CF-netCDF and CDM

Ethan Davis, John Caron,
Ben Domenico, Stefano Nativi*
UCAR Unidata
Univ of Florence*

OGC in MetOcean, Toulouse, France, November 2009

CF-netCDF and CDM

- CF-netCDF
- CDM/netCDF-java
- TDS
- Using CF-netCDF in OGC Standards
  - Mapping between CF-netCDF and GML
  - Status of CF-netCDF in OGC

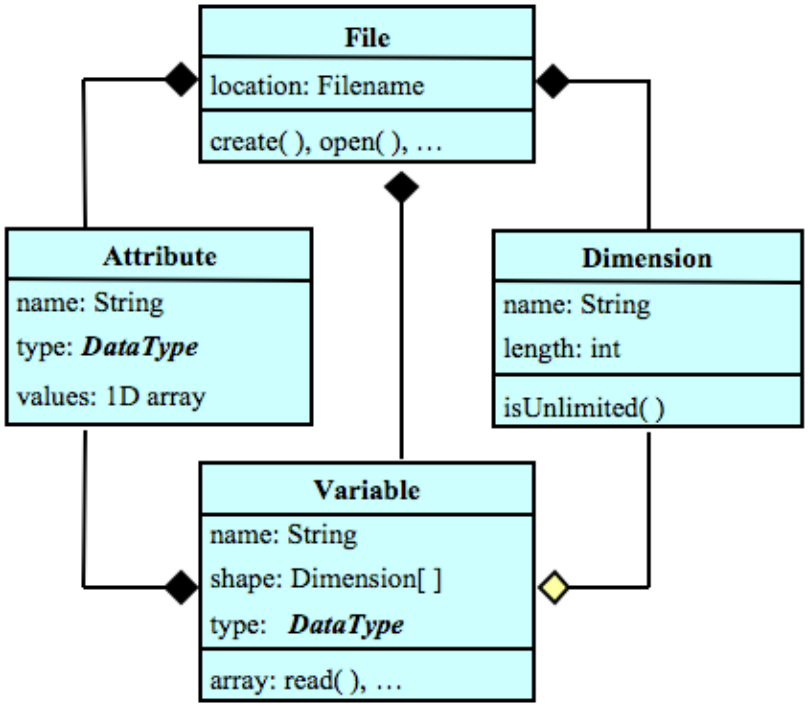
CF-netCDF

- Widely used encoding for gridded data with proposed additions for point, profile, trajectory, etc
  - Storage standard for many large projects
    - AMIP
    - CMIP5
    - IOOS
    - MERSEA
    - …
  - Read by many visualization/analysis packages
    - Ferret
    - GrADS
    - IDV
    - …

netCDF

- Libraries, data models, and formats for array-oriented scientific data.
  - Simple and general data model
  - Simple encoding / file format (for netCDF-3)

netCDF-3 Data Model



A file has named variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One dimension may be of unlimited length.

Variables and attributes have one of six primitive data types.

DataType
char
byte
short
int
float
double

netCDF-3 File Format

```
netcdf_file       = header  data
header            = magic  numrecs  dim_list  gatt_list  var_list
magic             = 'C'  'D'  'F'  VERSION
VERSION           = \x01 |                  // classic format
                    \x02                    // 64-bit offset format
numrecs           = NON_NEG | STREAMING      // length of record dimension
dim_list          = ABSENT | NC_DIMENSION  nelems  [dim ...]
gatt_list         = att_list                // global attributes
att_list          = ABSENT | NC_ATTRIBUTE  nelems  [attr ...]
var_list          = ABSENT | NC_VARIABLE   nelems  [var ...]
ABSENT            = ZERO  ZERO              // Means list is not present
ZERO              = \x00 \x00 \x00 \x00     // 32-bit zero
NC_DIMENSION = \x00 \x00 \x00 \x0A          // tag for list of dimensions
NC_VARIABLE   = \x00 \x00 \x00 \x0B         // tag for list of variables
NC_ATTRIBUTE = \x00 \x00 \x00 \x0C          // tag for list of attributes
nelems            = NON_NEG       // number of elements in following sequence
dim               = name  dim_length
name              = nelems  namestring
                        // Names a dimension, variable, or attribute.
                        // Names should match the regular expression
                        // ([a-zA-Z0-9_]|{MUTF8})([^\x00-\x1F/\x7F-\xFF]|{MUTF8})*
                        // For other constraints, see "Note on names", below.
```

...

