Integrating an OLAP Data Cube with ArcMap

Lt. (j.g.) William Winner, NOAA
Deb Agarwal, UC Berkeley
Catharine van Ingen, Microsoft

Project Origins

- Began at UC Berkeley
  - Proposal to NSF for a watershed data synthesis center
  - Proposal to Microsoft to create a demonstration centered around California water issues (2006)

Original Objectives

- Database requirements
  - Needed large storage capacity
  - Needed ability to merge data from different agencies in different formats
  - Needed to handle different types of data
  - Needed the ability to perform analyses quickly

What is a Data Cube?

- An N-dimensional database
  - Traditional DBs can do this too, but…
    - Data must be flattened
    - Done using Aggregations, Group By's and Cross Tabs

What is a Data Cube?

SALES			
Model	Year	Color	Sales
Chevy	1990	red	5
Chevy	1990	white	87
Chevy	1990	blue	62
Chevy	1991	red	54
Chevy	1991	white	95
Chevy	1991	blue	49
Chevy	1992	red	31
Chevy	1992	white	54
Chevy	1992	blue	71
Ford	1990	red	64
Ford	1990	white	62
Ford	1990	blue	63
Ford	1991	red	52
Ford	1991	white	9
Ford	1991	blue	55
Ford	1992	red	27
Ford	1992	white	62
Ford	1992	blue	39

From J.Gray, et al.: "Data cube: A relational operator generalizing group-by, crosstab and sub-totals", *ICDE 1996*, pages 152-159, 1996

What is a Data Cube?

SALES			
Model	Year	Color	Sales
Chevy	1990	red	5
Chevy	1990	white	87
Chevy	1990	blue	62
Chevy	1991	red	54
Chevy	1991	white	95
Chevy	1991	blue	49
Chevy	1992	red	31
Chevy	1992	white	54
Chevy	1992	blue	71
Ford	1990	red	64
Ford	1990	white	62
Ford	1990	blue	63
Ford	1991	red	52
Ford	1991	white	9
Ford	1991	blue	55
Ford	1992	red	27
Ford	1992	white	62
Ford	1992	blue	39

Aggregate

Sum

Group By (with total)

By Color

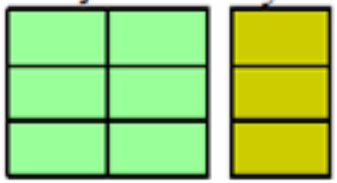
RED
WHITE
BLUE

Sum

Cross Tab

Chevy Ford By Color

RED
WHITE
BLUE



By Make

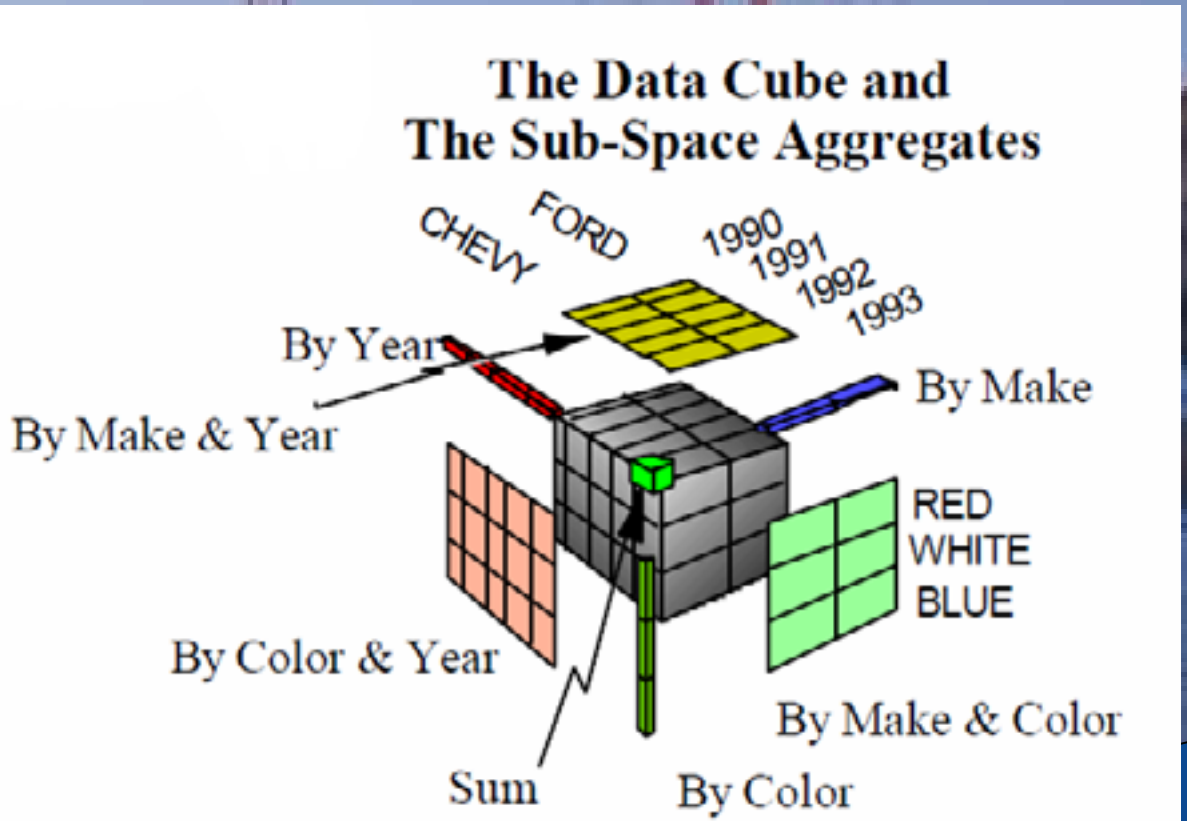
Sum

From J.Gray, et al.: "Data cube: A relational operator generalizing group-by, crosstab and sub-totals", *ICDE 1996*, pages 152-159, 1996

NOAA Berkeley WATER CENTER UC BERKELEY Microsoft

What is a Data Cube?

