

**Deepwater Horizon Oil Spill (DWHOS)
Water Column Technical Working Group**

Image Data Processing Plan: Holocam, DAVPR, VPRII

Principal Investigator: Cabell Davis, WHOI

June 7, 2012

Prepared by:

Cabell Davis (WHOI), Deborah French McCay, Eileen Graham, Melanie Schroeder (ASA), and Sandra Arismendez (NOAA) on behalf of the Trustees

Reviewed by:

NOAA: Dan Hahn, John Quinlan

Louisiana: Amanda Vincent (LDEQ)

BP/Cardno ENTRIX: William Graeber, Laura Riege, Tim Thompson, Jeffrey Simms, Jodi Harney, Craig Rice, and Mark Benfield (LSU)

Period of Performance

December 1, 2011 – October 1, 2012

Objectives and Approach

This plan describes processing of imaging data collected as part of the NRDA cooperative sampling efforts in 2010-2011, where the principal investigator is Dr. Cabell Davis (Woods Hole Oceanographic Institute, WHOI). Imaging and particle size data were collected by a variety of instruments on various cruises. This plan covers the data collected by:

- Holographic Camera (Holocam) (Attachment 1),
 - particle size range: few microns to several cm
 - maximum depth: 1200 m
- Digital Autonomous Video Plankton Recorder (DAVPR) (Attachment 2),
 - particle size range: 50 microns to several cm
 - maximum depth: 1200 m
- Video Plankton Recorder II (VPRII) (Attachment 2).
 - particle size range: 50 microns to several cm
 - maximum depth: 150 m

Other work plans describing analysis of data from Shadowed Image Particle Profiling and Evaluation Recorder (SIPPER, Principal Investigator (PI): Andrew Remsen), In Situ Ichthyoplankton Imaging System (ISIIS, PI: Robert Cowen), and FlowCAM (PI: Malinda Sutor, ZOOSCAN) involve similar analytical procedures and the processing for these data will be included in separate processing plans.

As part of the NOAA/NRDA Water Column TWG sampling effort, digital images for the identification and quantification of plankton, marine snow (suspended particulate material, SPM), and oil droplets were collected using a Holocam, DAVPR, and/or a VPRII on seven cruises between June 2010 and April 2011 (*Jack Fitz 3*, *Ocean Veritas 9*, *American Diver 1* (Legs 1 and 2), *HOS Davis 2*, *Walton Smith 2*, *Arctic 6*, and *Oceanus*) (summarized in Tables 1-3, depicted in Figures 1-3). During these cruises, raw images were collected and stored to archival disk drives for subsequent shore-based analysis. These archived data require post-processing to extract sub-images, classify and measure the objects imaged, compute abundance and size distributions, and identify location (latitude, longitude, depth) of the imaged objects. The end data products will include the extracted sub-images (in a common digital photo format), spreadsheets of image information (classifications, measurements, locations, etc.), summary tables and figures, and narrative descriptions of the data.

Table 1. Cruises where Holocam data were collected.

Cruise	Date	Approximate Location
<i>Jack Fitz 3</i>	June 11 – June 20, 2010	Near wellhead, CTD-mounted cast deployments
<i>Ocean Veritas 9</i>	July 13 – July 17, 2010	Near wellhead, CTD-mounted cast deployments
<i>American Diver 1</i>	Leg 1: July 20 – July 21, 2010 Leg 2: July 29 – August 6, 2010	27-81 NM (50-150km) SW of the wellhead
<i>HOS Davis 2</i>	August 25 – September 3, 2010	54-224 NM (100-360km) to the SW of the wellhead
<i>Walton Smith 2</i>	September 19 – September 25, 2010	12 NM (22 km) transects along shelf break

Table 2. Cruises where DAVPR data were collected.

Cruise	Date	Approximate Location
<i>American Diver 1</i>	Leg 2: July 29 – August 6, 2010	27-81 NM (50-150km) SW of the wellhead
<i>HOS Davis 2</i>	August 25 – September 3, 2010	54-224 NM (100-360km) to the SW of the wellhead
<i>Walton Smith 2</i>	September 19 – 25, 2010	12 NM (22 km) transects across shelf break
<i>Arctic 6</i>	January 14 – January 24, 2011	12 NM (22 km) transects across shelf break

Table 3. Cruise where VPRII data were collected.

