

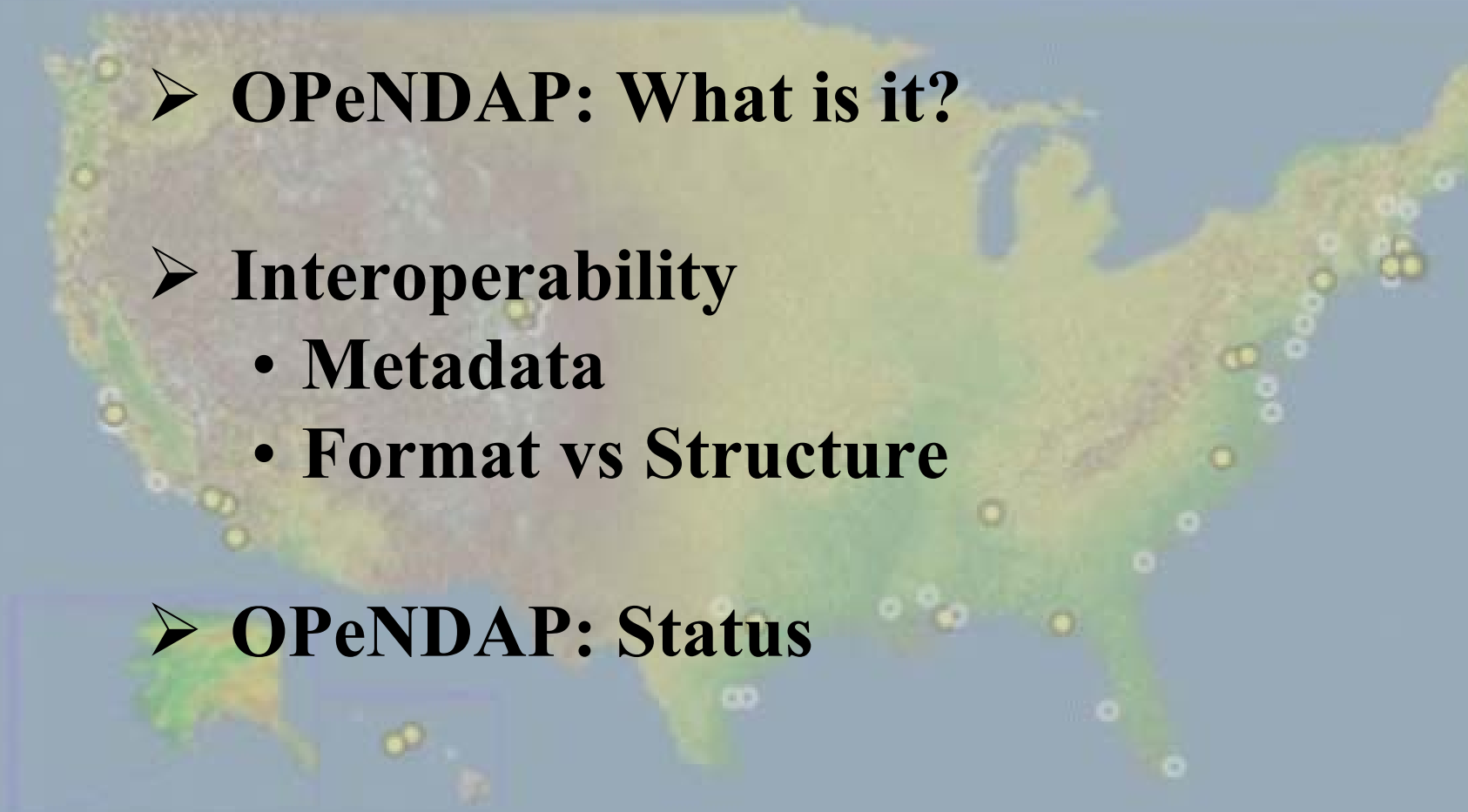
Interoperability in a Distributed, Heterogeneous Data Environment: The OPeNDAP Example

