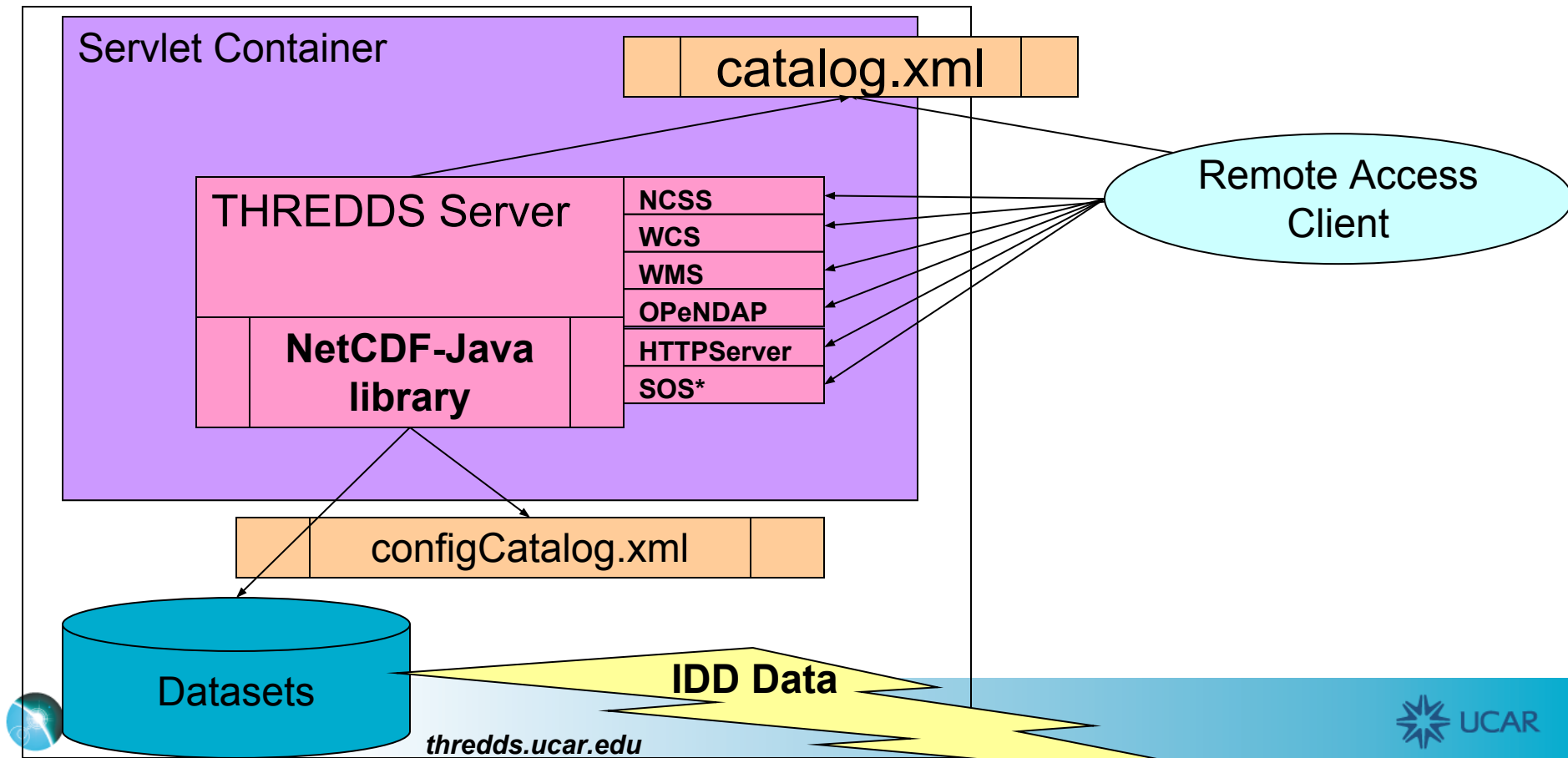


GRIB to NetCDF/CF as part of Unidata's THREDDS project

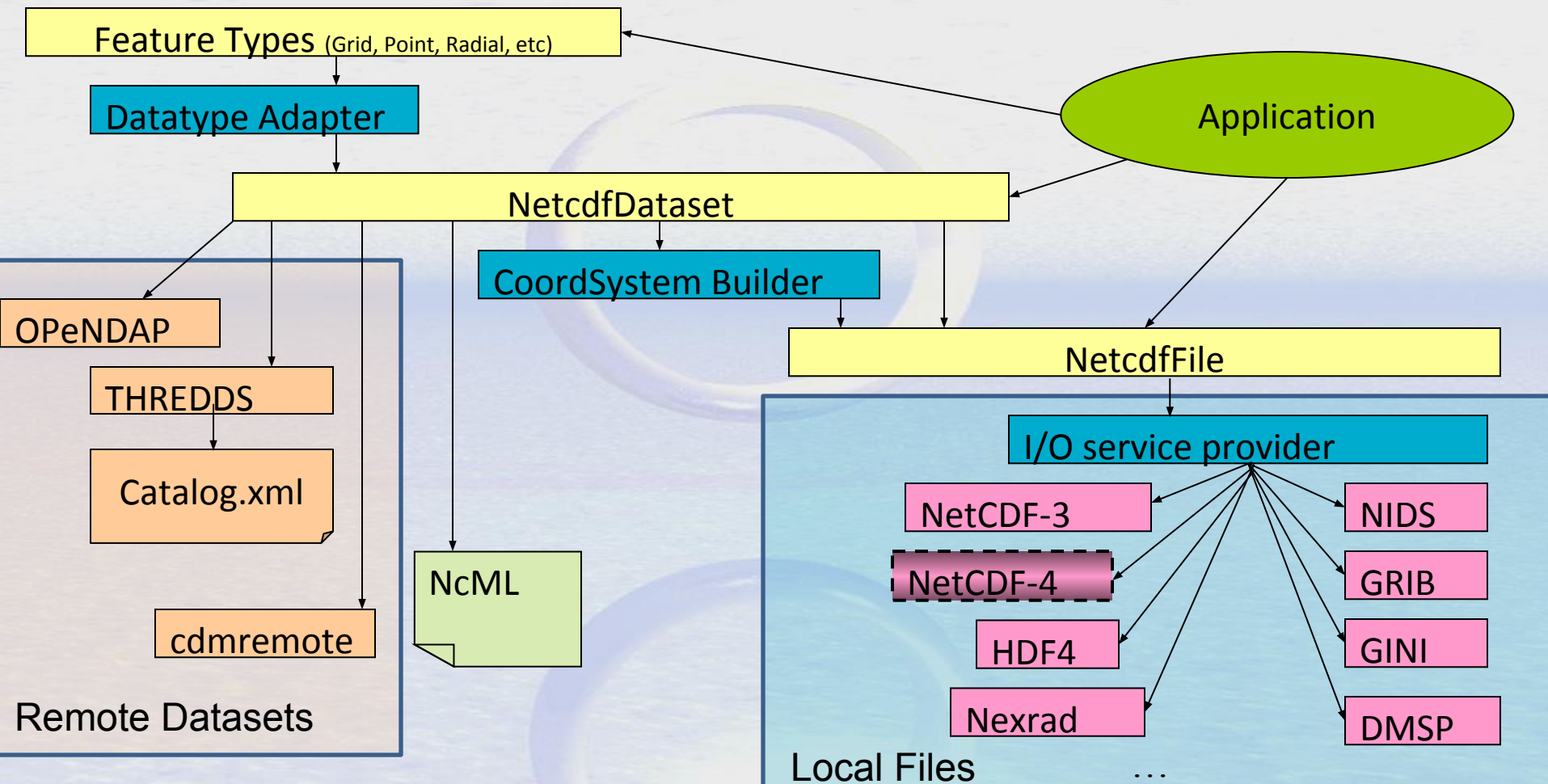
John Caron, UCAR/Unidata

Sep 24, 2014

THREDDS Data Server



NetCDF-Java library



Dataset: /thredds/ncss/grib/NCEP/GFS/Global_0p5deg/Best (Gridded Dataset Description)