Cruise	Date	Approximate Location
<i>Oceanus</i>	March 9 – April 8, 2011	Shelf and offshore: NW FL to LA-TX border

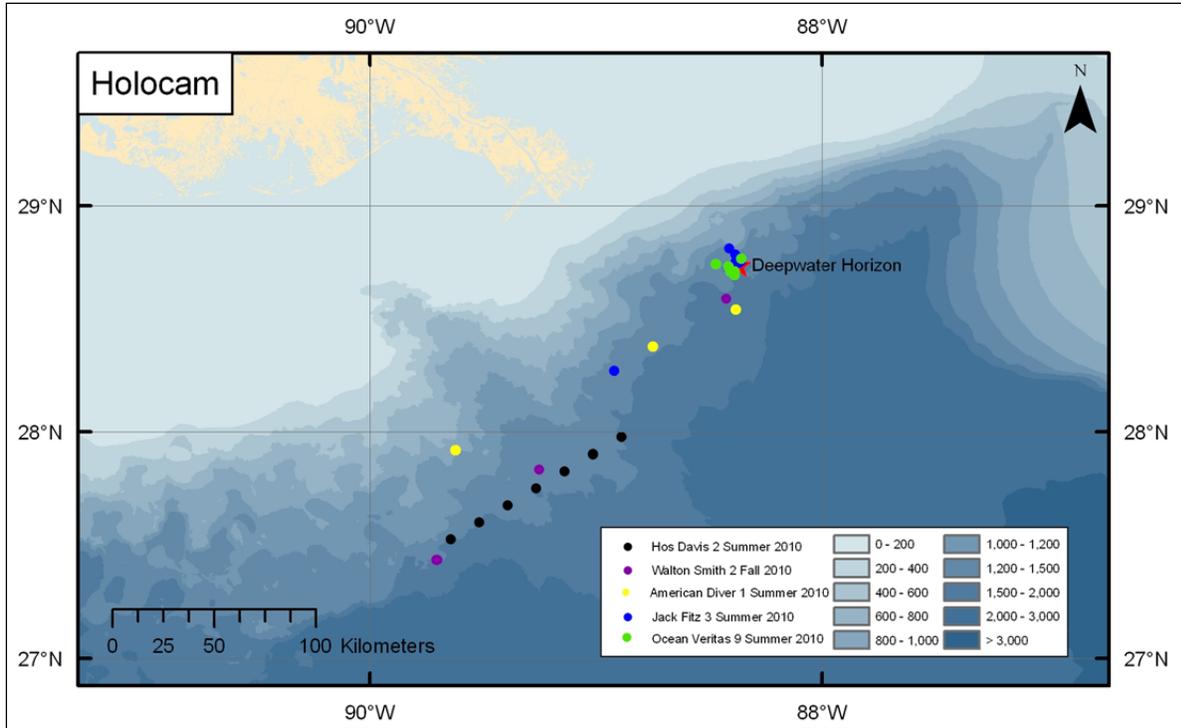


Figure 1. Locations of NRDA Holocam data collected in the Gulf of Mexico summer and fall of 2010.

