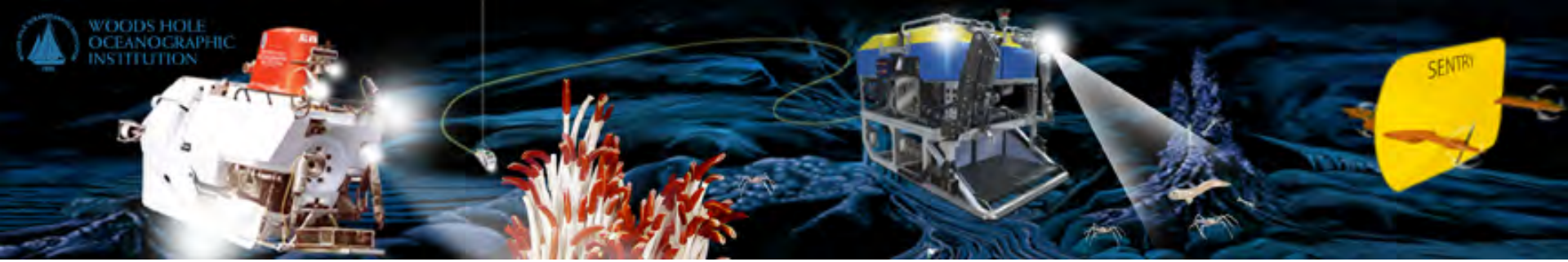


Data Management Update

Vicki Ferrini

IEDA Data Facility

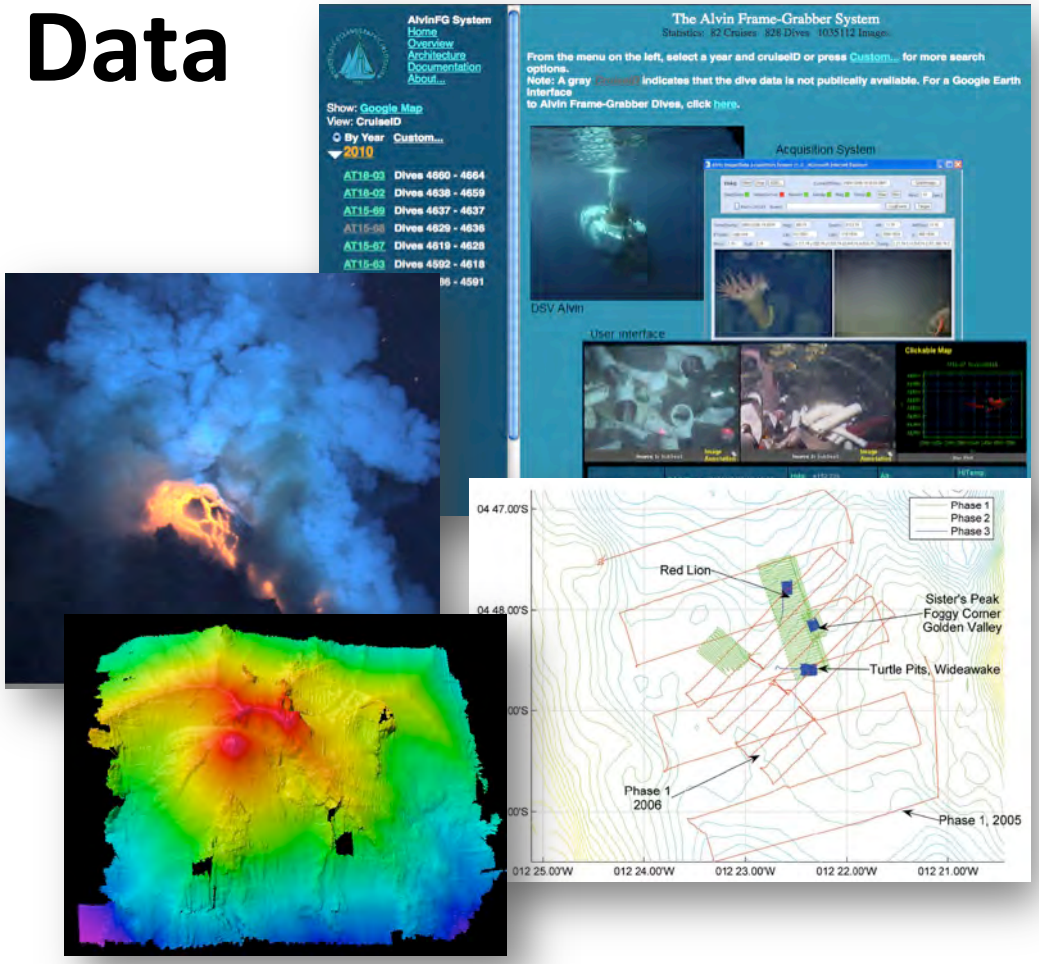
Lamont-Doherty Earth Observatory

IEDA
INTERDISCIPLINARY
EARTH DATA ALLIANCE

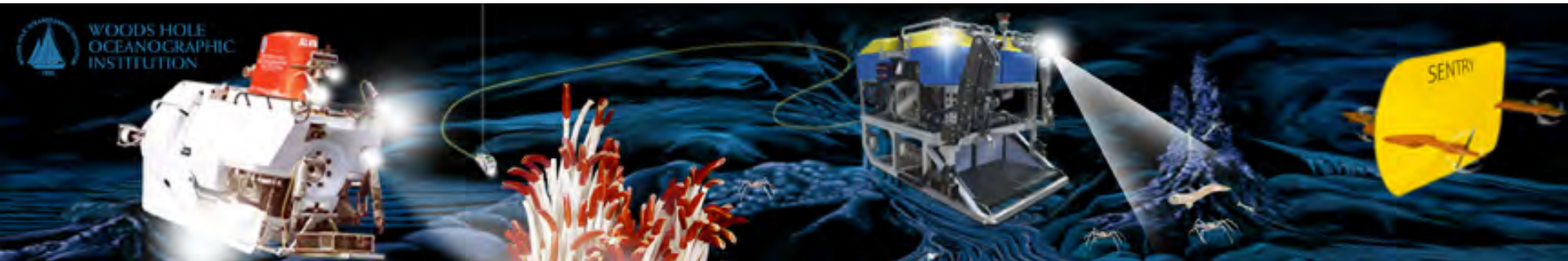
MGDS
marine-geo.org

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.)

