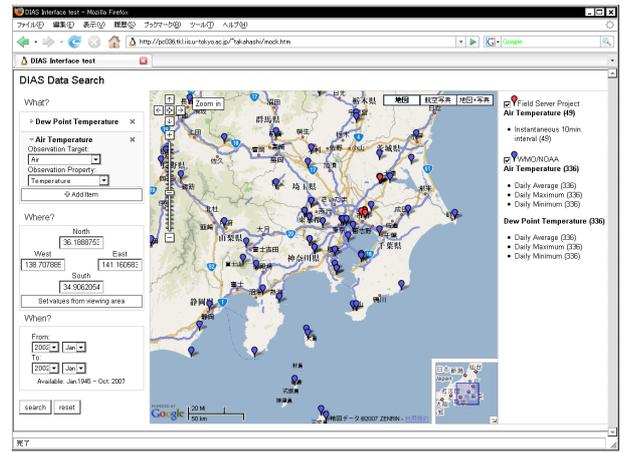


Figure 8: The Data Retrieval Interface

data model used in the underlying storage. Users can retrieve data by specifying conditions for the quintuple, with no aware of the data model schema used in the actual data storage. Various data formats used in the DIAS project, therefore the schema mapping is currently created manually. Automatic processing of the schema mapping is one of our future works.

Many earth observation data products has restrictions for their usage, and users may want to control access to metadata they have created. Therefore we incorporate an access control module to manage user accounts and how they can access each data. Client softwares and applications can be built upon this access control module.

#### 3.2 Application Interfaces

In this section, we describe some implementation of user interfaces. By utilizing metadata, users can obtain supplemental information of datasets, and encourage further understanding of datasets. However, we must motivate users to enter metadata to gain enough information. This is a difficult challenge for designing user interactions. If users are not aware of how they can benefit from entering metadata, no one will provide metadata, and the system will be forced to a halt. Our policy for designing user interfaces is to show how effective metadata annotation is as soon as possible.

##### 3.2.1 Data retrieval interface

We have previously implemented an data retrieval system [23] for earth observation data. The design policy of the interface is to effectively combine searches from three aspects: item, region, and time. In addition it aims to achieve a sight grasp of the data with availability of data shown on a map-based interface.

Figure 7 gives the screen of the web-based interface we have developed. The left column is used to input queries, and the right column show summary information of the query. The center column is a map interface using the GoogleMapsAPI[24], and sites corresponding to the query will be visualized on the map.

In the query input part, you can specify observation item, region, and period to narrow users result. Specifying item is done by specifying observation target, property and inter-

val respectively. Users can specify two or more observation items, and also decide whether the items should be contained in the result site (AND retrieval) or at least one of the alternative is contained (OR retrieval). We can search observation point within a rectangular area obtained by specifying the upper bound and the lower bound latitude and longitude. When either of three parameters is specified, the alternative of the other parameters is automatically limited, according to the specified value, to make sure we can at least obtain one site as a result. The system will automatically query the database to prefetch results and show the summary on the right and the center column. Users can easily see what kind of data is available to what extent, therefore the user's efficient retrieval behavior are supported.

We are planning to incorporate annotating function into this interface. With the prefetching function, we see the query result in the center map column of the interface. By simply clicking the marker representing a result, we may see every annotation data, annotated to the particular dataset at the right column. Users can then investigate the annotation, choose to download the data, or annotate another information to the corresponding dataset.

### 3.2.2 Metadata annotation interface with document metaphor.

Since the data providers often understands the data the most, large portions of structured metadata are generated by the data providers. However, conventional metadata publishing tools, such as GeoNetwork<sup>5</sup>, NOAA ArcView Extension<sup>6</sup>, requires profound understanding of the metadata schema. In order to avoid extra studying of metadata schema, we provide a document metaphor annotation interface. Data providers often manage and publish documents to explain the datasets they have produced. By providing tools for generating such document within our framework, users can save trouble of describing both the document and the input metadata. Figure 9 gives the overview of the function of the interface. The document metaphor annotation interface is a form style editor for metadata. Users can access the interface through conventional web browsers. The interface provides the section titles of the document, and users fill out the content of each sections (Fig. 10). If the dataset the user is going to refer is described in NetCDF or GrADS formats, some portions of the sections can be filled out with metadata extracted from the data. Also, some of the information about the data provider may be automatically filled out, if the user information is registered in advance. Inputs are stored in databases, and users can either download the document in Portable Document Format(PDF), or publish the document in HTML.

