Metadata Automation:
Survey Results and Ideas

Shawn R. Smith
Center for Ocean-Atmospheric Prediction Studies
The Florida State University, Tallahassee, FL 32306-2840
smith@coaps.fsu.edu

Overview

- Motivation: SAMOS has interest in automated metadata transfers from ship-to-shore for underway data collection
- Purpose: To understand current metadata practices in use by UNOLS operators
- Survey:
  - 8 questions
  - Conducted via email and on MMI web site
  - 10 responses received
    - 7 direct answers to questions
    - 3 provided comments, some of which could be fit to questions
  - An excel table of the responses is available

Metadata Classification

- Vessel
  - Call sign, IMO number
  - Dimensions
  - Institution contact information
  - Photos, schematics
- Cruise
  - Cruise identifier, name
  - Chief scientist, personnel
  - Dates, ports
- Instrument
  - Sensor make, model
  - Sensor location
  - Calibration info, etc.
  - Parameter units, sampling rate, precision, etc.

- All metadata are temporal in nature
  - Must document time period for which they are valid
  - Instruments, people, and even ship dimensions change

(1) How are metadata stored for your vessel?

Results

- Hard copy - 4
- Flat files - 5
  - Includes ASCII, PDF, DOC
- HTML - 6
  - Includes all forms of web access
  - One group using a Wiki
  - Some restrict access
- Database - 2

Comments

- Wide variety of methods employed
- Hard copies of sensor calibrations seem common
  - Some felt that having an easy way to apply digital calibration info would be very helpful
- I assume some of the variety is resource driven

(2) What events result in an update? Procedure for updating metadata?

Results

- Sensor swaps, new calibration, repairs are most common indicators to update metadata
- Not clear that many groups maintain information on personnel changes

- Most metadata (digital or hardcopy) updated by ship technical staff
- Cruise metadata updated by PI before and after cruise, auto-update of database

Comments

- Generally the weak point in many systems
  - Changes need to be made by techs or science party
  - Documentation never a high priority or desired task
- Some newer systems have sensors auto-update their metadata when installed
- **Key issue:** How to ensure metadata are updated on sensor swap or other change (e.g., personnel swap).

(3a) Where are metadata stored?
(3b) How are transfers achieved?

Results (3a)

- A combination of shipboard and shore-side storage existed for all who responded (8)

Results (3b)

- Manual (hand carry) - 5
  - Includes DVD transfers
- Electronic - 3
  - Includes email, ftp, or scripts
  - Mostly done at home port, sometimes use satellite comm.
- None - 1

Comments

- Advances in ship-to-shore communications should allow for more electronic transfers
  - Satellite systems
  - Wireless internet when near shore or in port
- **Key Point:** Automated transfers require some formal electronic storage of information on shore
  - Not all institutions use databases or have dedicated digital resources on shore

(4) How are metadata disseminated to users?

Results

- DVD/CD - 5
  - Most are post cruise summaries or metadata files provided to Chief Scientist
- Online - 6
  - Web access does exist, but usually is only a sub-set of metadata
  - Some restrict access to information
  - Some institutions also maintain metadata on hard disk onshore, but these are not linked to web or ftp
- On request - 3

Comments

- Many expressed interest in universal on-line access

(5) Could routine electronic metadata reports be sent to interested users?

Results

- Yes - 3
- No - 2
- Maybe - 1

- Some were not clear as to why you would want routine reports
- Most consider post-cruise reports sufficient

Comments

- Important to differentiate information needed prior to and after cruise (user dependent)
  - More real-time use of underway data (e.g., SAMOS, salinity) require pre-cruise metadata
- Need to be able to send routine (all cruises) and one-off (single cruise) reports

(6) Preferred electronic format and transfer protocol

Results

- Responses were a bit hard to decipher. Different issues for collection and exchange of metadata
- Preferred format (exchange)
  - Text (ascii) - 3
  - XML - 4
- Acquisition method
  - Email - 3
  - Web based form - 2

- Clearly more details to work out here

Comments

- Users may desire different formats
  - Simple text summaries for general user
  - Web summaries for easy access
  - Database access for advanced users (automated processing)

(7) What technology is needed for metadata management by vessel technicians/operators?

Results

- Html interface - 7
  - Wiki - 1
- Other - 1
  - E.g., MS Access or freeware application

- Several commented that developing a community-wide GUI for the web would be helpful
- Suggested that GUI ties into a Central Metadata Repository
  - Access by ship, cruise, dates, ocean region

Comments

- Several groups are working on web based metadata apps.
  - SAMOS - FSU
    - Alpha version exists
    - Demo on request
  - WHOI
  - UCSD/UNOLS
    - Cruise metadata
    - http://data.unols.org/reports/cruise-level-metadata
  - MBARI
  - U. Hawai'i

Summary

- Current technology will allow automated metadata transfers
- Need a system-wide approach
  - Central repository would be great
  - Could be a distributed metadata archive
- Way forward:
  - Take advantage of groups currently working on on-line metadata management systems
  - Form a task group to develop prototype web/email collection system and repository
  - Build upon UNOLS efforts to automate cruise metadata collection
  - Ensure system will be accessible to large and small operators
  - THIS MUST BE A FUNDED TASK!!!
    - Must be a UNOLS priority.

Great Quote

"Keeping accurate track of metadata is a real pain, but it's worse when you don't have it."

- B. Appelgate, 2007