SALES			
Model	Year	Color	Sales
Chevy	1990	red	5
Chevy	1990	white	87
Chevy	1990	blue	62
Chevy	1991	red	54
Chevy	1991	white	95
Chevy	1991	blue	49
Chevy	1992	red	31
Chevy	1992	white	54
Chevy	1992	blue	71
Ford	1990	red	64
Ford	1990	white	62
Ford	1990	blue	63
Ford	1991	red	52
Ford	1991	white	9
Ford	1991	blue	55
Ford	1992	red	27
Ford	1992	white	62
Ford	1992	blue	39

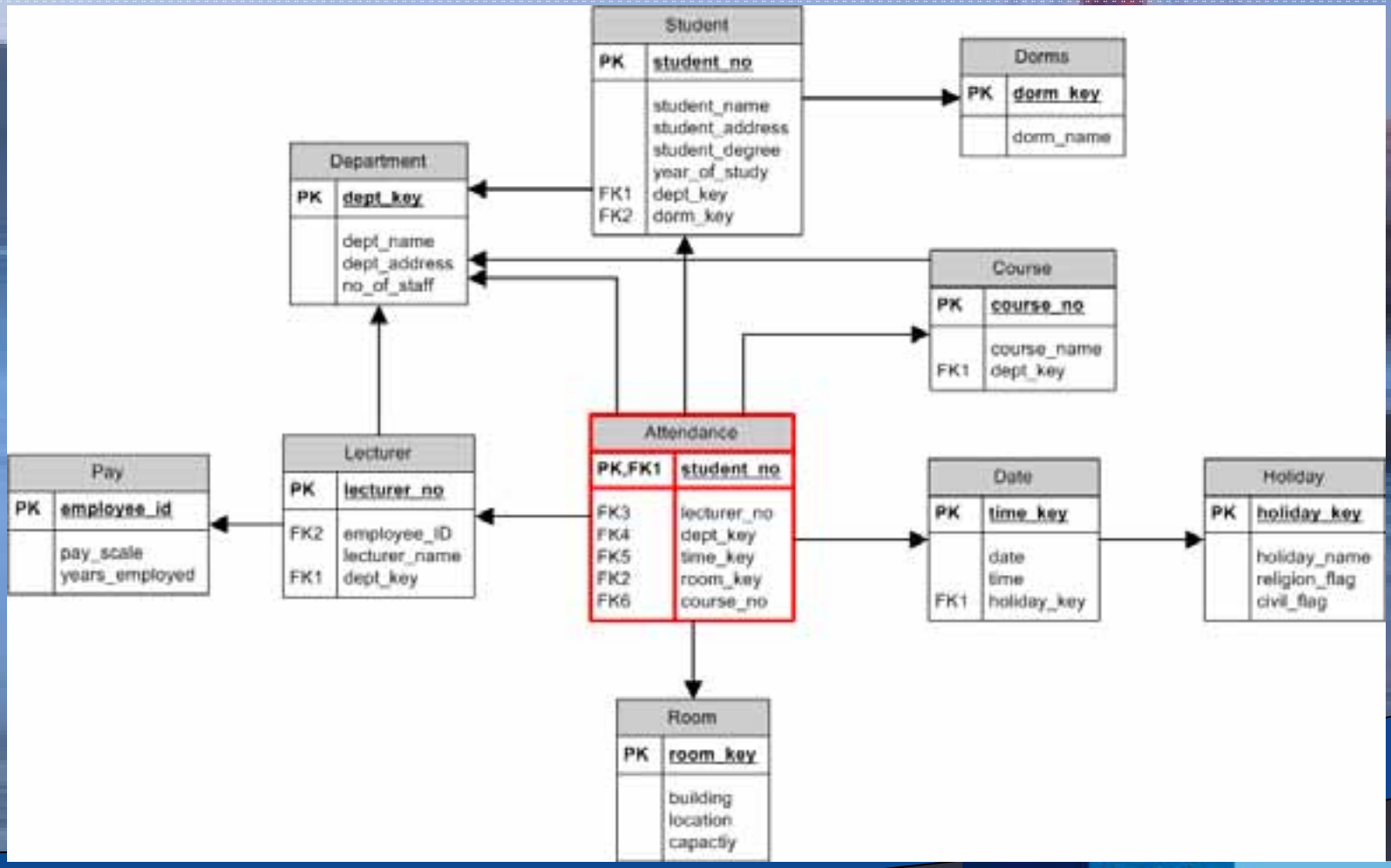


From J.Gray, et al., "Data cube: A relational operator generalizing group-by, cross-tab and sub-totals", *ICDE 1996*, pages 152-159, 1996

How is the Cube stored?

