Data Management

- **Data acquisition rates are skyrocketing**
  Contributors:
- High definition video and stills
- Inclusion of *Sentry* into NDSF
- Imminent replacement of *Jason* DVDs
- Reson MB on all vehicles –
  midwater research/plume hunting

Data Type	2009	2010	2011	2012	2013
Navigation (GB)	222	276	134	260	390
Bathymetry (GB)	132	2,800	2,800	5,600	5,700
Imaging (GB)	330	24,900	47,000	105,000	149,000

Data Management

A server to house these data long-term has not been identified but this issue is being actively addressed in advance of 2012 operations

- WHOI DLA does not have sufficient on-line storage
- NDSF engineering server is currently acting as a temporary repository, but it's a short term solution
- Current/budgeted equipment sufficient approximately through 2011
- Legacy media (film, digital tape) = 1,000 TB

<2008	All cruise data -> physical media -> WHOI DLA
2008-present	All cruise data on physical media -> WHOI DLA
	Non-video data -> NDSF engineering server
	Video on physical media (DVCAM, DVD) -> WHOI DLA
2010-2011	*Sentry,* HD video -> media neutral -> NDSF repo
2012	All data + video -> media neutral -> Unidentified repository

AB1

NDSF Operator's Report

Data Management

Impact of NSF's data management thrust on NDSF Archive

May '11 OCE Sample and Data Policy updates '04 policy

No changes mandated by policy to NDSF treatment of data

- PIs are responsible for archiving and disseminating data
  - Water column data must be archived to NODC
  - Geophysical data must be archived to NGDC
- Protections of up to two years

Slide 14

AB1 may have an edit on this. had asked scotty for input
Andy Bowen, 6/15/2011

Current NDSF/WHOI DLA Treatment

- WHOI DLA houses a copy of water column & geophysical data
- NDSF also houses a copy of these data (since 2008) on eng server
- WHOI DLA houses physical video/photo media
- WHOI DLA & NDSF cooperate to service community requests

WHOI DLA archives ship data packages to NODC

- *Alvin* data is part of the *Atlantis* package and is archived to NODC
- No archival of ship or *Alvin* data to NGDC

No national center listed in the policy for video and photos; network bandwidth is inadequate to copy the quantities of video data away from WHOI

Data Management

Option:

WHOI DLA & NDSF expand the ship data forwarding mechanism to archive *Alvin, Jason, and Sentry* cruise data packages to the national data centers, incorporating lessons learned from the Rolling Deck to Repository (R2R) program, with which the WHOI DLA is already connected

Data Management

Standard Protocols for Multibeam & Photomosaic Data Acquisition

- *Jason* data deliverables: 3 online documents
  - Standard data products white paper
  - Pre-cruise brief, describes on-line docs
  - Template for post-cruise summary

- System-specific cookbook/checklists
  - *Jason* Reson 7125 multibeam checklist/cookbook
  - *Jason* photo survey checklist/cookbook
  - *Alvin* multibeam primer/cooker

Renavigation
Overview

- Develop common software and procedures for post-processing navigation data
- Automated fusing with standard science sensor data
- Develop NDSF wide protocols for navigation data quality control and post-processing
- Preserve metadata on the source of the navigation data and parameters used in post-processing

Background – Hardware

- All vehicles now use common navigation instruments

- Subsea: DVLs, fiber-optic gyros standard on all vehicles

- External: USBL and LBL
- USBL is standard on all vehicles
- LBL is still frequently used by *Sentry*. Capability is retained by *Jason* and *Alvin*.

Background – Software

- *Jason* and *Alvin* continue to use the DVLNav software developed by Whitcomb and Kinsey in 2001
- Since 2008, *Sentry* and *Nereus* employ NavEst, a new navigation package developed at WHOI
- NavEst merges DVLNav and the ABE navigation code developed by Yoerger
- *Alvin* and *Jason* are in the process of migrating to NavEst resulting in common real time navigation software for all vehicles
- New real time Nav GUIs are being developed by Howland

Present State

- Navigation post-processing has traditionally lagged behind other areas, with each vehicle taking its own approach

