NetCDF and Related International Standards

Ben Domenico

October 2012

Outline

- Brief historical context
- Unidata/partners have established a solid foundation: Standard data access interfaces enable other Earth System Science communities to access Unidata datasets
- Recent progress: data model, interface, encoding standards
- Current focus, web services brokering layer:
  - new architecture with broker between clients and web servers
  - CF-netCDF data model makes brokering possible
- Future directions: data analysis and display via publications, educational modules and other documents

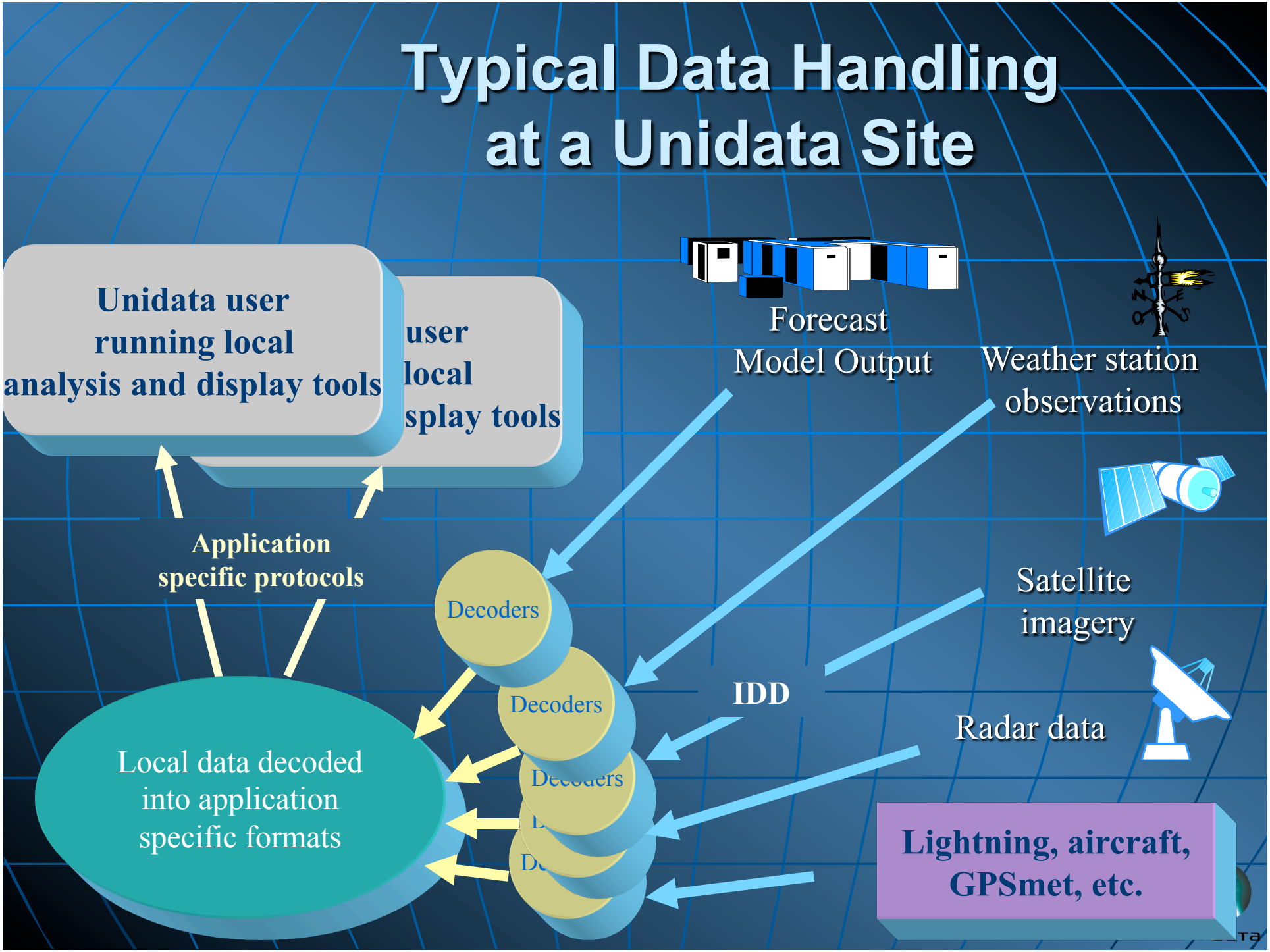
Navigating Earth System Science Data vs. service to Unidata core community?

- Evolution of Unidata data systems
- Guidance from proposal review panel regarding outreach vs. service to core community
- 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
  - Unidata does NOT support other communities' tools
  - Unidata does NOT provide expertise in other domains
  - We get others to do most of the hard work

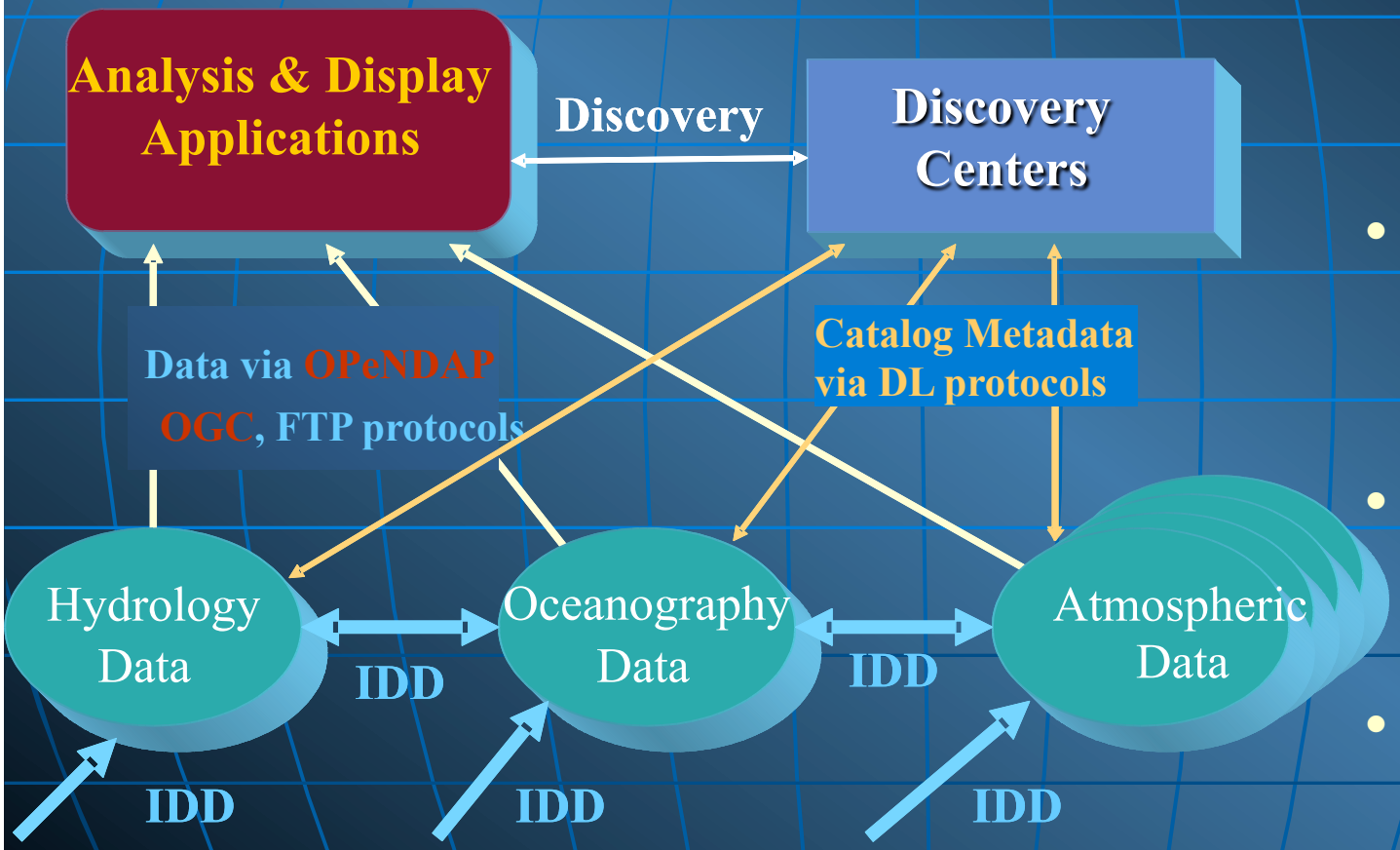
Unidata Data Systems Evolution

- **IDD/LDM:** Publish/subscribe (push) system for real-time data, i.e.,
  move the data to the user sites and provide tools to analyze and display the data on student and faculty workstations
- **THREDDS Data Server:** Client/server (pull) for retrospective and cross-disciplinary, i.e.,
  data, analysis and display client tools access data from servers (local or remote)
- **What next?**

Typical Data Handling at a Unidata Site



Thematic Real-time Environmental Distributed Data Servers (THREDDS)



- Integrate environmental data and tools into the world wide web
- Combine IDD "push" with several forms of pull and DL discovery
- About 25 data providers are THREDDS partners
- Connecting people, documents, and data

Note: *Initial THREDDS grants from NSF Education and GEO directorates (not Atmospheric Science)*

unidata

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.

