A DISTRIBUTED CATALOG AND DATA SERVICES SYSTEM FOR REMOTE SENSING DATA

Ramachandran Suresh[*], Liping Di[*], Kenneth McDonald[**]
[*]NASA/RITSS
4500 Forbes Blvd, Lanham, MD 20706, USA
suresh@rattler.gsfc.nasa.gov
lpd@rattler.gsfc.nasa.gov
[**] NASA GSFC Code 423, Greenbelt, MD. 20770 USA
kenneth.mcdonald@gsfc.nasa.gov

KEY WORDS: WWW, geospatial data, metadata, catalog, Remote Sensing, visualization.

ABSTRACT

Data and Information Access Link (DIAL) is web based package software tools for Remote sesning and geo-spatial data applications. Using DIAL, scientists can set up a web based data server, organize data, build metadata catalog and distribute data. Users with WWW browser can access DIAL. DIAL has implemented EOS Data Gateway EDG and CIP protocols through which projects can set up a distributed data and metadata catalog system. The use of this system enhances data interoperability.

1. INTRODUCTION

Before the advent of WWW technology, the process of setting up a data system or center to distribute and exchange scientific data was a difficult and expensive process. Currently, remote sensing data flows from producers to users mainly through data centers by way of specialized data systems that are expensive and involves significance maintenance costs. The developments in WWW technology has made it possible to set up low cost data distribution systems for satellite remote sensing data and field campaign data. These technologies and enhanced hardware and software products enable users with moderate levels of computer knowledge to establish information distribution sites. The web will enable scientists and projects to set up metadata catalogs, and provide browsing, visualization and on-line downloading of data. The web will make it easy for scientists to collaborate and exchange scientific data on-line.

In order to take advantage of Web technology for Remote Sensing data, NASA funded the Data and Information Access Link (DIAL) project to provide a low-cost web-based solution for managing and distributing Earth remote sensing and geo-spatial data. The DAIL has made it possible for small data producers and projects to set up a data system on their desktop and provide data services. The DIAL system is also capable of setting up distributed data systems that can inter-operate and exchange scientific data by the implementation of certain standard protocols. This paper describes the DIAL system.

2. DIAL DESCRIPTION

DIAL is a web based package of software tools that will help to set up a catalog and data services system for remote sensing data. DIAL was developed as a NASA technology prototype system to assist data producers and projects to organize and distribute their data and metadata.

DIAL is a web-based software system for distributing scientific data through the Internet (Di et.al., 1997, 98, McDoanld et.al.,1998). DIAL is designed to be a client-server based data and information system that, while powerful, is also compact, easy to setup and use, and has minimal computer power and maintenance requirements. It permits data producers to set up a server rapidly on their desktop computers, making data available via the Internet. Any users with a regular Web browser can interactively search, browse, subset, subsample, reformat, and download data.

2.1 Architecture

The DIAL architecture is modular, extendible, and is based on standards. Figure 1 shows the architecture. The DIAL system consists of three components: the end user interface, the DIAL server system, and the data management tools. The end user interface on the client side includes web browser with a built in Java or HTML interface for querying the system to locate datasets of interest and for interactively manipulating the datasets. The user requests are relayed to DIAL server system through http protocols. DIAL end users can easily integrate other application software as helper applications. The DIAL server system is CGI-based programs which responses to users' requests and sent the reply to the end user interface through an HTTP server. The DIAL server system works with the data in HDF or HDF-EOS formats. The metadata is stored in a binary table or a database either in Object Description Language (ODL) or Parameter Value Language (PVL) format. The interoperability layer in the server provides interoperations with other data systems through standards. The data management tools help the data providers to prepare the data and metadata for distribution through the DIAL system.

DIAL uses WWW standards such as http, html and Java applets in implementation. It can work with any metadata standards, such as FGDC, since its metadata catalog system is totally configurable and a mapping mechanism between different metadata systems is provided.

The DIAL software package is compact (less than 7 Megabytes) and does not need programming knowledge to use. Since it is completely developed from public domain software and protocols, it is not dependent on any COTS package. It is available on multiple platforms such as SGI, SUN, DEC, and PCs (Windows 95/98/NT). The source code is written in C and Java.