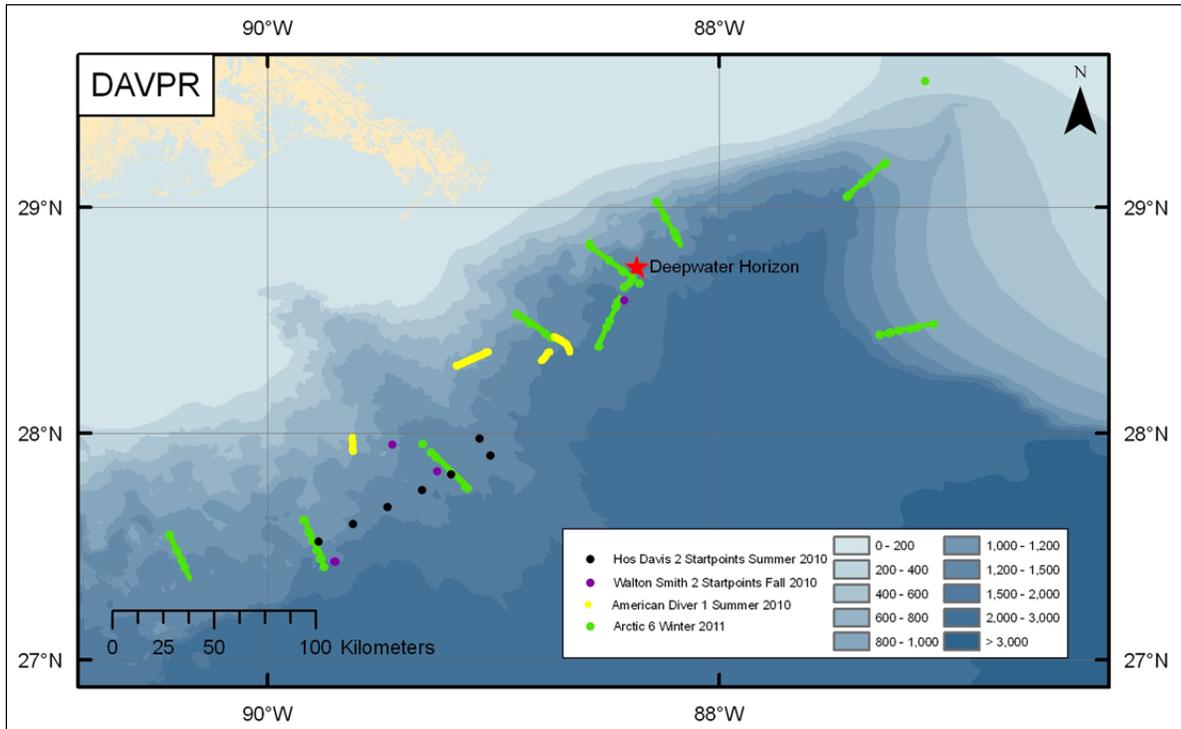


Figure 2. Locations of NRDA DAVPR data collected in the Gulf of Mexico summer 2010 – winter 2011.

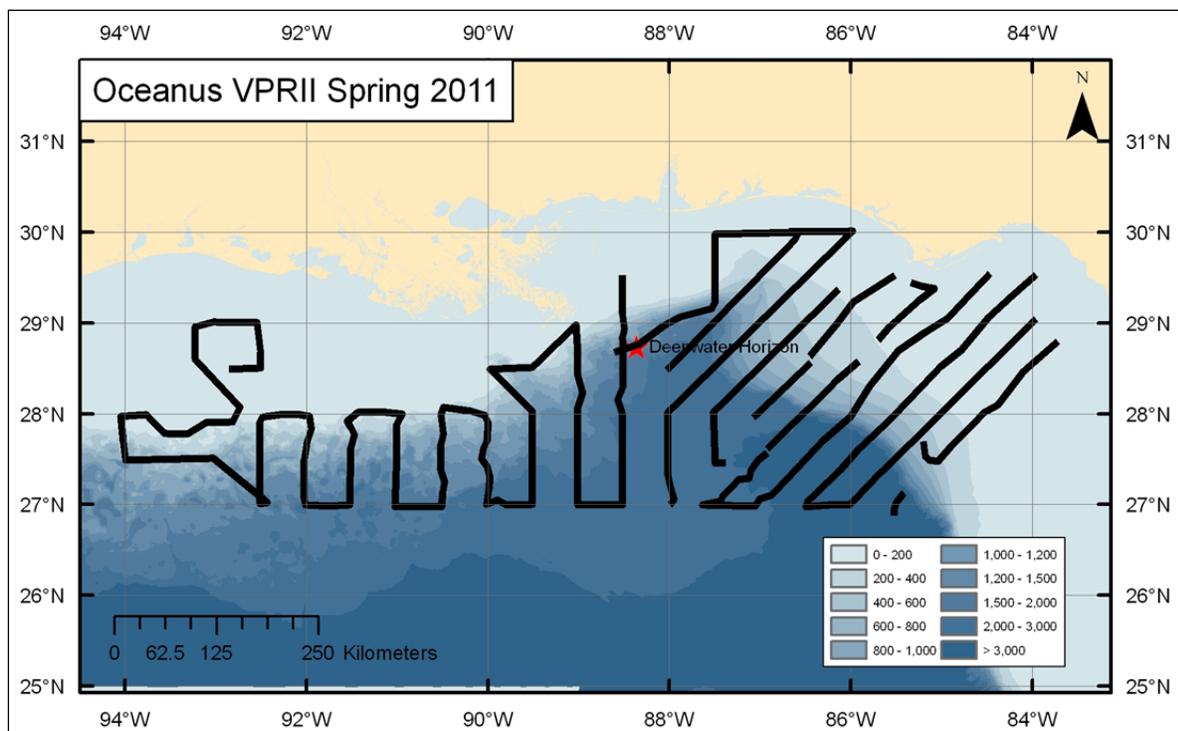


Figure 3. Locations of NRDA VPR II data collected in the Gulf of Mexico spring 2011.

