Unidata Outreach Activities

focusing on

Evolving Standards for Delivering Atmospheric Data into the GIS Realm (mainly OGC GALEON)

Ben Domenico, Unidata
Stefano Nativi, CNR/IMAA
Jeff Weber, Unidata
With Input from the GALEON Team
September 2009

Outreach Activities (Winding Down)

- KNMI ADAGUC
  - Royal Dutch Meteorological Institute
  - Atmospheric Data Access for the GIS User Community
- NSF NSDL: THREDDS 2G
  - Finished up last CU CIRES evaluation
  - THREDDS now part of Unidata core
- NASA: Gateway to Oceans Land Air Collaboration with George Mason
  - OGC Standard CS-W search of TDS
  - Final report in press
- NSF GEO: AccessData (formerly DLESE Data Services)
  - Final workshop was last June
  - Possible follow on evaluation activity

Ongoing Outreach Activities
(not the focus of this presentation)

- NCAR GIS Program (official program of NCAR as of a couple months ago)
- Marine Metadata Interoperability Project
- IOOS DMAC Steering Team
- CUAHSI Standing Committee
- Oceans Interoperability Experiment sponsor
- UCAR wide representative to OGC Technical Committee
- AGU (and EGU) ESSI Focus Group
- ESIN Journal Editorial Board
- Liaison to OOI Cyberinfrastructure Project

Working Together on A Mosaic for Atmospheric Data

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

Outline

- General Description of the Issues
- GALEON background and Progress
- Different atmospheric data types and established community data systems
- Collections of non-gridded datasets as standard “coverages”
- Which community and formal standards apply?
- CF-netCDF as a separate encoding standard
- Work to be done
- References

Background

What's the problem?

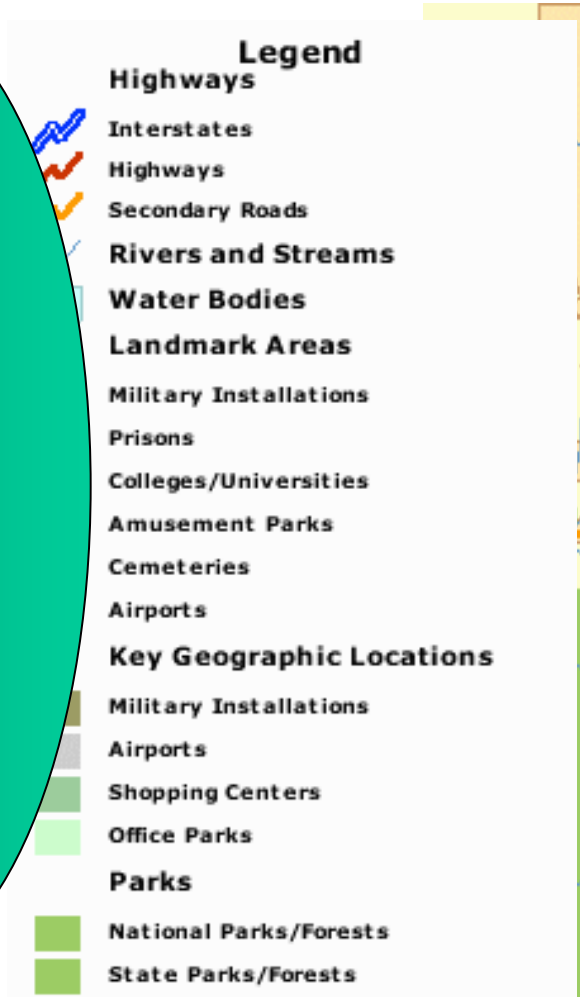
What are we trying to accomplish?

