

So, what is THREDDS anyway?

THREDDS (Thematic Real-time Environmental Distributed Data Services) is **middleware** to bridge the gap between data providers and data users. The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data.

THREDDS' initial focus was to allow data **users** to find datasets that are pertinent to their specific education and research needs, access the data, and use them without necessarily downloading the entire file to their local system.

To achieve this, we needed a way for data **providers** to publish lists of what data are available and to describe their data to enable discovery and use.

Catalogs are the heart of the THREDDS concept. They are XML documents that describe on-line datasets. Catalogs can contain arbitrary metadata, and we have also defined a standard set of metadata to bridge to discovery centers like GCMD, DLESE and NSDL.

Unidata's Common Data Model

Unidata's Common Data Model (CDM) is an ambitious project to unify scientific **data access**. It merges the OPeNDAP, netCDF, and HDF5 data models to create a common API for scientific data. As implemented by the NetCDF Java library, it can read netCDF, OPeNDAP, HDF5, HDF4, GRIB 1 & 2, BUFR, NEXRAD 2 & 3, GEMPAK, MCIDAS, GINI, among others. A pluggable framework allows other developers to add readers for their own specialized formats. The CDM also provides standard APIs for **georeferencing coordinate systems**, and specialized queries for **scientific feature types** like *Grid*, *Point*, and *Radial* datasets.



Fact Sheet

The THREDDS Data Server

The current focus of THREDDS development is the THREDDS Data Server (TDS), a web server that provides catalog, metadata, and data access services for scientific datasets. TDS configuration files control which local and/or remote dataset collections are scanned for the datasets to be served.

The TDS uses the Common Data Model (see sidebar) to read datasets in various formats, and serves them through *OPeNDAP*, *OGC Web Coverage Service (WCS)*, *OGC Web Map Service (WMS)*, *NetCDF subset service*, and bulk *HTTP file transfer* services. The first four allow the user to obtain **subsets** of the data, which is crucial for large datasets.

The TDS has the ability to aggregate many files into **virtual datasets**, which insulates users from the details of file storage and naming, and greatly simplifies user access to large collections of files.

Much of the real-time data available over Unidata's **Internet Data Distribution (IDD)** system is available through a THREDDS Data Server hosted at Unidata on <http://motherlode.ucar.edu:8080/thredds/>. We recommend Unidata's IDV as one way to browse and visualize these datasets.

The TDS is open source, 100% Java, and runs inside the open source Tomcat Servlet container.

THREDDS is a highly collaborative project involving universities, government entities, and private industry partners.

