Data Management Update

Vicki Ferrini
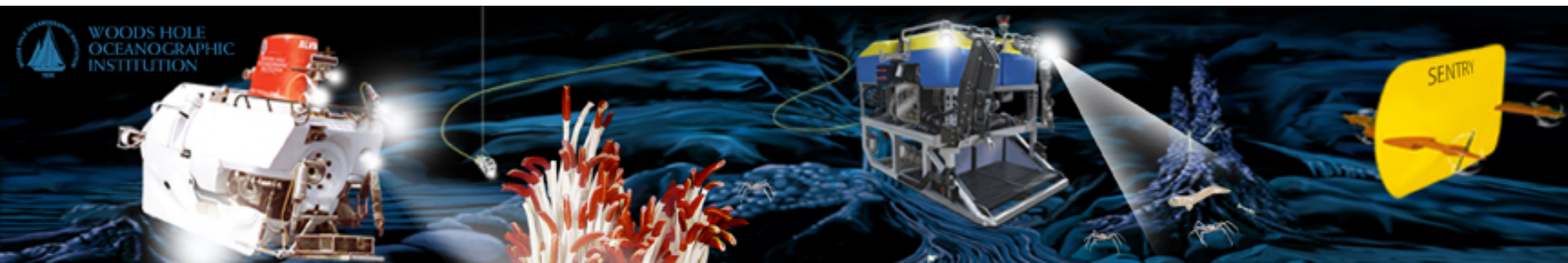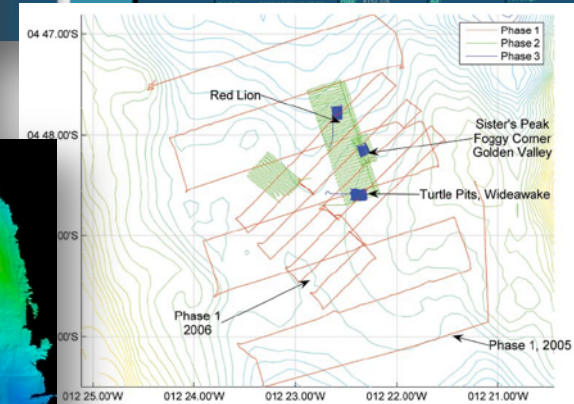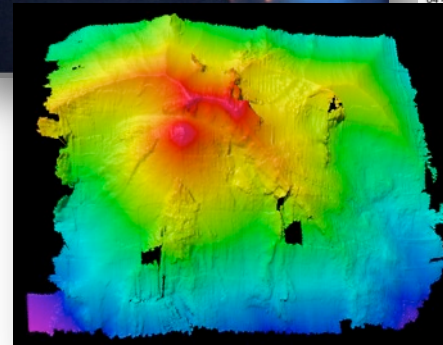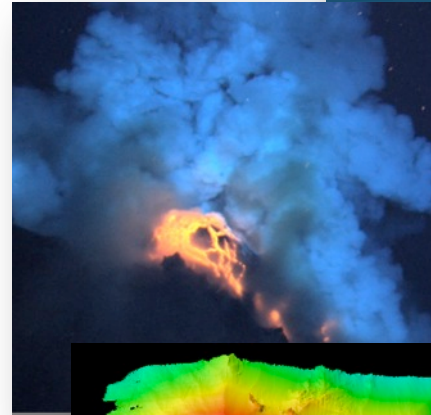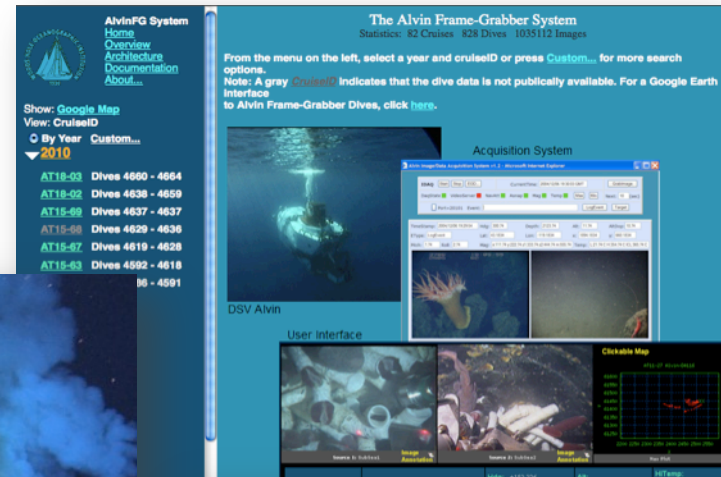
IEDA Data Facility