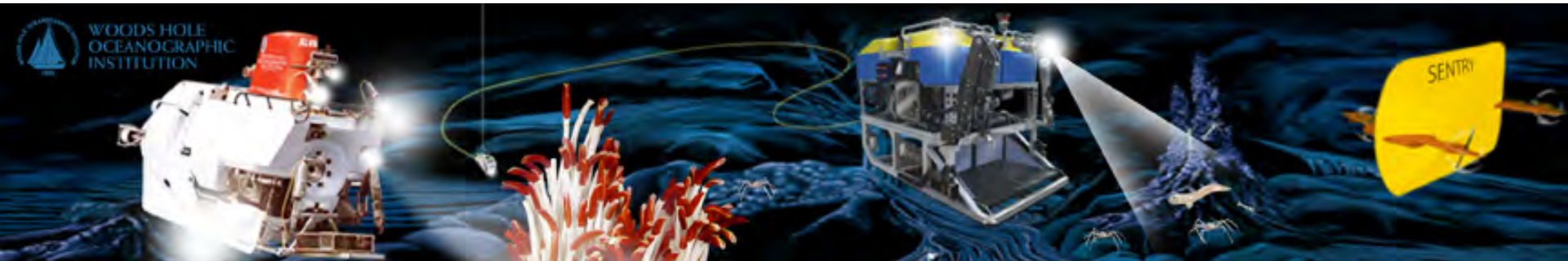


MGDS
marine-geo.org

IEDA: MGDS

IEDA
INTERDISCIPLINARY
EARTH DATA ALLIANCE

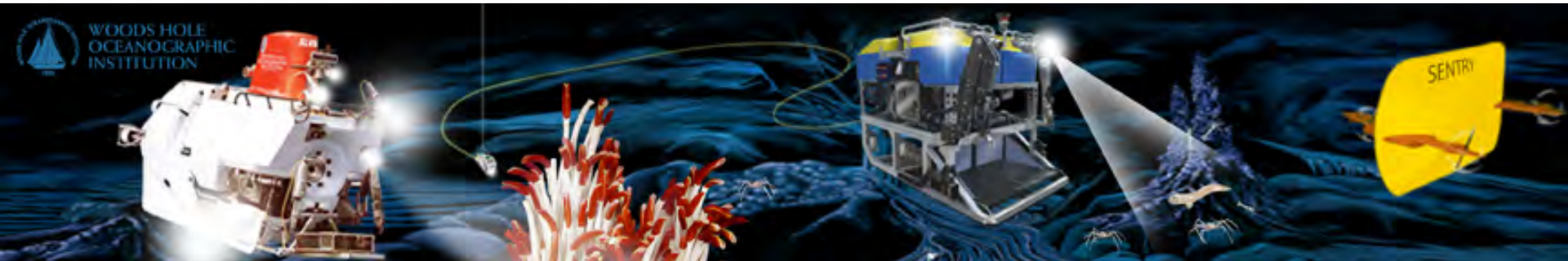
- Part of IEDA Data Facility
- Metadata catalog with access to data files
  - Primarily marine geophysical & complementary data types
- NSF OCE data policy-recommended
- Existing Data Collections include NDSF Data
  - Ridge 2000 & MARGINS Data Portals
  - NDSF Data Portal (deployed 2008)
  - Map-based 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*

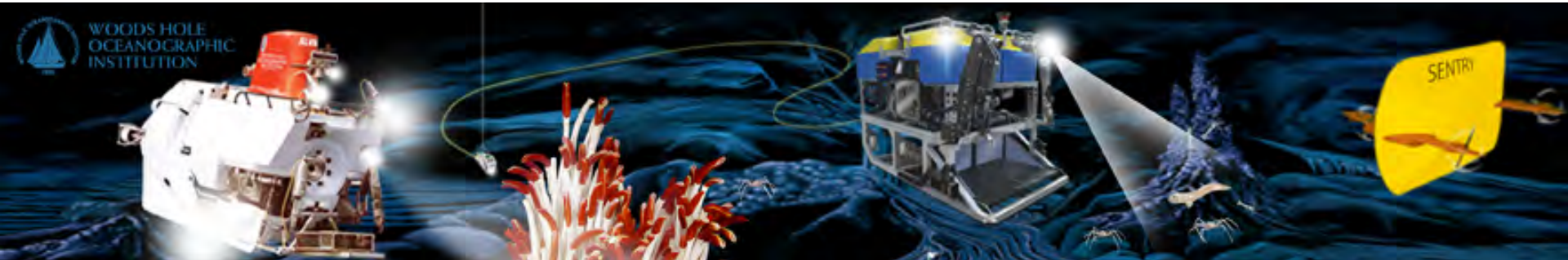
- Metadata from > 1,200 Dives
- > 46K Downloadable Data Files, 1.7 TB
  - Backscatter, Bathymetry, Magnetics, Nav, Photos, Photomosaics, Sidescan
- 14 Platforms (AUV, HOV, ROV)
- Variety of Contributors
  - WHOI/NDSF, MBARI, Schmidt Ocean Institute, Waitt Institute & Individual Scientists

