Vega: A Flexible Data Model for Environmental Time Series Data

L. A. Winslow, B. J. Benson, K. E. Chiu, P. C. Hanson, T. K. Kratz

Storing High Resolution Sensor Data in a Relational Database

- Deploy system
- Create data table
- Date/Time column
- Each variable is unique column

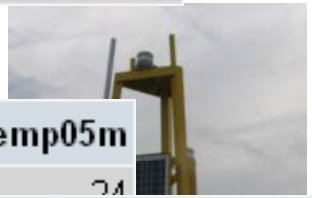
Mendota_Buoy_Table:

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Accommodate Additional Site

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Create Additional Table

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Name

- What about 5 sites?
- Or 10?


SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Mendota_

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Long_Lake_Buoy_Table:

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	25	24

Changes in Measured Variables

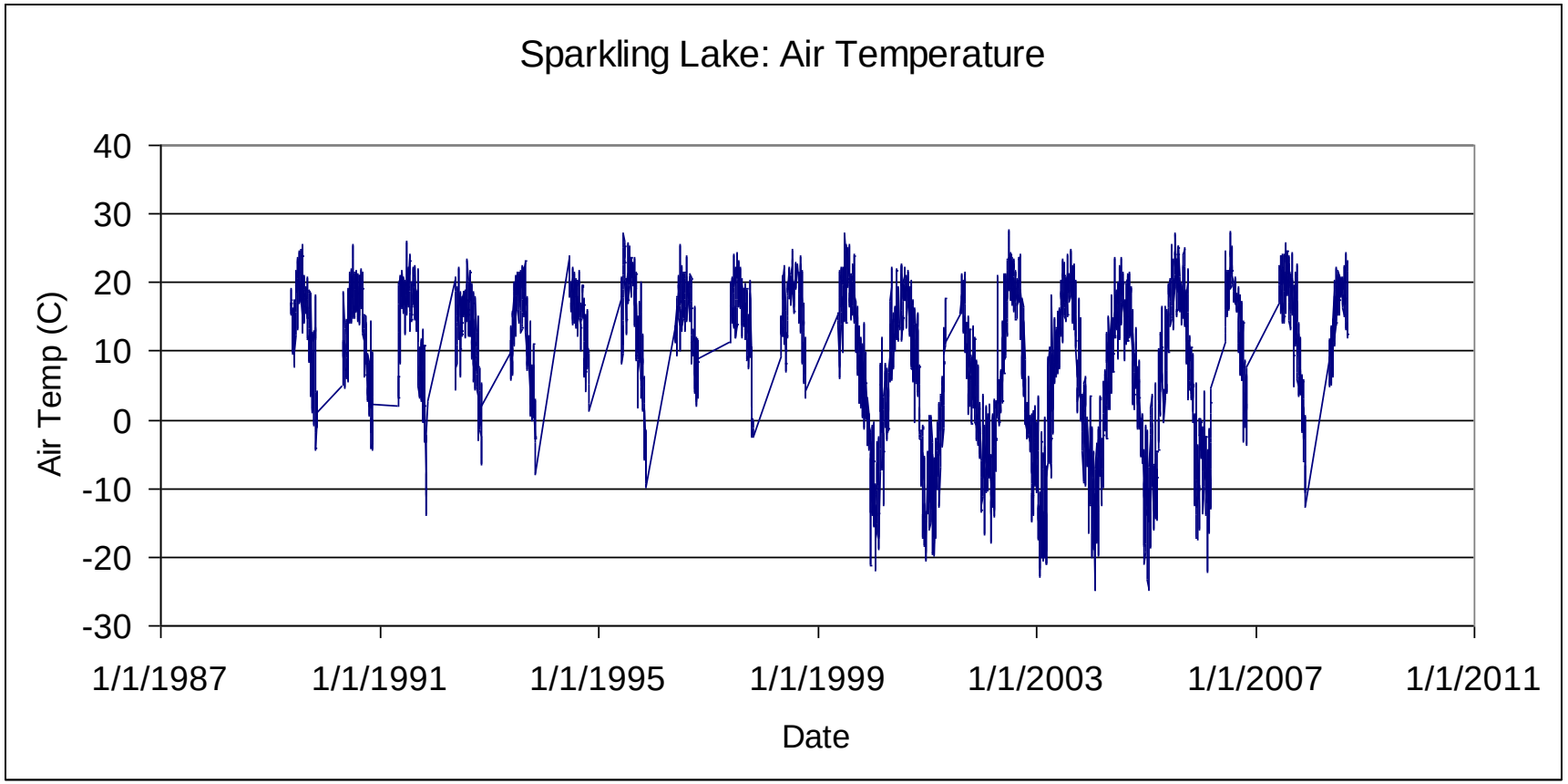
- Add or remove variables
- End up with many NULL fields
- ‘Legacy Structure’