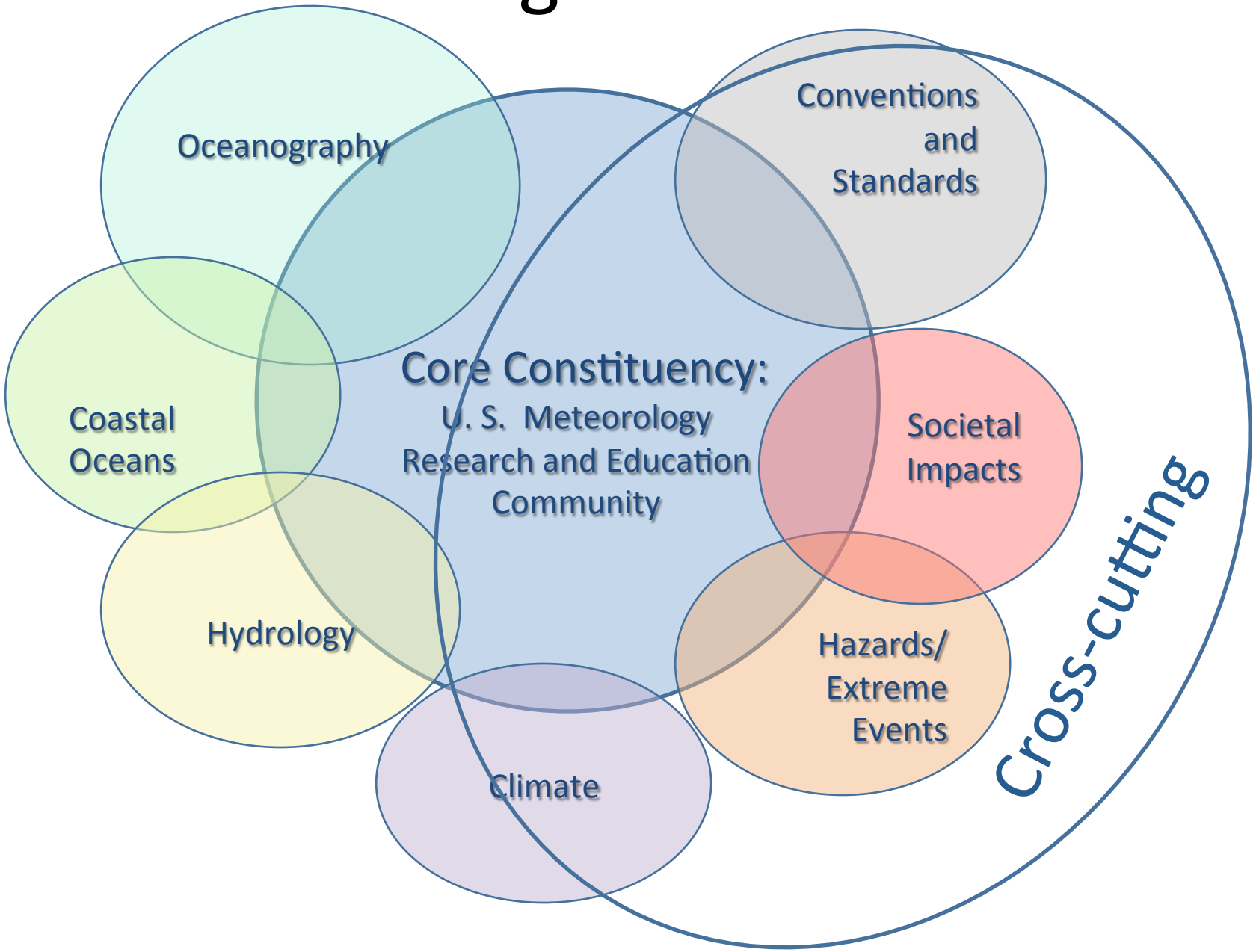
- Data access with supported analysis and display tools for our community
- Data access via standard interfaces to enable other communities to use our data with their own tools (arcGIS, IDL, Matlab, etc.)

Excerpts from Review Panel Report

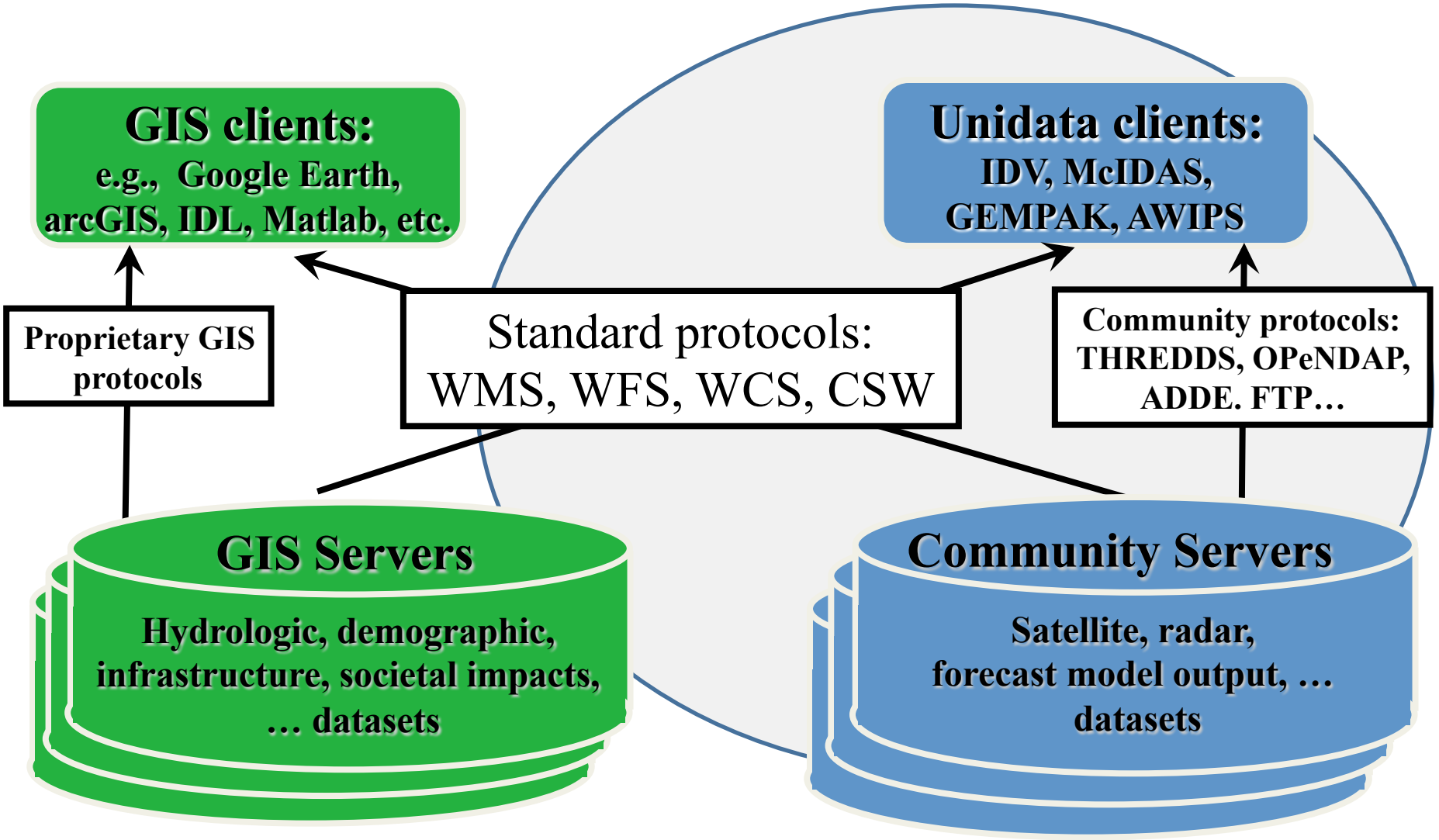
The UPC could play a significant leadership role within committees and consortiums like OGC seeking to address the need to develop standards and technologies for data discovery

