

A Query Language for Handling Big Observation Data Sets in the Sensor Web

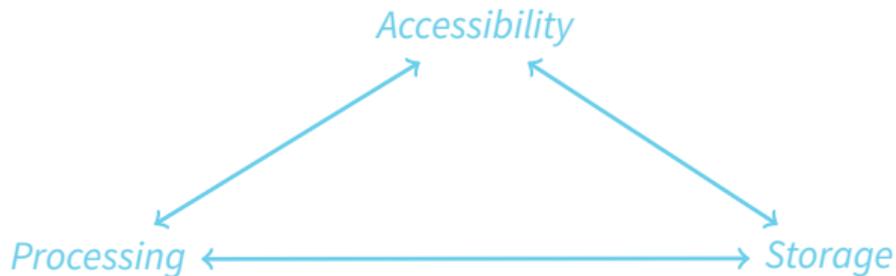
Christian Autermann
c.autermann@52north.org



September 8, 2017

Introduction

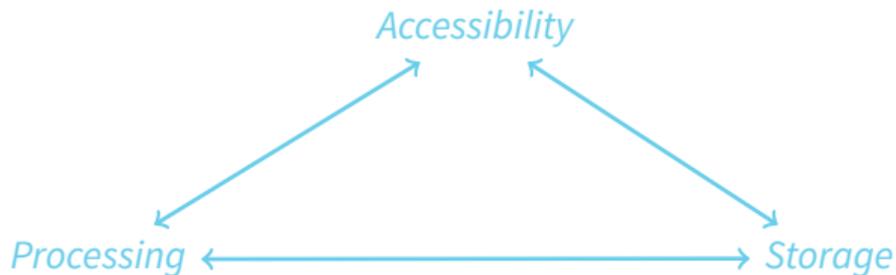
- How can big heterogeneous spatio-temporal datasets be organized, managed and provided to Sensor Web applications?
- How can views on big data sets and derived information products be made accessible in the Sensor Web?
- How can big observation data sets be processed efficiently?





Introduction

- How can big heterogeneous spatio-temporal datasets be organized, managed and provided to Sensor Web applications?
- How can views on big data sets and derived information products be made accessible in the Sensor Web?
- How can big observation data sets be processed efficiently?





Citation:

Patrick, Andrew; Thunell, Robert C (1997): Stable isotope ratios of planktonic foraminifera from LGM sediments of the Pacific Ocean. Dataset #868814 (*DOI registration in progress*), *Supplement to: Patrick, A; Thunell, RC (1997): Tropical Pacific sea surface temperatures and upper water column thermal structure during the Last Glacial Maximum. *Paleoceanography*, 12(5), 649-657, doi:10.1029/97PA01553*

Always quote above citation when using data! You can download the citation in several formats below.

[RIS Citation](#) [Bibtex Citation](#) [Text Citation](#) [Facebook](#) [Twitter](#) [Google+](#) [Show Map](#) [Google Earth](#)



Abstract:

Using two cores from the eastern and western Pacific, we have attempted to better quantify tropical ocean temperatures during the last glacial in order to determine how this climatically-important region responds to large scale changes in climate forcing. By analyzing the oxygen isotopes of surface dwelling (*G. sacculifer*, *G. ruber*), thermocline dwelling (*N. dutertrei*, *G. menardii*, *P. obliquiloculata*) and sub-thermocline dwelling (*G. inflata*) planktonic foraminifera, both relative and absolute estimates of the changes in the temperature gradient over this depth interval have been made. Owing to poor carbonate preservation in the Holocene section of both cores, relative temperature estimates suggest only a slight glacial cooling (~2°C) at these locations, similar to that reported by CLIMAP [1976, 1981]. However, absolute temperature estimates determined from calcite-seawater paleothermometry indicate the eastern equatorial Pacific (EEP) was ~3°C cooler during the last glacial maximum (LGM), while the western equatorial Pacific (WEP) was ~4°C cooler. The upper water column appears to have been less stratified in the EEP, with a steeper thermocline, interpreted as indicating an increase in upwelling during the LGM. The WEP maintained a well developed mixed layer and deep thermocline, similar to today. These results are consistent with a variety of recent tropical temperature estimates for the LGM.

