Mid-course Progress and Opportunities for Unidata Outreach

Ben Domenico

May 2011

Outline

- Overview
- Relationship to Unidata 2013 Proposal
- Collaborating Communities
- Recent Progress
- Current Focus
- Future Directions

Overview

- Focus near the boundary of our core community
- Make our data available more broadly and gain access to other data source in useful forms
- Implement/support up to standard interface
- Do NOT support other communities' tools
- Do NOT provide domain expertise
- Employ bottom-up approach to standards
- Get others to do most of the work

Unidata 2013 Review Panel Question

Question: Is the UPC prepared to provide the same quality of support to the newly engaged communities as it provides to its current constituents?

Response: While the support for all users will remain at a very high level, that does not mean it will be exactly the same.

- Data access with supported analysis and display tools for our community
- Data access via standard interfaces to enable other communities to use our data with their own tools (arcGIS, IDL, Matlab, etc.)

Excerpts from Review Panel Report

The UPC could play a significant leadership role within committees and consortiums like OGC seeking to address the need to develop standards and technologies for data discovery

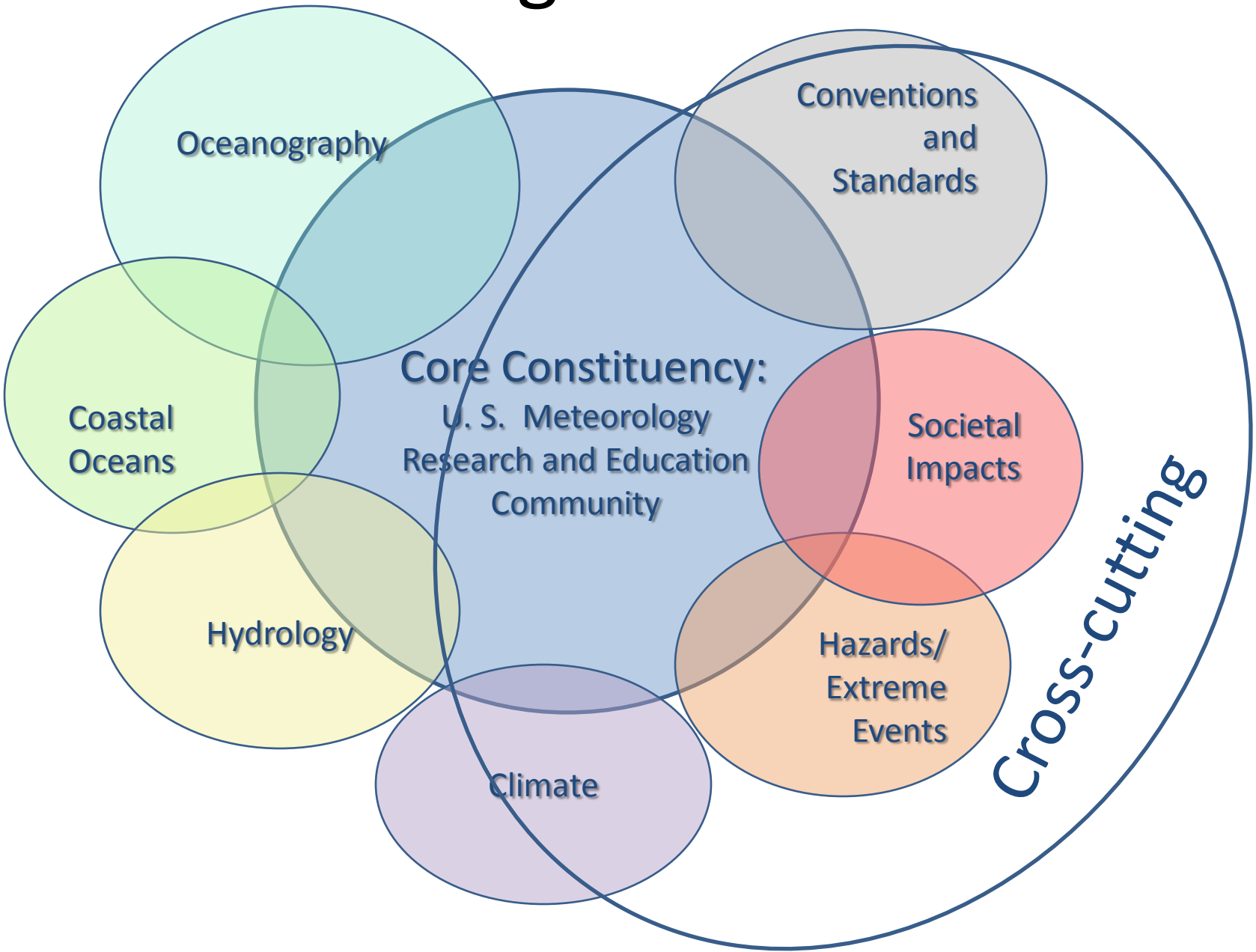
Unidata leadership and advocacy in this area could facilitate expanded utilization of Unidata information resources for other research areas like climate and provide Unidata users with easier access to other data sources like NASA satellite information

Unidata 2013 Thematic Areas

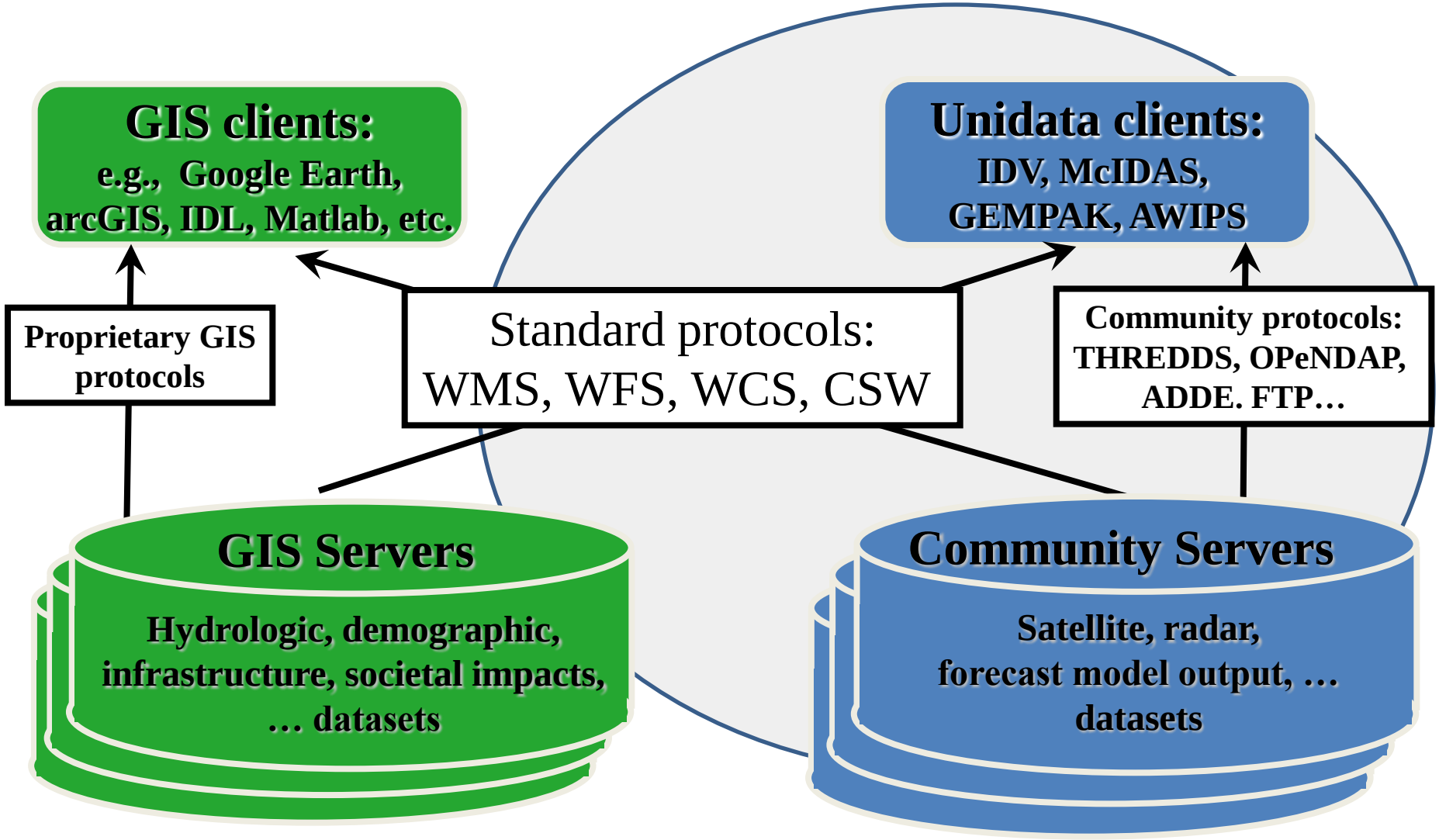
1. *Broadening participation and expanding community services*
2. *Advancing data services*
3. Developing and deploying useful tools
4. Enhancing user support services
5. *Providing leadership in cyberinfrastructure*
6. *Promoting diversity by expanding opportunities*