Base Time: 2014-08-31T00:00:00Z

Select Variable(s):

Variables with Time coordinate time

- Convective_precipitation_surface_Mixed_intervals_Accumulation = Convective precipitation (Mixed_intervals Accumulation) @ Ground or water surface
- Total_precipitation_surface_Mixed_intervals_Accumulation = Total precipitation (Mixed_intervals Accumulation) @ Ground or water surface

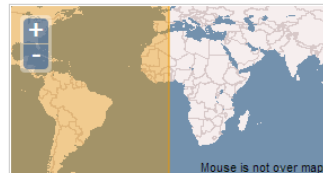
Variables with Time coordinate time1

- Potential_Evaporation_Rate_surface = Potential Evaporation Rate @ Ground or water surface
- Pressure_convective_cloud_bottom = Pressure @ Convective cloud bottom level
- Pressure_convective_cloud_top = Pressure @ Convective cloud top level

Variables with Time coordinate time2

- Best_4_layer_Lifted_Index_surface = Best (4 layer) Lifted Index @ Ground or water surface
- Cloud_water_entire_atmosphere = Cloud water @ Entire atmosphere layer
- Convective_available_potential_energy_surface = Convective available potential energy @ Ground or water surface
- Convective_inhibition_surface = Convective inhibition @ Ground or water surface
- Field_Capacity_surface = Field Capacity @ Ground or water surface
- Geopotential_height_highest_tropospheric_freezing = Geopotential height @ Highest tropospheric freezing level
- Geopotential_height_maximum_wind = Geopotential height @ Maximum wind level
- Geopotential_height_surface = Geopotential height @ Ground or water surface
- Geopotential_height_tropopause = Geopotential height @ Tropopause
- Geopotential_height_zeroDegC_isotherm = Geopotential height @ Level of 0°C isotherm
- Haines_Index_surface = Haines Index @ Ground or water surface
- ICAO_Standard_Atmosphere_Reference_Height_maximum_wind = ICAO Standard Atmosphere Reference Height @ Maximum wind level
- ICAO_Standard_Atmosphere_Reference_Height_tropopause = ICAO Standard Atmosphere Reference Height @ Tropopause
- Ice_cover_surface = Ice cover @ Ground or water surface
- Land_cover_0_sea_1_land_surface = Land cover (0 = sea, 1 = land) @ Ground or water surface
- MSLP_Eta_model_reduction_msl = MSLP (Eta model reduction) @ Mean sea level
- Planetary_Boundary_Layer_Height_surface = Planetary Boundary Layer Height @ Ground or water surface
- Precipitable_water_entire_atmosphere = Precipitable water @ Entire atmosphere layer
- Pressure_maximum_wind = Pressure @ Maximum wind level
- Pressure_reduced_to_MSL_msl = Pressure reduced to MSL @ Mean sea level
- Pressure_surface = Pressure @ Ground or water surface
- Pressure_tropopause = Pressure @ Tropopause
- Relative_humidity_entire_atmosphere = Relative humidity @ Entire atmosphere layer
- Relative_humidity_highest_tropospheric_freezing = Relative humidity @ Highest tropospheric freezing level
- Relative_humidity_zeroDegC_isotherm = Relative humidity @ Level of 0°C isotherm
- Sunshine_Duration_surface = Sunshine Duration @ Ground or water surface
- Surface_Lifted_Index_surface = Surface Lifted Index @ Ground or water surface

Choose Spatial Subset:



Lat/Lon subset Coordinate subset

Bounding Box (decimal degrees):

north

 west east

 south

- Disable horizontal subsetting
[reset to full extension](#)

Horizontal Stride:

