MBARI’s SSDS
Data Management for Ocean Observatories

Brian Schlining
Overview

• About MBARI
• Overview of MOOS
• Functional Requirements of SSDS
• Metadata use by SSDS
• Dataflow through SSDS
Monterey Bay Aquarium Research Institute

Established in 1987
MOOS History

Monterey Ocean Observing System

• 1989: Began mooring-based observations.
• 1999: Adopted the term “MOOS” (Monterey Ocean Observing System)
• 1999: Began development of an integrated ocean observatory.
MOOS – The Present Concept

**Mooring**

**Autonomous Underwater Vehicle (AUV)**

**Benthic Node**

MBARI
MOOS – The Present Concept

Mooring

AUV

Benthic Node

Plug and Work’ Instruments
Future of Ocean Observing

Variety of platforms
Data Management

Challenges:

• Large number of data sources
• Large variety of data sources
• Dynamic system
  ➢ Data sources may appear and disappear
• No standard data format
  ➢ Data can be instrument ‘native’
SSDS

Shore-side Data System

• Data management system for MOOS
SSDS – Functional Requirements

1. Capture and store data from MOOS data sources
   ➢ Includes files and ‘streams’

2. Capture information (i.e. metadata) about the stored data
   ➢ Location, instrument, platform, data format, etc.

3. Capture and store data products
   ➢ Derived products, quality controlled data, plots, etc.
4. Provide access to the original ‘raw’ data

5. Convert data to common formats for user application tools (Excel, Matlab, ARCView, …)

6. Present simple plots of any well-described data.

7. Provide access to data through an application program interface (API) and a web interface.
Metadata — What we have learned

• Metadata must accompany the data
  – A system’s power relies on good knowledge of its data
  – Without metadata, the data quickly becomes unusable

• Metadata must accompany the instrument
  – Every connector between the two increases error rates
  – Once data and metadata detached, reattaching is painful

• Metadata must be flexible and structured
  – Flexible: you’ll need to define new kinds of data sources
  – Structured: consistency => automation => value
‘Deployment’ information.

SSDS tracks:
- Where the data was collected.
- When it was collected.
- What other data was used.
- Relation to other deployments
The data source.

SSDS tracks:
- Software or hardware source
- Unique identifier
- Manufacturer information
- References to documentation
References to the data.

SSDS tracks:
- The data storage location.
- The deployment that produced this data.
SSDS – Metadata (Object View)

Format and contents of a DataContainer.

SSDS tracks:
- The contents of a data-set.
- The data format (To allow parsing by software).
SSDS – Metadata Standards

SSDS does not currently use metadata standards such as ESML, ESRI’s Marine Data Model, FGDC, Dublin Core or Marine XML.

Why not?

• Multiple overlapping standards. Which to choose?
  – None have achieved broad community support
  – Some are not yet complete (Marine XML)
  – Excessive documentation requirements (FGDC)

• May use with SSDS later (Dublin Core)
Example SIAM to SSDS Data Flow
A **device** is connected to a **platform**, such as a Mooring.
The mooring retrieves the metadata from the device.
The metadata is packaged and sent to a portal on shore before any data is sent to shore.
Example SIAM to SSDS Data Flow

The portal forwards the metadata to SSDS.

<?xml ...>
<Metadata>
 ...
</Metadata>
Example SIAM to SSDS Data Flow

SSDS stores the metadata in a database.
This allows applications to query for and use data.
Example SIAM to SSDS Data Flow
Example SIAM to SSDS Data Flow

The device produces a data record.
The data is packaged and sent to SSDS.
SSDS uses information in the **packet** to sort and store the data in a ‘raw’ format.

SSDS

- VersionID,
- DeviceID,
- MetadataID,
- RecordType,
- PlatformID,
- SystemTime,
- SequenceNumber,
- DataBuffer(34,56.234,0.0023,…)
Example SIAM to SSDS Data Flow

The ‘raw’ data is parsed and stored as netCDF for easier access.

- VersionID,
- DeviceID,
- MetadataID,
- RecordType,
- PlatformID,
- SystemTime,
- SequenceNumber,
- DataBuffer(34,56.234,0.0023,...)
Software applications allow users to discover and obtain data in formats useful to the typical MBARI user. (netCDF, text, etc.)

```
netcdf parosci {
  dimensions:
    time = UNLIMITED ; // (17761 currently)
  variables:
    double time(time) ;
      time:long_name = "Time (GMT)" ;
      time:units = "seconds since 1970-01-01 00:00:00" ;
    double depth(time) ;
      depth:long_name = "depth" ;
      depth:units = "UNKNOWN" ;
  // global attributes:
    :title = "AUV data" ;
    :created = "2003-06-12T23:34:58Z" ;
    :history0 = ": Deployment information for parosci.log" ;
    :deploymentName = "2003.099.10" ;
    :instrumentId = "3699" ;
}
```
Software applications also provide simple visual representations of data.
Example SIAM to SSDS Data Flow

- Device
- Mooring
- Portal
- SSDS
- Web Pages
- MBARI Software
- DB
- Provide internet access
- Serialized
  - netCDF
  - ________
  - ________
  - ________
Example SIAM to SSDS Data Flow

Save development time by using existing software applications.
Development Team

• John Graybeal
• Kevin Gomes
• Michael McCann
• Brian Schlining (_______)
• Rich Schramm
• Daniel Wilkin
Science and Technical Advisors

- John Ryan
- Francisco Chavez
- Ken Johnson
- Drew Gashler
- Mark Chaffee
- Tom O’Reilly
- Mark Chaffee

Funding provided by the David and Lucile Packard Foundation