Disparate Data Models: Different Ways of Thinking about Data

- To the GIS (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

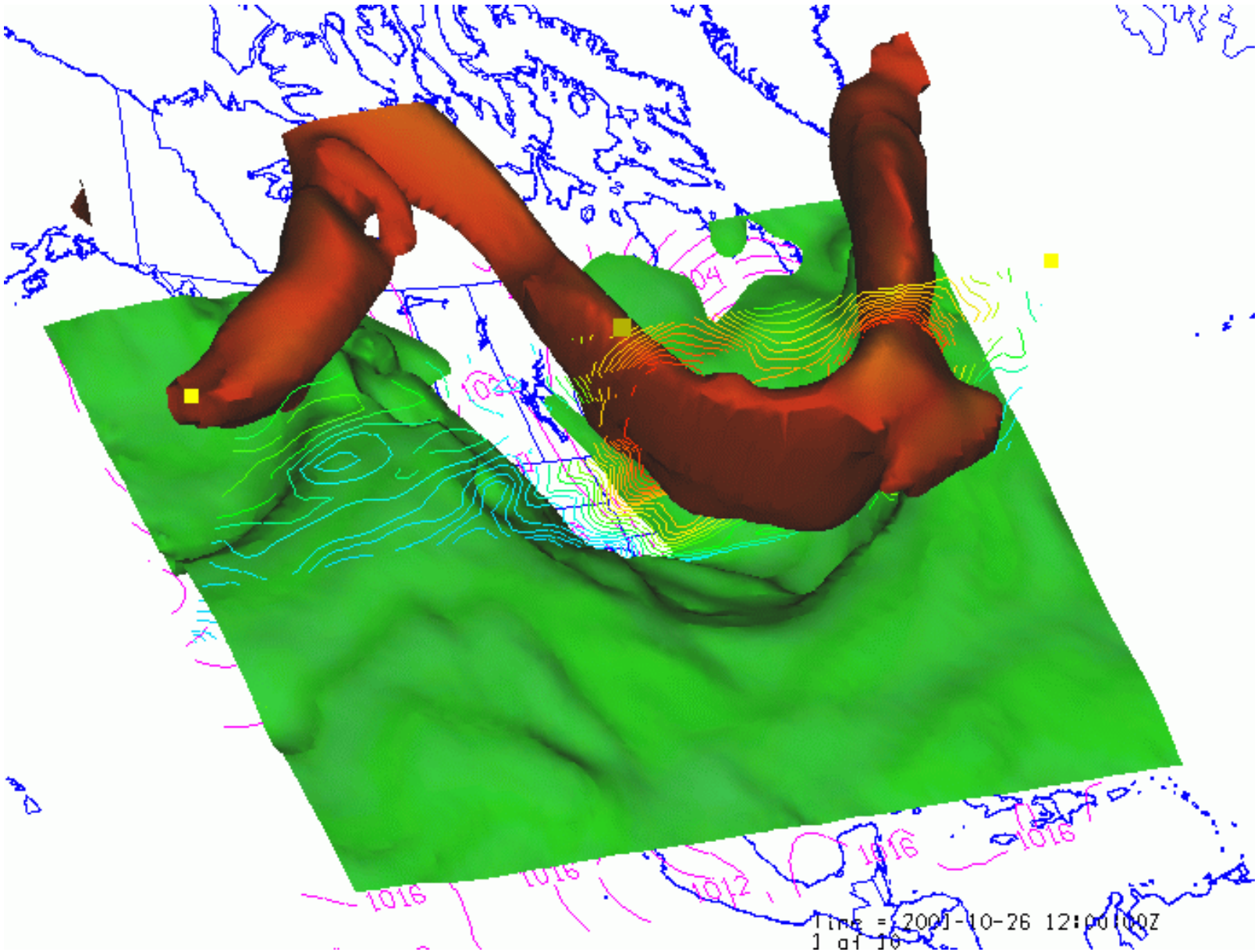
Traditional GIS view

Attributes in DBMS tables

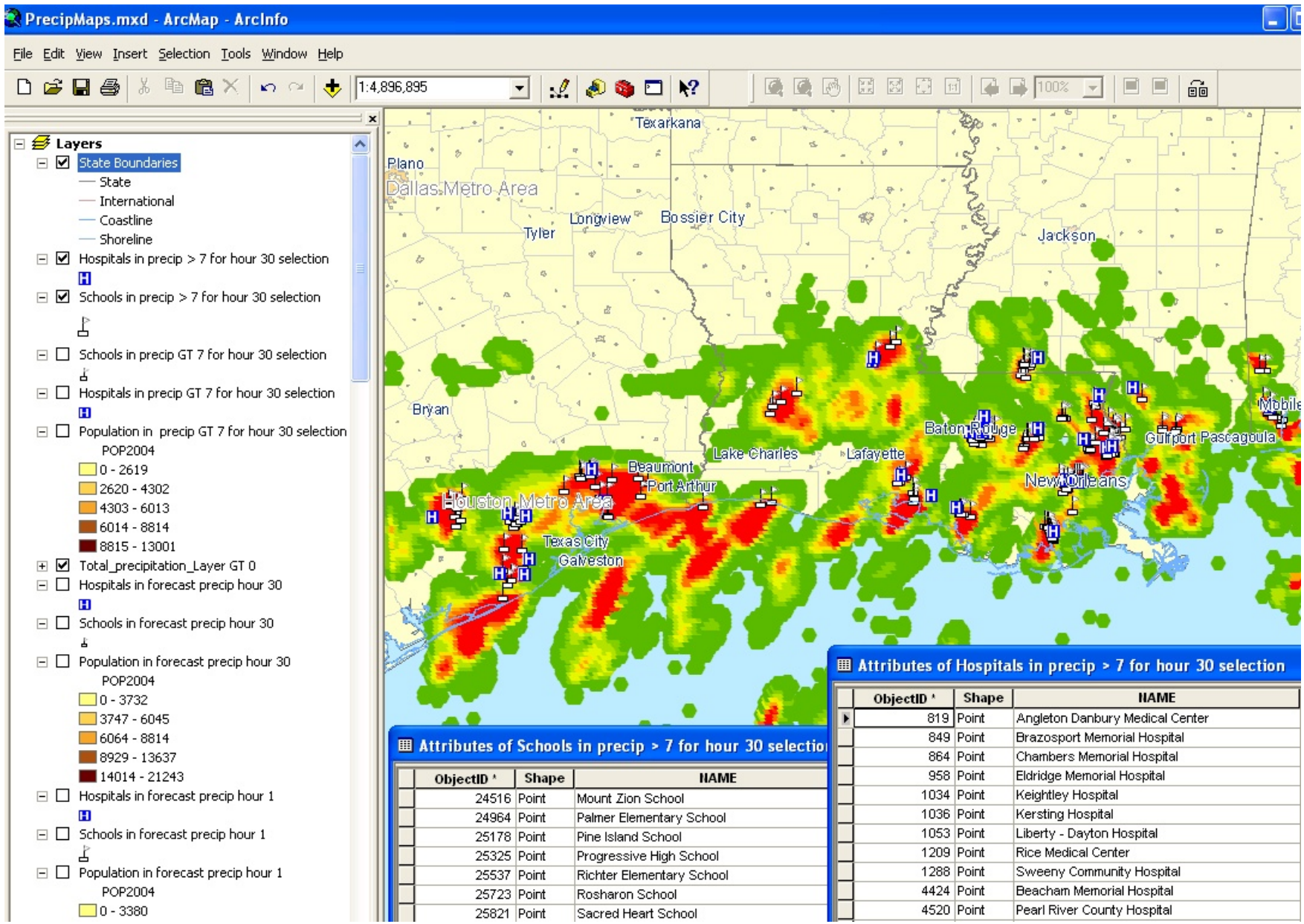


Features as points, lines, polygons

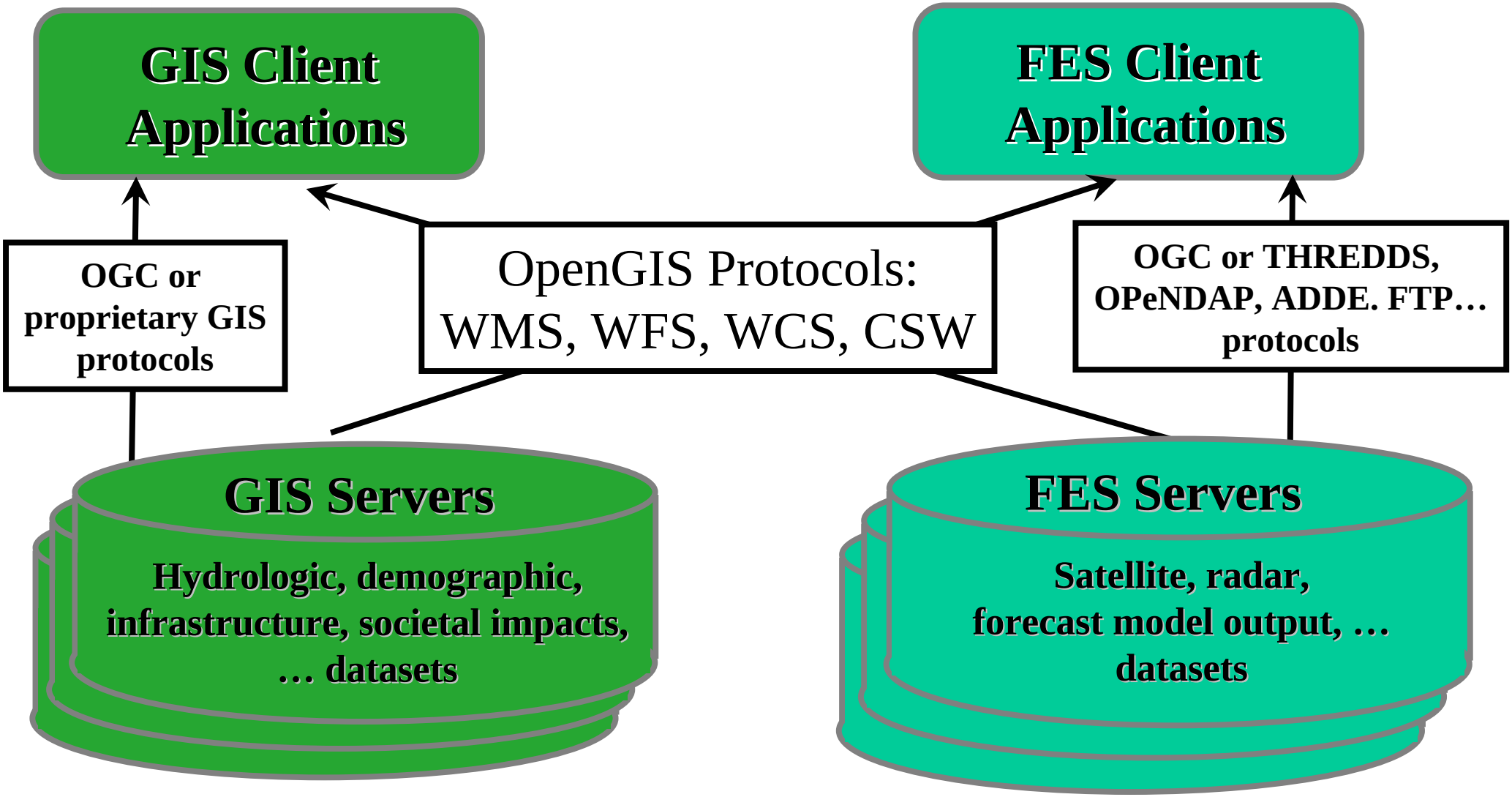
Typical NetCDF Visualization



Apply GIS Tools To Atmospheric Science Data



Taking Advantage of Web Services for Data System Interoperability



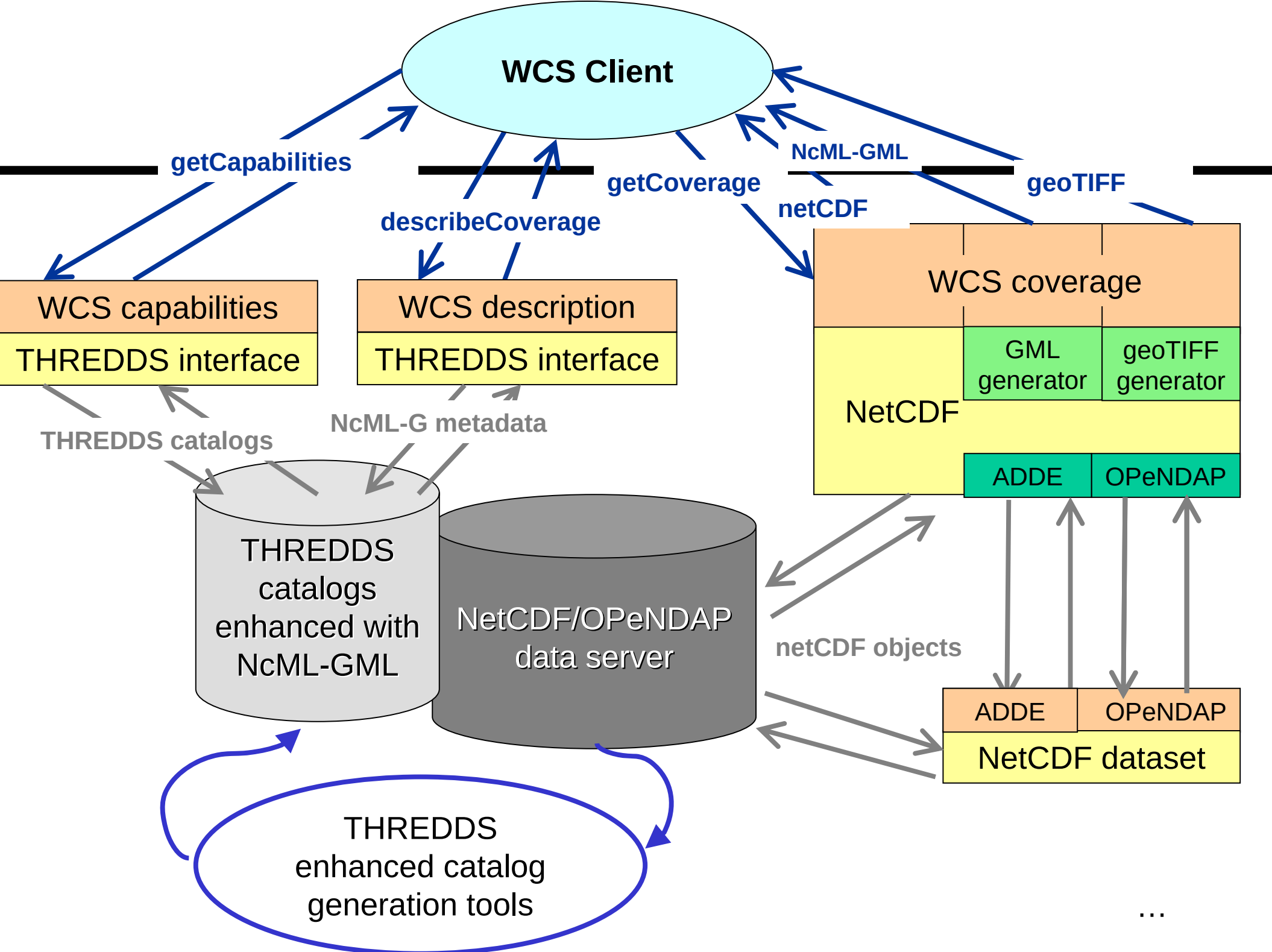
GALEON Background

What has been done so far?

GALEON (Geo-interface for Air, Land, Earth, Oceans NetCDF)

- Provide standard interfaces, e.g.,
  - Web Coverage Service (WCS)
  - Web Feature Service (WFS)
  - Web Map Service (WMS)
  - Catalog Services for the Web (CSW)

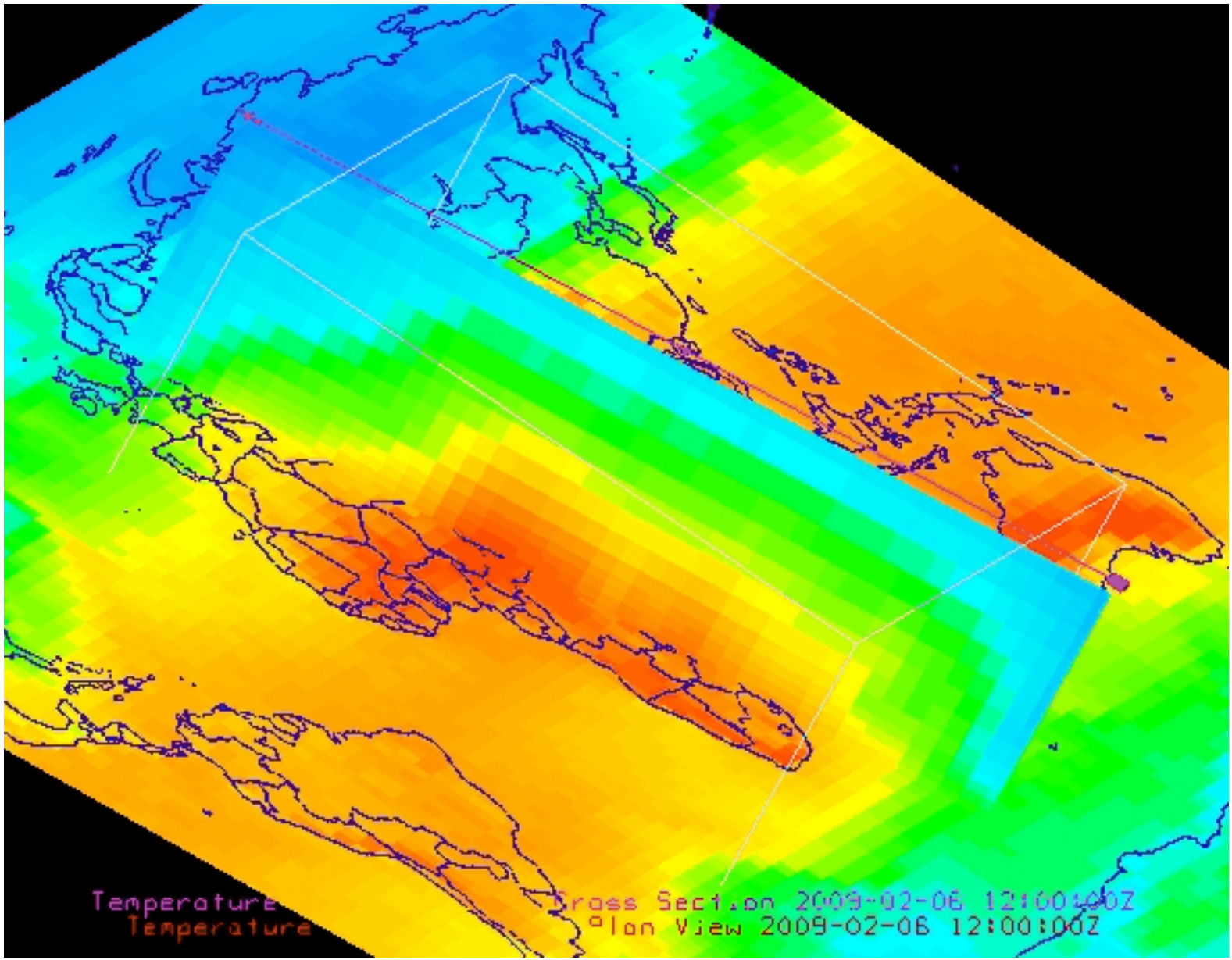
- To existing THREDDS services, e.g.,
  - HTTP access to netCDF
  - OPeNDAP client/server protocol
  - THREDDS catalogs
  - Delivering netCDF binary files


