

A HYBRID APPROACH TO DATA ACQUISITION AND MANAGEMENT

*WEBB PINNER: OER, SHARON MESICK: NESDIS, BRENDAN RESER: NCDDC,
COLLEEN PETERS: SOI, LEIGHTON ROLLEY: SOI*

INTRODUCTION:

Data from shipboard sensors are collected in various ASCII, binary, open and proprietary formats. Acquiring all of these varieties using a monolithic data acquisition system (DAS) can be cumbersome and often complex. The NOAA Office on Ocean Exploration and Research (OER) has collaborated with the National Coastal Data Development Center (NCDDC) to develop a hybrid data acquisition and management system that standardizes data management functions from disparate systems to meet both short and long term data management objectives.

By taking a holistic approach to data acquisition and management the team has developed a system that is flexible, straightforward to manage, readily accessible by the cruise participants and can be adapted to support other vessels. Initially developed to support the NOAA Ship Okeanos Explorer, the approach has also been successfully deployed aboard the R/V Falkor operated by the Schmidt Ocean Institute (SOI).

PROBLEM:

The challenge is how to organize and manage all data types, regardless of format, DAS software and collection model. The solution must also provide cruise participants with safe and intuitive access to data while minimizing the effort needed by the operator to maintain the system.

VARIOUS DATA TYPES:

- Open ASCII (NMEA: csv)
- Proprietary ASCII (SBE CTD: HEX)
- Open Binary (Imagery: JPG)
- Proprietary Binary (EK60: RAW)
- Shipboard Products (Reports: pdf, docx)

VARIOUS DATA ACQUISITION SYSTEMS:

- Freely Available DAS (OMAO's SCS)
- Proprietary (Kongsberg's SIS)

VARIOUS COLLECTION MODELS:

- Always on (GPS, GYRO)
- Discrete Events (CTD Casts, ROV Dives)

SOLUTION DEVELOPED TO SUPPORT THE *OKEANOS EXPLORER*:

Using the *NOAA Ship Okeanos Explorer* as a test bed, the team developed a philosophy for how all data needed to be collected, managed and dissemination. The solution relies on equally on technology, documentation and training.

STANDARD OPERATING PROCEDURES:

Standard Operating Procedures (SOPs) define how to operate the equipment, how to configure the corresponding data acquisition systems, name and store files.

STANDARDIZED NAMING CONVENTIONS AND DIRECTORY STRUCTURES:

Filenames include a cruise identifier, start date/time, data type identifier and additional information that uniquely identifies the data set. Directory names identify the sensor type and/or collection date

CENTRALIZED STORAGE:

All collected data is copied to a centralized shipboard data warehouse. Cruise participants are granted read-only access to data stored on the shipboard data warehouse.

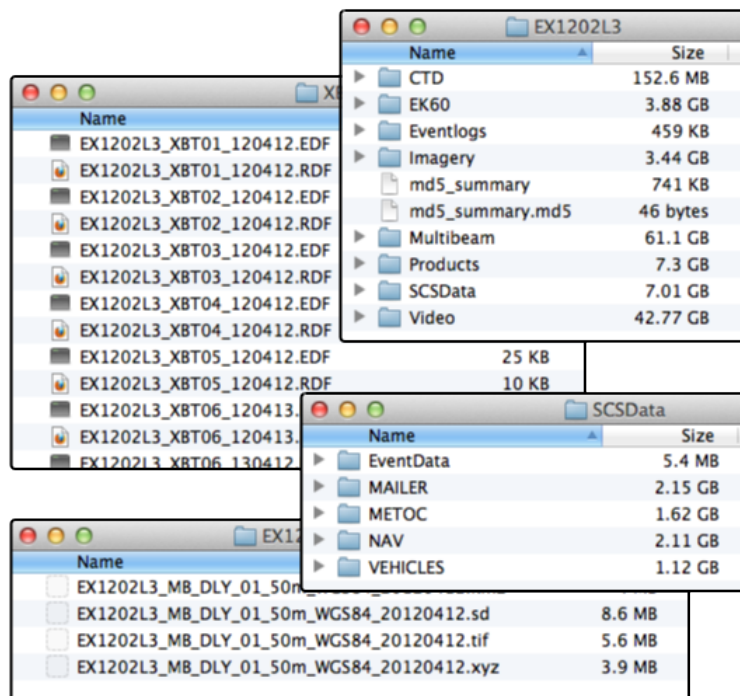


Figure 1 Examples of Standardize Directory Structure and Naming conventions.