**as of 12/12/2015*



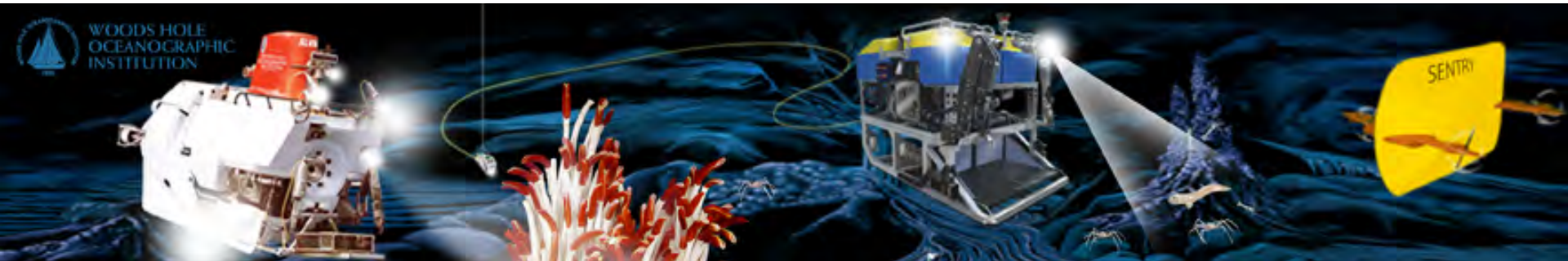
NDSF-MGDS Data Transfer Pilot (2013)

- 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



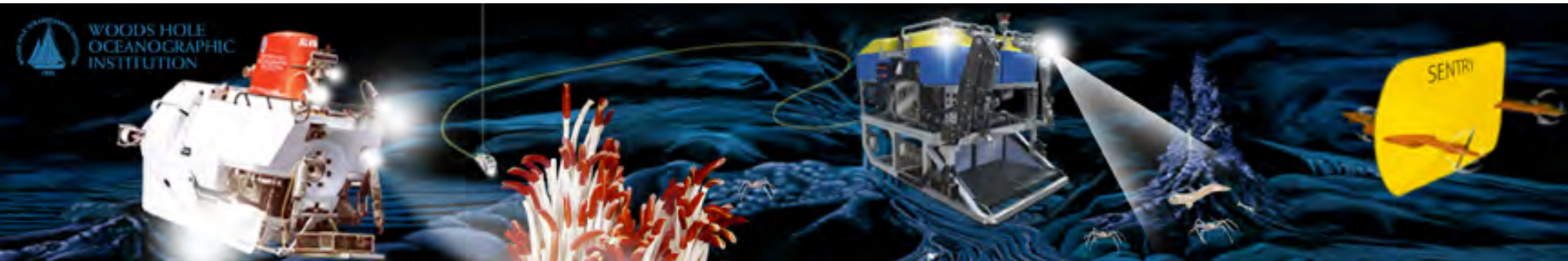
Review of Data Transfer Pilot (2014)

- 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



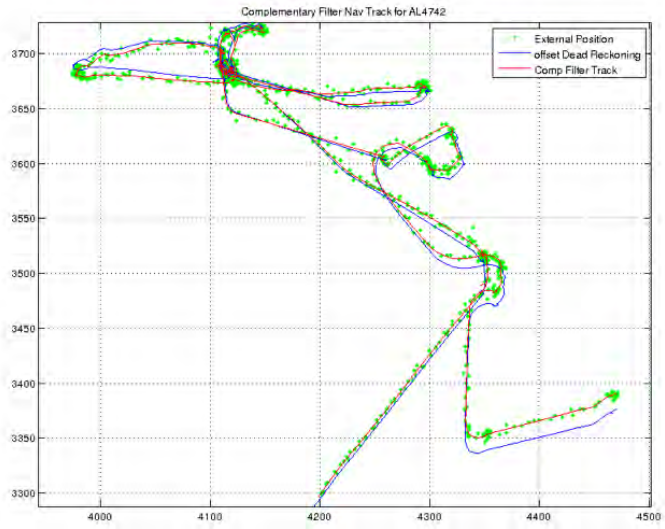
Review of Data Transfer Pilot (2014)

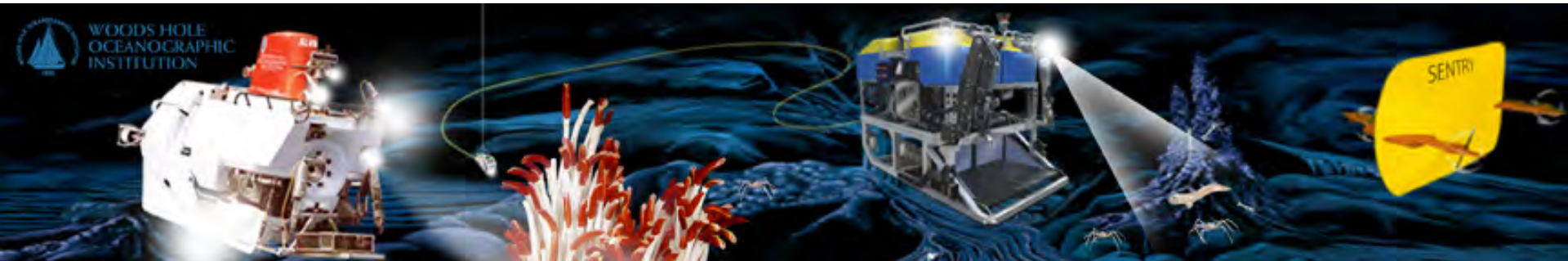
- Technical challenges to be addressed by NDSF
  - Standardize data deliverables facility-wide (metadata, naming conventions, formats, versions, directory structure)
- Additional features to be developed by MGDS
  - Enhance Web Services & data access services
  - Data Download Reports by vehicle
  - Ongoing development of access tools & interfaces