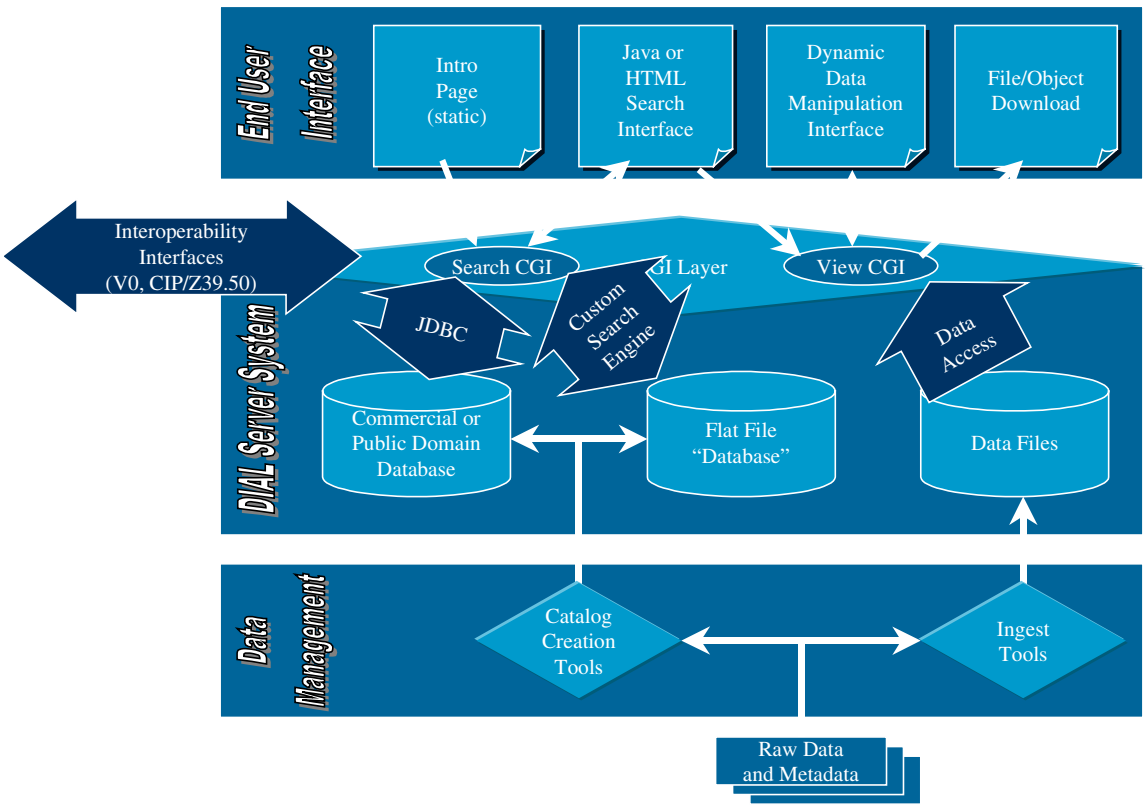

Figure 1. Detailed Architecture

2.2. Functionality

The DIAL system provides two major types of user services: the catalog service and the data service. The catalog service allows data users to find individual data objects (granules) in an archive by entering some search criteria. Once a user finds the needed individual data objects in the archive, the data

service will allow the user to manipulate and download the objects. The DIAL software is available for users free of charge at http://dial.gsfc.nasa.gov

Catalog services:

- Spatial, temporal, and parameter-based search
- catalog search at inventory and directory levels
- Automatic creation of catalog based on metadata
- Supports of ODBC/JDBC compatible databases for storing metadata/catalog
- User-friendly Java search interface with query preview
- HTML search interface for slow network connection

Data services

- On-line access to data and metadata
- Single- and multi-granule subsetting and subsampling based on array coordinates or record numbers
- Single- and multi-granule subsetting and subsampling based on geographic/map coordinates and physical parameters
- Browse and dynamic/interactive visualization of data
- Animation of time series or high dimensional data
- Interactive color composites of multi-spectral data
- Overlay of coastal and political boundaries
- On-line downloading of data in multiple formats
- X-Y plotting for non-image data

Although data users can download data managed by DIAL in multiple formats (e.g., binary, ASCII, HDF-EOS, HDF, and HTML), in the server side it currently works with HDF and HDF-EOS and netCDF formatted data sets. Data translation programs, which convert data between HDF-EOS and commonly used GIS formats, such as Arc/Info Exchange format, GeoTIFF, ERDAS LAN, and ArcView Shape, are available. In addition, DIAL's architecture allows extension for accessing data in other formats.