Data Processing Prioritization and Schedule

Holocam data sets collected by Cabell Davis have been prioritized for processing on the basis of the potential for identifying oil droplets and measuring the size distribution of such droplets. Holocam data from the *Jack Fitz 3* (June 11 – June 20, 2010), in particular, are expected to yield information regarding oil droplet size distribution during the active release from the Macondo well. Holocam data from *Ocean Veritas 9* (July 13 – July 16, 2010) will also be examined for possible presence of oil droplets (the release having stopped on July 15, 2010).

Holocam, DAVPR and VPR II can also provide valuable information on the distribution and abundance of plankton in the study area. The DAVPR data were prioritized over the VPR II data because DAVPR covers a longer timeframe and samples deeper in the water column. Plankton analyses on the Holocam data will be conducted on selected cruises only; specifically, those during which other plankton imaging systems were not deployed. Only if time and resources allow will the full suite of Holocam data be analyzed. Holocam is given the lowest priority for plankton analyses because data processing is completely manual and time intensive.

Given the above, the priority sequence of cruises for processing the Holocam, DAVPR, and VPR II data is as follows.

Priority 1: Holocam – oil droplet data

1a. *Jack Fitz 3*: Holocam – oil droplet data only (no plankton)

1b. *Ocean Veritas 9*: Holocam – possible oil droplet data (no plankton)

Completion Expected: March 31, 2012

Priority 2: DAVPR – plankton and marine snow (SPM)

3a. *American Diver 1* and *Walton Smith 2*

3b. *HOS Davis 2*

3c. *Arctic 6*

Completion Expected: May 15, 2012

Priority 3: VPRII – plankton and marine snow (SPM): *Oceanus*

Completion Expected: July 15, 2012

Priority 4: Holocam – plankton

4a. *Jack Fitz 3* and *Ocean Veritas 9*

4b. *HOS Davis 2* (if time and resources allow)

4c. *American Diver 1* and *Walton Smith 2* (if time and resources allow)

Completion Expected: August 15, 2012

Methodology

Background

Holocam, DAVPR, and VPRII image data were collected in conjunction with data from other sampling systems as part of the overall NRDA effort. The image data require post-processing and time matching with global positioning satellite (GPS) and conductivity, temperature, and depth (CTD) data to determine locations and depths at which images were captured. Raw image files themselves require post-processing to extract sub-images within regions-of-interest (ROIs) for resolving individual objects including plankton, suspended particulate matter (SPM), and oil droplets. Observed objects in these ROIs then need to be identified using automated identification methods, which in turn require compilation of training data sets and classifiers. ROIs typically contain only one object of interest but multiple objects can sometimes be found in one ROI. However, ROIs with more than one object, or overlapping objects, are classified in a “multiple objects” category or as “unknown”. Both of these categories require manual review to properly classify all objects within the ROI.

Once images have been classified, they can be time-matched with CTD data to obtain location and depth information for each image. The total number of images captured for a given data category (e.g., SPM) per unit time is converted to number of objects per unit volume using the calibrated image volume and sampling rate for each instrument (volume per unit time). In addition to the number of objects per unit volume, size of objects will be determined by counting the pixels within the outline of the object and multiplying by the area per pixel (Davis et al., 2004, 2005) and then converting to equivalent spherical diameter (Sheldon et al. 1972, Herman 1992) and used to estimate size distributions. These methods have been described in detail in peer-reviewed publications (Davis et al., 1992a,b, 1996, 2004, 2005; Davis and McGillicuddy, 2006; Hu and Davis, 2005, 2006; Li et al., 2007; Loomis, et al., 2007).

All of the Holocam, DAVPR, and VPRII data collected as part of the NOAA/NRDA water column TWG sampling effort are shown in Tables included in the attachments. A total of 28,385 raw Holocam files (1,296 GB), 784 raw DAVPR files (344 GB), and 10,566,971 VPRII files (415 GB) were collected during NRDA cruises.