Choose Time Subset:

Time range Single time

Starting:
 Ending:
 Stride:
[reset to full extension](#)

Choose Vertical Level:

Single Level Vertical Stride

Level:

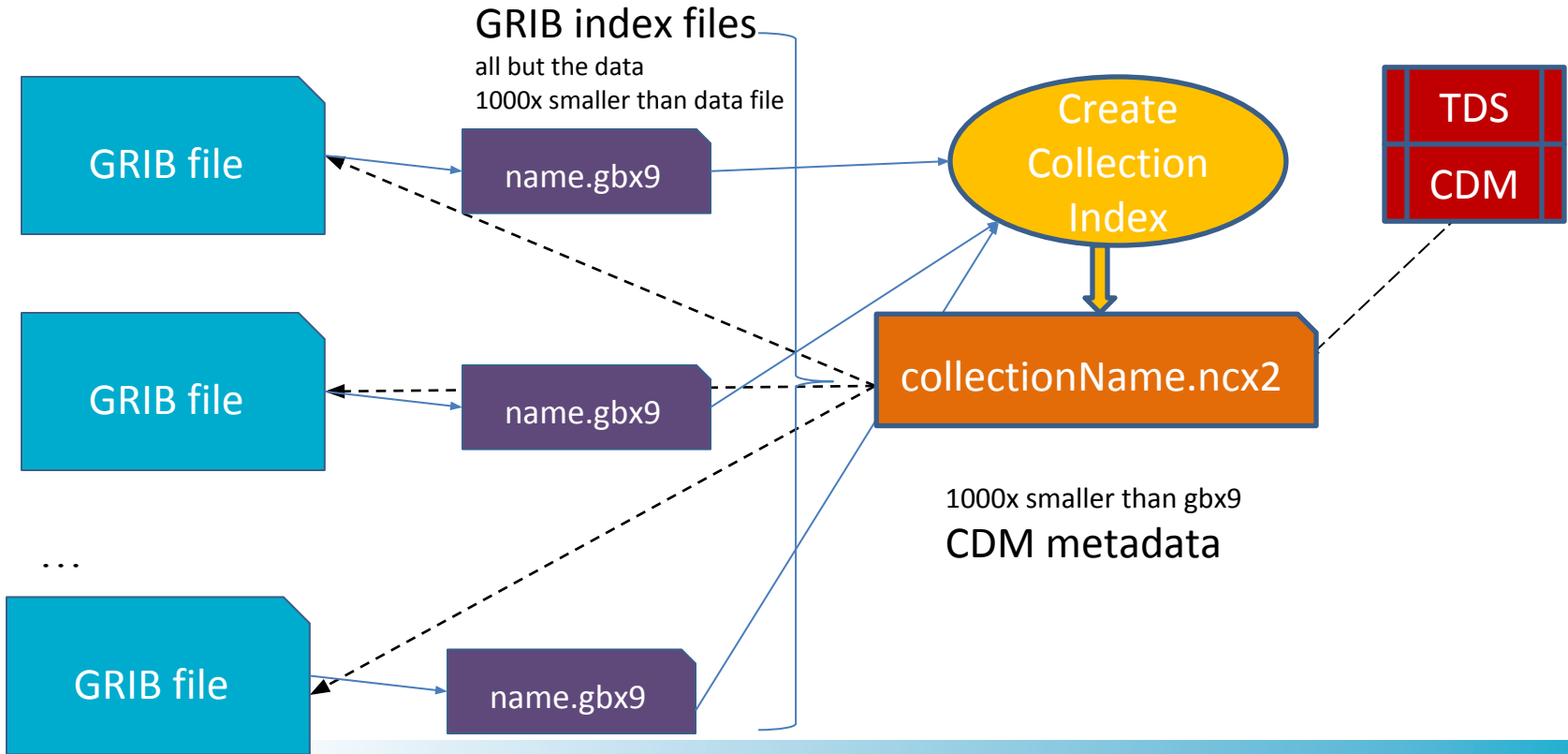
Add 2D Lat/Lon to file (if needed for CF compliance)

