

Microsoft® Research

# Faculty Summit 2010

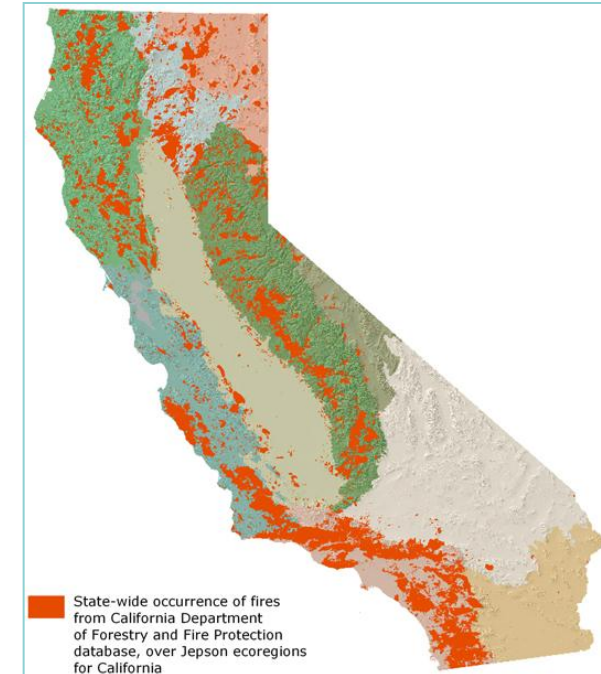
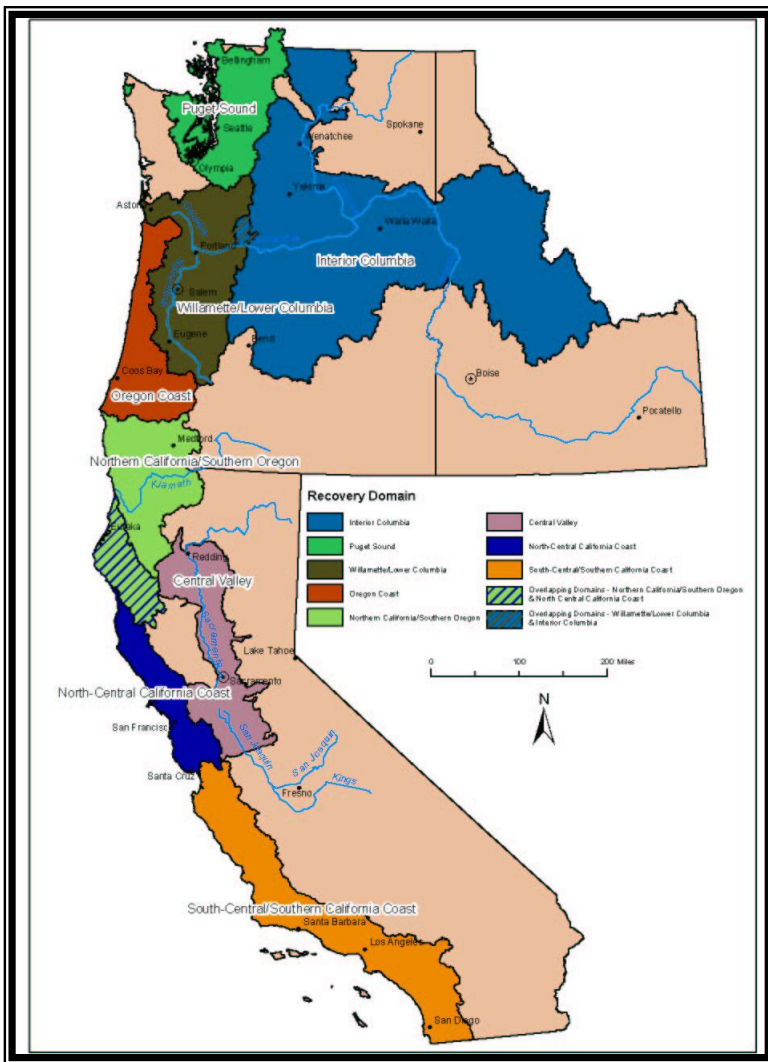
## Environmental Data Management

Deb Agarwal

Berkeley Water Center - University of California and  
Lawrence Berkeley National Laboratory

# Global and Regional Eco-Science

- A major shift is happening in the way eco-science is done.
  - Moving from individual studies of local processes to collaborative studies of regional and global processes. (e.g. studying the impact of climate change)
- Studying global scale environmental processes requires:
  - Integration of local, regional, and global spatial scales.
  - Integration across disciplines, e.g., climatology, hydrology, forestry, etc., and across methodologies (field observations, remote sensing, and modeling).