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

Collaborating Communities



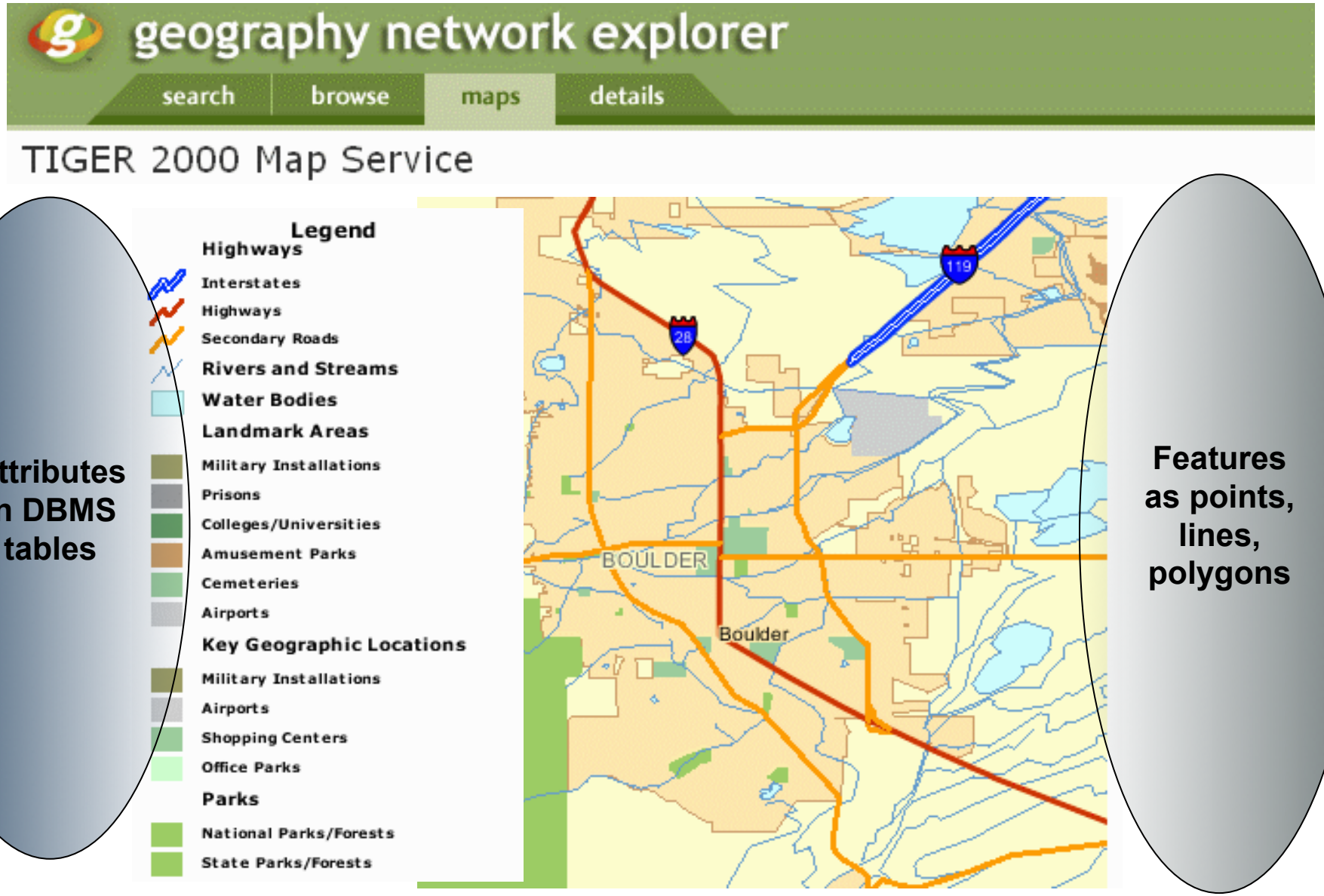
Standard Interface Web Services for Interoperability with GIS Community



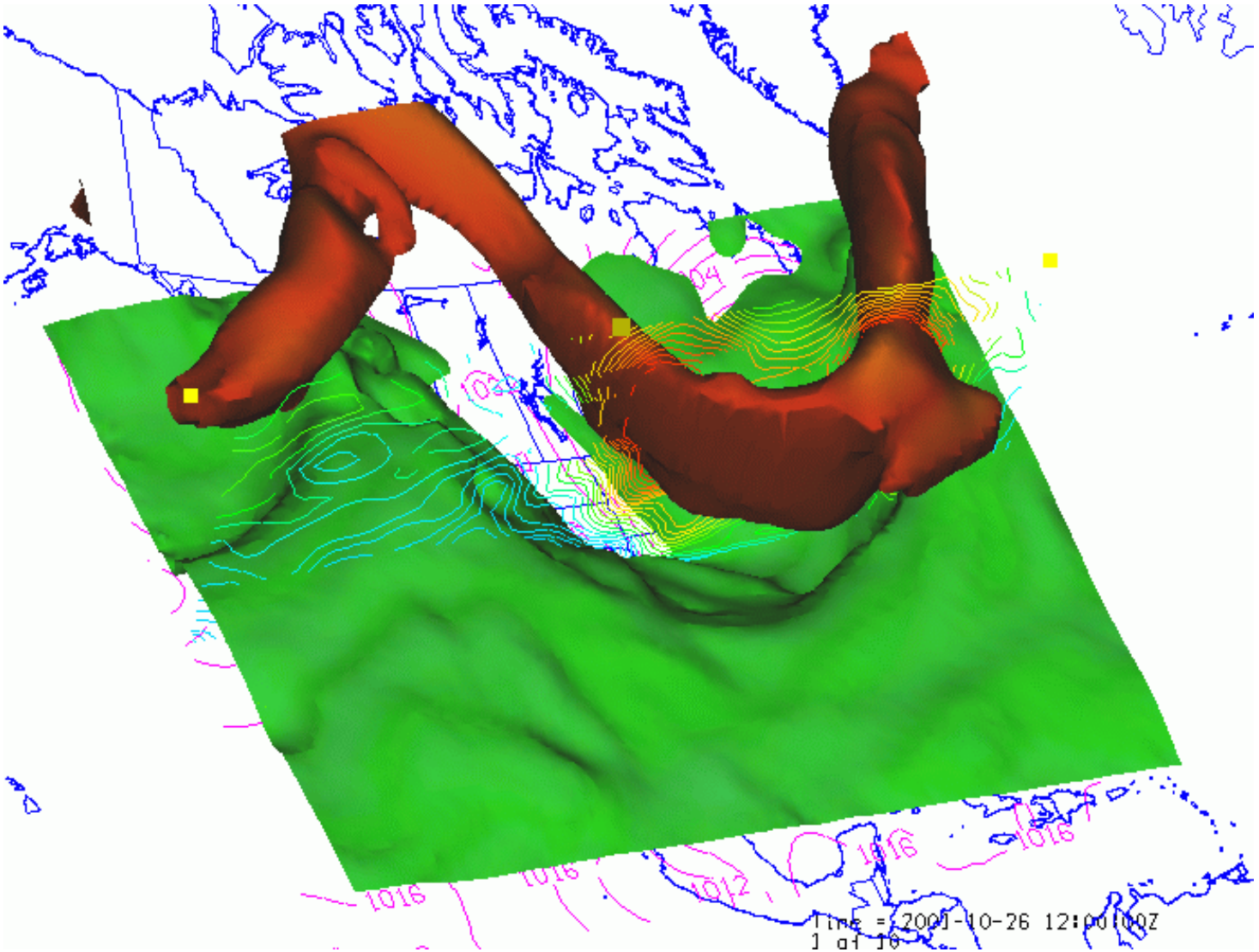
Disparate Data Models: Different Ways of Thinking about Data