Lamont-Doherty Earth Observatory

NDSF-produced Data

- Images & Video

- Raw data streams
  - Navigation
  - Sonar, CTD, subbottom

- First-order products
  - Dive Metadata
  - Grids & Images
  - Processed Navigation
  - Event Logs

*+ Science Party Instrumentation & Final Products (PI resp.)*

IEDA: MGDS

- Part of IEDA Data Facility
- Metadata catalog with access to data files
  - Primarily marine geophysical & complementary data types
- NSF OCE data policy-recommended
- Existing Data Collections include NDSF Data:
  - Ridge 2000 & MARGINS Data Portals
  - NDSF Data Portal (deployed 2008)
  - Map-based access to FrameGrabber & Virtual Van Images
  - Data Rescue – Digitized Alvin slides & provide map-based access
- Data Curation – preservation, online search & access
- Data DOIs & direct links with publications
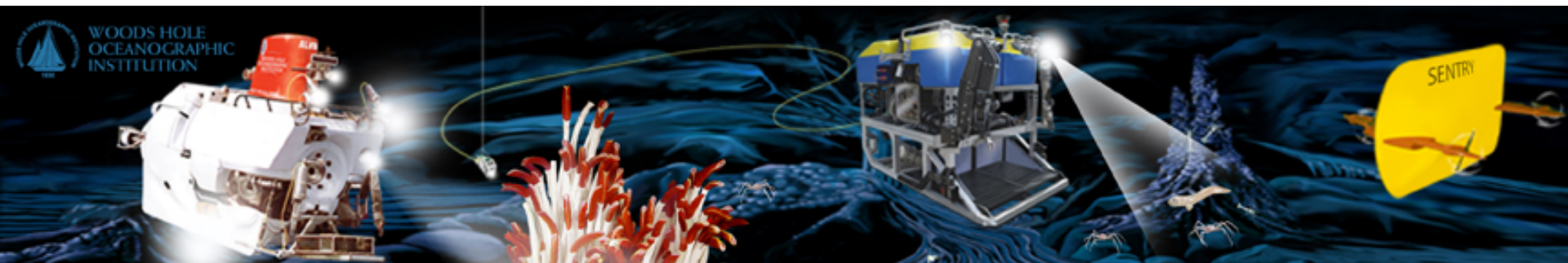
Submersible data in MGDS*

- Metadata from ~ 1,200 Dives
- > 1 Million downloadable data files, 1.7 TB
  - Backscatter, Bathymetry, Magnetics, Nav, Photos, Photomosaics, Sidescan
- 12 Platforms (AUV, HOV, ROV)
- Variety of Contributors
  - WHOI/NDSF, MBARI, Schmidt Ocean Institute, Waitt Institute & Individual Scientists