# Data Rich Environment

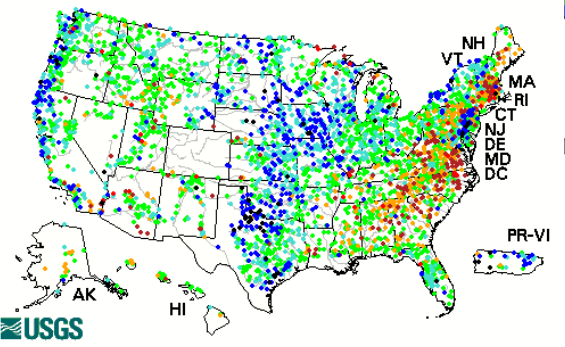
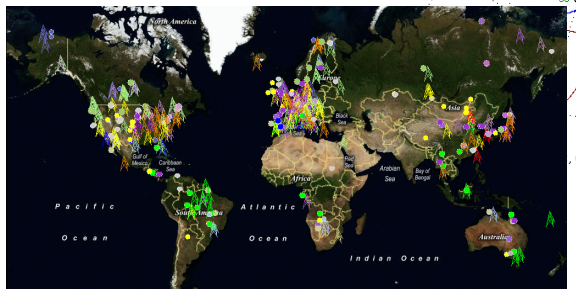
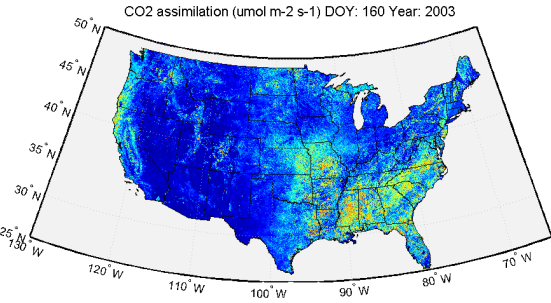
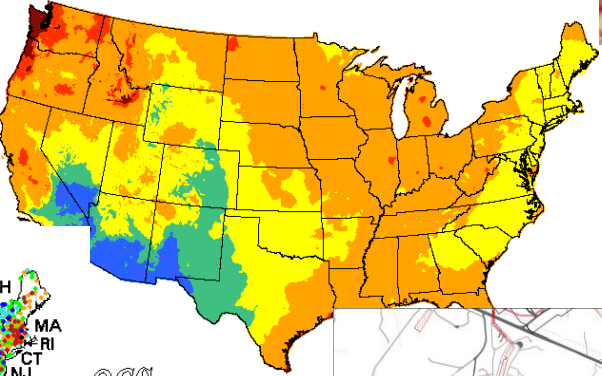
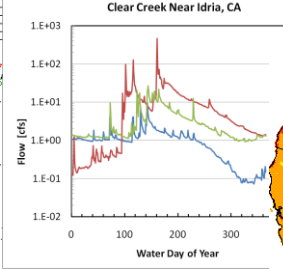
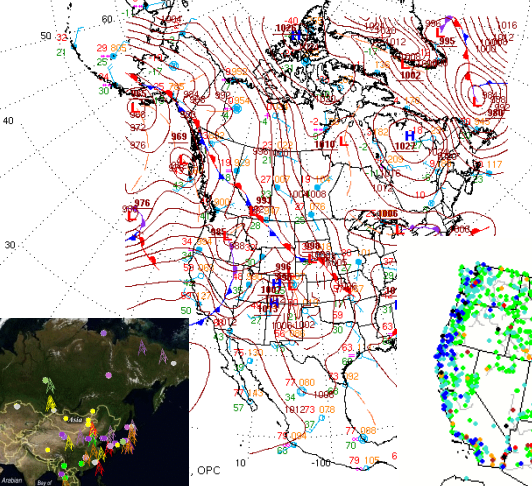
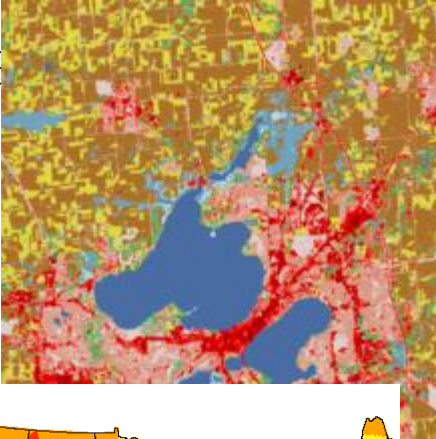
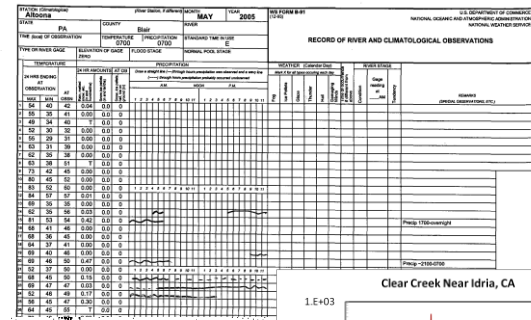
- National and International Datasets

- USGS National Water Information System
- NOAA National Climatic Data Center
- FLUXNET Network
- Satellite data (e.g. MODIS)

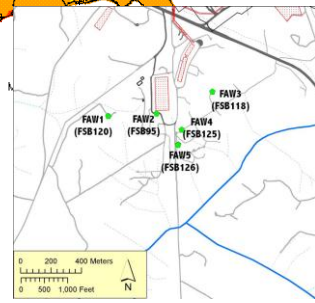
- Local Datasets

- Local Agencies
- Companies (e.g. Timber)
- Ecology Organizations
- Individual Researchers

