METADATA PUBLICATION AND SEARCH SYSTEM IN JAMSTEC

Y Hanafusa[1], H Saito[2] and Y Abe[3]*

[*1] *Data Research Center for Marine-Earth Sciences, Japan Agency for Marine-Earth Science and Technology,*
3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan
Email: hanafusay@jamstec.go.jp
[2] *Data Research Center for Marine-Earth Sciences, Japan Agency for Marine-Earth Science and Technology,*
3173-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan
Email: saitoh@jamstec.go.jp
[3] *Marine Works Japan Ltd., 2-16-32 Kamariyahigashi, Kanazawa-ku, Yokohama, Kanagawa 236-0042, Japan*
Email: abe@mwj.co.jp

ABSTRACT

Japan Agency for Marine-Earth Science and Technology (JAMSTEC) provides users of its data with comprehensive search services which enable users to find data from JAMSTEC's various data dissemination sites. These are the "JAMSTEC Data Search Portal" which helps users to search for observational data on a map, and the "JAMSTEC Data Catalog" which enables users to find data sites by selecting science keywords. The "Data Search Portal" and the "Data Catalog" have been developed and operated as dedicated metadata publication and search services collaborating with data sites in JAMSTEC.

Keywords: Metadata, Web GIS, DIF, Data Search Service, JAMSTEC

1 INTRODUCTION

JAMSTEC has performed very wide variety of observation and research both on the Ocean and the Earth. In 2007 JAMSTEC established the data policy (http://www.jamstec.go.jp/e/database/data_policy.html) on handling of data and samples from its research activities, and developed rules and management systems based on the policy. JAMSTEC operates over 100 research cruises a year and disseminates observational data from these cruises in the "Data Site for Research Cruises (http://www.godac.jamstec.go.jp/cruisedata/e/)", images in the "E-Library for Deep Sea Images (http://www.godac.jamstec.go.jp/jedi/e/index.html)", rock sample information in the "GANSEKI (http://www.godac.jamstec.go.jp/ganseki/)", sediment core sample information in the "Core Data Site (http://www.godac.jamstec.go.jp/coredata/e/) and biological sample information in the "Marine Biological Sample Database (http://www.godac.jamstec.go.jp/bio-sample/index_e.html)". Additionally various research projects operate their own data dissemination sites.

As data management system in JAMSTEC was developed, the numbers of data sites and the amount of opened data increased rapidly. On the other hand users have to search for data in various data dissemination sites designed for the specific data type and there were many contacts to the data management office to inquire whether the specific data is opened or not, and how to find the data. Therefore, JAMSTEC has developed two data search services which enable users to find data by themselves easily.

One is the "Data Search Portal (http://www.godac.jamstec.go.jp/dataportal/index_eng.html)" which helps users to search for observational data by specifying the area of interest on a map. The other is the "Data Catalog (http://www.godac.jamstec.go.jp/catalog/data_catalog/index_en.html)" which leads users to the appropriate data dissemination sites by selecting science keywords. "Figure 1" shows the schematic flow chart in finding data using these two services. Both systems provide just search service based on the metadata. Users are able to learn the details of the data on the linked data dissemination page and, if needed, to download the data or apply for the use of the off-line data.
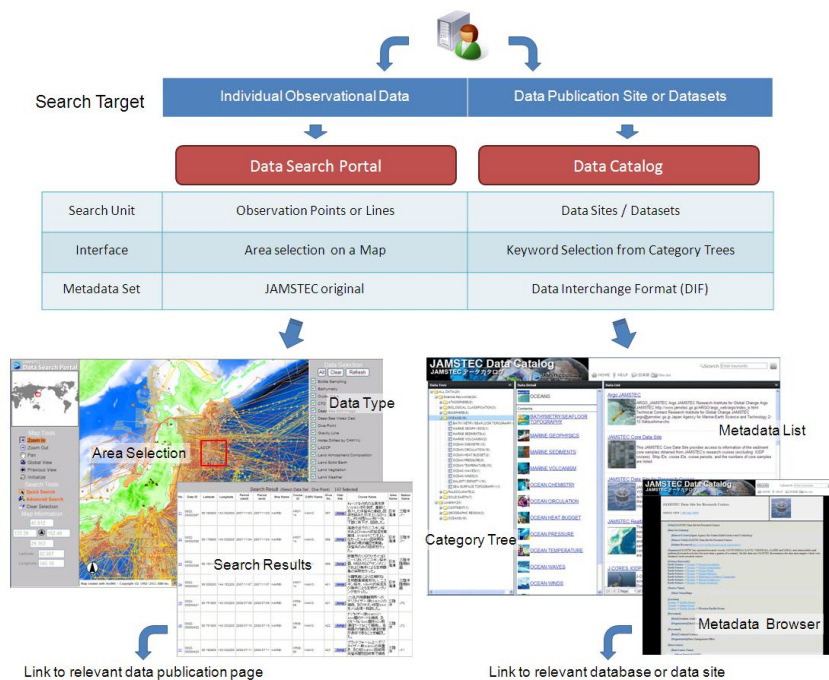
Figure 1. Schematic flow of data search using "Data Search Portal" and "Data Catalog".