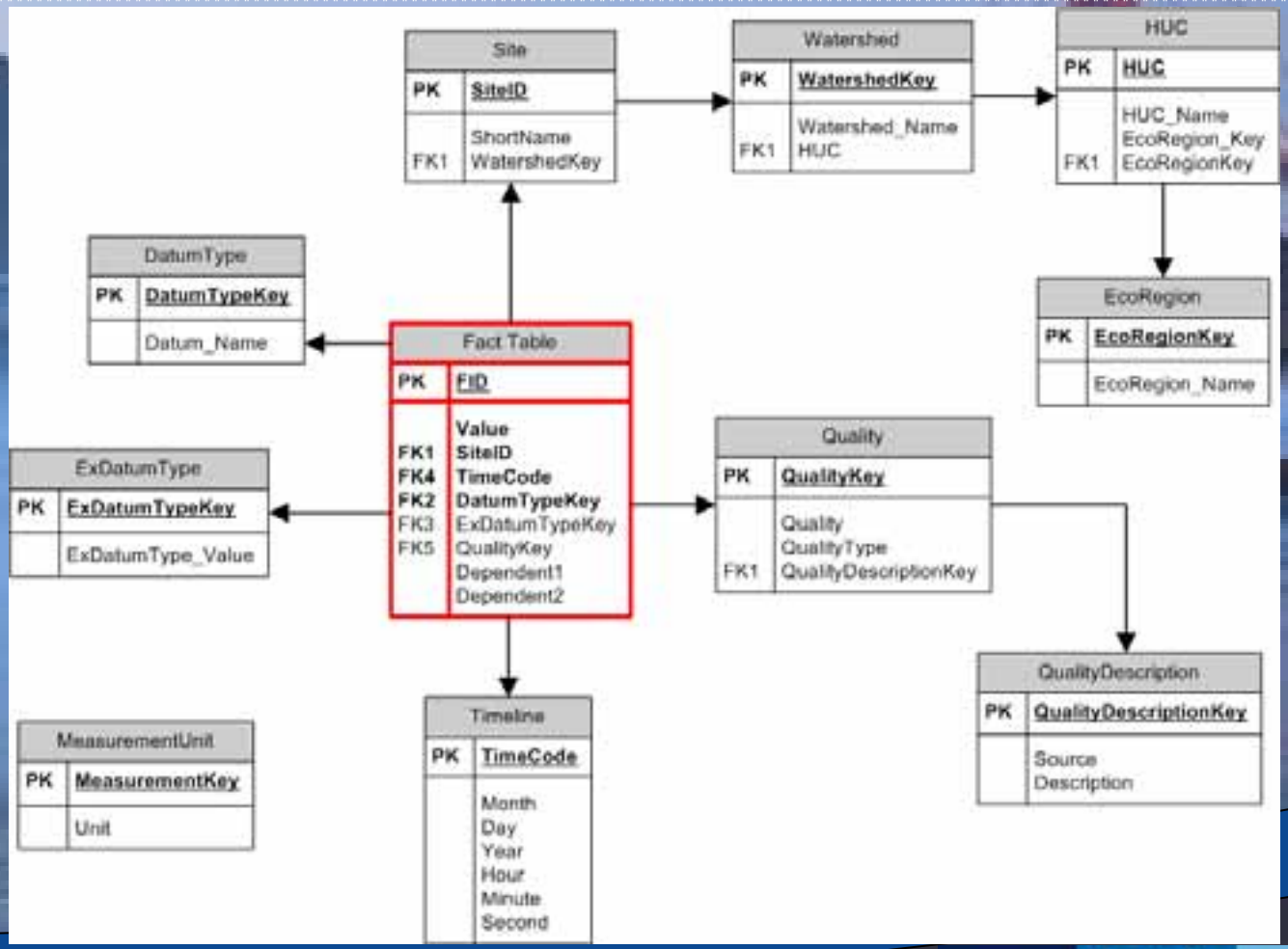
- The Cube is a computed product from a series of relational tables
- Our Cube is based on a Context-Dependent Snowflake Schema

Snowflake Schema

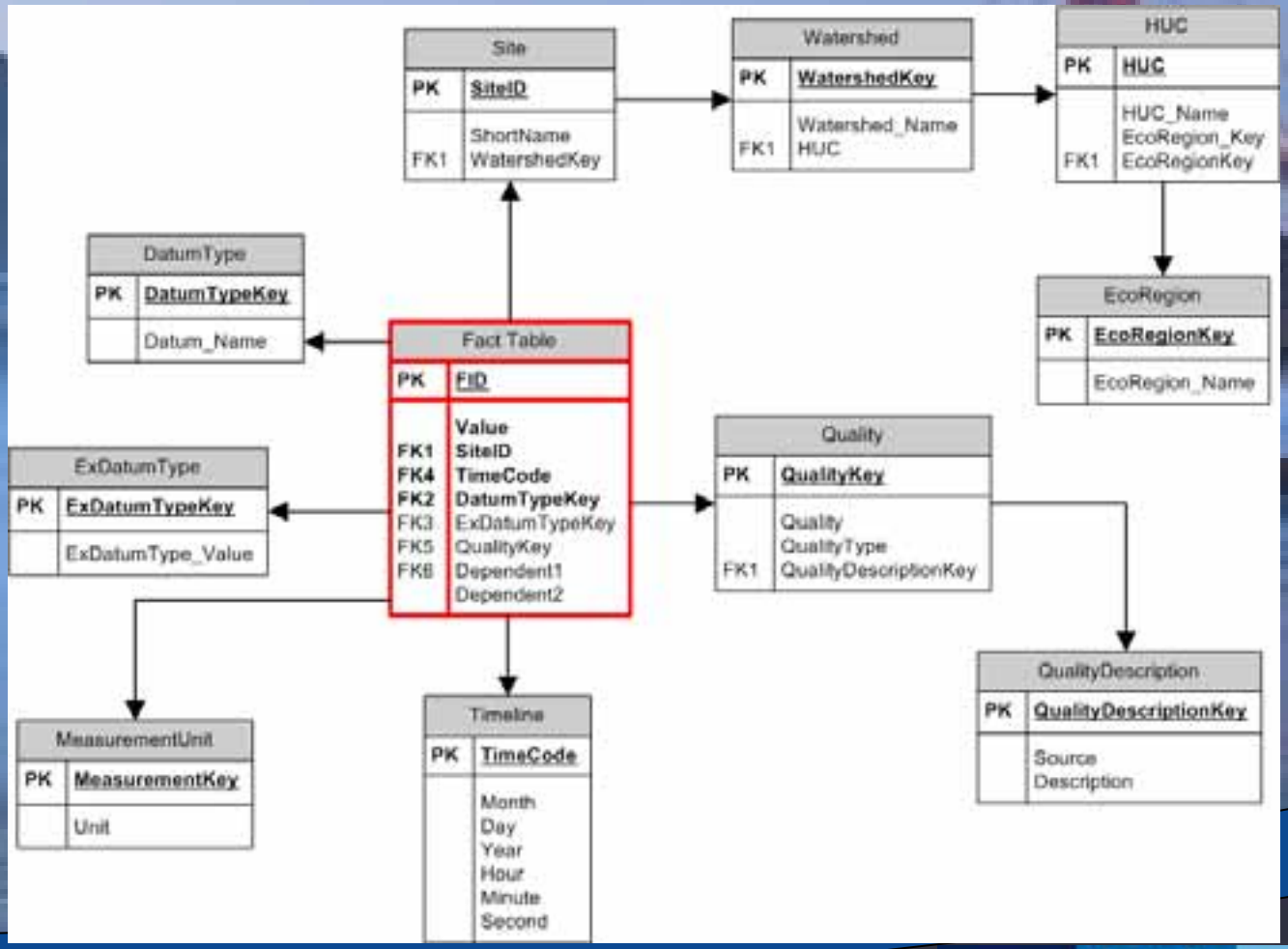


From Levene, M. and George Loizou. "Why is the snowflake schema a good data warehouse design?". *Information Systems*, v.28, pps 225-240, 2003

