INFORMATION TECHNOLOGY AND SYSTEMS CENTER
UNIVERSITY OF ALABAMA IN HUNTSVILLE
HUNTSVILLE, AL-35899

EARTH SCIENCE MARKUP LANGUAGE:

TRANSITIONING FROM DESIGN

TO

APPLICATION

ESDIS PROTOTYPE/STUDY REPORT

PROTOTYPE LEAD
KAREN MOE

TECHNICAL DEVELOPMENT
DR. SARA GRAVES
RAHUL RAMACHANDRAN

June 7, 2002

PROJECT OBJECTIVES

The primary objective of the proposed Earth Science Markup Language (ESML) research is to transition from design to application. The resulting schema and prototype software will foster community acceptance for the "define once, use anywhere" concept central to ESML. Supporting goals include:

- Refinement of the ESML schema and software libraries in cooperation with the user community.
- Application of the ESML schema and software libraries to a variety of Earth science data sets and analysis tools.
- Development of supporting prototype software for enhanced ease of use.
- Cooperation with standards bodies in order to assure ESML is aligned with related metadata standards as appropriate.
- Widespread publication of the ESML approach, schema, and software.

METHODOLOGY

In order to meet the research goals outlined above, ITSC will leverage the numerous tools available for authoring, parsing, and validating XML in a variety of programming languages (C/C++, Java, Perl, etc.). In addition, the team will participate in appropriate conferences in order to promote the utility of the tools to potential users and to gather input from the broader science community. Proposed research tasks include:

Schema Evolution. The ESML schema will continue to evolve throughout the period of research. This continued refinement of the ESML design will be required, based on the analysis of additional data formats, the need for dataset-independent software, and recommendations from the Earth science community, including data producers, software developers, and general users. The work on this prototype was divided into two distinct phases. In phase I, the team surveyed and researched various Earth science dataset and metadata conventions. The actual design and implementation occurred in phase II.

Software libraries. An early indicator of the value of ESML is the ability to develop a generic data reader based on it. Prototyping of a basic C++ software library providing this generic read capability for ASCII and binary data files is already underway. In the research proposed here, the ESML library will be updated to provide links to the software tools available for supported data formats such as GRIB, McIDAS, HDF, HDF-EOS, and netCDF. A Java version of the ESML library is also planned to provide for greater software portability to platforms such as Windows, Linux, Solaris, and Irix.

Additional user tools. The primary rationale for ESML is to ease data handling by Earth scientists, thereby freeing their time for research. User tools will be essential in order to make ESML an easy to use and attractive option. Most important will be a metadata generation tool (ESML Editor), which will allow the data producer to fully describe his data for independent access. Data consumers will also value the metadata generator, as it will allow them to quickly provide the documentation needed to integrate data files, even those in non-standard formats.

Collaborations. Collaboration with the Earth science community has been central to the initial design of ESML. As software libraries and related tools are prototyped, additional work with Earth science applications developers will be required. ESML can be particularly useful for data access interoperability, so collaborations in this arena will be sought during the proposed research.

Standards. Throughout the research period, the team will track advances in XML and related technologies, as well as the evolution of metadata standards in general. In addition to monitoring posted developments on XML-related web pages, research partners will attend selected conferences such as the Open Forum on Metadata Registries, the IEEE Metadata Conference, and others.

Outreach. The activities outlined above will create a prototype ESML toolkit, ready for promotion to the Earth science community. To this end, ITSC will continue to publish results on the ESML web site, submit papers for publication, and make presentations at such major Earth science conferences as AMS, American Geophysical Union (AGU), and Investigators' Working Group (IWG). In addition, ITSC plans to institute one or more training seminars at interested universities, NASA centers, and/or other sites.

PROJECT MILESTONES

- **Extend ESML Library**

The ESML library provides applications with a programmatic way to read and interpret data files based on the ESML descriptions. A logical data model has been used by the ESML library to make it more intuitive. Access to data using the ESML library can be achieved by creating an ESML_OO object. This ESML_OO contains an array of ESML_Collections. An example of an ESML_Collection is a swath orbit containing data and navigation fields. It holds collections of similar data/navigation fields.