Outreach initiatives contribute directly to thematic areas in red

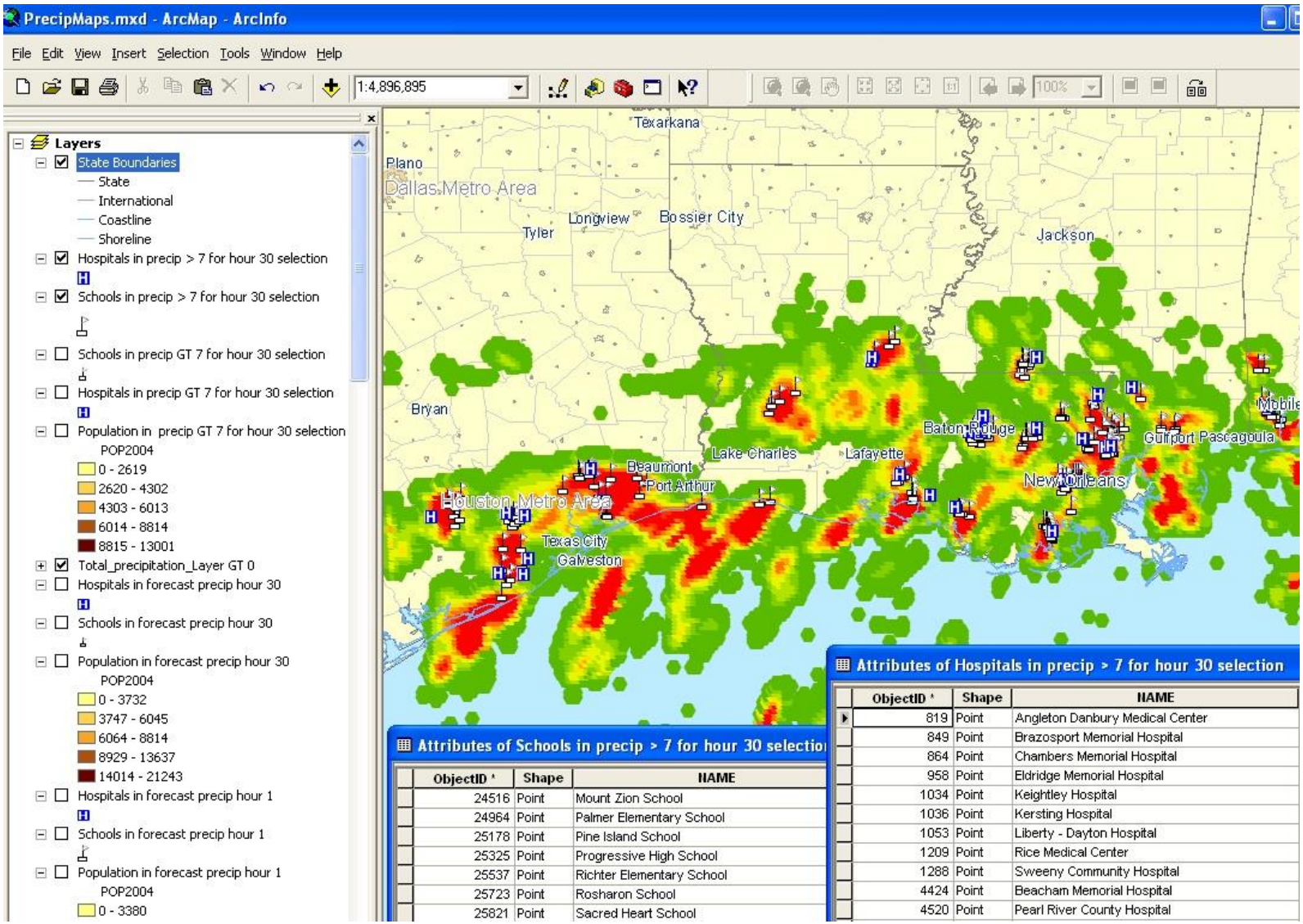
Collaborating Communities



Standard Interface Web Services for Data System Interoperability



GIS Tools Applied to Atmospheric Science Data



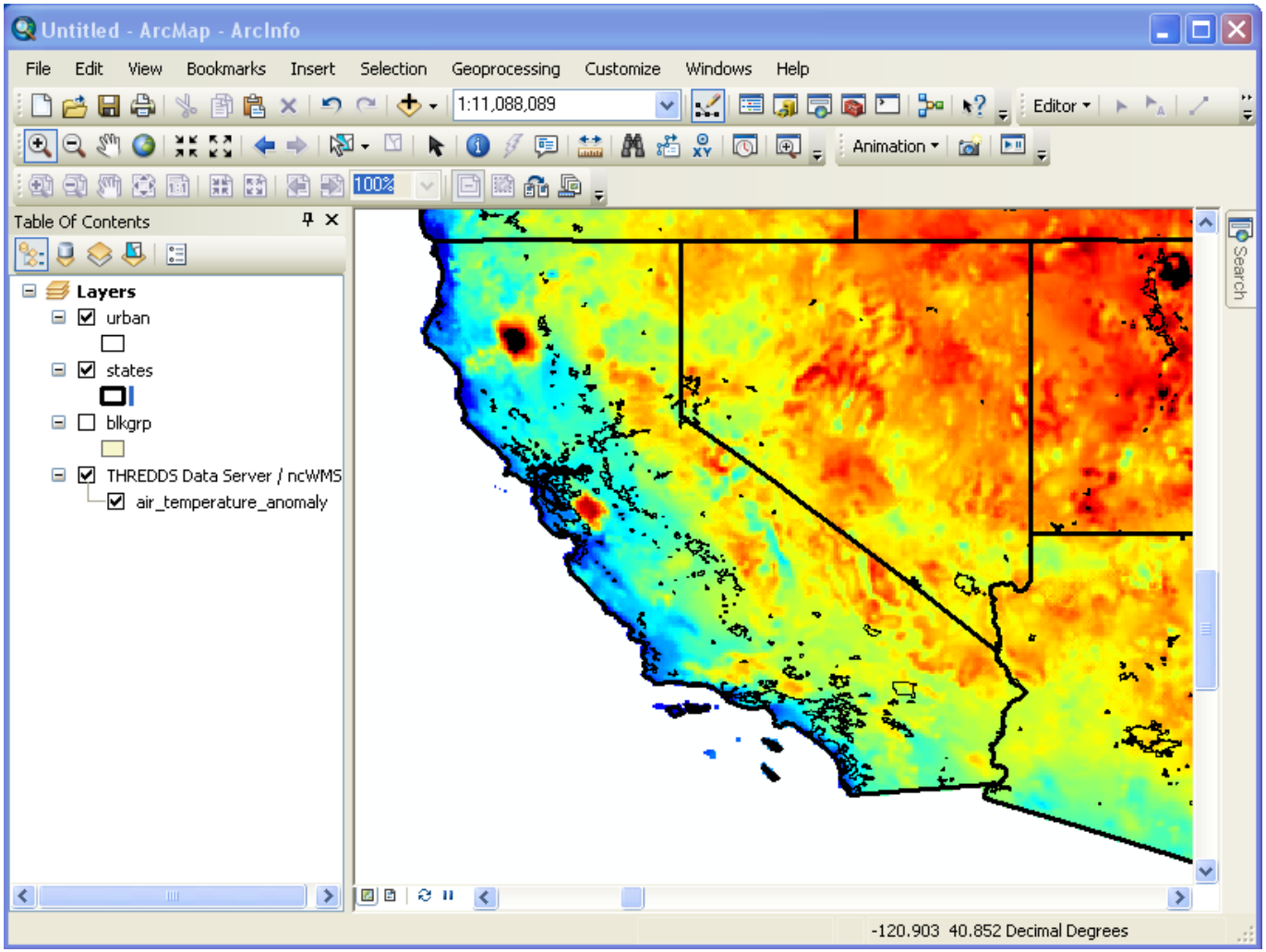
A Few of the Collaborating Organizations

- NOAA (e.g., PMEL, PFEL, NCDC, NGDC, IOOS?)
- NASA
- USGS
- OGC (Open Geospatial Consortium)
- CF (Climate and Forecast) Community
- CUAHSI (Hydrology)
- NCAR (GIS Program, CISL, …)
- NEON (National Ecological Observatory Network) Data Services
- International (Italian National Research Council, British Atmospheric Data Center, University of Reading E-science Center, Meteo France, Australian Bureau of Meteorology, …)
- Industry (ESRI, ITTvis, Applied Science Associates, …)

Collaboration Illustration: ArcMAP Display of NCAR Data from THREDDS Web Map Server