TOOLS

Simple, freely available and cross-platform tools are employed to enforce the naming conventions, and directory structures specified in the SOPs.

AUTOMATED DATA RETRIEVAL

Scheduled tasks and scripts on each collection workstation push data from the local workstations to the data warehouse. Data is pushed rather than pulled to enforce network security policies protecting the ship's mission network from the public network.

The scripts installed on the data collection workstations (SCS Server, CTD Workstation, EK60 workstation, etc.) enforce the standardized naming conventions for that data set and ensure the data is uploaded to the appropriate directory. Files that do not meet the naming requirements are not uploaded to the data warehouse. Scripts and installation procedures exist for the three major platforms: Windows, Linux, Mac OS X.

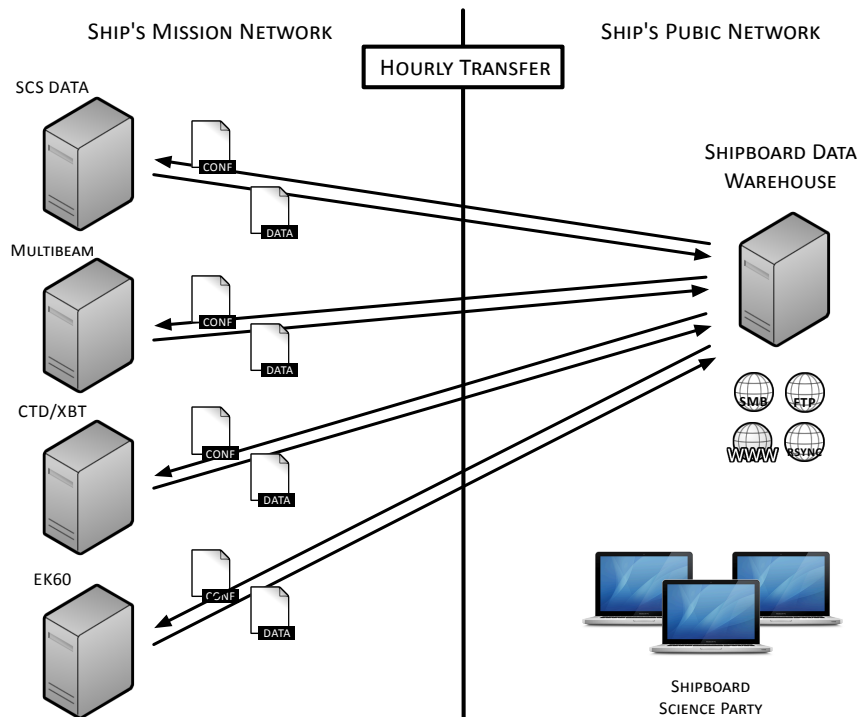


Figure 2 Backup scripts pulls configuration data from the data warehouse and use that information to push datasets to the data warehouse

WEB-BASED MANAGEMENT

The behaviors of each script (source and destination directories) are controlled via a centralized web-interface hosted on the data warehouse. The web-interface also includes a master switch that can disable all transfers. This is used to quickly disable the automated retrieval system while the vessel is in port.