as of 09/08/2014
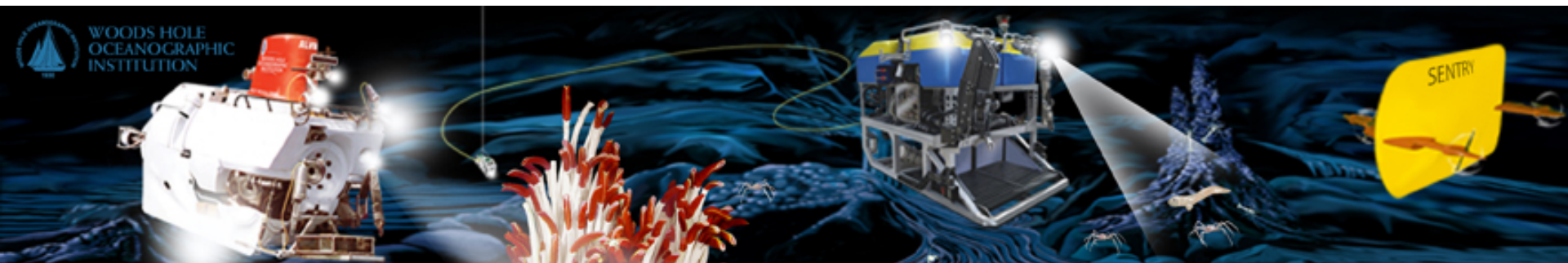
NDSF-MGDS Data Transfer Pilot (2013)

- Leverage IEDA/MGDS infrastructure tools and services
  - Search, Visualize, Download + Web Services
  - Data Provenance, Attribution, Data DOIs, Links to Publications
  - Web Services
- Select data types from select cruises
  - Underway and first-order (at-sea) data products
    - Metadata, Bathymetry, Navigation, Sidescan, Subbottom
    - Raw sensor formats, ascii, grids, images…
  - PI approval requested prior to each transfer
 - Data transfer via hard drive
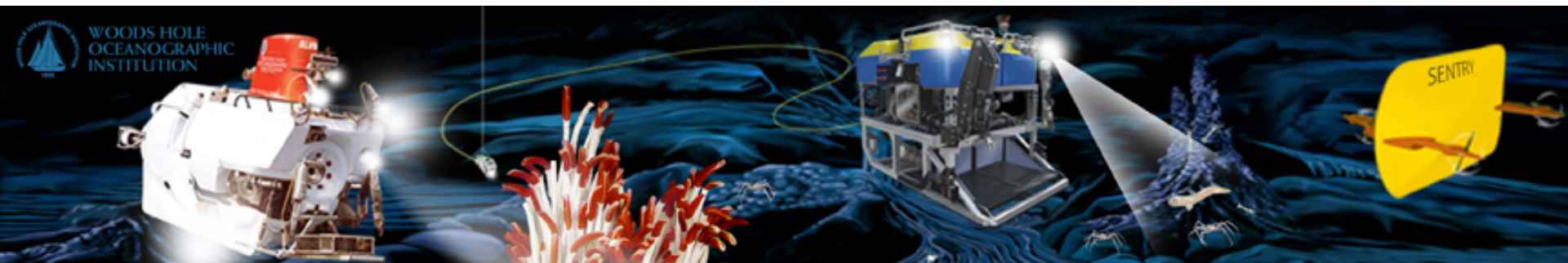
Review of Data Transfer Pilot (2014)

- Pilot data transfer to MGDS was successful

- DeSSC recommends routine transfer of data to MGDS

- New pre-cruise data release form developed
  - authorize the transfer of NDSF data to MGDS
  - uphold PI data embargoes

- Some challenges & short-term development priorities were identified
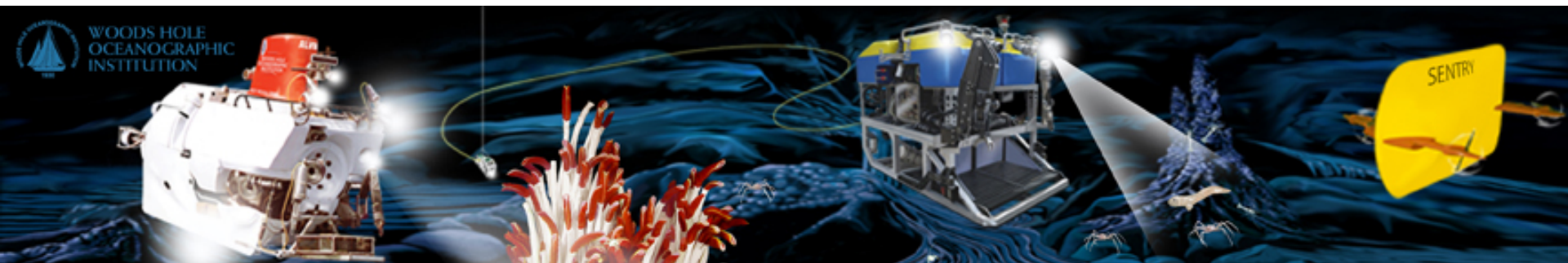
Review of Data Transfer Pilot (2014)