Snowflake Schema



Context-Dependent Snowflake



NMFS Involvement

- Came aboard at the end of 2008
- Saw the potential to use this to help inform Salmonid Recovery Plans
- Saw the potential to include a spatial component with the data

Toolbar Requirements

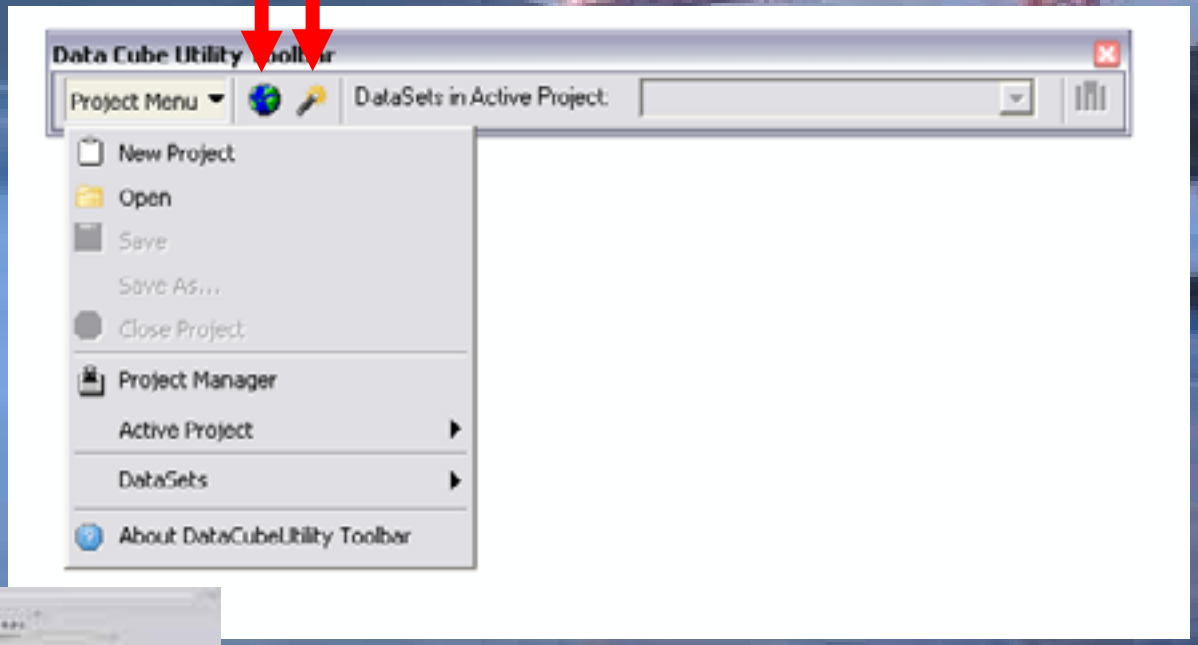
- Allow for spatial queries
- Flatten data to fit within a shapefile's attribute table
- Provide some way to visually move through the data

How did we do it?

- The toolbar was written using Visual Studio 2008 along with the ArcGIS API's.
- Also included project management and storage
- Written with just shy of 10,000 lines of designer- and user-generated code