2015 NDSF Data Accomplishments

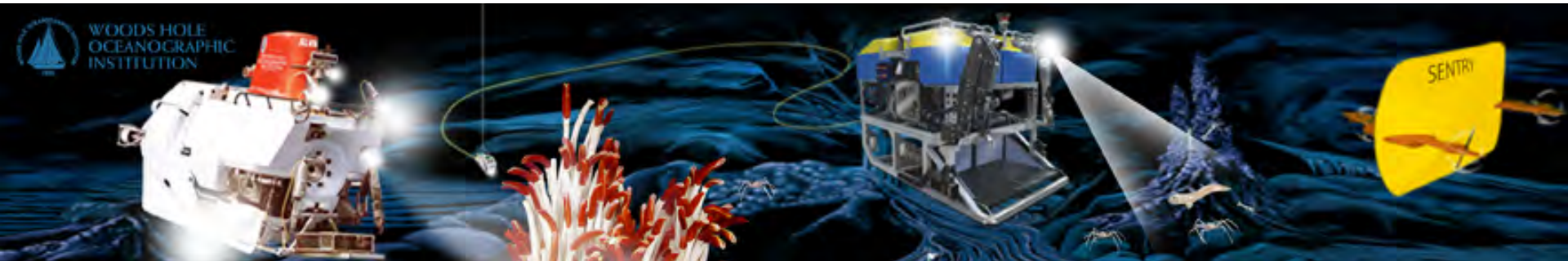
- Developed & deployed middle-ware to apply Sentry processing pipeline to Alvin data
  - Processed navigation
  - Consistent file formats
  - Consistent organization
- Streamlined NDSF->MGDS data transfer pathways (no need to copy data to hard drives)





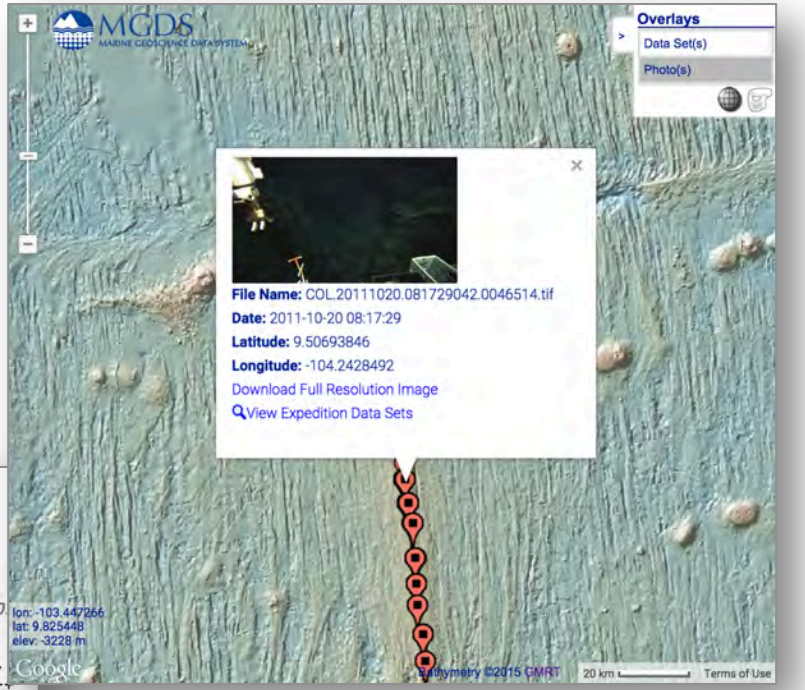
2015 NDSF Data Accomplishments

- 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 Accomplishments

- Map-based access to geo-located images
- Full precision bathymetry in GMRT Synthesis
- Download Reports by platform *(incl. PI-submitted data)*



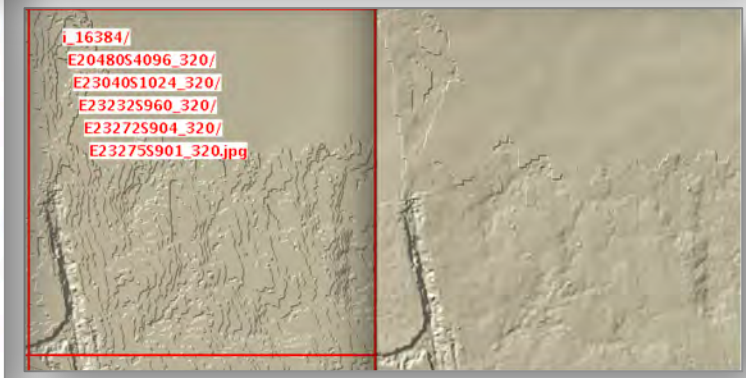
Data Download Report for Sentry

Data Download Reports are prepared bi-annually and are sent to all Scientists associated with downloaded data including Chief/Co-Chief Scientists of field programs, project PIs and Co-PIs, and data set Investigators. Please contact us with any questions or concerns.

