WHOI Ocean Informatics Project

- Four person team (Andrew Maffei, Art Gaylord, Peter Fox, Jen Schopf)
- Working with scientists to identify important institution-level data and information technology needs
- Team members also:
    - help with informatics-related science proposals
    - support introduction of new technologies, approaches, resources for data/informatics related ocean science projects
- Close collaboration with RPI “Tetherless World Constellation”
    - Peter Fox, Chair
    - data-interoperability
    - data-intensive applications
    - semantic web technologies
- Helping on the NDSF data archive image conversion proposal to NSF/STIC is one of our important efforts

NDSF Image Conversion Proposal

- Problem Statement
  - much of the image data is at-risk as the recording media ages
  - images data is not readily accessible to scientists
    - lack remote access
    - fragility of media limits viewing opportunities
    - descriptive data, nav data is separated from image data
- Pilot program funding sought through a proposal to NSF STCI
  - initial focus upon image/video data from EPR 9-10° N datase AM1
  - kickoff workshop to gather expert input and refine methodology and technologies for pilot.
  - periodic reviews to measure progress and resolve issues
  - pilot scope is intended to produce a useful, complete archive suitable to address a significant science question
  - pilot will provide the model for further digitization of existing holdings and processing of new image and video data

Slide 3

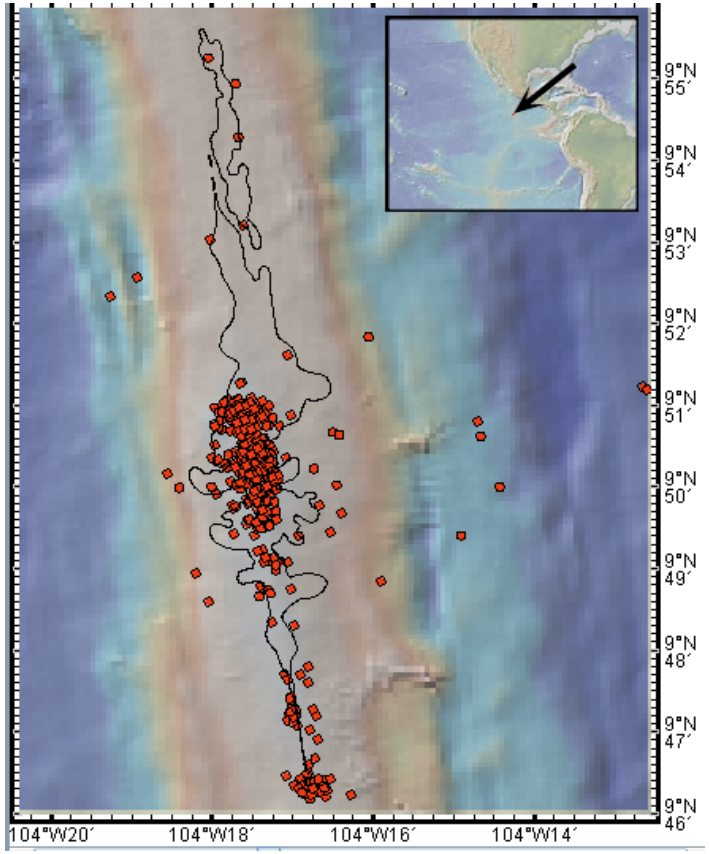
AM1 suggest that we say "Initial focus" here
Andrew Maffei, 6/16/2009

Project Plan

- Initial Assessment of media and nav data
- Gather expert advice and DESSC guidance
    - science needs
    - media preservation experts
    - informatics and data storage experts
- Experimentation
    - technical approaches to media conversion
    - user access testing and interoperability
- Bulk conversion of prototype dataset (EPR)
- Use of the dataset as demonstration
- Pilot project assessment
- success in meeting scientific and preservation goals
    - readiness to move to a proposal to NSF for full digital archiving of image/video holdings

Why the EPR dataset?

- High scientific interest in this dataset
- Area of extensive past and current research activities (~600 Alvin dives)
- It involves many media types
- It is well defined in scope
- Serves as a good model for future work



Prioritizing additional digitization

- Possible criteria for selection
  - historical importance
  - scientific value
  - nav data existence
  - nav data quality and resolution
  - image quality
  - fragility of media
- Guidance is needed from DESSC both with respect to the extent and the order in which collections are preserved