The Olympus Data Information Sharing System ™

By Marius Necsoiu, Ph.D.

Electronic information technology has become a crucial component in research, business, and government. We have come to rely on the ability to retrieve information quickly, accurately, and easily. Because of this, many enterprises no longer treat their information repositories as only tools for decision-making. The knowledge base itself has become an increasingly important corporate asset.

Scientists and engineers at the Center for Nuclear Waste Regulatory Analyses (CNWRA), located at Southwest Research Institute (SwRI), faced the difficult task of managing more than 15 years of information and creating a knowledge-based system to allow quick and easy access and searches of this data. The CNWRA was established in October 1987 as a federally funded research and development center to provide the Nuclear Regulatory Commission an independent assessment capability related to the licensing of a proposed underground geologic repository for high-level radioactive waste. The facility is intended to safely house and isolate 70,000 tons of high-level radioactive waste from the environment for at least 10,000 years. The CNWRA provides geoscience expertise in structural geology, volcanism, seismology, hydrology, geomorphology, geophysics, general geology, and physical geography — all of which require extensive use of maps and other spatial and analytical data.

The CNWRA identified the requirements for designing a data management system to allow full access for its staff to stored and archived data. A newly developed, integrated approach to solving data management problems was needed to facilitate interdisciplinary collaboration. CNWRA and U.S. Nuclear Regulatory Commission researchers needed quick and easy access to comprehensive, spatially referenced data such as maps and data points. The system also had to allow for extensive data handling, analysis, and map preparation with minimal training.

The Olympus Data Information Sharing System™ (Olympus DISS™) is an intranet web-based geographical data and information sharing system that allows access to data, including the original source material (Necsoiu, 2001). The system uses established data standards and allows for the output of geographic data in multiple formats.

Background

Flexibility was a key element of the system design because it had to ingest a variety of data formats such as raster (aerial and satellite images), vector (maps or illustrations), and tabular (table) data. SwRI also wanted to maximize its investment: The system needed to be "self-enriching." Collecting data is expensive and time-consuming. The new system had to be simple and user friendly – as simple as accessing the Internet so researchers could focus on analysis or algorithm development rather than format conversion.

The new system accesses data over the Intranet and, if necessary, combines it with data downloadable from the web. Olympus DISS™ also provides a secure way to

control how and to whom data and services are delivered. Each researcher can control whether the data holdings are accessible to only a limited number of people or to the whole data-sharing community. This allows the safe storage of proprietary data. The system is also easily scaled to meet changing demands.

Change is the essence of geospatial data in a networked environment. Such data uses spatial descriptors, organization, and relationships taken from remote sensing, mapping, surveying, and global positioning systems. Once created, data is accessible almost instantaneously through a network and can be used for many kinds of spatial analysis. Information can then be reused and analyzed for a new situation or retransmitted to another user.

How does one track changes? A simple answer would be to associate a metadata file to each geographic dataset. Metadata, or "data about data," typically contains more information about the data, offering multiple sources of information in one larger file. More specific metadata contains descriptive information about the content, quality, condition, and origin as well as other characteristics.

For the Olympus geo-information infrastructure, spatial metadata records are essential. Metadata facilitates data identification by search and retrieval mechanisms based on the user's selection criteria. It allows a user to fully understand the content and evaluate its usefulness by providing downloadable geographic datasets.

The required level of detail of metadata depends on its purpose. Data managers need very specific information on data format, internal structures and data definitions. Users generally require a kind of "catalogue" of information as to where to find certain

data, how to use it, and who originated it. Olympus' architecture accommodates both of these purposes — it can ingest both incomplete and complete sets of records. The incomplete set can be updated as more data becomes available.

Metadata Production

The information needed to create metadata is often readily available at the time data are collected. A small amount of time invested at the beginning of a project may save significant expenses later. Data producers and users cannot afford to be without documented data. The initial expense of documenting data clearly outweighs the potential costs of duplicated or redundant data generation. Recently developed metadata standards such as Federal Geographic Data Committee Content standards for Digital Geospatial Metadata (FGDC, 1998) provide a systematic way to collect metadata.