**Peter Cornillon
Graduate School of Oceanography
University of Rhode Island**

Presented at CODATA

2 October 2002

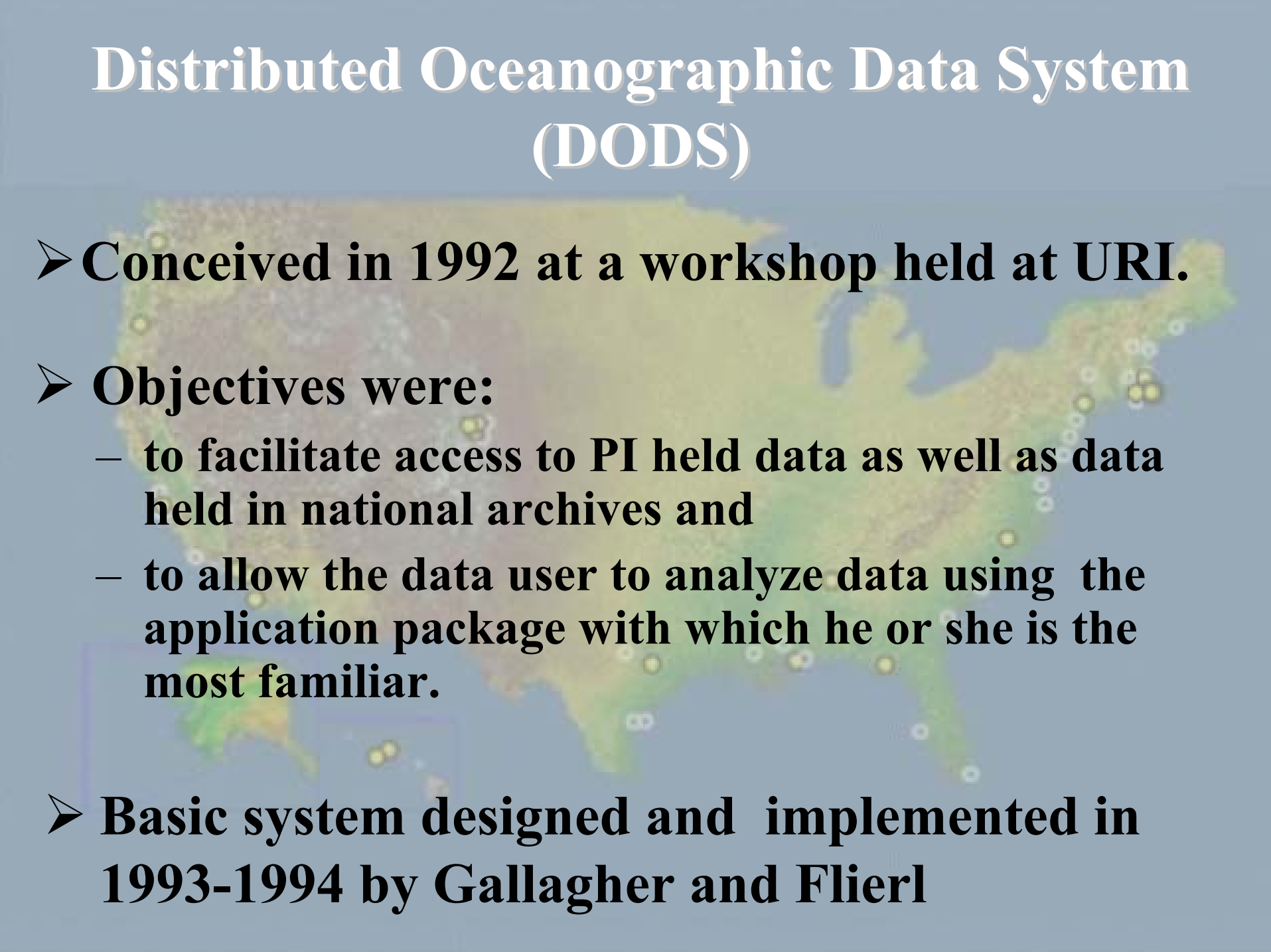
Outline

- 
- **OPeNDAP: What is it?**
 - **Interoperability**
 - **Metadata**
 - **Format vs Structure**
 - **OPeNDAP: Status**

OPeNDAP ← DODS



Distributed Oceanographic Data System (DODS)

- 
- A faint background map of the United States is visible, overlaid with numerous small, semi-transparent circles representing oceanographic data points. These points are scattered across the map, with a higher concentration along the East and West coasts.
- **Conceived in 1992 at a workshop held at URI.**
 - **Objectives were:**
 - to facilitate access to PI held data as well as data held in national archives and
 - to allow the data user to analyze data using the application package with which he or she is the most familiar.
 - **Basic system designed and implemented in 1993-1994 by Gallagher and Flierl**

Distributed Oceanographic Data System

DODS consisted of two fundamental parts:

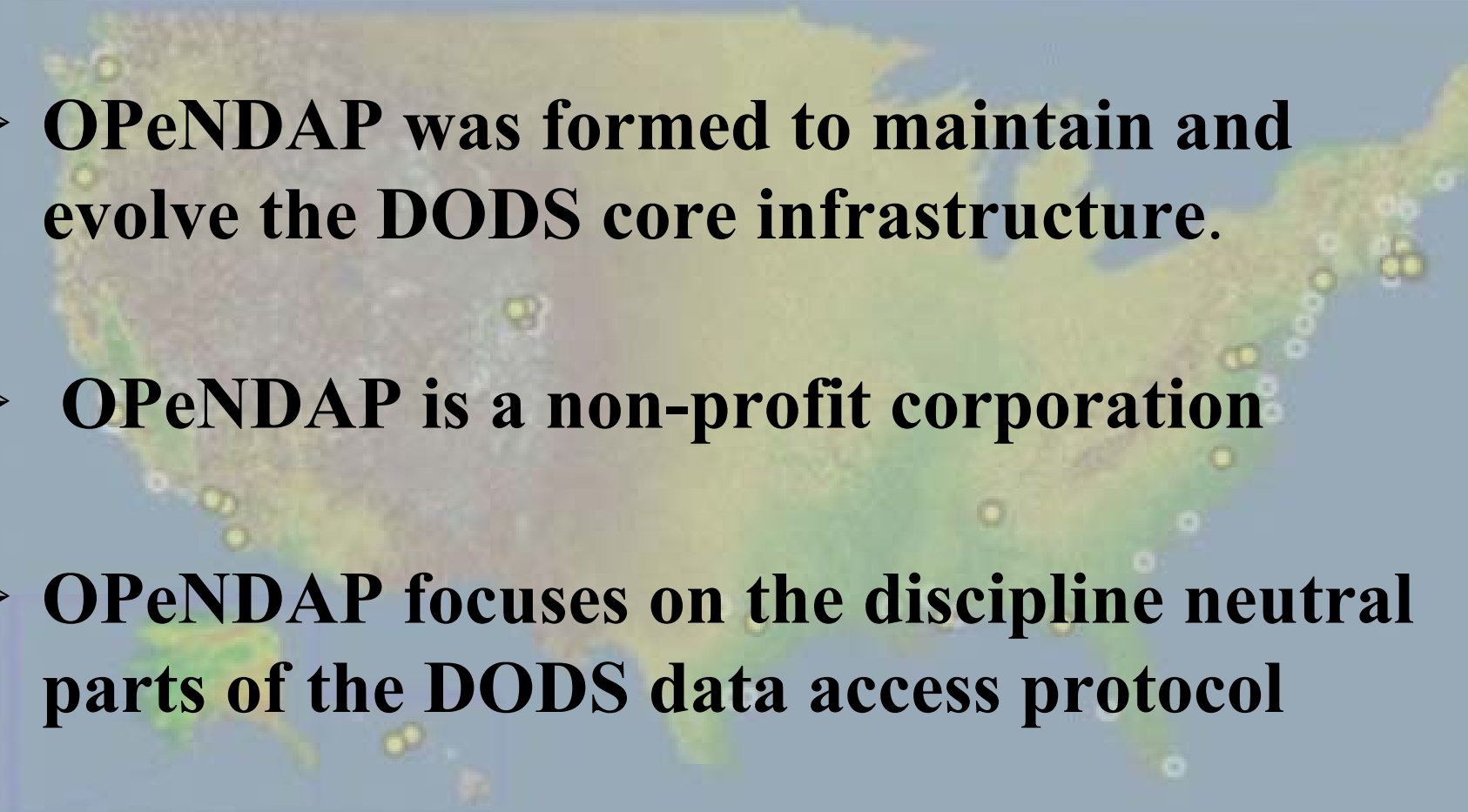
- **a discipline independent core infrastructure for moving data on the net,**
- **a discipline specific portion related to data – data population, data location, specialized clients, etc.**

DODS \Rightarrow OPeNDAP & NVODS

To isolate the discipline independent part of the system from the discipline specific part, two entities have been formed:

- **Open Source Project for a Network Data Access Protocol (OPeNDAP)**
- **National Virtual Ocean Data System (NVODS)**

OPeNDAP

- 
- **OPeNDAP was formed to maintain and evolve the DODS core infrastructure.**
 - **OPeNDAP is a non-profit corporation**
 - **OPeNDAP focuses on the discipline neutral parts of the DODS data access protocol**

Objective of OPeNDAP

- **To provide a data access protocol allowing for machine-to-machine interoperability with semantic meaning**
- ➔ **The scripted exchange of data between computers, without human intervention.**