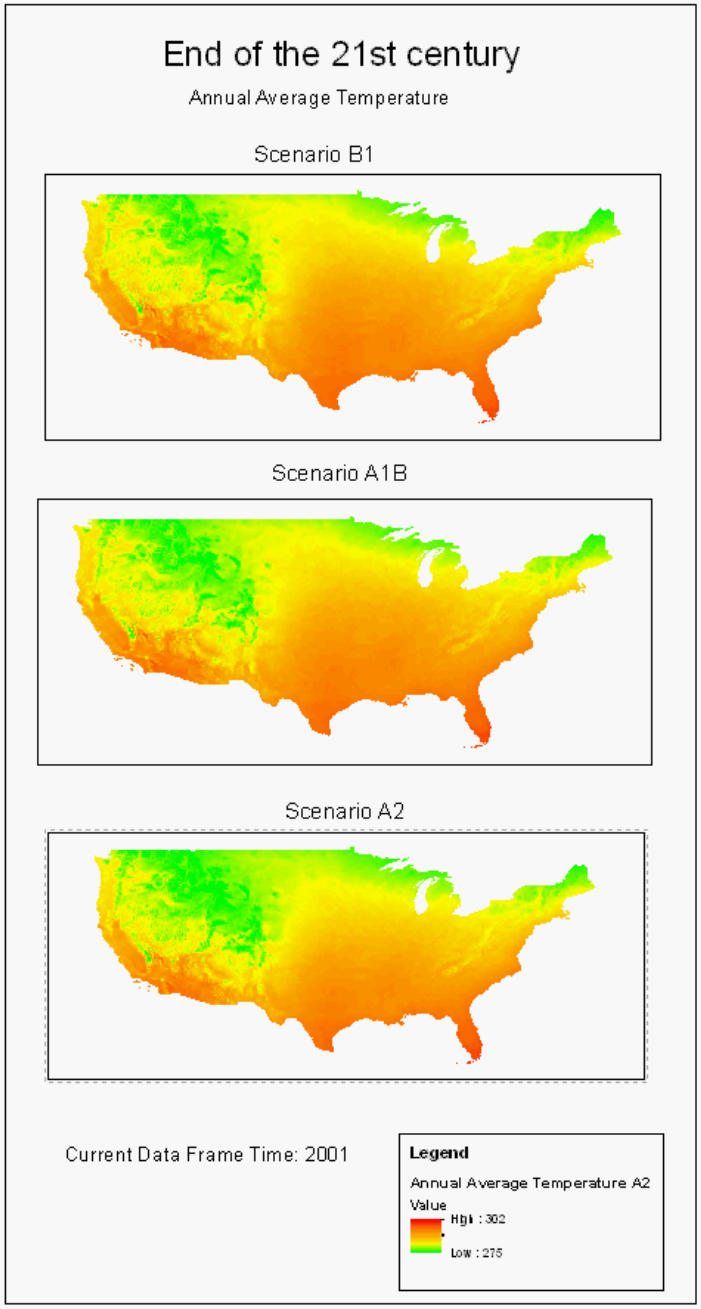
NCAR model air temperature anomaly display

- Data from NCAR climate model
- TDS software from Unidata
- WMS software from University of Reading
- Server run by NCAR GIS program
- GIS map and analysis software from ESRI



“Collateral Successes”

- Annual average temperature directly read from netCDF file into ESRI arcMAP
- Images can be accessed via WMS
- Data can be accessed via WCS
- Temporal animation now much more convenient in new versions of ESRI tools
- Similar to earlier successes with IDL, Matlab, Ferret, GRaDs and many other analysis and display tools.

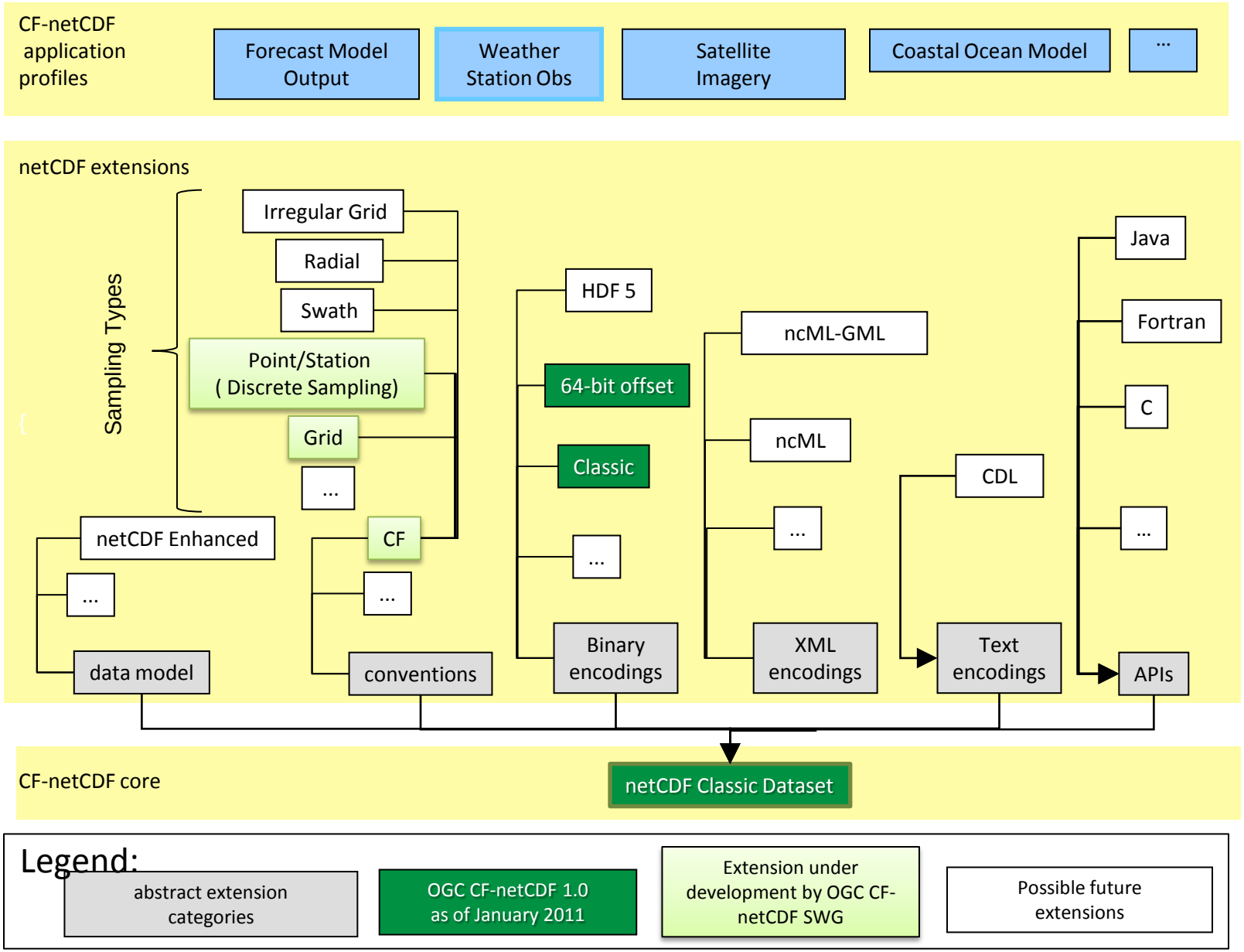


Recent Progress

- netCDF adopted as (first) OGC binary encoding standard
- netCDF read/write incorporated direction into ESRI arcGIS
- Draft OGC spec for CF extension to netCDF core with Italian National Research Council (CNR)
- Draft OGC spec for CF-netCDF extension to WCS with CNR
- Working on extension spec for netCDF enhanced (netCDF4) data model
- “Crossing the Digital Divide” data discovery experiment with CUAHSI and CNR
- User community incorporating GI-cat catalog metadata harvesting for THREDDS Data Servers (Rich Signell created a configuration video.)
- GEOSS arctic climate and weather demonstration with USGS, NCAR GIS, GMU, Michigan Tech
- Harmonization of CSML and CDM scientific feature types with BADC and CF community
- NCPP (NOAA Climate Projection Pilot) with NOAA, USGS, NCAR GIS, Michigan Tech)
- WMO community joins OGC and MetOceans working group formed
- OPeNDAP considering joining OGC
- Discussed plans with CF community at recent GO-ESSP meetings
- HDF5 is part of OGC Web Services phase 8 testbed
- Hosting OGC Technical Committee and GEOSS Climate Workshop in Sept.
- Participating in University Consortium for GIS and CUAHSI GIS in June
- Participating in International Coastal Oceans Network 5 Workshop in August
- Invited to international workshop on Networking of Air Quality Data Systems

Status of CF-netCDF OGC Standards Development

Stefano Nativi, Italian National Research Council, Institute of Methodologies for Environmental Analysis, Prato, Italy
Ben Domenico, Unidata Program Center, University Corporation for Atmospheric Research, Boulder, USA



Current Focus in OGC

- CF extension to OGC netCDF core spec
- CF-netCDF extension to WCS core spec
- Enhanced data model (netCDF4) extension to netCDF core spec
- CSML and CDM/CF Feature Types
- Hosting September OGC Technical Committee Meetings