- Technical challenges to be addressed by NDSF
  - Standardize data deliverables facility-wide (metadata, naming conventions, formats, versions, directory structure)

- Additional features to be developed by MGDS
  - Enhance Web Services & data access services
  - Data Download Reports by vehicle
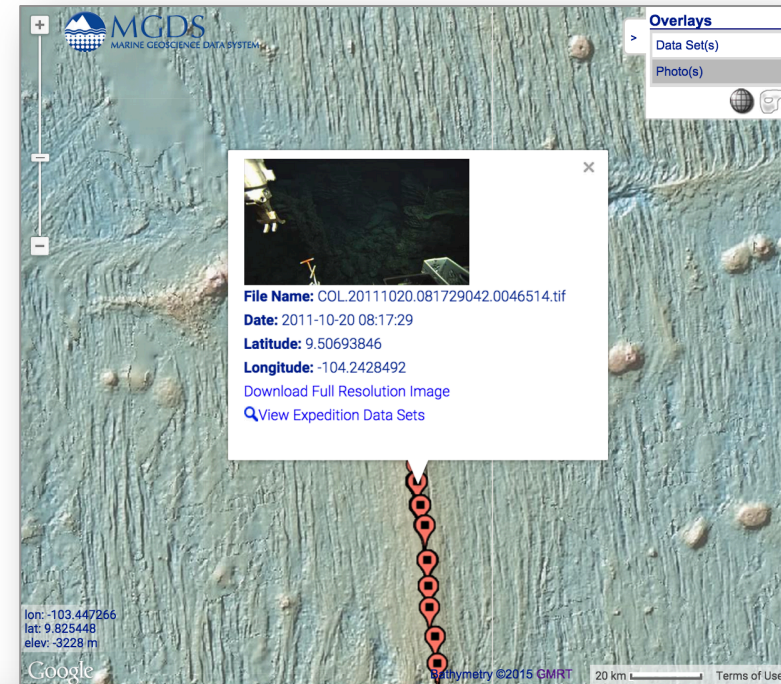 - Ongoing development of access tools & interfaces
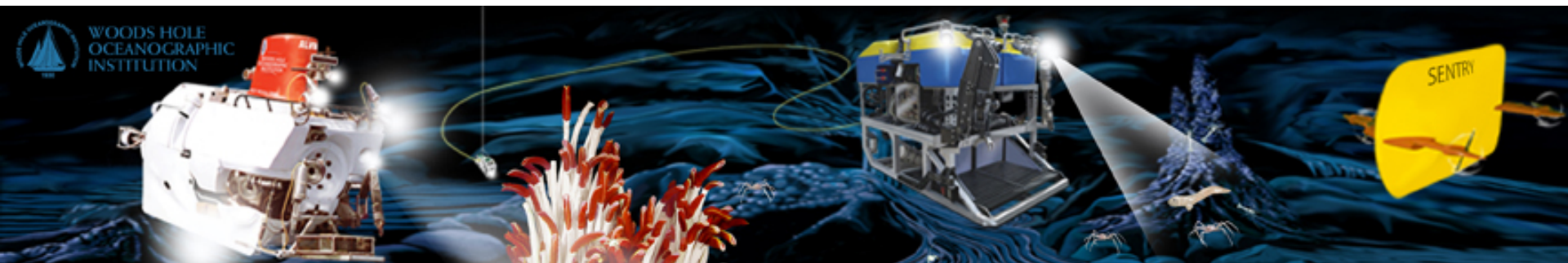
2015 NDSF Data Accomplishments