Interoperability

- **The degree to which machine-to-machine interoperability is achieved depends on the metadata associated with the data.**
- **Metadata is information about the data.**
- **We divide metadata into two basic classes**
 - **Syntactic metadata**
 - **Semantic metadata**

Syntactic **Metadata**

- **Information about the data types and structures at the computer level - the syntax of the data;**
 - e.g., variable T represents a 20x40 element floating point array.
- **Required as part of the transport protocol for data in a network based data system.**
- **The transport protocol is characterized by a (syntactic) data model**

Syntactic Data Model

- **The organizational description of data as they are moved between elements within the system.**
 - generally consists of data types, and
 - groupings of these data types
- **Also includes operations that may be performed on the data – e.g., subsetting**

Semantic **Metadata**

- **Information about the contents of the data set.**
 - e.g., variable T represents
 - sea surface temperature
 - with units of $^{\circ}\text{C}$
 - collected by so-and-so and
 - extends from 40°W to 30°W and 10°N to 30°N

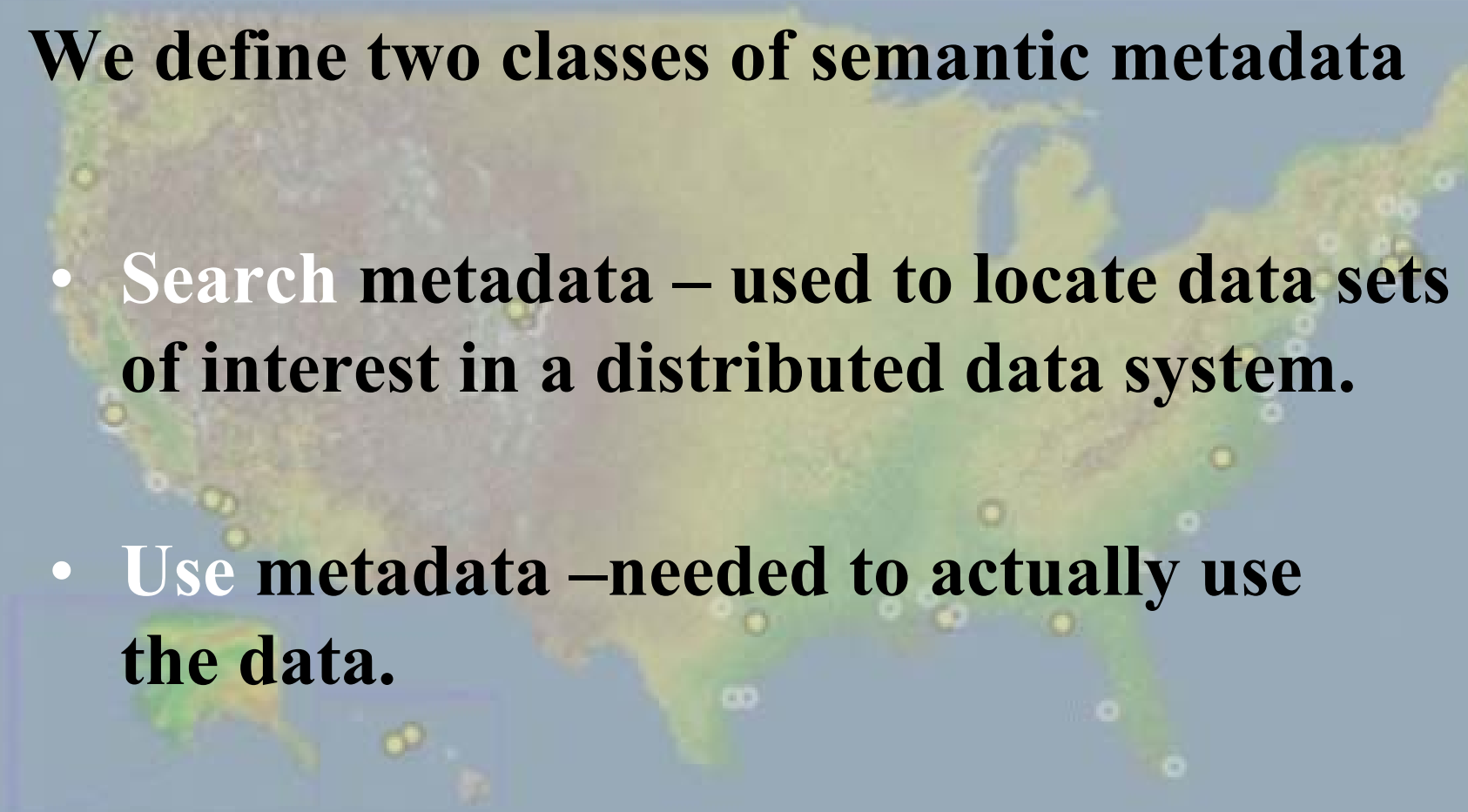


Syntactic vs Semantic Metadata

- **Syntactic metadata are easily constrained.**
- **Semantic metadata are not.**
 - they depend on the use that is to be made of them