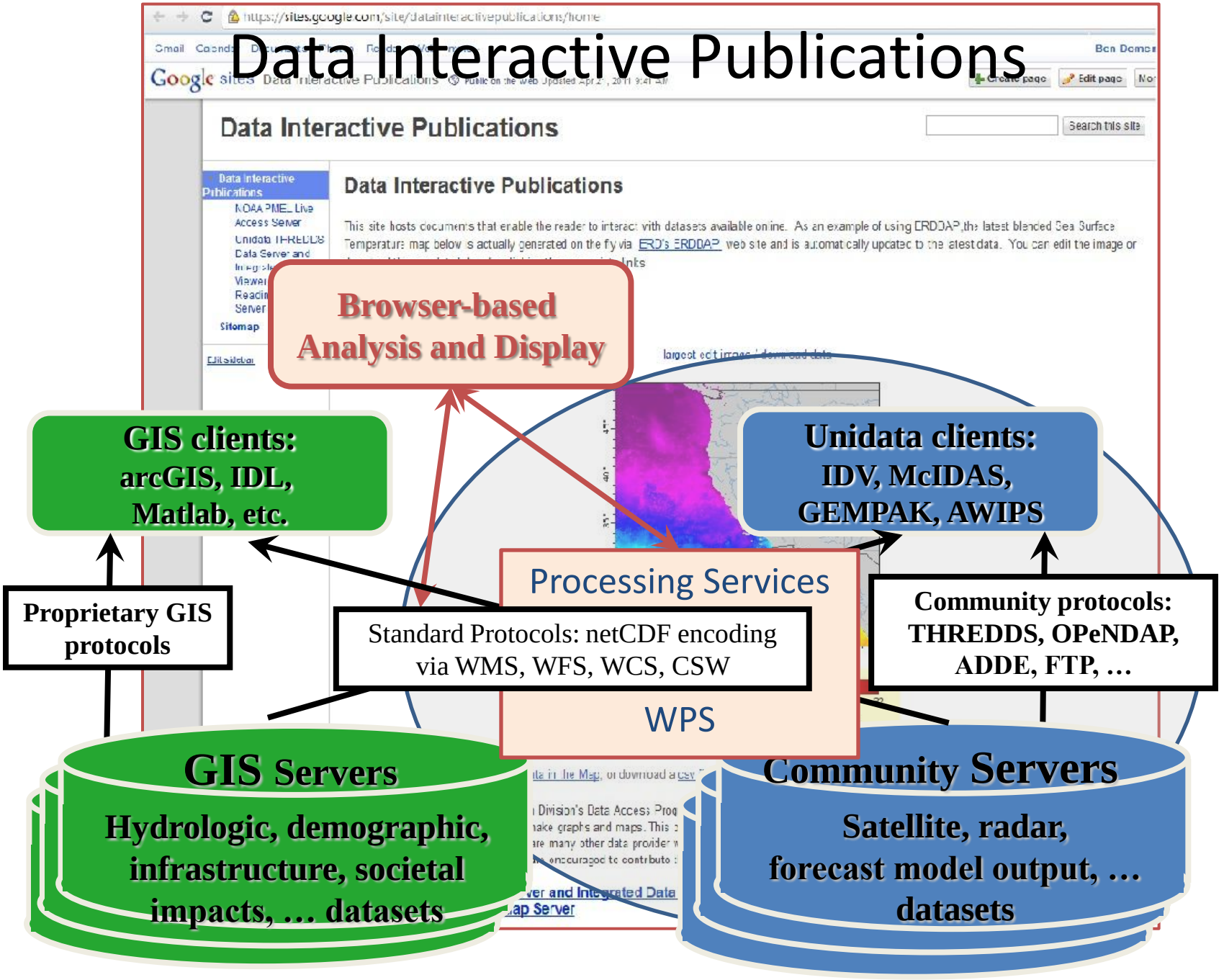
So What?

- Activities foster awareness and adoption of Unidata infrastructure
- More data systems compatible with Unidata tools and services
- More data from different, organizations, disciplines, and regions discoverable – with documentation and in usable forms
- Some organizations, e.g., in European Union, required by law to use standard-compliant software
- Valuable international collaborations on development of tools and provision of services
- New opportunities and work paradigms possible

New Possibility: Data Interactive Publications

- Publications are the valued “exchange medium” of the academic community
- Data citation (for transparency, reproducibility, etc.) is an emerging issue for scientific publications
- Data access and analysis are the basis of nearly all those publications
- In the present technological environment, can we bridge the gap between the publications and the data analysis?
- Having data and processing on same system is efficient.
- Will this work in an era of server side (cloud?) computing and mobile client devices?

Data Interactive Publications



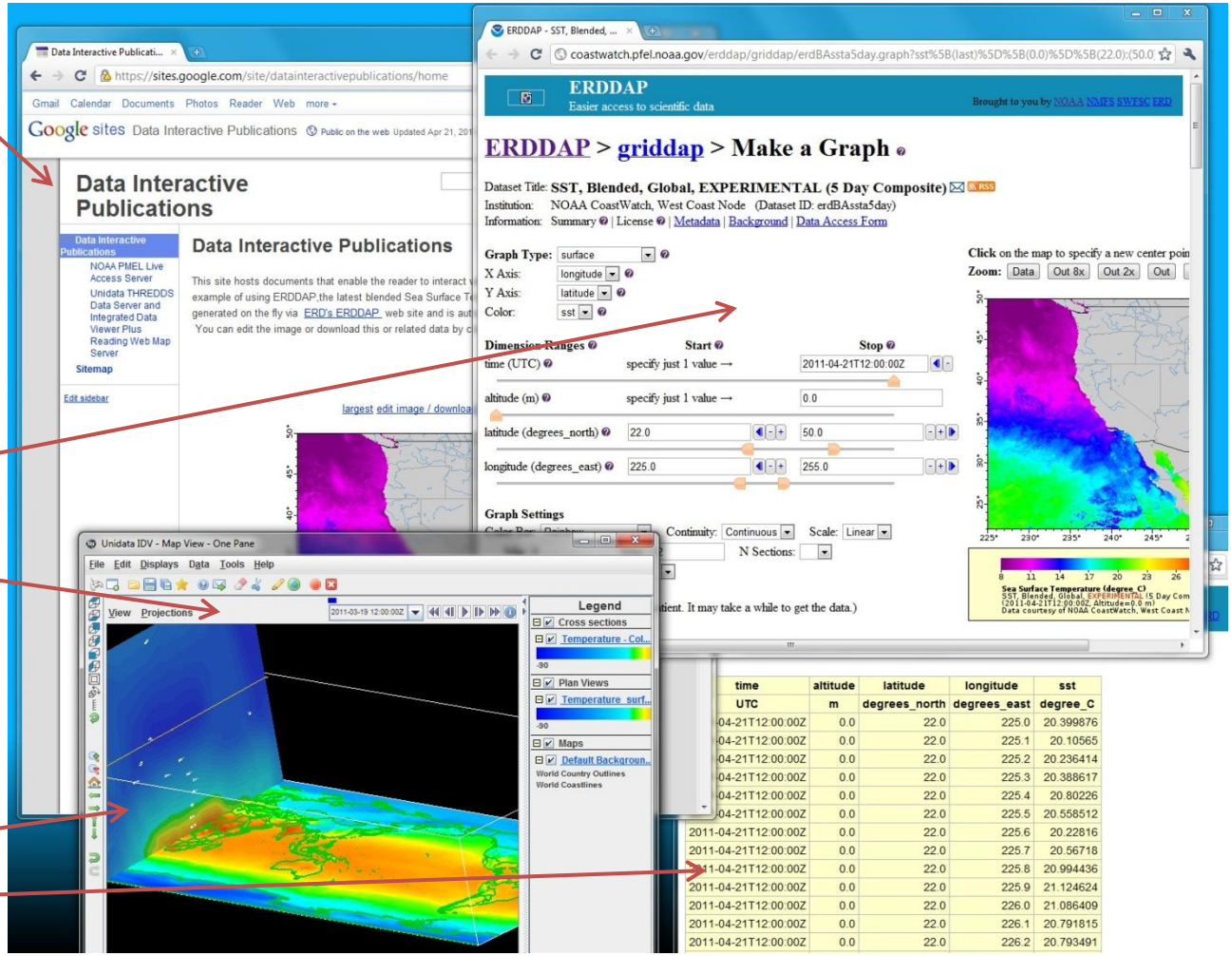
Via Standard Interfaces, Publications Enable Access to Distributed Processing, Tools, Data

Straightforward HTML documents can be created on public wiki or local web server.

Via embedded links, reader accesses data and analysis and display tools.

Tools can be:
- thin web clients,
- rich desktop applications.

From within the document, reader can interact with data via:
- tools,
- "live" tables
- downloads.



Advantages

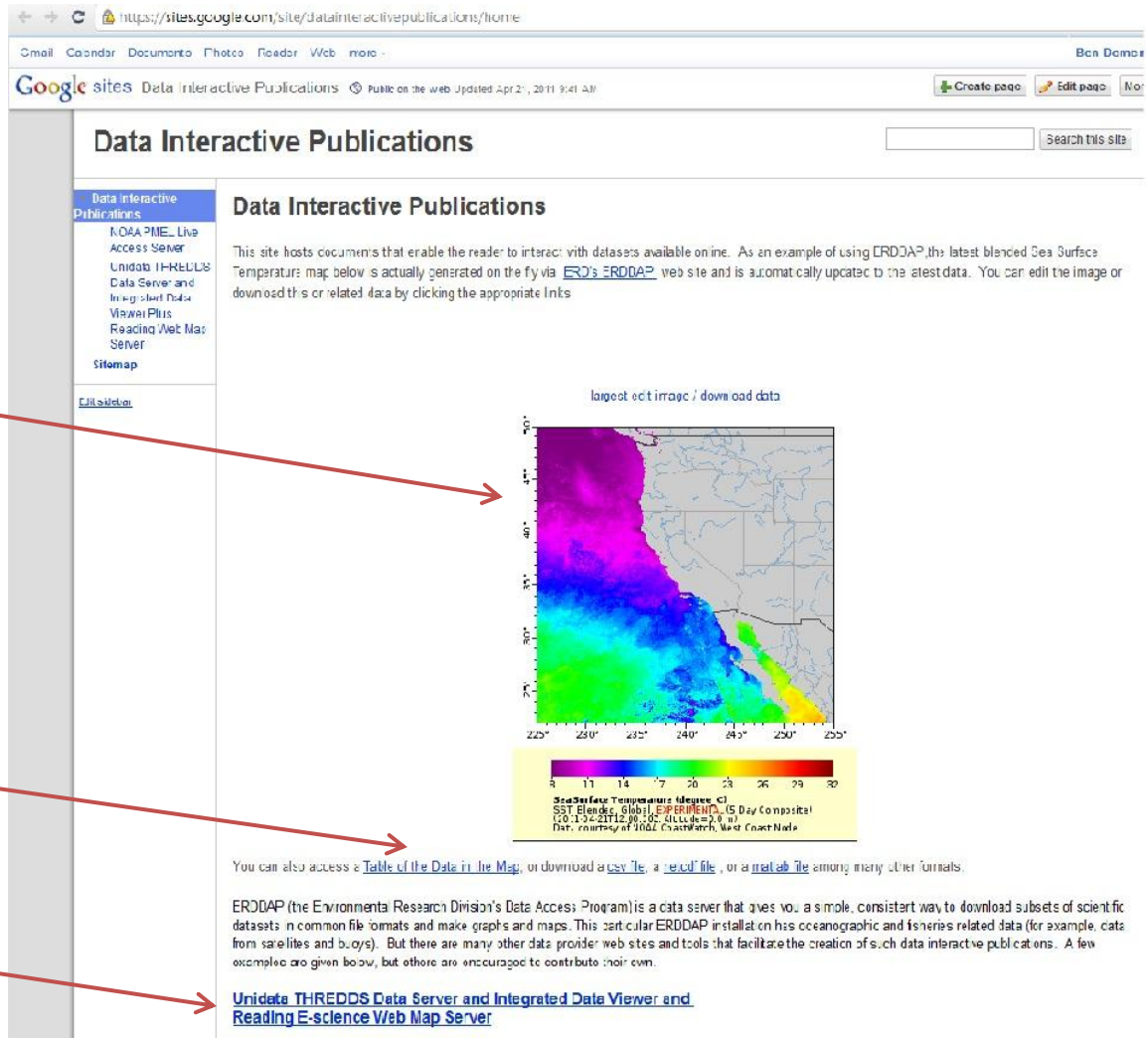
- Convenient authoring of online publications that enable the reader to access and analyze the data cited in the publication
- Community Incentive: makes “documentation” of data a rewarded part of an academic’s job
- Transparency: fosters openness by providing broader and easier access to processing and data as well as products
- Server flexibility: services can be on local workstation/cluster, on central server, in the cloud, or in Wyoming
- Efficiency: processing can be performed near “big data” stores
- Collaboration: could foster more NCAR/Unidata collaboration
- Internal Cooperation: would involve both engineering and non-engineering staff within Unidata
- Client flexibility: data analysis can be driven from mobile devices as well as desktop