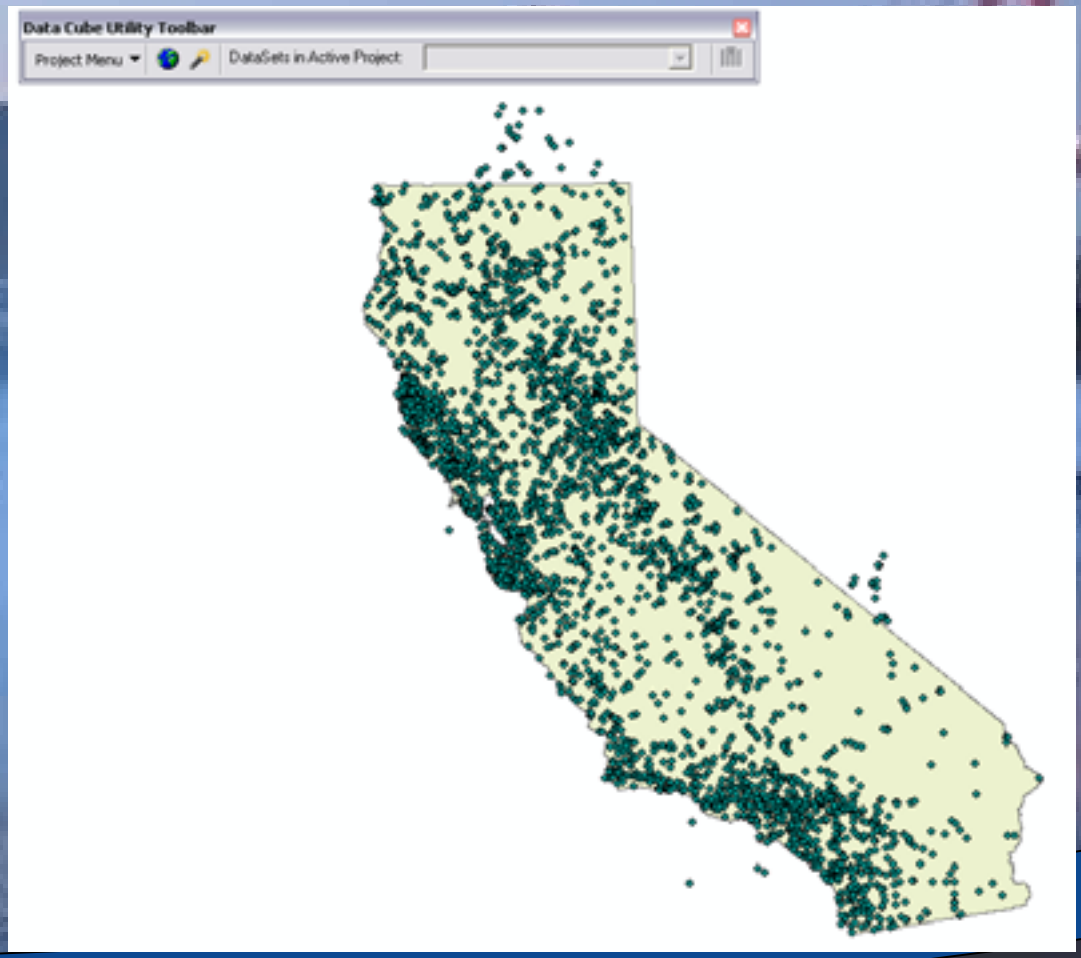
Data Cube Utility Toolbar

View Sites Site Selection Wizard

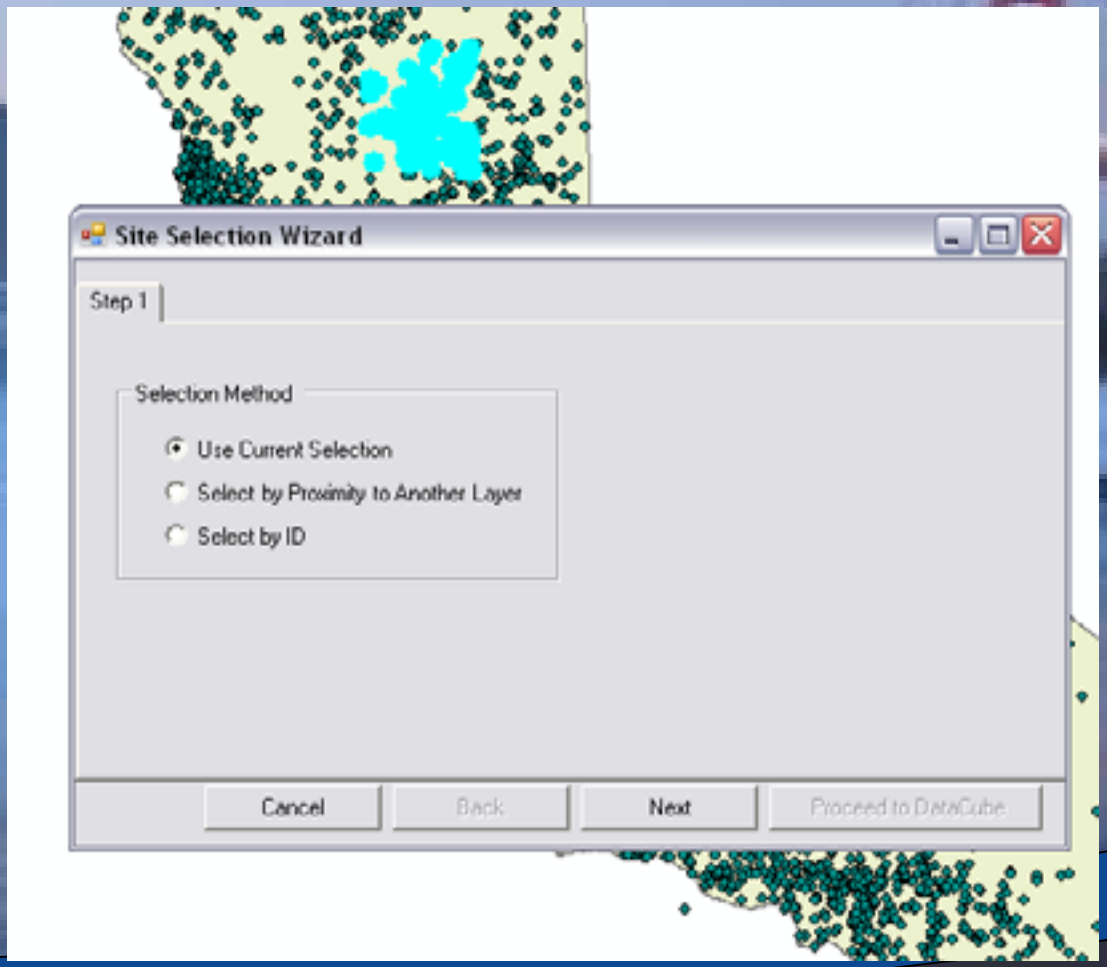


NOAA Berkeley WATER CENTER UC BERKELEY Microsoft