- 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

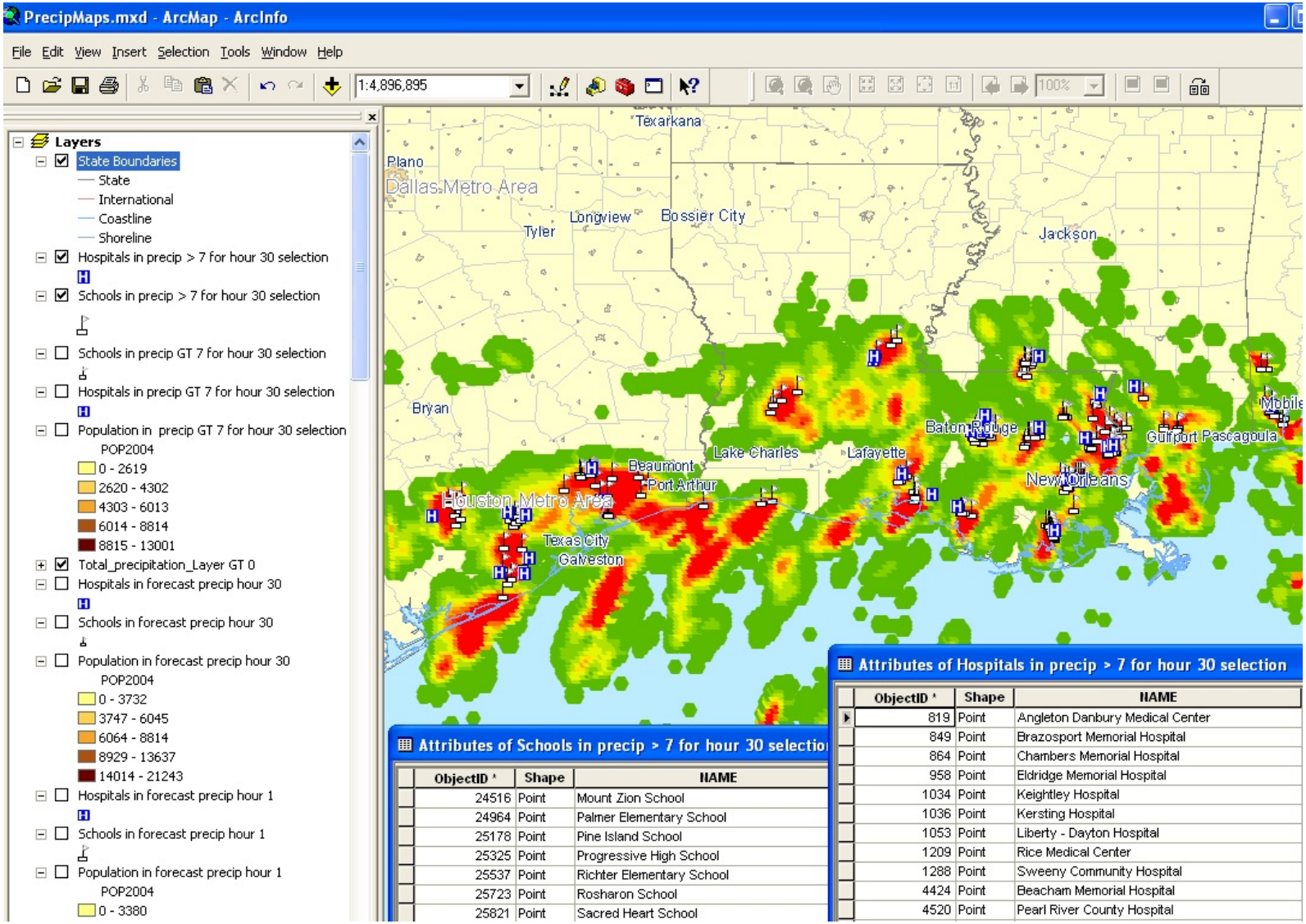
Traditional GIS view



Typical NetCDF Visualization



GIS Tools Applied to Atmospheric Science Data



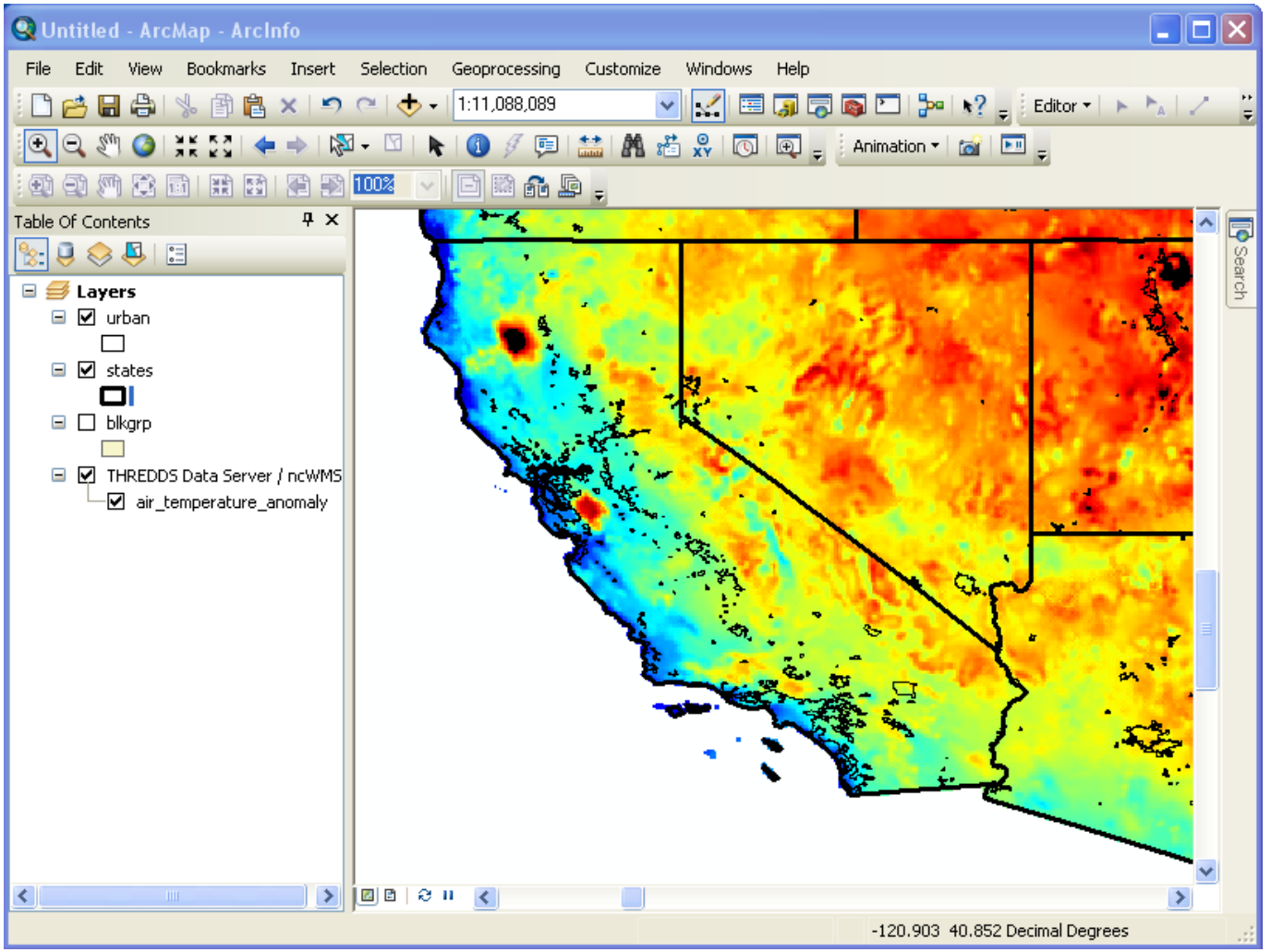
A Few of the Collaborating Organizations

- 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

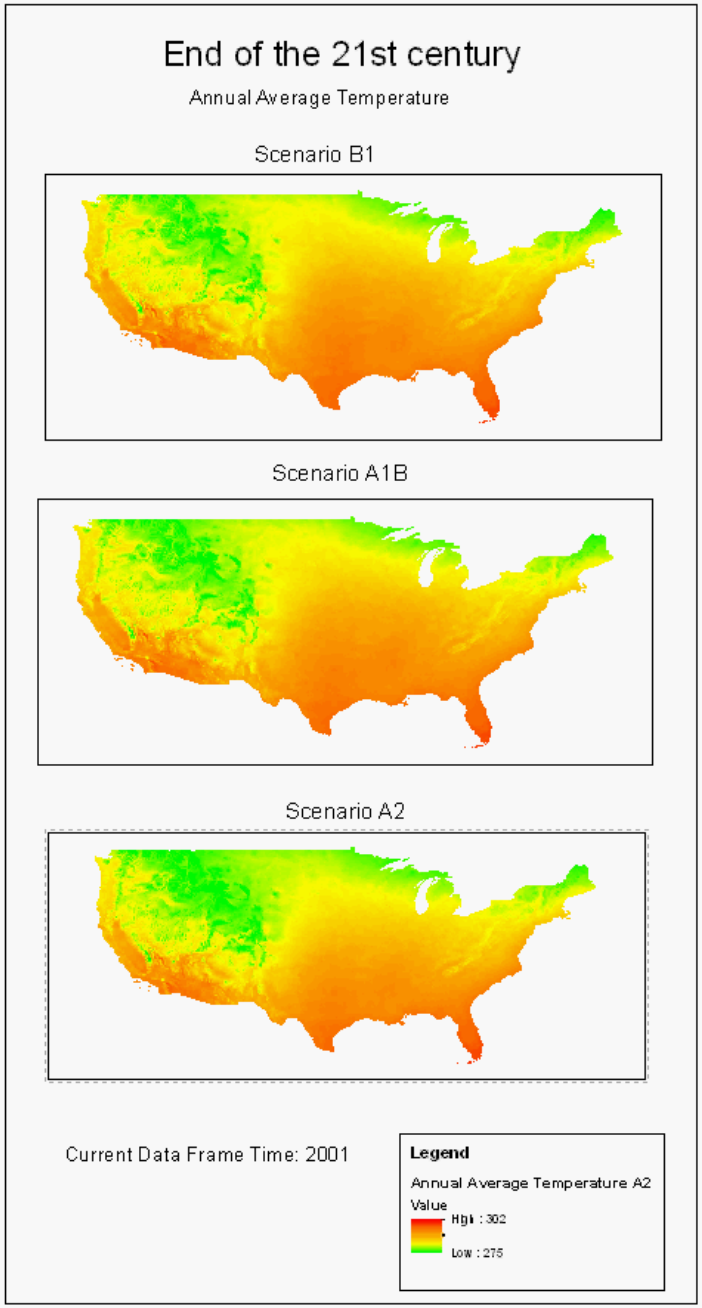
NCAR model air temperature anomaly display

- 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.