Challenges

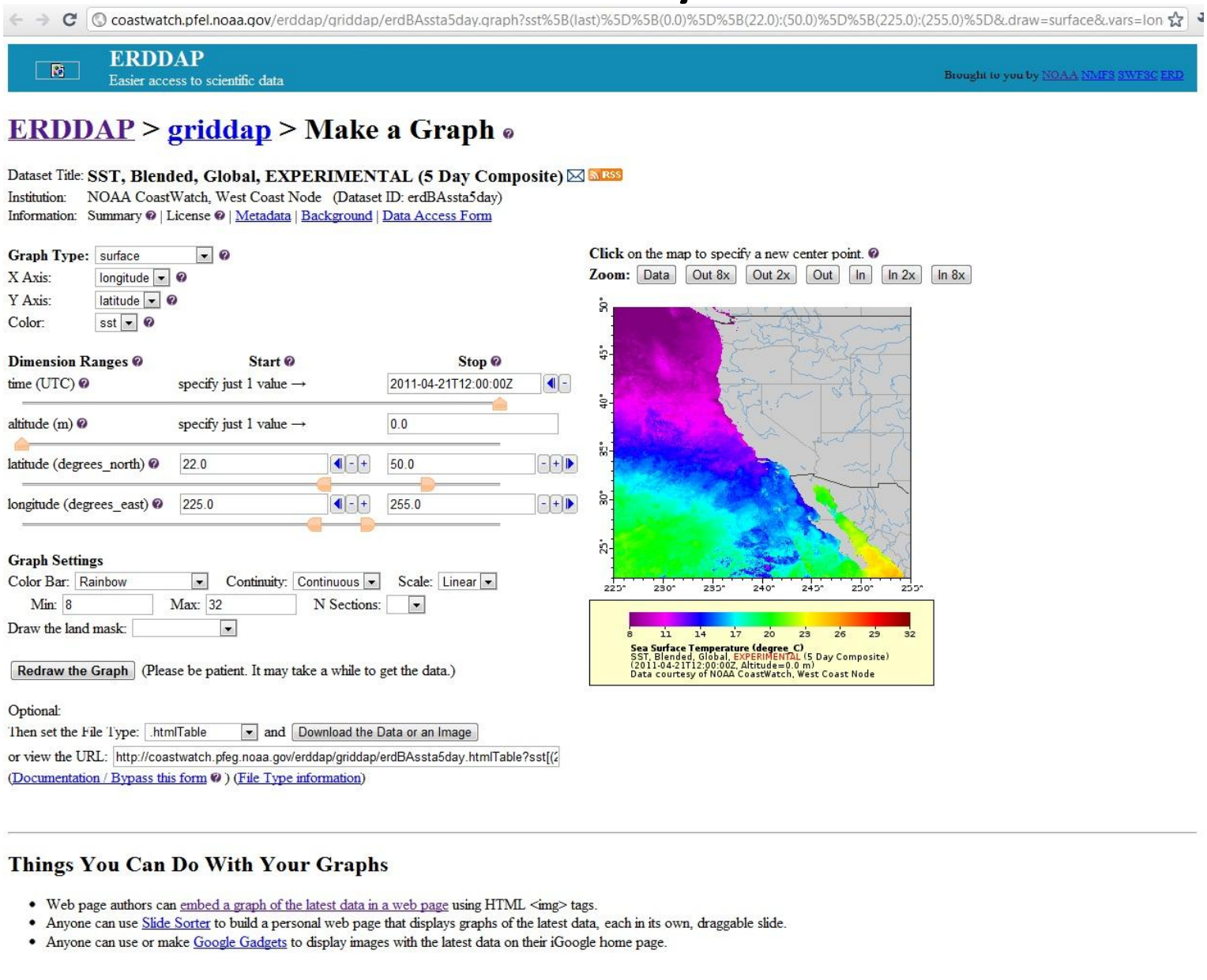
- New way of thinking for developers and users
- Additional effort: new or re-programmed staff
- Persistence of all components: datasets, processing services, and interfaces
- Computing resources: distribution and limitations
- Security: authentication and authorization
- Primitive web processing services at present
- Interfaces are new and generally untried in an environment where people are relying on them
- Need to engage publications industry

Publication on Google Sites Wiki

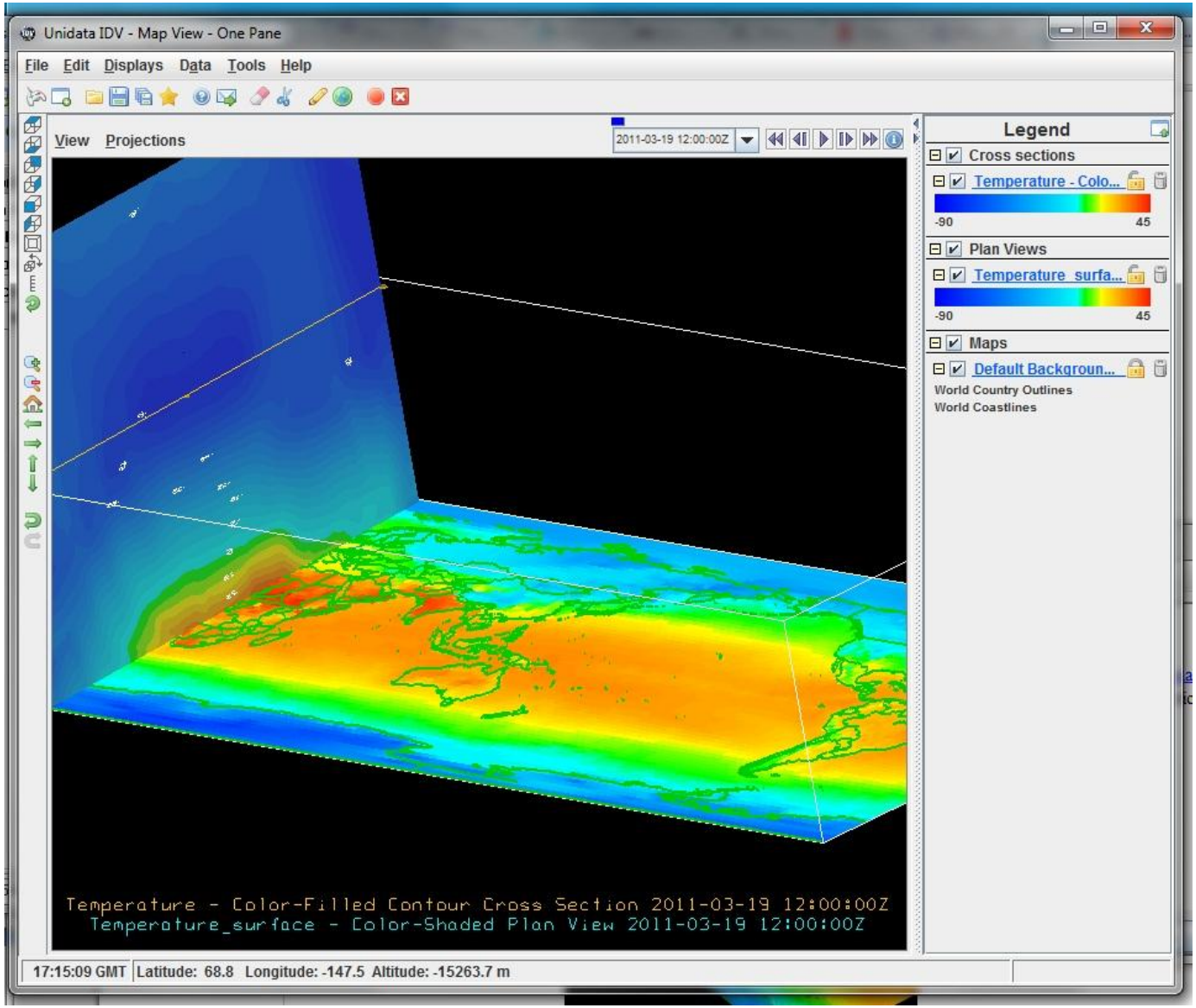
- Online document created with simple wiki tools
- Embedded images generated “on the fly” from another web site
- Embedded links enable reader to:
  - access related datasets
  - perform one’s own analysis and display