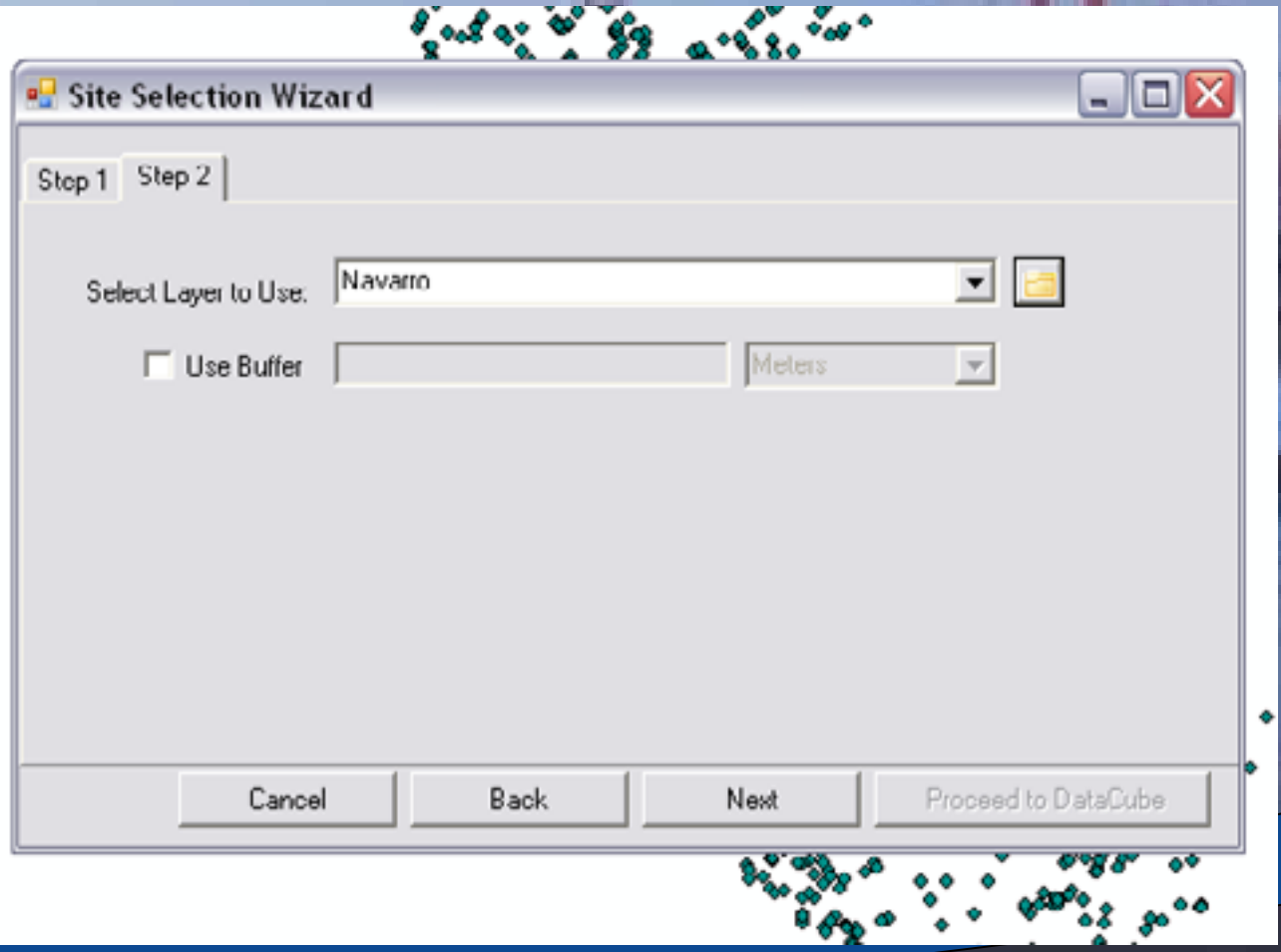
Data Cube Utility Toolbar



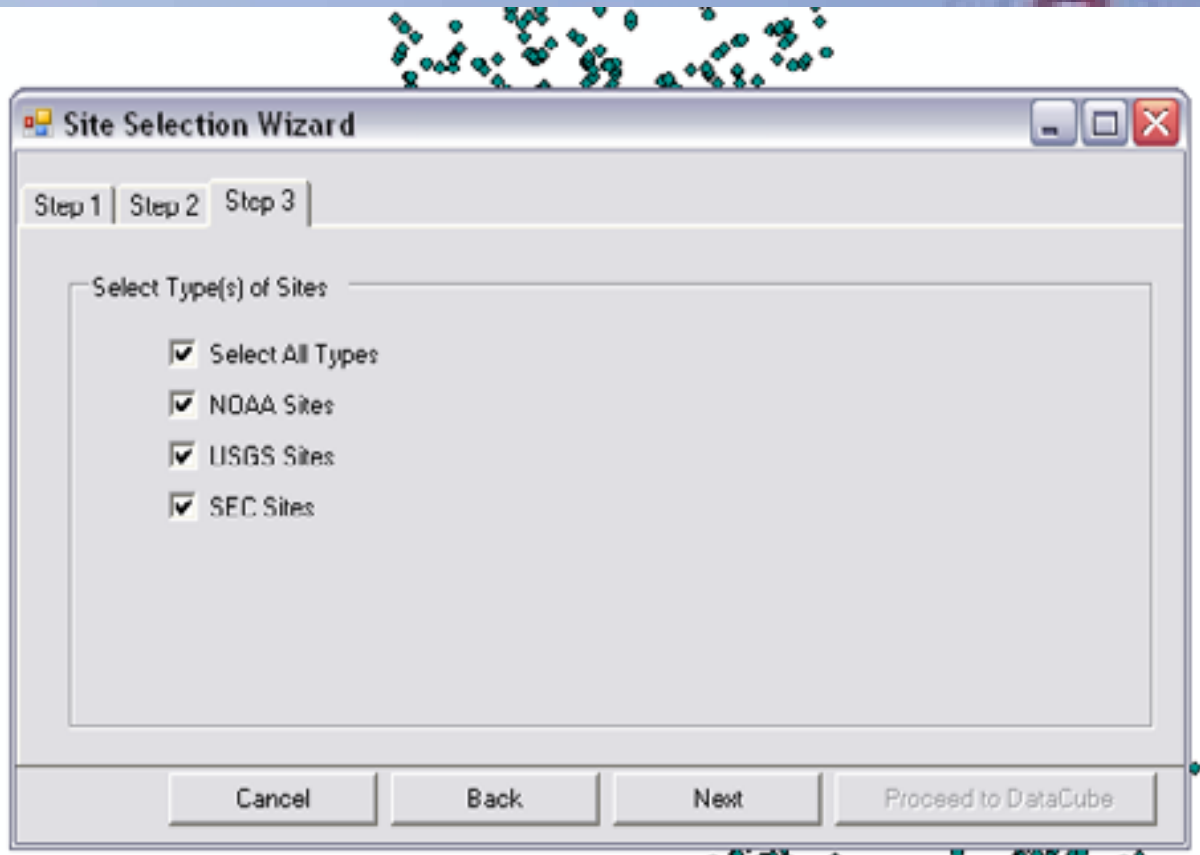
Data Cube Utility Toolbar



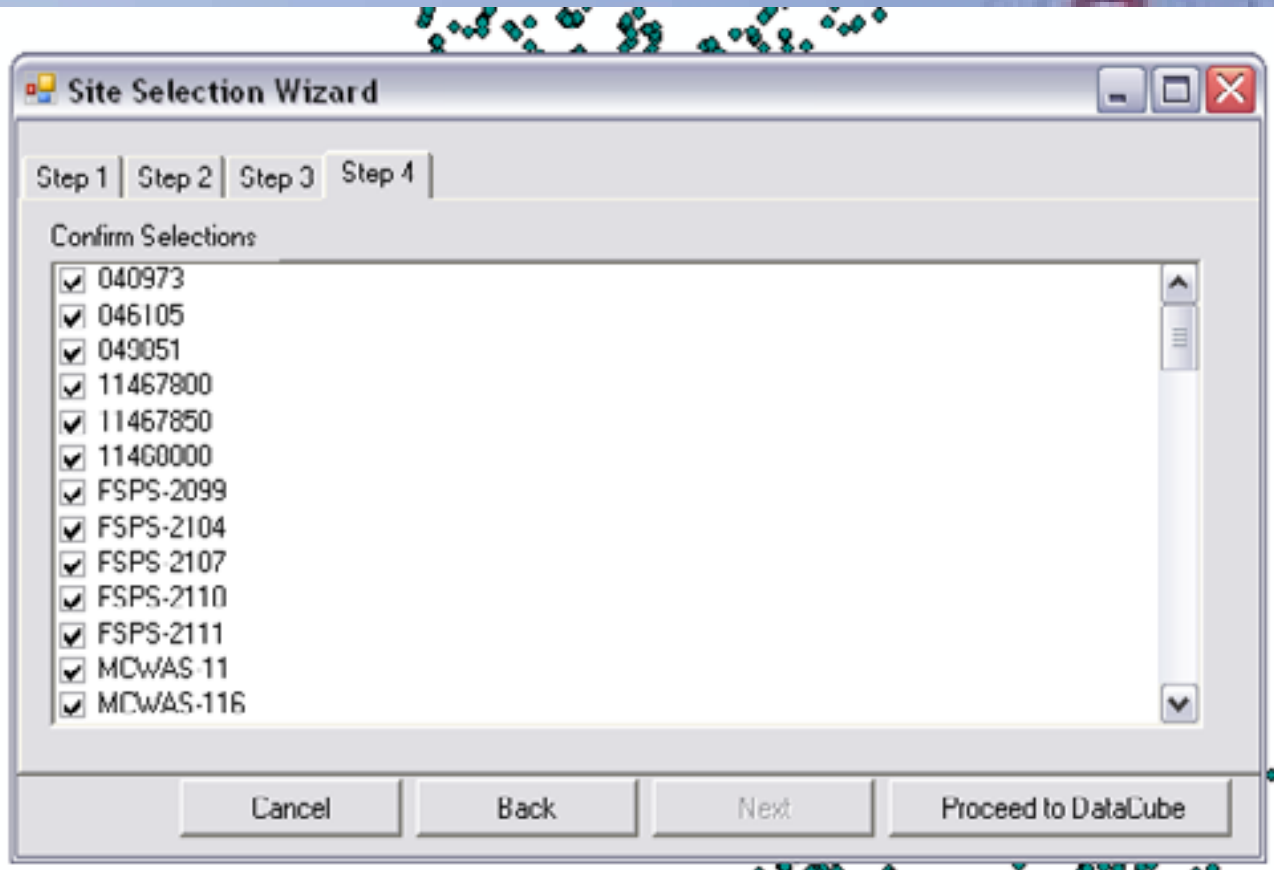
Data Cube Utility Toolbar



