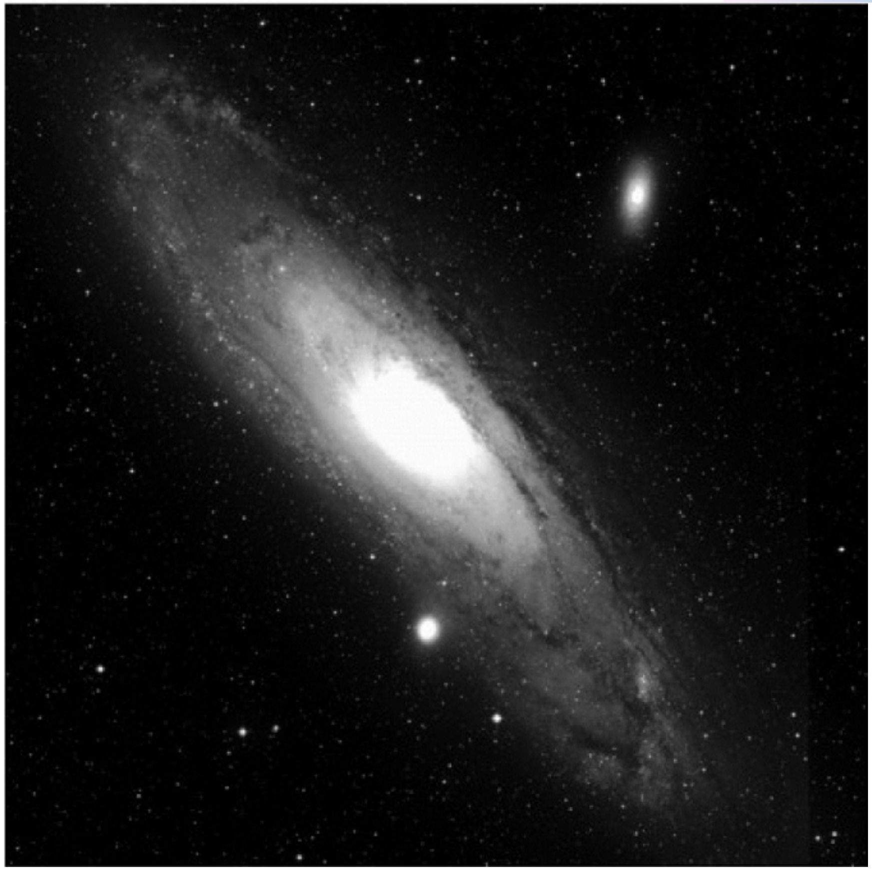
Making Data Make Sense:
Accessing and Visualizing Very Large Data Sets



Daniel S. Katz

Parallel Applications Technologies (PAT) Group

Gary Block, Jim Collier, (Dave Curkendall,) Laura Husman, Joe Jacob, Peggy Li, Craig Miller, Lucian Plesea, Herb Siegel

http://pat.jpl.nasa.gov/

The Scientist's Problem

- Amount of available data is ever expanding
- Gaining knowledge from data is still hard
- One very effective method is to visualize the data
- The Parallel Applications Technologies (PAT) group has been working on helping scientists understand data for many years
- This talk will cover general methods for accessing and visualizing data, as well as highlighting specific examples

General Topics

- ➜ Visualizing your data
  - Examples: Terrain data w/ DLT and RIVA
- Providing images to others
- Accessing and visualizing other people's data

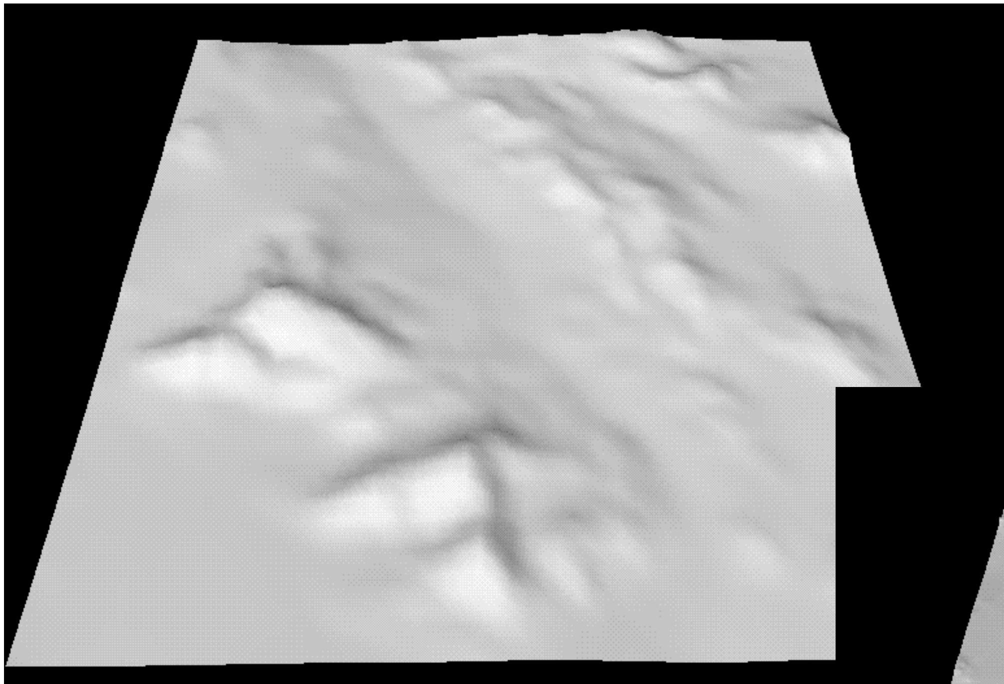
Visualizing Local Data

- Assume a very large data set exists on a local supercomputer
  - It's on the supercomputer because that's where it was generated
- Example: synthetically-enhanced Martian terrain data
  - Single Processor Algorithm by Bob Gaskell/Sec. 312
  - Parallelized by Richard Chen, Craig Miller, Herb Siegel
  - Used as a component of Terrain and Environmental Data Server (TEDS)

TEDS and Terrain Generation

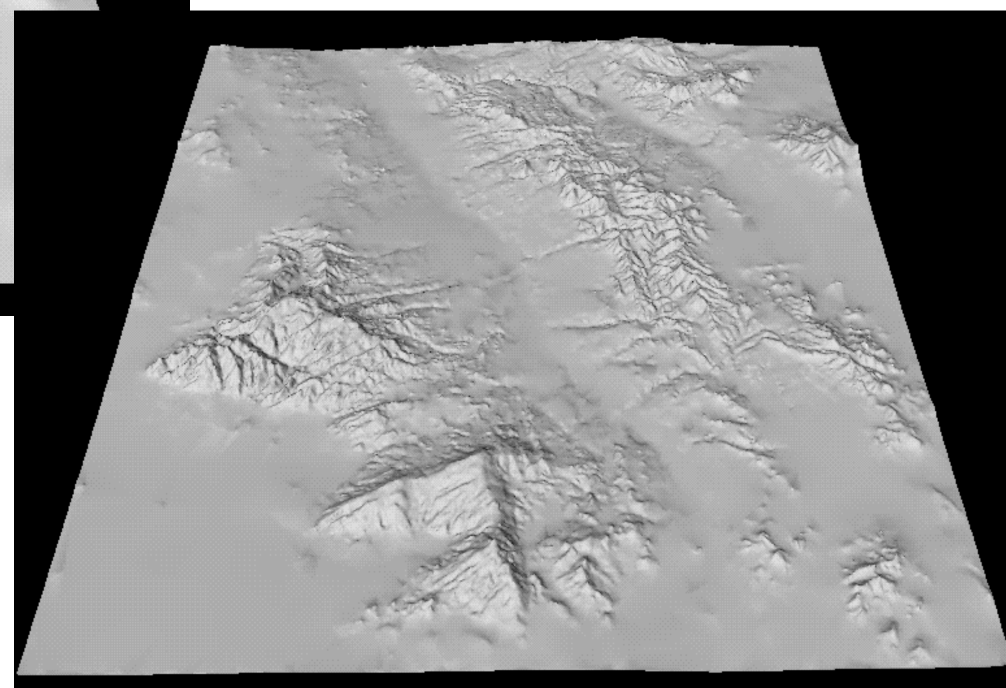
- PAT group is building a Terrain and Environmental Data Server (TEDS)
- Intended to be a 24/7 service for anyone doing Mars simulations involving terrain, that will:
  - Include terrain storage, generation, enhancement, and access
    - Storage and access includes:
      - Measured terrain (Mars Yard, field sites, etc.)
      - Other modeled terrain (Mars sites, etc.)
  - Interface with various tools, including MarsTERM (ROAMS), instruments models (from M. Lee), Mission Simulation Framework (Ames), etc.
- Parallel computing incorporated as needed, specifically for generating terrain and enhancing terrain

Ideal Enhancement

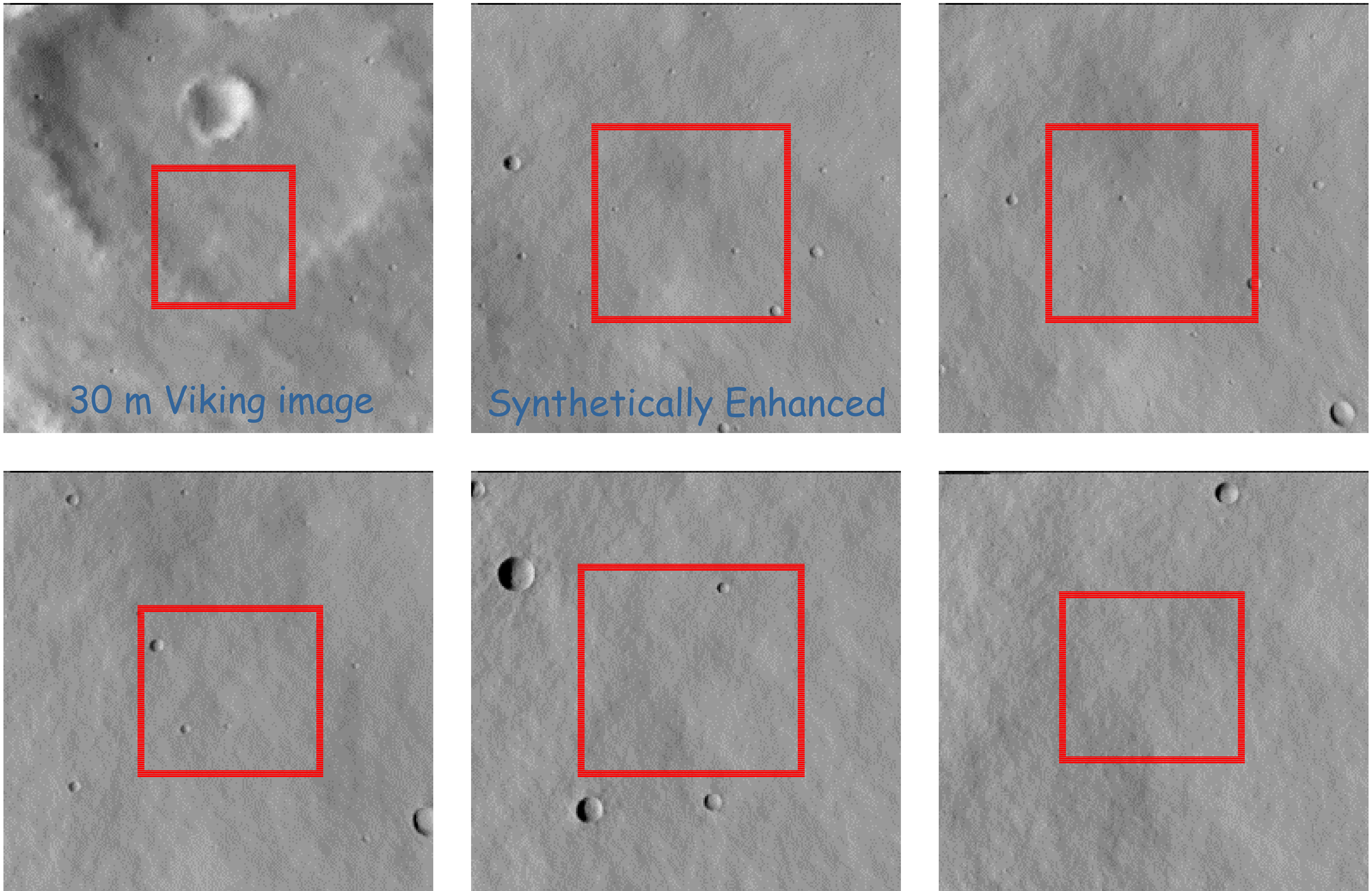


Input: Coarse Data (i.e., MOLA)