	A	B	C	D	E	F	G	H	I
	Page	Target	Habitat Attribute	Indicator	Method	States	Poor	Fair	
1	8	Spawning Adults	Estuary	Passage at Mouth	Prof. Opinion		<30 days	30-60 days	
2	8	Spawning Adults	Hydrology	Passage Flow	Flow Panel Results	DDME			
3	8	Spawning Adults	Passage	Physical Barriers	Passage Database	FINAL	<50% of IP-km	50-70%	
4									
5	11	Spawning Adults	Viability	Freshwater Harvest	Review Regulations	Status?			
6	11	Spawning Adults	Viability	Density Target	MMFS Calculation	Apply TRT Criteria	Watershed Specific		
7		Spawning Adults	Sediment	Spawning Gravel	take all labors with emb. rating < 5, multiply by sq width of riffle squared	Hopland Dong	Queries		
8									
9									
10									
11									
12		Egg	Hydrology	Instantaneous Condition	Flow				



OCS



# Data Comes in Many Forms



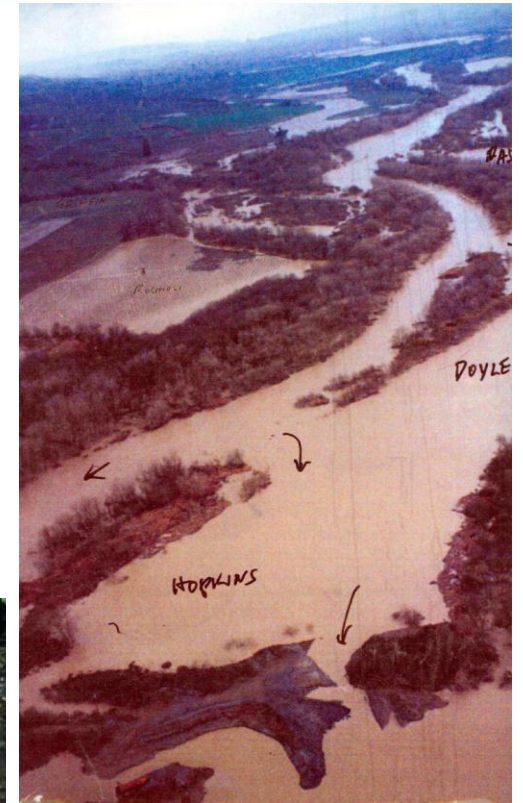