Typically, providing metadata sets is very labor-intensive and expensive. For the Olympus DISS™, however, all CNWRA technical staff can contribute in generating information records. Thus, each researcher can process and organize data and create searchable keywords while developing the geodata or shortly afterwards, if desired. By distributing the workload, the expenses and time are efficiently distributed.

The Olympus DISS™

Distributed DISS has been successful for other research facilities, but the CNWRA system distinguishes itself with its flexible design, simplicity, and minimal cost.

Olympus DISS™ provides the necessary organization and search capabilities for an intranet web-based system. Because Olympus uses established data standards, such as the FGDC, it provides a flexible mechanism for future work with the data.

The system software is centralized, using software components such as ArcCatalog, a commercial off-the-shelf software component produced by ESRI, Inc.; metadata parser (MP), and the Isite information system, two public domain packages produced by the USGS and the Center for Networked Information Discovery and Retrieval (CNIDR), respectively; and Harvester, a software component produced by the CNWRA.

Applications

Olympus DISS™ provides search and retrieval mechanisms for querying a metadata database, containing records associated with each specific geographic dataset or geodata. Each metadata file contains information that describes geodata in the same way a card in a library card catalog describes a book.

Olympus DISS™ can create and "ingest" metadata from a variety of geodata formats: ERDAS IMAGINE, TIFF, MrSID, JPEG, ERDAS 7.5 LAN, ERDAS Raw, ESRI GRID Stack File, ESRI Shapefile and ESRI Arc/Info coverage. Once metadata and the associated geodata are created, the user places it in a designated repository area. Daily, the Olympus system "harvests" the metadata in the repository and automatically builds an index and a relational database with metadata information. Olympus DISS™ provides geographic (spatial), keyword, and temporal search and retrieval capabilities

for the repository of geographic data. Search and retrieval are done through a web-based graphical user interface consisting of a login page, a search page, and results and metadata pages.

Spatial searches allow three methods of entering data for queries: to enter geographic coordinates in the provided text fields; to simply draw a box around the desired geographic location on a map of either the United States or the world; or to select, from a dropdown menu, the state (in the United States) that the user wishes to search. If a map is used, the select tool can be personalized by customizing both the color and style from two dropdown menus. The selectable style could be Point, Point (Compressed X and Y), or XY Plane. In all methods, the geographic coordinates could be the values queried in the database.

The result page lists the term(s) queried, the number of matching records found, the number of records currently being viewed, and the titles and links to metadata describing the available data. From this page, the user can go to each geo-dataset metadata page. Each metadata page provides information of how to download data.

Recently, a new feature was added that allows storing and retrieving non-spatial data such as photos, graphs and engineering drawings.

Future Applications

The Olympus DISS™ uses a standard desktop computer with an Intranet connection and web-browsing tool. Data is stored in a server. The system contains more than 1,500 records and is continuing to grow. Many of these records consist of legacy data that could be quickly retrieved for use in multiple projects.

The CNWRA anticipates benefiting from the Olympus DISS™ as a central repository for all CNWRA geospatial data, allowing for better data management. This will allow individual desktop machines to be used for research and analysis, rather than long-term data storage. This system provides better quality control and reduces redundant data. The Olympus DISS™ is a fast, convenient way of accessing geospatial data. Data backups are more efficient in time and space requirements, because only the current research data is handled. The system has an open architecture that can be configured to interface with other data management systems that uses the Z39.50 information retrieval protocol standard and the FGDC metadata standard. It could easily be implemented for data sharing across the Institute or across a company. Implementation costs are minimal.

Comments about this article? Contact Necsoiu at (210) 522-5541, or e-mail at mnecsoiu@swri.org.

Acknowledgments

The author acknowledges the contributions of the Olympus DISS™ Team: Brandi Winfrey for web design and interface coding issues, and Katherine Murphy for compiling metadata records. Dr. Larry McKague, Dr. English Pearcy, and the Geology and Geophysics and the Geohydrology and Geochemistry teams provided helpful critiques and constructive reviews of the manuscript. Special thanks to Dr. Wesley Patrick and Dr. Budhi Sagar for their support in developing this system.

Bibliography

Necsoiu, Marius. Capability Development: Olympus CNWRA's Distributed Data and Information Sharing System (White Paper). June 26, 2001.

Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998). Federal Geographic Data Committee. Washington, D.C.