Embedded Links to Online Data and Analysis Tools



Embedded Links to Datasets and Desktop Applications



Interoperability and Collaboration

GFS 2.5 degree Surface Temperature

Forecast model output from US National Centers for Environmental Prediction

Real-time delivery via Unidata IDD

Stored in Unidata THREDDS Data Server motherlode

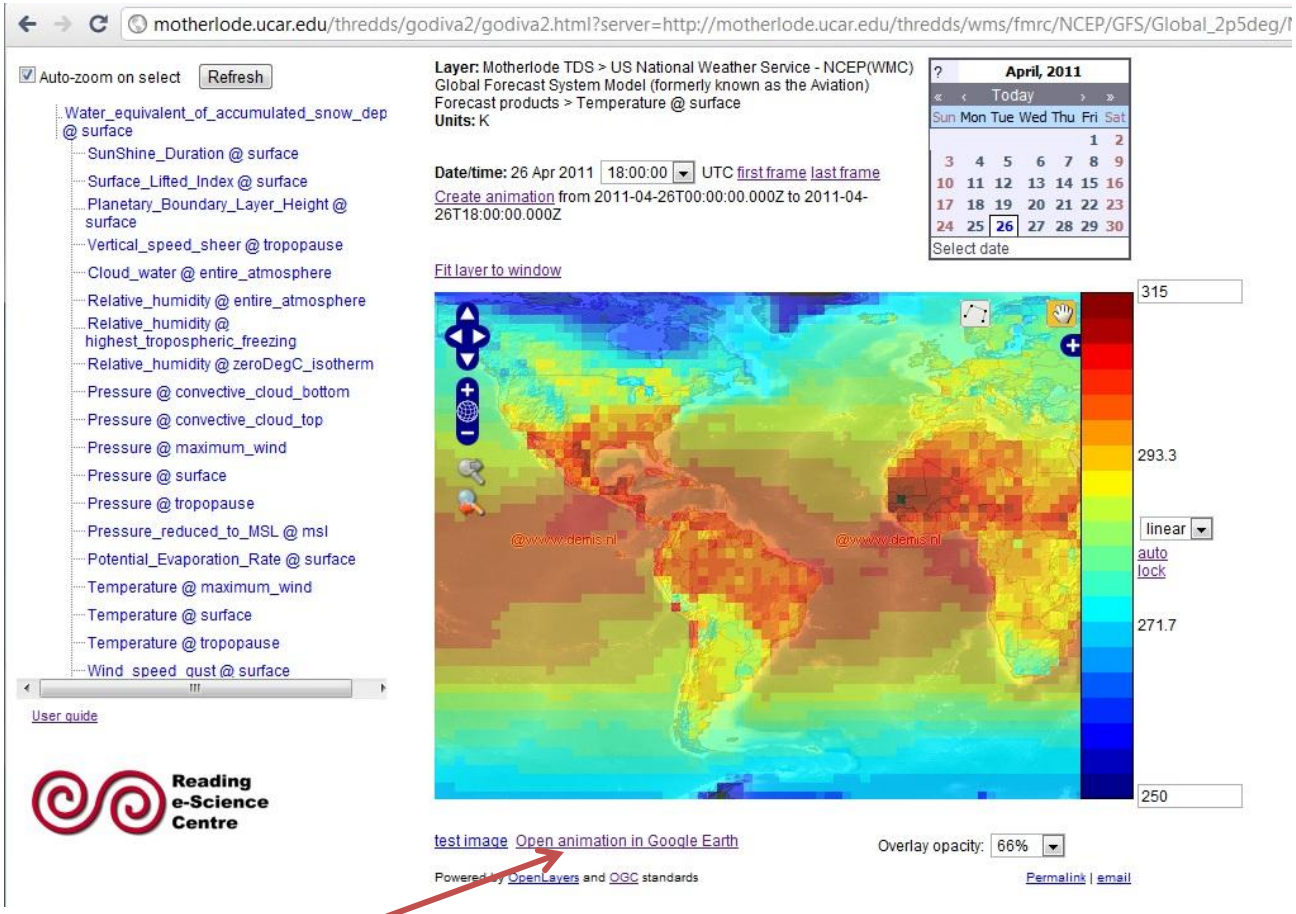
Dataset in WMO GRIB2 files

Access via OGC WCS (Unidata), WMS (Reading e_Science), OPeNDAP (OPeNDAP Inc.), Unidata NetCDF Subset Service …

Unidata netCDF access via Unidata Common Data Model

Visualization via Reading GODIVA2

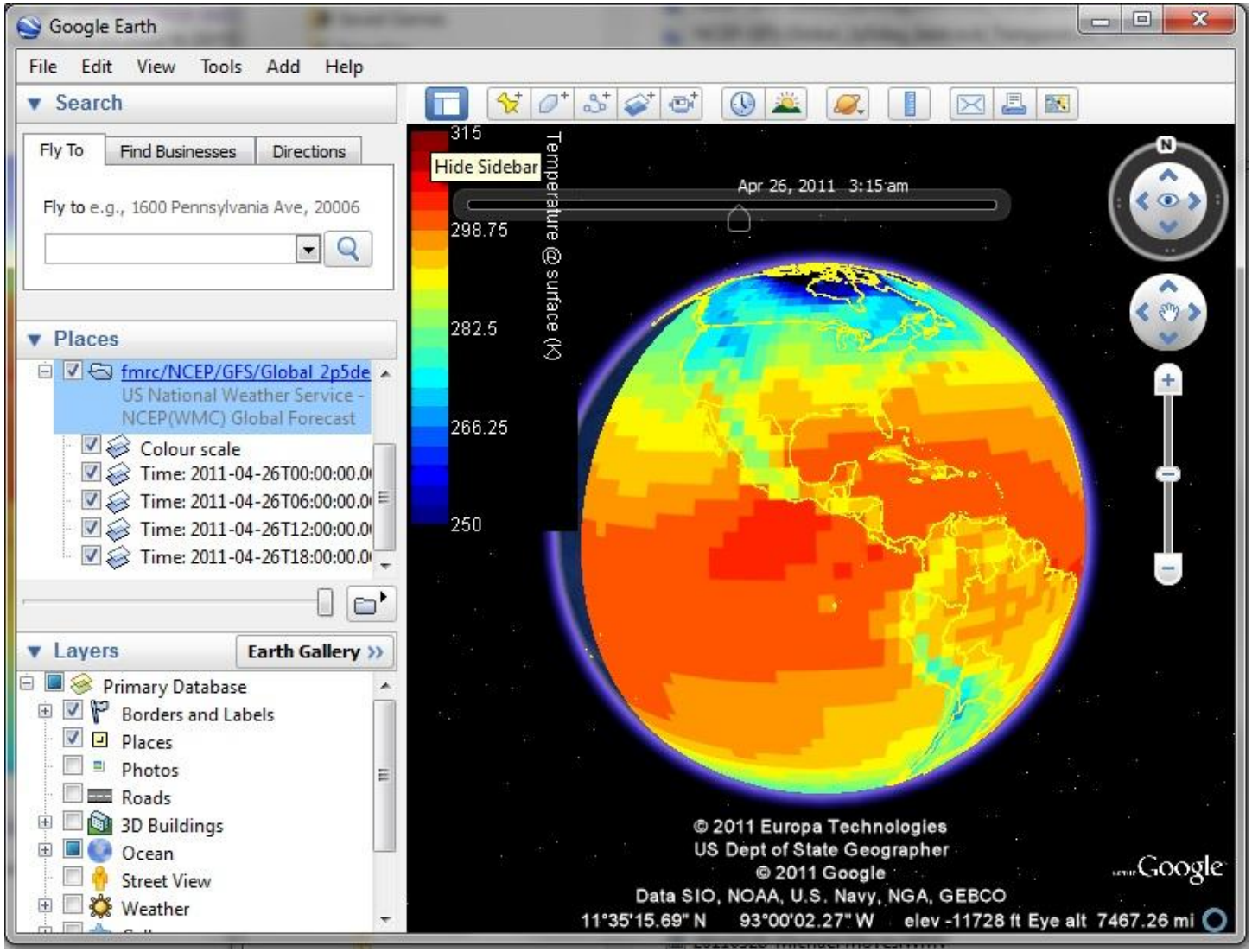
Hundreds of analysis and visualization tools work with Unidata netCDF datasets



Next slide

Google Earth Visualization

Created on Unidata TDS with e-Sciences WMS/GODIVA



Implications of Recent Developments

Earlier barriers are being overcome:

- Browser-based access to interactivity is becoming more common (Live Access Server, TDS/WMS/GODIVA, ERDDAP, FerretTDS…)
- Server-side (cloud?) processing is a viable platform for interactivity
- RAMADDA/TDS provides mechanism for persistent storage of datasets by user community
- Clients with modest processing/storage ability come into play (tablets, smart phones)
- Documents, data, storage, data access, processing capabilities, user devices can all be distributed
- Processing can be co-located with large datasets.

