

Ethan R. Davis\*  
John Caron  
UCAR/Unidata, Boulder, CO

## 1. INTRODUCTION

Thematic Real-time Environmental Distributed Data Services (THREDDS) is a project, spearheaded by Unidata and funded by the National Science Foundation, to simplify the access and use of Earth system data by faculty, students, and researchers.

The THREDDS project is developing an end-to-end data/metadata system that will allow users to access a web of scientific data through a variety of software applications. The THREDDS system will concentrate on "use" metadata and be developed hand-in-hand with client applications, initially focusing on data visualization and analysis applications. The THREDDS system will also facilitate the building of the scientific data web, providing data/metadata publishers with the means to represent the relationships between datasets and add meaning to datasets. Through this web of meaning and relationships, data users will be able to more easily use and understand data. The project will result in a working prototype that includes data publishers, 3rd-party metadata publishers, and data users.

By building a community of data publishers, data users, software engineers, and metadata experts, the THREDDS project will get a broad range of viewpoints on this complex problem. For more details on the THREDDS community and how to become part of the THREDDS community, see [5].

### 2.0 THREDDS Technical Underpinnings

Many efforts are under way to standardize metadata requirements for scientific data, especially georeferenced data. Most of these efforts concentrate on descriptive metadata that can be used for search and discovery of the datasets. The THREDDS project will concentrate on metadata that will allow users to more fully understand and use the data. For instance, metadata that describes dependent and independent variables or that describes how a grid is geolocated on the earth.

We will be working with other groups on linking this use metadata with metadata standards. Hopefully, the THREDDS metadata framework will lend itself to linking to existing metadata in these other standards.

---

*\*Corresponding author address:*

Ethan Davis, UCAR/Unidata, P.O. Box 3000, Boulder, CO 80307, email: edavis@ucar.edu.

The development of the THREDDS framework will include:

- 1) The definition of a metadata framework for communicating use metadata between data providers, 3rd party metadata publishers, and data consumers. This metadata framework is referred to as PICats (Publishable Inventories and Catalogs).
- 2) Tools for building a data web with PICats
- 3) Tools for communicating existing metadata as PICats
- 4) Tools and APIs for accessing the data/metadata web from data analysis and visualization applications and from data search sites.
- 5) Work with others on search and discovery capabilities (e.g., DLESE)

### 2.1 Metadata Framework

The THREDDS metadata framework allows data and metadata to be distributed over the network, with the intention that those who are interested in adding meaning to existing datasets can do so. The responsibility for adding metadata will be on those with the time and interest in having access to more meaningful data (not necessarily the data publisher).

The current implementation uses XML documents over HTTP. We will be looking at other possibilities, including RDF and Topic Maps. Our current plan for the metadata framework is divided into three parts: the Inventory Catalog which organizes datasets into collections; the Dataset Description which provides georeferencing and additional information; and the Standard Quantities which provide a standard vocabulary for the fields available in a dataset.

The Inventory Catalog allows datasets to be organized into collections. Each dataset or collection can link to textual documentation or additional metadata in the form of Dataset Descriptions. An HTML version of the catalog can be generated using our example XSLT document. For a detailed look at the Inventory Catalog definition, please see our THREDDS Inventory Catalog specification [9].

The Dataset Descriptions are referenced from the Inventory Catalog by a particular dataset or collection. Due to the THREDDS concentration on data access applications, we will initially concentrate on "use" metadata though they will also contain catalog level metadata.

Standard Quantities allow users to define a controlled vocabulary of physical quantities and describe their

semantics as clearly as possible. The flexible structure of the Standard Quantities allows different communities to define their own vocabularies in a community-based development. The goal is then to allow data/metadata publishers to map dataset variable names to a standard name with a well-defined meaning.

