The Data System Integrator

Presented at the AGU/AMS Geosciences Department Heads and Chairs Meeting 15 October 2004

by

Peter Cornillon

Graduate School of Oceanography University of Rhode Island Narragansett, RI

and

OPeNDAP 165 Dean Knauss Drive, Narragansett, RI
Data systems generally involve some combination of the following elements:

- Archive
- Access/Delivery
- Discovery
- Analysis/Visualization
These elements have been developed and managed as monolithic systems by the groups assembling the elements.
Trend in Data System Development

- Away from centrally designed, implemented and maintained *systems*
toward

- The integration of independently designed, implemented and maintained *system elements*. 
With a Plethora of Unconnected System Elements
Defining a chaotic environment for those interested in accessing these resources
In an environment consisting of highly distributed system elements, the notion of a *data system* is amorphous at best.
The Data System Integrator

- Defines the data system by bringing order to a disordered array of system elements.
The Data System Integrator

- Is an element which assembles a suite of data system elements that together provide seamless access to the data from discovery to use.

- Generally ‘speaks’ several different protocols.
OPeNDAP
The Open source Project for a Network Data Access Protocol
The ODC - a Data System Integrator
The ODC - a Data System Integrator

- GFDL netCDF
- URI HDF
- GSFC Binary

GCMD

NVODS

Matlab
Ferret
VisAD
GrADS
IDV
Access
IDL
Excel
ncBrowse

OPeNDAP

ODC

GSFC

URI

HDF

GCMD
The Beauty of the Data System Integrator

Is that it can be built for specific applications:

- Oceanography
- Education
- Hydrology
- Discipline neutral - ODC

i.e., you can tailor your data system to your application without having to also manage all of the system elements; *most of the work has already been done.*
And there is nothing that precludes different data system integrators from sharing data system elements.
Location of the Integrator

The data system integrator may reside locally or remotely.
Conclusion

- The data system integrator is what will define the data system in the future.
- It offers an enormous amount of flexibility.
Metadata
The Ultimate Objective of a Data System

- To provide requested data to the user’s analysis/visualization package in a consistent, readily useable form.

For example: A user might want all ocean temperature values (with associated times and locations) that lie between 90 and 110 m and have uncertainties less than 1°C.
Interoperability

To achieve this objective, system elements must 
**interoperate**; i.e., the system must:

- Be capable of finding all data of interest.
- Know the format of these data objects.
- Be capable of transforming from these formats to that required by the application software.
- Understand the semantics of the data.
These interoperability requirements require in turn descriptions of the data; i.e., metadata.

The degree of system interoperability is determined by the associated metadata.
The required metadata falls in two classes:

- **Syntactic** metadata – Information about the data types and structures at the computer level - the syntax of the data; *e.g.*, variable T is a 20x40 element floating point array

- **Semantic** metadata – Information about the contents of the data set. *e.g.*, variable T is sea surface temperature with units of °C
Syntactic and Semantic Metadata

- **Syntactic** metadata provides the information needed to read and plot the data, but in general not to label the axes.
Semantic metadata provides the information needed to label the axes in a plot.
Interoperability and OPeNDAP

The two types of metadata suggest two levels of interoperability:

- **Syntactic** interoperability – Consistent format representation across data sets.

- **Semantic** interoperability – Consistent semantic interpretations of the data.

**OPeNDAP** mandates syntactic interoperability via a strict syntactic description of all data available via the system.
http://opendap.org

http://nvods.org

http://unidata.ucar.edu/packages/dods
And it is getting more
And it is getting more and more chaotic