

# DODS Aggregation and THREDDS Catalog Services

John Caron\*

Unidata Program Center, UCAR

## 1. Introduction

The DODS Aggregation/Catalog Server (AS) provides three services: 1) it is a DODS server for netCDF files; 2) it allows DODS datasets to be logically "aggregated"; and 3) it provides a list of available datasets in a standard format, called a "Dataset Inventory Catalog".

The AS is a 100% Java servlet extending the Java/DODS Server library (1). It uses multithreading and resource pools to achieve good performance, and can be added to any web server that supports Java servlets.

This development is part of the ongoing technical collaboration between Unidata and DODS. The prototype Catalog service is part of the Unidata THREDDS initiative (2).

## 2. DODS Aggregation

An aggregation dataset looks like a normal DODS dataset to the client, but is actually a logical view of a group of other datasets. As currently implemented, the underlying datasets may be either netCDF files (local to the AS) or datasets obtained from other DODS servers, called DODS datasets. Other file types may be added in the future. There are three kinds of aggregation:

(Type 1) Datasets in which a new coordinate variable is created, where one dataset corresponds to each coordinate value. For example, a series of satellite data, with each file containing the data for a single time.

(Type 2) Datasets with different variables in separate files. For example, COADS at <http://www.cdc.noaa.gov/cgi-bin/nph-nc/Datasets/coads/1degree/equatorial/enh/>

(Type 3) Datasets that are to be concatenated on an existing coordinate dimension. For example, the NCEP reanalysis at <http://www.cdc.noaa.gov/cgi-bin/nph-nc/Datasets/ncep.reanalysis.dailyavgs/pressure/> has daily averages for each day in the year, with one file for each year from 1948-2001. The AS allows this to be viewed as a single dataset, with the time coordinate spanning all 53 years.

The AS can serve netCDF files, aggregation datasets, or a combination of both. The client will not see a difference between the two, except perhaps in performance.

## 3. Catalog Services

The Dataset Inventory Catalog is the "public view" of the available datasets. It specifies hierarchical collections of datasets, and allows in-line comments and references to external URLs. The Catalog itself is an XML document, which makes it well defined and thus machine-readable, and can be easily parsed by freely available XML libraries.

These qualities make Catalogs suitable for building more complex services on top of. Some of the services we envision coming out of the THREDDS initiative include: combining Catalogs from different servers; simple searching by time, space or quantity; and "third-party" construction of dataset collections for specialized purposes, such as for educational materials.

Another important component of Catalogs are "Dataset Descriptors", which specify the time and space dimensions of a dataset, along with a mapping of dataset variable names to a "Standard Quantities" database. These are optional for an AS to provide, but necessary to provide some of the services mentioned above. We envision that Dataset Descriptors might be added by someone other than the data providers, and seamlessly integrated into client applications.

## 4. More Information

A prototype AS is running at <http://thredds.unidata.ucar.edu:8080/dodsC/>. The Catalog can be obtained from an AS by appending the path "catalog.xml".

Technical status of the AS is at <http://www.unidata.ucar.edu/projects/THREDDS/tech/DODSAggServer.html>

Technical status of THREDDS is at <http://www.unidata.ucar.edu/projects/THREDDS/tech/index.html>.

Thanks to Ethan Davis, James Gallagher, Steve Hankin, Nathan Potter, and Roland Schweitzer for important contributions to this project.

## References

(1)<http://www.unidata.ucar.edu/packages/dods/home/swJava1.0/>

(2)<http://www.unidata.ucar.edu/projects/THREDDS/Overview.htm>

---

\*Correspondence: John Caron, Unidata/UCAR,  
PO Box 3000, Boulder, CO, 80307-3000.  
email: caron@ucar.edu.