SampleDate	AirTemp	WindSpeed	WindDir	DO_05m	DOSat	Phycocyanin	Water_temp00m	Water_temp05m
2008-09-08 12:44:24	25	5	355	8	95	*NULL*	25	24
2008-09-08 23:00:00	20	4	345	8	100	150	25	24

Add Complex Metadata

- Add Metadata
  - Sensor Info
  - Data steward
  - Offset (depth, height)
  - Sampling Method
- Combine in Field Name
  - DO_05M
  - DO_DOPTO_05M
  - DO_YSI_10M
  - DO_YSI_CALIBRATED_10M
  - WIND_SPEED_VECTOR_AVG

Long-term datasets are becoming more common



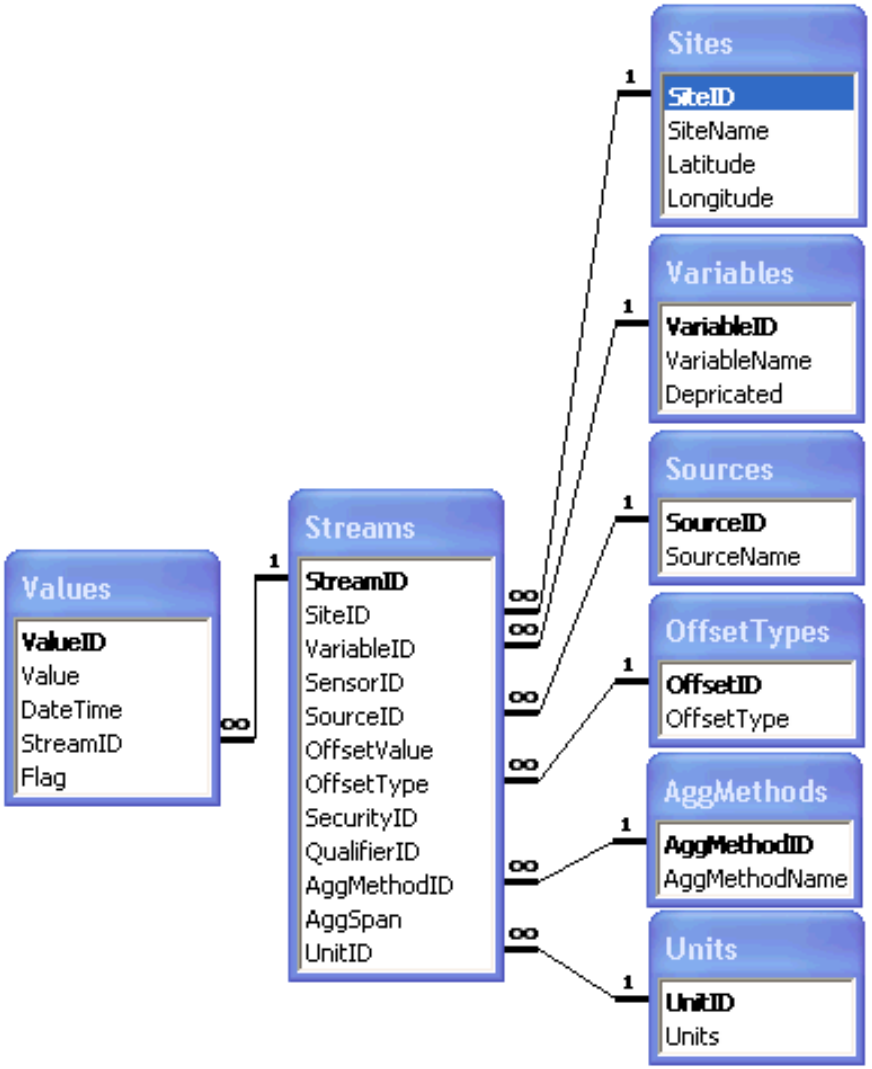
Vega Data Model

- Goals
  - Accommodate dataset changes over time
    - Eliminate legacy structure
  - Easy to understand and develop software
  - Maintain rapid query times
- Inspired by the CUAHSI ODM