The ESML_Data object holds all the data fields and methods to access metadata information about the fields. It also provides methods to access the "data" contained within the object, field by field. The ESML_Navigation object holds the navigation fields (latitude, longitude, time etc) and methods to access both metadata and navigation data contained in it. The ESML_FieldSize object provides the size information (rank and dimensions) for a specified field used by both ESML_Data and ESML_Navigation. The ESML library update includes incorporation of HDF-EOS library and GRIB reader code and addition of new features such as handling wild cards, symbols and equations.

- **Extend ESML Schema**

The ESML schema contains rules for writing valid ESML files. These rules allow users to describe three important aspects of the data file: Content, Structural and Semantic Metadata

Content metadata describes the dataset in human-readable terms. Although this metadata may be parsed by the ESML reader, it is not necessarily "understood" by the computer. The content metadata documents the origin and pedigree of the data, information that is not necessary for automated data manipulation.

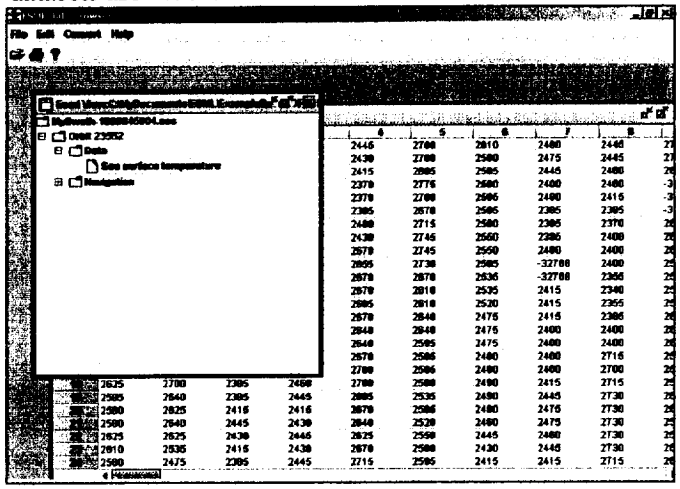
The structural (or syntactic) metadata can be used to describe the details of the structure of the data stored in a data file in term of bits, bytes and records. The initial formats described included free formats, such as Binary and ASCII text, and structured formats, such as HDF-EOS. The ESML schema has been extended to add other data format elements such as GRIB and McIDAS.

Semantic metadata is used to give meaning to the elements described by the syntactic metadata. These metadata elements allow the parser to "understand" the actual meaning of the data in terms other than bits and bytes. The ESML schema also provides means to allow users to specify an equation for data conversion.

- **Prototype an ESML "enabled" Application**

ESML Data Browser

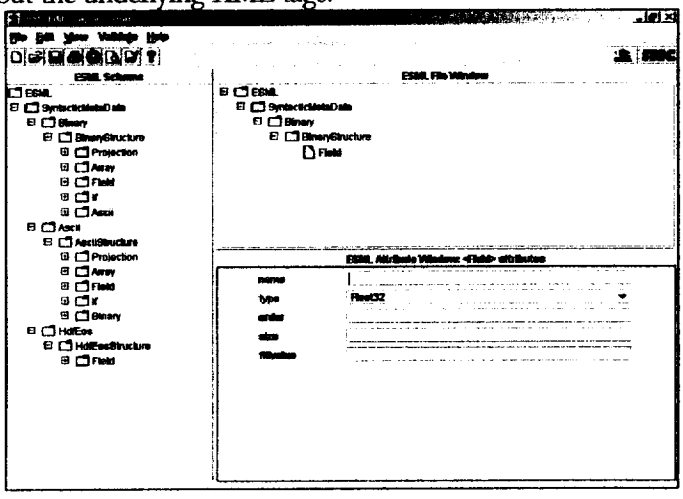
The ESML Data Browser is an example of an application that utilizes the ESML library. This application has been designed for scientists to give them the ability to view any data file that has an ESML description. This tool can also be used by scientists to test whether the ESML descriptions they have created for datasets are indeed correct.



- **Application that allows users to write ESML description file**

ESML Editor

For scientists not familiar with XML, writing ESML tags can potentially be as complicated as writing a new data decoder. To alleviate this problem, an intelligent, easy to use Editor has been created. This ESML Editor provides scientists a user-friendly interface to create an ESML file without worrying about the underlying XML tags.