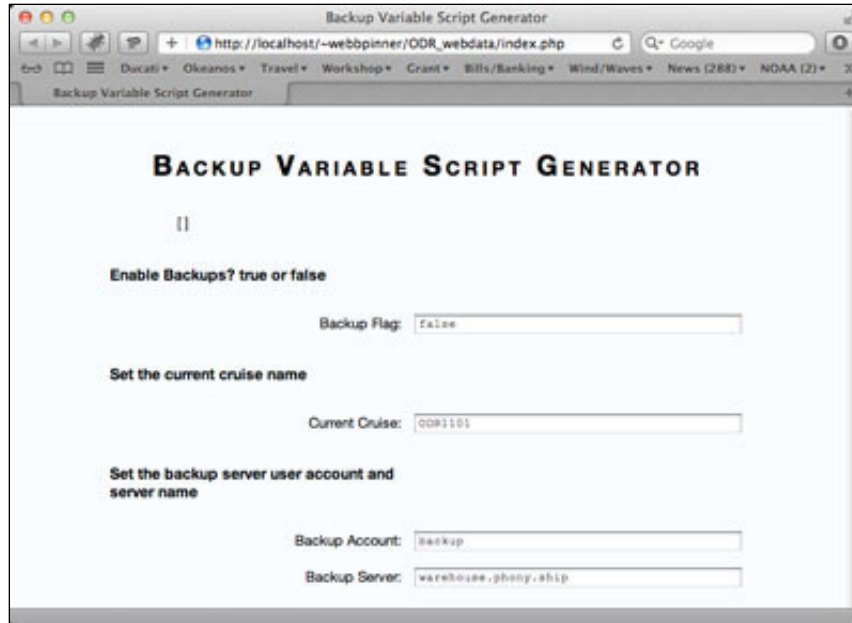


Figure 3 - Web-based control of the automated data retrieval system

HEALTH MONITORING

All file transfers produce log files that list all files that were successfully uploaded. Additional log files are generated showing all files in a source directory that did not meet naming requirements. Both log file types are translated into RSS feeds that are viewable using RSS readers or web-browsers. This enables cruise participants to quickly discover when and what files were recently uploaded to the data warehouse and informs technicians of any files that need to be addressed.

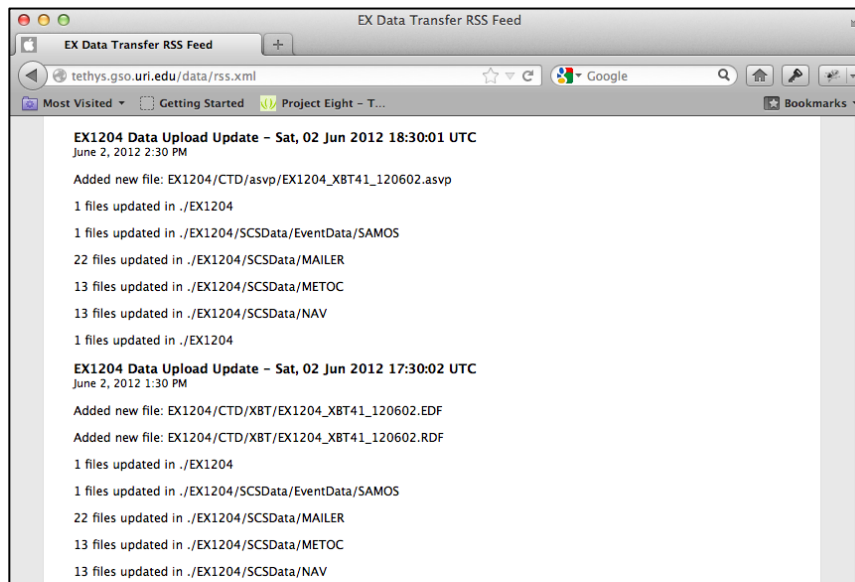


Figure 4 RSS Feed generated from the data file transfer logs

ON-SCREEN NOTIFICATION

On-Screen notification tools are installed at each collection workstation. These notification tools are used to gracefully and unobtrusively inform operators that an upload was successful or if there are errors that require further attention.