Central Concepts

- Values
  - Individual observation (floating point format)
  - Air temp at airport at 12:00 1-1-2007 (-5.1° C)
  - Individually linked to metadata
- Data Streams
  - Group of Values which vary only in time
  - Individual time series
  - All air temp sampled at airport
    - Wind speed is different ‘Data Stream’

Vega: Simple


Sites
SiteID
SiteName
Latitude
Longitude
Variables
VariableID
VariableName
Depricated
Sources
SourceID
SourceName
OffsetTypes
OffsetID
OffsetType
AggMethods
AggMethodID
AggMethodName
Units
UnitID
Units
Streams
StreamID
SiteID
VariableID
SensorID
SourceID
OffsetValue
OffsetType
SecurityID
QualifierID
AggMethodID
AggSpan
UnitID
Values
ValueID
Value
DateTime
StreamID
Flag
1
∞

Indexing

- Speeds up searching through large tables
  - Vega impossible without it
- Similar to an alphabetized phonebook
- With Index:
  - Time ~ Log(number of rows)
- Without Index:
  - Time ~ number of rows
- Values Index (also Unique)
  - DateTime
  - StreamID

Performance

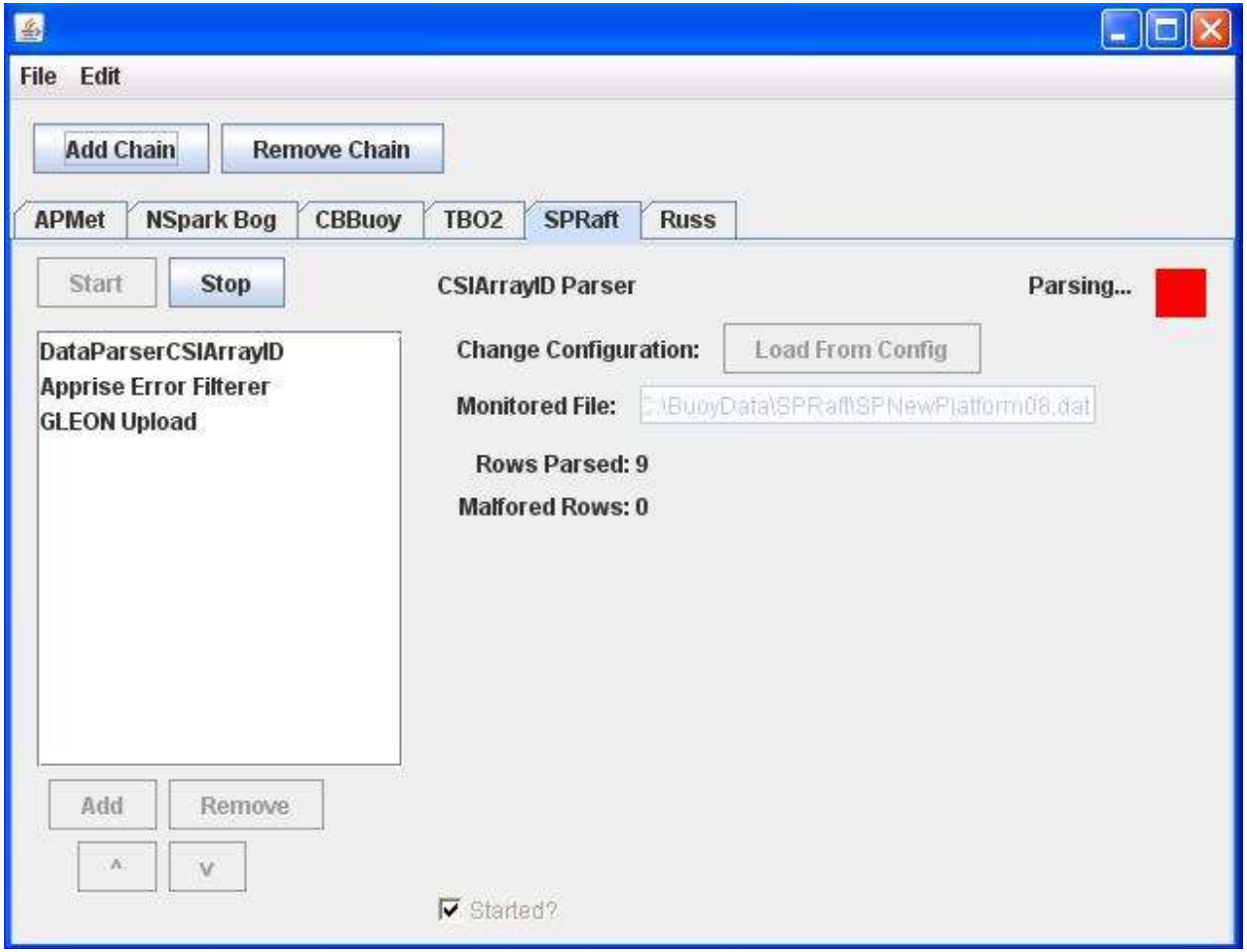
- 40 million Value Database Time to Query
  - One Value: 0.07 Sec
  - ~20k Values: 0.5 Sec