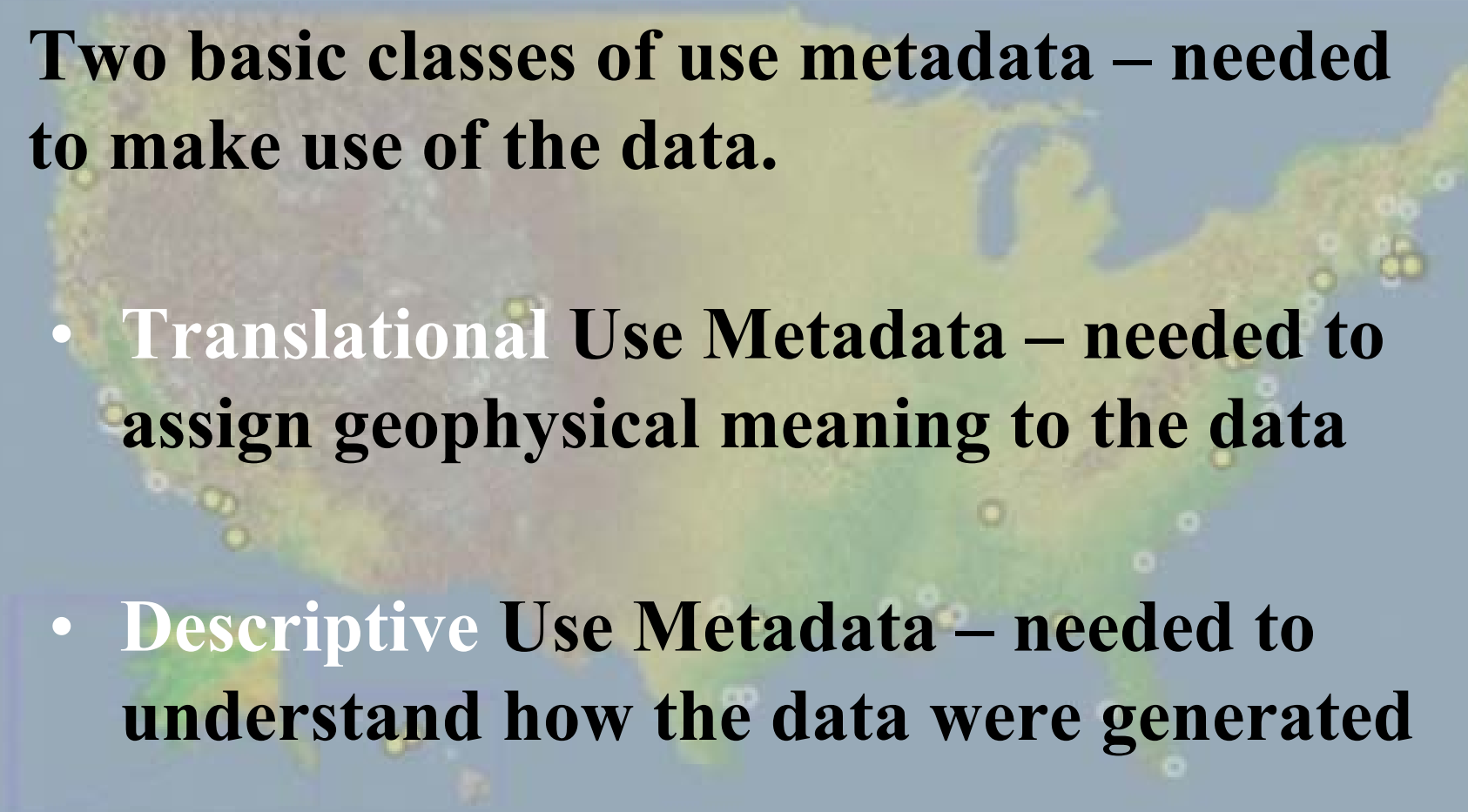
Semantic Metadata Types

- 
- **We define two classes of semantic metadata**
 - **Search metadata – used to locate data sets of interest in a distributed data system.**
 - **Use metadata –needed to actually use the data.**

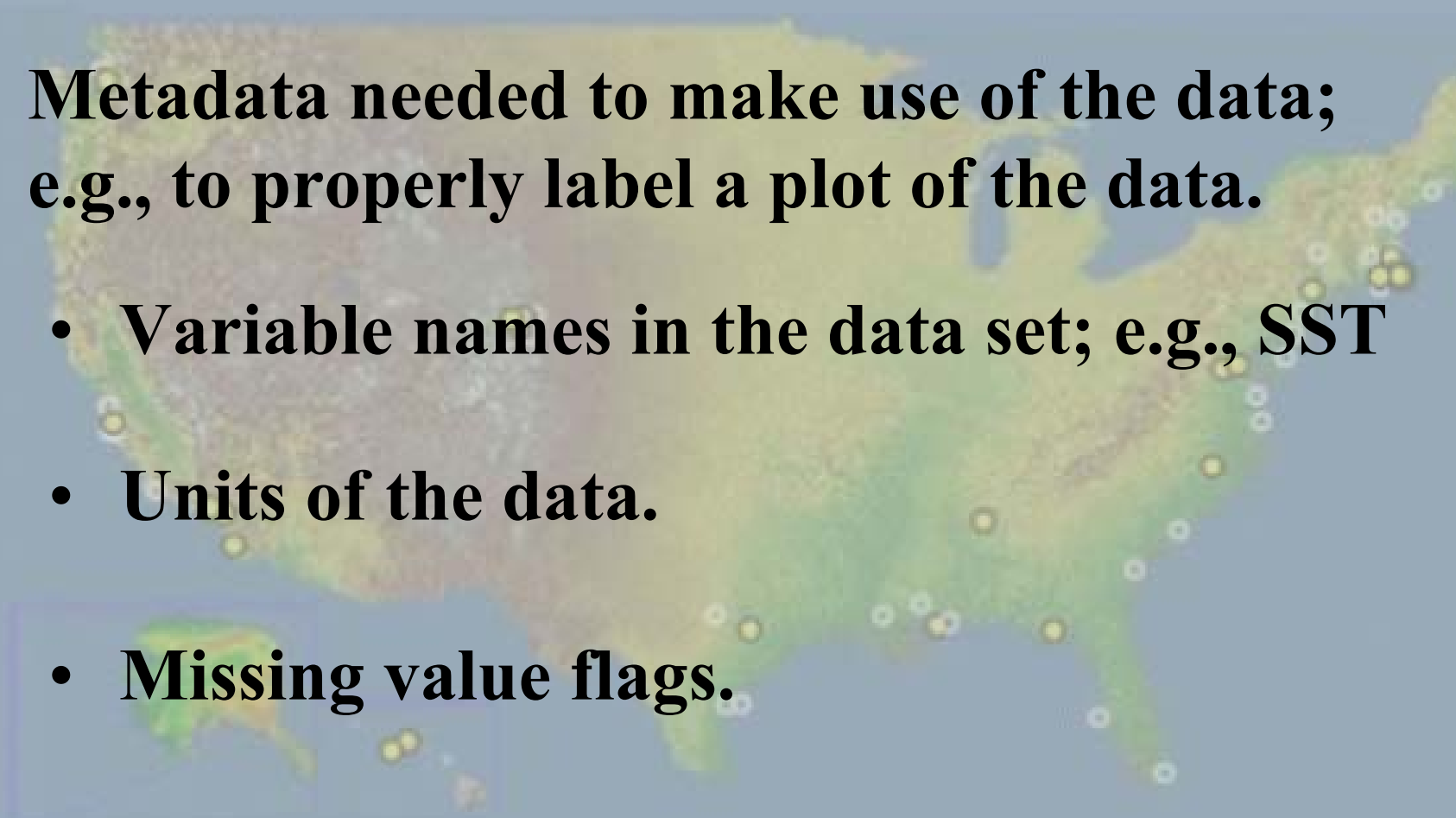
Semantic Search Metadata Types

- We define three search metadata categories
 - **Parameter** – variables in the data set
 - **Range** – the range of values, lat, lon,...
 - **Descriptive** – additional information not generally required for a search such as:
 - Ship from which the data were collected
 - Campaign
 - PI


Semantic Use Metadata Types

- 
- **Two basic classes of use metadata – needed to make use of the data.**
 - **Translational Use Metadata – needed to assign geophysical meaning to the data**
 - **Descriptive Use Metadata – needed to understand how the data were generated**