- Add Lat/Lon variables

Choose Output Format:

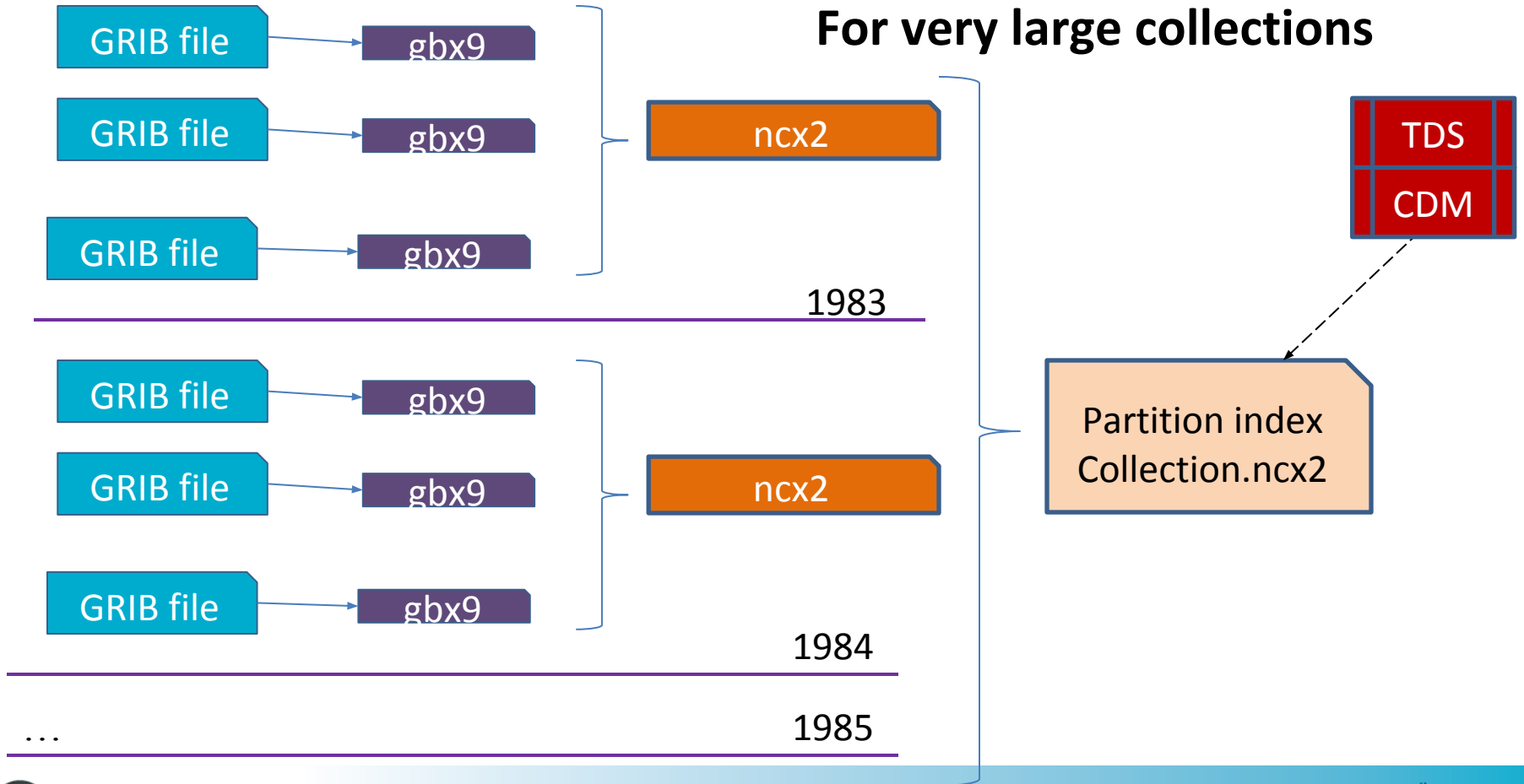
Format:

GRIB collection indexing



GRIB time partitioning (nested hierarchies)

For very large collections



GRIB -> NetCDF Issues

1. External Tables
2. What is a Variable (and what is its name)?
3. CF Semantics
4. File Sizes (tomorrow)

External Tables

- Official WMO tables are still in Word/PDF
 - ◆ not machine parseable
 - ◆ “Official” tables may have mistakes
 - ◆ OTOH GRIB-2 now also published in XML (Yay!)
- Local tables in wide use
 - ◆ No canonical format or place to find them (BAD)
 - ◆ Many centers do not correctly version their tables (very BAD)
 - ◆ Many centers override WMO tables, esp for GRIB-1 (disaster)
- No foolproof way to know correct table
 - ◆ “Expert-only” format; must know who wrote it
 - ◆ GRIB is not a reliable long term archive until problem is fixed
 - ◆ Proposed solution: web registry, embed table id in GRIB record

What is a Variable?

aka Field, Parameter, Dataset, ...

NetCDF :

- a container for a multidimensional array of data
- same data type, same attributes(units, description...)
- has a unique name, typically “human readable”
- arbitrary metadata to clarify meaning

NetCDF /CF:

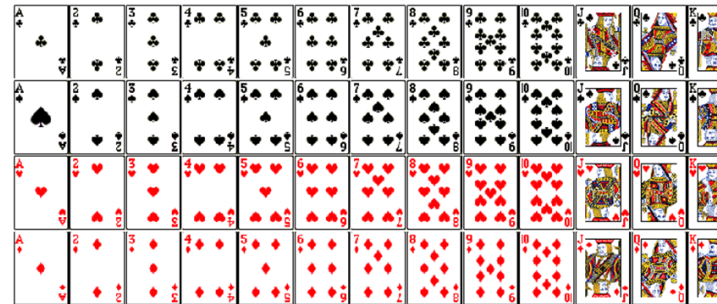
`float windSpeed(reftime, time, vertical, lat, lon);`

- non-spacetime dimensions: ensemble, wavelength, vector component, ...

How to make Variables from collection of GRIB2 records?



- Each 2D slice is stored independently in a GRIB record
- Imagine each GRIB record as one row in a database
- GRIB-2 has ~30 PDS templates, each with 10-20 attributes
- Which of these attributes should be used for “variable key” ?
- CDM currently uses :
 - PDS Template
 - Parameter Discipline, Category, Number
 - if local tables are used, the Center and Subcenter ids
 - the Level Type; if its a vertical layer
 - if its a time interval, the Statistical Process Type (Code table 4.10)
 - if its a probability, the Probability Type (Code table 4.9)
 - the Derived forecast Type (Code table 4.7)
 - if the generating process type is 6 or 7 (error)
 - the GDS hashcode
 - Allow user to override (expert level)
- “Dataset schema” not able to be encoded in GRIB



Variable Naming

CDM Variable names = parameter name X level name [X layer] [X statistic] [X error] ...

