4.9 USING A LDAP DIRECTORY SERVER FOR ENVIRONMENTAL DATA DISCOVERY

Donald W. Denbo*
Joint Institute for the Study of Ocean and Atmosphere, University of Washington, Seattle, WA

1. INTRODUCTION

The initial step in any distributed system for data browsing, visualization, and access is data discovery. In small systems the process of data discovery may be to pick data sets of interest from a scrolling list of those available. However, as the number and type of data sets available becomes large a completely manual selection process becomes unwieldy. We have developed a data discovery service that enables users to find data sets based on geographic region, observed quantities, or keyword. The data discovery service obtains its metadata directly from the data sever, thus insuring that the information in the service is complete and up-to-date.

We have based our data discovery service on the Netscape Directory Server a Lightweight Directory Access Protocol (LDAP) server. LDAP servers provide the robustness and scalability, via server replication and distributed servers, required of a data discovery service. Initial clients for this service are a Java application and Web server that access the Climate Data Portal (http://www.epic.noaa.gov/cdp).

2. DESIGN GOALS

The directory server and clients will need the following capabilities to create an effective system for environmental data discovery.

- Expandable, extensible, and scalable.
- Use standards when possible.
- Platform independent
- Automated population of directory directly from data servers.
- Tools to aid in the creation of queries.
- Work both with applications and web servers.

3. APPROACH

The development of the directory server leverages existing Open Source and commercial products that implement the Lightweight Directory Access Protocol (LDAP) and extensive experience gained during the NOAAServer project (Daddio et al., 1999; Soreide and Daddio, 1998).

Netscape Directory Server (4.1) was chosen as the LDAP server used in this project because it is a robust commercial product that implements LDAPv3 (the latest LDAP protocol) and NOAA has obtained a license to support of NOAA's Enterprise Messaging System.

Programs to load the directory schema into the server, load and update metadata from the Climate Data Portal systems, and client software to construct queries were written in Java using Netscape's Directory Software Development Kit for LDAPv3.

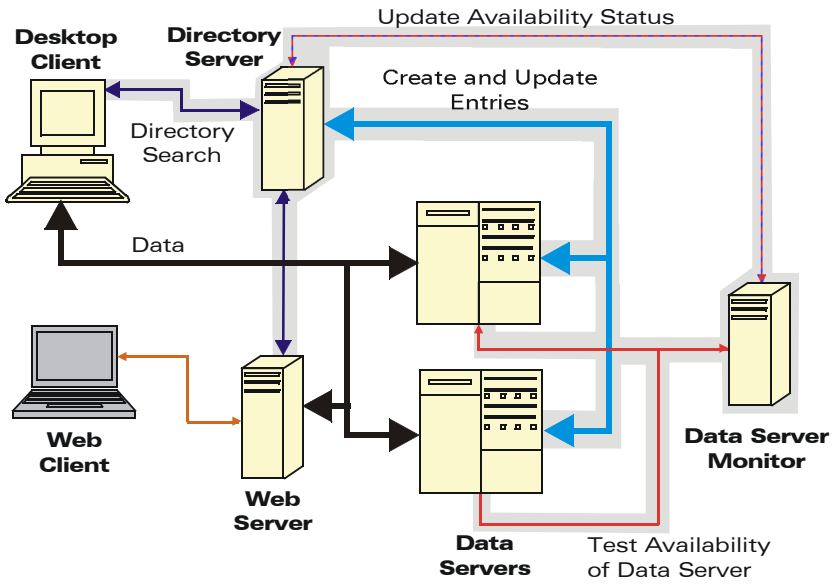

Figure 1. Architecture of Climate Data Portal and directory service.

4. ARCHITECTURE

The directory service three major pieces to its architecture. The directory server and the flow of information, the schema, the relationship between directory objects and parameters, and the construction of a LDAP query.

4.1 *System Components*

The complete Climate Data Portal system (Soreide et al., 2000), currently implemented and planned, is shown in Figure 1. Those components outlined in gray are part of the directory service. The directory service works by having each data server create and update entries in the directory server whenever a new data collection is added or any of the metadata has changed. Both the desktop client and web server, query the directory server to find the data server and collection that has data of interest. Data is then directly transferred from the data server to the desktop client or web server.