## 4. SUMMARY AND FUTURE WORKS

In this paper we proposed an conceptual data model for annotating earth observation data. Utilizing the conceptual model enables users to state metadata without concerning data models used in the actual data storage, and preserve user semantics of the annotations. We also introduced our metadata management system. We are currently implementing additional user interfaces for utilizing metadata. Our future work includes collaboration of annotation

<sup>5</sup><http://geonetwork-opensource.org/>

<sup>6</sup><http://www.csc.noaa.gov/metadata/download.html>

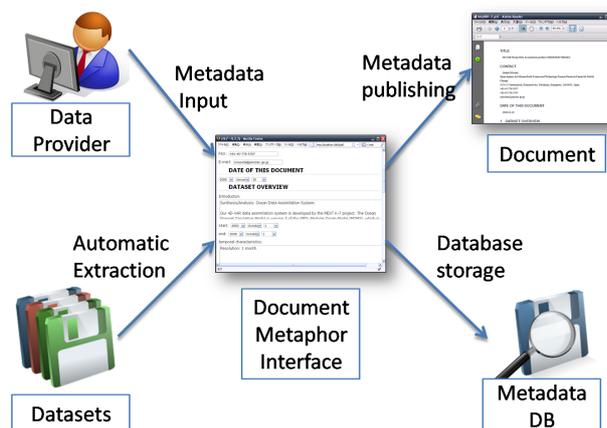


Figure 9: Document Metaphor interface

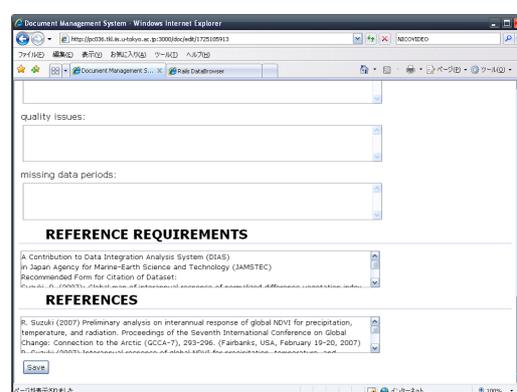


Figure 10: Metadata registration interface

between users, and management of data provenance information within our framework.

## Acknowledgements

This research is supported by “Data Integration and Analysis System” funded by the National Key Technology, Ministry of Education, Culture, Sports, Science and Technology, Japan.

## 5. REFERENCES

- [1] International Organization for Standardization. Iso 19115:2003, geographic information metadata.
- [2] Federal Geographic Data Committee. Content standard for digital geospatial metadata. fgdc-std-001-1998,, June 1998.
- [3] Youtube. <http://www.youtube.com/>.
- [4] Wikipedia. <http://en.wikipedia.org/wiki/>.
- [5] Facebook. <http://www.facebook.com/>.
- [6] delicious. <http://delicious.com/>.
- [7] Netcdf(network common data form). <http://www.unidata.ucar.edu/software/netcdf/>.
- [8] Grid analysis and display system (GrADS). <http://www.iges.org/grads/>.

- [9] José Kahan, Marja-Riitta Koivunen, Eric Prud'hommeaux, and Ralph R. Swick. Annotea: an open rdf infrastructure for shared web annotations. *Computer Networks*, Vol. 39, No. 5, pp. 589–608, 2002.
- [10] Daniel LaLiberte and Alan Braverman. A protocol for scalable group and public annotations. *Computer Networks and ISDN Systems*, Vol. 27, No. 6, pp. 911–918, 1995.
- [11] Matthew A. Schickler, Murray S. Mazer, and Charles Brooks. Pan-browser support for annotations and other meta-information on the world wide web. *Computer Networks*, Vol. 28, No. 7-11, pp. 1063–1074, 1996.
- [12] biodas.org. <http://biodas.org/>.
- [13] Robin D. Dowell, Rodney M. Jokerst, Allen Day, Sean R. Eddy, and Lincoln Stein. The distributed annotation system. *BMC Bioinformatics*, Vol. 2, p. 7, 2001.
- [14] Yingwei Cui and Jennifer Widom. Lineage tracing for general data warehouse transformations. *VLDB J.*, Vol. 12, No. 1, pp. 41–58, 2003.
- [15] Peter Buneman, Adriane Chapman, and James Cheney. Provenance management in curated databases. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 539–550, 2006.
- [16] Rajendra Bose. A conceptual framework for composing and managing scientific data lineage. In *SSDBM*, pp. 15–19, 2002.
- [17] Trio, a system for integrated management of data, uncertainty, and lineage). <http://infolab.stanford.edu/trio/>.
- [18] National oceanic & atmospheric administration. <http://www.noaa.gov/>.
- [19] Semantic web for earth and environmental terminology (sweet). <http://sweet.jpl.nasa.gov/ontology/>.
- [20] Jet propulsion laboratory (jpl), <http://jpl.nasa.gov/>.
- [21] Resource description language (rdf), <http://www.w3.org/rdf/>.
- [22] Collaboration, knowledge representation and automatability, w3c. <http://www.w3.org/Collaboration/>.
- [23] Akira Takahashi, Hiroko Kinutani, and Masatoshi Yoshikawa. Research for an integrated earth observation data retrieval framework based on metadata development with ontologies. In *Proceedings of the IPSJ Symposium on DataBase and Web (DBWeb2007)*, 2007.
- [24] Google maps api. <http://www.google.com/apis/maps/>.