Climate Data for Non-experts: Standards-based Interoperability

Ben Domenico

Unidata Program Center
University Corporation for Atmospheric Research
April 2010
Working Together on
A Mosaic for Atmospheric Data

This presentation describes and draws on the work* of many collaborating individuals and institutions.

* Unidata’s contribution supported by the U.S. National Science Foundation

Ostia Antica circa 7 BC
Acronym Glossary

- GALEON (Geo-interface for Air, Land, Environment, Oceans NetCDF)
- FES (Fluid Earth Systems, aka “metoceans” mainly the data systems of the atmospheric and ocean sciences)
- [http://www.unidata.ucar.edu/content/publications/acronyms/glossary.html](http://www.unidata.ucar.edu/content/publications/acronyms/glossary.html)
Outline

• Challenge: make climate data understandable beyond the climate science community
• Weather data lessons learned: different ways of thinking about data
• Interoperable data systems enable use of familiar and appropriate tools
• Model output: multi-dimensional grids
• Verifying models with observations
• Existing tools in different communities
• Transformations: data are often not available in the right form
• References
Other Communities

- Experts in other disciplines
- Impacts on society and infrastructure
- Decision makers
- Educators at all levels
- The general public
- The GIS (Geographic Information System) community in general
Organizations and Projects Addressing Interoperability for Climate Data

- OGC (Open Geospatial Consortium) -- establishing netCDF data model, encoding, Climate and Forecast Conventions as international standards
- GEOSS (Global Earth Observing System of Systems): Climate and Biodiversity focus
- MetOceans Domain Working Group within the OGC defining climate data use cases
- CCLI (Climate Challenge Integration Project) associated with the Free and Open Software for Geosciences (FOSS4G)
- Looking to NCAR, UCAR, Unidata for contributions and sometimes leadership
The Challenge of Disparate Data Models: Different Ways of Thinking about Data

- To the GIS (includes solid earth and societal impacts) community, the world is:
  - A collection of static features (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).
  - The features are discrete objects with attributes which can be stored and manipulated conveniently in a relational database.

- To the fluids (atmosphere and oceans) communities, the world is:
  - A set of parameters (e.g., pressure, temperature, wind speed) which vary as continuous functions in 3-dimensional space and time.
  - The behavior of the parameters in space and time is governed by a set of equations.
  - Data are simply discrete points in the mathematical function space.

- Each community is making progress in understanding and adapting to needs and strengths of the other. Progress areas will be highlighted.
Traditional GIS view

Attributes in DBMS tables

Features as points, lines, polygons
Atmospheric Data Visualization
Apply GIS Tools To Atmospheric Science Data
Taking Advantage of Web Services for Data System Interoperability

GIS Client Applications

OGC or proprietary GIS protocols

OpenGeospatial Protocols: WMS, WFS, WCS

GIS Servers

Hydrologic, demographic, infrastructure, societal impacts, ... datasets

MetOceans Client Applications

OGC or THREDDS, OPeNDAP, ADDE, FTP... protocols

MetOceans Servers

Model output, ensembles and observational datasets
Special Characteristics of Atmospheric Data

- full 3D in space
- multiple times (forecast run time and valid time)
- time relative to the present (e.g., latest)
- non-regularly-spaced grids
- observational datasets that are not gridded at all
- non-spatial elevation (e.g. pressure) coordinate
- Data interpolation depends on physics (and data) whereas GIS world is concerned mainly with geometry
- agreement on Coordinate Reference System specifications
Output of Forecast Models:
Time-varying, Multi-dimensional

WCS is ideal for this scientific data type.
Building on Existing Systems that Work

- Solid set of established data systems serving the MetOceans community
- Many of these data systems and tools can serve climate sciences as well
- Climate data systems have many of the same interoperability issues
Taking Advantage of Web Services for Data System Interoperability

GIS Client Applications

OGC or proprietary GIS protocols

OpenGIS Protocols: WMS, WFS, WCS

GIS Servers

Hydrologic, demographic, infrastructure, societal impacts, ... datasets

FES Client Applications

OGC or THREDDS, OPeNDAP, ADDE, FTP... protocols

FES Servers

Satellite, radar, forecast model output, ... datasets
Working Systems in MetOceans Community

- Unidata IDD/LDM “pushes” many GB/hr of real-time data to hundreds of sites 24x7
- netCDF provides common interface to many file formats (HDF5, GRIB, and many others via TDS)
- OPeNDAP delivers many dataset types via client/server pull interface
- THREDDS provides catalog data framework for its own community
- THREDDS Data Server (TDS) integrates service interfaces and on-the-fly conversion to netCDF objects
- CF conventions:
  - available for gridded data, coordinate system specs are more explicit now
  - proposed for point, station, trajectory -- including means for specifying locations for non-gridded data collections.
NetCDF/OPeNDAP data server

WCS Client

getCapabilities

describeCoverage

getCoverage

NcMLGML

geoTIFF

WCS capabilities

THREDDS interface

THREDDS catalogs

NetCDF dataset

NetCDF/OPeNDAP data server

THREDDS enhanced catalog generation tools

THREDDS enhanced catalog generation tools

THREDDS catalogs

NcML-G metadata

WCS coverage

GML generator

geoTIFF generator

NetCDF objects

NetCDF dataset

THREDDS interface

WCS description

THREDDS interface

THREDDS enhanced catalog generation tools

getCapabilities

describeCoverage

getCoverage
Datasets Not in Convenient Form:
NCAR GIS Climate Science Examples

- Have: netCDF files with daily min / max temperature
- Need to find: hottest 3 consecutive daily min temperatures over a period of years
- With GIS tools, this necessitated exporting each time as a separate GIS “layer.”
- This is not convenient as we are working with 365 layers for each year
More Climate Examples:

• Using the daily temperature data we want to find all occurrences of when a threshold is exceeded
  – Currently working with scripts to extract each time as a layer is very cumbersome
• We have 20 years of daily data and need to find the 97\textsuperscript{th} percentile temperature
  – In order to get a percentile you first need to get the min and max values then calculate the percentile
Need Processing Services

- Interoperable data services are a good beginning
- Re-usable, web processing services are a next step
- Examples of automated processing needs, e.g.,
  - Gridding/assimilation
  - Forecast models themselves
  - Transformations between pressure and height
  - Basic climate examples from previous slide
Summary

• MetOceans datasets fully 3D in space with multiple time dimensions and special coordinate systems
• Many data types: grids, station/point, profile, trajectory, radar, swath, irregular grids
• Internal systems and community standards: IDD, THREDDS, OPeNDAP, NetCDF, CF Conventions
• Standards-based web services for serving and processing data for use in other communities.
• NetCDF and OGC standard interfaces bridge community data systems
References

- GALEON document with more details
- GALEON Wiki
- Unidata NetCDF
- CF Conventions
- OGC WCS Specification
- OGC Observations and Measurements:
- ISO 19123 Coverage Specification
- GML
  - CSML
  - NcML-GML
- ISO 19111: Geographic Information: Spatial Referencing by Coordinates
- CS-W
- Interoperability Day Presentations
  - Andrew Woolf
  - Stefano Nativi
  - Wenli Yang
  - Stefan Falke
  - ESIN Paper
- Proposed CF conventions for non-gridded datasets