Specific work tasks:

1) *Holocam Data – Jack Fitz 3* – Processing of holograms collected during the *Jack Fitz 3* cruise is required to generate size distributions of objects classified as oil droplets versus depth in a standard format. Depth- and station-dependent changes in size distribution will be determined. Custom software (Loomis & Davis, unpublished) has been developed to classify objects as oil droplets and automatically measure, and count such objects captured on the holograms. The computational algorithm will be modified to extract plankton and marine snow size and abundance data from the holograms. If automation is not possible or requires an effort exceeding time available for analysis, a subset of holograms (Priority 4a) will be manually sorted for these categories. The abundance and sizes of each category will be determined and reported by location and depth. See Attachment 1 for a more detailed description of the processing protocol.

2) *Holocam Data – Other cruises (not Jack Fitz 3)* – The computational processing (or manual methods) will be used to extract regions-of-interest (ROIs) from holograms collected on *Ocean Veritas 9*, *American Diver 1* (Legs 1 and 2), *HOS Davis 2*, and *Walton Smith 2*. In addition, the holograms will be time-matched to the CTD data so that the depth and location coordinates of each hologram image can be established. The ROIs will be computationally sorted into classes (e.g. various plankton taxa, SPM, oil droplets). The abundance and size information for each category will be determined and reported by location and depth. See Attachment 1 for a more detailed description of the processing protocol.

3) *DAVPR and VPRII Data* – The raw DAVPR data collected during the *American Diver 1* (Leg 2), *HOS Davis 2*, *Walton Smith 2*, *Arctic 6*, and *Oceanus* cruises will be analyzed for plankton, SPM, and oil droplets. Color ROIs will be extracted from the raw DAVPR files and converted to grayscale. The VPRII ROIs were extracted in real-time during the cruise, generating $\sim 10^7$ grayscale ROIs. The Visual Plankton software program (Davis et al., 2005) will be used to train and use automated digital classifiers to identify and sort the ROIs into different categories including various taxa, SPM. The Visual Plankton software will be used to time-match the results of the DAVPR/VPRII and CTD data, plot the distributional patterns, and save the processed data for further analyses. The abundance and sizes of each category will be reported by location and depth. See Attachment 2 for a more detailed description of the processing protocol.

Quality Assurance and Control

Quality assurance of data products will be achieved via a phased QA/QC process to be implemented across the NRDA plankton/particle imaging PI's (Cabell Davis (Holocam, DAVPR, VPRII), Andrew Remsen (SIPPER), Robert Cowen (ISIIS), and Malinda Sutor (FlowCAM, ZOOSCAN)). The first phase of QA/QC will involve the sharing of computerized classification methods among the PIs working with digital image analyses of plankton and other classification categories. This activity will assess the degree to which the PIs are classifying imaged objects to a similar level (e.g., SPM and major taxa of plankton such as copepods, etc.). If needed, taxonomic resolution will be adjusted to achieve the best possible agreement between image classification systems, while recognizing the differences in size classes evaluated and discrimination capabilities such that identification is to the lowest possible taxon for each instrument. It is envisioned, but not required, that the initial discussion between the PIs will occur at an in-person meeting. Following this initial cross-coordination, communication is

expected to occur on an as-needed basis, as decided by the PIs and the Trustees. A report will be compiled and provided to all parties to this agreement which will document the common classification schemes that provides the basis for integration across the listed imaging systems. The report will also provide a comparison of the level of taxonomic resolution between the instruments/systems used. If any major updates or changes are made to the classification systems as the data processing progresses, an amended report will be developed and distributed.

A second phase of QA/QC will be performed at the software level. This includes identification of training images, the creation of training image sets, and conducting routine manual checks of image processing software. These 3 tasks will be carried out by Cabell Davis and technicians during the processing of each data set (see descriptions in Attachments 1 & 2). Trained technicians will perform the manual identifications, and Dr. Davis will check their work by identifying a randomly selected subset of at least 5% of the images. Additionally, for the DAVPR and VPRII images, detection and false positive rates will be determined via confusion matrices generated when building the automatic classifiers from sets of training images (Hu and Davis, 2006). A confusion matrix compares the classifications made by a human expert (columns) to those made by the computer (rows). The accuracy, precision and comparability of the identifications varies with taxonomic group. By using equal numbers of images from the training set for each category classified, classification accuracy is quantified by the confusion matrices (i.e., the confusion matrices show the proportion of false positives, true positives, false negatives, and true negatives). Abundances (e.g., number of organisms per liter) are corrected using the information from the confusion matrix. Abundances generated by this automated method are similar to those based on human sorted images (Hu and Davis, 2006). Confusion matrices and associated abundance corrections (Hu and Davis, 2006) will be shared among PI's to ensure the different software systems are performing similarly, and will be included in the official data record released to the Trustees and BP. Any identifications completed for which there is no automated process (e.g. Holocam plankton analysis) will also be subject to a 5% check of the images by Dr. Davis.