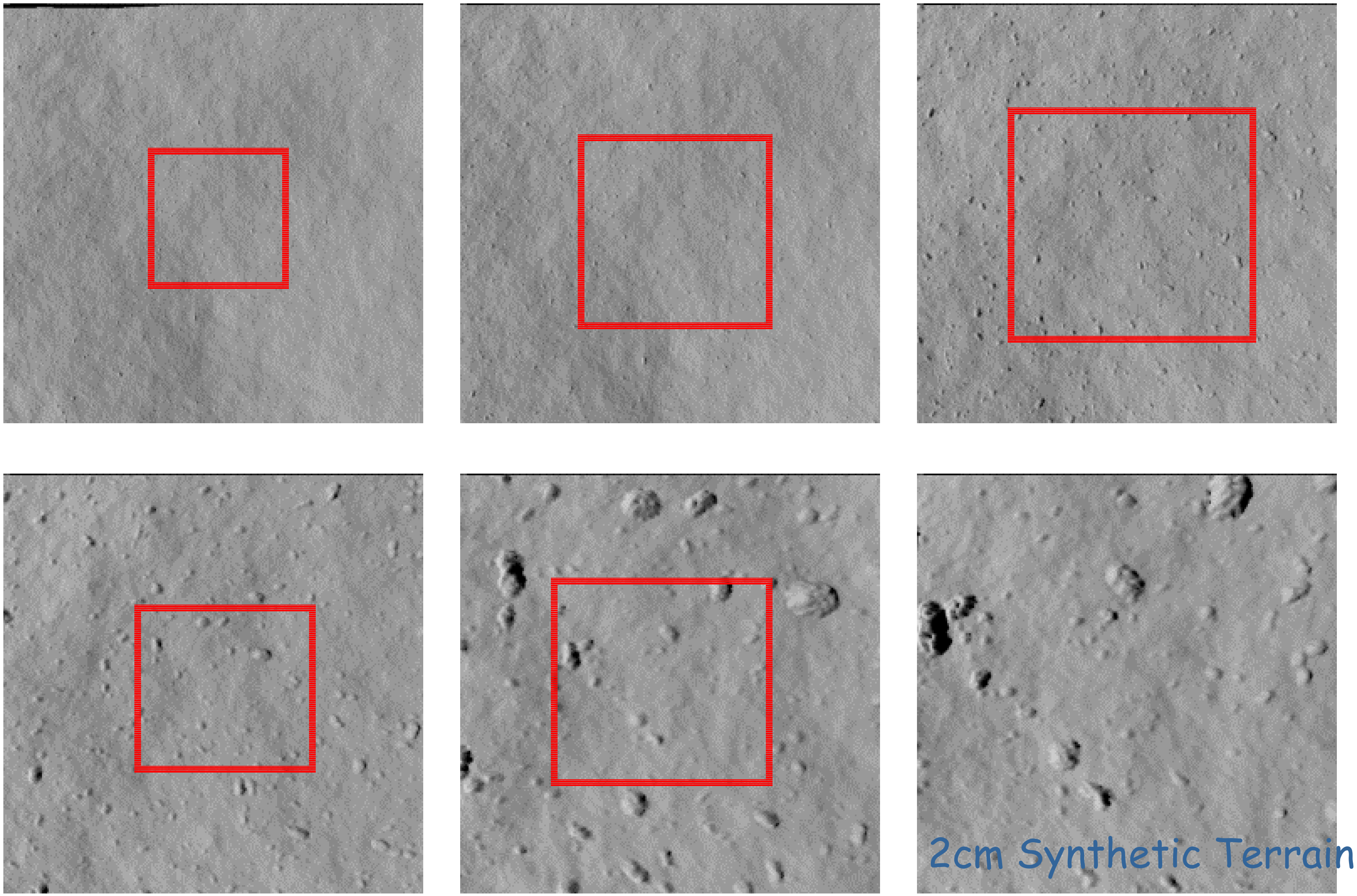
Output: Synthetically Enhanced



Synthetic Terrain: Starting with what we know…



…and adding detail

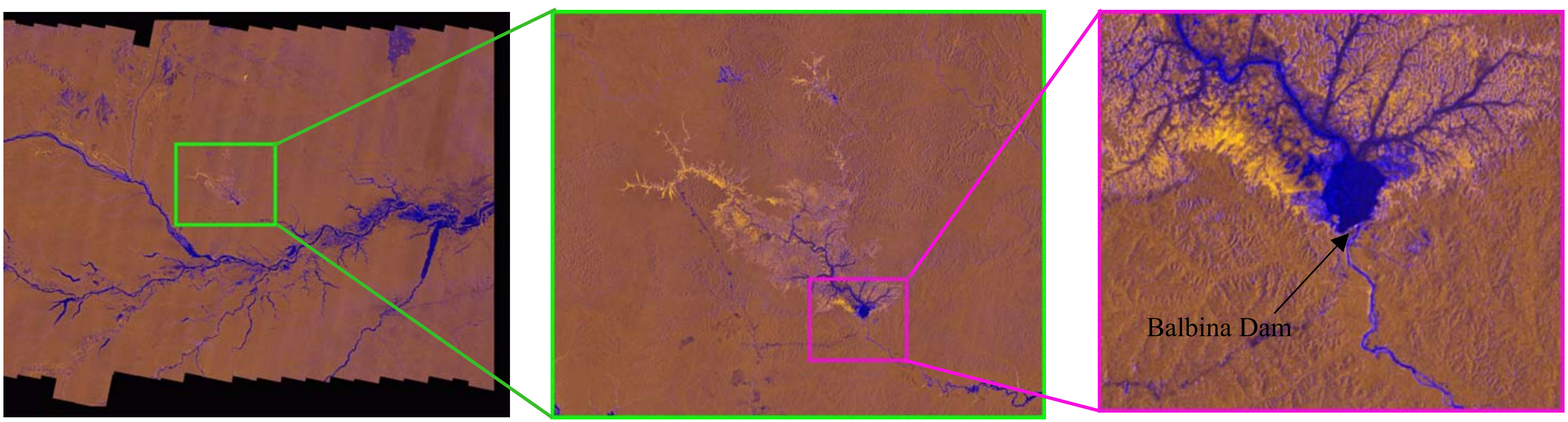


A Sample Synthetic Terrain

QuickTime™ and a YUV420 codec decompressor are needed to see this picture.

Digital Light Table (DLT)

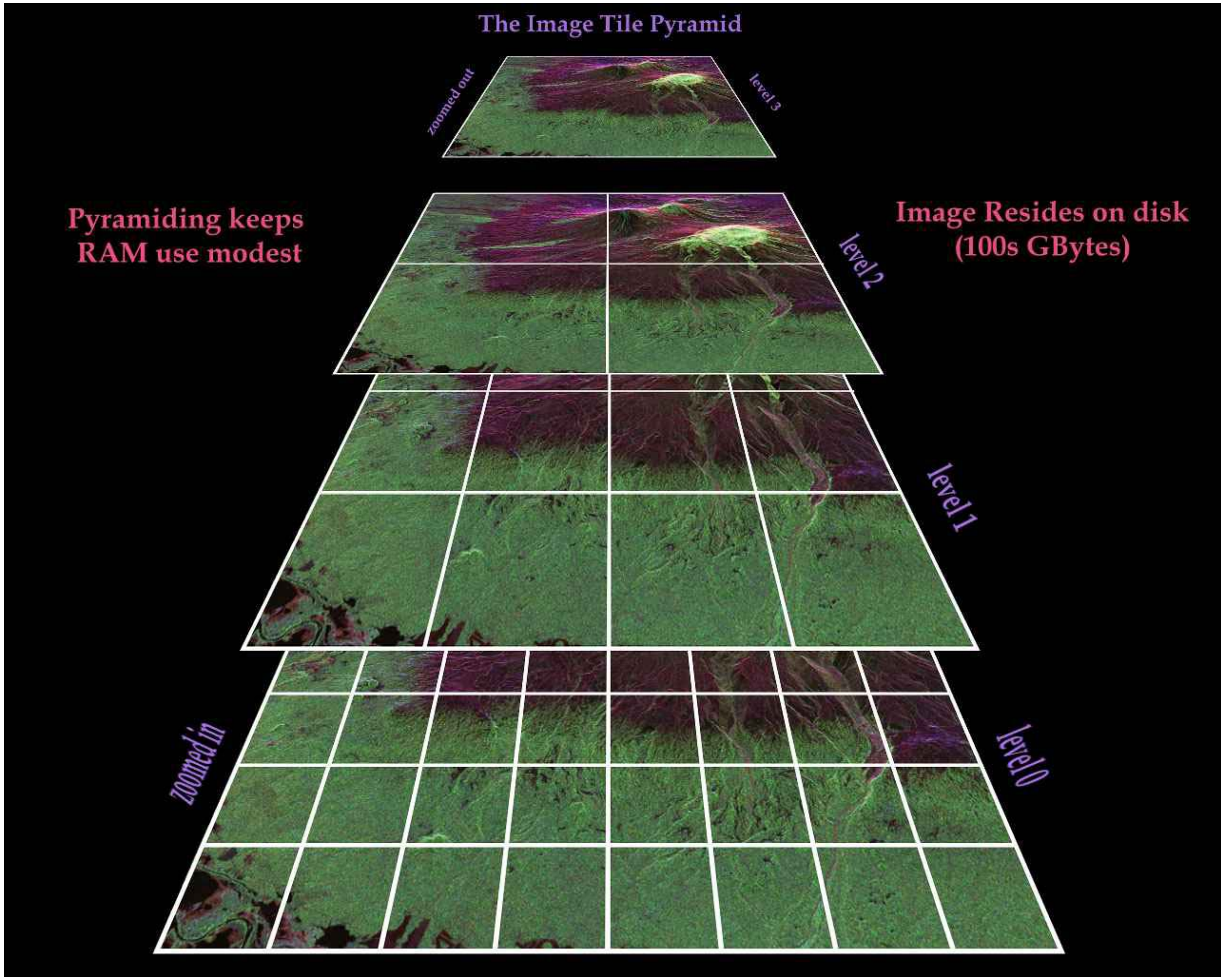
- Previous animation built using Digital Light Table (by Herb Siegel and Craig Miller)
- Built for fast visual interactive access to very large data sets, including terrain data w/ elevation
- Originally built for viewing JERS-1 Amazon mosaic (mosaic created by Paul Siqueira and Bruce Chapman)
- Uses graphics hardware for fast pan and zoom



DLT Disk Storage

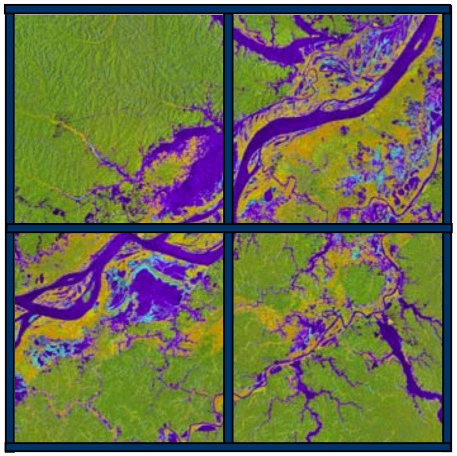
To permit smooth pan and zoom, the DLT needs to be able to quickly access data at the appropriate resolution

- The DLT uses the Image Tile Pyramid is to store the input data on disk:
- The original data is tiled.
- Each level has tiles 1/4 the resolution of prev. level
- This tiling **allows the DLT to smoothly pan and zoom** by only using the proper level's data in order to keep the output screen(s) updated



DLT Views and Collaboration

Three Ways to Use the DLT:



Multi-Screen Auditorium

One Scientist

Multiple Scientists Discussing an Image (Anyone Can Control the DLT, Other Displays are Mirrored)

DLT Architecture

For multiple displays, each graphic engine selects the tiles and performs the rendering for its display:

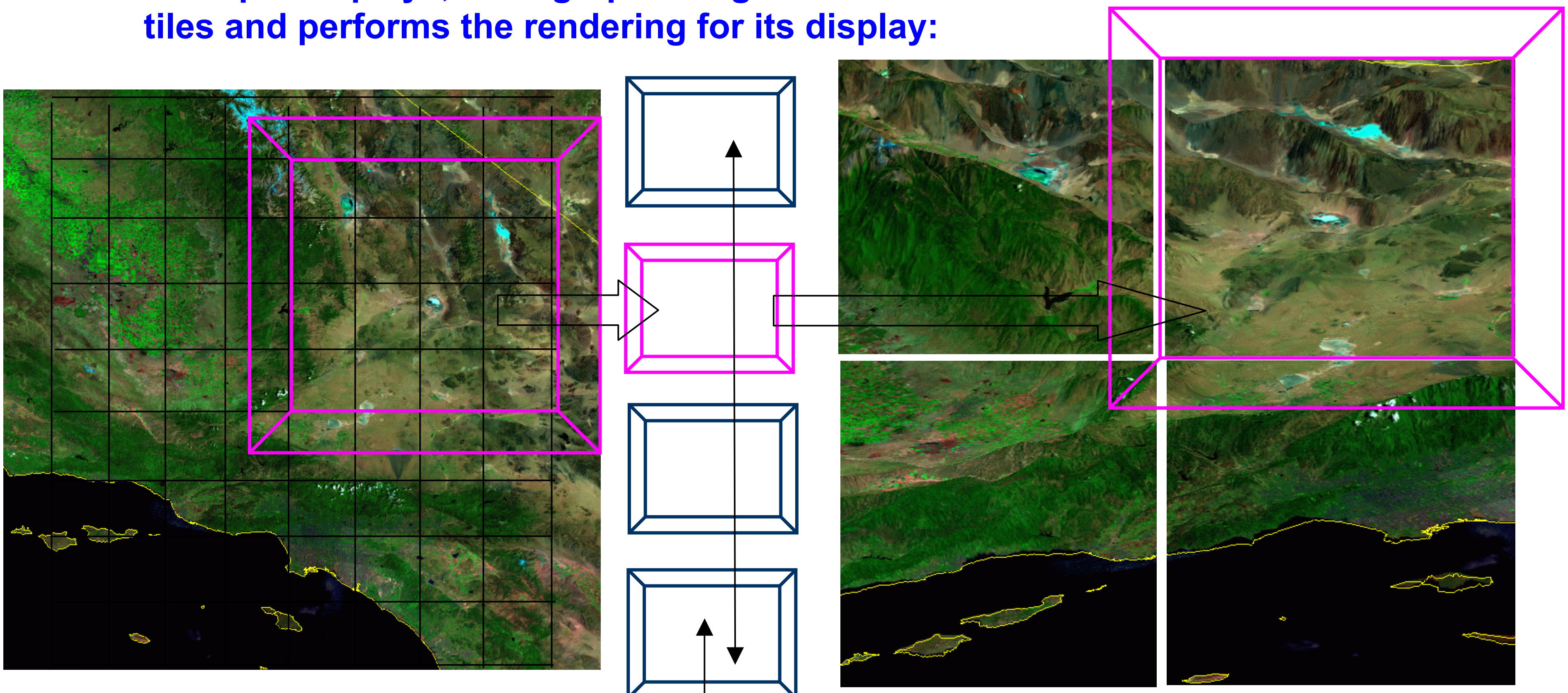
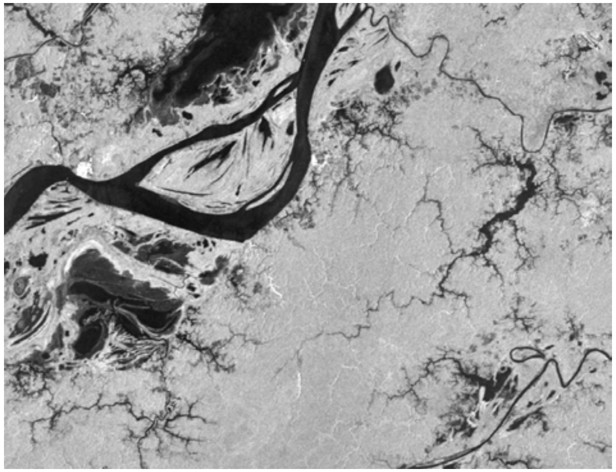


Image and Map layers, tiled at multiple resolutions: Appropriate tiles are selected & then read by graphic engines.