...

Conventions for netCDF

- netCDF is intended to be self-describing
  - However, to fully understand the description, conventions must be agreed upon and used.
- Many netCDF attribute conventions have been developed over the years
  - CF, COARDS, GDT, CDC
  - ARGO
  - NUWG
  - …

CF Convention

- Widely used, stable governance
- With CF can describe
  - Coordinate systems
  - Standard names (for geophysical quantities)
  - Other
    - Projections (and more general CRS to some degree)
    - Dimensionless vertical coordinates (sigma, hybrid, etc)
    - Climatological statistics
- Has been mainly for gridded data
  - Proposed changes for point, station, profile, trajectory, etc
  - Possible future proposals: radar, image, swath, …

CF-netCDF Strengths

- Simple data model
- Dimensions are first-class objects, variables that share dimensions share coordinates
- Separation of data model and convention
- Not serialization of an applications internal data model

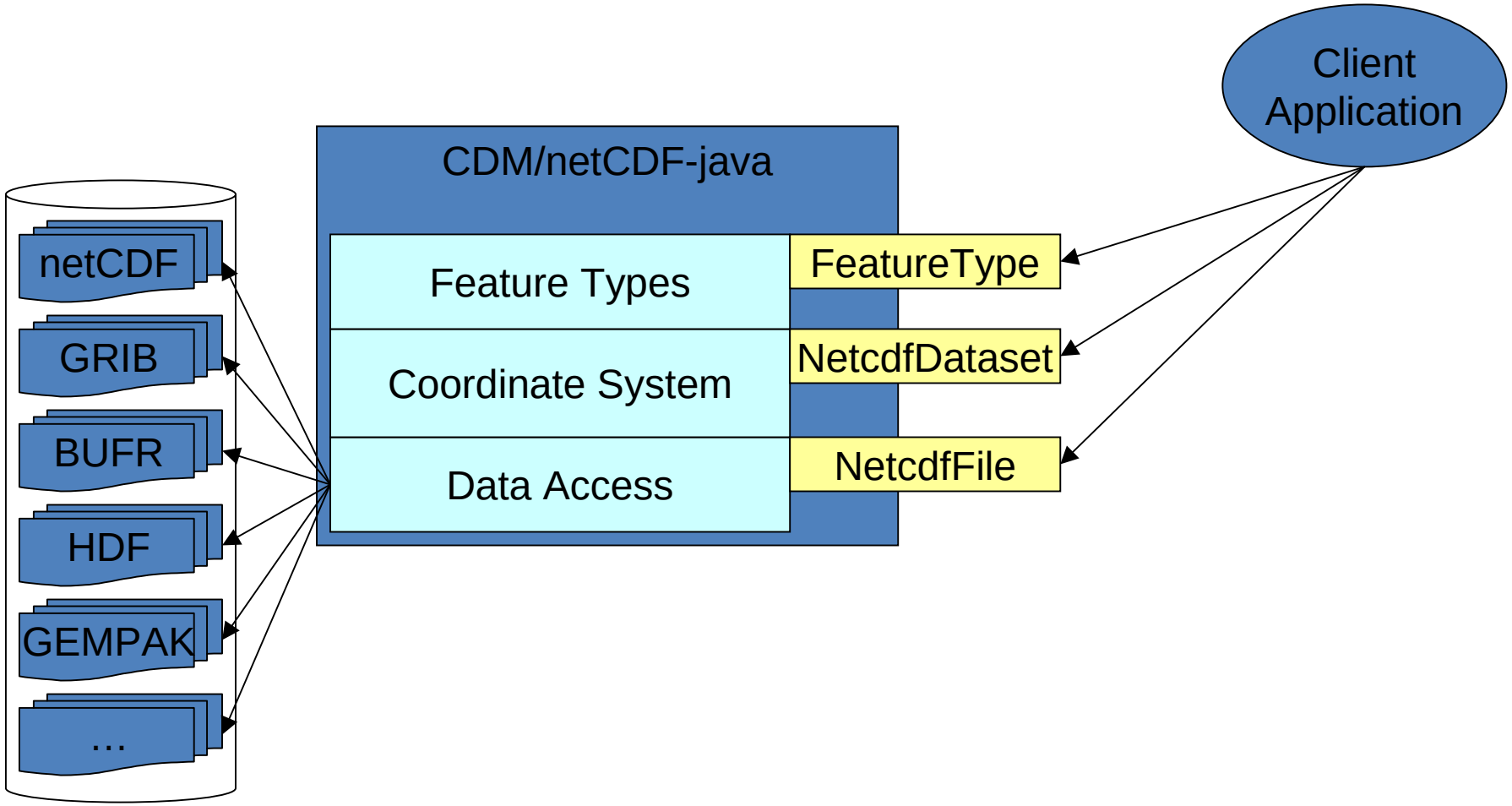
CDM/netCDF-java

- Data access layer
  - netCDF-3, netCDF-4
  - BUFR, GRIB-1, GRIB2
  - HDF, HDF-EOS, HDF5, HDF5-EOS
  - Radar: NEXRAD level II & III, DORADE, …
  - GEMPAK, McIDAS
  - OPeNDAP
  - …

CDM/netCDF-java

- Coordinate System layer
  - Recognizes a variety of conventions including
    - CF, COARDS
    - NUWG
    - HDF-EOS
    - …
- CDM Feature Types
  - Gridded data
  - Point Observation: point, station, profiles, trajectory, etc.
  - Radar
  - Image (not yet implemented, super-class of grid?)
  - Swath (not yet implemented)
  - Others ?