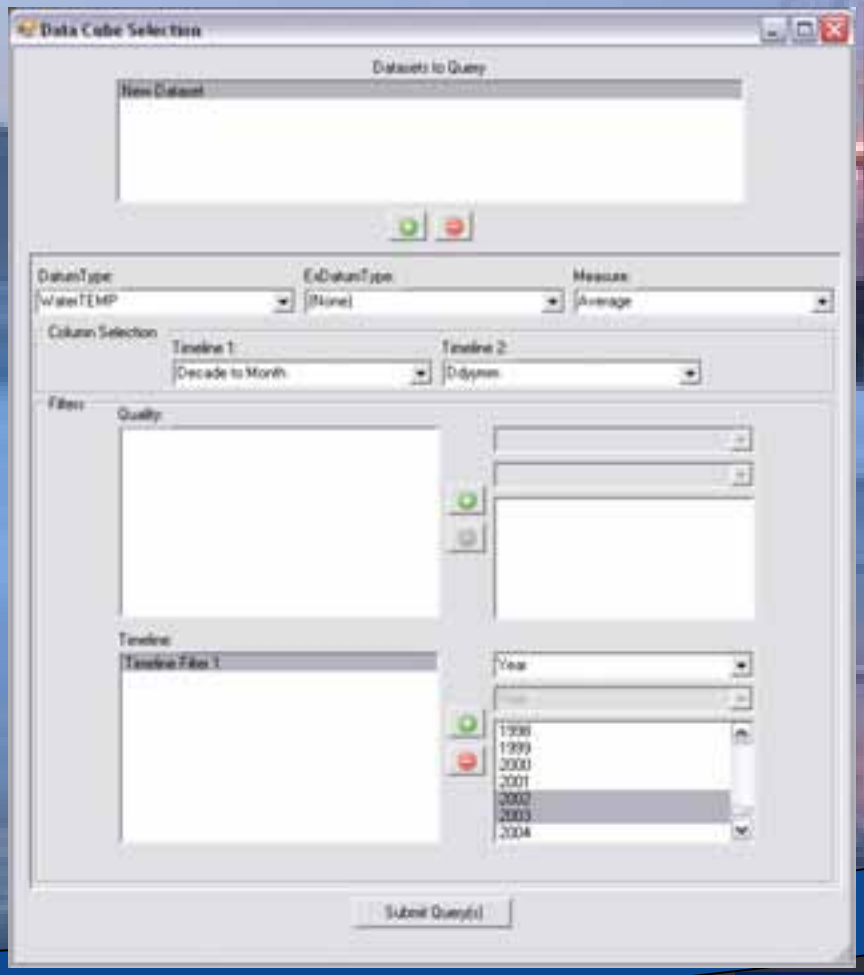
Data Cube Utility Toolbar



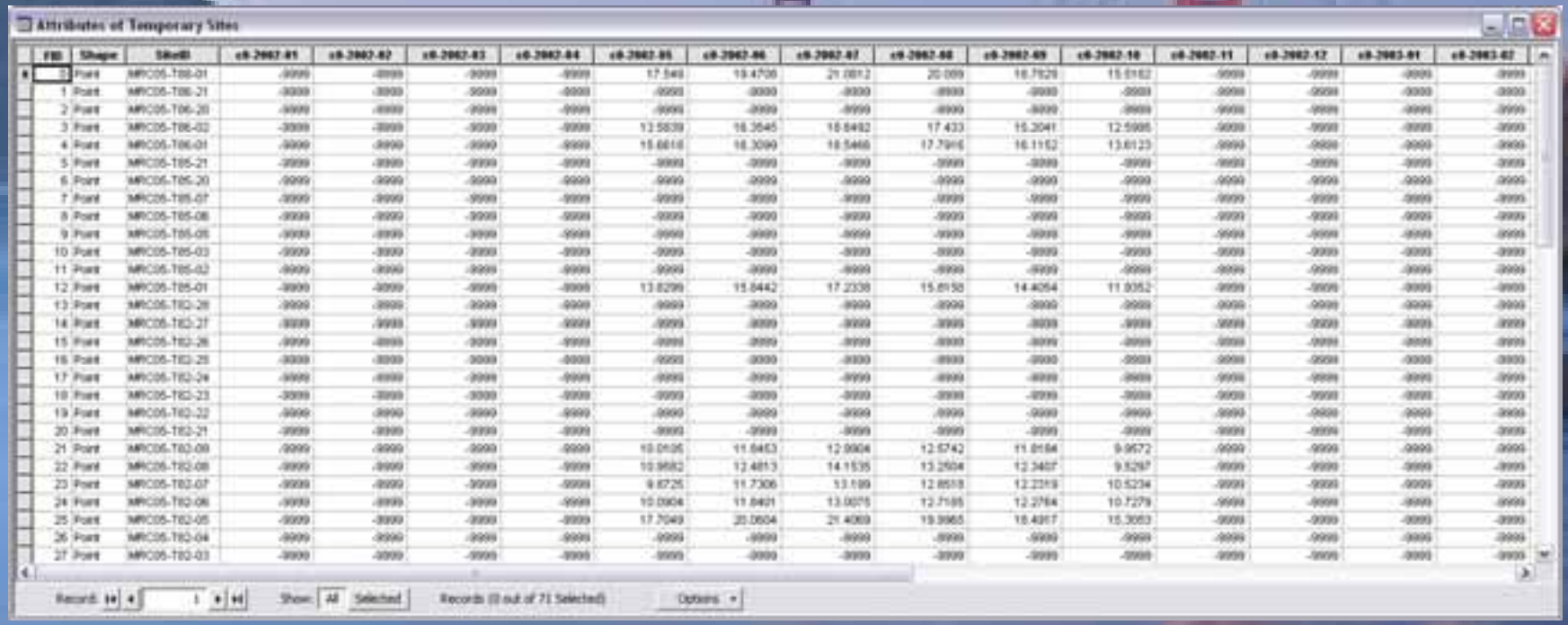
Data Cube Utility Toolbar



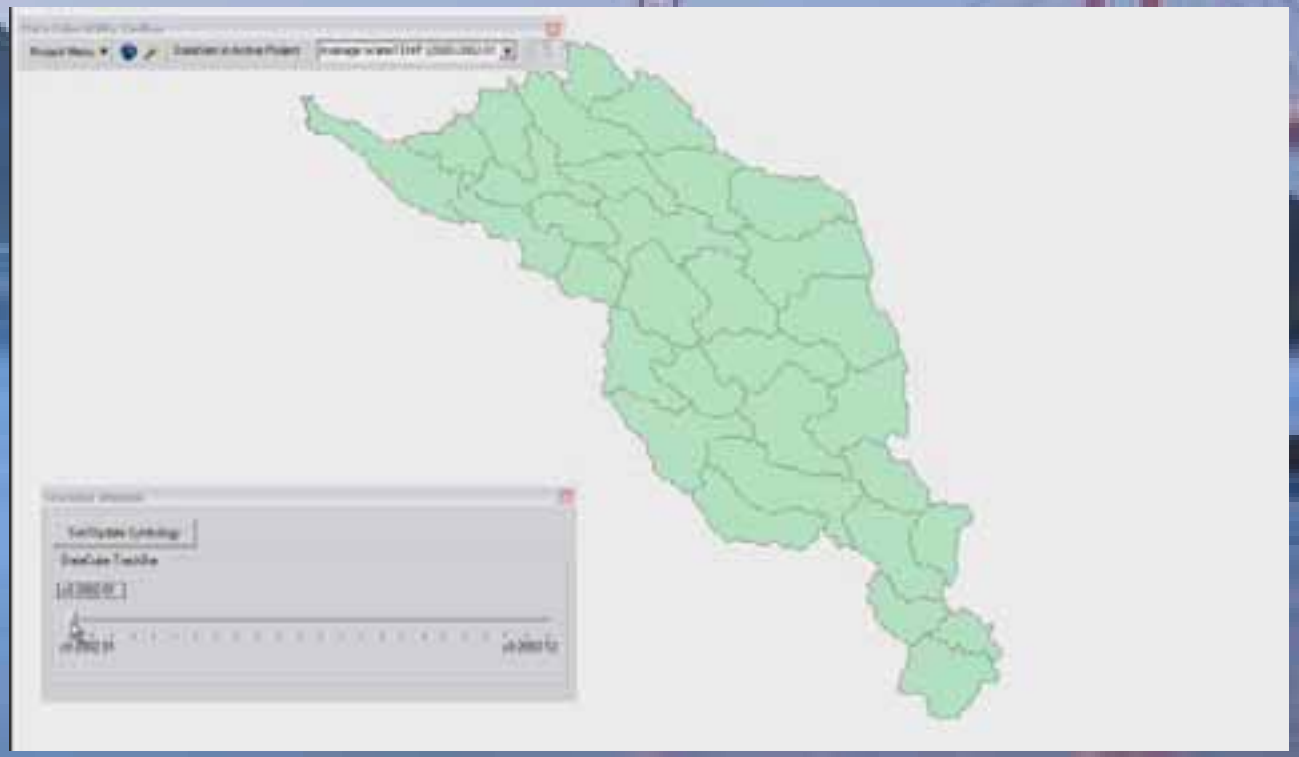
Data Cube Utility Toolbar



Data Cube Utility Toolbar



Data Cube Utility Toolbar



Future Goals

- Integrate spatial attributes directly into the cube
- Improve post-query speed
- Allow for other cubes
- Incorporate other data types