Parallel Graphic Engines: Commands & live data sent to any engine are automatically shared by all graphic engines.

Displays

DLT's Change Detection Capability



Dry season image

Mapped to blue

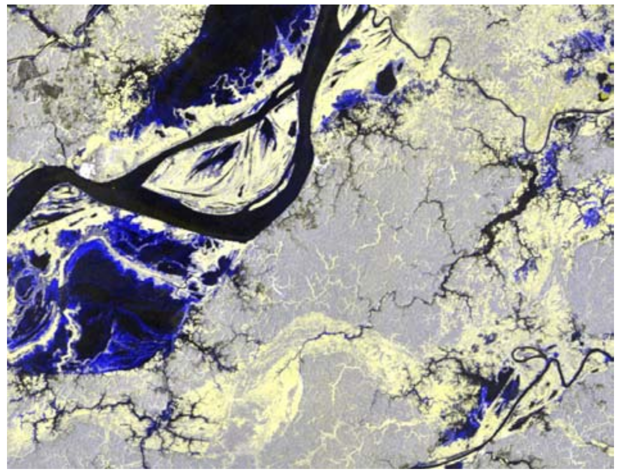
Multi-season image, where

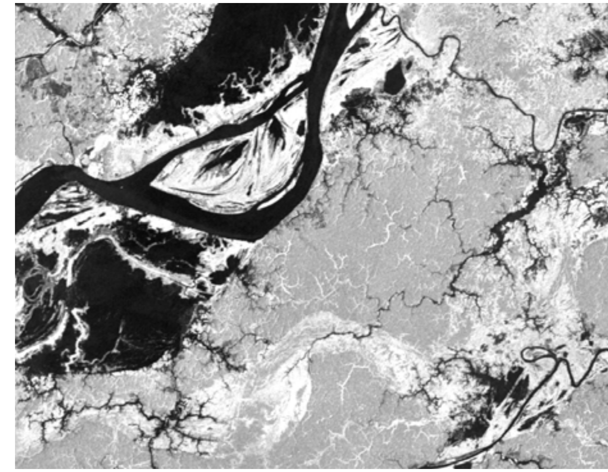
Black = covered with water in both seasons

Grey = not covered with water in either season

Yellow = inundated in wet season

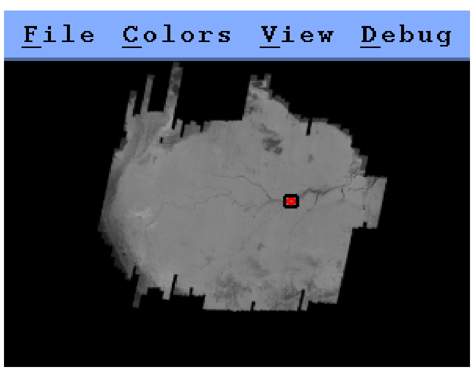
Blue = marshy in dry season



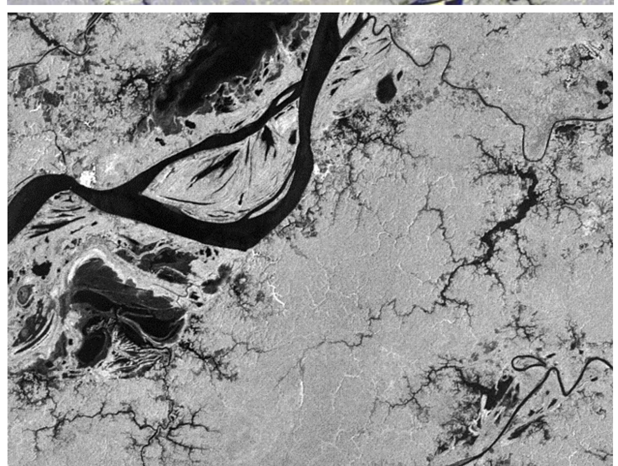


Wet season image

Mapped to green and red (yellow)



Viewer GUI tool, showing the JERS-1 Amazon SAR mosaic



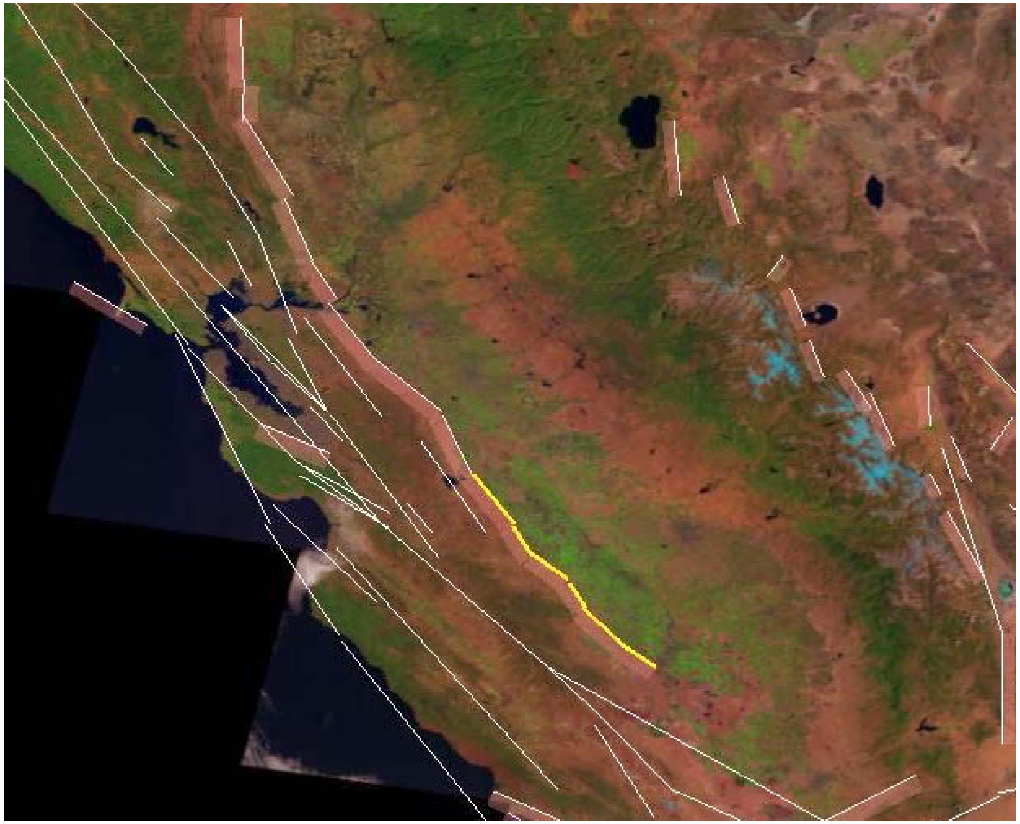
Alternating the wet and dry Images - flash capability

Correspondence of Image Data and Metadata

Catalog Data

Fault Name string	end1-x degree	end1-y degree	end2-x degree	end2-y degree	len km	wid km	dip deg
Great Valley 9 (r, 15 W)	-120.75	36.81	-120.99	37.10	39	10	15
Great Valley 10 (r, 15 W)	-120.65	36.63	-120.76	36.80	22	10	15
Great Valley 11 (r, 15 W)	-120.45	36.50	-120.65	36.64	25	10	15
Great Valley 12 (r, 15 W)	-120.35	36.34	-120.44	36.48	17	10	15
Great Valley 13 (r, 15 W)	-120.15	36.14	-120.37	36.34	30	10	15
Great Valley 14 (r, 15 W)	-119.95	36.00	-120.16	36.14	24	10	15
Antelope Valley (n, 60 E)	-119.48	38.51	-119.57	38.78	31	15	60
Genoa (n, 60 E)	-119.81	38.67	-119.85	39.10	50	15	60
Hartley Springs (n, 60 E)	-119.00	37.64	-119.08	37.85	25	15	60
Mono Lake (n, 60 E)	-119.10	37.93	-119.19	38.15	26	15	60
Robinson Creek (n, 60 SE)	-119.32	38.21	-119.23	38.33	17	15	60
Great Valley 1 (r, 15 W)	-122.28	39.28	-122.30	39.68	44	10	15
Great Valley 2 (r, 15 W)	-122.29	39.09	-122.28	39.29	22	10	15
Great Valley 3 (r,15 W)	-122.00	38.67	-122.28	39.12	55	10	15
San Andreas - Coachella (rl-ss	-115.71	33.35	-116.48	33.92	95	12	90
San Andreas - San Bernardino (	-116.48	33.92	-117.53	34.31	107	18	90
San Andreas (southern) (rl-ss)	-115.71	33.35	-117.53	34.31	203	12	90
San Andreas - Mojave (rl-ss)	-117.53	34.31	-118.50	34.70	99	12	90
San Andreas -Carrizo (rl-ss)	-118.51	34.70	-119.86	35.31	145	12	90
San Andreas-Cholame (rl-ss)	-119.86	35.31	-120.29	35.75	62	12	90
San Andreas Parkfield Segment	-120.29	35.75	-120.56	36.00	37	12	90
San Andreas (1857 rupture) (rl	-117.53	34.31	-120.56	36.00	345	12	90

California fault database

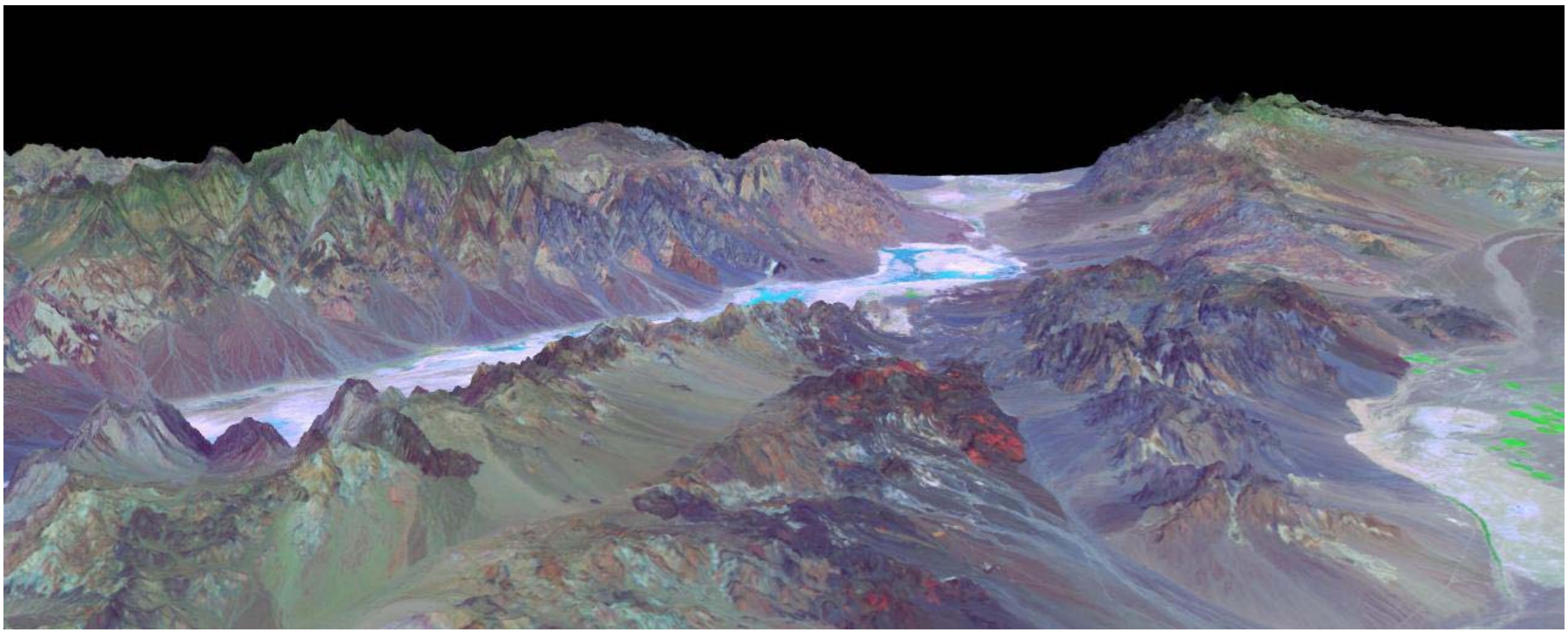


California Landsat with fault segment overlay

- Information in a catalog is tied to associated locations in an image
- Image to Catalog: User may select a region and see the catalog entries for those objects in that region highlighted both in the image and in the catalog window
- Catalog to Image: User may select a catalog entry and see that object highlighted in the image or jump to the position of that object in the image