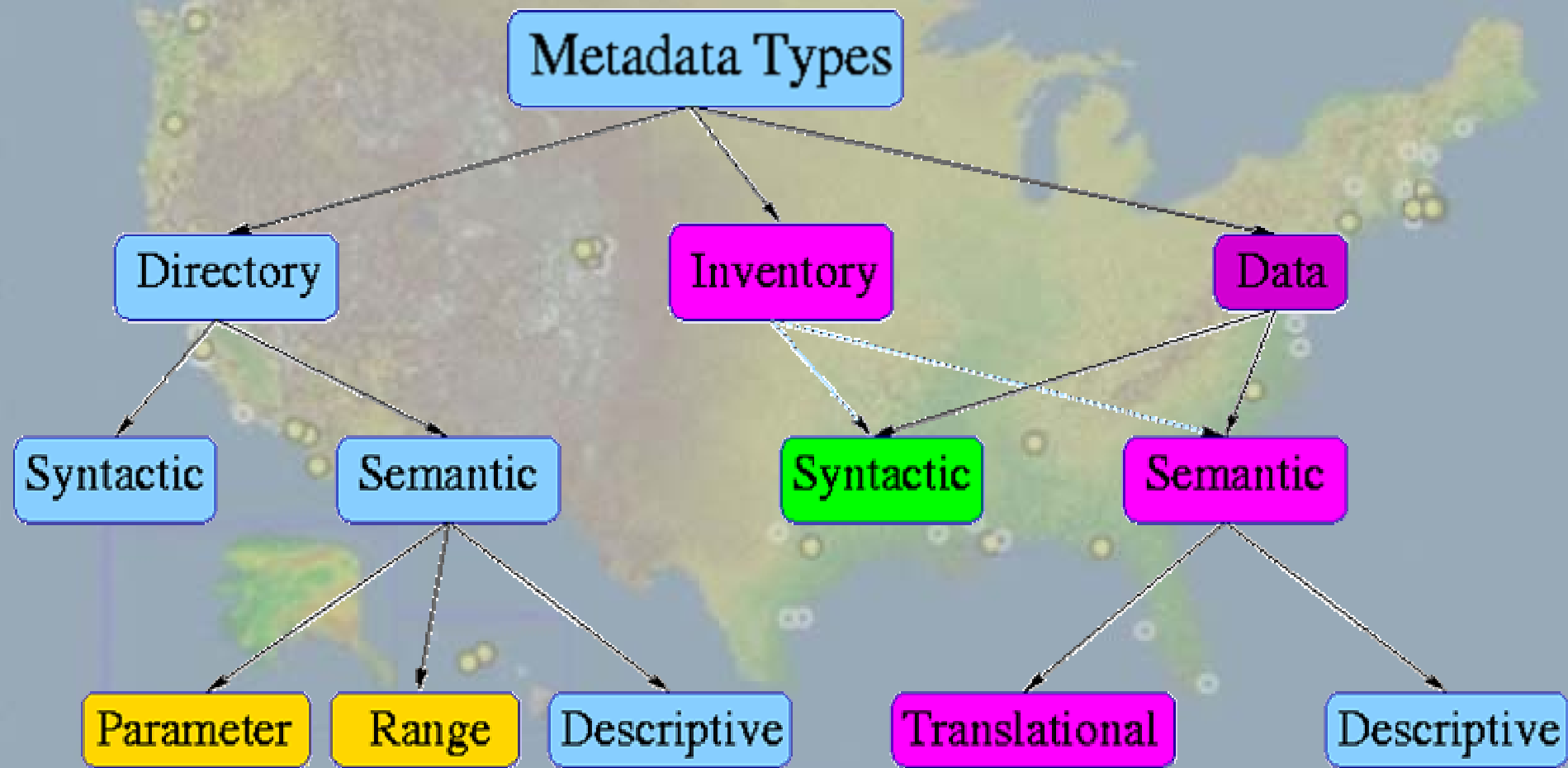
Translational **Use Metadata**

- 
- **Metadata needed to make use of the data; e.g., to properly label a plot of the data.**
 - **Variable names in the data set; e.g., SST**
 - **Units of the data.**
 - **Missing value flags.**

Descriptive **Use Metadata**

- 
- **Metadata needed to understand how the data were generated**
 - **Algorithm used to generate the data**
 - **Calibration of instruments**
 - **Data quality**

