4.1 The Importance of Metadata Within NOAA and its Vital Link in Providing Access to Environmental Information

Howard J. Diamond[1]*
National Environmental Satellite, Data, and Information Service (NESDIS)
National Oceanic and Atmospheric Administration (NOAA)
Silver Spring, Maryland

Ted Habermann
NOAA/NESDIS/National Geophysical Data Center
Boulder, Colorado

1. INTRODUCTION

In short, metadata can be defined as data about data. Metadata is the key to documenting environmental datasets and in being able to provide access to them through the use of keyword searches. Within NOAA, metadata is an important activity that provides a vital link in providing access to environmental information. NOAA's relationship to other environmental metadata through such other entities as the Federal Geographic Data Committee (FGDC) and Global Change Master Directory.

2. DISTRIBUTED DATA ACCESS

The number of distributed databases in the federal government is tremendous. These databases have a diverse set of holdings that span an incredible range of every imaginable element of environmental information made available as a result of the work of practically every federal department. However, without high-quality and up-to-date metadata to describe these data, it is nearly impossible to find the data that are necessary to support research and public policy initiatives such as climate change. Metadata supports science; it extends data usefulness, and permits the replication of results. Metadata design can be a checklist for research planning, and are an investment in current and future science. Some of these thoughts have been extracted from an excellent presentation on metadata given by Dr. Raymond McCord from Oak Ridge National Laboratory; Dr. McCord's presentation can be found at http://globalchange.gov/workshop2001/proceedings/McCord/

NOAA has been involved in many metadata efforts over the last decade. Most of these efforts have been aimed at easing the data discovery process by creating a "one-stop shop" for some type of data. The "one-stop shop" concept emerged early in the history of the World Wide Web. Since that time we have gained considerable experience with a long list of "one-stop shops" or "portals", as they have come to be called. We have also seen the emergence and eventual dominance of web search engines as the data discovery tool of choice for end users. This dominance has created an environment in which one-stop shops, by themselves, can not motivate data providers to create high-quality metadata. We believe that high-quality metadata remains incredibly important, particularly in the emerging world of distributed data systems. We will discuss aspects of distributed data systems that require high quality metadata and propose that participation in these systems needs to motivate improved metadata in the future.

3. DATA DISCOVERY TO DATA EXPLORATION

The first step in using distributed data systems is data discovery. The goal of data discovery is to produce a group of links that are related to the subject that the searcher is interested in and that are related to one another. The most common mechanisms for data discovery are searching full text and traversing hierarchies. Many groups are developing spatial and temporal search paradigms that probably produce groups of data that are more clearly related to the searcher's needs but, to be successful, these searches require production and provision of meaningful metadata by data providers.

It is hoped that distributed data systems could enable data exploration rather than data discovery. Data exploration implies a more active and iterative process in which explorers create some knowledge (personalized information) along the way. The results of exploration might be thought of as a story or a travel journal documenting the path taken and the motivations for taking that path. These are actually similar to papers presented at science conferences. They recount the story of the knowledge building process. Metadata content is an important factor limiting us to data discovery rather than data exploration.

*Corresponding author address: Howard J. Diamond, NOAA/NESDIS, 1335 East-West Highway, Room 7214 Silver Spring, MD 20910; e-mail: howard.diamond@noaa.gov.

4. DATA DISCOVERY TO DATA USE

Many emerging systems, particularly those supported by web mapping standards, are changing the focus of data discovery from reading text descriptions to interactively examining and comparing datasets. These systems are actually using the data themselves in the data discovery process. They therefore require metadata that enables applications to use the data, even if only to present it on a map. Of course, once data use is enabled through high quality metadata, the boundaries between data discovery, data exploration, and applications become blurred. Making analysis easy by overcoming barriers to use makes much more sophisticated data discovery decisions possible. It also increases the quality of all data accessible in the system by making it easier for more users to examine and compare data from multiple sources.

