NDSF Data Management Update

Repair and upgrade of engineering/archiving server

- Multiple system failures subsequent to Feb blizzard "Nemo".
- RAID space added using refurbished equipment
- Total space now 85 TB or RAID6 storage, mirrored
- Funding will support the addition of another 40TB of capacity in 2013

Switch to Long Term FileSystem (LTFS) on Linear Tape Open (LTO) media

- NDSF archive includes a modest (undersized) LTO tape robot
- Switched to direct archival within LTFS from classic backup method (2102)
- Cruise RB13-03, Ross/Brooke
  - full time HD recording, incurring costs that exceed standad deliverables
  - extra costs covered by NOAA, archival directly to LTO (2 copies)

Installation of offload and video processing racks aboard Atlantis

- Two racks:
    - SSSG rack in comp lab
    - science video post-processing in main lab
- High performance offload system copies and mirrors original video, < 1hour
- Storage for ~20 dives
- Overnight generation of proxy video products
- Currently assessing digital asset mgmt sw for presentation and protection functions

NDSF Data Management Update

DVD replacement search - *Jason* constant video recording

- DVD technology continues to be available, but just barely
- Low impact, high quality replacement: .mp4 or .mkv files containing video compressed with h.264 codec.
- Systems from three manufacturers have been assessed
  - two complete recorders
  - encoder component

  Result: Neither recorder addresses all needs, and are prone to failure. Cost is 150x a DVD recorder.

Improved in-ship display of primary *Jason* control van video

- Two large monitors (46")
- Multiviewer places several sources on one monitor simultaneously

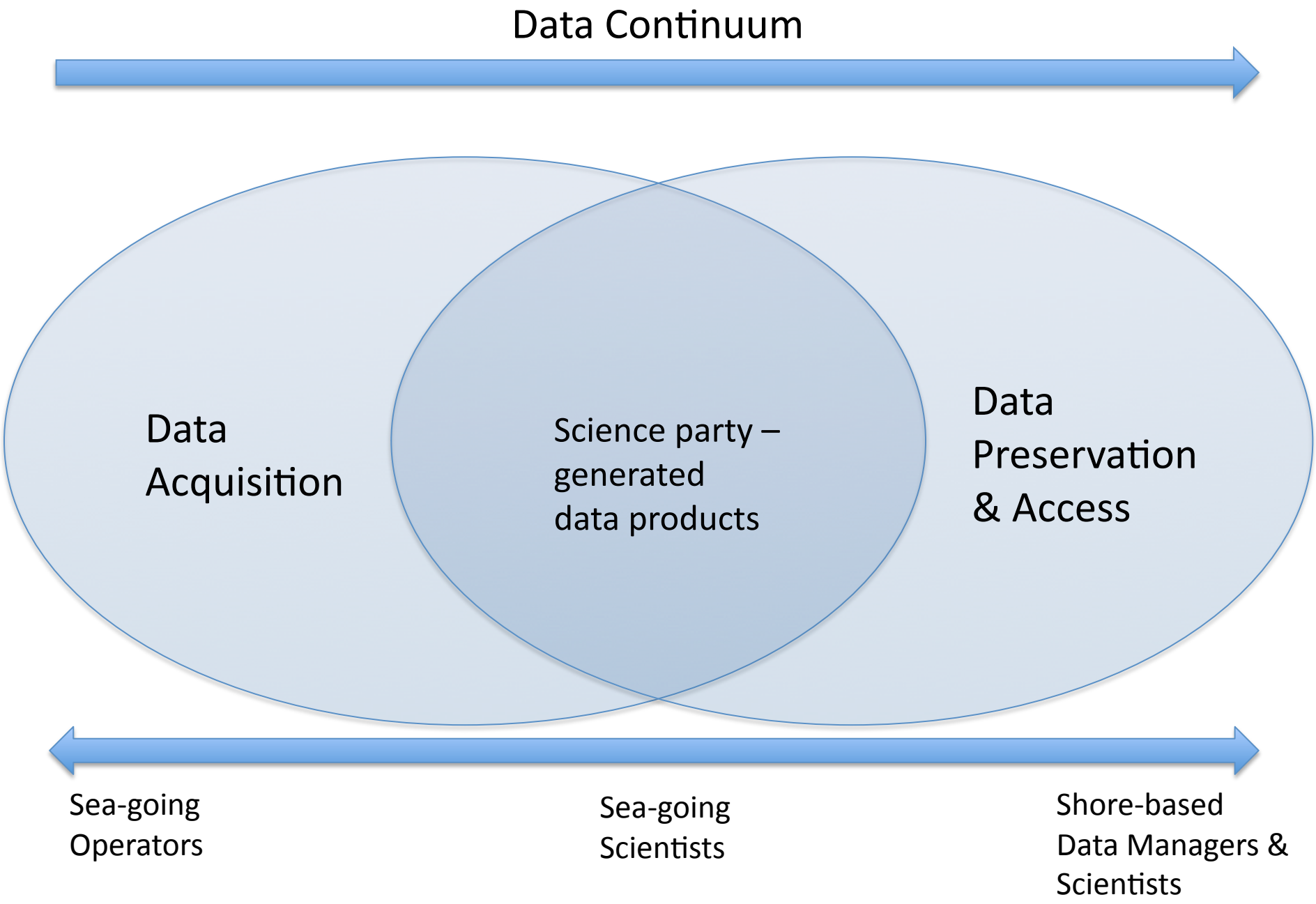
Duplicate VirtualVan/Framegrabber server for rollover on primary failure

Goals

- Define “data management”
- Strategy
- Identify places where DESSC can help
- Possible near-term vs. long-term solutions
- Identify potential partnerships/synergies

Importance of Data Management

- Support science and discovery
- Scientific reproducibility
- Optimizing operations
- Increasing volumes of data
- Data Policies with increasing focus on data sharing


Data Continuum
Data Acquisition
Science party – generated data products
Data Preservation & Access
Sea-going Operators
Sea-going Scientists
Shore-based Data Managers & Scientists

Data Acquisition

- Documentation
  - *Capture metadata contemporaneously*
  - *Data Quality*
  - *Sample metadata*
- User Interfaces (e.g. Fledermaus)
- Raw Underway Sensor Data
  - *Final nav embedded?*
- Time-stamped Video & Stills
- In-field Processed Sensor Data
  - *Navigation*
  - *Bathymetry grids**
- Guidelines
  - *Survey planning*
  - *Sample metadata capture*
- Cross-platform assessment & consistency

Post-cruise preservation/access

- Basic Dive Metadata
- Sample Metadata
- Raw underway sensor data (sonar, CTD, nav, eH, etc)
- Photo/Video
  - Frame Grabber/Virtual Van
  - Other?
- Operational data products
  - Bathymetry Grids
  - Processed Navigation

Science-party generated products

- Final geophysical data products
- Mosaics
- Results of Lab analyses

Archiving & Preservation

- NSF Data Policy
- Storage media
- Redundancy
- Metadata
- Data Discovery & Access

Opportunities

- Strategic relationships with complementary data efforts ?
  - *IEDA*
    - *MGDS*
    - *SESAR*
    - *EarthChem*
  - *BCO-DMO*
  - *R2R*
  - *MAC*
  - *NOAA*
  - *OET*
  - *Others?*
- EarthCube ?