Other version:

[Tropical Pacific SST reconstruction](#) 🔍

Coverage:

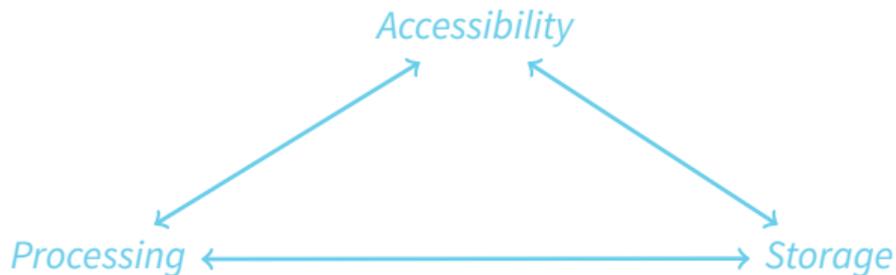
*Median Latitude: -1.808350 * Median Longitude: -142.983350 * South-bound Latitude: -3.616700 * West-bound Longitude: 158.000000 * North-bound Latitude: 0.000000 * East-bound Longitude: -83.966700*

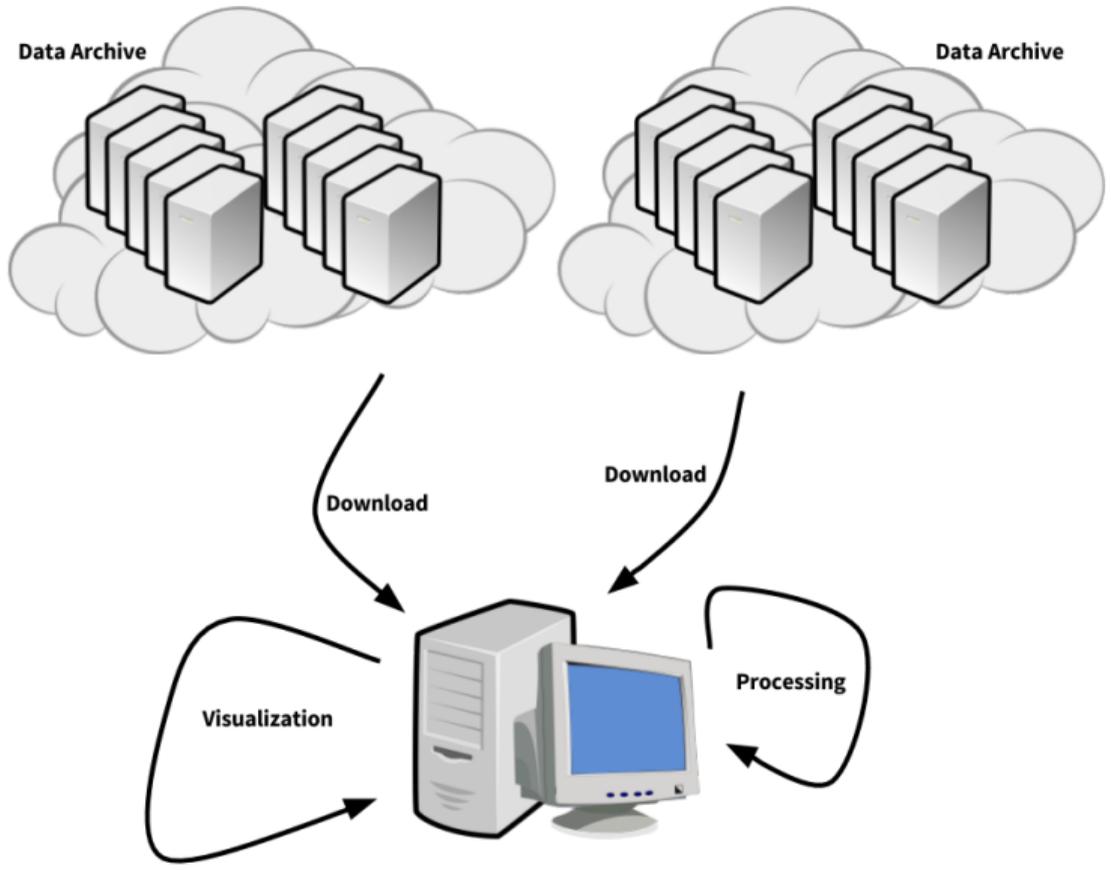
License:

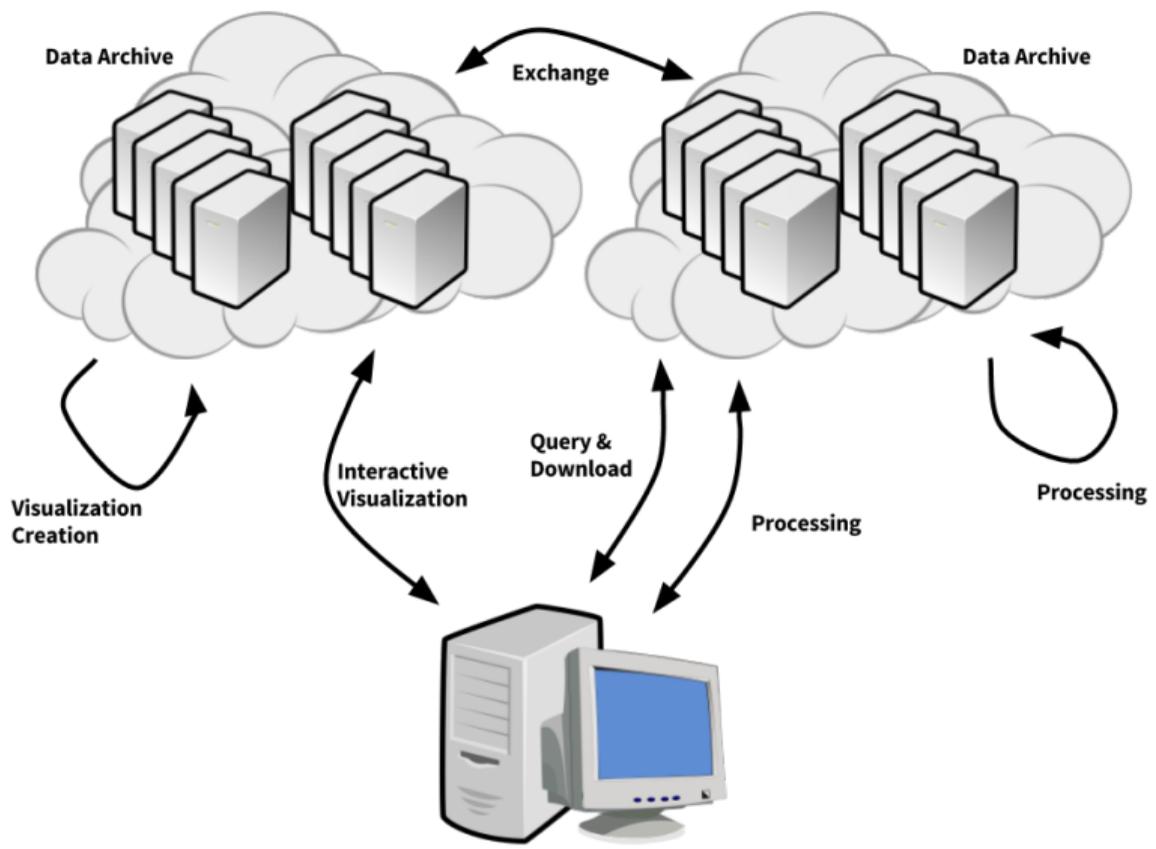
Creative Commons Attribution 3.0 Unported

Introduction

- How can big heterogeneous spatio-temporal datasets be organized, managed and provided to Sensor Web applications?
- How can views on big data sets and derived information products be made accessible in the Sensor Web?
- How can big observation data sets be processed efficiently?