Areas Needing Attention/Resources

- TDS/RAMADDA authentication enabling user data input for persistent storage
- Web processing services that enable more processing on server side (cloud?)
- Simple apps on modest client platforms that enable interaction with data access AND PROCESSING services

Examples

- General wiki-based publications at https://sites.google.com/site/datainteractivepublications/
- TDS/ncWMS motherlode example: http://motherlode.ucar.edu/thredds/godiva2/godiva2.html?menu=&layer=Temperature_surface&elevation=0&scale=205.9,306.3&bbox=-142.03125,-68.203125,37.96875,72.421875&server=http://motherlode.ucar.edu/thredds/wms/fmrc/NCEP/GFS/Global_2p5deg/NCEP-GFS-Global_2p5deg_best.ncd
- Live Access Server example: http://ferret.pmel.noaa.gov/NVODS/getUI.do?dsid=woa01_monthly&catid=0B541688EA4ACDF44451F0623AE315CF&varid=t0112an1
- ERDDAP Example: http://www.pfeg.noaa.gov/~cwilson/bloom/BW_14.html

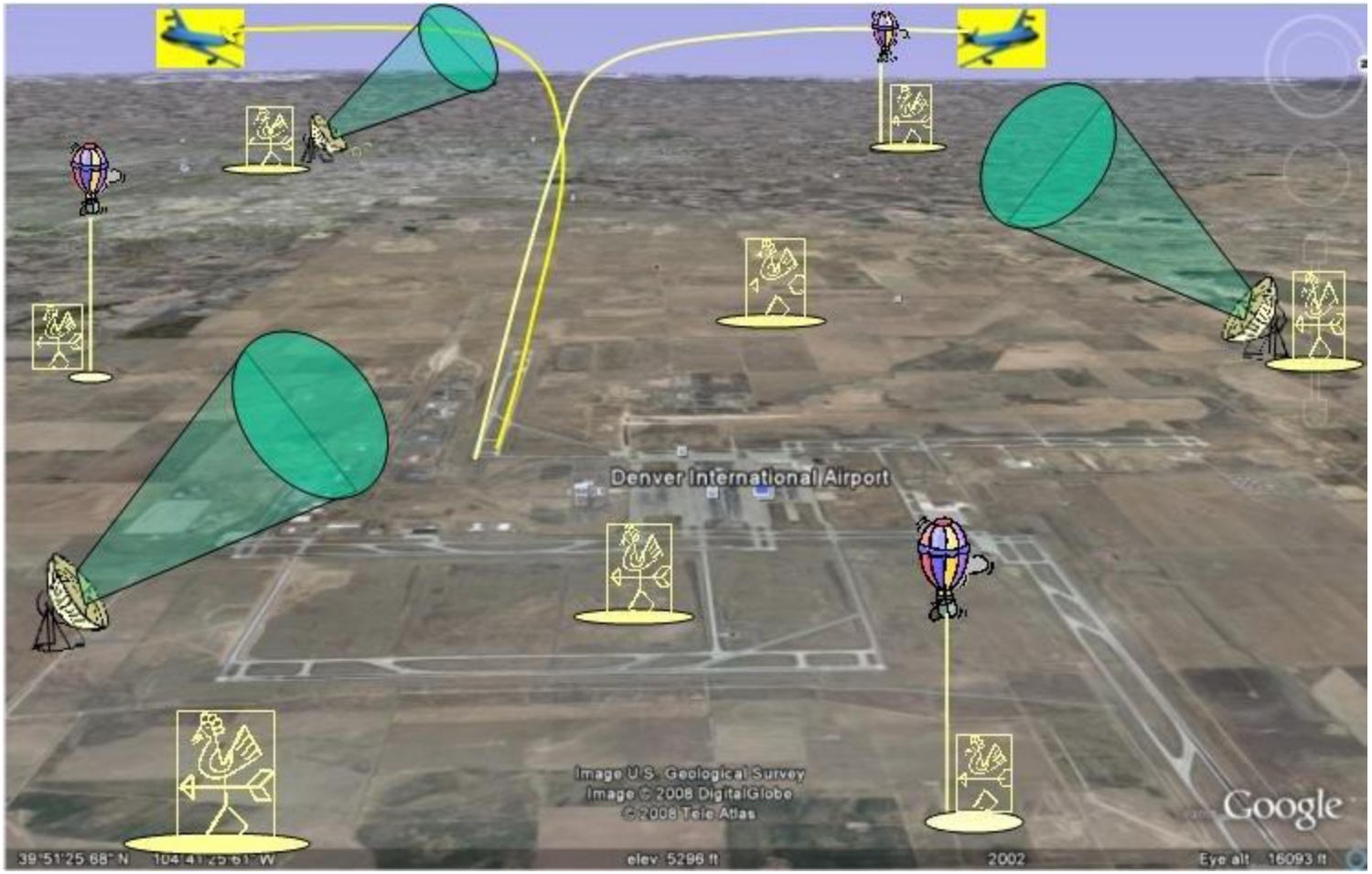
Remaining Slides Optional

- If there is time, the rest of the slides can be the basis of a technical discussion of how to deal with different scientific data types.

How to Deal with ***Collections*** of Various Scientific Data 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.

Airport Weather Use Case

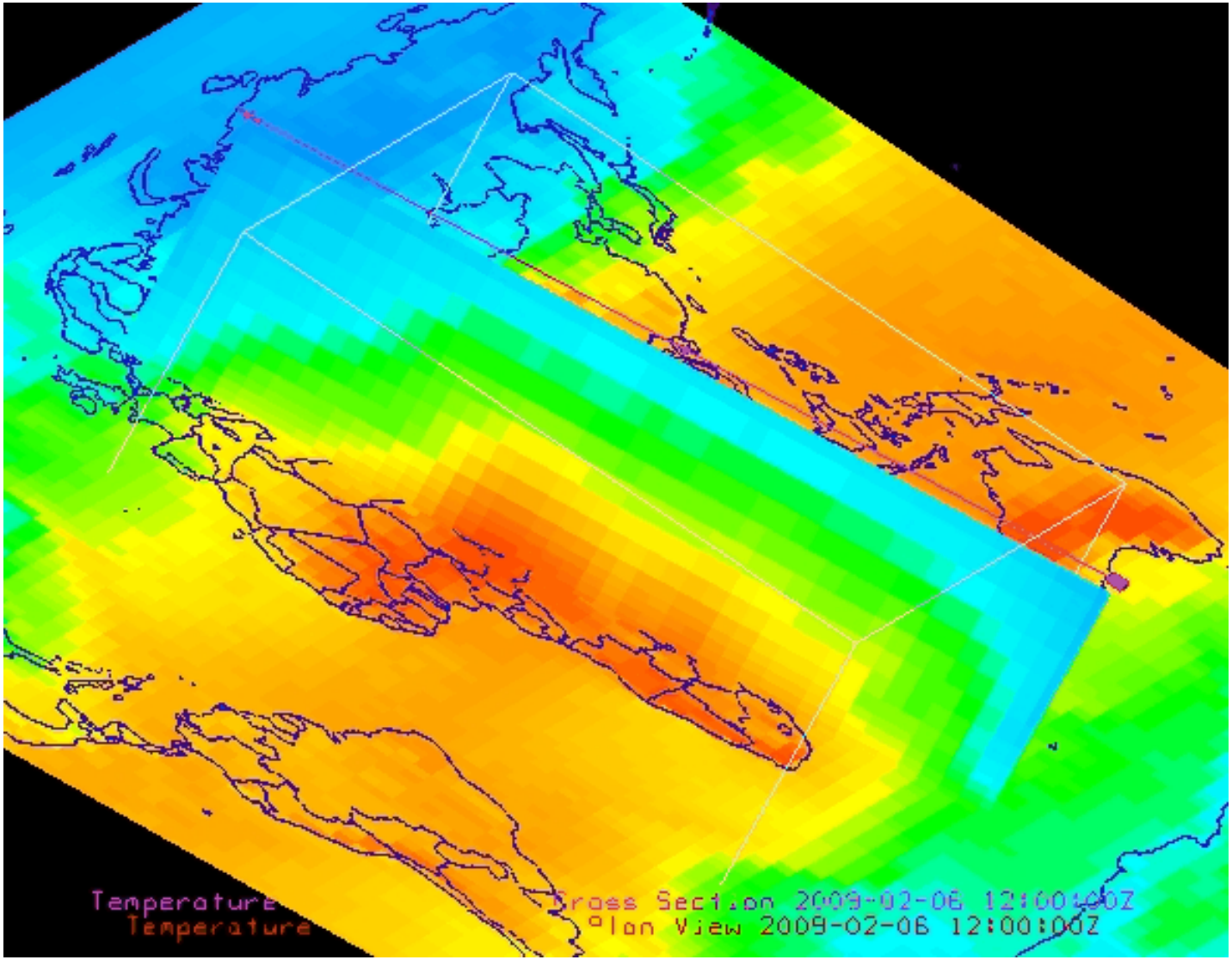


Give me all the atmospheric observations and forecasts in a 3D bounding box around the airport during the time frame of a recent storm

