

INTRODUCTION

Data from shipboard oceanographic sensors are collected in various ASCII, binary, open and proprietary formats. Acquiring all of these flavors using a monolithic data acquisition system (DAS) can be cumbersome and often complex. The NOAA Office on Ocean Exploration and Research (OER) and a team from the National Coastal Data Development Center (NCDDC) developed a hybrid data acquisition and management system that standardizes data management functions for disparate systems that meets 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 that can be adapted to support other vessels. In December 2012 this approach was successfully applied to 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 straight-forward 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 (Konsberg's SIS)

VARIOUS COLLECTION MODELS:

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

SOLUTION DEVELOPED ABOARD THE NOAA SHIP 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 disseminated. The solution relies on equally on technology, and documentation and training.

STANDARD OPERATING PROCEDURES:

Standard Operating Procedures (SOPs) define how to operate equipment, configure the corresponding DAS, name and store data 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 AND SHIPBOARD ACCESS:

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.

TOOLS

Simple, freely available and cross-platform tools are employed to enforce the naming conventions, and the 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 data is uploaded to the correct directory on the warehouse. Files that do not meet the naming requirements are not uploaded to the warehouse. Scripts and installation procedures exist for Windows, Linux, and Mac OS X platforms.

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 new files are available on the data warehouse and informs technicians of any files that need to be addressed.

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 easily adapt to support new vessels as was proven aboard the R/V Falkor

ACCESSIBILITY:

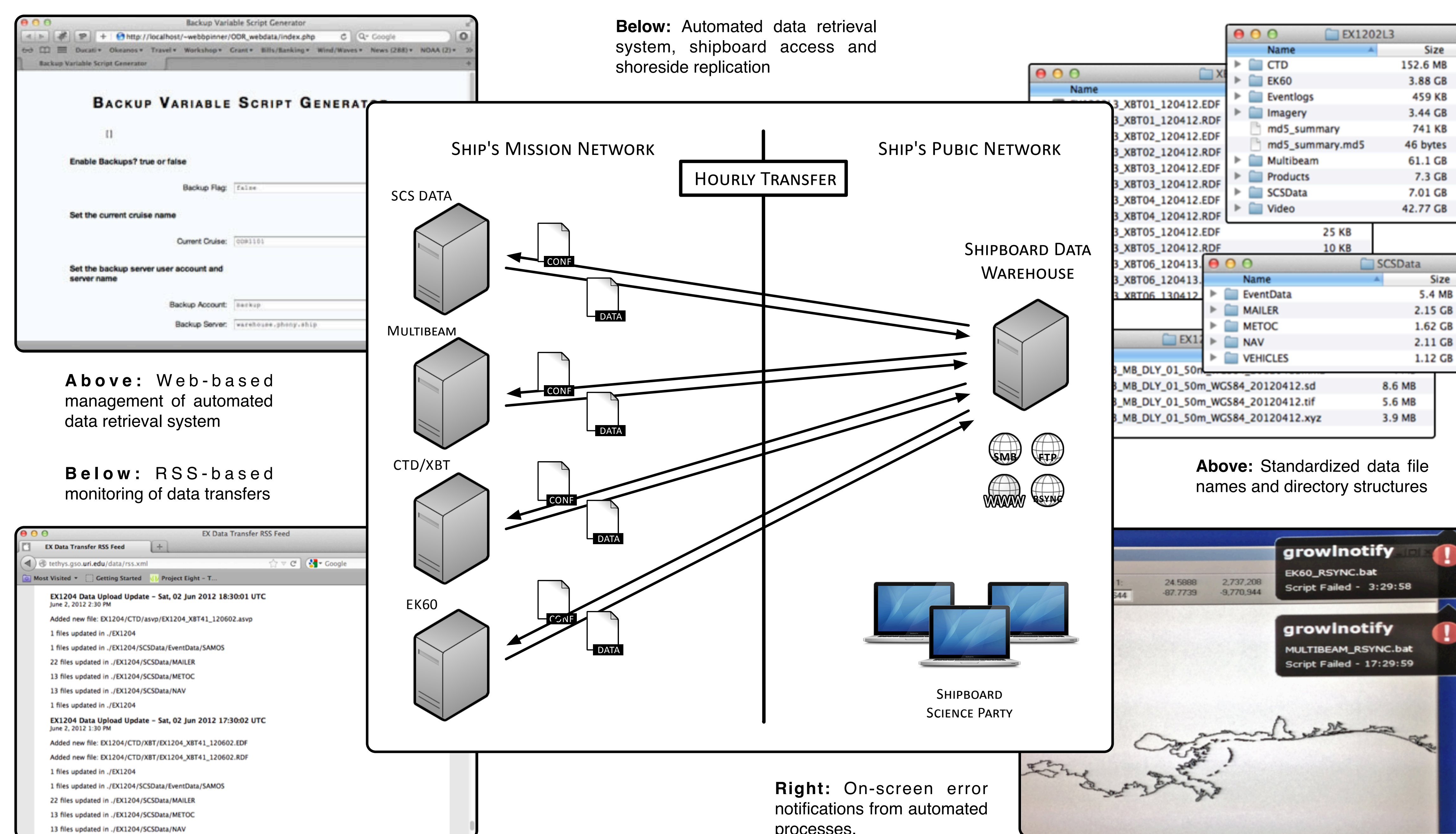
This approach provides cruise participants with safe access to data and with minimal effort from ship's technician. Having a consolidated repository for all data collected by the vessel provides a straightforward means for ship's technicians to providing complete copies of cruise datasets to participants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

REPEATABILITY:

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

SOFTWARE PACKAGES USED:

- Linux (Debian 5/6)
- RSYNC
- CYGWIN
- UNIX tools: bash, awk, sed, grep
- Apache Webserver
- RSS
- Samba
- ProFTPd
- Grawl/Grawl for Windows
- OpenSSH
- NameMangler/Better File Renamer



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.

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.

EXTENSIBILITY:

By standardizing the way data is collection 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.