A third phase of QA/QC will involve an independent cross-check between the principal investigators working with imaging systems (Davis, Remsen, Cowen, and Sutor). At least 50 images will be randomly selected from each taxonomic group and sent to at least one other PI for their independent identification by manual inspection. This will serve as a quality control check for Dr. Davis' systems and will be part of an integrated system of QA/QC checks across the plankton program. The listed PIs will independently review and identify images and will confer with each other about classifications. As necessary, Davis will also make adjustments to software, as possible, to ensure maximum agreement among investigators.

The final phase of QA/QC will be to determine that the data are reported in geographically correct locations, as compared to field-recorded vessel and station locations for the imagery collections. This task will be performed by the NOAA NRDA Data Management Team after the lab data for a priority level are received, and before the data are released to the Trustees and BP. Since the whole process is based on electronic data, transcription from paper data sheets to electronic media is not anticipated. However if transcription does occur, a cross-check of 100% of all transcriptions will be conducted by someone other than the person who completed the original transcription.

Distribution of Results

Upon completion of the data processing and inter-PI QA/QC procedures (QA/QC phases (1) – (3) as identified above) at each priority level, the principal investigator (Dr. Cabell Davis) will deliver all products generated as part of this work plan (e.g., processed data sets) to the NOAA NRDA Data Management Team. Once the Data Management Team has completed their QA/QC review of the station metadata for that priority level (the final phase above), the data and metadata for that priority level will be made available to the parties of this agreement by means appropriate to the data type as determined by the NOAA NRDA Data Management Team. NOAA and the Louisiana Oil Spill Coordinator's Office (LOSCO) on behalf of the State of Louisiana and BP (or Cardno ENTRIX on behalf of BP) will be alerted when these data become available.

In the interest of maintaining one consistent data set for use by all parties, only the verified and validated data set made available to the parties by the NOAA NRDA Data Management Team shall be considered the consensus data set. In order to ensure reliability of the consensus data and full review by the parties, no party shall publish consensus data until 14 days after such data have been made available to the parties. Any questions raised about the consensus data set, as it was made available to the parties by the NOAA NRDA Data Management Team, shall be handled consistent with the procedures in Section 7.2 of the Deepwater Horizon NRDA Analytical Quality Assurance Plan.

- The trustees and BP shall each designate an individual responsible for raising questions, if any, on the consensus data set.
- If questions are raised, the two designated individuals will meet to determine the source of the difference and resolve.
- The questions raised and their resolution shall be distributed to all parties.
- No changes to the consensus data set will be made if the differences are considered immaterial by both designated individuals, acting on behalf of the parties.
- If the parties agree that changes to the dataset should be made, the dataset will be updated in accordance with the resolution and reposted with a notation that the dataset has been revised.
- If the designated individuals do not agree on how to resolve the difference concerning the consensus data set, the designated individuals shall request assistance from the Assessment Managers for the trustees and BP.

Retention of Materials

All imagery will be retained, along with any changes in processing software or results. All of this information will be maintained during all review steps in the process and stored in secure locations under Trustee control and will be provided to all parties as part of the data release process.

All materials associated with the collection or analysis of samples under these protocols or pursuant to any approved work plan, including any remains of samples and, including remains of extracts created during or remaining after analytical testing, must be preserved and disposed of in accordance with the preservation and disposal requirements set forth in Pretrial Orders (“PTOs”) # 1, # 30, #35, # 37, #39 and #43 and any other applicable Court Orders governing tangible items

that are or may be issued in MDL No. 2179 IN RE: Oil Spill by the Oil Rig "DEEPWATER HORIZON" (E.D. LA 2010). Destructive analytical testing of oil, dispersant or sediment samples may only be conducted in accordance with PTO # 37, paragraph 11, and PTO # 39, paragraph 11. Circumstances and procedures governing preservation and disposal of sample materials by the trustees must be set forth in a written protocol that is approved by the state or federal agency whose employees or contractors are in possession or control of such materials and must comply with the provisions of PTOs # 1, # 30, # 35, 37, #39 and #43.