2 DATA SEARCH PORTAL

2.1 Objective

The search target of the "Data Search Portal" is an individual observational data from research cruises, moorings and terrestrial observations. It was developed to serve comprehensive data search service through various data dissemination sites by specifying data types and the area of interest. The "Data Search Portal" was set in operation since November 2008.

2.2 Method

Using the "Data Search Portal" users are able to confirm the distribution of observations of specific data type and then select the area for search. If needed, additional criteria (research vessel name, cruise period, observation variable, etc.) can be added. The list of search results includes the common metadata (date, location, data id, etc.) and a link to the data dissemination site which enables users to move directly to the relevant data page. This is a simple URL link and the "Data Search Portal" can cooperate with any data system as far as a data page can be specified with URL.

The supposed main users of the "Data Search Portal" are scientists who are searching for the data in the same area at different time or other data type in the same area. So authors developed a map-based spatial retrieval function using a web GIS server (ArcIMS 9.2).

2.3 Metadata

Location information in points or lines is extracted from data files and merged into shape files with observation metadata. The metadata includes several common metadata for cruise or dive (cruise number, period, observation variable, etc.) and optional metadata for a specific data type (chief scientist, water depth, area name, etc.). The vocabulary of observation variable is determined originally based on the frequently observed variables in JAMSTEC. Data types in the "Data Search Portal" come mainly from research cruises or dives but some data types come from mooring observations, ocean bottom stations and terrestrial observations. At the end of 2011

over 30,000 observations shown in "Table 1" are opened on the "Data Search Portal".

Table 1. Data types and number of observations on the "Data Search Portal"

Research Field	Data Type	Number of Records
General	Cruise Track, Dive Point	5,764
Oceanography	Temperature and Salinity Profile, Water Chemical Analysis, Current Profile, Primary Production, Sediment Core, etc.	16,635
Meteorology	Marine Meteorology, Terrestrial Meteorology, Atmospheric Composition, Fixed Point Observatory, etc.	196
Solid Earth	Bathymetry, Gravity, Magnetics, Rock Sample, Drilling Hole, etc.	2,981
Biology	Marine Biological Sample, Vegetation	1,175
Images	Still Image, Video	4,824
Total		31,575

3   DATA CATALOG

3.1   Objective

While the target of the "Data Search Portal" is individual observation, the target of the "Data Catalog" is a larger data dissemination unit (data sites, databases, datasets, etc.). JAMSTEC aims to provide comprehensive metadata publication and search service for all of the data in JAMSTEC with these two systems complementally. The "Data Catalog" is also designed to be a data publication system for scientists who do not have data dissemination sites by themselves. JAMSTEC has opened the "Data Catalog" on September 2011.

3.2   Method

The user interface of the "Data Catalog" is a classification tree of hierarchical keywords. Supposed users of the "Data Catalog" are scientists searching for data not in their main research field. They are not familiar with the appropriate keyword to search for data in the field. It is easy to understand for those users to select keywords from the tree confirming the structure of the keywords.

The hierarchical keywords of research fields or observations are listed in the category tree. Users are able to narrow down the list of metadata by selecting keywords in the category tree. Going down the tree, related sub categories and metadata of the keyword will appear as shown in "Figure 2".

The content of each data site is not fully described in the common metadata format. Metadata in the "Data Catalog" is just for finding data sites. Details of the data site and data download depend on the each data site. Users are also able to search for metadata by free text.

Figure 2. Selecting a keyword in the category tree in the "Data Catalog" brings a list of relevant metadata.

3.3   Metadata

Authors adopted the "Directory Interchange Format (DIF)" in the "Global Change Master Directory (GCMD)" as the metadata standard of the "Data Catalog". The structure of DIF metadata standard are shown in the "Directory Interchange Format (DIF) Writer's Guide (http://gcmd.nasa.gov/User/difguide/)". The vocabulary of the keywords in the category tree is also based on the "Earth Science Keywords (Olsen, Major, Shein, Scialdone, Vogel, Leicester et al., 2007)" of GCMD. Because the DIF and its "Earth Science Keywords" cover a wide variety of research fields in the Ocean and the Earth sciences and keywords are controlled by GCMD, it is suitable to describe various research data in JAMSTEC uniformly. The metadata in the "Data Catalog" is also registered to the GCMD.

For domestic users Japanese version of metadata and a set of Japanese keywords of the "Earth Science Keywords" were made. In the Japanese page of the "Data Catalog" users are able to search for and browse metadata in Japanese. A specific metadata page in English is linked to the relevant Japanese page, and vice versa.

4   TOWARD THE "ONE STOP DATA SHOP"

JAMSTEC is going to develop easier and more effective data systems for users. One supposed solution is a "One Stop Data Shop", which enables users to search for, browse and get all data in JAMSTEC from a single site. We think "Data Search Portal" and "Data Catalog" are possible prototypes of search system in the "One Stop Data Shop". Integration or cooperation with other data systems in JAMSTEC or other institutions will be expected.

5   ACKNOWLEDGEMENTS

6   REFERENCES

Olsen, L.M., G. Major, K. Shein, J. Scialdone, R. Vogel, S. Leicester, H. Weir, S. Ritz, T. Stevens, M. Meaux,

C.Solomon, R. Bilodeau, M. Holland, T. Northcutt, R. A. Restrepo, 2007 .NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 6.0.0.0.0