Manual Measurement



Automated Measurement



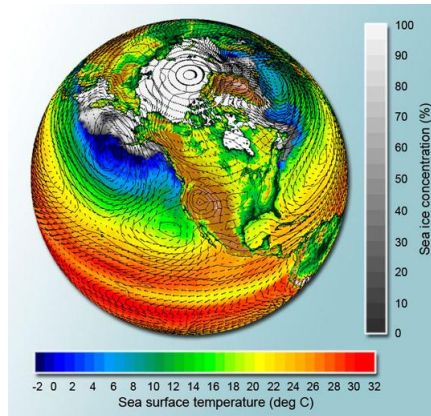
Sample Collection



Historical Photographs



Typing



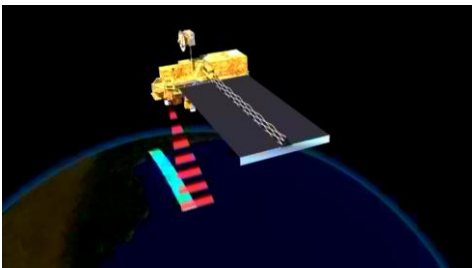
Model Output



Counting

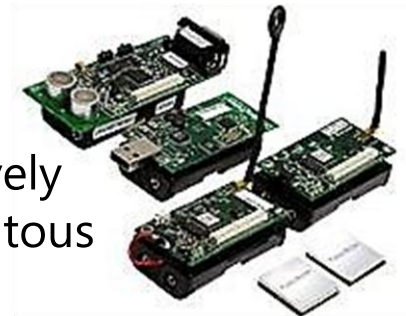


Aircraft Surveys



Satellite

Relatively Ubiquitous Motes



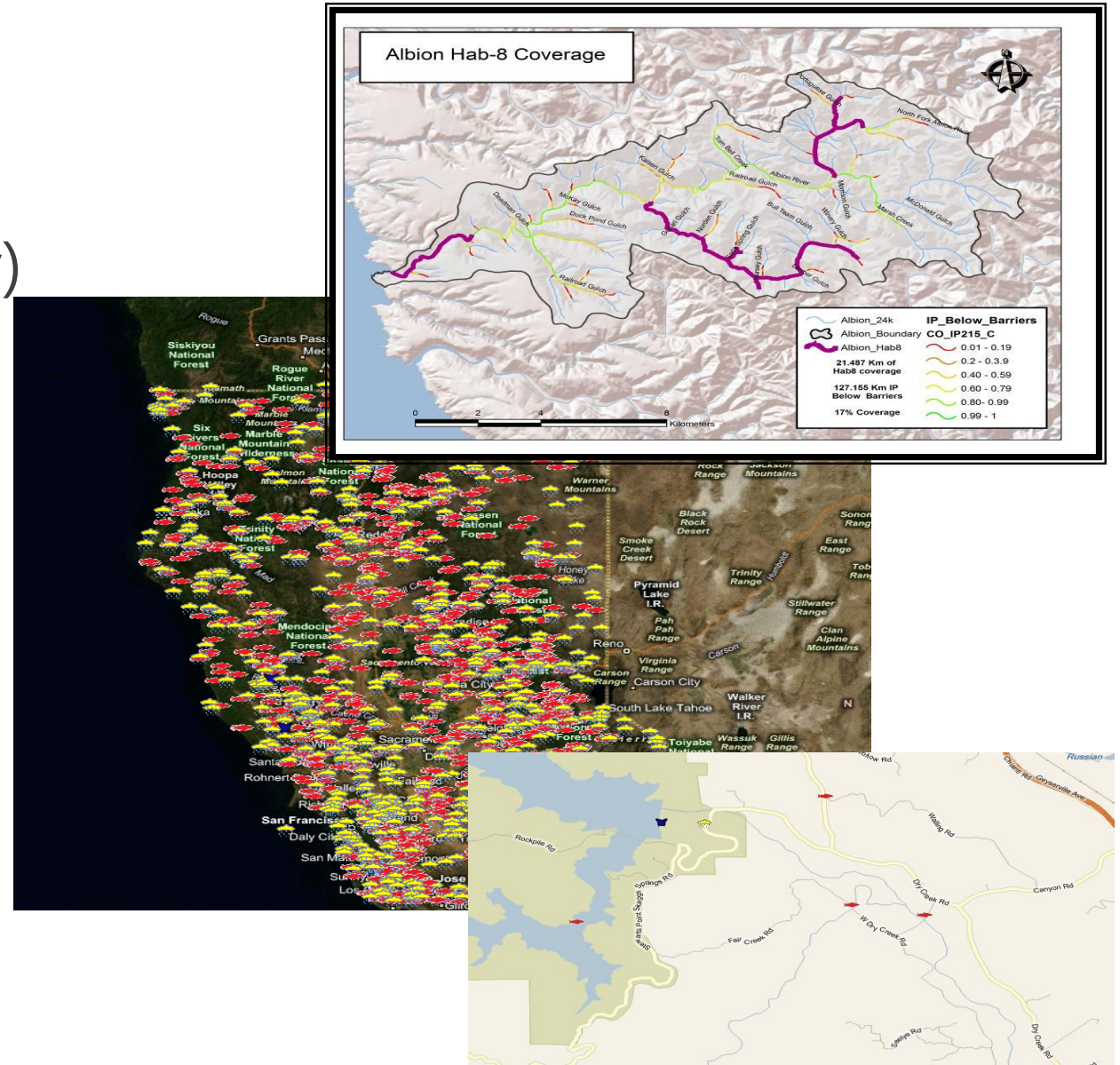
# Data Synthesis – The Challenge of Bringing it all Together

USGS stream gauges  
Coho presence/absence data  
MODIS evapotranspiration  
Cross section locations  
Samples from 1997  
Stream temperature gauges  
Water system  
Obstructions and other  
Human activities



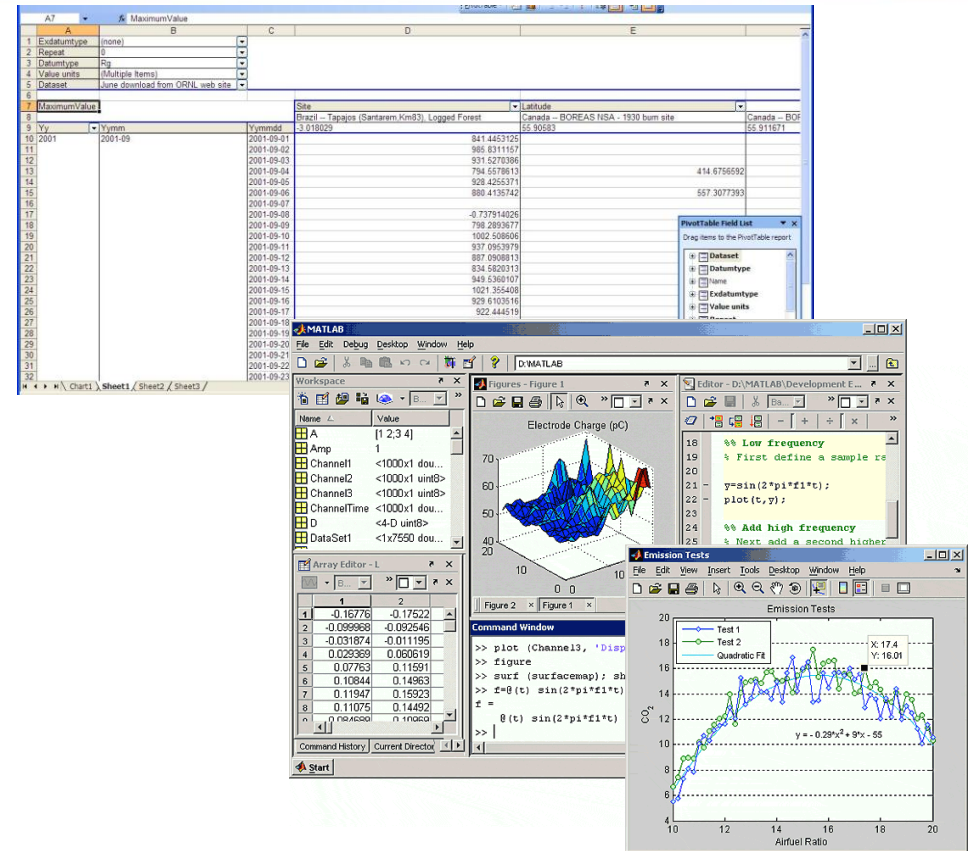
# Data Integration Challenges

- Point and spatial
- Time series and low frequency
- Vocabulary meanings (ontology)
- Sparse in time, duration, or location