- Data Volumes
  - GLEON ~90,000 new values per day
  - Currently storing 30 million values
  - Values table 2.6 GB

Software Development Gains

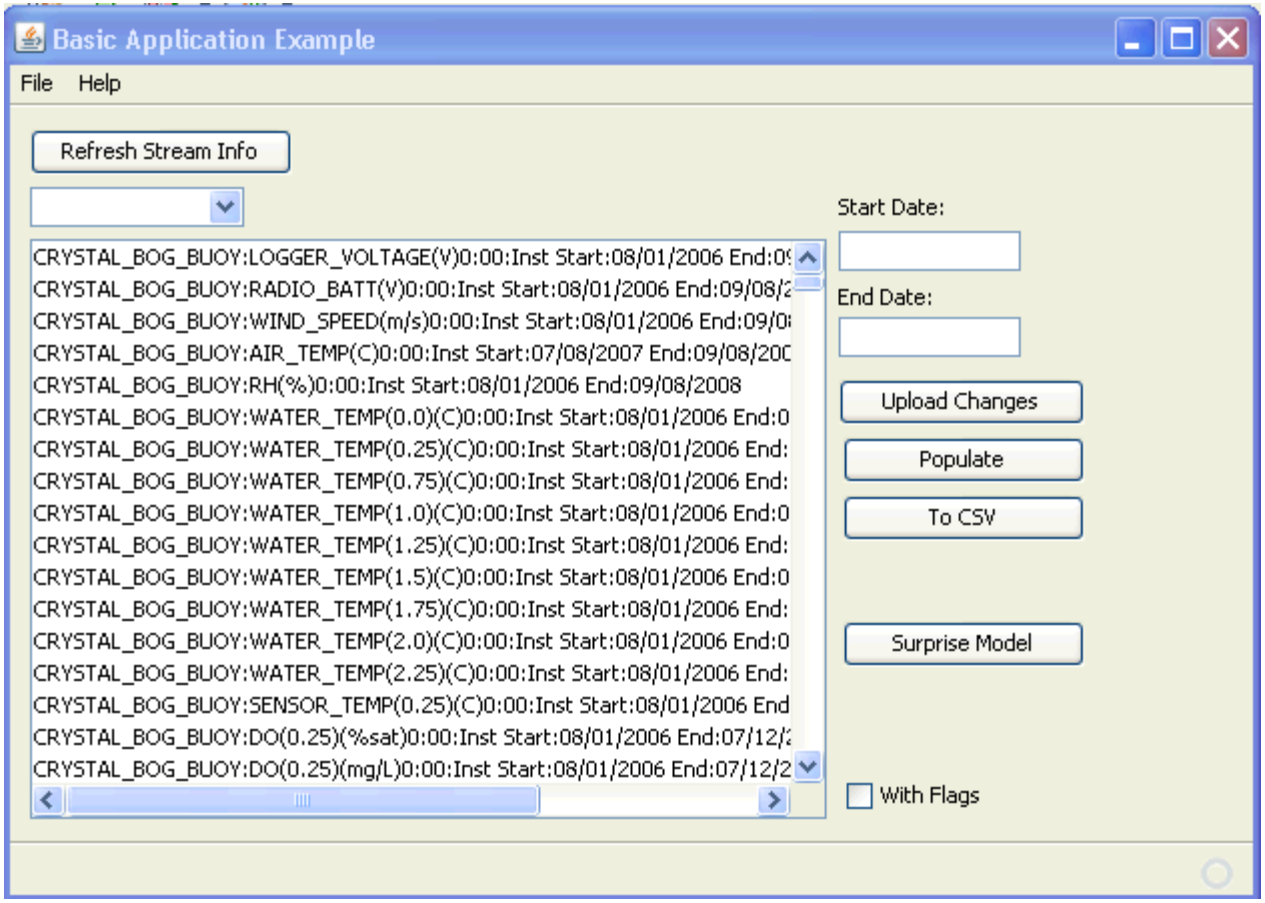
- Software for one site works for all sites
- Example: HTML
  - Many document formatting standards
  - HTML emerged as standard
  - Millions of websites can be read by one browser