MGDS Data Downloads

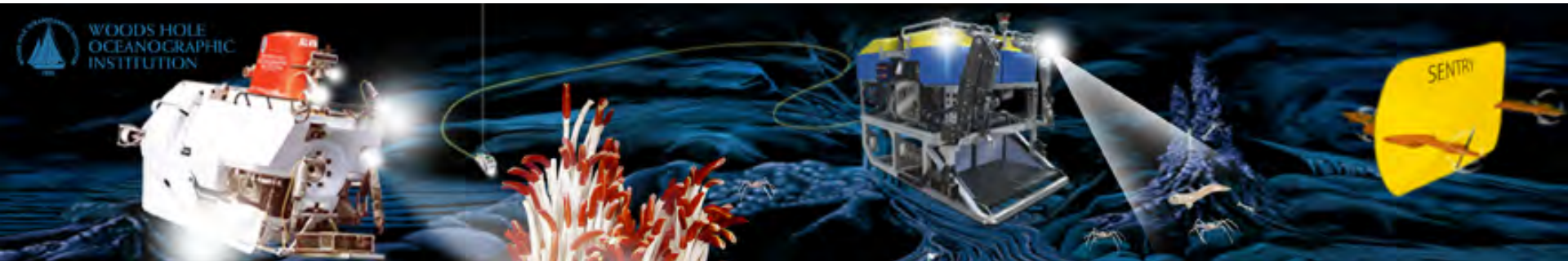
Report covers January 01, 2010 to June 30, 20

Collection ID	Chief Scientist	Data Set Investigator(s)	Device Info	Total Size	Total File Downloads	Data Type(s)	Download Purpose(s)
AT15-63	Sinton	White, Sinton	*AUV: Sentry* Sonar:Multibeam	8 GB	243	Bathymetry *(Grid)* doi: 10.1594/IEDA/318210	
AT18-03	Fisher	Fisher	*AUV: Sentry* Sonar:Multibeam	865 MB	20	Bathymetry *(Grid)* doi: 10.1594/IEDA/320811	
AT18-03	Fisher	Fisher	*AUV: Sentry* Sonar:Multibeam	4 MB	13	Bathymetry *(Image)*	
AT18-11	Valentine	Valentine	*AUV: Sentry* Sonar:Multibeam	111 MB	1	Bathymetry:Swath Backscatter:Acoustic *(Swath)*	
						Bathymetry:Swath	



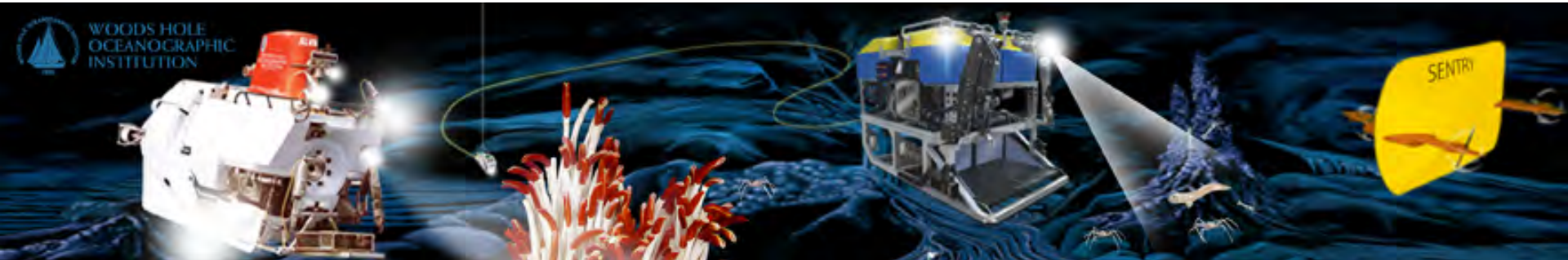
2015 MGDS Accomplishments

- MGDS Web Services
  - Basis of Search Interfaces (UIs)
  - Enhance Discovery & Access
    - Machine-machine queries
- Dive Web Feature Service
  - Returns metadata (including Lat/Long for map display) + links to data based on Vehicle Name, Type, Dive ID, etc
  - Easy/customizable map integration (Google Maps, etc)



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



Underwater Video Workshop

Establishing Community Standards for Underwater Video Acquisition, Tagging, Archiving and Access

Dates TBD:
Late May/Early June 2016 @ URI

PIs: Ferrini (LDEO), Coleman (URI), Soule (WHOI)

Video Workshop - GOALS

- Understand current state of video logging & data management across community *(including industry)*
  - pre-workshop surveys
- Summarize technical and science-based needs, challenges, existing & desired solutions
- Develop Community Consensus on:
  - Formats, logging standards, metadata standards, storage/archiving
  - How to best move forward
- Report to be made available on UNOLS website

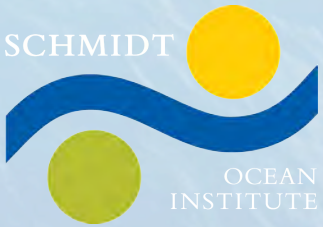
Video Workshop – Steering Committee

- PIs (Ferrini, Coleman, Soule)
- Brian Schlining
  - Monterey Bay Aquarium Research Institute (MBARI)
  - Software Engineer, creator of MBARI VARS system for video annotation data management
- Webb Pinner
  - The Global Foundation for Ocean Exploration (GFOE)
  - Oceanographic Data Management System Architect

Video Workshop Planning

- Pre-workshop survey (early 2016)
- ~25-35 in-person participants
  - Operators
  - Science Users
  - TV/Video Professionals
  - Data Managers/Metadata Specialists
- Remote participation
  - *Lots* of interest already!!

