



Fact Sheet

So, what is THREDDDS anyway?

THREDDDS (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.

THREDDDS' 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 THREDDDS 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.

Dynamic Catalog Generation

The THREDDDS **Catalog Generator** produces THREDDDS catalogs by scanning or crawling one or more local or remote dataset collections. Catalogs can be generated periodically or on demand, using configuration files that control what directories get scanned, and how the catalogs are created.

The THREDDDS Data Server

The current focus of THREDDDS development is the THREDDDS Data Server (TDS), which actually serves the **contents** of the datasets, in addition to providing catalogs and metadata for them.

The TDS uses the Common Data Model (see below) to read datasets in various formats, and serves them through *OPeNDAP*, *OGC Web Coverage Service*, *NetCDF subset*, and bulk *HTTP file transfer* services. The first three 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 THREDDDS Data Server hosted at Unidata on <http://motherlode.ucar.edu:8080/thredds/>. You are welcome to browse and access these meteorological datasets.

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

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 many types of data. As currently implemented by the NetCDF Java library, it can read (besides OPeNDAP, netCDF, and HDF5) GRIB 1 and 2, BUFR, NEXRAD, and 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 data types* like *Grid*, *Point*, and *Radial* datasets.



unidata

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



WWW.UNIDATA.UCAR.EDU