OPeNDAP and Metadata

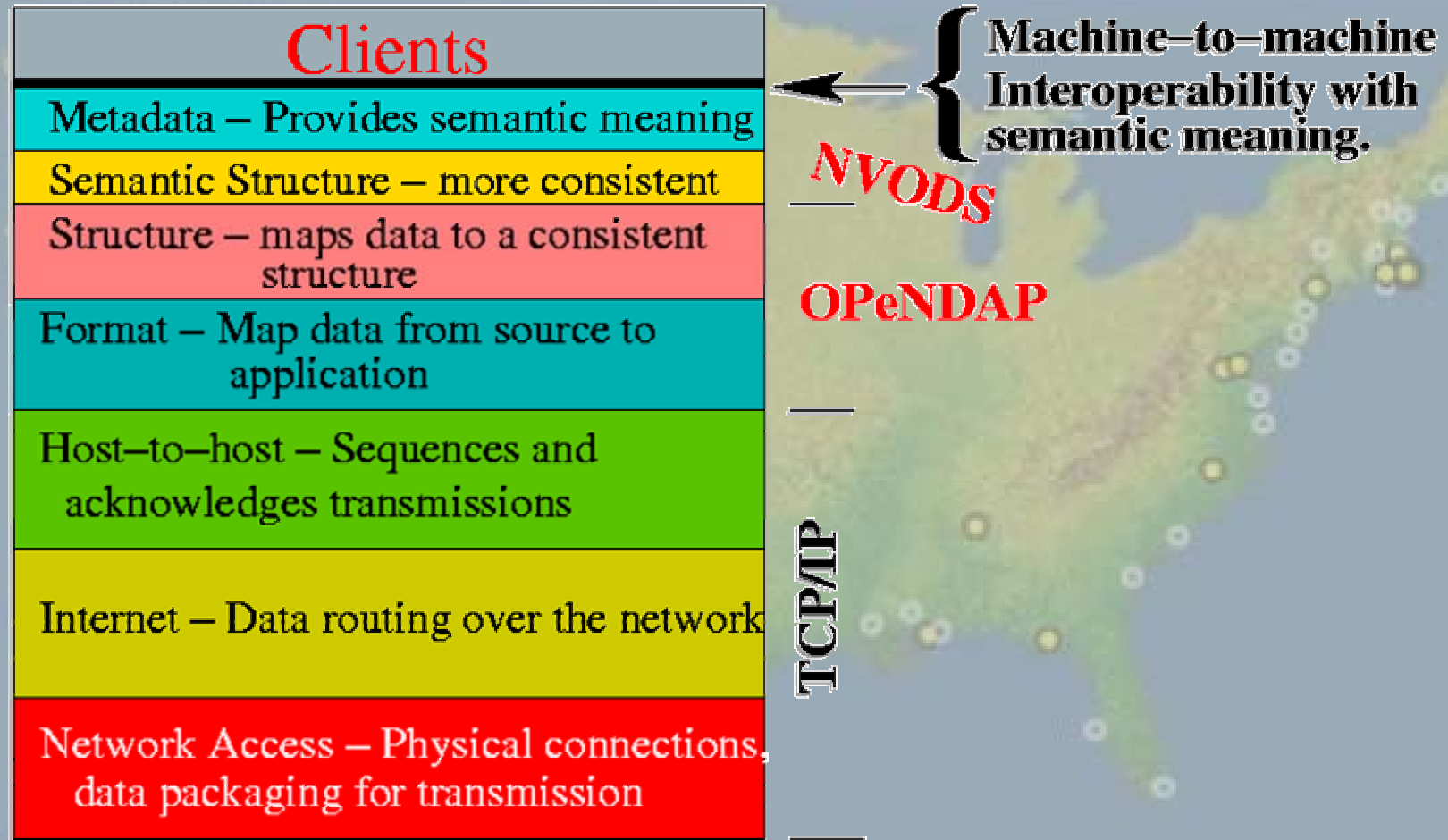


Interoperability

Interoperability may be defined at any one of a number of levels ranging from:

- **the lowest (hardware) - how computers are linked electronically, to**
- **the highest – semantically meaningful, machine-to-machine exchanges (Level 3 Interoperability).**

Layers of Interoperability

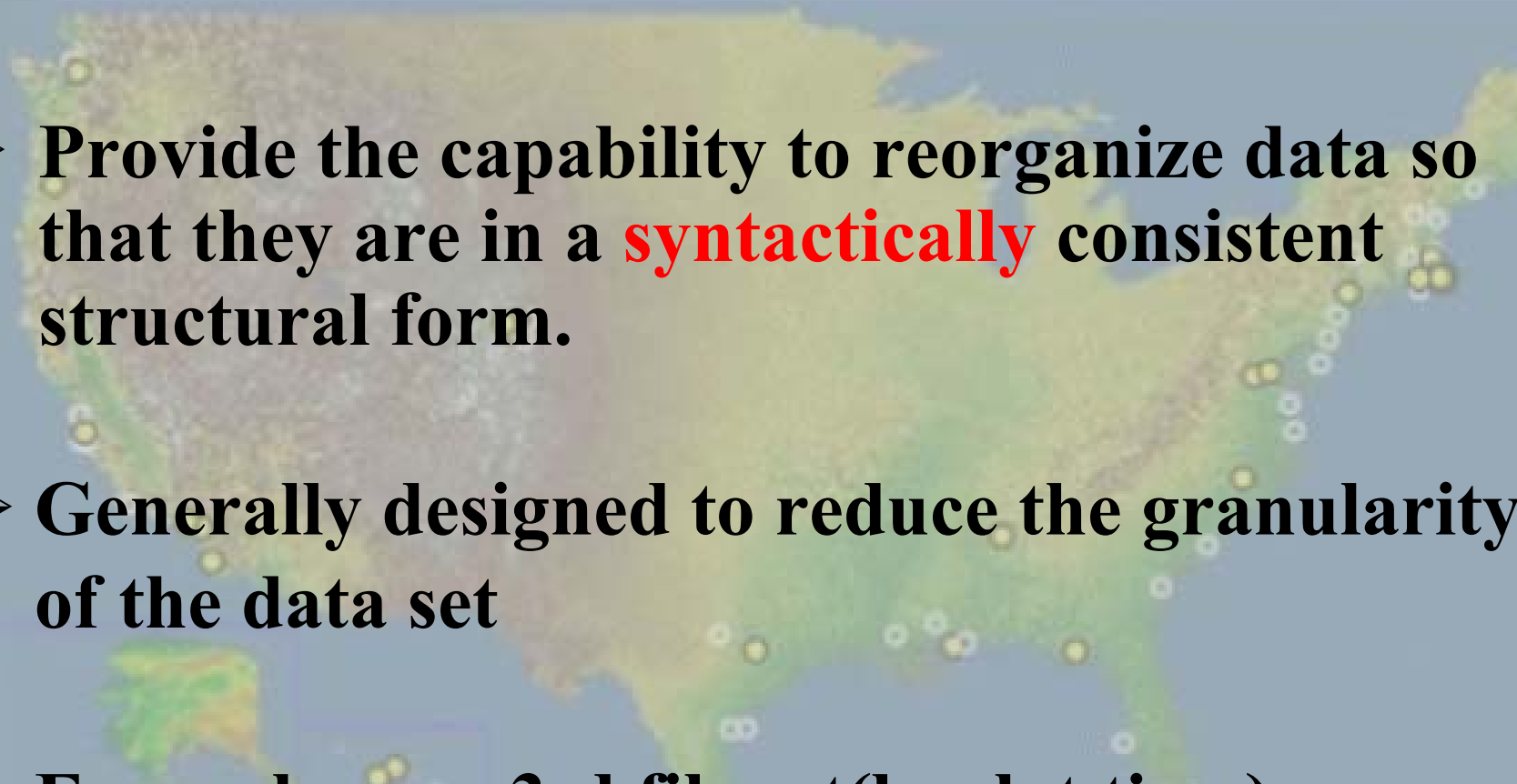


Organizational Complexity


Example: Consider the different ways of organizing a multi-year data set consisting of one global sea surface temperature (SST) field per day:

- **one 2-d file per day $\text{sst}(\text{lat}, \text{lon})$ - URI**
- **one 3-d file $\text{sst}(\text{lon}, \text{lat}, \text{time})$ - PMEL**
- **one file per year with one variable per day $\Rightarrow 365$ variables per file, n files for n year - GSFC**

STRUCTURE LEVEL

- 
- Provide the capability to reorganize data so that they are in a **syntactically** consistent structural form.
 - Generally designed to reduce the granularity of the data set
 - Example: one 3-d file `sst(lon,lat,time)`

SEMANTIC STRUCTURE LEVEL

- 
- Provide the capability to reorganize data so that they are in a **semantically consistent structural form**.
 - Example: one 4-d file `sst(lon,lat,depth,time)`, depth being a null dimension.