OceanVideoLab



oceanvideolab.org

Ocean Video Lab – *Overview*

- Create a searchable online database of underwater video observations
  - Science
    - exploration, ground truth, machine learning, big data
  - Outreach
    - public engagement, quick access to highlight content for outreach efforts
- Enable broad discoverability
  - YouTube
  - API for accessing tagged information and integrating with other ocean image tagging efforts
  - Link to related data resources that complement imagery where possible (e.g. MGDS)

SCHMIDT OCEAN INSTITUTE

OVL POC- Technology

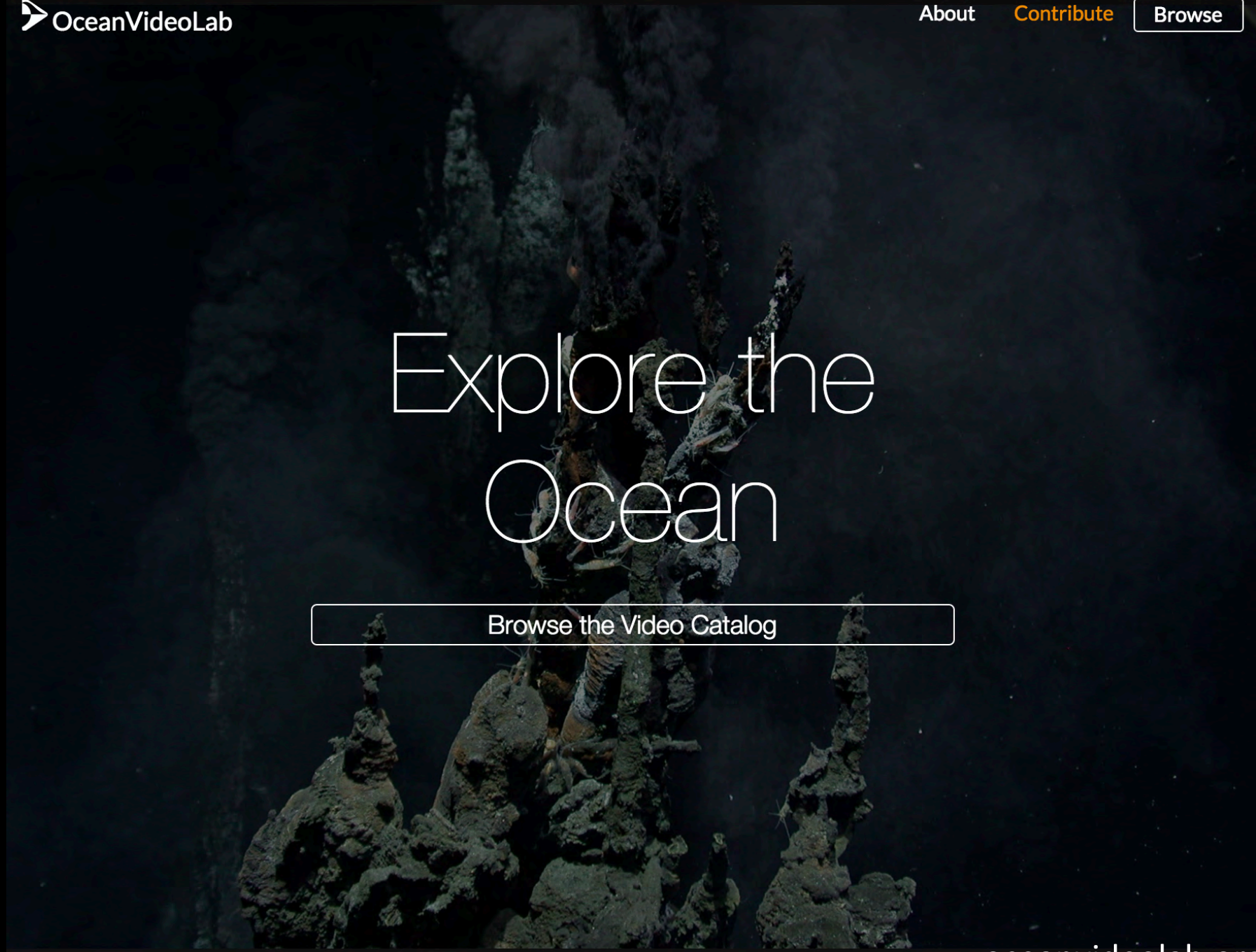
- YouTube
  - Broad Usage vs. Vimeo
    - YouTube ~1 Billion Users/Month
    - Vimeo ~ 100 Million Users/Month
  - Provides API for interacting with content
  - Excellent Performance
    - Bandwidth limitations handled by YouTube
  - SOI live streams to YouTube
- PostgreSQL Relational Database
- PostGIS
- Leaflet Map Interface (GMRT basemap)

SCHMIDT OCEAN INSTITUTE

OVL POC– *Design Goals*

- Registration page for inputting URL to YouTube video
- Accommodate *any* underwater video
  - temporally continuous
  - online
  - openly accessible
- YouTube Player
- Simple annotation interface
- Integrate with Start Time + Navigation
  - Display vehicle position in map interface
  - Geo-locate observations
  - *Encouraged* (not required)