WCS Client
getCapabilities
describeCoverage
getCoverage
NcML-GML
geoTIFF
netCDF
WCS capabilities
THREDDS interface
WCS description
THREDDS interface
WCS coverage
GML generator
geoTIFF generator
NetCDF
ADDE
OPeNDAP
THREDDS catalogs
NcML-G metadata
THREDDS catalogs enhanced with NcML-GML
NetCDF/OPeNDAP data server
netCDF objects
ADDE
OPeNDAP
NetCDF dataset
THREDDS enhanced catalog generation tools
...

GALEON 1 Lessons:

- WCS works well
- Simple space-time bounding box request is useful in many cases
- CF conventions well defined for gridded data
- CF-netCDF via WCS useful for wide range of clients -- from arcGIS to IDL to IDV
- Special MetOcean Community Needs
  - 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 coordinate
  - agreement on CRS (Coordinate Reference System) specifications

GALEON Initial Focus: Gridded Output of Forecast Models

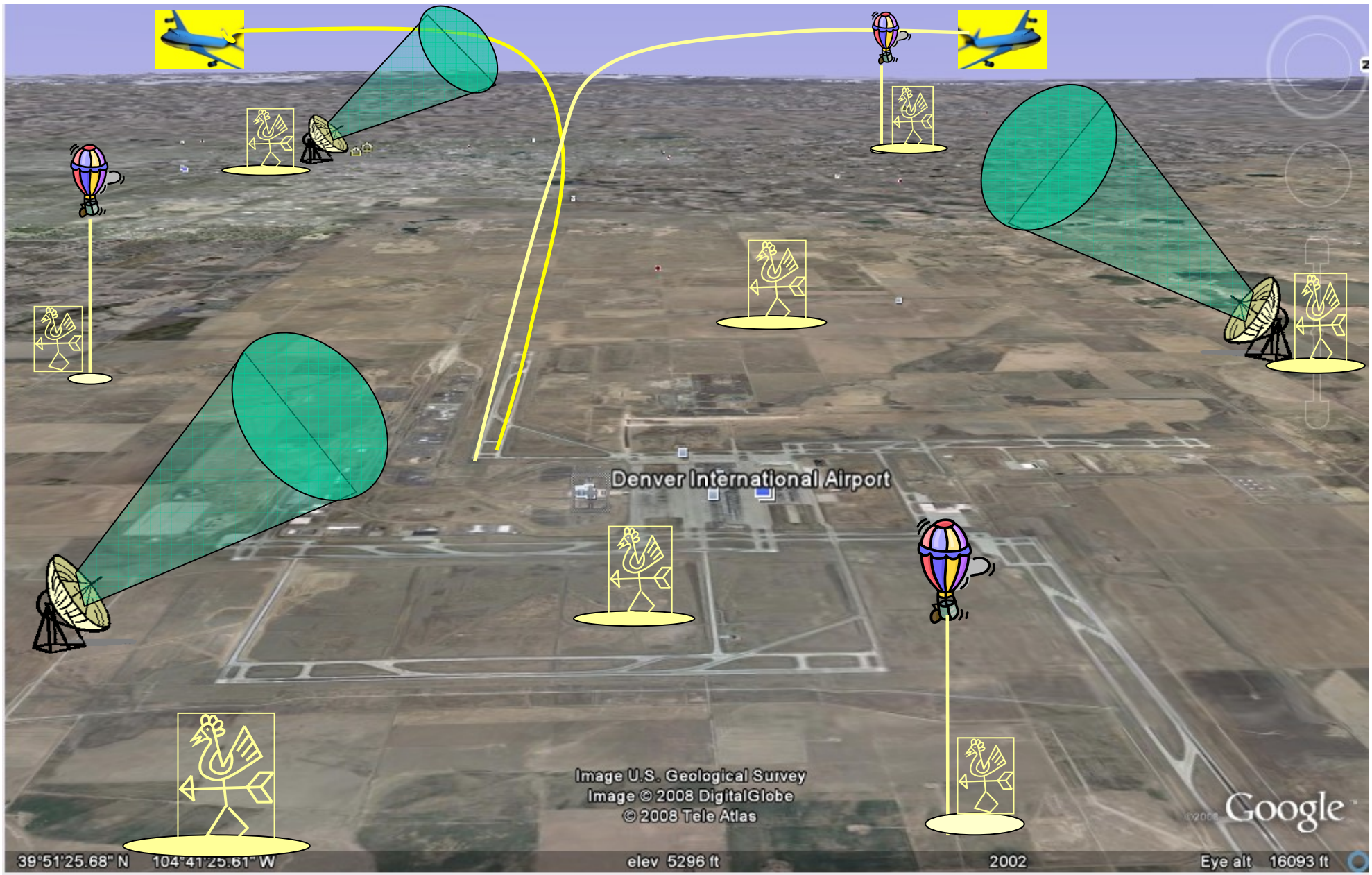


More General Problem: Collections of MetOcean Datasets

How do we deal with

1. collections of
2. many different data types ?
   - in our own MetOceans community?
   - in the world of formal standards?

Airport Weather Use Case:
Multiple Platforms Sampling the Atmosphere



Airport Weather Use Scenario: More than Forecast Model Output

- Integrate and compare model output and observation data near airport
- Specify 3D bounding box centered on airport
- Specify time frame of interest (e.g., periods of severe storms)
- Request observed and forecast atmospheric parameter values
- In GALEON 1, WCS worked well for gridded data from forecast model output and some satellite imagery

Airport Weather Data Types:

Examples of Unidata “Common Data Model” Scientific Data Types
and Climate Science Modelling Language Scientific Feature Types

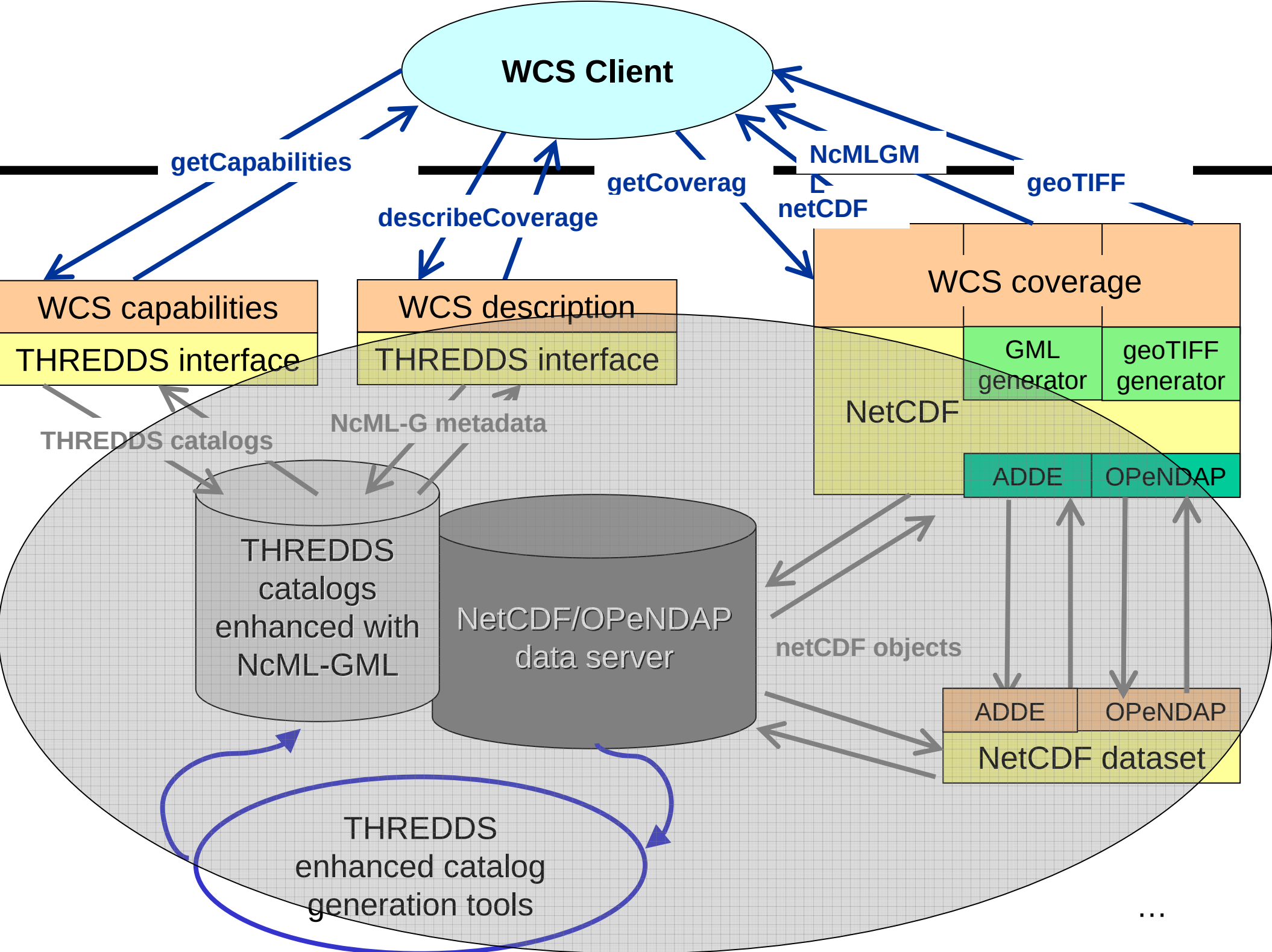
- point data from lightning strike observations
- "station" observations from fixed weather stations
- vertical profiles from balloon soundings and wind profilers
- trajectory data obtained from instruments onboard aircraft which have taken off and landed recently
- volumetric scans from ground-based radars
- visible, infrared, and water-vapor (and possibly other wavelength) **satellite imagery**
- ***gridded output from national or hemispheric weather forecasts*** (typically run at centers like NCEP and ECMWF) -- sometimes used as boundary conditions for a higher-resolution local forecast model.