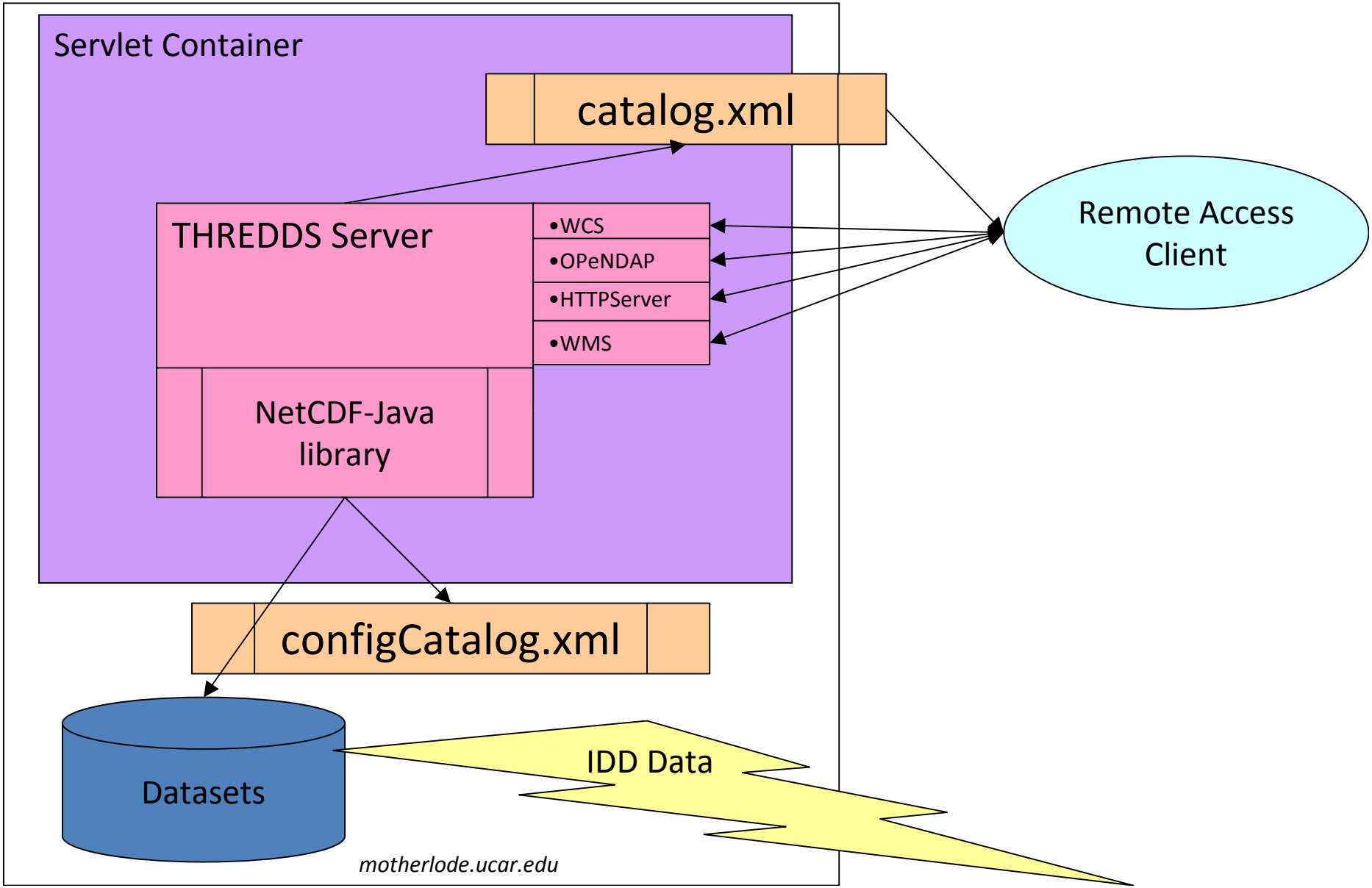
CDM/netCDF-java



THREDDS Data Server

- Built on the netCDF-java library:
  - Can read various data formats
  - Understands various conventions
  - Uses NcML to
    - Correct, modify, and add metadata
    - Construct virtual aggregated datasets
- Supports various data access services:
  - OPeNDAP
  - OGC WCS and WMS
  - Several experimental protocols
    - Collection and Feature Type subsetting
    - Streaming