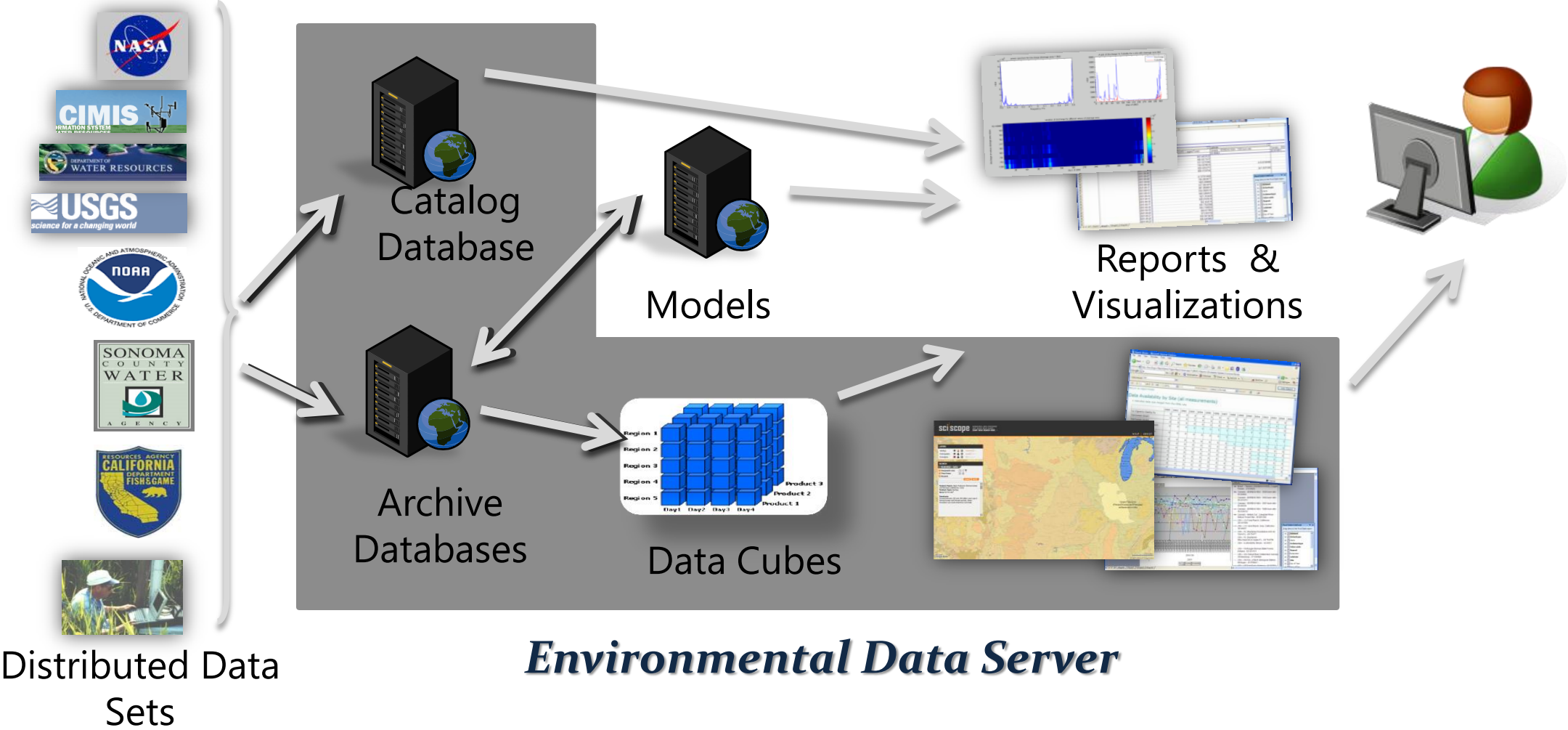


# Data Access and Mining – Time Series Data

- Organize data by its natural dimensions
  - Watershed
  - Stream
  - Data type
  - Time
  - etc
- Select
  - Aggregations – yearly, monthly, stream, etc
  - Filters – watershed, years,
- Search for specific characteristics
- Browse and display data in the way that makes sense to the viewer