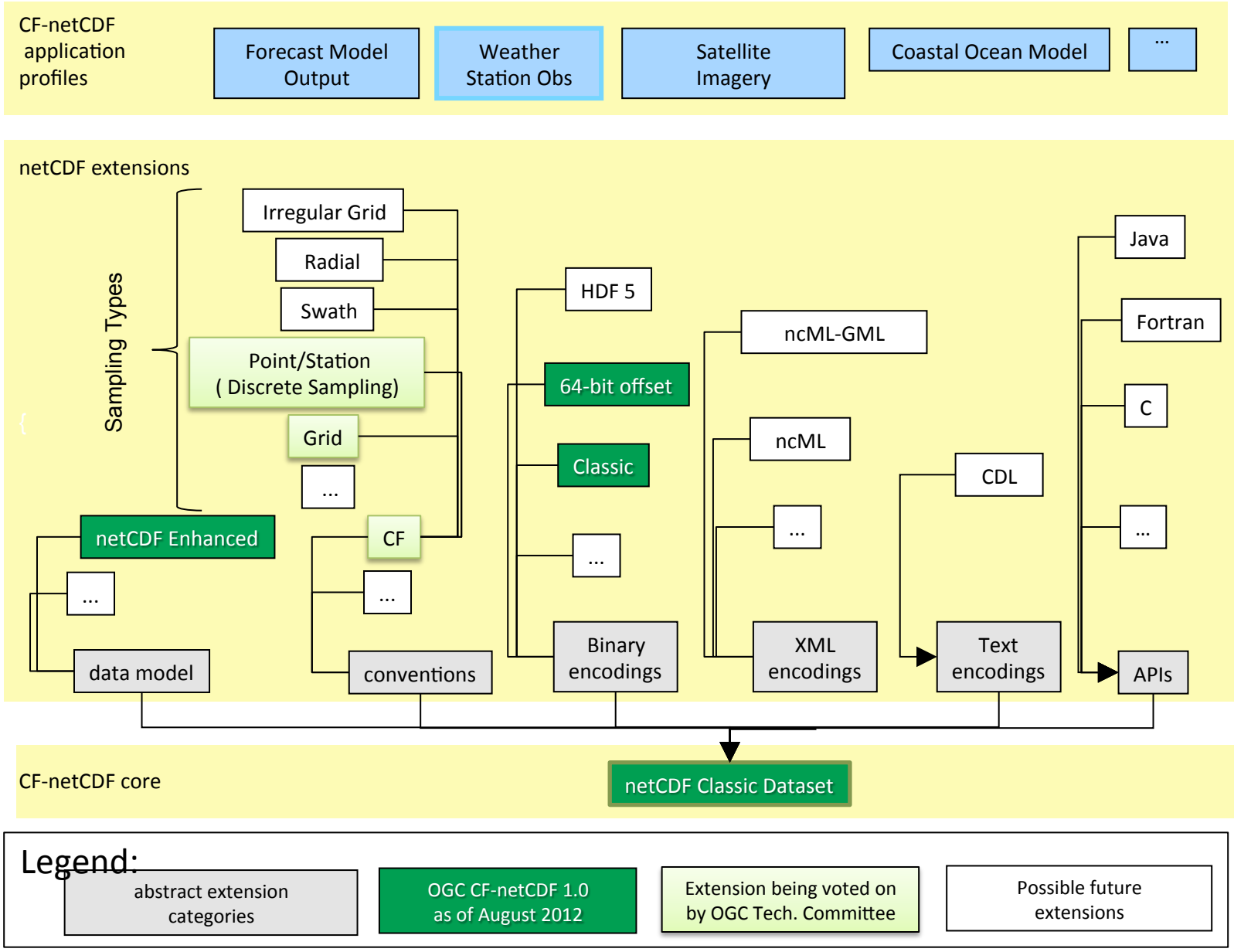


Recent Progress

- netCDF (classic and enhanced) 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 has joined OGC
- Discussed plans with CF community to coordinate data model developments
- 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



Recent Focus in OGC

- Enhanced data model (netCDF4) extension to netCDF core spec (formally adopted last month)
- CF extension to OGC netCDF core spec (being voted on at this moment)
- CF-netCDF extension to WCS core spec (in draft form)
- CSML and CDM/CF Feature Types
- Hosted September 2011 OGC Technical Committee Meetings

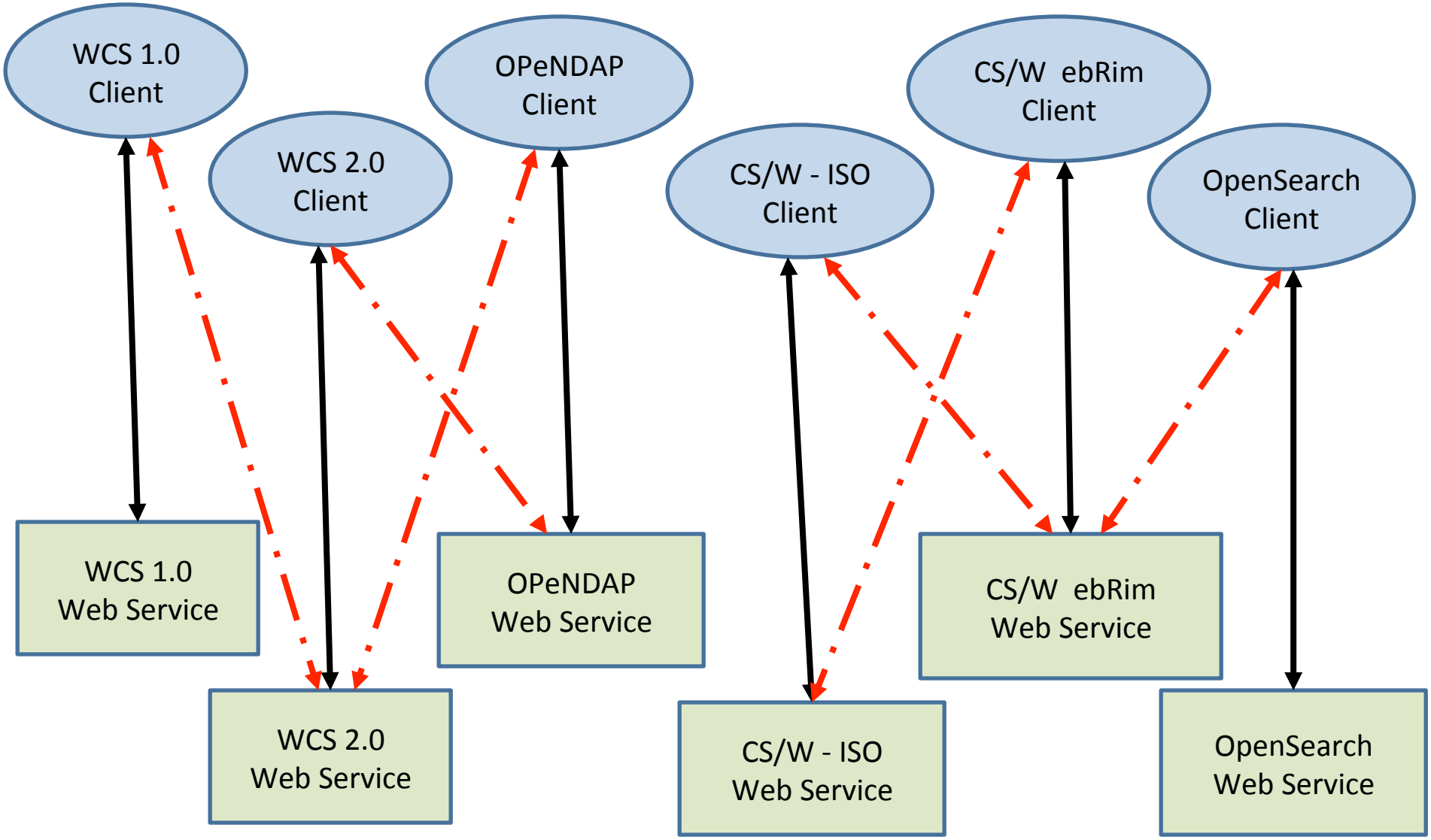
So What?

- Enables cross-disciplinary use of data without undo drain on support for core constituency
- Fosters awareness and adoption of Unidata infrastructure
- Makes more data systems compatible with Unidata tools and services
- Provides access to more data from different, organizations, disciplines, and regions discoverable – with documentation and in usable forms
- Satisfies legal requirements for some organizations, e.g., in EU, required by law to use standard-compliant software
- Encourages valuable international collaborations on development of tools and provision of services
- Makes new work paradigms and system architectures possible

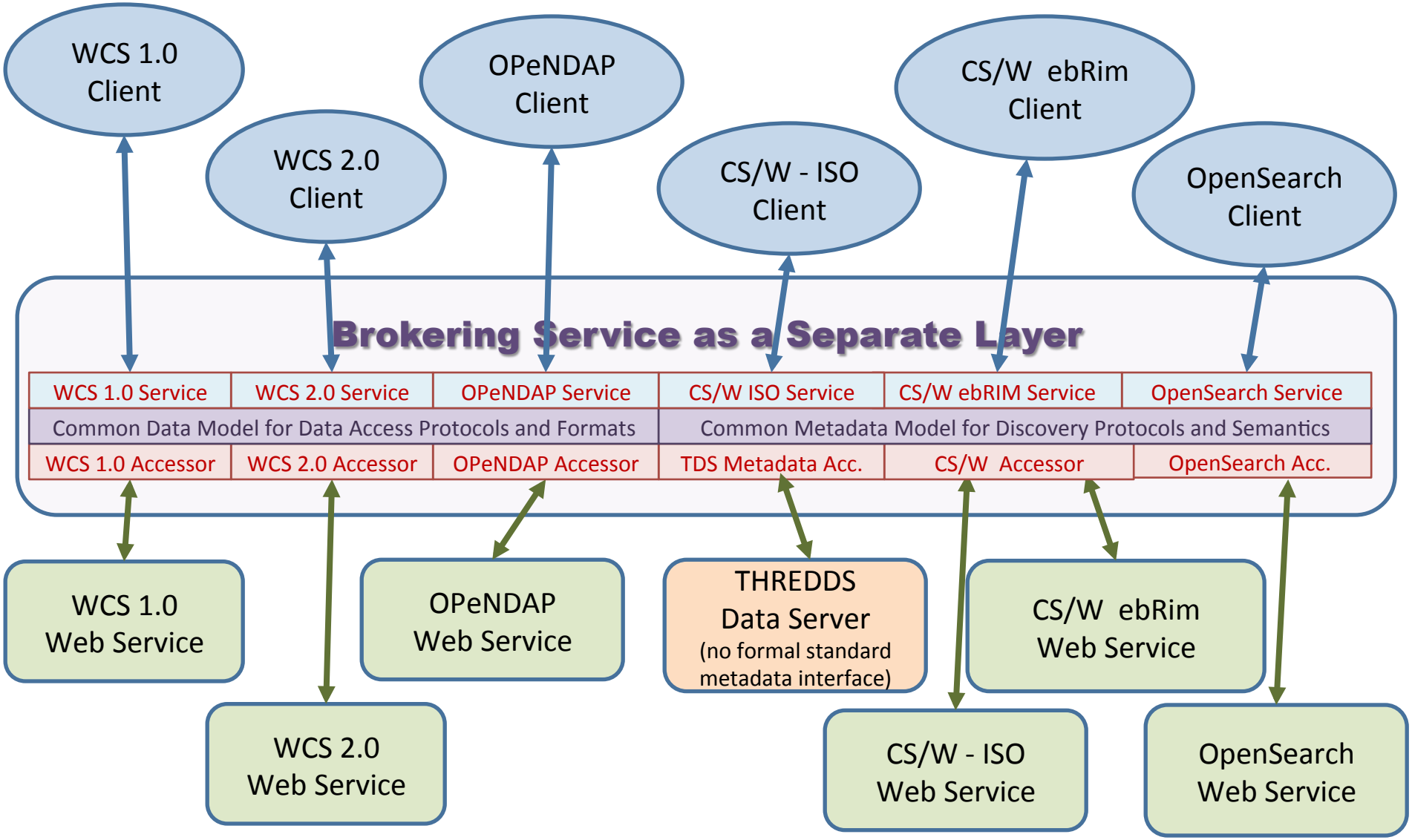
Advantages and Limitations of Web Client / Server Architecture

- Agreeing on protocols, enables communication between programs:
  - developed by different groups,
  - in different languages
  - running on different machines on different networks
- In practice:
  - Clients and servers support limited number of protocols and encodings.
  - Too often, clients only communicate well with servers developed by same group.