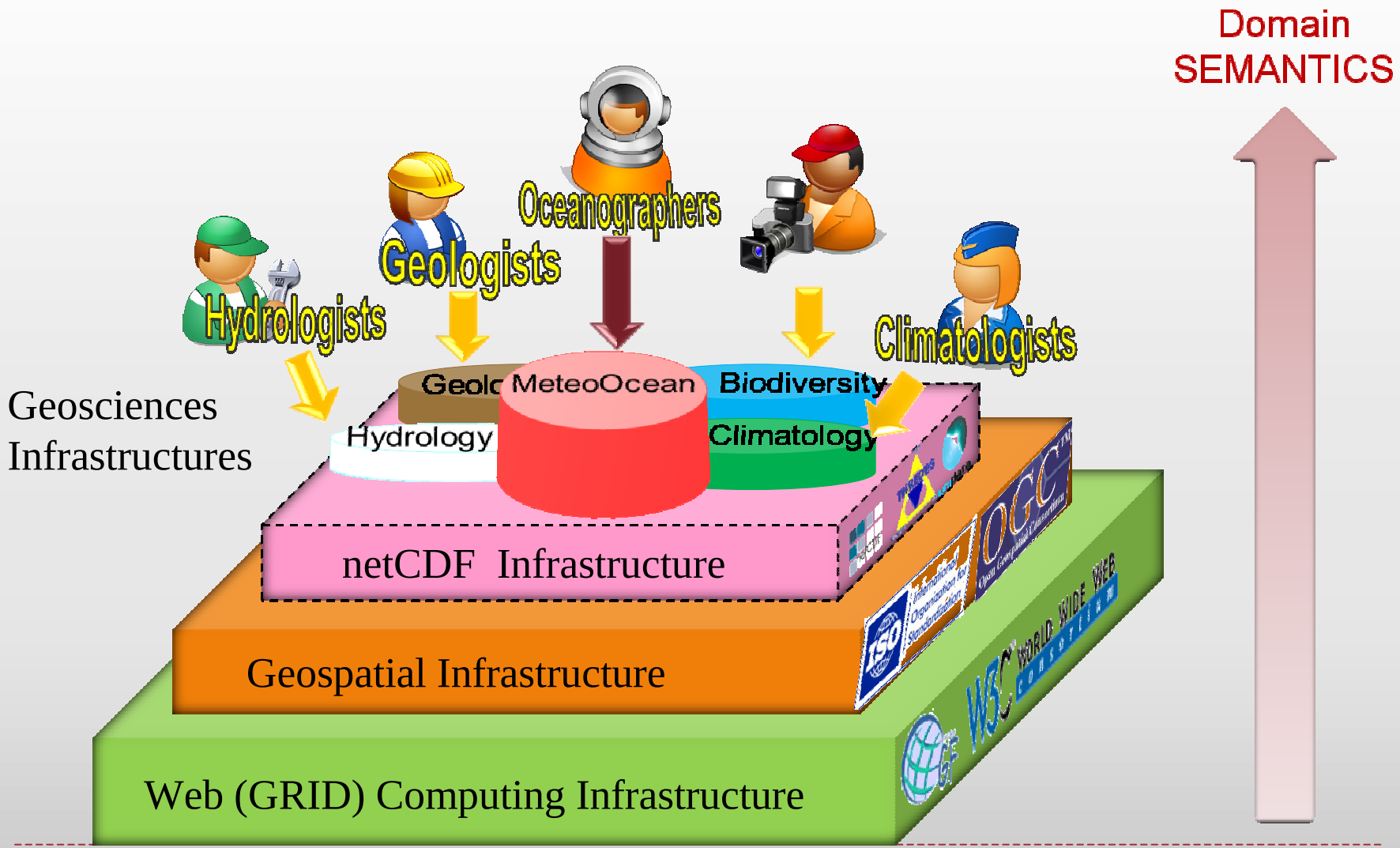
THREDDS Data Server



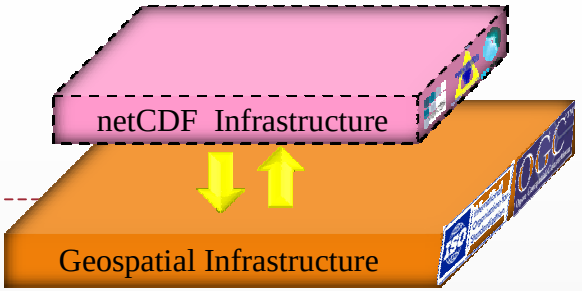
Using CF-netCDF in OGC Standards

- Mapping between CF-netCDF and GML
  - CSML
  - NcML-GML
  - ???
- CF-netCDF as an OGC Encoding Standard
  - Status

The Geosciences interoperability Infrastructure/resources



Standardizing the netCDF Infrastructure



Some on-going initiatives to make netCDF and Geospatial infrastructures/resources interoperable

Data Model

- netCDF-CF standardization (NASA, OGC)
- CDM Coverage/ Feature Types

Encoding language

- ncML-G+ (GML Commnity Pofile)