DIAL provides two options to store the searchable catalog: 1) a binary table on which DIAL's own search engine will search directly; and 2) ODBC-compatible databases on which DIAL will search through its generic JDBC interface. The binary table option provides the same search capabilities as the ODBC-compatible databases, but the search speed will be slowed down when the table is very large.

3. DISTRIBUTED DIAL SYSTEM

Currently, DIAL can be used either as a standalone server or as a distributed server. DIAL can work as a distributed system connecting several DIAL sites through the EOS Data Gateway (EDG), (Suresh et.al, 1999). DIAL supports EOSDIS "Version 0" protocol, extending its catalog interoperability to EOSDIS V0 community. DIAL will support Catalogue Interoperability Protocol (CIP) by September 2000 as it is the de-facto international standard for catalog interoperability of geo-spatial data. Supporting CIP protocol will enhance the data interoperability to a wider community of remote sensing data. Figure 2 shows the interoperability architecture of DIAL.

3.1 Catalog Interoperability with CIP and EDG

The Catalogue Interoperability Protocol (CIP) was developed by the Committee for Earth Observation Satellites (CEOS) for facilitating the catalog interoperability among the international space agencies. Major space agencies have agreed to use this protocol for providing access to their catalogues. The CIP is based upon the international search and retrieval protocol Z39.50-1995.

The main objectives of CIP are to provide transparent distributed search, retrieval, browse, order and other services against Earth Observation catalogues. Using CIP, search and retrieval can occur transparently and independently of the physical structures or locations of the catalogues. It enables

clients to dynamically learn about services and features. It also allows the creation of and access to collections of data.

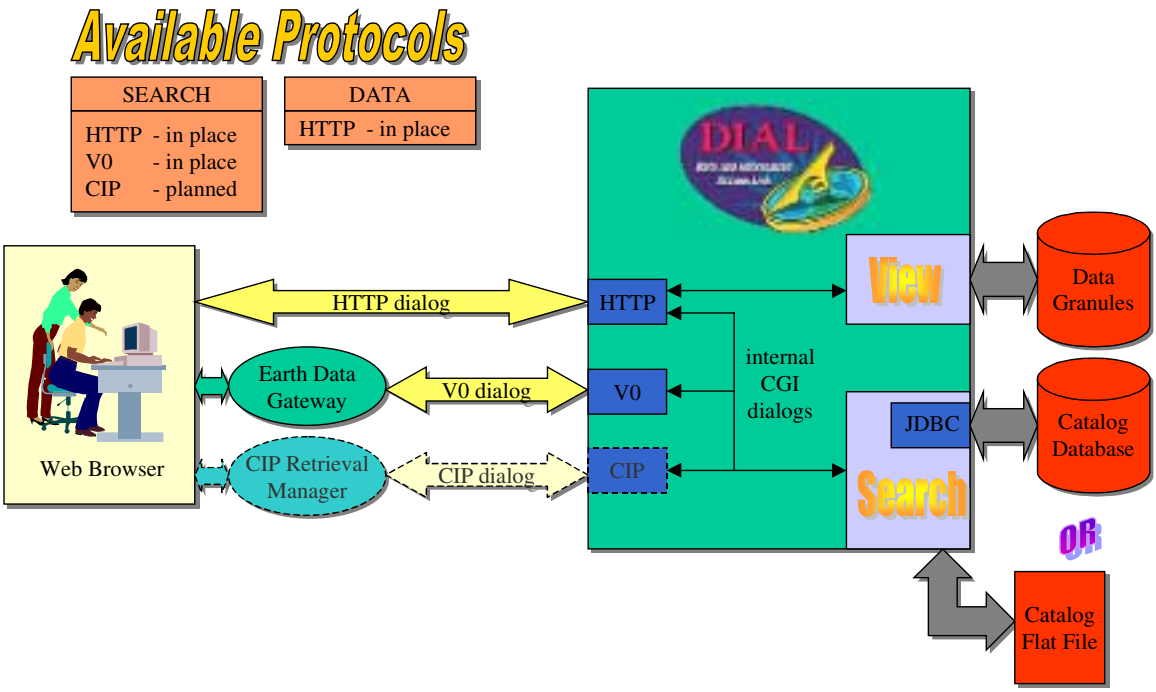

Figure 2. High Level Architecture

EDG is a distributed information management system that can search multiple distributed data catalogs through its query interface. EDG provides a consistent view of more than 900 data products held at each of the EOSDIS DAACs and several international data centers. Users without specific knowledge of the data can search science data holdings, retrieve detailed data inventories and high-level descriptions of data sets, view sample browse imagery, and place orders for data products. Both free-text simple search and advanced data attribute search interfaces are available. Data descriptions of NASA Global Change Master Directory (GCMD) are integrated into the Directory search function, improving the consistency and efficiency of providing useful extension of the information in the system.