Practical Limitations of Web Client / Service Architecture



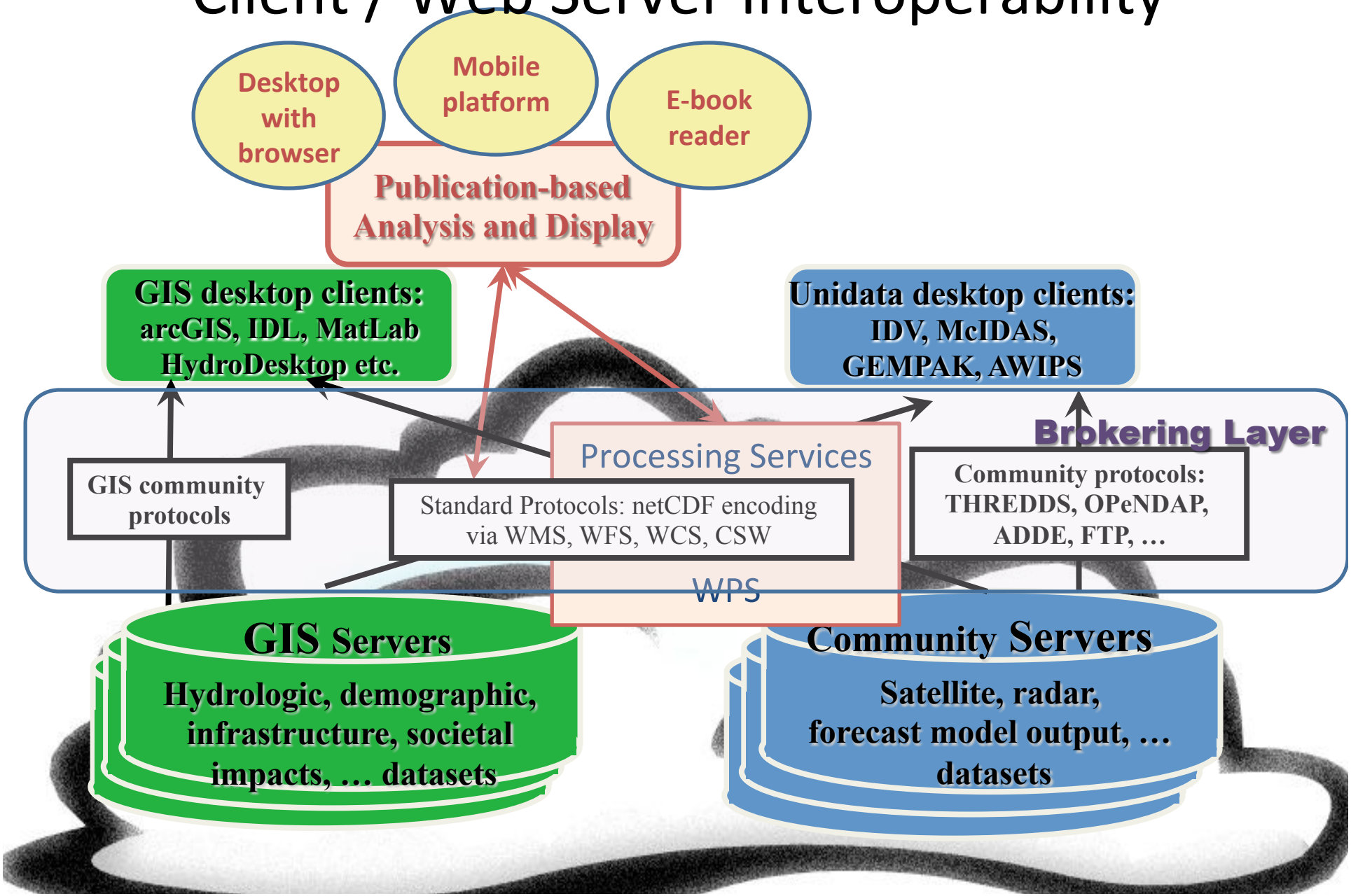
Three-tiered Architecture with Brokering Layer



All this Lead to a 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 these publications
- Bridges the gap between publications and data analysis
- For "big data," it's important to have processing take place near the data store.
- Approach fits nicely into era of server side (cloud?) computing and mobile client devices?

Data Interactive Publications: Client / Web Server Interoperability



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

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

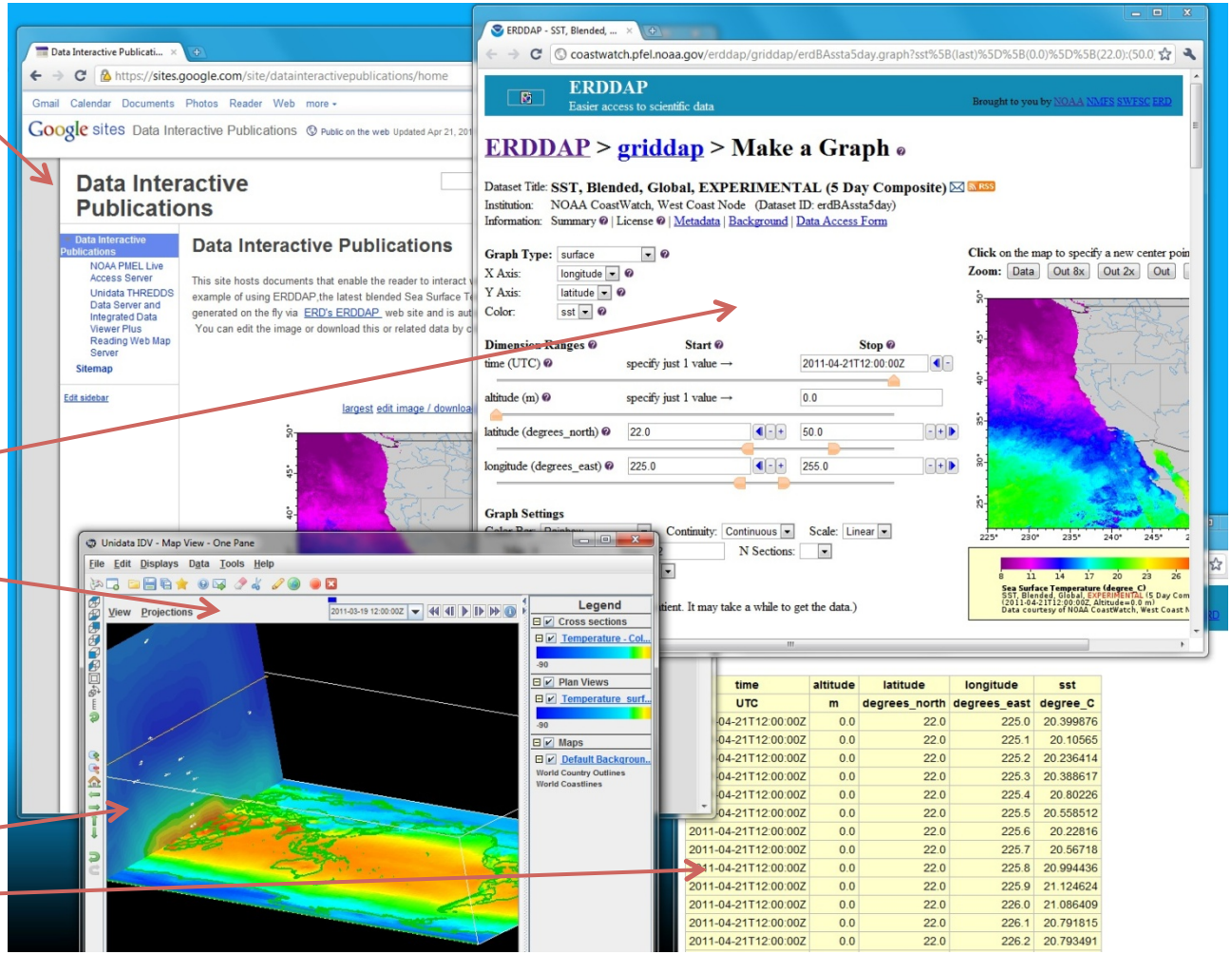
Via embedded links, reader accesses data and analysis and display tools.