Special Requirements for Weather Data

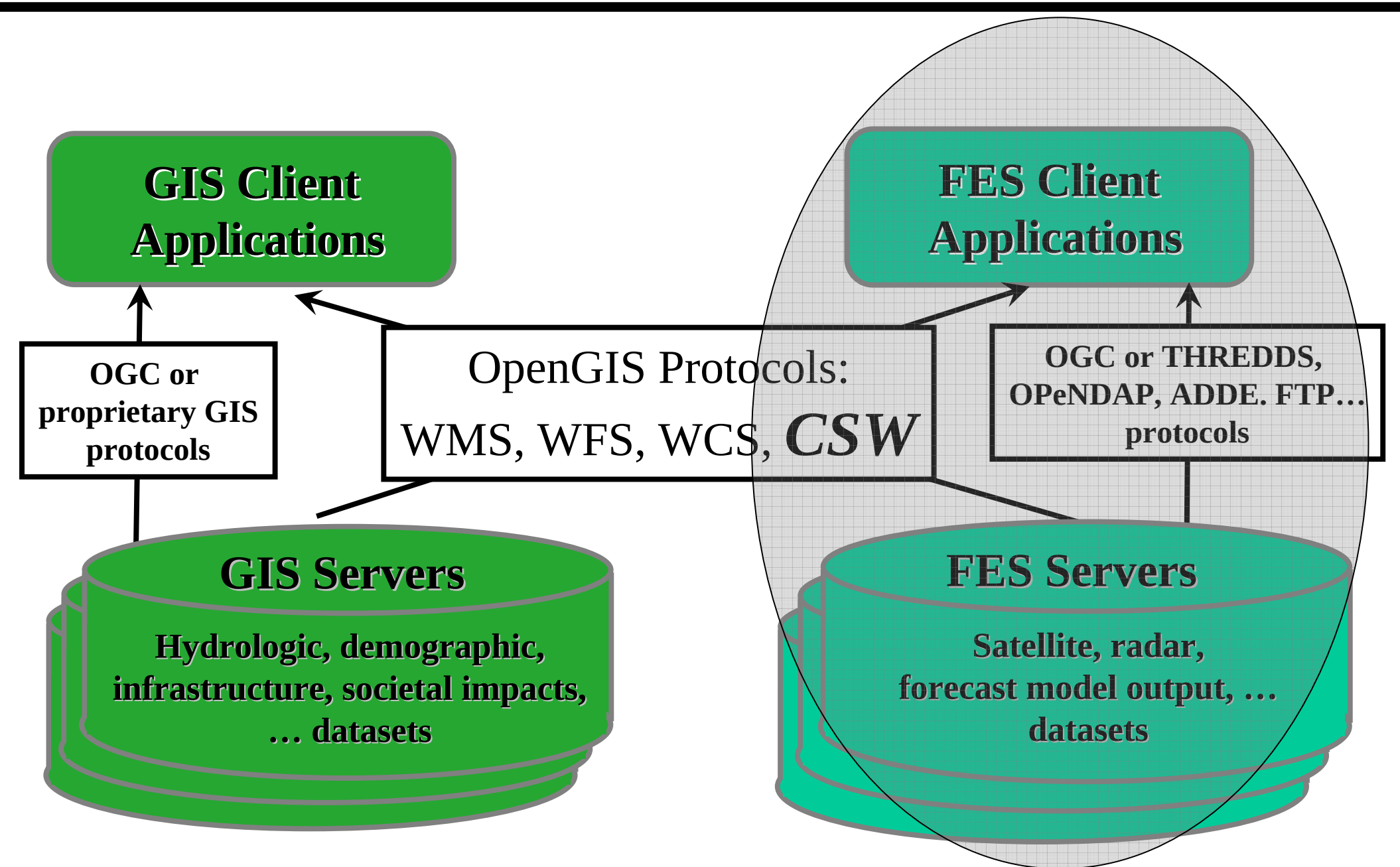
- Real-time access
- Elevation/altitude dimension is important
- Elevation dimension often given in terms of pressure
- Range value interpolation depends on physics (and data) whereas GIS world is concerned mainly with geometry
- Automated processing components, e.g.,
  - Gridding/assimilation
  - Forecast models
  - Transformations between pressure and height

Existing Systems that Work

We have a solid set of
established data systems serving
the MetOceans (or FES) community


WCS Client
getCapabilities
describeCoverage
getCoverag
NcMLGM
netCDF
geoTIFF
WCS capabilities
THREDDS interface
WCS description
THREDDS interface
WCS coverage
GML generator
geoTIFF generator
NetCDF
ADDE
OPeNDAP
THREDDS catalogs
NcML-G metadata
THREDDS catalogs enhanced with NcML-GML
NetCDF/OPeNDAP data server
netCDF objects
ADDE
OPeNDAP
NetCDF dataset
THREDDS enhanced catalog generation tools
...

Taking Advantage of Web Services for Data System Interoperability



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, trajectory, radial, unstructured grids?

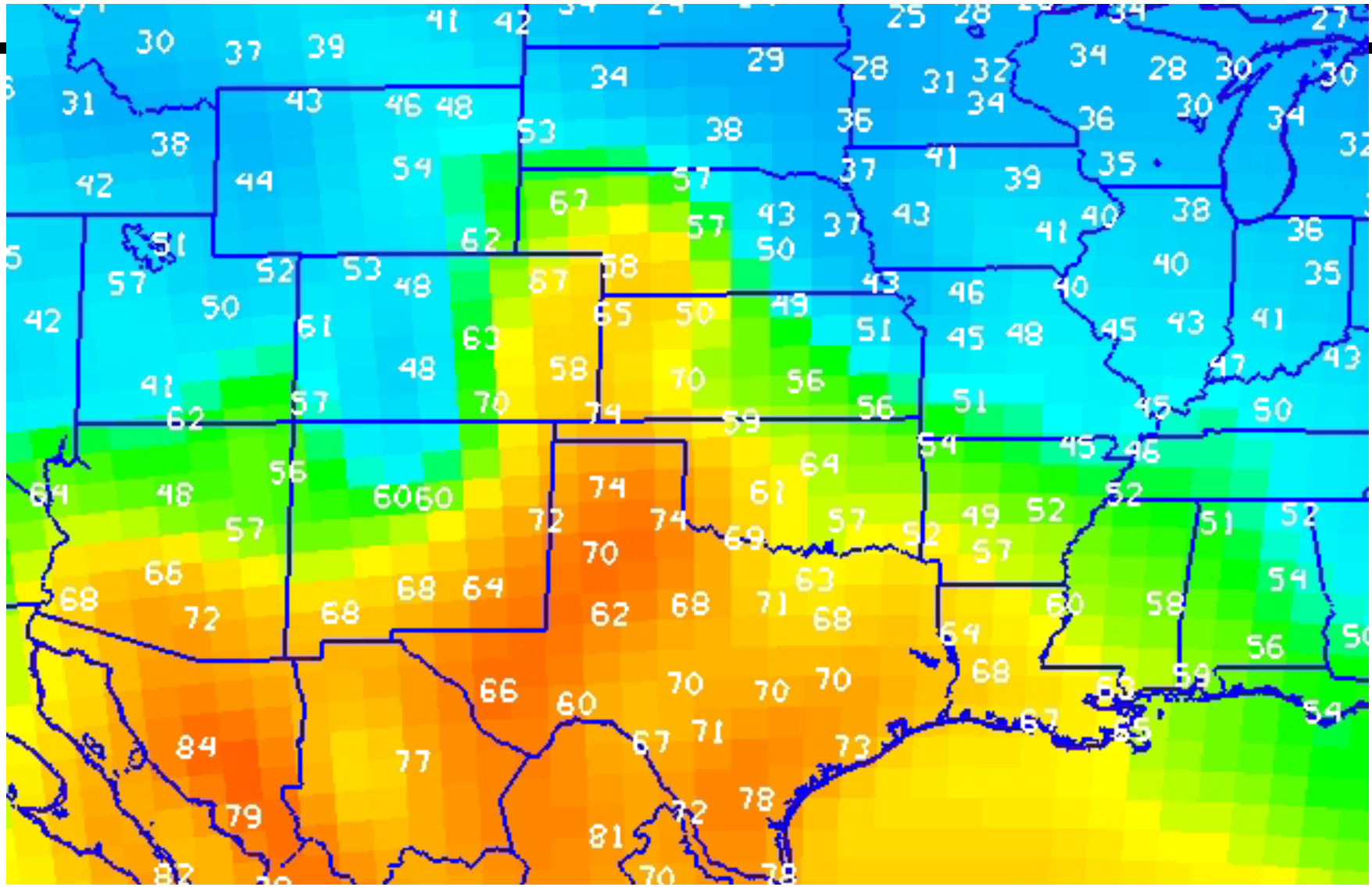
Standard Interfaces for Serving ***Collections*** of Different Data Types

How do we serve collections of different MetOceans data types via standard interfaces and protocols?

Are These Collections Coverages?