* *Corresponding author address:* Donald W. Denbo, NOAA/PMEL/OCRD, 7600 Sand Point Way NE, Seattle, WA 98115; e-mail: dwd@pmel.noaa.gov

4.2 *Directory Schema*

One significant advantages to using LDAP directory servers is their standardized approach to creating schemas. Basic objects and attributes have been standardized and are recognized by all LDAPv3 compatible servers. These objects and attributes can be easily expanded in a standardized way, increasing the opportunity for reusability and compatibility. When creating our specialized objects, *dataserver* and *datacollection*, we used existing attributes whenever possible for example, *owner* and *labelduri*. And created new attributes when necessary, for example, *spatialDomain* and *dataCount*. A graphical representation of the directory server schema is show in Figure 2.

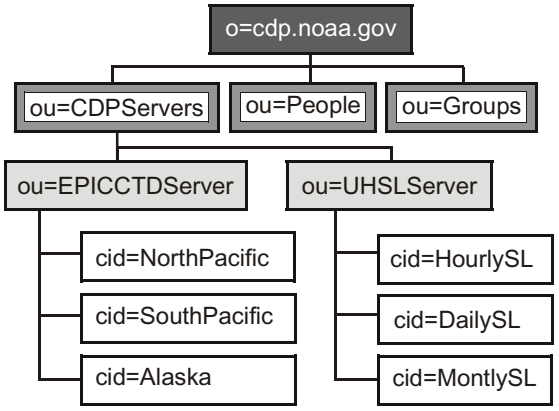

Figure 2. Simplified directory schema.

4.3 *Query Translation*

LDAP queries have a very “Reverse Polish” flavor. While this may make writing software to parse the queries straight forward it make for sensible keyword entry by a human difficult. We have developed a JavaCC parser, a lexical analyzer and parser that produces Java code from a production language, that takes a readable syntax and translates it to the format required by the directory server. For example, the input string

temperature AND salinity AND (pacific OR atlantic)

is translated to

&(|(keyword=atlantic)(keyword=pacific))(&(keyword=salinity)(keyword=temperature))

The graphical user interface uses a list of keywords downloaded from the server to aid the user in constructing a query (Figure 3).

5. FUTURE DIRECTIONS

The data server monitor, responsible for increasing the robustness of the Climate Data Portal, will be implemented in the near future. The monitor will be responsible for periodically testing the availability of all the data servers. When a data server becomes unavailable the monitor will notify the server administrator, marking the

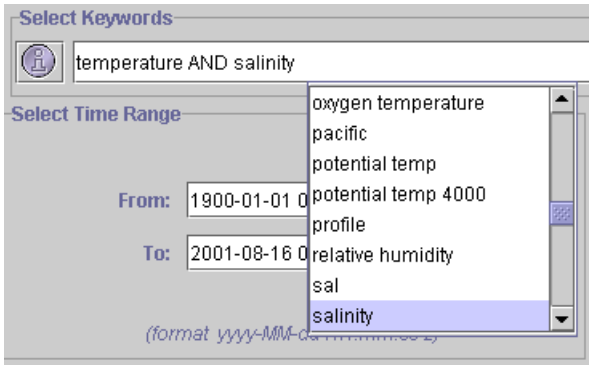

Figure 3. Keyword entry for directory search.

data server as unavailable in the directory, and attempting to restart the data server.

Continuing to improve the quality of the metadata, to ensure better query results, is a on-going goal. Adding support for multiple time ranges and polygonal spatial boundaries would substantially improve the directories usefulness by better describing a data collections temporal and spatial domains.

Acknowledgment. This publication was supported by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement #NA67RJ0155, Contribution #870. PMEL contribution 2406. The views expressed herein are those of the author(s) and do not necessarily reflect the views of NOAA or any of its subagencies. This work was funded by NOAA’s HPCC program.

6. REFERENCES

Daddio, E., S. Hankin, N. Soreide, D. Denbo, W. Zhu, S. Roberts, J. Sirott, and S. Rosenberg, 1999. NOAAServer: Unified access to distributed NOAA data. In *15th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Dallas, Texas, AMS, 10-15 January 1999, 430-433.

Soreide, N.N., C. Sun, B.J. Kilonsky, D.W. Denbo and W. Zhu, 2001. A Climate Data Portal. In *17th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Albuquerque, NM, 15-18 January 2001, 191-193.

Soreide, N.N., and E. Daddio, 1998. NOAAServer: Unified access to distributed NOAA environmental data. In *Marine Technology Society/Ocean Community Conference '98*, Baltimore, MD, 16-19 November 1998.