Accessibility

How can views on big data sets and derived information products be made accessible in the Sensor Web?

- Typical request scenarios for search, download, visualization and processing?
- Sensor Web standards capable of and suitable for handling massive observation data sets?
- Conceptual models and encodings? O&M or NetCDF?
- Service Standards? SOS natural choice, what about WPS, WCS/WCPS?

Processing

How can big observation data sets be processed efficiently?

- How does the underlying storage structure influence performance?
- How does the WPS handle situations in which transferring datasets is hard to achieve?
- Can the WPS be used as a Rich-Data-Interface for big observation databases?
- How can predefined, parameterized or even interactive analyses be realized?
- How could a query language that enables on-demand analysis of time series data look like?
- How could a combined analysis of multiple datasets of different origins be accomplished with such high volumes of data?

Storage

How can big heterogeneous spatio-temporal datasets be organized, managed, and provided to Sensor Web applications?

- Requesting a subset of a timeseries vs. requesting the measurements of a phenomenon across multiple sensors at a single point in time
- How can queries in both request dimensions be realized?
- Are existing database technologies (e.g. distributed, array or object databases) applicable?
- Does there even exist a solution that offers acceptable trade offs between the different requirements?

Focus

- What are the core building blocks needed to bring big observational data sets to the SensorWeb?
 - Development of an optimal storage structure for time series data
 - Development of a gradual migration strategy for existing data
 - Development of an SOS interface extension that enables on-demand analysis



Legend

- BETR001
 - 81102 - PM10 (µg/m³)
 - 8576
 - 81102 - PM10
-
- ▶ First value at 29.08.12 03:00 (23 µg/m³)
 - ▶ Last value at 28.04.17 13:00 (18 µg/m³)
 - 📄 Data as CSV (Zip-Archiv)
 - [Show Capabilities of the SOS](#)

01.07.12 - 31.07.17

A Query Language for Handling Big Observation Data Sets in the Sensor Web

- Algebra to describe the analysis of timeseries data
 - Based upon the relational algebra
 - Similar to what the WCPS does for the WCS
- What do experts really need?

Requirement Analysis

- Spatio-temporal interpolation, extrapolation and aggregation
 - statistical aggregates like mean, sum, ...
 - more complex temporal aggregations like weekdays, day phases, ...
- Joins between timeseries based upon time and space
 - Do two timeseries observed the same feature of interest?
 - When two observation describe the same instant in time?
 - fuzzy matching
- Temporal Predictions
- Identification of trends or seasonal patterns
- Persisting of results as new timeseries and/or observations

A Query Language for Handling Big Observation Data Sets in the Sensor Web

- Using plain SQL with well known tables?
- Using some sort of SQL derivate?
- Using Python?
- Using R?
- Something else?

Timeseries Identification

- Combersome to identify a timeseries by it's parameters
 - WHERE t.procedure = '...' AND t.observedProperty = '...' AND t.featureOfInterest = '...' AND ...

→ An identifier is needed

- Actually: what is a timeseries?
- Do different cruises of a research vessel represent the same timeseries?
- Does replacing a sensor on a platform constitute a new timeseries?

→ SOS offering as the basic timeseries identifier

- Not as flexible in SOS 2.0, but the best we got
- unfortunately not part of O&M

Outlook

- Currently in early design phase
- Inputs from potential users is needed
- First implementation as an SOS extension

Thanks. *Questions?*