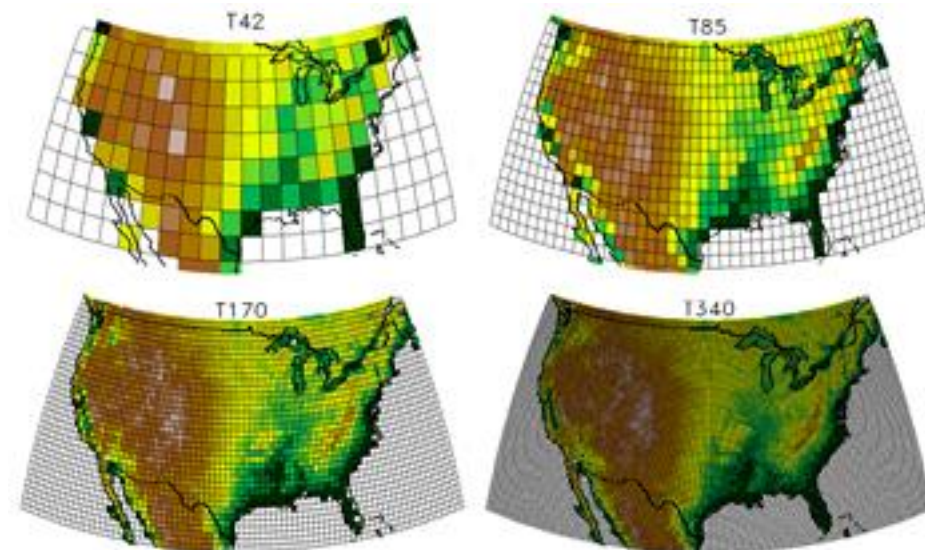
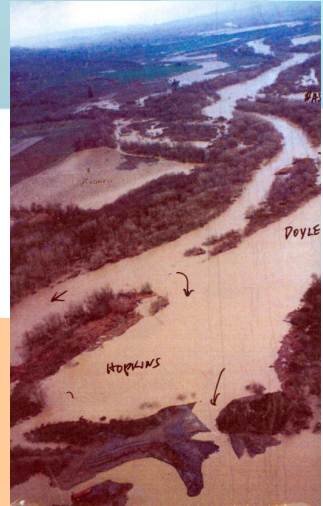
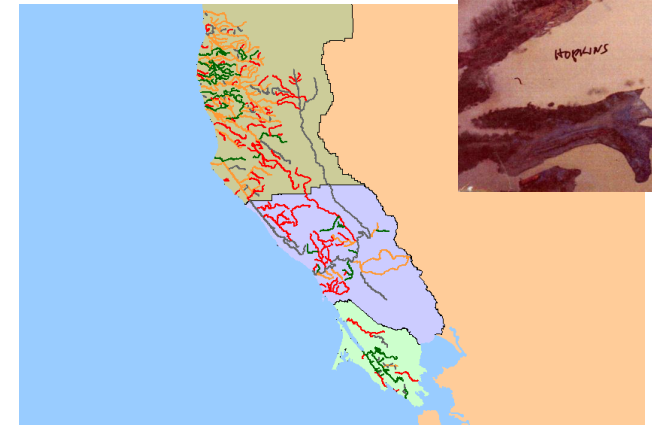
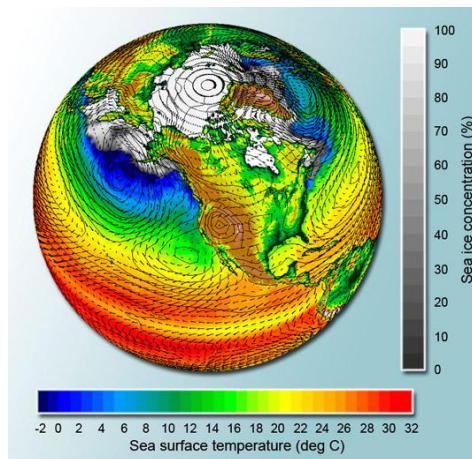
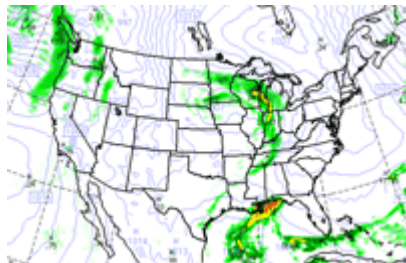
# Data Analysis Infrastructure For Time Series Data





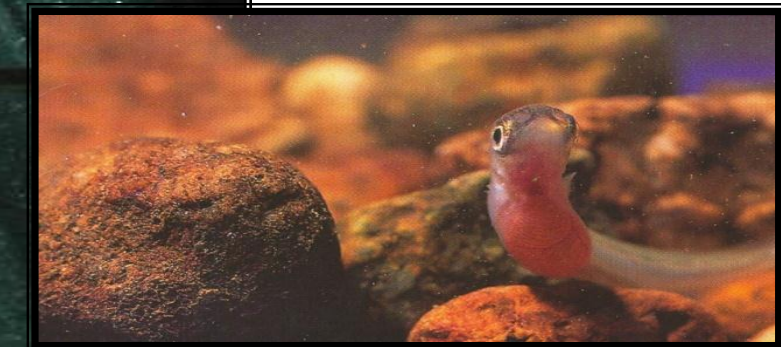
# Non-Time Series Data

- One-time or infrequent event information
- Approximate measurements
- Raster images
- Photographs
- Video
- GIS information and shape files
- Model output