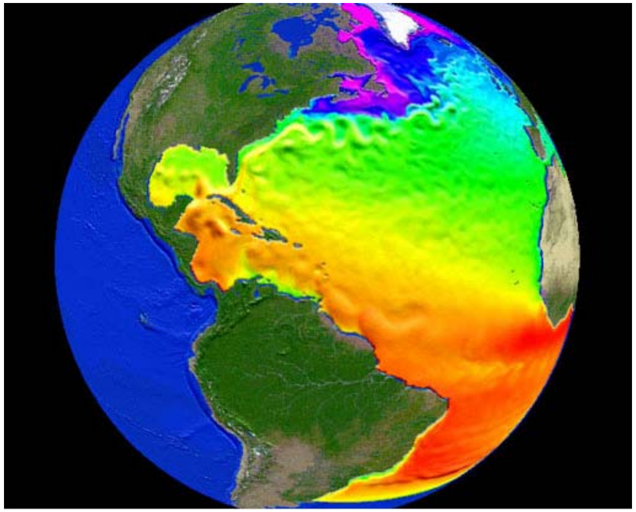
RIVA: Another Terrain Data Visualization Tool

- Peggy Li's Remote Interactive Visualization and Analysis (RIVA) System
- Scalable, parallel software rendering for 3-D planetary data

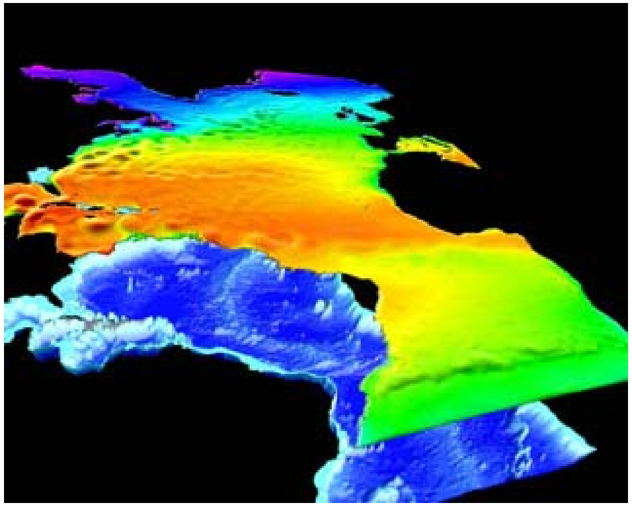


RIVA Features

- Spherical data representation capable of global data sets and regional terrain data sets
- Scalable to large number of processors, large input data sets and large output images
- Generate high-resolution still images and animations (IMAX and HD formats)
- Support out-of-core rendering for data sets bigger than the physical memory



Ocean surface temperature overlay on top of a global earth image



Ocean surface temperature separated from the ocean bottom topo map using zbuffer

RIVA Features (2)

- Multiple surface rendering with different resolution, different format, and different coverage, compositing surface using zbuffer or alpha-blending
- Distributed and interactive data exploration and visualization
- Animation of time-varying simulation data set using out-of-core rendering technique
- Batch mode movie production

(a)

30 meter Landsat, bands 7, 4, and 2

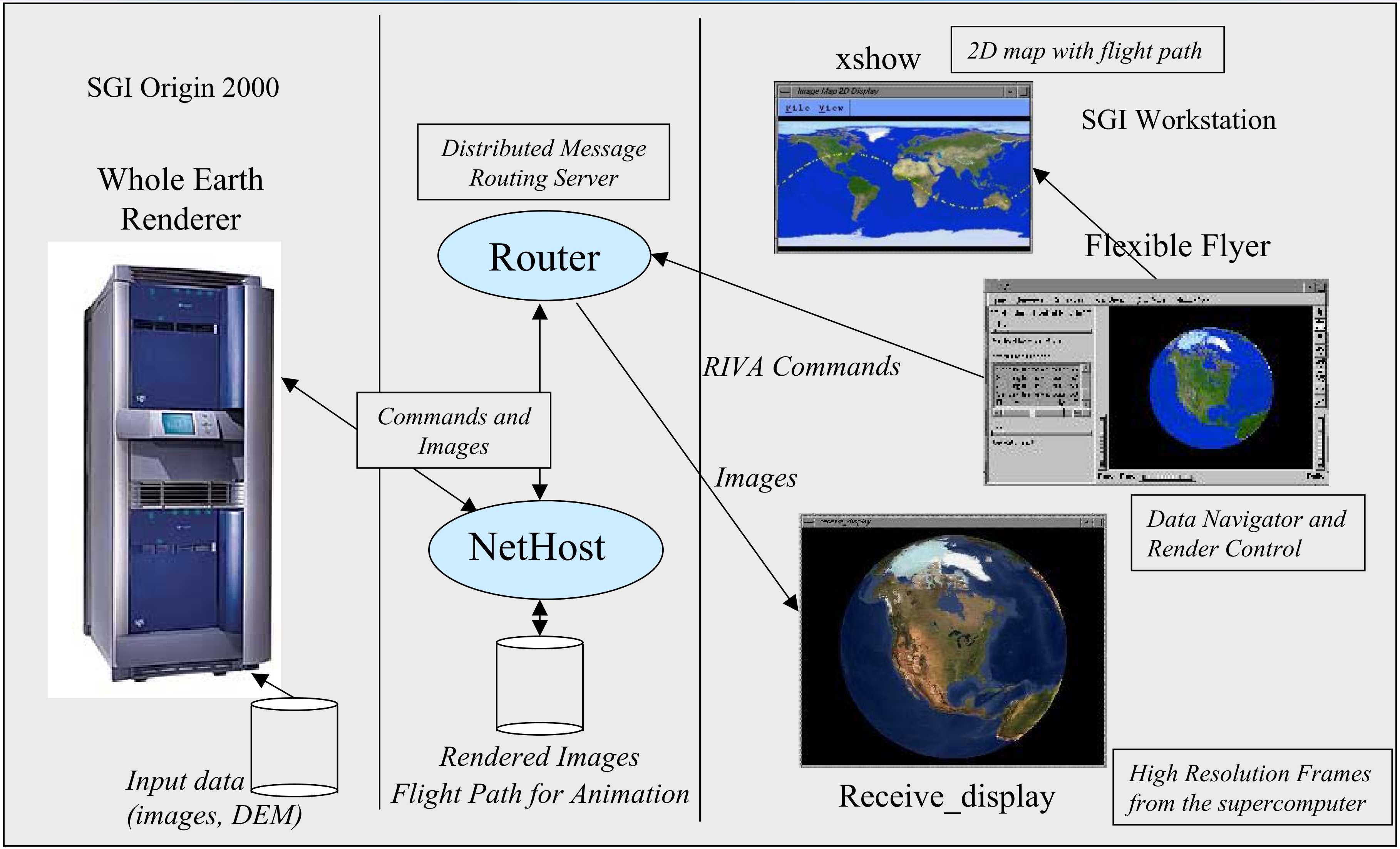
(b)

Two images blended with the opacity of 2.25 meter image set to 0.58

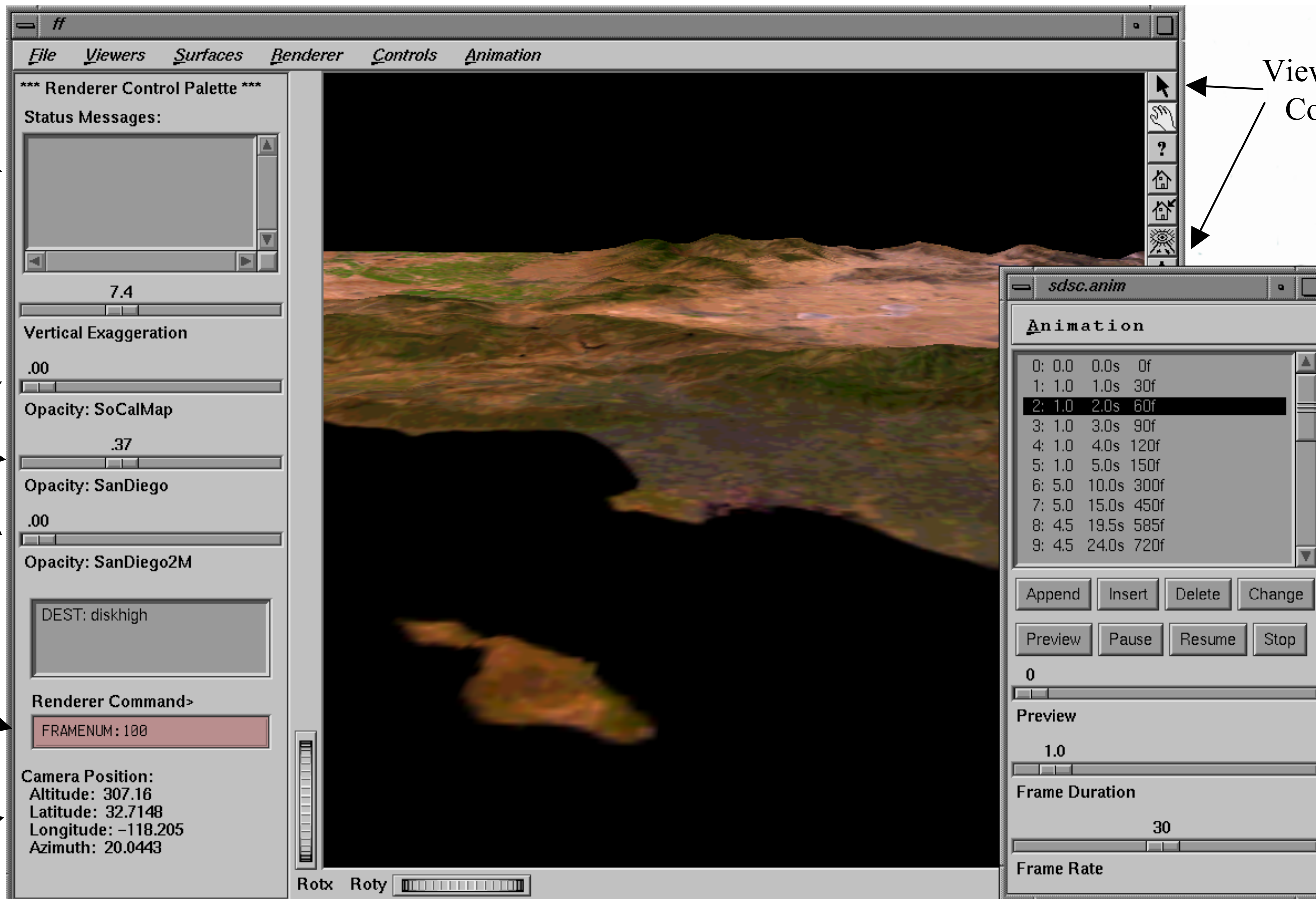
(c)

2.25 meter grayscale satellite image

RIVA System Architecture



RIVA User Interface – Flexible Flyer



Status message from the renderer

Vertical Exaggeration Slider

Opacity control – One per dataset

Typing rendering command directly

Current camera position and look angle

Viewpoint Control

Key Frame List

Preview & Edit Control

Preview Frame Slider

Frame duration In seconds

Frame rate per second

Control Panel ***Navigation Window*** ***Key Frame Editor***

RIVA Movie

A Fly-over movie using four datasets: a southern California map, a 154 meter resolution Southern California image and a 30 meter LandSAT image of San Diego, ending with 2.25 meter resolution data of San Diego.