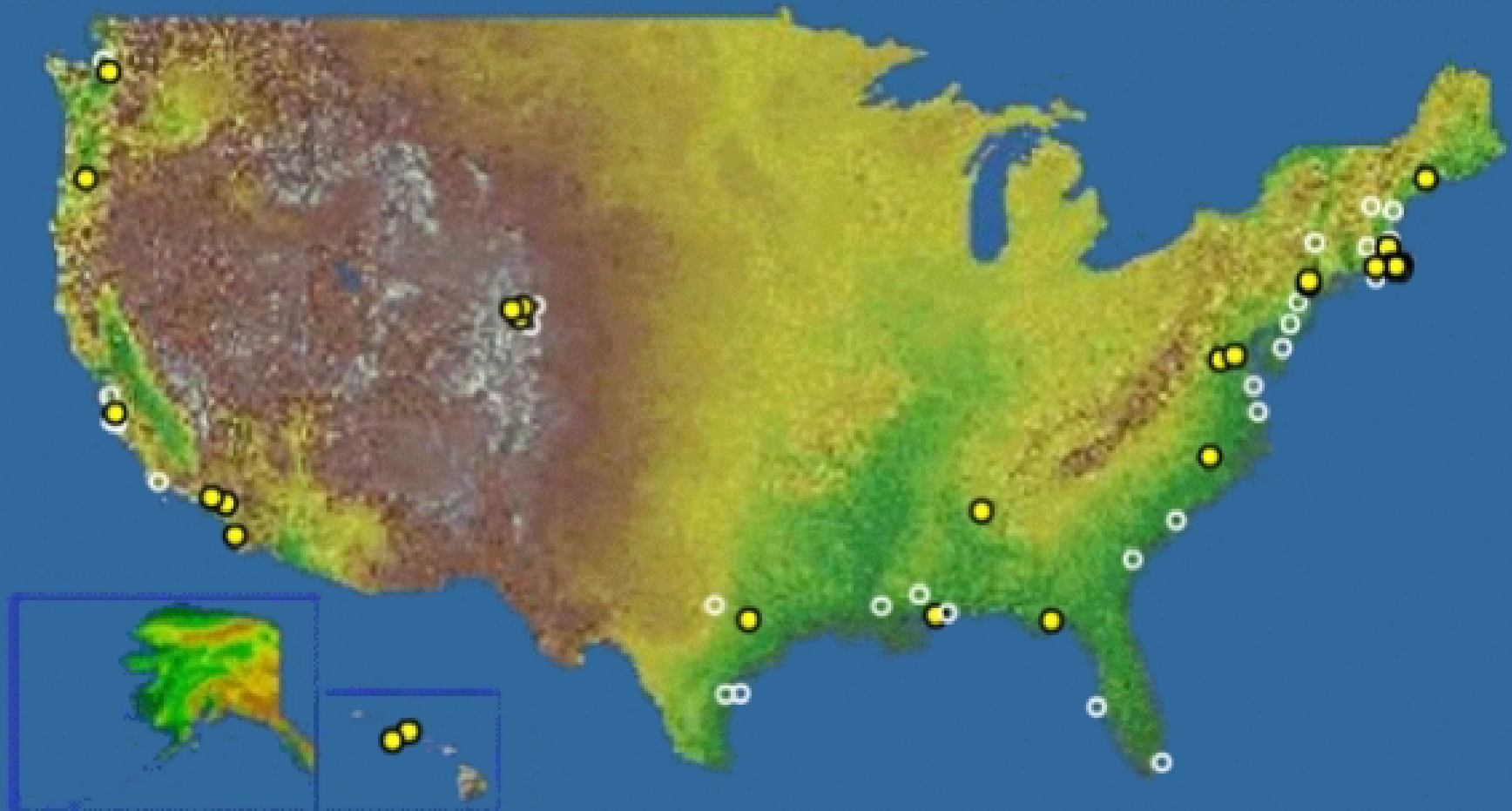
Aggregation Server

- **Developed by John Caron of Unidata with input from CDC, PMEL, URI and OPeNDAP**
- **Is for the aggregation of grids and arrays only**
- **Operates at the syntactic level**
- **Handles three types of aggregation**