GALEON 1 Lessons

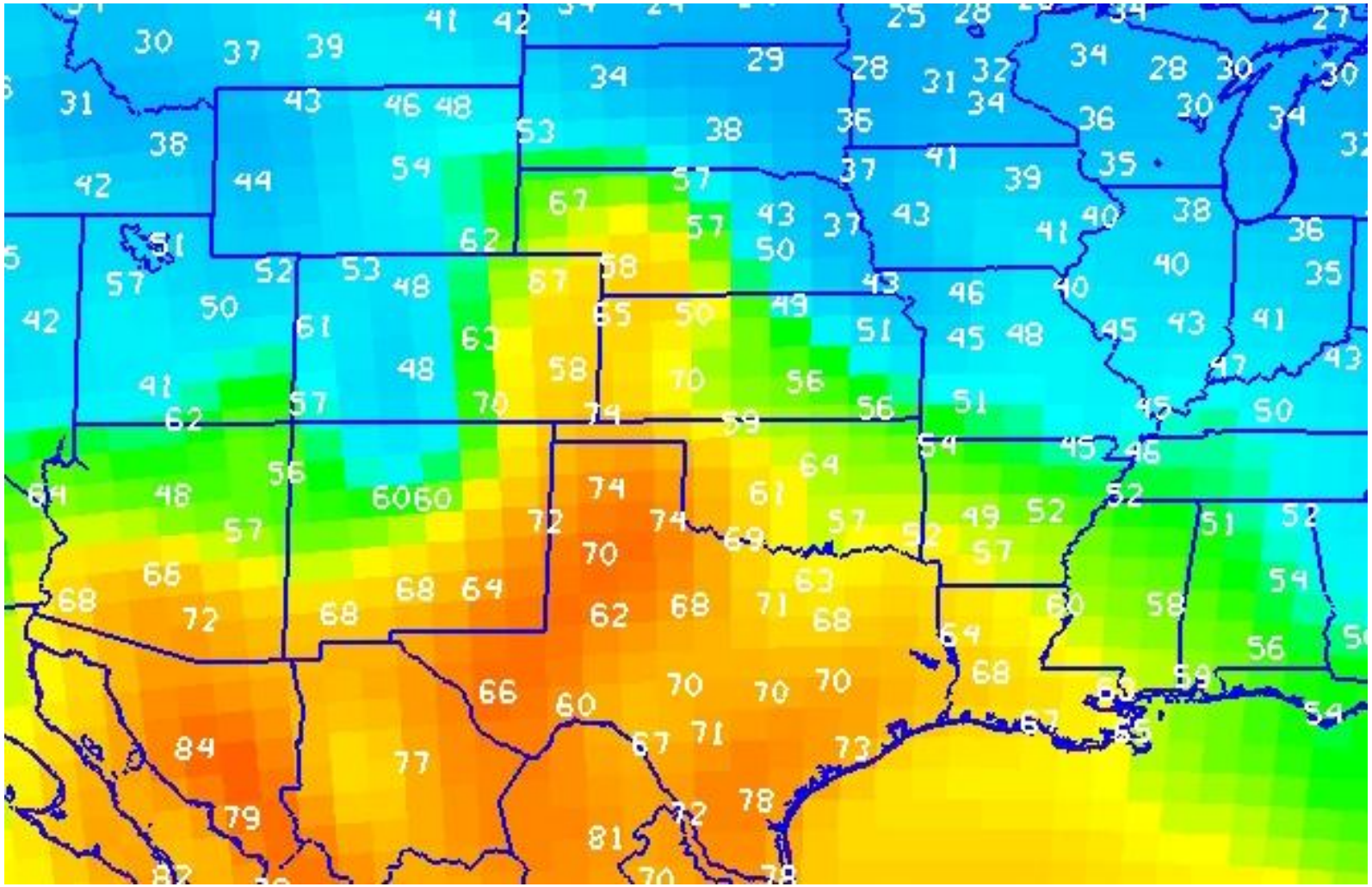
- Relatively simple WCS use case is valuable:
  - Bounding box (in coordinate space), time frame, coverage name (e.g., surface temperature) subsetting is practical
  - CF-netCDF payload works for many clients
- WCS limitations:
  - gridded data (regularly spaced in some projection
  - WCS 1.1 complicated (all things to all people)
  - WCS going toward modular core/extensions approach

Gridded Output of Forecast Models



WCS is ideal for this scientific data type

Collections of Station Observations



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

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

Climate Science Modeling Language and Common Data Model Feature Types

Scientific Feature Types	
CSML	CF/CDM
Point	Point
PointSeries	StationTimeSeries
Trajectory	Trajectory
Profile	Profile
ProfileSeries	StationProfile
Swath	Swath
ScanningRadar	StationayrRadialSweep
Section	Collection of Profiles
Grid	Grid (single time)
GridSeries	Grid

Alignment of CSML with CF Point Obs.

(Quote from BADC CSML Group)

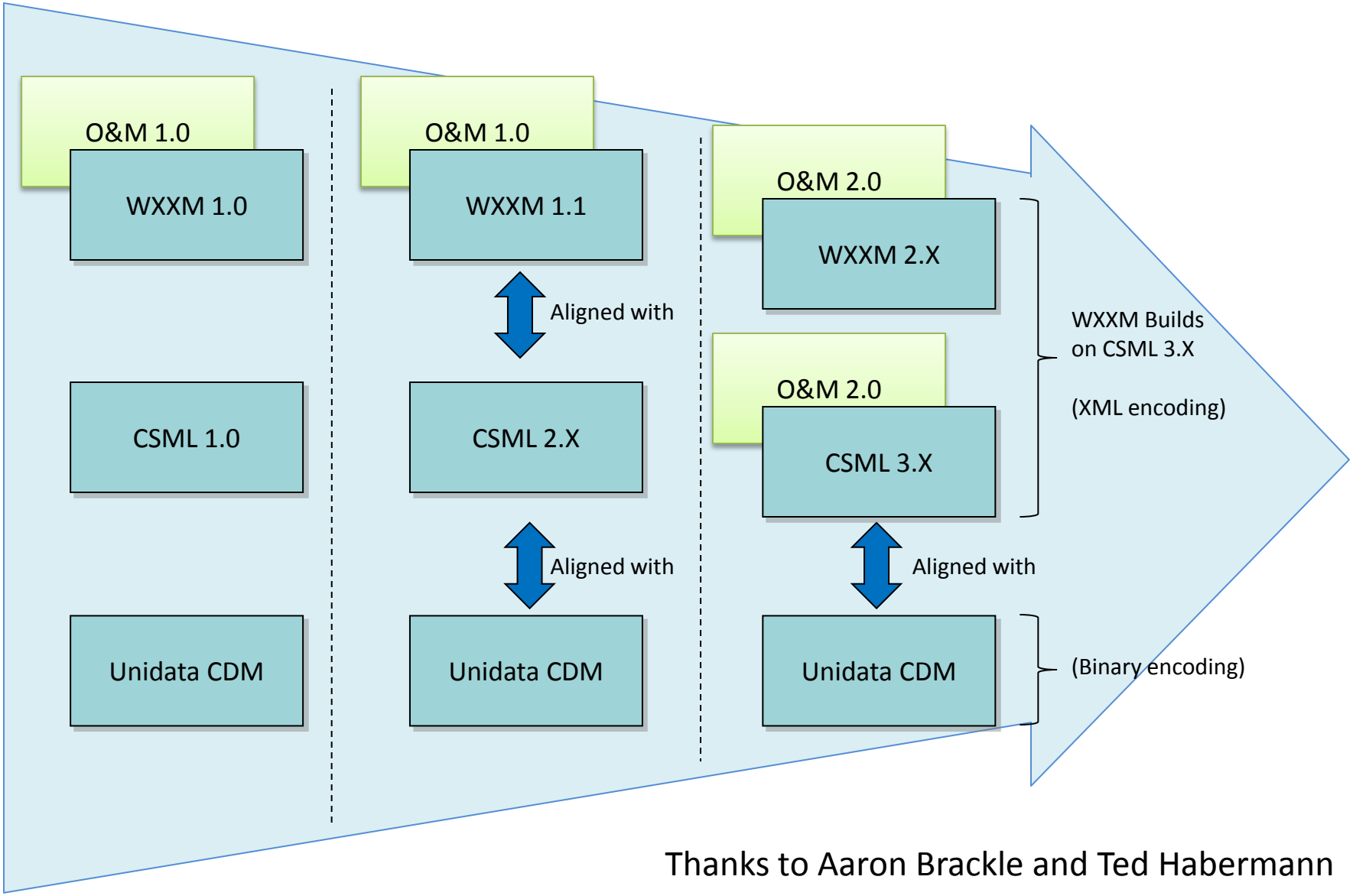
For several iteration now the Climate and Forecast (CF) conventions for Point Observations and CSML have been converging on similar feature models. The latest version of CSML sees very direct mappings between CSML and CF Point Observations (see also the Unidata Common Data Model). This will help bridge the gap between the netCDF and file view of the world and the OGC feature and service view.

Scientific Data Types
Mapping to ISO Coverages (from 2009)

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***

Convergence of Scientific Data and Feature Types



Thanks to Aaron Brackle and Ted Habermann

References

- CSML
  http://csml.badc.rl.ac.uk/
- CF Conventions
  http://cf-pcmdi.llnl.gov/
- Unidata Common Data Model
  http://www.unidata.ucar.edu/software/netcdf-java/CDM/

Future Opportunities

- Develop “final” agreement on scientific feature types with OGC CSML, CDM, CF
- Finalize CF conventions for station/point (discrete sampling) features
- Establish collections of station/point data as recognized ISO/OGC coverages.
- Participate in/guide CF conventions for radar, swath, unstructured grids
- Bring additional data types into OGC
- Style layer descriptors for MetOceans datasets