General Topics

- Visualizing your data

➜ Providing images to others
  - Example: MAPUS

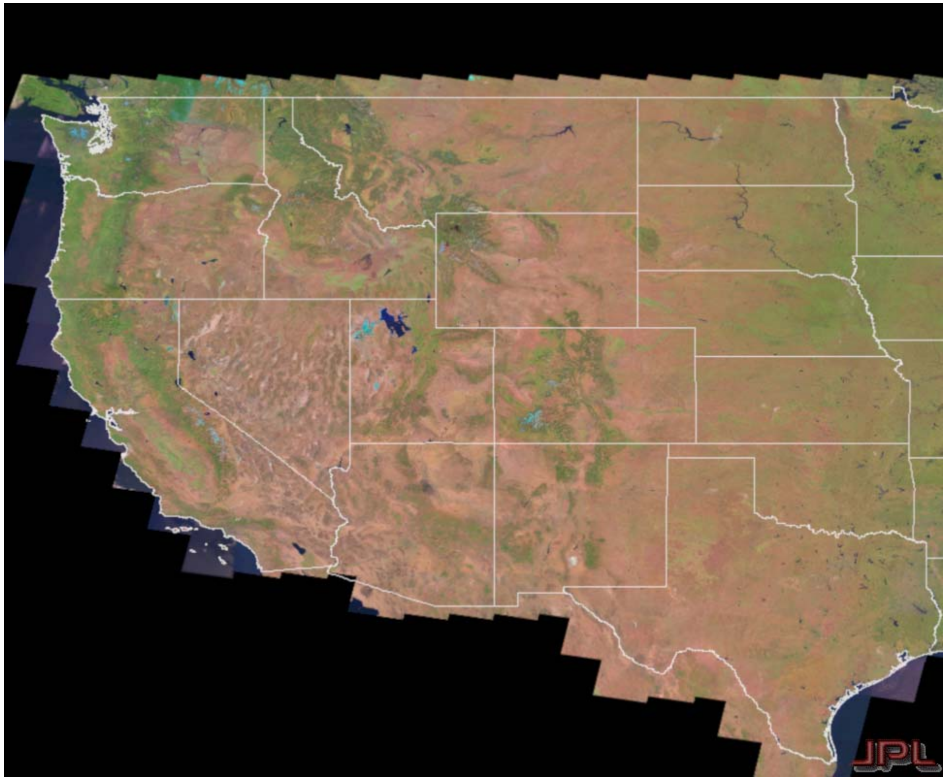
- Accessing and visualizing other people’s data

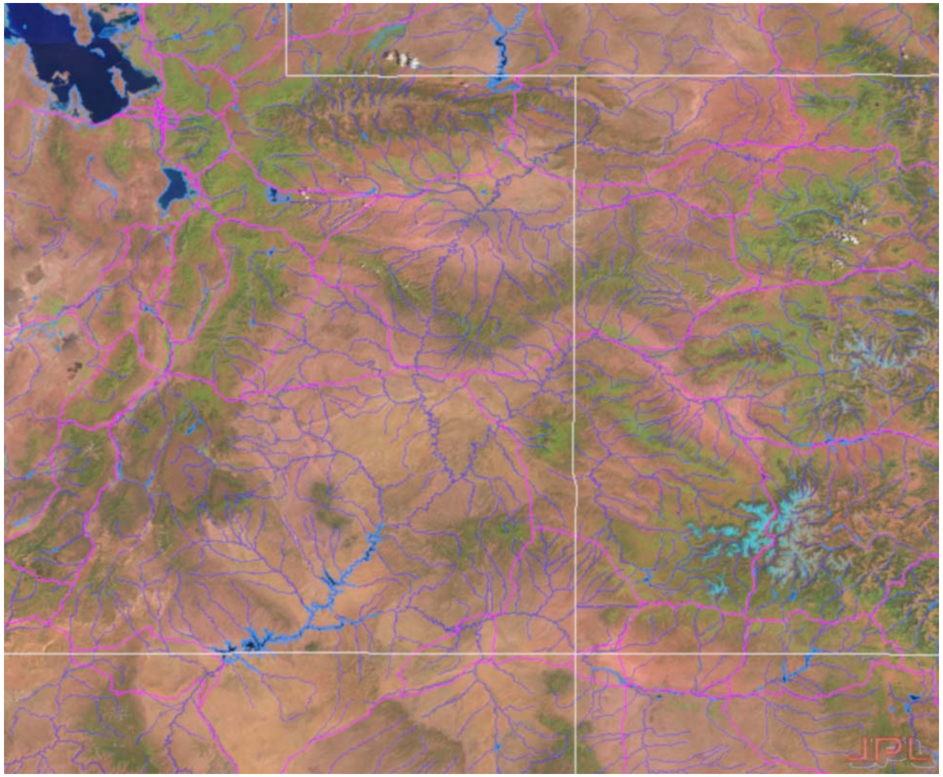
Sharing Data Visually

- The obvious choice today:
  - Deliver images through a web browser
  - Allow users to decide what images to build
- Example: Lucian Plesea's MAPUS (http://mapus.jpl.nasa.gov/)
  - Starting with a 180 GB mosaic of 30 m Landsat data, including:
    - Access to all 6 bands
    - Digital Elevation Model (DEM)
    - Can overlay political boundaries, roads,rivers…
  - GIS Web Map Server (WMS)
  - Changing into MapEarth
    - Global Landsat 7 mosaic being built now…
    - Currently has ~5% of Earth

MAPUS Custom Image Examples (1)

- Progressive overlay capability
- More geopolitical info added as user zooms in

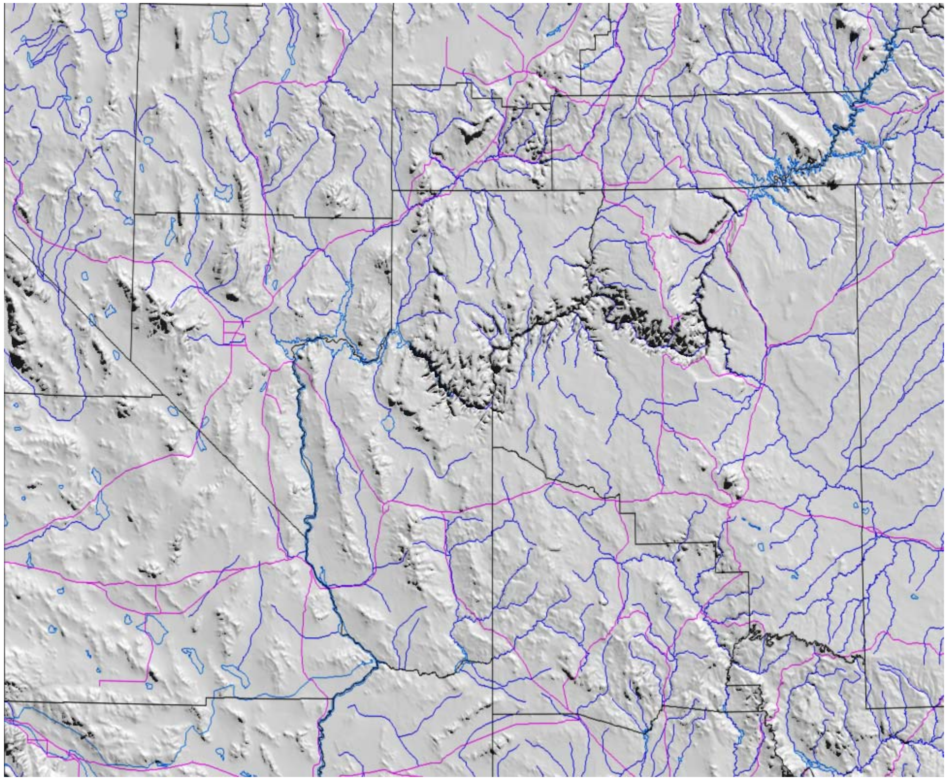


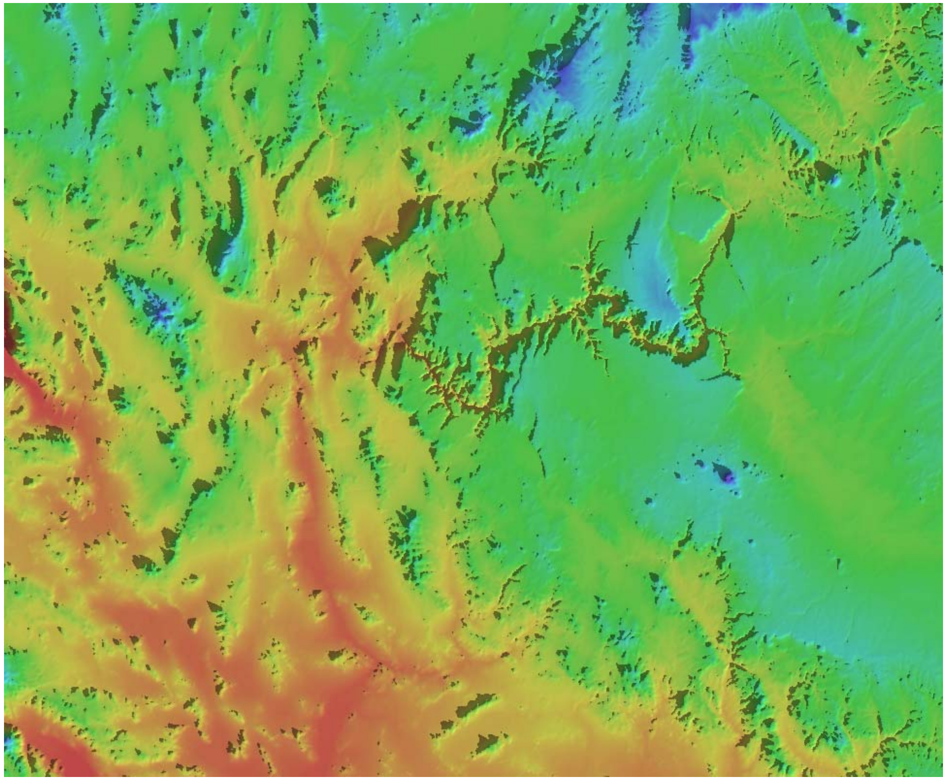


MAPUS Custom Image Examples (2)

Different ways of looking at the topology:

1. Gray slate lit from a controllable angle (NW in this case), which generates shadows
2. Color coded image, also with shadows. Light is from SE. Color generated at run-time by mapping DEM to hue

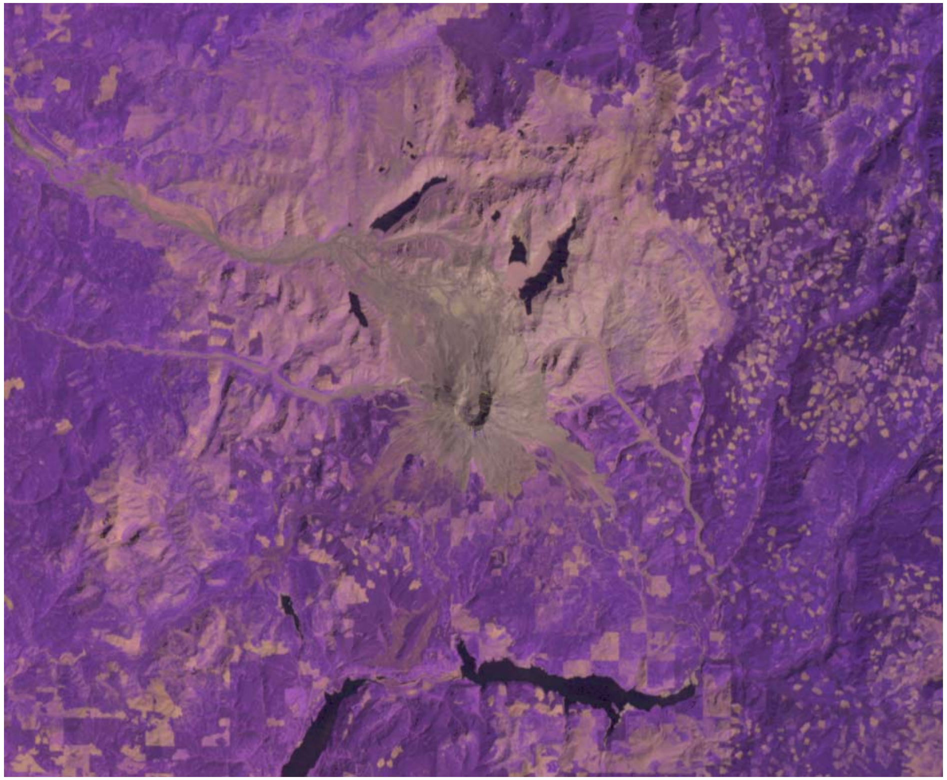




MAPUS Custom Image Examples (3)

Different bands and projections:

- Pseudo-color and cylindrical projection vs. IR color mapping, sinusoidal projection (w/ enhanced shadows)