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
- Cross-disciplinary: enabling readers of scientific publications and educational modules to work directly with data from other disciplines
- Professional 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 documents and data products
- Search: data interactive publications can be found via Google, Bing, etc. which in turn:
  - leads the searcher to the datasets and analysis tools the publication points to
  - makes it possible to do Google-type rankings based on pointers into and out of the documents and datasets
- 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 productive collaborations
- 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
- 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
- Additional UPC effort: "skunk works?" with collaborators, new or re-programmed staff
- Need to engage publications industry

Implications of Recent Developments

Earlier barriers are being overcome:

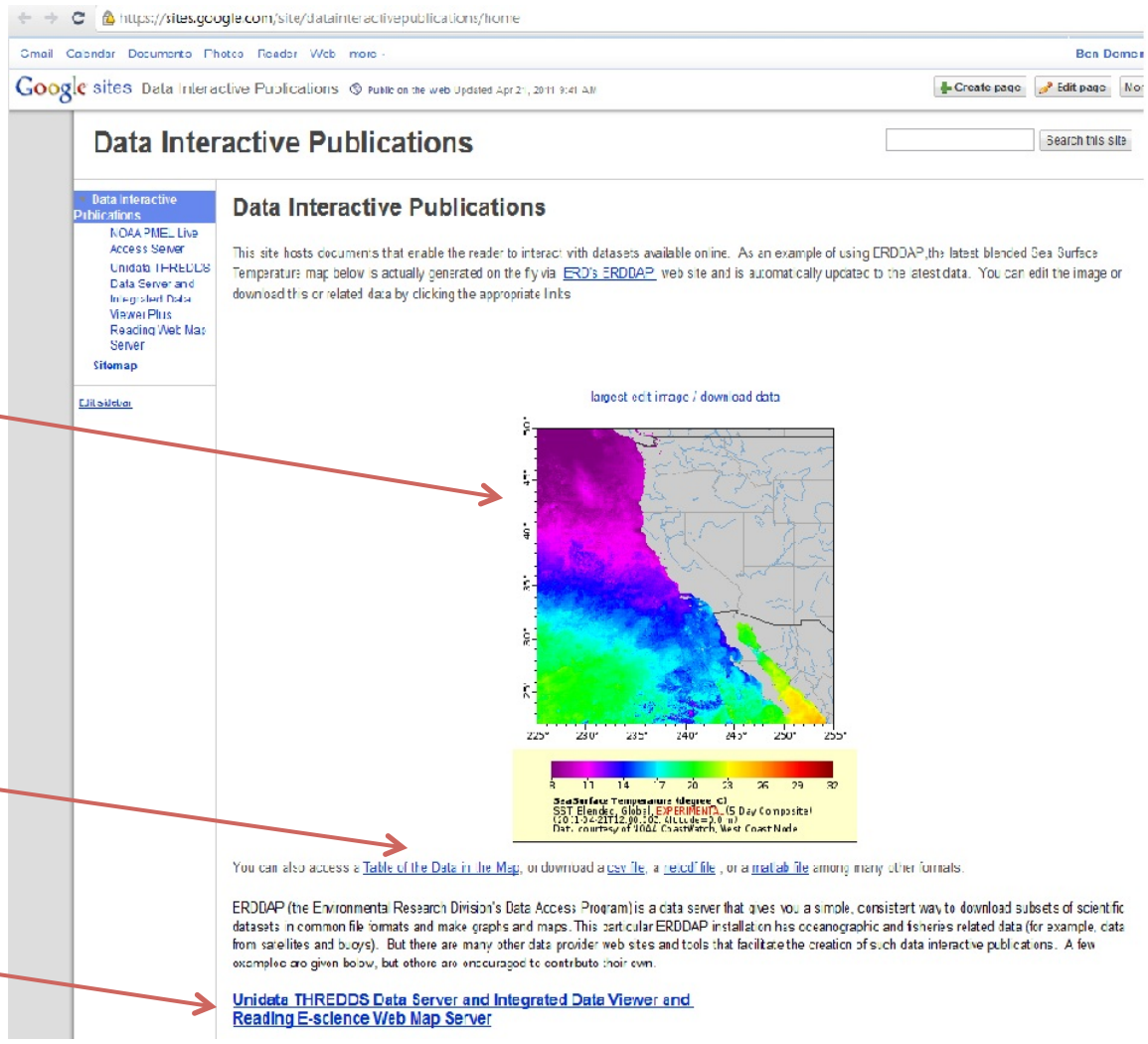
- Browser-based access to interactivity is becoming more common (Live Access Server, TDS/WMS/GODIVA, ERDDAP, FerretTDS…)
- Server-side (cloud?) processing is a viable platform for interactivity
- RAMADDA/TDS provides mechanism for persistent storage of datasets by user community
- Clients with modest 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.

Issues for Unidata Community

- Where does the processing belong?
  - In a desktop application (IDV, McIDAS, AWIPS)?
  - In an app on mobile devices?
  - As a web brokering service in a departmental server?
  - As a web brokering service at the UPC?
  - In the cloud?
  - In Wyoming?
- How much emphasis on Earth System Science as opposed to Mesoscale Meteorology?
- How many different approaches can be supported?

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

coastwatch.pfel.noaa.gov/erddap/griddap/erdBAssta5day.graph?sst%5B(last)%5D%5B(0.0)%5D%5B(22.0):(50.0)%5D%5B(225.0):(255.0)%5D&.draw=surface&.vars=lon

ERDDAP
Easier access to scientific data

Brought to you by NOAA NMFS SWFSC ERD

ERDDAP > griddap > Make a Graph

Dataset Title: **SST, Blended, Global, EXPERIMENTAL (5 Day Composite)** RSS
Institution: NOAA CoastWatch, West Coast Node (Dataset ID: erdBAssta5day)
Information: Summary | License | Metadata | Background | Data Access Form

Graph Type: surface
X Axis: longitude
Y Axis: latitude
Color: sst

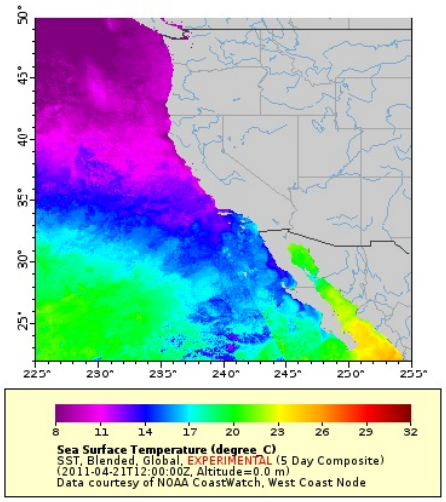
Dimension Ranges	Start	Stop
time (UTC)	specify just 1 value →	2011-04-21T12:00:00Z
altitude (m)	specify just 1 value →	0.0
latitude (degrees_north)	22.0	50.0
longitude (degrees_east)	225.0	255.0

Graph Settings
Color Bar: Rainbow Continuity: Continuous Scale: Linear
Min: 8 Max: 32 N Sections:
Draw the land mask:

Redraw the Graph (Please be patient. It may take a while to get the data.)

Optional:
Then set the File Type: .htmlTable and Download the Data or an Image
or view the URL: http://coastwatch.pfeg.noaa.gov/erddap/griddap/erdBAssta5day.htmlTable?sst[(2
(Documentation / Bypass this form) (File Type information)

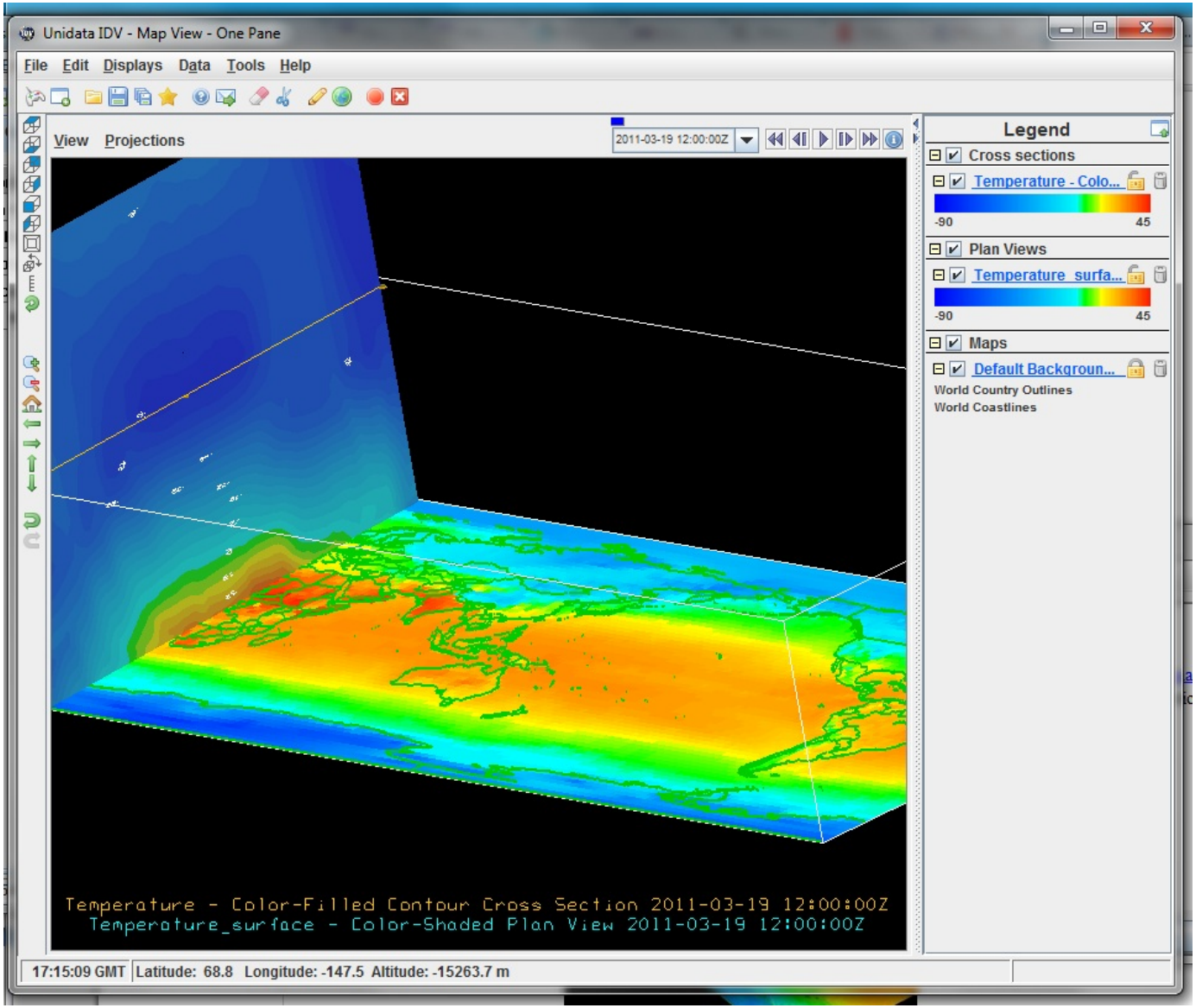
Click on the map to specify a new center point.
Zoom: Data Out 8x Out 2x Out In In 2x In 8x