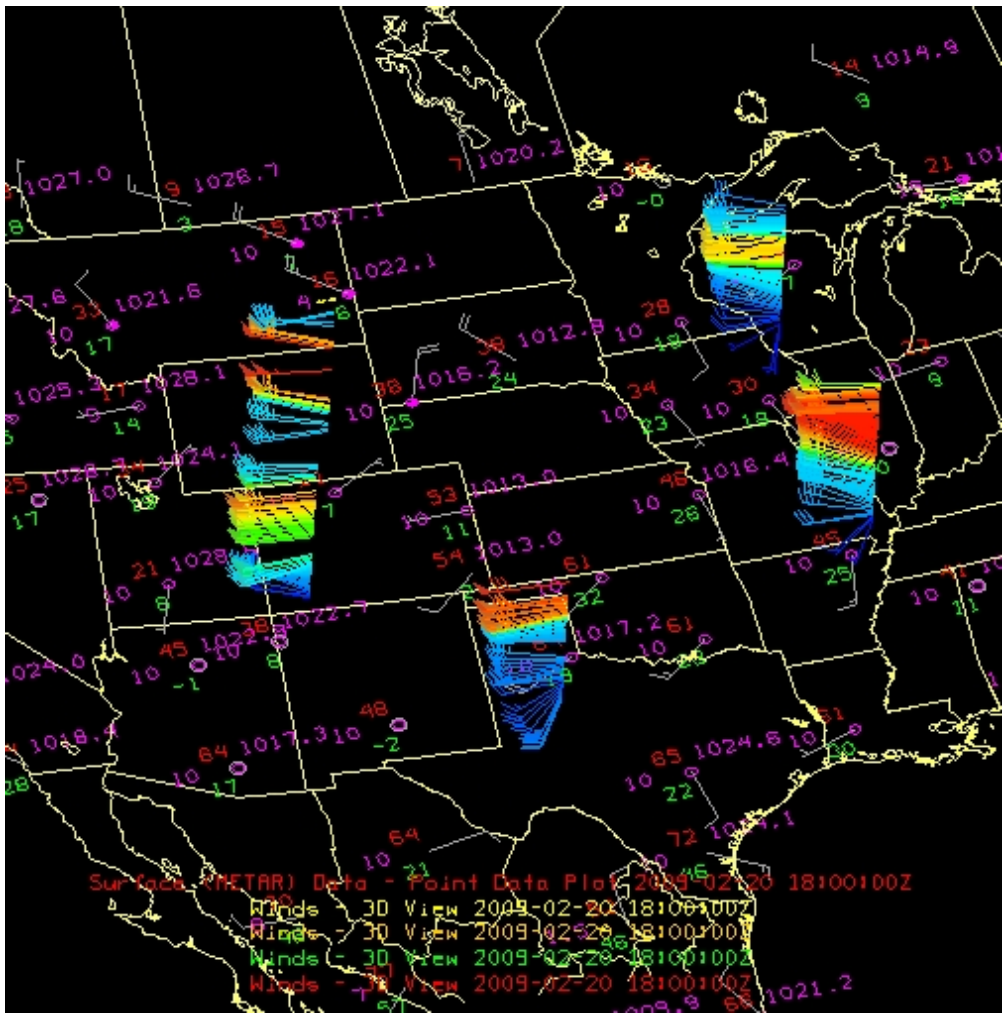
- Data request similar to that of WCS is useful for cases comparing forecasts and observations
- ISO general feature model calls them “aggregations”
- ISO 19123 definitions of coverage includes:
  - grid,
  - point
  - curve
  - surface
  - solid
- But WCS only serves regular grids at this point

Collections of Station Observations

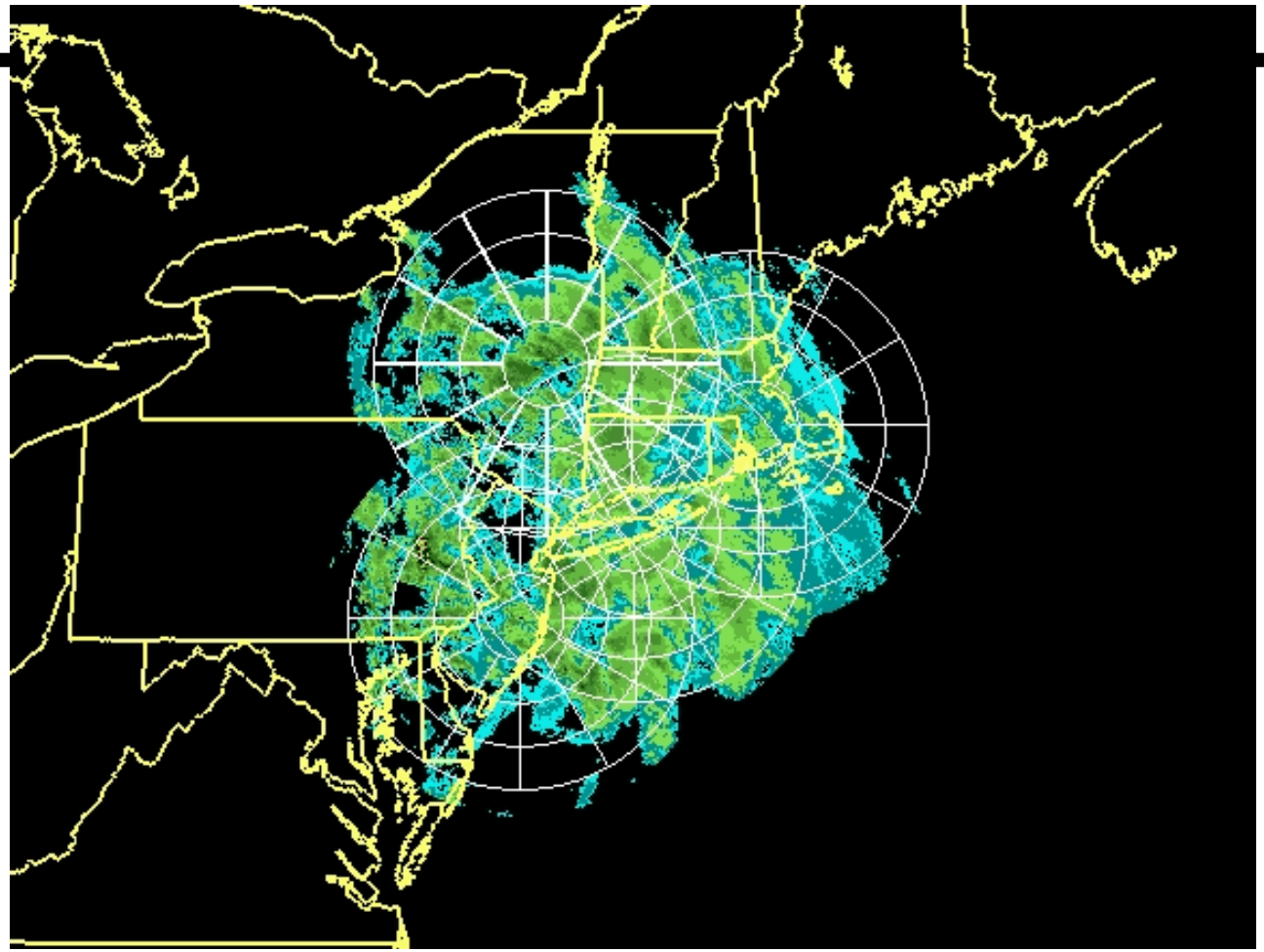


Common Use Case:
comparing forecast and observations for the same region and time

Different Types of Weather Station Obs



Radar Data Collections



Collections of data from individual radars look a lot like the gridded coverages output from weather forecast models or satellite imagery. But the “range rings in the animated illustration show clearly that determining the locations of individual data points is more complicated than for regularly spaced grids.

ISO 19123 Coverage Definition: Background Information

- A coverage is a feature that associates positions within a bounded space (its domain) to feature attribute values (its range). In other words, ***it is both a feature and a function.***
- Examples include a raster image, a polygon overlay or a digital elevation matrix.
- A coverage may represent a single feature or a set of features
- A coverage domain is a set of geometric objects described in terms of direct positions.
- The direct positions are associated with a spatial or temporal coordinate reference system.
- *Commonly used domains include **point sets**, grids, collections of closed rectangles, and other collections of geometric objects.*

Which Standards Apply?

For collections of: lightning strike point observations, weather station observations, vertical profiles, onboard aircraft observation trajectories, volumetric radar scans, satellite swath images

- If these are coverages, should **WCS** apply for non-gridded datasets?
- Fit with Sensor Web Enablement (**SWE**) Observations and
- Measurements (**O&M**)?
- Relationship to **ISO 19123 Coverage** specification?
- Delivery via **WCS, WFS, SOS**?
- **ISO 19111 Coordinate Reference System** for collections
- Web Processing Services (**WPS and WCPS**)
- **GML** role: CSML, NcML-GML, GML-JP2K?
- **CS-W** (Catalog Services for the Web) cataloging

WCS and SWE O&M

- *Feature of Interest* – bounding box and time frame in WCS
- *Sampling Feature* (FES data sets are discrete samples of continuously varying properties of the feature of interest)
- Collections of Sampling Features as "*Sampling Coverages*"?
- Observations and Measurements Documents (up for revision) http://www.opengeospatial.org/standards/om

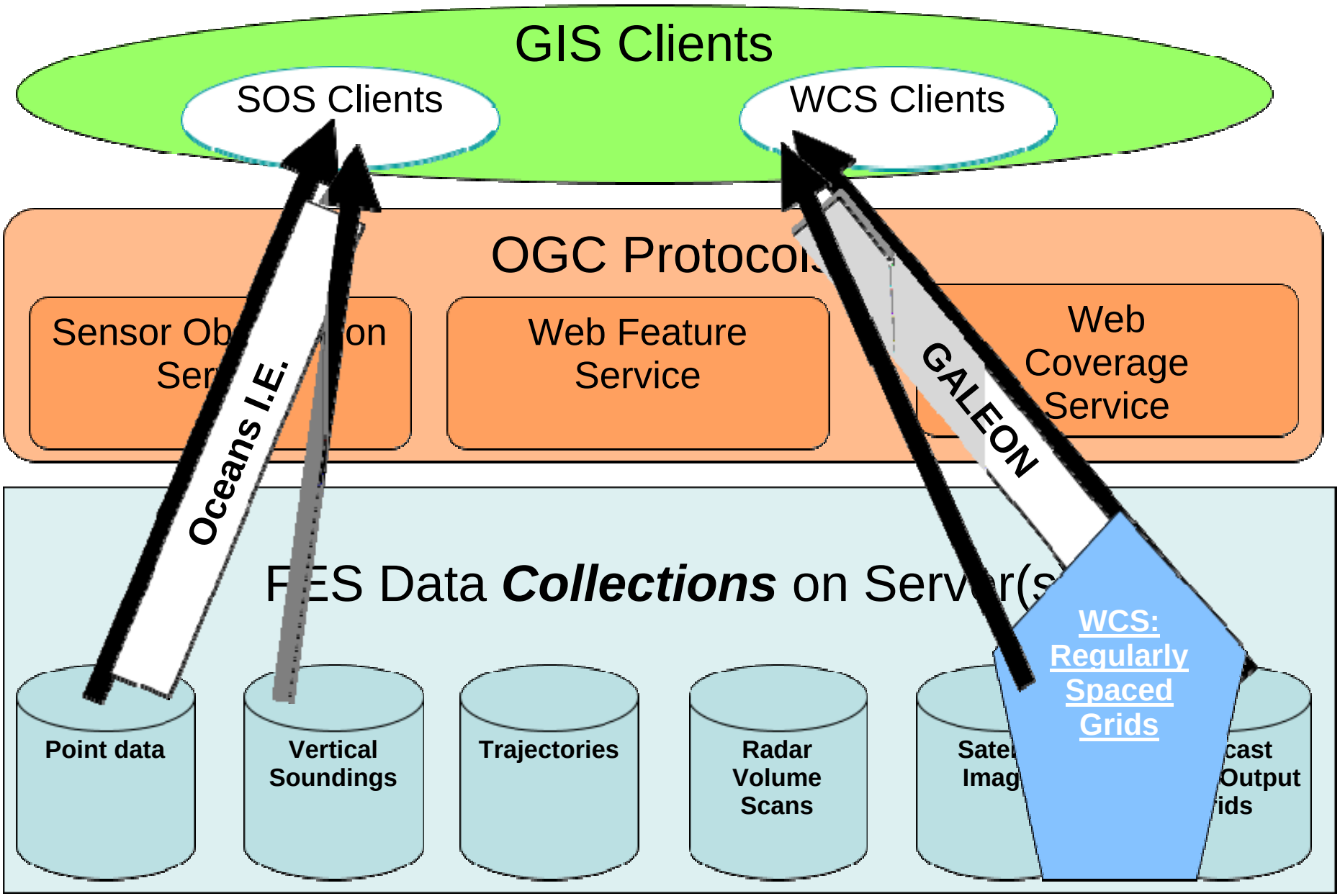
Service Protocols

So what’s the proper protocol for serving these many and varied data types?