General Topics

- Visualizing your data
- Providing images to others

➜ Accessing and visualizing other people's data

- Examples: *yourSky* and *OurOcean*

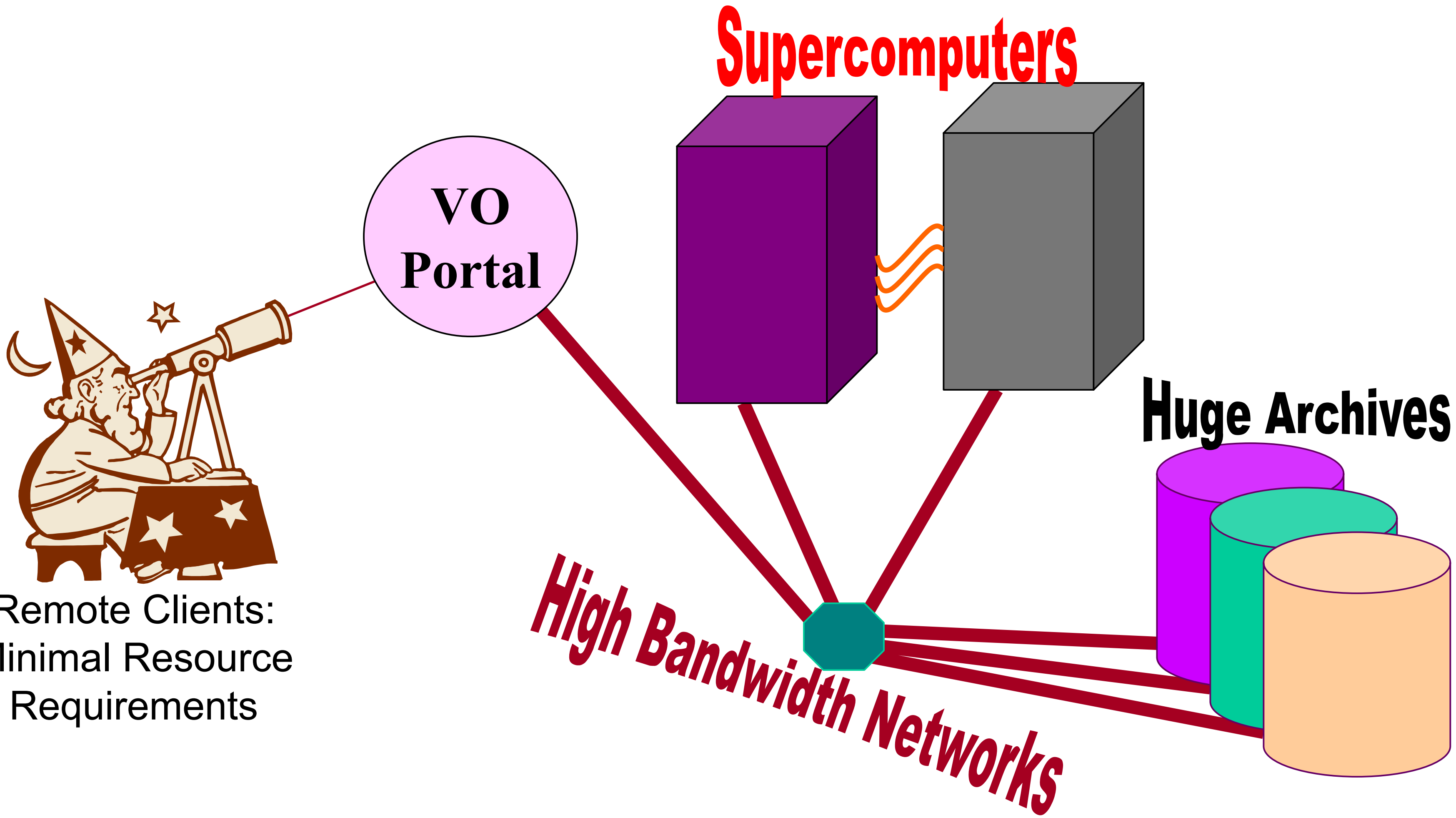
Accessing and Visualizing Remote Data

- Two examples:
  - *yourSky* (http://yourSky.jpl.nasa.gov/)
    - By Joe Jacob, Dave Curkendall, and Gary Block
    - Builds custom image mosaics using astronomical survey data
  - *OurOcean* (http://OurOcean.jpl.nasa.gov/)
    - By Peggy Li, Joe Jacob, Yi Chao, and Zhijin Li,
    - Supports data retrieval, data archiving, data processing, data distribution
      - Focuses on East Pacific Ocean wind
      - Supports difference output data formats, including images

yourSky (http://yourSky.jpl.nasa.gov/)

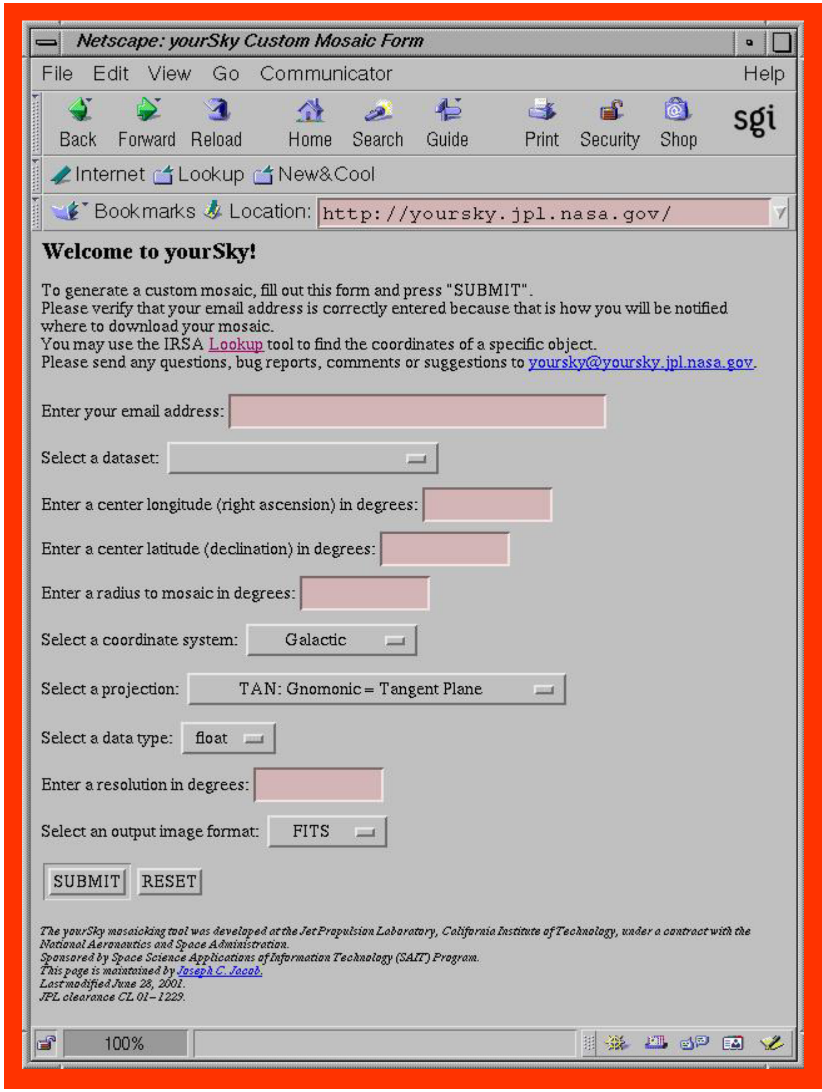
- Huge data sets of astronomical data, such as 2MASS, DPOSS, SDSS are (coming) available
- NSF National Virtual Observatory (NVO) project aims to use such repositories to form a virtual observatory
  - Research results from on-line data will be just as rich as that from “real” telescopes
- Need tools with simple interfaces to allow custom access to data, including derived products
  - Simple Interface = web browser
  - Derived product = image mosaics
  - Custom access = user specified dataset, location, size, resolution, coordinate system, projection, data type, and image format

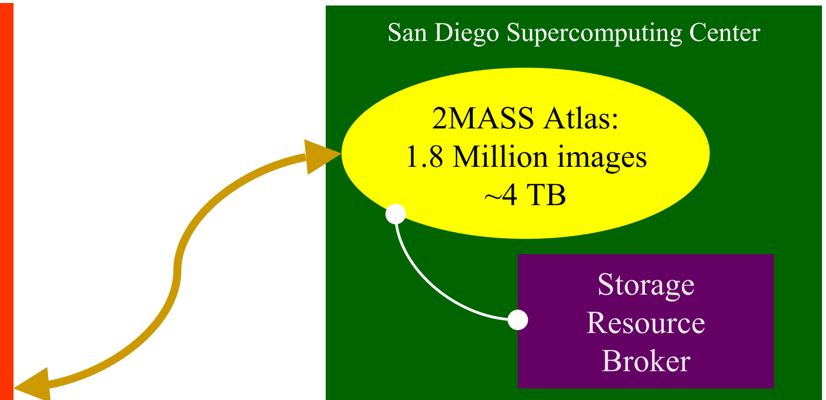
Making the Virtual Observatory Easier to Use Than a Telescope…



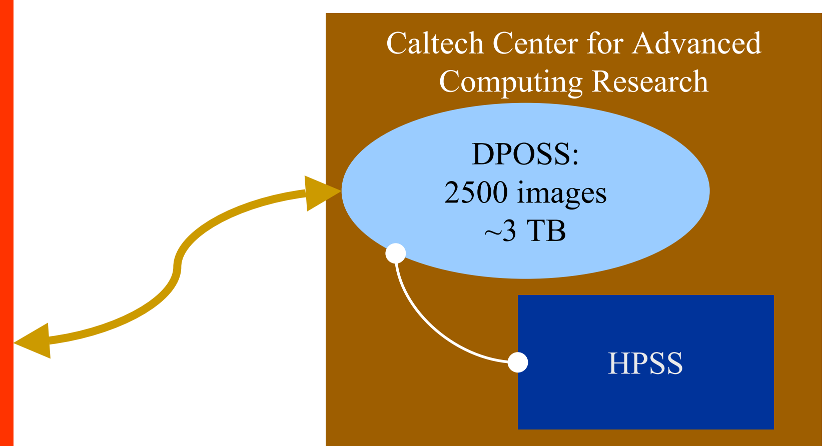
Remote Clients:
Minimal Resource
Requirements

yourSky's Current Form Interface

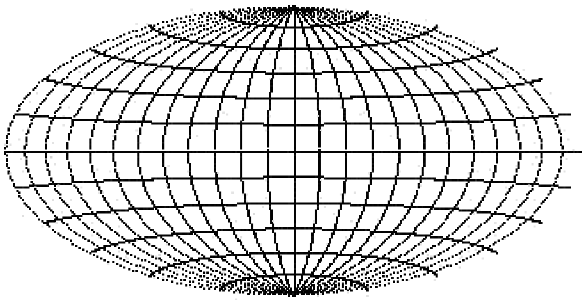




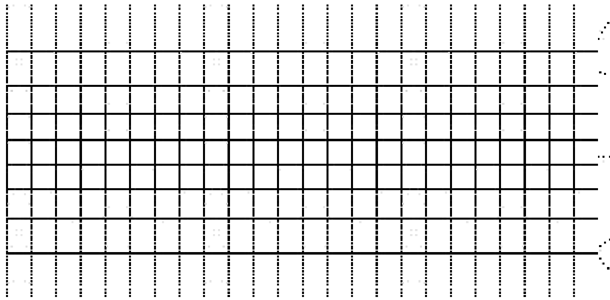
yourSky can access all of the publicly released DPOSS and 2MASS images for custom mosaic construction

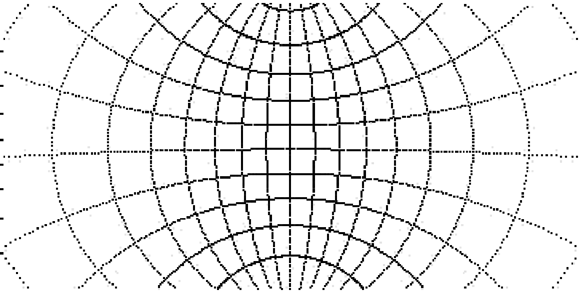


Customization



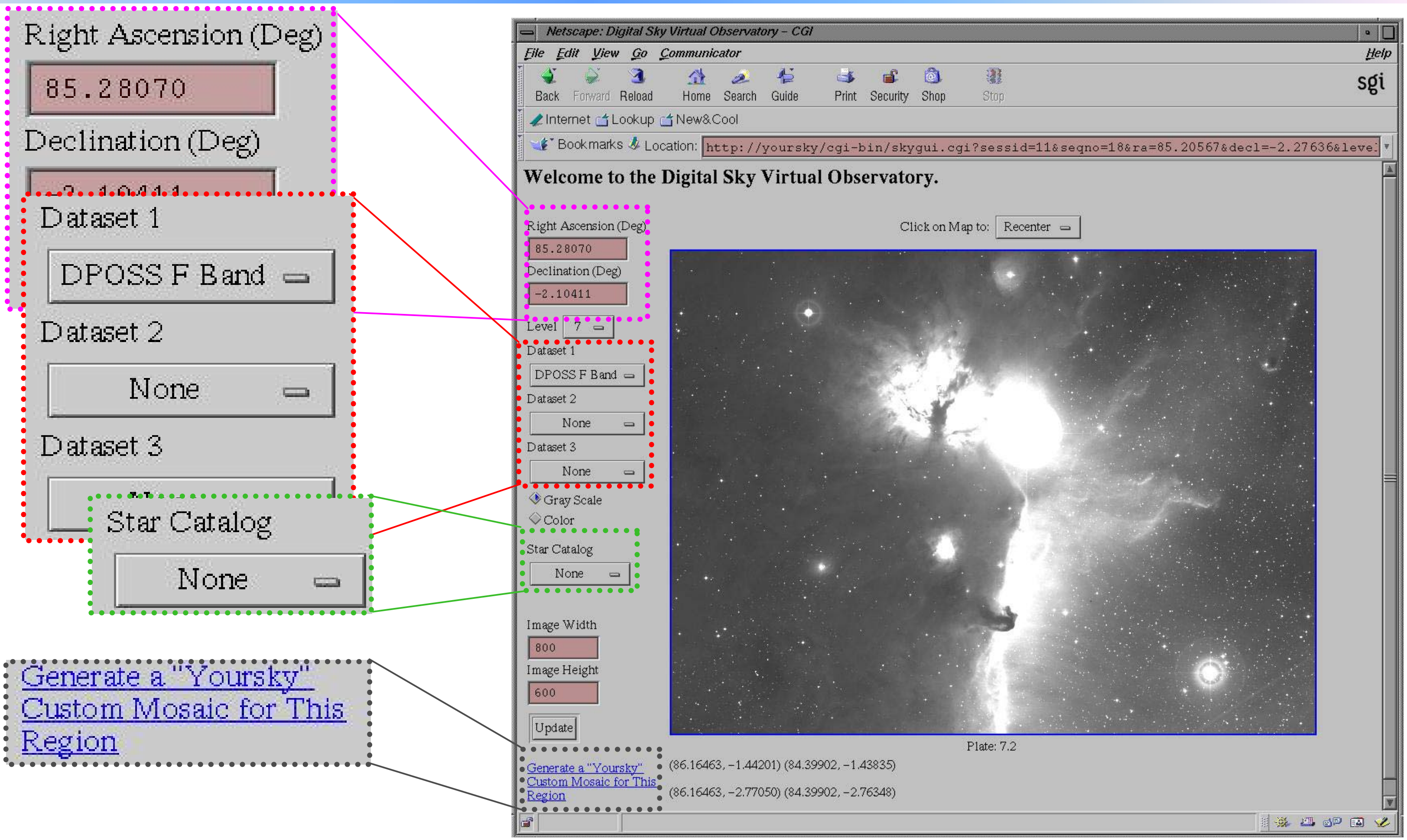
- Coordinate Systems: Galactic, Ecliptic, J2000 Equatorial, B1950 Equatorial
- WCS projections: LIN, TAN, SIN, STG, AZP, ARC, ZPN, ZEA, AIR, CYP, CAR, MER, CEA, COP, COD, COE, COO, BON, PCO, SFL, PAR, AIT, MOL, CSC, TSC, DSS, PLT
- Image Formats: FITS, JPEG, PGM, PNG, TIFF, Raw Data
- Data Type: {8, 16, 32}-bit unsigned integer, {8, 16, 32}-bit signed integer, {single, double}-precision floating point





yourSky's Future Graphical Interface

A web-based pan and zoom engine:

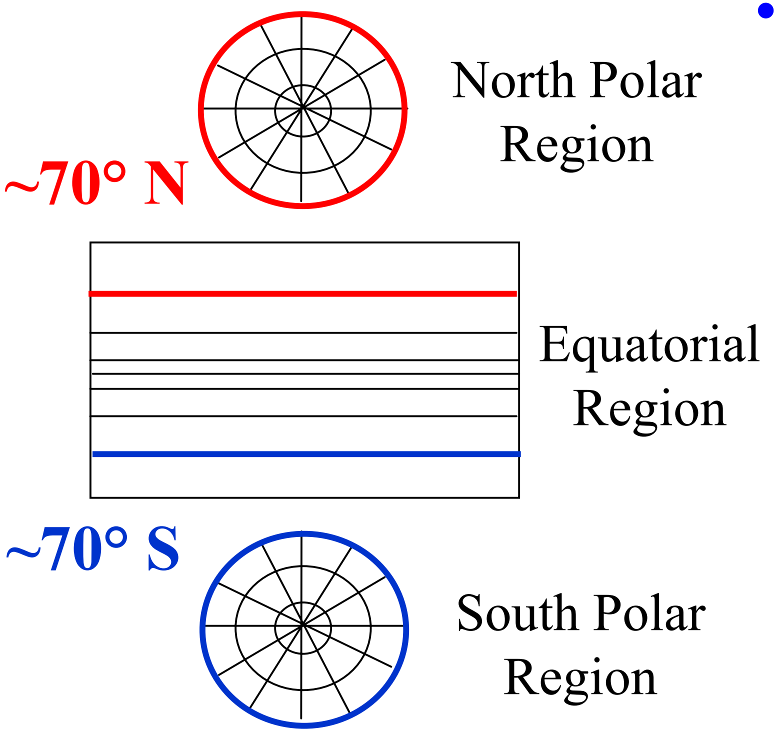


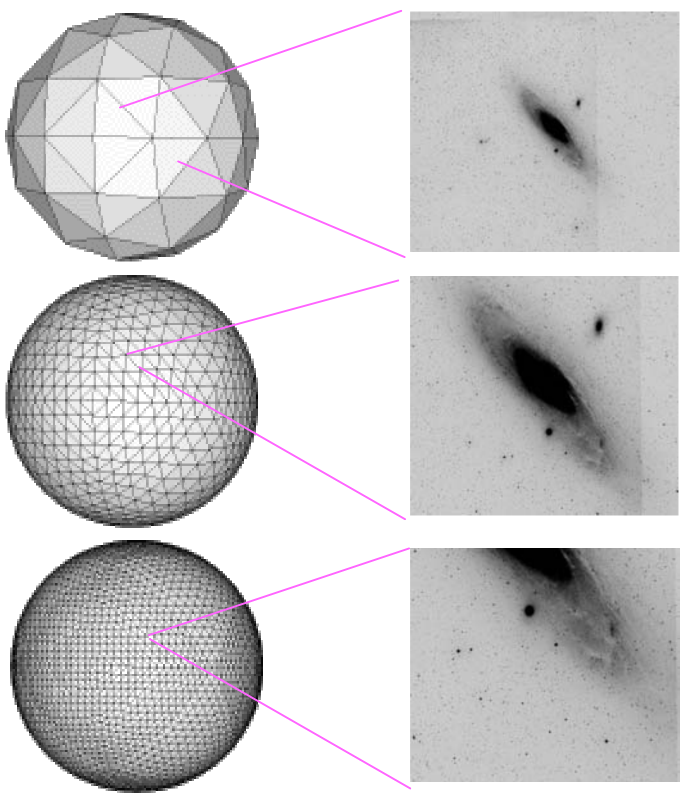
Images being populated - completion expected ~ Nov 15

Graphical Front-End: Architecture

- Hierarchical Triangular Mesh
- Single tangent plane at each vertex

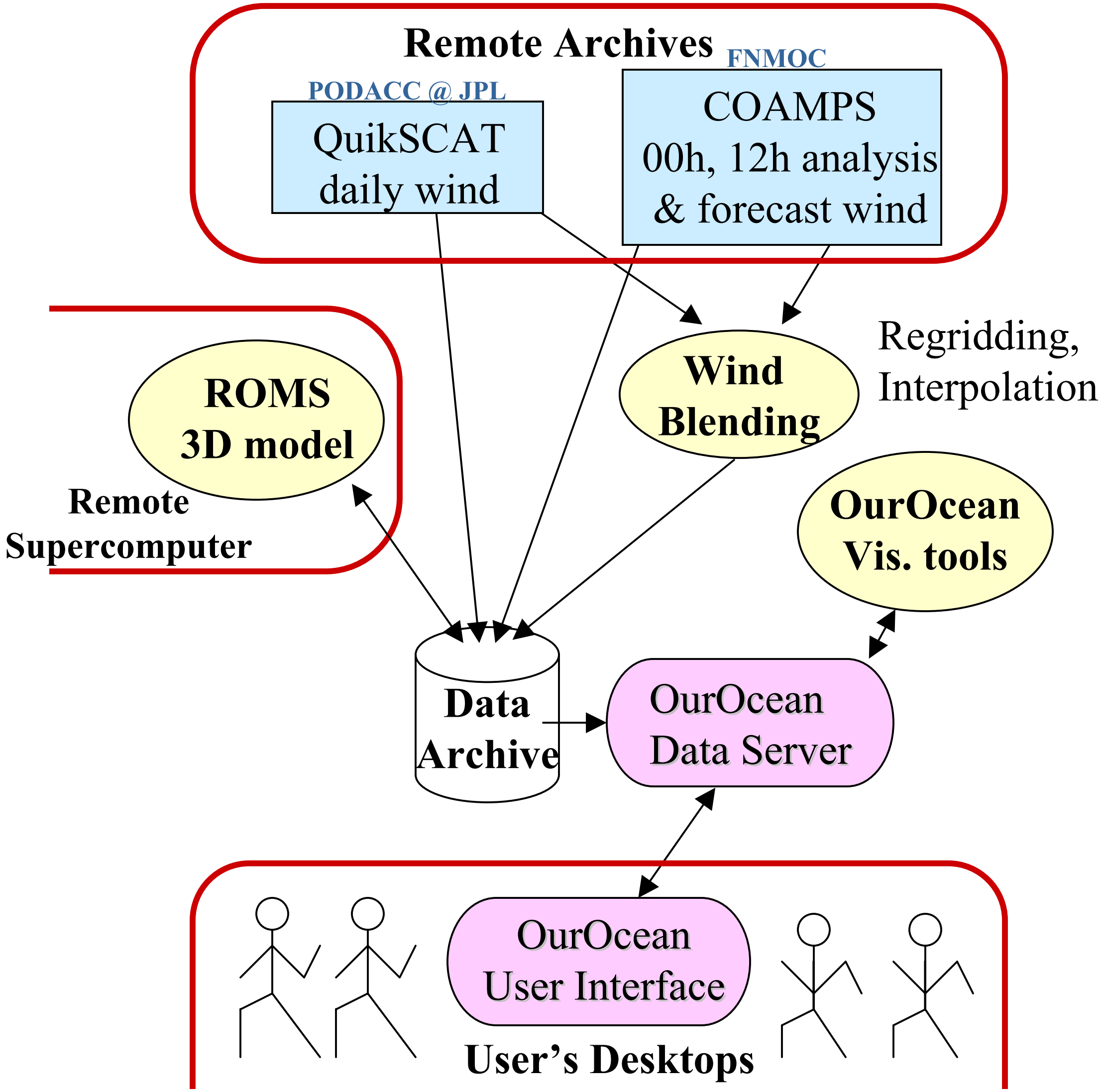
Synoptic View



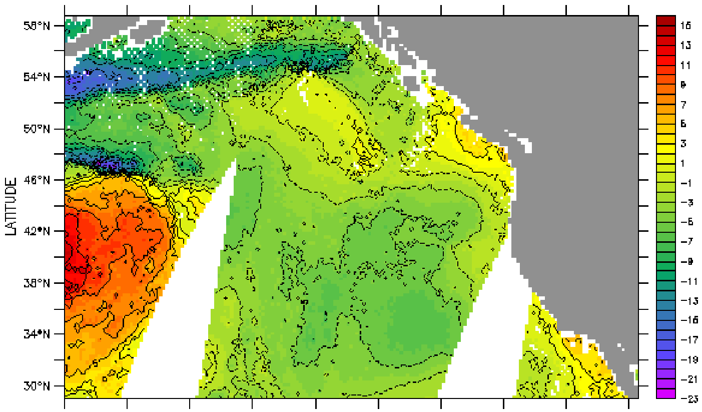


OurOcean (http://OurOcean.jpl.nasa.gov/)

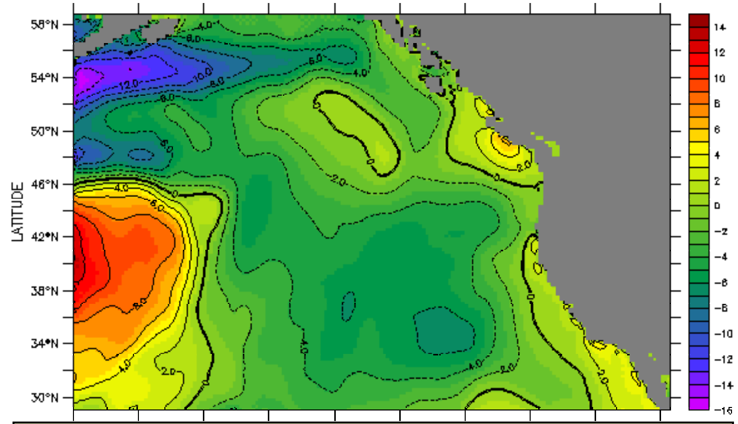
- A portal to serve near real-time coastal ocean datasets
- Focus on East Pacific Ocean winds
- Provides user access to data from archives, processed data, outputs of super-computer runs
- Outputs can be visual



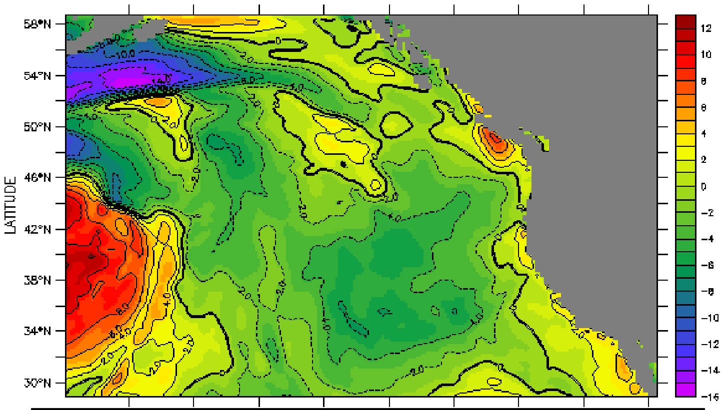
OurOcean Data Products



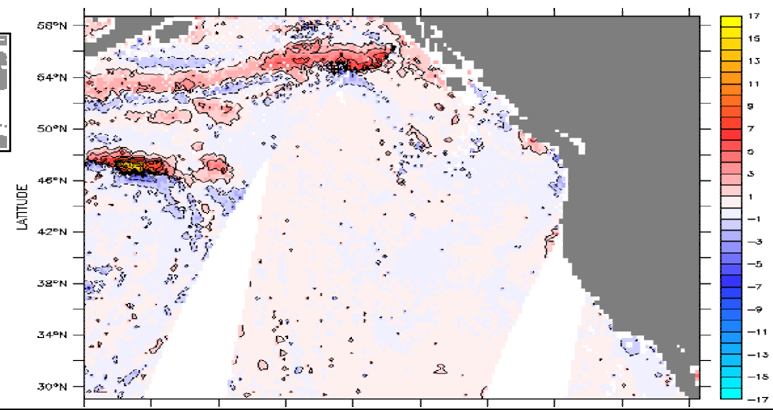
The QuikSCAT wind, u component, descending pass, 9/21/2002