5. DATA PRESENTATION TO DATA COMPARISON

Most websites are created by single data providers and focus on presenting that provider's data. In most cases these websites include significant amounts of structure and ancillary stuff that facilitates the transition of these data into information. The ancillary stuff adds information content and value, but at the same time it may limit the capability of the users to compare and contrast the data with that from other providers. Multiple views and comparisons are critical to the knowledge building process.

It is hoped that distributed data systems would include environments that enabled users to compare and contrast data from different sources and to understand variations in the quality of the data. In the beginning these comparisons could be simple maps, time-series plots, or descriptive statistics. As our data sharing capabilities improve, more sophisticated comparisons might become possible. Once again, meaningful metadata is required to make these comparisons possible.

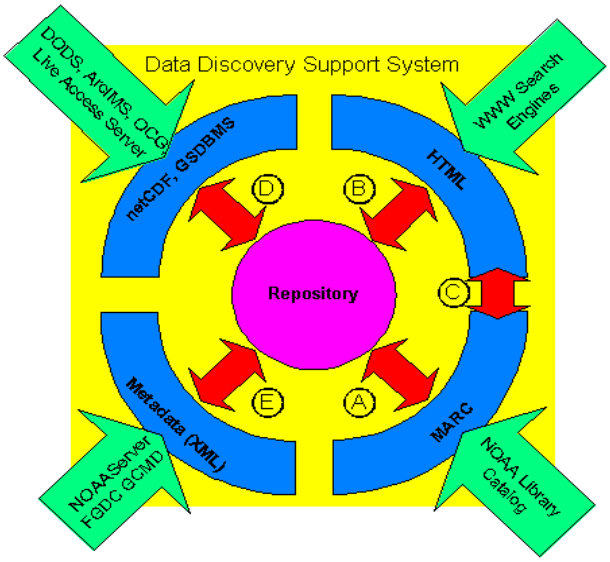
6. KNOWLEDGE TO WISDOM

As soon as distributed systems make it easy to cross traditional website boundaries or integrate data from different providers we have begun to approach transitioning knowledge to wisdom. Making this happen in distributed systems is going to require widespread adoption of various metadata and data standards as well as high-quality metadata. A decade ago, this would have been unthinkable. Consider the problems that NASA had in convincing earth scientists that a standard data format was required if the vision of EOSDIS was to be achieved. It never happened. Consider at the same time that over the last five years many hundreds of data providers at many levels have created FGDC clearinghouse nodes that can be searched from a single page. Also consider that during the last year over 600 data providers from all over the world have joined the Geography Network. There may be reason for hope.

7. ISSUES

There are a number of issues facing NOAA regarding metadata. First, many of our descriptions are severely out-of-date and not all parts of the organization are providing metadata descriptions (a cultural problem). We need to foster a better distributed process for handling metadata and give people the tools they need to accomplish this. Second, we have the issue of a transition from the current FGDC standard for metadata to the newer ISO 19115 standard that is ready for implementation sometime in the 2002 timeframe. While it is based primarily on the FGDC standard, there will be nuances we will have to take into account in order to make this as easy a transition as possible for end-users.

The FGDC is working on tools for converting to the new format, and will be developing a US national profile (using the current FGDC standard as a baseline) in order to make this as easy as possible. One thing we are looking at in NOAA is moving to a more comprehensive discovery system with metadata that uses a relational database concept in order to increase access to a diverse array of holdings (e.g., the NOAA Library, scientific net-CDF files, HTML web page tags, as well as the traditional FGDC metadata holdings) in order to increase the ability to discover data. The above schematic for a such a data discovery support system was developed by the co-author of this paper to better represent this concept in a graphical form.



8. CONCLUSION

Understanding our environment and the human impacts on it will require unprecedented access to data distributed across a wide variety of systems all over the world. Data providers must facilitate the transitioning of those data into information, knowledge, and ultimately wisdom. Requirements for supporting these transitions are different than those for adding data and information to the web. The latter can be done without useful or meaningful metadata; the former cannot.