The goal of EDG is to facilitate Earth science research through improved access to existing data and to serve as the search and order tool for data products produced by EOS. EDG uses Version 0 protocols and a data dictionary for providing catalog interoperability. Recently, EDG has been assigned a new role as the operational search and order tool for Landsat-7, Terra and other data sources. This service is available at http://eos.nasa.gov/imswelcome

3.2 Approaches for Setting up a Distributed DIAL System with EDG

EDG-DIAL application as distributed system: Project Amazon has 20 different scientists/data providers worldwide. The group studies the environment of Amazon River basin using satellite remote sensing, aircraft, and field campaign data. The group uses data from heterogeneous sources, processes them into high-level data products, and shares the products with colleagues and end users. Each scientist has a DIAL Version 2.4 server installed at his site and generates metadata describing his products. Most of the data products produced by scientists are in HDF or HDF-EOS formats. The field campaign data are in native or GIS formats. The scientists use translation software converting data in GeoTIFF and Arc/Info Exchange formats into HDF-EOS format for DIAL system to distribute.

To provide end users with a capability of one-stop search to all distributed data holdings in the project and a common interface for accessing the data, the data providers can choose distributed approach. To accomplish this, the Amazon project should do the following:

- Host EOS Data Gateway (EDG). Needs to be hosted on one machine only (either Sun or SGI)

- Establish a data dictionary or use the existing V0 data dictionary as the baseline.
- Define the data attribute types and possible values.
- Gather information (data attributes and their values) about the data holdings from the participating DIAL servers (Amazon data providers).
- Integrate the data attribute information of all the participating DIAL servers with EDG.
- Each scientist/data provider has to do the following in order to provide access to his data through the EDG:
    1. Create data catalog and metadata/attribute tables using DIAL tools
    2. Provide information about the data holdings to EDG
    3. Maintain EDG-DIAL server

Alternative approach
The Amazon project can also become part of the existing EDG without installing a gateway. In order to implement this approach the Amazon project must adopt the data dictionary of EDG and contact EDG staff and work with them.

4. Users

Current DIAL installations include NASA EOS Principal Investigators, NASA Earth Science Information Partners (ESIP), NOAA's Pacific Marine Environmental Laboratory, JPL, and international partners (Japan, UK, China, IGBP START South East Asian Regional Center). It is also in operational use by the Advanced Composition Explorer (ACE) satellite project at JPL.

5. CONCLUSIONS

Web based data systems can take advantage of standard protocols to create a distributed system that enhances scientific data exchange and interoperability. DIAL system with EDG and CIP provides a distributed system for users.

ACKNOWLEDGEMENTS

This research is funded by NASA Technology prototype under NASA contract NO: 50-007

REFERENCES

CEOS/WGISS/PTT/CIP-B, June 1998 Issue 2.4, Catalog Interoperability Protocol (CIP) Specification - Release B, Committee on Earth Observation Satellites

Specification, ANSI/NISO Z39.50-1995, Official Text for Z39.50-1995, July 1995Information Retrieval (Z39.50): Application Service Definition and Protocol

Suresh, R., Di. L., Kumar.L., Ilg.D.,McDonald, K., 1999. A Web-based Access to Distributed Heterogeneous Geospatial Data Archives using NASA's EOSDIS V0 Protocol. In Chen, J., Zhou, Q., Li Zhilin., Jiang Jie (eds) Dynamic and Multidimensional GIS, Vol 32, Part 4W12, Beijing, China, P 371.

Di, L, Suresh,R., Doan, K., Ilg. D., McDonald, K.,1998. DIAL-an Interoperable Web-based Scientific Data Distribution System. In M. Goodchild, M. Egenhofer, R. Fegeas, and C. Kottman (eds), Interoperating Geographic Information Systems, Section 4. System Experiences. Kluwer Academic Publishers, Norwell, MA.

Di, L, Suresh, R., Doan, K., Ilg, D., McDonald, K.,1997. A Web-based Scientific Data Server for Accessing and Distributing Earth Science Data. In Proceedings of Interop'97 PP 250-261

McDonald, K., Suresh, R., Di. L., Kumar. L 1998, Data and Information Access Link (DIAL), Fall meeting American Geophysical Union