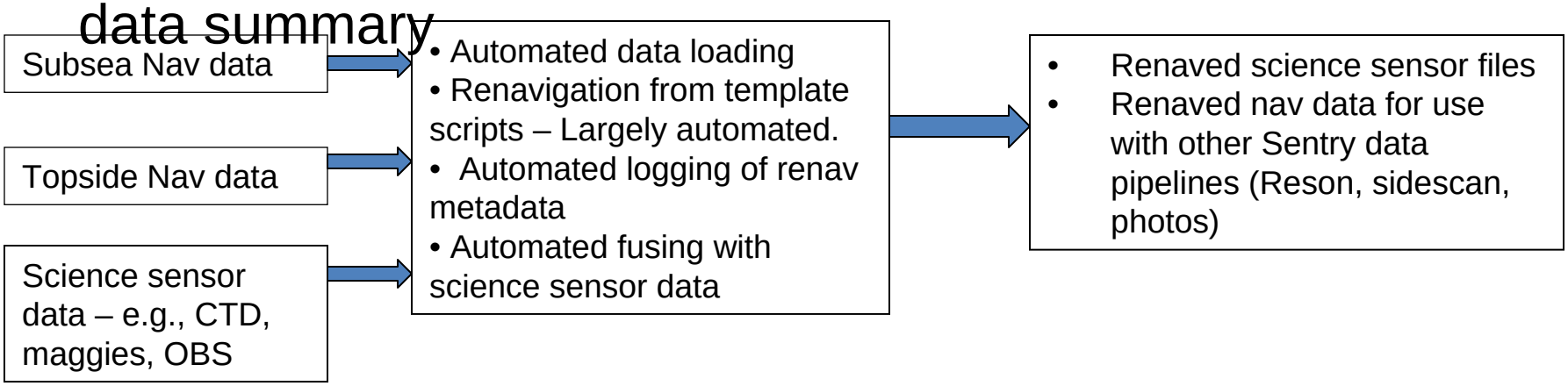
- For example, *Jason* and *Alvin* employ one set of tools while *Sentry* employs another

- *Sentry* tools have been rapidly evolving:
  - Motivated by rapid increase in data capacity and shorter vehicle turnarounds
  - Need to co-register nav data with other science products and feed into other data pipelines (e.g., Reson data, photos)
  - Archives metadata of navigation post-processing
  - Enables easy renavigation and fusing with data at a later date

Renavigation

Existing *Sentry* Pipeline

- Topside and subsea fused in Matlab
- Renavigation scripts are generated based on templates and the user edits a limited number of dive-specific parameters
- The resulting renav data is saved along with all of the metadata and the renav script
- Code is mostly automated although some human interaction is always required
- Automated scripts provide data for the cruise report and data summary



Renavigation
Goals

- Migrate NDSF navigation processing to the *Sentry* model
- Automate and simplify the process as much as possible
- Improve LBL and USBL post-processing tools
- Automate fusing of post-processed navigation data and standard science sensor data
- Ensure the source of the post-processed nav data is kept with the renavigated data
- Improved documentation and automated generation of cruise data

Ongoing Work

- Incorporate *Jason* and *Alvin* into this data pipeline
- Develop new protocols for at-sea navigators to quality control and post-process navigation data
- Develop new techniques for renavigating dives where the vehicle does different tasks – e.g., vent sampling, bathymetric map survey, etc.
- Timeline:
  - *Sentry* – in use; continued improvements
  - *Jason* – work in progress; shore-side development; at-sea testing as opportunities are available
  - *Alvin* – ready at the end of the overhaul

Imaging Revenue Report

Total billed **$33,242.50**
Received (to date) **$18,660.00**
(June 2010 – June 2011)

Total NDSF requests = 72
Commercial use requests = 49
Educational use requests = 23
(Total image requests to WHOI (NDSF and non-NDSF = 120)
Total still images licensed = 69
Total video licensed = 1180 seconds (19.6 minutes)

Textbook Requests for NDSF images
N = 36

Museums requesting NDSF stills/video
Museum Victoria, Mystic Aquarium, Newport Harbour Nautical Museum, UK Natural History Museum

Networks airing programs using NDSF footage
Animal Planet, BBC, Discovery Channel, History Channel, National Geographic, NBC, NHK (Japan), PBS – Miami, PBS – NOVA, ZDF/Germany

R2R Update Report – NDSF/R2R Synergies

- Time is right to review & renew NDSF's data repository and data archive responsibilities, services and capabilities

R2R Update Report – NDSF/R2R Synergies

- Time is right to review & renew NDSF's data repository and data archive responsibilities, services and capabilities

- Key principle: Data management responsibilities should reside with people most familiar (closest) to the data

R2R Update Report – NDSF/R2R Synergies

- Time is right to review & renew NDSF's data repository and data archive responsibilities, services and capabilities

- Key principle: Data management responsibilities should reside with people most familiar (closest) to the data

- R2R experience is of high value and can help NDSF to:
  - Create data quality reports for routine data
  - Provide NDC data submission services for scientists
  - Become "authoritative" for NDSF vehicle deployment identifiers (and associated metadata)
  - Identify controlled vocabularies for metadata and data
  - Coordinate NDSF & R2R DB schema development
  - More …