The blended wind data, u component, descending pass, 9/21/2002



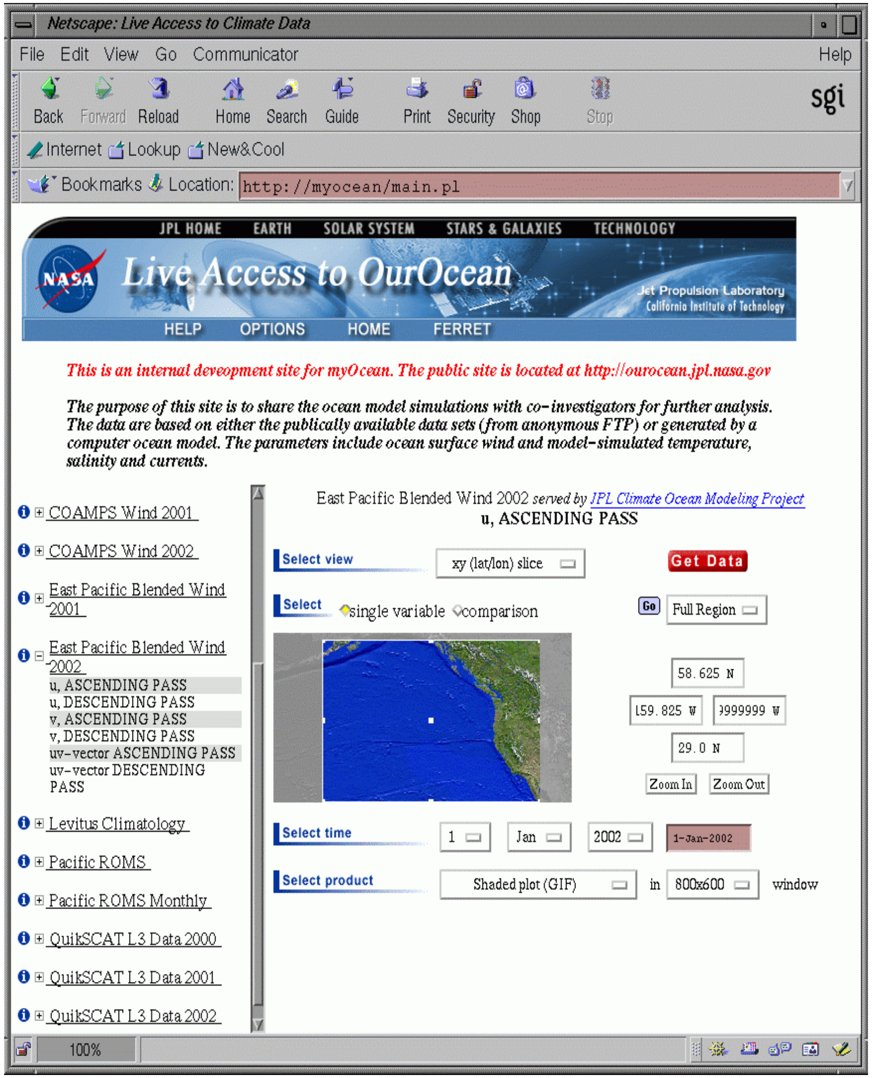
The COAMPS wind, u analysis data at 00h, 9/21/2002



The comparison of the blended wind and the QuikSCAT data, u component, 9/21/2002

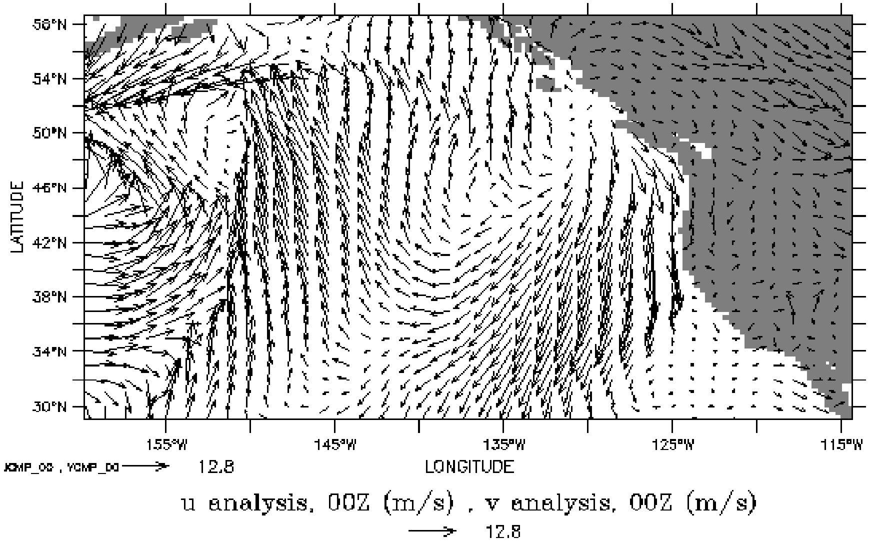
OurOcean User Interface

- Web browser running a Java applet
- Multi-level selection of datasets
- Clickable and zoomable map to select a sub-area of the dataset
- Pull-down menu to select a subset in depth or time dimension
- Various output formats– images at different resolutions, raw data, NetCDF file, text data
- 2D plots of cross sections from a 2D, 3D, or 4D data set.
- Configurable user interface using XML metadata

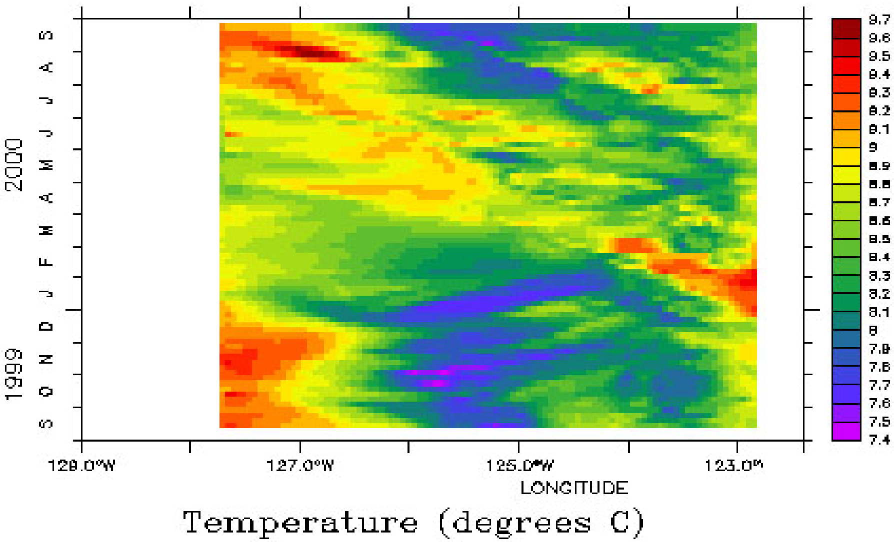


OurOcean Visualization Tools

- Ferret as default visualization tool
  - Produces 2D plots of large 4D gridded data sets:
    - Shaded plots
    - Contour maps
    - Vector plots
    - Land shade



Wind vector plot from COAMPS



The temperature profile from 9/99 –9/00 at 37.5º latitude, Data generated from a ROMS 3D model

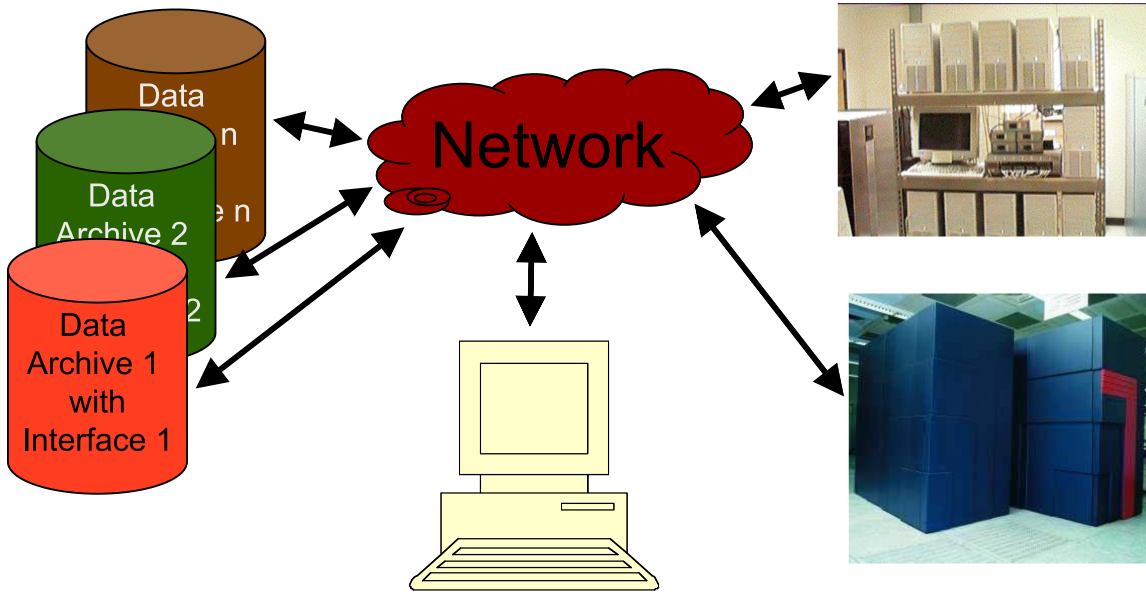
- Customized visualization tools such as Matlab, IDL or other 3D viz. tools can be added
- Capable of plotting compound variables (vectors) or difference of any two variables
- XML definition for plotting options

OurOcean Status & Plans

- *OurOcean* is serving 3 real time datasets (QuikSCAT, COAMPS, Blended Wind) and 3 ROMS datasets covering different regions (Atlantic, Central California Coastal and Pacific)
- *OurOcean* is currently open to selected collaborators outside JPL, will go public in December.
- Future plans:
  - Run East Pacific Coastal ROMS model on SGI Origin 2000 daily with the blended wind data as input, serve real time ROMS data on *OurOcean*
  - Add customized 3D visualization tool to *OurOcean*

Web Portals: Hiding Complexity

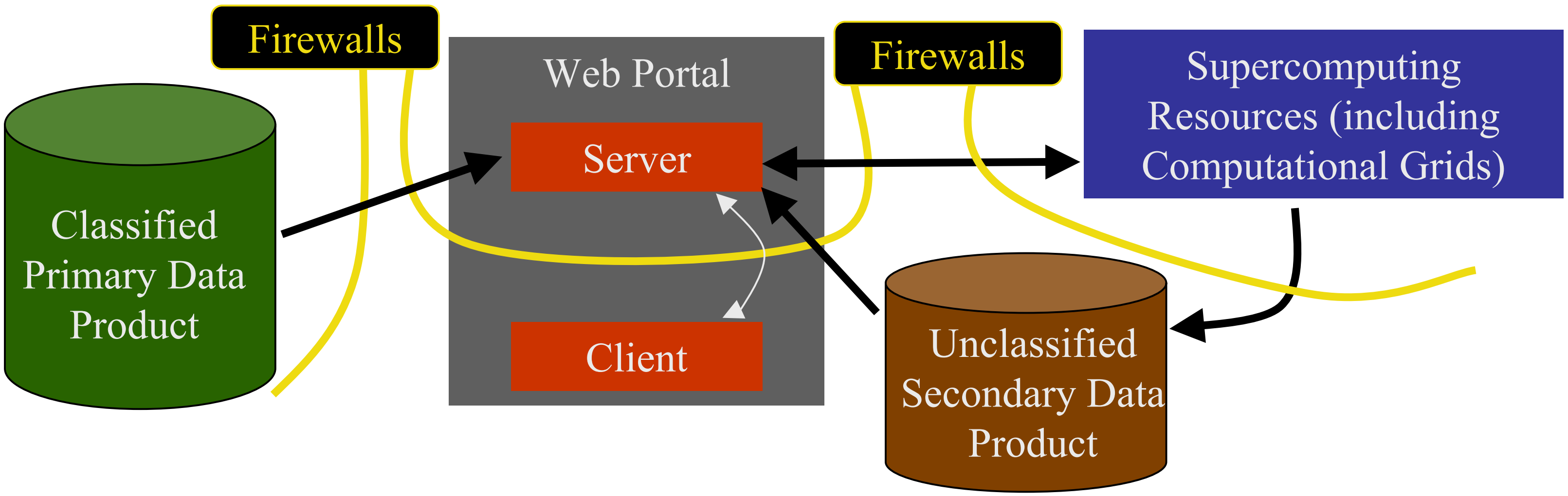
Web portals
hide the complexity
of accessing remote data archives and running jobs on supercomputers



- Users interact with the data via a web browser
- Users do not need to know how or where the data are stored or how to retrieve data from the different archives
- Users do not need to endure the complexity of scheduling applications on modern supercomputers
- Actual processing can be done anywhere using Grid technologies

Web portals **allow the user to focus on data**, not archives or computers

Web Portals: Security Advantages



Web portals can
control access to assets and data
of varied classifications

- Example: Delivery of unclassified secondary data products from classified primary data, using classified processing, a classified server, and an unclassified client

Conclusions

- Existing tools developed by the PAT group to make sense of large amounts of data:
  - Digital Light Table (DLT), RIVA, MAPUS, *yourSky*, *OurOcean*
  - All were developed in collaboration with scientists
  - They are a combination of:
    - Visualizing a scientist's data
    - Allowing others to view images generated from a scientist's data
    - Accessing and visualizing remote data
      - From archives and/or using supercomputers

Questions?