OPeNDAP²



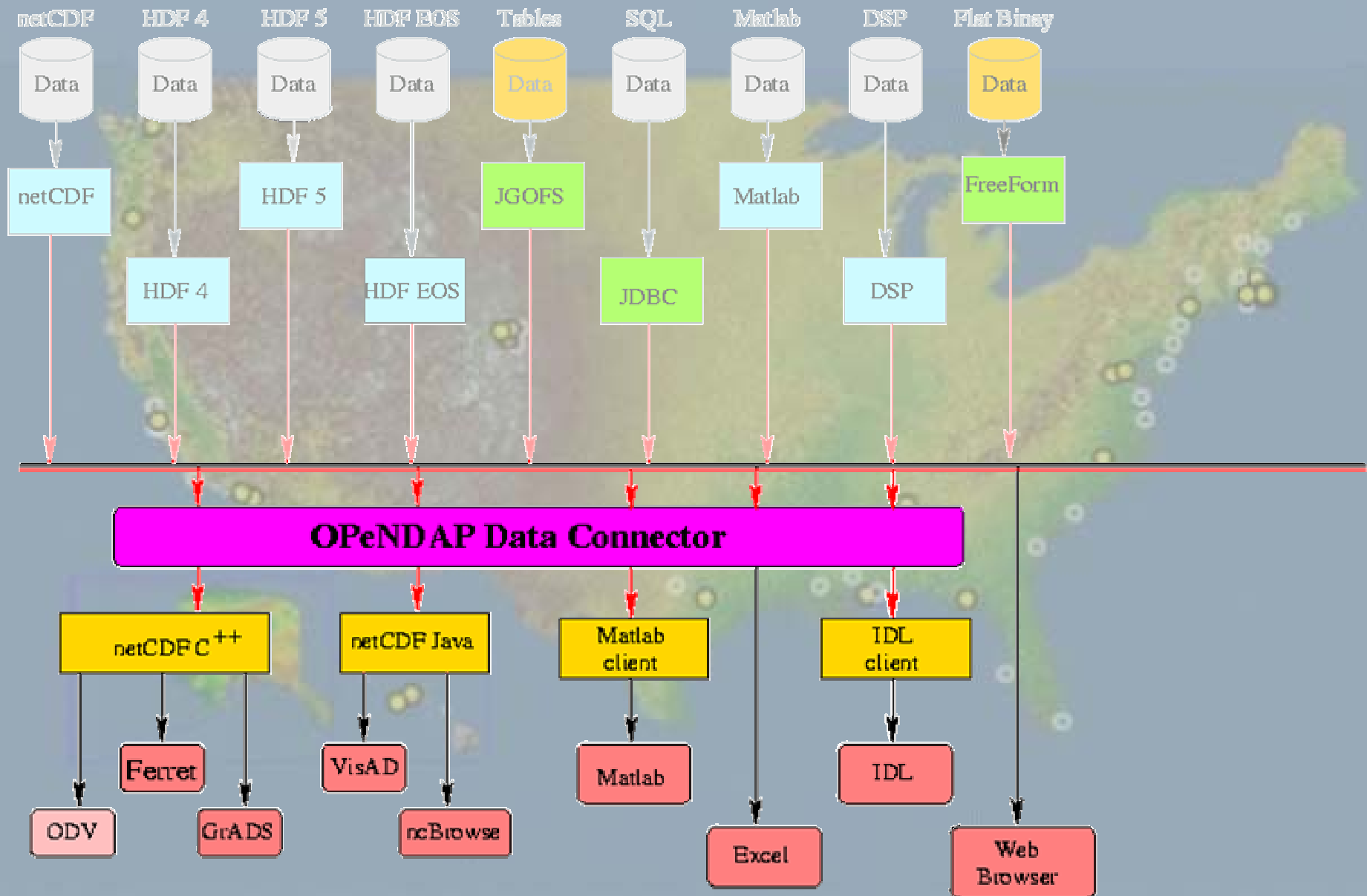
OPeNDAP/NVODS Server Sites



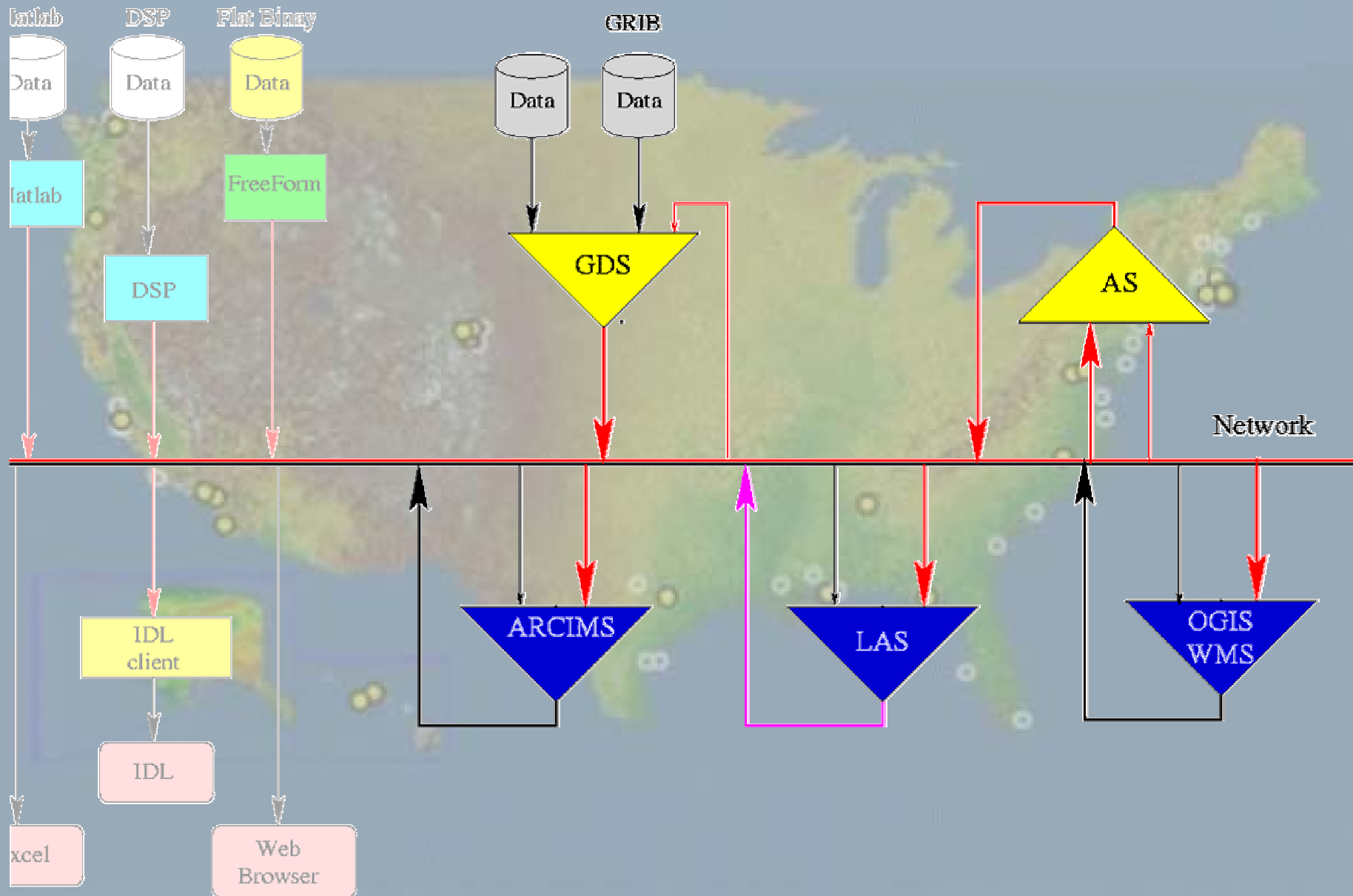
- Current DODS Sites
- Future NOPP-DODS Sites

[Back to
Main Menu](#)


OPeNDAP² Client and Server Status



Special Servers



Projects Using OPeNDAP²

- 
- **GODAE** (Global Ocean Data Assimilation Experiment)
 - **NOMADS** (NOAA Operational Model Archive and Distribution System)
 - **AOIMPS**
 - **ESG II - Earth System Grid II**
 - **Ocean. US** (US-GOOS)
 - **High Altitude Observatory Community**

Institutions Making Heavy use of OPeNDAP²

- 
- **Ingrid - Columbia University**
 - **COLA - Center for Ocean-Land-Atmosphere**
 - **Goddard DAAC**
 - **CDC - Climate Diagnostic Center**
 - **PMEL - Pacific Marine Environment Lab**

NVODS Monthly Accesses (2002)

Site	April	May	June	July	August
URI	4,856	19,504	3,691	26,693	7,440
LDEO	80,709	62,930	46,092	93,088	
CDC	102,518	153,362	62,395	181,974	107,512
JPL			73,516		
MBARI			3607		
TOTAL	188,083	235,796	189,301	301,755	114,952

NVODS Unique Users (2002)

Site	April	May	June	July	August
URI	73	68	72	44	69
LDEO					
CDC	124	105	91	116	111
JPL					
MBARI					

Interesting Statistics OPeNDAP2 Access

- **IRI data accesses for 1st quarter of 2002**

Type	Requests	%	Volume (gb)	%
DODS	191,611	8.5	375.2	69.4
Other	2,062,681	91.5	165.2	30.6
Total	2,254,292	100.0	540.4	100.0

- **PMEL OPeNDAP² ~ 35,000 with ~26,000 internal.**

Conclusions

➤ OPeNDAP

- **Is not a data system.**
- **Provides a base on which one can build a data system.**
- **The focus on syntactic data has lead to the rapid installation of OPeNDAP-enabled servers.**
- **Is discipline neutral**
- **Has been adopted in several disciplines as a data system core**