Current software for GLEON and Madison LTER: Data Acquisition



Data Retrieval: dbBadger.gleonrcn.org

Data QA/QC


Basic Application Example
File
Help
Refresh Stream Info
CRYSTAL_BOG_BUOY:LOGGER_VOLTAGE(V)0:00:Inst Start:08/01/2006 End:0
CRYSTAL_BOG_BUOY:RADIO_BATT(V)0:00:Inst Start:08/01/2006 End:09/08/
CRYSTAL_BOG_BUOY:WIND_SPEED(m/s)0:00:Inst Start:08/01/2006 End:09/0
CRYSTAL_BOG_BUOY:AIR_TEMP(C)0:00:Inst Start:07/08/2007 End:09/08/200
CRYSTAL_BOG_BUOY:RH(%)0:00:Inst Start:08/01/2006 End:09/08/2008
CRYSTAL_BOG_BUOY:WATER_TEMP(0.0)(C)0:00:Inst Start:08/01/2006 End:0
CRYSTAL_BOG_BUOY:WATER_TEMP(0.25)(C)0:00:Inst Start:08/01/2006 End:
CRYSTAL_BOG_BUOY:WATER_TEMP(0.75)(C)0:00:Inst Start:08/01/2006 End:
CRYSTAL_BOG_BUOY:WATER_TEMP(1.0)(C)0:00:Inst Start:08/01/2006 End:0
CRYSTAL_BOG_BUOY:WATER_TEMP(1.25)(C)0:00:Inst Start:08/01/2006 End:
CRYSTAL_BOG_BUOY:WATER_TEMP(1.5)(C)0:00:Inst Start:08/01/2006 End:0
CRYSTAL_BOG_BUOY:WATER_TEMP(1.75)(C)0:00:Inst Start:08/01/2006 End:
CRYSTAL_BOG_BUOY:WATER_TEMP(2.0)(C)0:00:Inst Start:08/01/2006 End:0
CRYSTAL_BOG_BUOY:WATER_TEMP(2.25)(C)0:00:Inst Start:08/01/2006 End:
CRYSTAL_BOG_BUOY:SENSOR_TEMP(0.25)(C)0:00:Inst Start:08/01/2006 End
CRYSTAL_BOG_BUOY:DO(0.25)(%sat)0:00:Inst Start:08/01/2006 End:07/12/
CRYSTAL_BOG_BUOY:DO(0.25)(mg/L)0:00:Inst Start:08/01/2006 End:07/12/2
Start Date:
End Date:
Upload Changes
Populate
To CSV
Surprise Model
With Flags

Vision

- Simple software package
  - No IT support required
  - Facilitate web-enabled data sharing
- Future
  - Expand to all GLEON sites
  - Include those with custom IM system in place

Acknowledgements

- This work was supported by awards from the National Science Foundation grants DEB-0217533, DBI-0639229, and DBI-0446017 and the Gordon and Betty Moore Foundation.

Performance