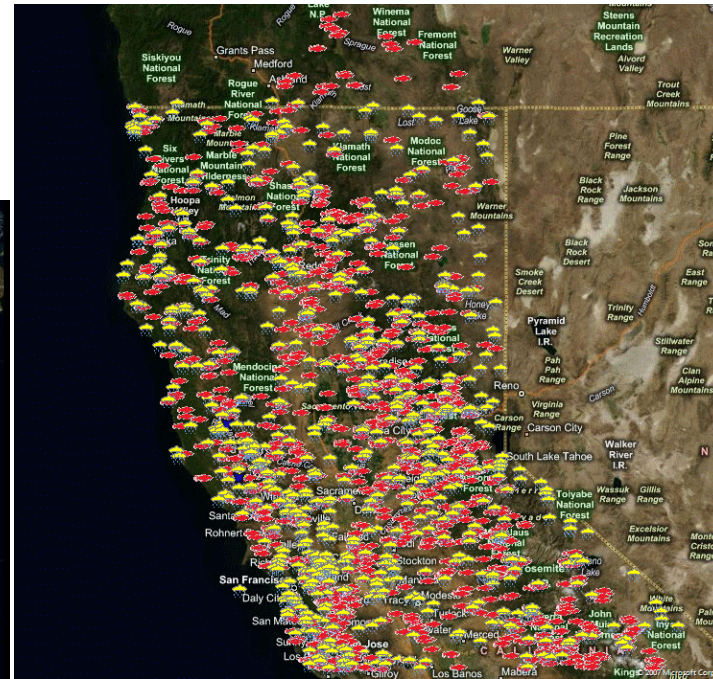
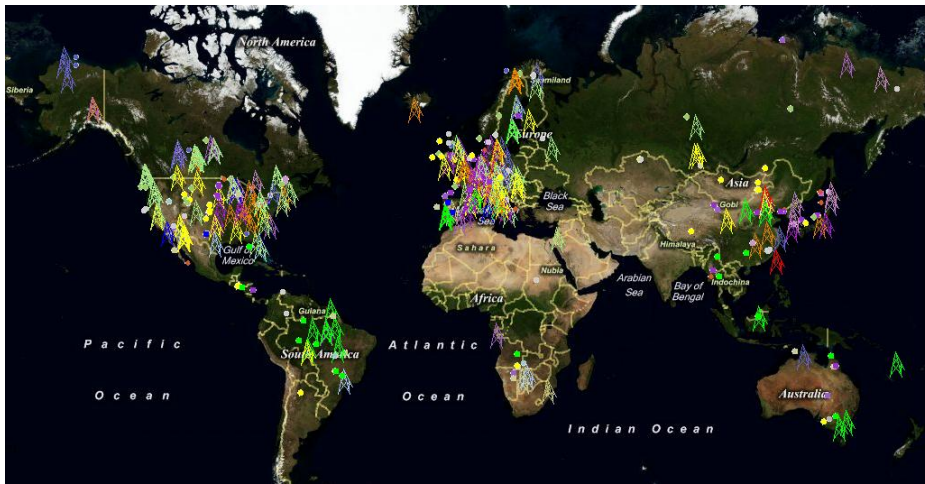
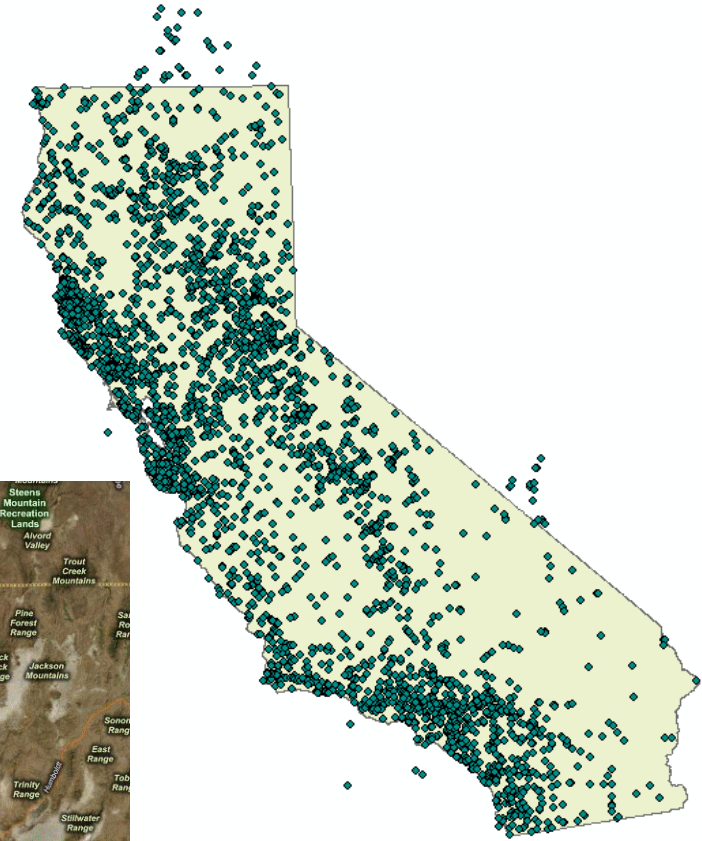
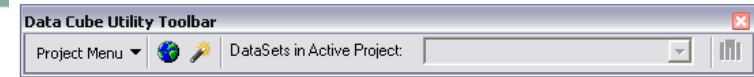
# Salmon Lifestages - Different Conditions

LIFE STAGE	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Adult migration	Grey	White	White	White	White	White	White	White	White	Grey	Black	Black
Spawning	Black	Grey	White	White	White	White	White	White	White	Grey	Black	Black
Egg Incubation	Black	Black	Black	Grey	White	White	White	White	White	White	Grey	Black
Emergence/ Fry	White	Grey	Black	Black	Black	Grey	White	White	White	White	White	White
Juvenile rearing	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black	Black
Emigration	White	White	Grey	Black	Black	Black	Black	White	White	White	White	White