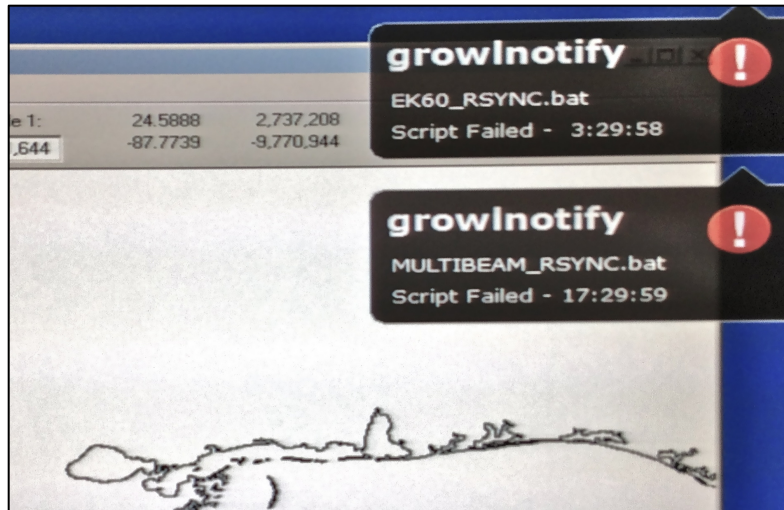


Figure 5 Real-time on-screen notifications showing problems with the transfer process

BENEFITS:

FLEXIBILITY

This approach is able to ingest data from new sensors including systems as complicated as Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs). It can also adapt to support new vessels as was successfully proven aboard the R/V Falkor.

ACCESSIBILITY

This approach provides cruise participants with safe access to data and with minimal effort from ship's technicians. Having a consolidated repository for all data collected by the vessel provides a straightforward means for technician to supply complete copies of cruise datasets to participants or for archive.

REPEATABILITY

This approach helps produce a consistent data product on every cruise independent of the cruise objectives and the cruise participants or technician.

EXTENSIBILITY

By standardizing the way data is collected and managed, additional tools can be developed that build on this first-level organization.

- Automated Metadata Generation of cruise datasets: Promotes standardized documentation for discover and access; rapid upload of dataset to National Archives. Promotes long-term preservation & stewardship, supports data discovery, direct public data access maximizing data reuse.
- The Okeanos Dashboard: An automatically populating dashboard-style web-portal providing an interactive interface to various collected datasets and a visually pleasing display of the data transfer RSS feeds.

SYSTEM REQUIREMENTS

HARDWARE

- 1 x86, Linux-based server
- Storage to accommodate all of the data collected during a single cruise, however it is recommended that there be enough storage to accommodate a field season.

SOFTWARE STACK

- Debian Linux – The underlying platform for the warehouse
- Apache Webserver – provides http access to collected data
- UNIX tools: bash, awk, sed, grep – The glue that holds everything together
- RSYNC – Transfer program used to move datasets to the warehouse
- CYGWIN – Unix tools for Windows
- Samba – provides access to collected datasets to Windows-based machines as Windows Shares
- ProFTPD – Provide FTP access to collected datasets
- Growl/Growl for Windows – Provides on-screen notifications
- OpenSSH – Provides security layer underneath all rsync file transfers
- NameMangler/BetterFileRenamer – Provides simplified batch file renaming on Windows-based and Mac-based systems.

CURRENT USERS

NOAA SHIP OKEANOS EXPLORER

- Prototype vessel used for the development and testing of the hybrid approach to data acquisition and management including its underlying technologies.
- Since 2010 the vessel has been refining this approach and performing hourly uploads of collected dataset to a shore-side server while underway. This includes cruises that do not include telepresence-related objectives.
- This approach has played a large part in the success of the vessel's telepresence-enabled cruises. It provides shore-based cruise participants with improved situational awareness through near-real-time access to data and data products.

SCHMIDT OCEAN INSTITUTE'S R/V FALKOR

- Newest vessel to implement the hybrid approach to data acquisition and management and its underlying technologies.
- The system was deployed aboard the vessel from start-to-finish with no prior preparations during a 10-day cruise. Work was completed using the ship's marine technicians, electronics technician and one additional contractor.
- SOI only uses the approach for management of data shipboard and is not currently pushing data to a shore-based server during cruises.