Data Access Alternatives

- **WCS** was shown to work well in GALEON 1 for straightforward data access use case, but only for regularly-spaced grids. (*GALEON focus*)
- Points, trajectories, vertical profiles are thought of as “features,” but **WFS** has limitations when it comes to collections of features and the time dimension. (*British Atmospheric Data Center CSML)*
- **SOS** works for time series of observations from sensors, but not for space-time bounding box requests in its present form. (*OGC Oceans Interoperability Experiment)*

Data Types and Service Protocols

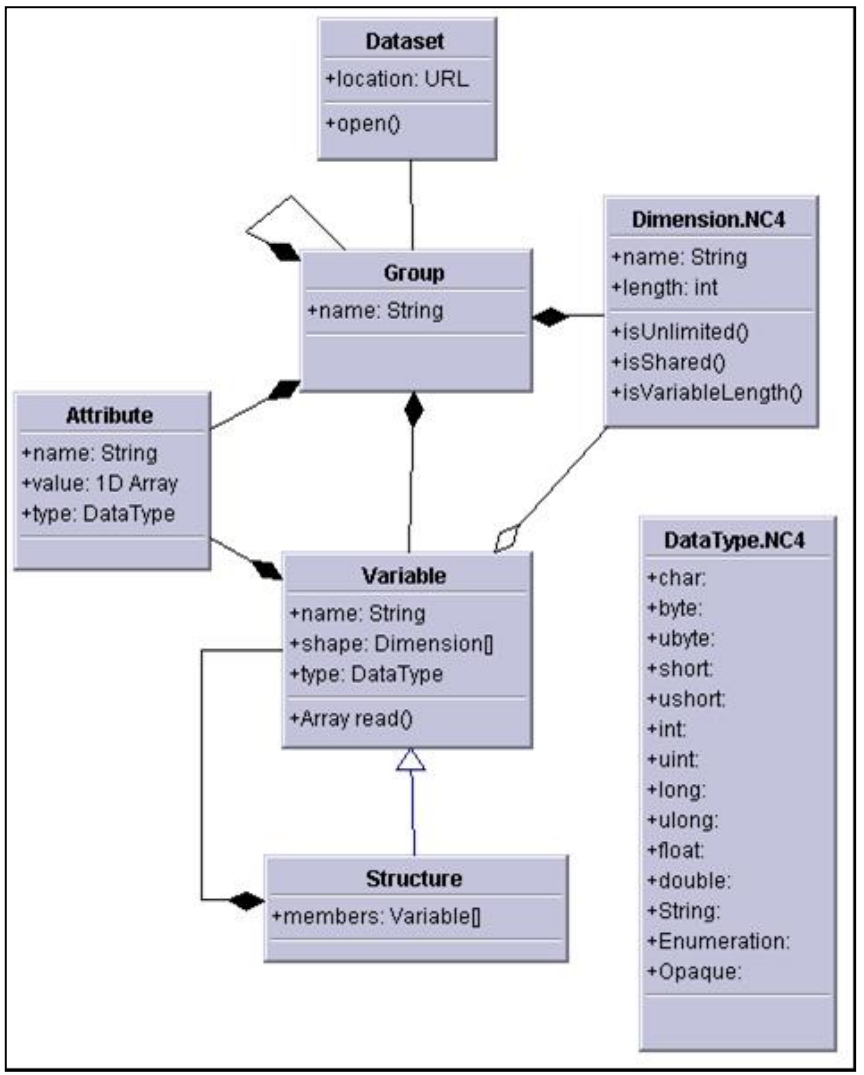


Data Models

What is a data model?

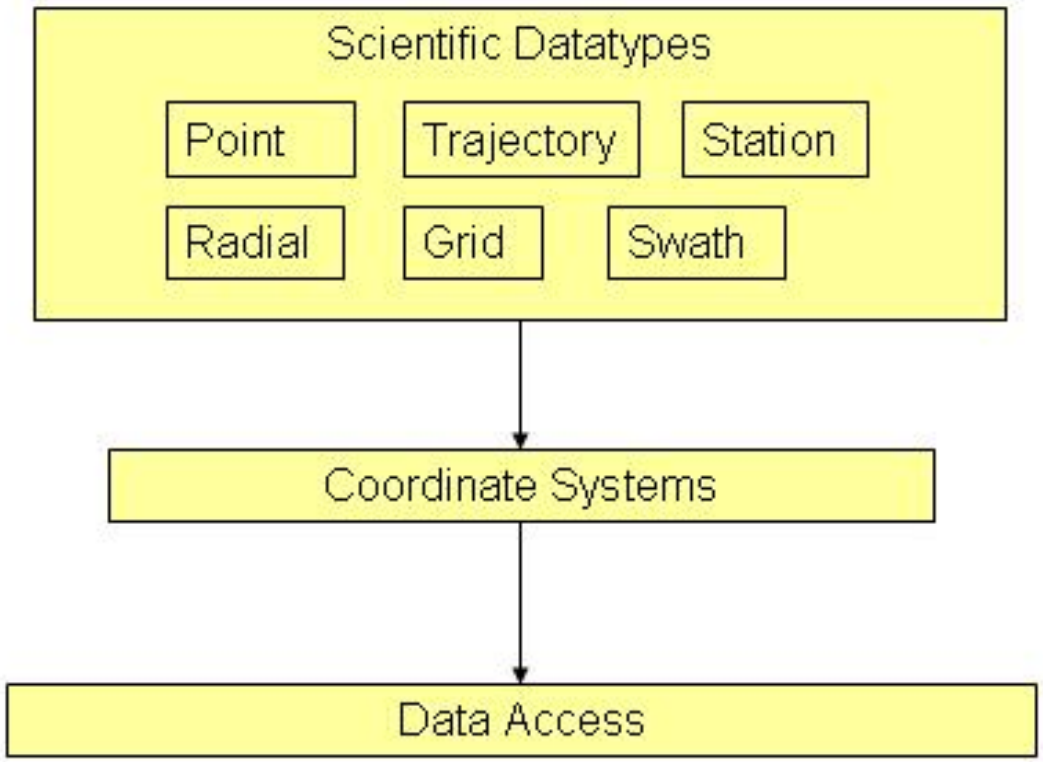
A database schema?

Something described by a UML diagram?



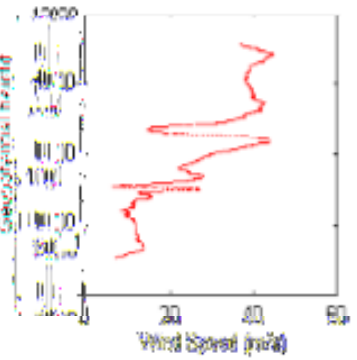
Unidata access layer CDM
(Common Data Model)

CDM Scientific Data Types

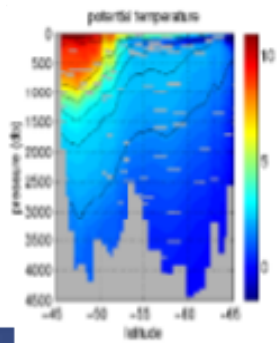


Climate Science Modelling Language Scientific Feature Types of BADC

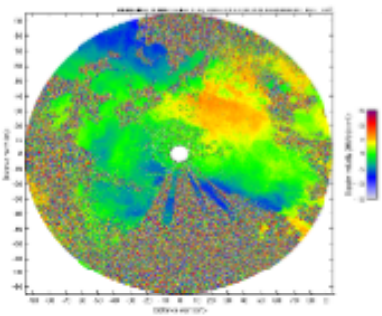
ProfileFeature



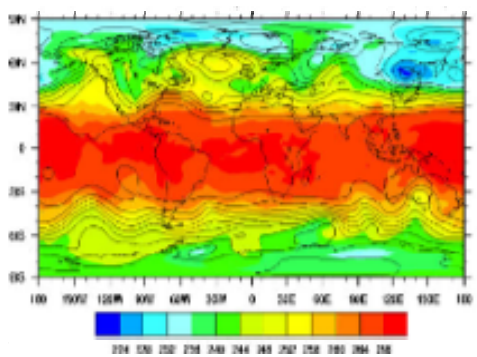
RaggedSectionFeature



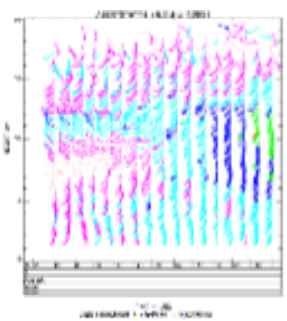
ScanningRadarFeature



GridFeature



ProfileSeriesFeature



Thanks to Andrew Woolf of BADC

CSML-CDM Mapping

CSML Feature Type	CDM Feature Type
PointFeature	PointFeature
PointSeriesFeature	StationFeature
TrajectoryFeature	TrajectoryFeature
PointCollectionFeature	PointFeature collection at fixed time
ProfileFeature	ProfileFeature
ProfileSeriesFeature	StationProfileFeature at one location and fixed vertical levels
RaggedProfileSeriesFeature	StationProfileFeature at one location
SectionFeature	SectionFeature with fixed number of vertical levels
RaggedSectionFeature	SectionFeature

At the Abstract Standard Level ISO 19123 Coverage Model

- Up for revision
- In most cases, a continuous coverage is also associated with a discrete coverage that provides a set of control values to be used as a basis for evaluating the continuous coverage.
- Evaluation of the continuous coverage at other direct positions is done by interpolating between the geometry value pairs of the control set (thiessen polygon, quadrilateral grid, hexagonal grid, TIN, segmented curve)* l
- Discrete coverage types can represent sampling features of O&M
- Collections of sampling features as sampling coverages*

*Possible candidates for revision that’s underway

Scientific Data Types
Mapping to ISO Coverages

Unidata CDM Scientific Data Type	**ISO 19123 Coverage Type**
Unstructured Grid	DiscretePointCoverage*
Structured Grid	DiscreteGridPointCoverage
Swath	DiscreteSurfaceCoverage
Unconnected Points	DiscretePointCoverage*
Station observation/Timeseries	DiscretePointCoverage
General Trajectory	DiscretePointCoverage* or DiscreteCurveCoverage
Vertical Profile	DiscretePointCoverage*
Radar Radial	DiscreteSurfaceCoverage or DiscreteCurveCoverage

*Generally, the domain is a set of ***irregularly distributed points***

Coordinate Reference Systems (CRS)

How do we specify where things are in space?