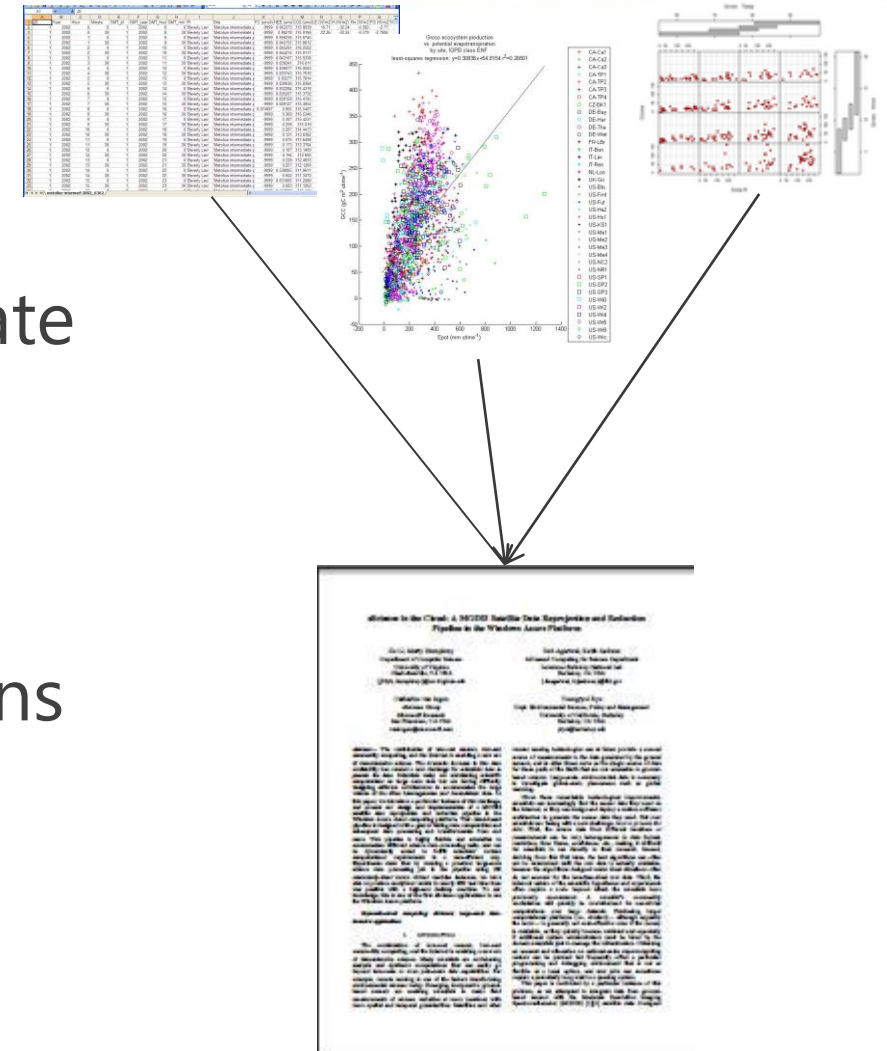
# A Start on this Issue

- Mashups
  - Browsing in a geographic view
  - Overlays of rasters
  - Popups with data
- Geographic Information System (GIS)
  - Connection to time series data
  - Operate in a geographic tool



# Synthesis Challenge

- Find the data needed – Agency sites, individual researchers, networks, . . .
- Convert it to a format and units appropriate to the analysis
- Interpolate in space or time as needed
- Gap fill and quality assess
- Understand appropriate use and limitations
- Perform the analysis
- Gather all the 'fair-use' criteria and acknowledgments
- Publish the paper







# Acknowledgements

## Berkeley Water Center, University of California, Berkeley, Lawrence Berkeley Laboratory

- Dennis Baldocchi
- Jim Hunt
- Monte Goode
- Susan Hubbard
- Keith Jackson
- Rebecca Leonardson (student)
- Carolyn Remick

## Microsoft Research

- Catharine van Ingen
- Tony Hey
- Dan Fay

## University of Virginia

- Marty Humphrey
- Norm Beekwilder
- Jie Li (student)

## University of Indiana

- You-Wei Cheah (student)

## Fluxnet Collaboration

- Dennis Baldocchi
- Youngryel Ryu (postdoc)
- Dario Papale (CarboEurope)
- Markus Reichstein (CarboEurope)
- Alan Barr (Fluxnet-Canada)
- Bob Cook
- Dorothea Frank
- Susan Holladay
- Hank Margolis (Fluxnet-Canada)
- Rodrigo Vargas (postdoc)

## Ameriflux Collaboration

- Beverly Law
- Tom Boden
- Mattias Falk
- Tara Hudiburg (student)
- Hongyan Luo (postdoc)
- Gretchen Miller (student)
- Lucie Ploude (student)
- Andrew Richardson
- Andrea Scheutz (student)
- Christophe Thomas

## National Marine Fisheries

- Charlotte Ambrose
- Maura Moodie
- Charleen Gavette
- William Winner

## Sonoma Ecology Center

- Alex Young
- Lisa Micheli
- Deanne DiPietro



Microsoft® Research

# Faculty Summit 2010