- **External ESML "enabled" Application**

In order to demonstrate the viability of ESML as a solution that allows data-application interoperability, several representative external applications were modified to use the ESML library. These applications are:

1. Web Map and Coverage Servers for Passive Microwave data sets
2. Algorithm Development and Mining (ADaM) system
3. Data servers for the Space Time Toolkit Visualization Package

- **Workshop/Training Session**

As a part of the ESML project objectives to promote this approach to the Earth Science community, several detailed presentations/training sessions on ESML and how to use it were given. These sessions listed below were given to:

1. Scientists and students at the National Space Science and Technology Center (NSSTC), Huntsville, AL in November, 2001
2. Workshop attendees at the 9th meeting of the Federation of Earth Science Information Partners, University of Maryland May, 2002.

OTHER ACCOMPLISHMENTS

- **Collaborations Established with other Projects**

1. THREDDS:
Unidata's Thematic Real-time Environmental Distributed Data Services (THREDDS) provides students, educators and researchers with coherent access to a large collection of real-time and archived datasets from a variety of environmental data sources at a number of distributed server sites. The datasets will be conveniently accessible from a collection of THREDDS-enabled data analysis and display tools.

2. NOMADS:
The NOAA Operational Model Archive and Distribution System (NOMADS) objective is to preserve and provide retrospective access to GCM's and reference quality long-term observational and high volume three dimensional data, NCEP NWP models and re-start and re-analysis information, as well as other NCDC available products and data.

3. UNITE:
This project focuses on the process of enabling independently developed ESIP applications and services to effectively utilize distributed heterogeneous data products through a proposed Universal Interchange Technology for Earth science data and services (UNITE).

4. ESML Pilot Project
This project intends to utilize ESML to develop two powerful applications that will demonstrate the usefulness of ESML for the science community. One of the applications will focus on satellite data sets and the other on numerical models.

- **ESML users technical support**

Due to the large volume of ESML users it become necessary to provide ESML technical support for not only writing ESML description files but also supporting application developer in utilizing the ESML library. This is an area that needs to be enhanced for future efforts.

- **Presentations and Publications**

Presentations:
In addition to the Training/Workshop Sessions, ESML was presented at different forums. These were:

1. Earth Science Technology Conference, August 2001
2. Earth Science Technology Conference, June 2002
3. NOMADS Meeting, NCDC, Asheville, NC, October 2001
4. NVODS/DODS Workshop, Boulder, Colorado, January 2002
5. OGC's 39th Technical Committee and OGC Planning Committee Meeting - WWW MAPPING SIG, February 2002
6. NASA Science Data Processing Workshop, February, 2002
7. Interchange Technology for applications to facilitate generic access to heterogeneous data formats", IGARSS, Toronto, Canada, June 24-28, 2002

Papers:

1. "ML: What's Next for Earth Science", Jeanne Behnke, Rahul Ramachandran, Helen Conover and Sara Graves, NASA Science Data Processing Workshop, February, 2002
2. Ramachandran R., M. Alshayeb, B. Beaumont, H. Conover, S. Graves, X. Li, S. Movva, A. McDowell and M. Smith, "Earth Science Markup Language: A Solution for Generic Access to Heterogeneous Data Sets," Earth Science Technology Conference, Maryland, August 28-30, 2001
3. Ramachandran. R., H. Conover, S. Graves, and K. Keiser," Interchange Technology for applications to facilitate generic access to heterogeneous data formats", IGARSS, Toronto, Canada, June 24-28, 2002

- **Collaborations expected in Future Projects**

1. NSF-ITR with Oklahoma University, UI-Urbana Champaign, UNIDATA and others
2. NSF-ITR with NCAR, UCLA and others

- **Other Important Metrics**

ESML Users:
Around 60 users have signed on to ESML mailing list and are interacting with the development team

TRL advancement:
The Software developed for this project was advanced from TRL 3 to TRL 5

All the latest ESML products (ESML Schema, Library, Editor and Data Browser) can be downloaded from the URL:

esml.itsc.uah.edu