Things You Can Do With Your Graphs

- Web page authors can embed a graph of the latest data in a web page using HTML <img> tags.
- Anyone can use Slide Sorter to build a personal web page that displays graphs of the latest data, each in its own, draggable slide.
- Anyone can use or make Google Gadgets to display images with the latest data on their iGoogle home page.

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

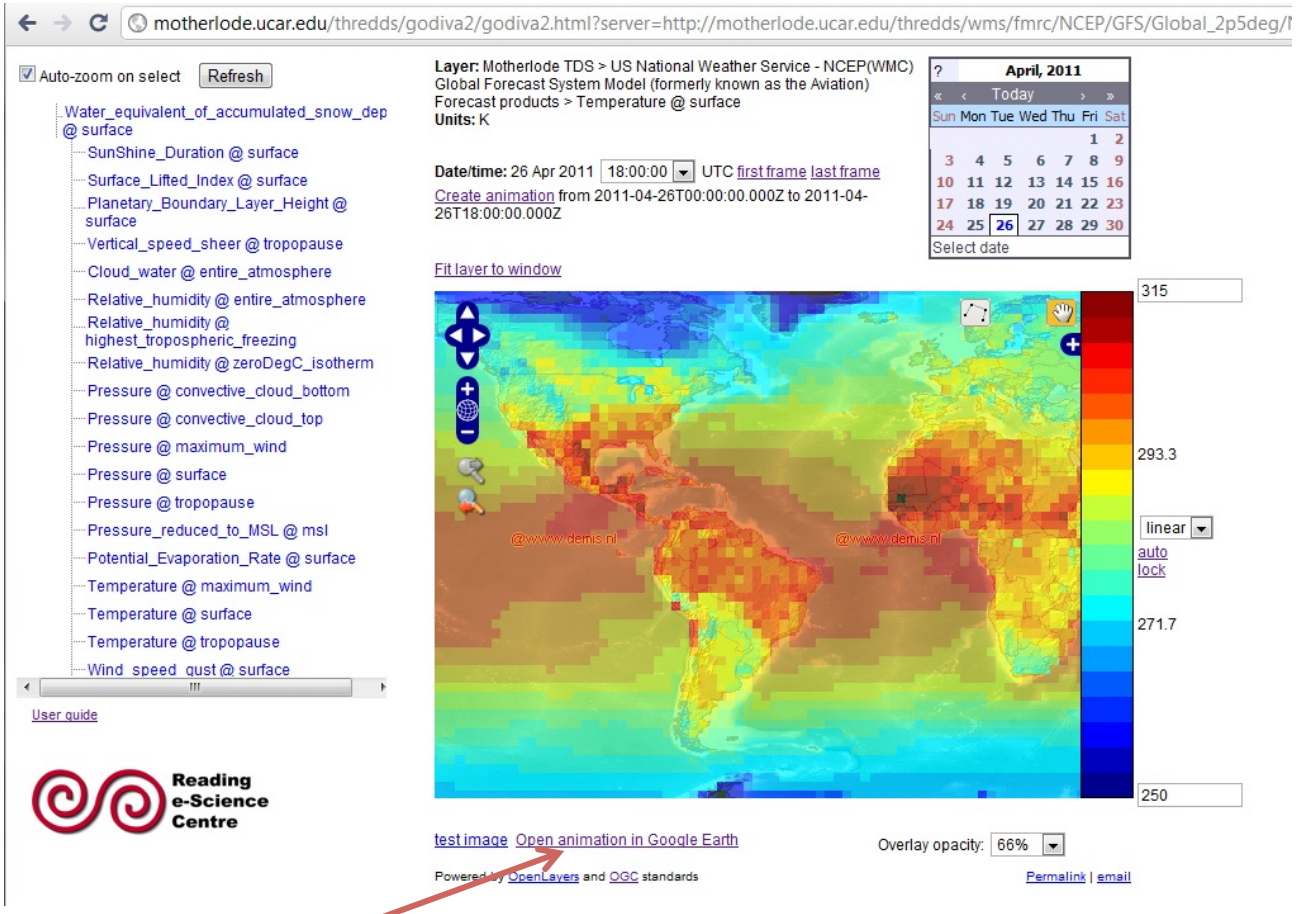
Dataset in WMO GRIB2 files

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

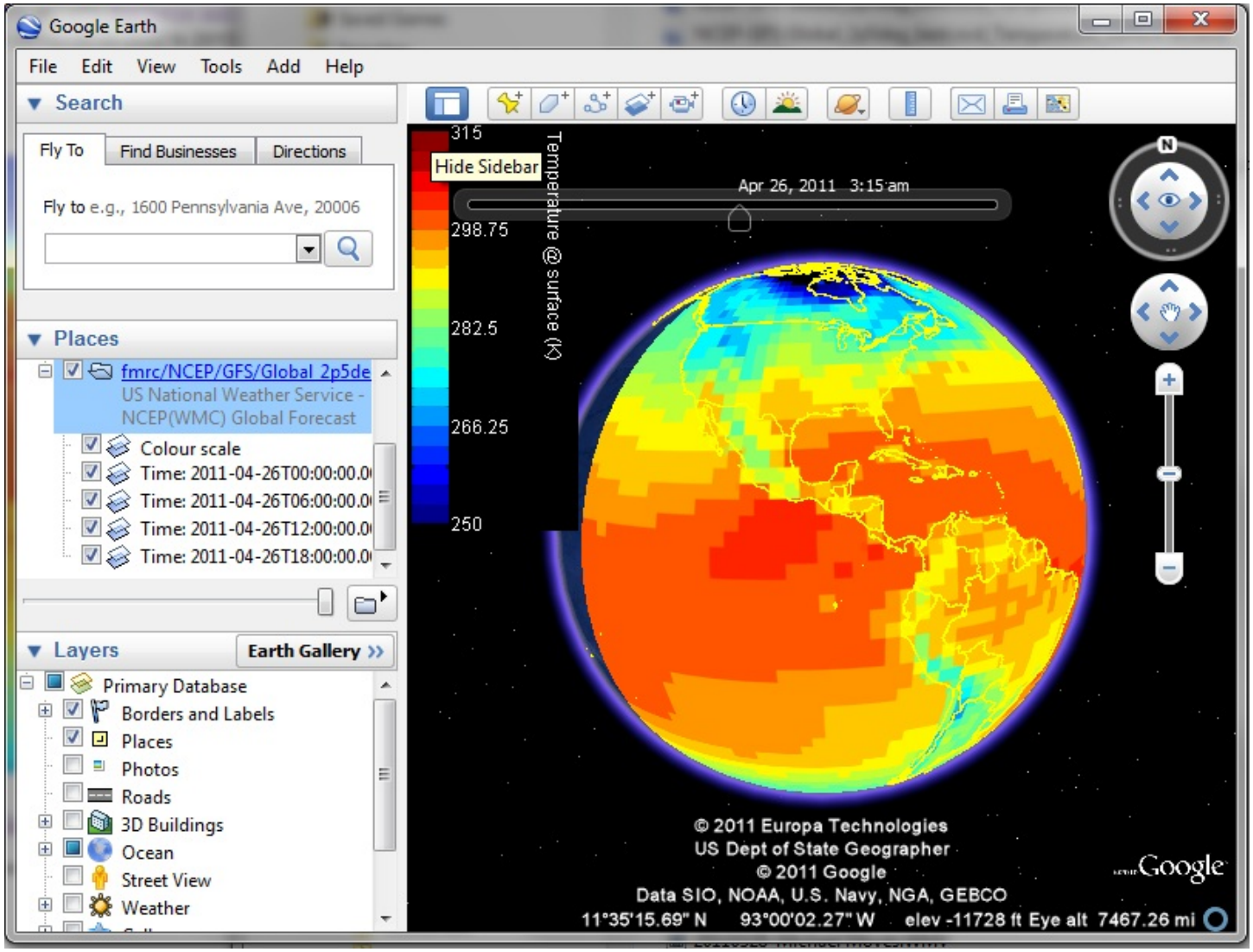
Hundreds of analysis and visualization tools work with Unidata netCDF datasets



Next slide

Google Earth Visualization

Created on Unidata TDS with e-Sciences WMS/GODIVA



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