- Developed & deployed middle-ware to apply Sentry processing pipeline to Alvin data
  - Processed navigation, consistent formats, etc
- NDSF Facility-Wide Data Convergence Project
  - Standardize processing & data delivery architecture
  - Development of at-sea network accessible portal
    - Data Products
    - Metadata & Status Tracking
  - Optimizes at-sea data management & streamlines shore-side management and transfer
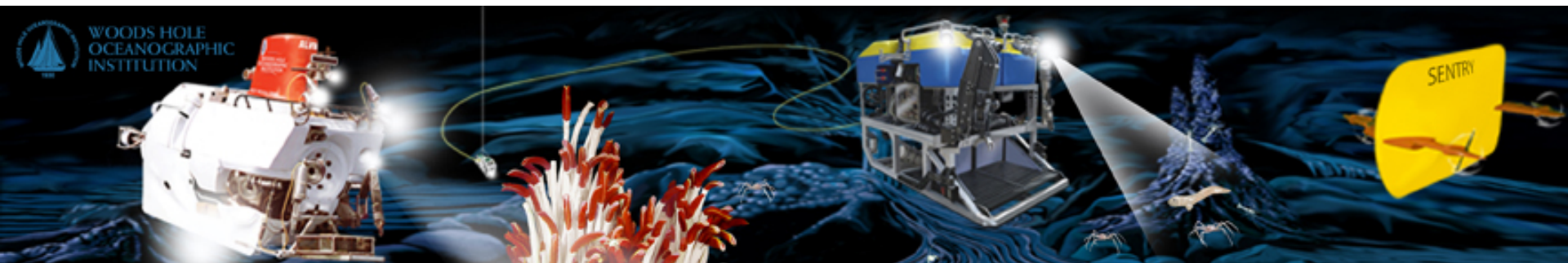
2015 MGDS – New Features

- Map-based access to geo-located images
- Full precision bathymetry in GMRT Synthesis
- Download Reports by platform *(incl. PI-submitted data)*
- Enhanced Web Services
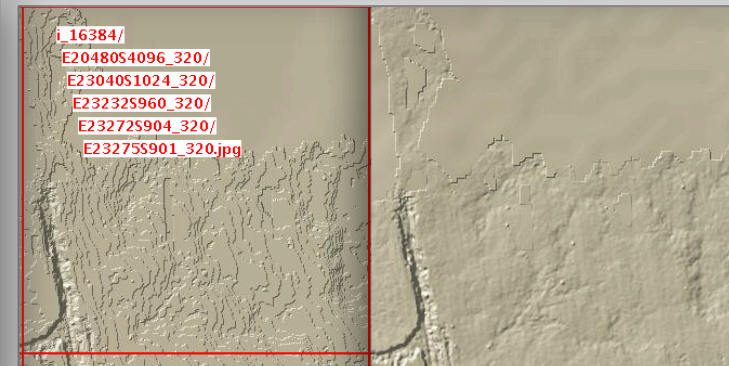 - Data Access & Map Interfaces

MGDS – NDSF Data Discovery

- Search Interface
  - Cruise ID, Platform ID (e.g. Sentry), Platform type (e.g. ROV) Dates, Investigator(s), Reference(s), Location/Extent of Data, Data Type, File Format + more
- Access all available files related to an individual Dive
- Web Services
  - Basis of Search UI
  - Improve access
    - Machine-machine
    - IEDA Data Browser
  - File Service (EC GeoWS)
  - Event (Dive) Service
- Presentation at 2015 AGU

EarthCube Integrative Activity

- *Interdisciplinary Earth Data Alliance as a Model for Integrating EarthCube Technology Resources and Engaging the Broad Community*
  - Funded: Summer 2015
  - Partner Workshop: Fall 2015
- NDSF Components (Collab. with MGDS)
  - Enhance Dive Metadata
    - Community consensus on metadata needs
    - Efficient capture of high-quality metadata
    - Well-aligned with WHOI Data Convergence Effort
  - Deploy Web Services for NDSF
    - Dive metadata service
    - Web Service for expanding access to Virtual Van/FrameGrabber