SCHMIDT OCEAN INSTITUTE


OceanVideoLab
About
Contribute
Browse
Explore the Ocean
Browse the Video Catalog

OVL – Video Browser

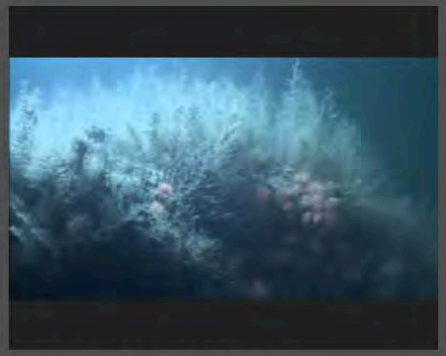
OceanVideoLab

- Browse
- Site Map
- Contribute
- About
- Home

Alvin Dive 3974
0 annotations

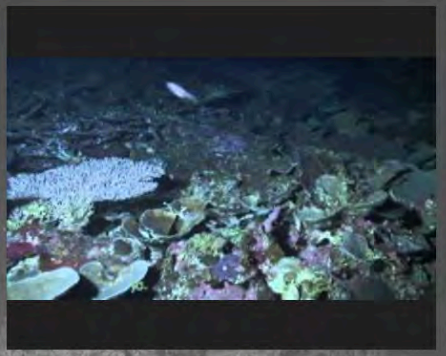
LIS Dive 02 - Video 3
0 annotations

LIS Dive 02 - Video 4
1 annotations

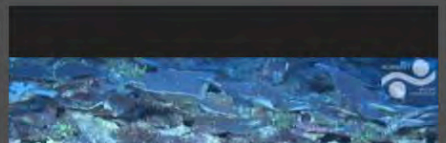


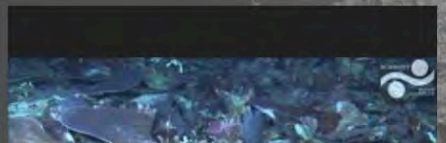
LIS Dive 07 - Video 6
0 annotations

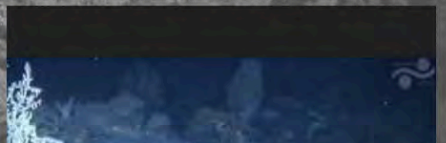
Scott Reef - coral spawning ROV Dive 1
3 annotations



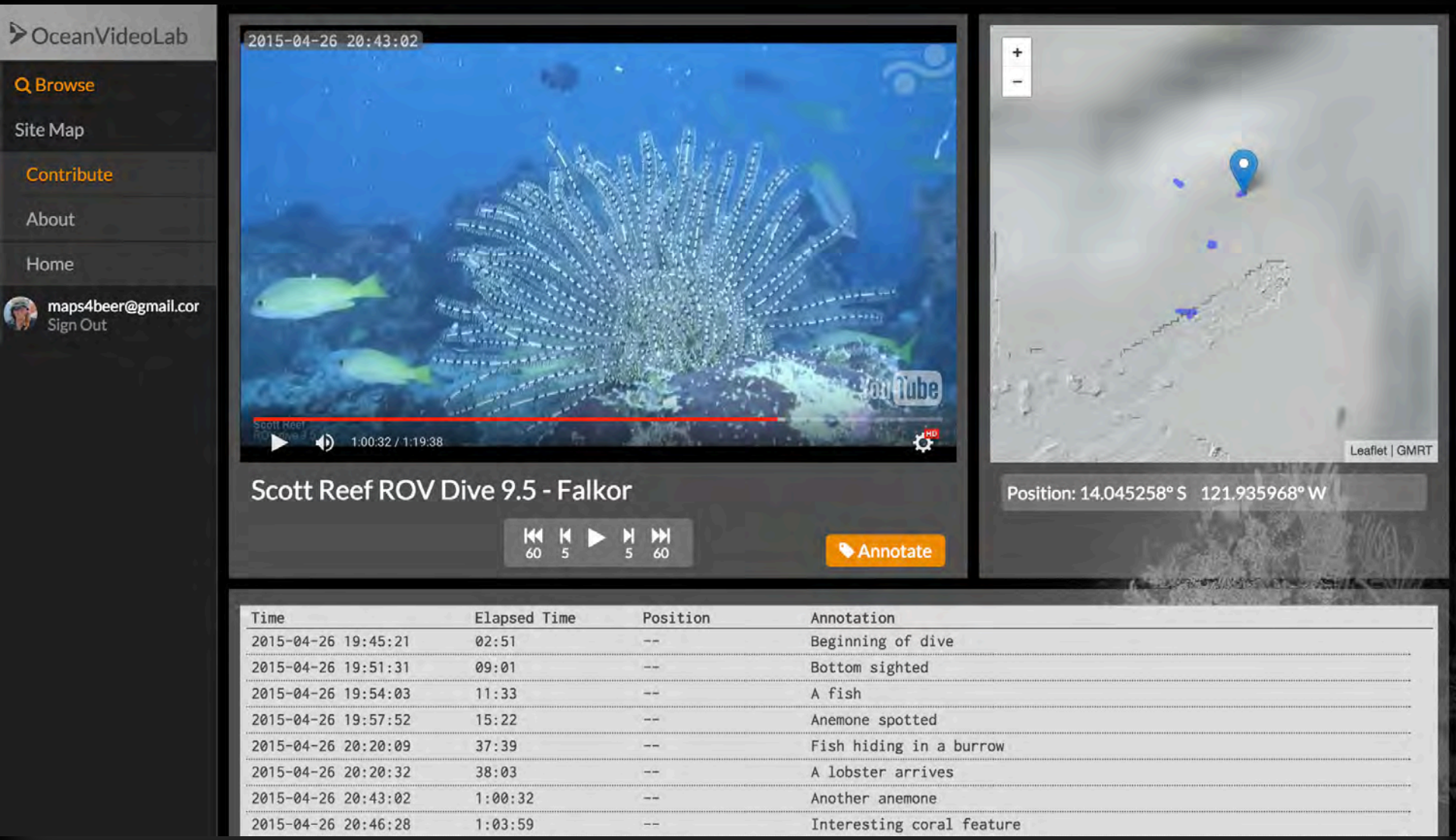
Scott Reef - coral spawning ROV Dive 1.5
3 annotations