Interfaces

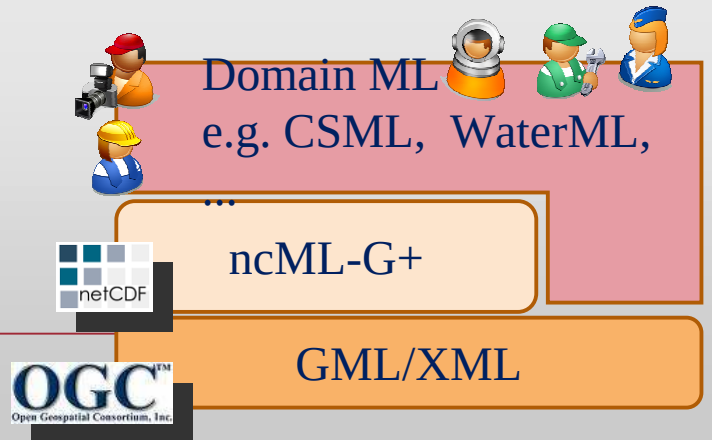
- CF-netCDF/ OPeNDAP extension for WCS
- Other OWS extension packages for CF-netCDF

Why NcML-G+

- To explicitly encode the full dataset complexity supported by CDM-netCDF
- Web Services (SOA and ROA) require XML-based encodings (i.e. infoset encodings)
  - A promising technology to underpin the most advanced functionalities (e.g. semantics)
- Need for a complete, well-defined, flexible, and relatively "low-level" (neutral) GML-based encoding.
  - The GML encoding of the netCDF/CDM data model seems to be a possible answer
- Bottom-up approach to complement the top-down approach followed by domain MLs

NcML-G+ possible benefits

- Standardization & Interoperability
  - it is a full GML profile;
  - it implements the netCDF data model generalizing and extending it applying the CDM;
- Data Model Harmonization and Multi-disciplinarity
  - it supports the encoding of different coverage & feature types (i.e. the CDM types);
  - it is neutral as far as application domain semantics is concerned;
  - it can be easily used by existing Community profiles (e.g. WaterML, GeoSciML, CSML, etc.)
    building specific semantics
  - it can be used in cross-domain applications;



CF-netCDF as OGC Standard

- Status of CF-netCDF encoding standard
  - Core: netCDF-3 file format
  - Extension: CF Convention
- Status of CF-netCDF as WCS extension

OGC CF-netCDF Standard

- Early December: vote on draft core CF-netCDF encoding standard
- Core specification based on NASA SPG standard for netCDF classic
  http://www.esdswg.org/spg/rfc/esds-rfc-011/ESDS-RFC-011v2.00.pdf
- Work underway on CF conventions standard
- Proposed to NASA SPG
- Targeted as first extension to OGC core CF-netCDF encoding standard.

Remaining Steps

- In early December, release candidate standard to the OGC Architecture Board for review.
  This review usually takes two weeks or less.
- Assuming a positive OAB review, the SWG votes to release the document for a 30 day public comment period.
- SWG works with OGC staff to prepare a press release.
- After 30 days, collate the comments (if any) and work the comments and edit the document
- SWG recommends a Technical Committee adoption vote.
- Adoption vote happens. Requires 60 days.

Extension for CF Conventions

- CF standard recently proposed to NASA
- http://www.esdswg.org/spg/rfc/esds-rfc-021
- Plan to base OGC CF extension on what has been submitted to NASA
- NASA submission takes middle ground between:
  - Spec that is simply a pointer to the CF docs
  - Spec that is simply a snapshot of CF docs

Relationship to WCS CF-netCDF

- Initial effort to establish CF-netCDF as a binary encoding spec is continuing
- WCS encoding spec includes specifics of WCS client server interactions as well as the CF-netCDF encoding
- Ultimately the goal is to harmonize the two initiatives so the WCS encoding points to the independent CF-netCDF spec.
- Other protocols may also cite the CF-netCDF encoding spec

Summary

- CF-netCDF encoding format
  - Simple data model and file format
  - Widely used
- TDS and CDM
  - Building upon existing systems and data
  - Bridging to developing standards
- Using CF-netCDF in OGC Standards