Earth Coordinate System Basics

- Coordinates relative to mean sea level (MSL) ellipsoid or geoid (gravity irregularities)
- 2D position on surface
  - geographic (latitude, longitude) or
  - projected (onto x, y coordinates)
- Elevation relative
  - spatial elevation relative to MSL
  - elevation relative to actual surface of Earth (digital elevation model relative to MSL)
  - data dependent proxy (e.g., air pressure, data-dependent physics, e.g., hydrostatic equation, relative to MSL)

ISO 19111 Coordinate Systems

- Earth referenced coordinate reference system (CRS)
- Engineering coordinate system (with point in Earth-referenced CRS as origin
- Image coordinate system
- ISO Document 19111: Geographic Information: Spatial Referencing by Coordinates
- ISO 19111-2 allows for non-spatial elevation dimension

Engineering Coordinate Systems

- Not directly Earth referenced
- Most remote sensing systems
- Examples:
  - Wind profiler
  - Surface radar scanning
  - Satellite scanning algorithms
  - Aircraft-borne radar

Compound CRS

(Ben's simplified version to illustrate atmospheric data use cases)

	Earth referenced horizontal	**Earth referenced vertical**	**Remote sensing or engineering**
Lightning	Explicit random	Implicit surface	N/A
Station observations	Tabular station	Tabular or implicit surface	N/A
Aircraft or ship observations*	Explicit trajectory	Explicit	N/A
Model output	Fixed grid	Fixed grid (often not spatial)	N/A
Vertical Profiles	Tabular station	Explicit or fixed grid	Vertical "scan"
Ground-based Radar	Tabular station	Tabular	Radar scan
Aircraft or ship remote sensing*	Explicit trajectory	Explicit	Instrument scan
Satellite*	Algorithmic trajectory	Algorithmic trajectory	Instrument scan
GOES Satellite	Explicit or algorithmic trajectory	Explicit or algorithmic trajectory	Instrument scan

**Moving observation platform.*

Data point locations

- Explicit with each data point, e.g., lightning
- Tabular, e.g., repeated observations at fixed* station locations
  (*Note that station locations may change, but not often compared to data value changes)
- Fixed algorithmic grid, e.g., output of forecast models
- Moving platform - explicit locations, e.g. aircraft-borne observations along flight paths (trajectories)
- Moving platform – algorithmic location, e.g., satellite position given by orbital mechanics

Image CRS

- Recent focus of OGC WCS and CRS working groups
- Specifies coordinates in terms of indices
- Can be related to Earth referenced CRS via an algorithm, projection ID, or table look up
- Many similarities to netCDF and OPeNDAP means for specifying CRS

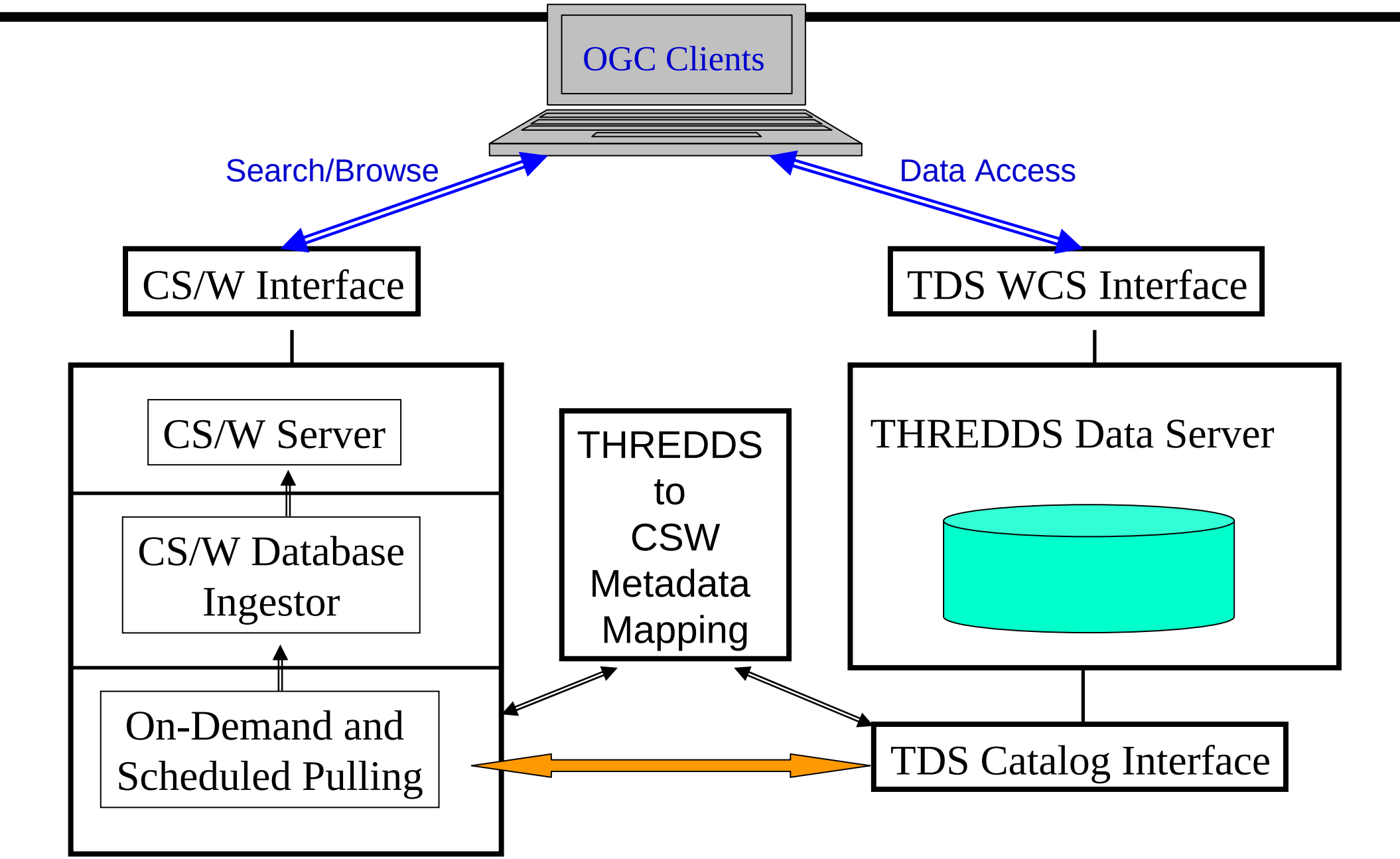
Other Related Standards

There are several other standards specifications that are related to our efforts but beyond the scope here.

Web Processing Services

- Interpolating gridded data to points
- Assimilating observed data samples to grid
- Converting from pressure to height and back
- Most transformations depend on physics (and data as well)
- WCPS available as well as WPS
- References?

CS/W-THREDDS Gateway



GML

- Beyond scope here
- OGC Document
- Core plus extensions approach
- Special focus of BADC collaborators
- Related to GALEON
  - WCS manifest
  - CSML
  - NcML-GML
  - GML-JP2K

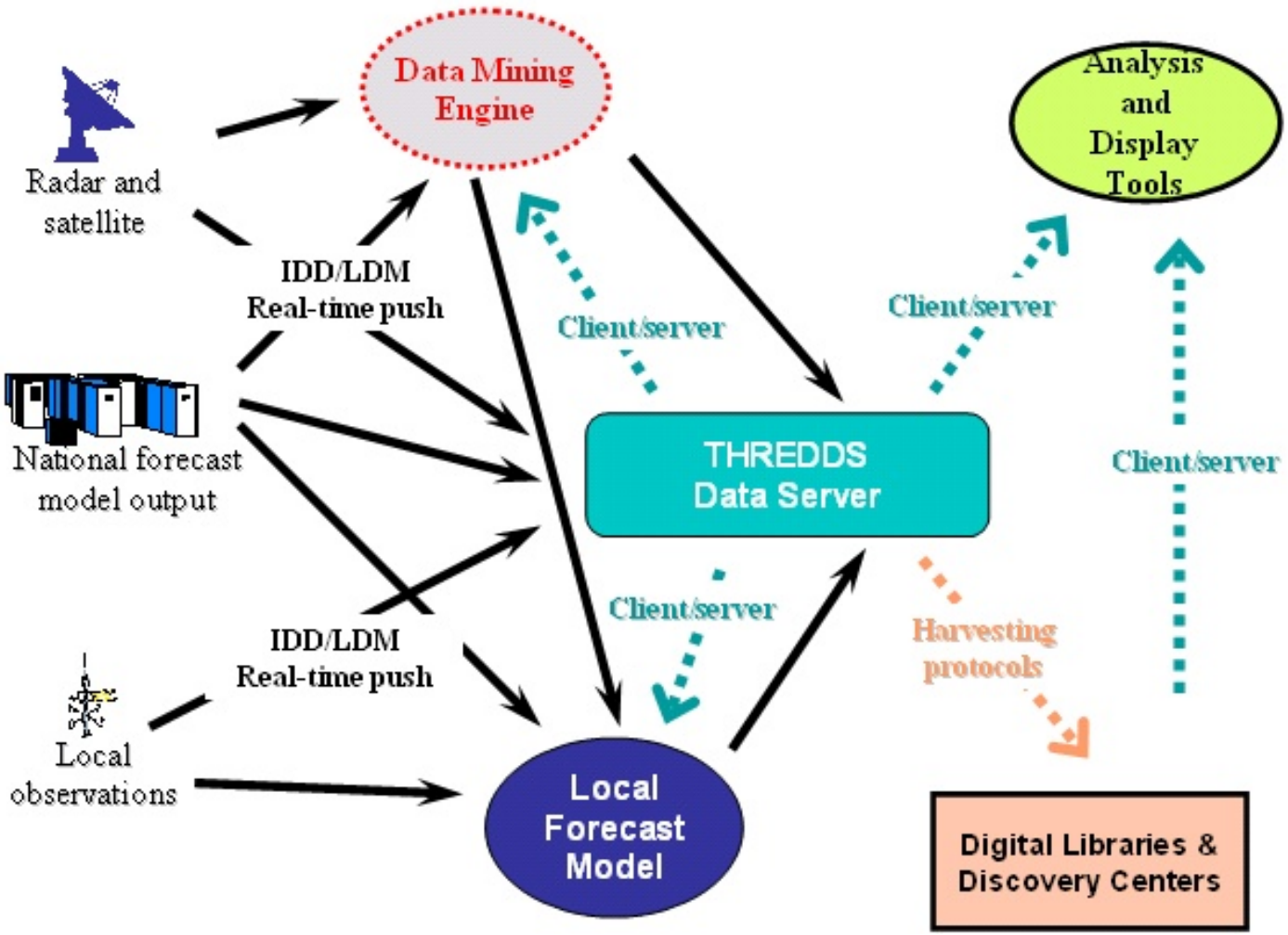
CS-W Cataloging

- CS-W Specification
- U of Florence Gi-GO Client
- ESRI Client
- GMU CS-W service for THREDDS Data Server