VAR_0-0-10_L1_Imixed_S0	Latent heat net flux (Mixed_intervals Average) @ Ground or water surface	64,361,720	time1,lat,lon
VAR_0-0-11_L1_Imixed_S0	Sensible heat net flux (Mixed_intervals Average) @ Ground or water surfa...	64,361,720	time1,lat,lon
VAR_0-1-0_L103	Specific humidity @ Specified height level above ground	65,1,361,720	time,height_above_grou
VAR_0-1-0_L108_layer	Specific humidity @ Level at specified pressure difference from ground to...	65,1,361,720	time,pressure_differenc
VAR_0-1-1_L4	Relative humidity @ Level of 0°C isotherm	65,361,720	time,lat,lon
VAR_0-1-1_L100	Relative humidity @ Isobaric surface	65,25,361,720	time,pressure3,lat,lon
VAR_0-1-1_L103	Relative humidity @ Specified height level above ground	65,1,361,720	time,height_above_grou
VAR_0-1-1_L104_layer	Relative humidity @ Sigma level layer	65,4,361,720	time,sigma_layer,lat,lon
VAR_0-1-1_L104	Relative humidity @ Sigma level	65,1,361,720	time,sigma,lat,lon
VAR_0-1-1_L108_layer	Relative humidity @ Level at specified pressure difference from ground t...	65,1,361,720	time,pressure_differenc
VAR_0-1-1_L200	Relative humidity @ Entire atmosphere layer	65,361,720	time,lat,lon
VAR_0-1-1_L204	Relative humidity @ Highest tropospheric freezing level	65,361,720	time,lat,lon
VAR_0-1-3_L200	Precipitable water @ Entire atmosphere layer	65,361,720	time,lat,lon
VAR_0-1-7_L1_Imixed_S0	Precipitation rate (Mixed_intervals Average) @ Ground or water surface	64,361,720	time1,lat,lon
VAR_0-1-8_L1_Imixed_S1	Total precipitation (Mixed_intervals Accumulation) @ Ground or water sur...	65,361,720	time2,lat,lon
VAR_0-1-10_L1_Imixed_S1	Convective precipitation (Mixed_intervals Accumulation) @ Ground or wat...	65,361,720	time2,lat,lon
VAR_0-1-13_L1	Water equivalent of accumulated snow depth @ Ground or water surface	65,361,720	time,lat,lon
VAR_0-1-22_L100	Cloud mixing ratio @ Isobaric surface	65,21,361,720	time,pressure1,lat,lon

1. NetCDF Variable names need to be unique, short and stable (but not GRIB parameter names)
2. Some centers have a “Short name” in the parameter table - wrong place for it

How to describe dataset schema?

NetCDF has CDL and NcML, how to do this in GRIB?

- Should be intuitive to scientists
- Must be unambiguous, ie machine parseable
- Define which attributes in the GRIB records are used for Variable id
- Create a unique name
- You will either do this in code or externalize it to a table (better)

Name	Parameter	Vertical	value	Time interval	Statistic	Error	GenProcess
RH_at_2m	0-1-1	103	2				
RH_isobaric_levels	0-1-1	100	*				
Precipitation_rate_3hr_average	0-1-7	1		3 hour	Average		
Precipitation_rate_3hr_average_error	0-1-7	1		3 hour	Average	Yes	
Cloud_mixing_ratio_from_TIGGE	0-1-22	100	*				144
Cloud_mixing_ratio_from_ECMWF	0-1-22	100	*				145

CF semantics

- GRIB has created many important shared semantics
 - ◆ tables, standard names, etc
- These must be mapped to CF semantics by domain experts
- CDM has done some “easy” parts
 - ◆ coordinate systems, projections, time coordinates

GRIB2 <--> NetCDF/CF

In general, not lossless in either direction

- no place to store arbitrary attributes in GRIB
- each GRIB record could have different metadata, but you only can store attributes at file or variable

But if you work hard enough you could make
GRIB-> netCDF->GRIB lossless

- for your (ECMWF) data
- not all metadata would be CF

Manual on Codes, Volume 1.2

Code table 5 – Time range indicator

- complex averaging schemes

Code table 6 - Data representation type (projections)

- common ones are in CF
- missing spherical harmonic coefficients

Parameters <-> CF Standard Names

- endless amount of work to clarify semantics
- but probably can get 80/20

Conclusions

ECMWF can ignore GRIB table problem

- continue to make your GRIB data work correctly with your software

ECMWF can solve “what is a variable”

- in code or help develop a GRIB schema definition

ECMWF can make GRIB -> netCDF -> GRIB lossless

- encode GRIB semantics into (non-CF) metadata where needed
- CDM already does GRIB -> netCDF (lossy), ECMWF can improve

Hard work to get all (most?) of GRIB into CF

- GRIB parameters -> standard names