## 2.2 Creating the Data/Metadata Web

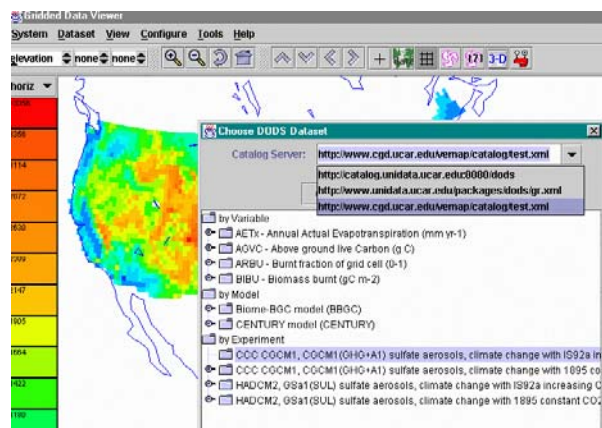
The THREDDS project is working on tools to help data publishers easily create PICats. Currently we have two pieces of software that produce PICats. First is the PICat AutoGenerator (picatAG) [10] that crawls local data directories to determine the datasets available. Second, the DODS Aggregation Server [1,3] produces THREDDS Inventory Catalogs of the datasets that it serves. The Aggregation Server configuration file is based on the THREDDS Inventory Catalog.

## 2.3 Accessing Existing Metadata

Though much of the existing metadata is more catalog centered than use metadata, it will be essential that existing metadata can be shared through the THREDDS metadata framework. Creating tools that sit on top of existing metadata stores and respond to requests for metadata in a THREDDS appropriate way will be important. This is similar to how DODS handles existing data that is in various data formats.

## 2.4 Tools and APIs for accessing data/metadata web

THREDDS uses the Unidata MetApps packages [6], e.g., the Gridded Data Viewer (GDV), for prototyping the client side of the THREDDS system. Developing THREDDS capabilities in a client-side application along with the metadata and server developments allows us to keep the data user viewpoint in focus. The data access client work will help us develop an API for THREDDS capabilities. The PICat menu/selector [11] (seen below in GDV) is a Java component that can be added to other applications to perform the low level work of connecting with and displaying PICats.



## 2.5 Search and Discovery

The first step in search and discovery for THREDDS will be to create a Catalog Server that has access to the PICats of a set of federated data servers. The Catalog Server will allow for quick searches of the datasets available from any of the servers rather than having to search each one separately. Catalog Servers will be able to respond to user queries with targeted PICats.

THREDDS will work with the Digital Library for Earth System Education (DLESE) [2] on the simple query language for these servers and also on extending the DLESE discovery system to include metadata from the THREDDS data/metadata web.

As we start on this path, THREDDS will work with other groups to generalize our approach to search and discovery.

## 3.0 More information

For more information on THREDDS, go to the THREDDS home page [7] and to the THREDDS technical overview and status page [8]. To become part of the THREDDS community, please go to the THREDDS home page and join our THREDDS mailing list.

## 4.0 References

- [1] Caron, John, "DODS Aggregation and THREDDS Catalog Services", AMS 2002, IIPS paper number 4.5
- [2] DLESE  
<http://www.dlese.org>
- [3] DODS Aggregation Server  
<http://www.unidata.ucar.edu/projects/THREDDS/tech/DODSaggServer.html>
- [4] DODS  
<http://www.unidata.ucar.edu/packages/dods/>
- [5] Domenico, Ben, "Thematic Real-time Environmental Distributed Data Services (THREDDS)", AMS 2002, IIPS, paper number 9.3
- [6] Unidata MetApps project  
<http://www.unidata.ucar.edu/projects/metapps/>
- [7] THREDDS home page  
<http://www.unidata.ucar.edu/projects/THREDDS/>
- [8] THREDDS technical overview and status  
<http://www.unidata.ucar.edu/projects/THREDDS/tech/>
- [9] THREDDS Inventory Catalog specification  
<http://www.unidata.ucar.edu/projects/THREDDS/tech/InvCatalog.html>
- [10] THREDDS picatAG  
<http://www.unidata.ucar.edu/projects/THREDDS/tech/picatAG>
- [11] THREDDS PICat menu selector  
<http://www.unidata.ucar.edu/projects/THREDDS/tech/javadoc/>