Budgeting:

The Parties acknowledge that this budget is an estimate, and that actual costs may prove to be higher due to a number of potential factors. The costs indicated in Budget Chart # 1 below and any additional reasonable costs within the scope of this workplan that may arise shall be reimbursed by BP upon receipt of written invoices submitted by the Trustees. The Trustees will make a good faith effort to notify BP in advance of any such increased costs. It is anticipated that the work will require approximately 6 months of the PI’s time to complete.

Budget Chart.

NOAA Contractor Labor:	
Cabell Davis, PI (6 months)	
Phil Alatalo, technician (6 months)	
Labor Total	\$165,275
Travel	\$2,885
Other Direct Costs	\$4,000
Indirect Costs	
Laboratory	\$95,078
General Administrative	\$48,114
Total Budget	\$315,352

References:

Davis, C. S., S. M. Gallager, and A. R. Solow. 1992a. Microaggregations of oceanic plankton observed by towed video microscopy. *Science*, 257, 230–232.

Davis, C. S., S. M. Gallager, M. S. Berman, L. R. Haury, and J. R. Strickler. 1992b. The Video Plankton Recorder (VPR): Design and initial results. *Arch. Hydrobiol. Beih.*, 36, 67–81.

Davis, C. S., S. M. Gallager, M. Marra, and W. K. Stewart. 1996. Rapid visualization of plankton abundance and taxonomic composition using the Video Plankton Recorder. *Deep Sea Res. II*, 43, 1947–1970.

Davis, C. S., Q. Hu, S. M. Gallager, X. Tang, and C. A. Ashjian. 2004. Real-time observation of taxa-specific plankton distributions: An optical sampling method. *Mar. Ecol. Prog. Ser.*, 284, 77-96.

Davis, C. S., F. Thwaites, S. M. Gallager, and Q. Hu. 2005. A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography. *Limnol. Oceanogr.: Methods*, 3, 59–74.

Davis, C. S. and D. J. McGillicuddy. 2006. Transatlantic abundance of the N2-fixing colonial cyanobacterium *Trichodesmium*, *Science*, 312, 1517 – 1520.

- Herman, A. W. (1992) Design and calibration of a new optical plankton counter capable of sizing small zooplankton. *Deep Sea Research Part A Oceanographic Research Papers* 39:395-415.
- Hu, Q. and C. S. Davis. 2005. Automatic plankton image recognition with co-occurrence matrices and support vector machine. *Mar. Ecol. Prog. Ser.*, 295, 21-31.
- Hu, Q. and C. S. Davis. 2006. Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. *Mar. Ecol. Prog. Ser.*, 306, 51-61.
- Li, W., N. C. Loomis, Q. Hu, and C. S. Davis. 2007. Focus detection from digital in-line holograms based on spectral l_1 norms. *J. Optical Soc. Am.*, 24(10), 3054-3062.
- Loomis, N. J. A. Dominguez-Caballero, W. Li, Q. Hu, C. Davis, J. Milgram, and G. Barbastathis. A compact, low-power digital holographic imaging system for automated plankton taxonomic classification. 4th Intl Zooplankton Symposium, Hiroshima, Japan, May 27-June 1, 2007.
- Sheldon RW, Prakash A, Sutcliffe WH, Jr. (1972) The Size Distribution of Particles in the Ocean. *Limnology and Oceanography* 17:327-340.

Attachments:

- Attachment 1. Holocam Processing Protocol
- Attachment 2. DAVPR and VPRII Processing Protocol

**Deepwater Horizon Oil Spill (DWHOS)
Water Column Technical Working Group**

Image Data Processing Plan: Holocam, DAVPR, VPRII

Principal Investigator: Cabell Davis, WHOI

June 7, 2012

Approvals

Approval of this work plan is for the purposes of processing (including identifying and classifying) digital image data for the Natural Resource Damage Assessment. Parties each reserve its right to produce its own independent interpretation and analysis of any data processed pursuant to this work plan.

BP Approval

Joyce Miley Joyce Miley 6/8/12
Printed Name Signature Date

Federal Trustee Approval

Panel Hahn Panel Hahn 6/7/12
Printed Name Signature Date

Louisiana Approval

KAROLIE H. DEUSCHEN [Signature] 7/6/12
Printed Name Signature Date