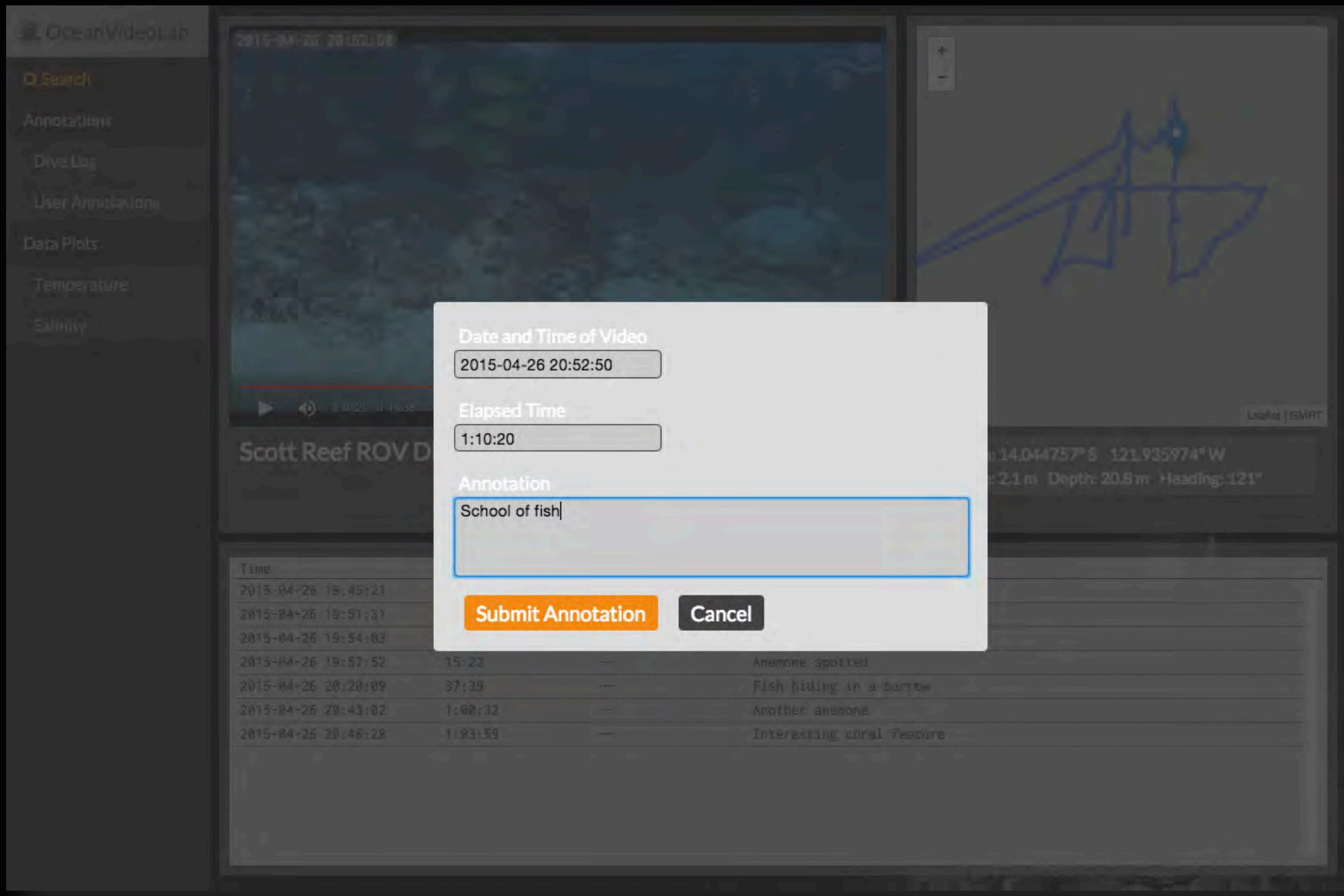


OVL – Video Player

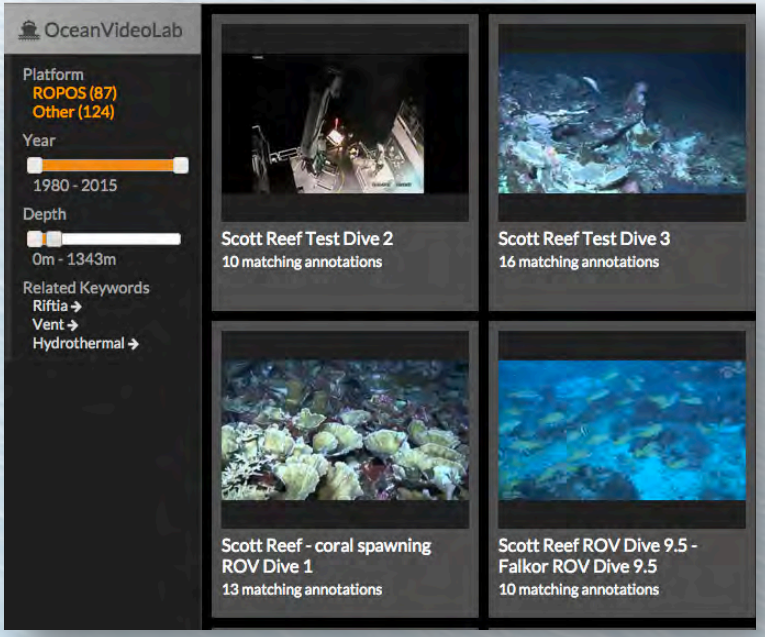


Time	Elapsed Time	Position	Annotation
2015-04-26 19:45:21	02:51	--	Beginning of dive
2015-04-26 19:51:31	09:01	--	Bottom sighted
2015-04-26 19:54:03	11:33	--	A fish
2015-04-26 19:57:52	15:22	--	Anemone spotted
2015-04-26 20:20:09	37:39	--	Fish hiding in a burrow
2015-04-26 20:20:32	38:03	--	A lobster arrives
2015-04-26 20:43:02	1:00:32	--	Another anemone
2015-04-26 20:46:28	1:03:59	--	Interesting coral feature

OVL – Annotation



OVL –Next Steps



- 2016
  - Beta Testers (NDSF Users ??)
  - Develop & Deploy API
  - Google Maps
  - Search Interface
  - Enhance access to related data (plots, map overlays, etc)
  - Expand code for Nav input/integration
  - FrameGrabs & detailed tagging within frame ?
- Beyond 2016…?
  - Integration of tags/imagery with distributed data ?
  - Expand to Vimeo ?
  - Code-share with complementary efforts (e.g. Squidle)
  - Machine Learning & Big Data ?