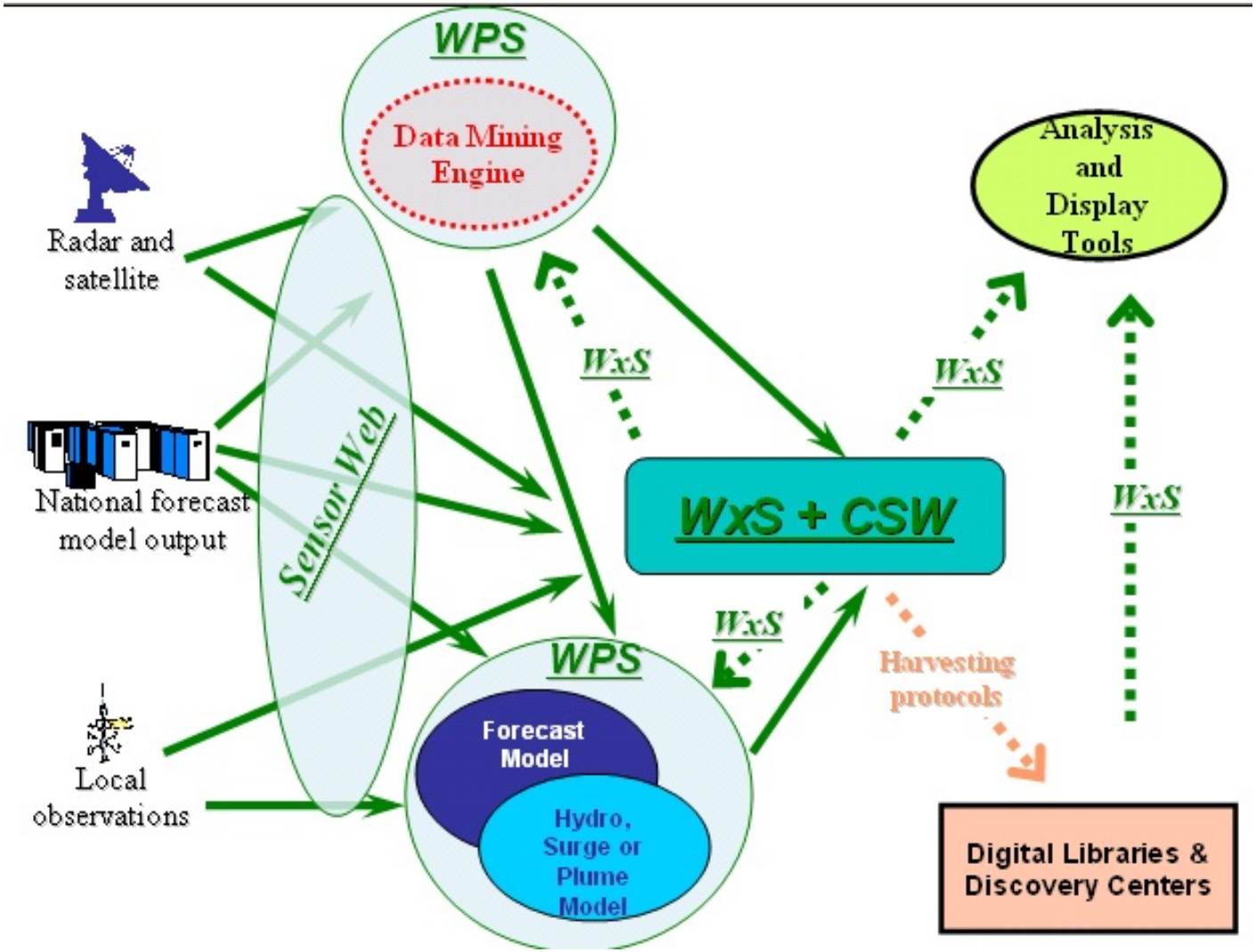
Where To From Here?

The key challenges are to select the right standardiztion areas for applying limited resources.

End-to-End Data/Forecast System



End-to-End Data/Forecast System via Standard Interfaces



Action Plan Outline

- Agree on high-level dataset categories
- Clarify relationships among:
  - Unidata CDM Scientific Data Types
  - CSML Scientific Feature Types
  - Obs. & Meas. Sampling Features
- Establish extensions to CF conventions for each dataset category
- Map CF-netCDF categories to ISO 19123 (possibly modifying ISO 19123)
- Establish metadata forms: CSML, ncML-GML
- Establish CF-netCDF as a separate OGC standard
- Experiment with CF-netCDF encoded coverages as payload for WCS, WFS, SOS

Working Together on A Mosaic for Atmospheric Data

Divide (Labor) and Conquer

- Coordinate individual efforts toward a whole (mosaic) greater than the sum of the parts
- Each group focuses on areas of expertise
- Work on tasks each group has funding for
- Stay aware of other groups' efforts
- Coordinate efforts wherever possible
- Results of lessons learned from implementation and experimentation feeds into OGC standard definition process
- OGC liaison takes recommended changes to ISO
- E.g., ISO 19111, Coordinate Reference System Part 2: Extension for parametric values

ESRI arcGIS Specifics

- CF-netCDF direct access is a powerful addition for local MetOcean datasets
- WCS access via python is an effective mechanism for remote access
- BADC addition of WCS client library to python OWSlib (also has WMS, WFS) makes python more generally useful
- WCS client implementation in arcGIS lacks netCDF access
- Limited success with CS-W catalog access
- Commitment to netCDF4 will be valuable – also provides bonus access to HDF5 files.

CF-netCDF Role

An important new development is the possibility of proposing CF-netCDF as a separate standard for binary encoding.

CF-netCDF as a Standard

- Previous efforts centered on CF-netCDF as an standard extension for WCS
- Considerable discussion of delivering CF-netCDF as a coverage feature for WFS
- Possible “out of band” binary payload for SOS
- Why not propose CF-netCDF as an OGC binary encoding specification independent of delivery protocol?
- Then propose extensions to WFS, WCS, SOS delivery protocols referring to CF-netCDF encoding spec

Advantages of Independent CF-netCDF Encoding Specification

- Fits with OGC Grid Coverage Common
- Need specifications for each protocol, but this approach simplifies each specification document
- No need to specify delivery specific details, e.g., getCoverage, getFeature, getObservaion with the encoding specification
- Delivery specifications (WCS, SOS, WFS) can point to the binary encoding specification
- Encoding spec version numbers not tied to delivery spec versions

CF-netCDF Standardization Issues

- Specifying file format, API or code base?
- netCDF3, netCDF4, ncML (netCDF Markup Language)?
- HDF5 file format for netCDF4
- netCDF control (not really an issue – stays with Unidata for netCDF)
- CF control (remains with current CF body)
- IP issues (under discussion but appears manageable)

GALEON Community Homework

- Establish CF-netCDF as an OGC standard
- Finish work to establish CF-netCDF as WCS extension
- Continue efforts to map non-gridded data collection types to standard coverages, features, observations
- Establish CF conventions for non-gridded data collections: e.g. https://cf-pcmdi.llnl.gov/trac/wiki/PointObservationConventions (upcoming GO-ESSP meeting)
- Work with WCS, WFS, SOS working groups to establish specs for accessing such data collections
- Figure out how IDD/LDM fits into all this (main source of personal frustration)

CF-netCDF as WCS Encoding:
in annexes of proposed WCS BP doc

- CF-netCDF describeCoverage respons
- Domain, range, field coverage data structures
- CF-netCDF getCoverage response
- GetCoverage response for CF-netCDF data
- OutputCoverage
- GridCoverageValues
- Manifest (Coverages data structure)
- RequiredOutputCoverageMetadata
- GridCoverageFile
- GridCoverageValuesURI association
- CF-netCDF file
- NcMLDataset
- OPeNDAP-URL
- Content model of the WCS complete GetCoverage response for CF-netCDF3 binary file
- Complete GetCoverage response for ncML document
- Partial GetCoverage response
- WCS GetCoverage response: Multipart data encoding
- SOAP with binary data and HTTP responses
- Proposed extensions for handling ncML Responses
- Examples
  - Content-ID generation netCDF 3 with CF1.1 convention
  - ncML dataset
  - GetCoverage response encoding examples
  - SOAP Request of two netCDF data items and metadata
  - HTTP Request of two netCDF data items and metadata
  - SOAP Response with binary and ncML data
  - Multipart section containing ncML with binary data included
  - Multipart section containing ncML with binary data extracted using XOP

CF netCDF Coverage Encoding
in body of proposed WCS BP document

- Overview of netCDF and CF conventions
- NetCDF-3 Data Model
- NetCDF Coordinate Variables
- NetCDF Standard Attribute Conventions
- NetCDF-3 Binary File Format
- NcML (netCDF Markup Language)
- CF Standard names
- CF Units
- CF Coordinate types and coordinate systems
- CF Grid Cells
- Code for Implementing the netCDF Interface
- Documentation, Support, Examples
- Compliance?
- CF-netCDF Mapping to WCS Coverage Data Model
- CF-netCDF grid data profile model and ISO DiscreteGridPointCoverage profile model
- Mapping Rules
- Limitations

Process (modeled on KML Approach)

- Start with Best Practice
- Form a team to do the RFC.
- Little "negotiation" because of broad use.
- Alignment with some OGC and ISO, mainly CRS and ability to easily extend CF-netCDF
- Entire KML RFC/SWG process took about 6 months - including
  - 30 day comment period
  - 60 day adoption vote.

Summary

- Our community has existing systems for serving its datasets internally
- Initial standardization efforts have been successful
- Even internally, work is needed (e.g., CF standards for observational datasets)
- Standards community is responding to our input in the specification of standard interfaces
- We need sound judgment in focusing our resources
- Continued and expanded collaboration is crucial
- There is a light at the end of the tunnel

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