To: WHOI Scientific Staff
Cc: WHOI Technical staff
From: Andrew Maffei, Ocean Informatics Working Group Phase I Leader
Date: August 19, 2013
Subject: Report of the Ocean Informatics Working Group (OIWG) and other Ocean Informatics Project Accomplishments and Activities (2009-2013)

WHOI's Ocean Informatics Working Group (OIWG) has achieved its initial core mission of initiating and expanding a basic Ocean Informatics (OI) infrastructure at WHOI. Created in 2009 with seed funding from the Directorate the OIWG, started to help build, fund and train a WHOI OI staff and initiate partnerships with groups like the NSF Biological and Chemical Data Management Office (BCO-DMO), the Gordon and Betty Moore Foundation (GBMF), US Geological Survey (USGS), and the National Marine Fisheries Service (NMFS). Together these collaborators have provided financial support, contributed staff hours to OI-related projects, helped to establish new collaborations focused on science results, and begun to produce a suite of software tools and achieve other accomplishments detailed below (and in Appendix A). The question now is "What Next?". It is now timely to evaluate the progress made and choose a role and priority of Ocean Informatics in the Institution.

The "Coping with your Data" meeting was held on July 26, 2013. We encouraged WHOI scientific and senior technical staff along with staff members at collaborating organizations interested in this topic to join in the definition of the next phase of this effort. We included live demonstrations of the work done to date and asked that participants consider the appropriate priority, scope, and role of WHOI in the emerging cyber-infrastructure for Earth and Ocean sciences (geo-informatics). This meeting acted as a marker of the transition from phase I to phase II of the Ocean Informatics program as the first leadership team (Andrew Maffei, Peter Fox, Art Gaylord, and Jen Schopf) handed the OI effort off to the next team. For OI phase II we envision Stace Beaulieu as the lead, supported by Cyndy Chandler, Joe Futrelle, and Lisa Raymond, as initial members of an ad-hoc advisory team.

Perspective and Outlook

During phase I of this effort we tried to direct the thrust of WHOI's several informatics-related projects to fit WHOI's culture. All our efforts have been strongly aimed at improving scientific outcomes – providing tools and approaches that support a scientist's ability to do better research in their own area of expertise. If an informatics effort cannot be directly linked to better science outcomes, we prefer to abandon it. By applying this attitude on a national/international scale, we see WHOI's role to be one of grounding geoscience informatics/cyber-infrastructure initiatives by demanding that they clearly produce significant science results.

Rather than concentrating directly on the development of "transformative" informatics technologies, we focus on the real and present informatics use-cases/challenges/needs/headaches of the ocean researcher. Efforts focus on the

immediately practical and science-centered actions – preferring the application of existing IT technologies towards *transformative science outcomes* rather than developing new technologies that might themselves be labeled as "transformative".

Somewhat surprisingly, in the process of doing this, we have stumbled upon some informatics approaches/technologies that are novel and applicable to a much broader scientific research community. In retrospect, we attribute these discoveries to the close interactions we have had with research scientists advocating for practical solutions to their immediate needs combined with the broad, inter-disciplinary experience and expertise of our informatics team. Software has been built with the idea to meet the needs of a larger scientific community after first fulfilling the needs of a smaller laboratory group. The methodologies that we have developed, resulting in these broader science informatics discoveries can be highlighted in future national-level initiatives/projects if WHOI scientists choose to go after Big Data solicitations.

With the appearance of the Research Data Alliance (http://www.nsf.gov/news/news_summ.jsp?cntn_id=126010), NSF's EarthCube program (http://www.nsf.gov/geo/earthcube/), and similar private foundation/agency initiatives the OIWG phase I team believes it is now the appropriate time to decide whether to escalate, maintain, or downsize WHOI's efforts towards the emerging field of ocean informatics. There are several possible roles WHOI scientists could choose for the Institution's participation in ocean informatics efforts at varying levels of complexity and commitment. Some of these include:

1. **Become a leader in ocean- and geo-informatics.** Many government agencies are strongly committing to solutions for "Big Data" in the coming years. Given WHOI's accomplishments in the field of informatics over the past few years and strong collaborations with academic partners with IT specialties (detailed below), we would likely be very competitive as a partner in one or more large data-centric solicitations being offered by NSF, NASA, NOAA, and the USGS. This course of action would require active interest and support of the WHOI scientific staff as well as taking advantage of existing and developing new strategic collaborative partnerships.

2. **Lead a Woods Hole area informatics effort.** A smaller scale effort might be to lead in the development of an informatics services organization within the Woods Hole scientific community to provide inter-organizational library, informatics, internet, programming, and other related services. The WHOI OI program could be a strong part of that organization, in collaboration with other Woods Hole area research organizations. Our OI initiative already has funded informatics efforts underway with NOAA NEFSC and USGS (see Appendix A). MBL and Woods Hole Research Center could also be approached.

3. **Focus primarily on Institutional work.** Instead of pushing to be a nationally competitive leader in ocean informatics, we could decide to concentrate only on work needed to support WHOI research data and informatics needs. Stace Beaulieu can lead the effort to "glue" scientific and technical staff into small teams that apply informatics approaches that prove successful. A small core team directs Institution-level efforts in support of direct research needs of WHOI scientists and their research staff in carrying out their work. Ideally, Stace and the informatics staff team members would support themselves through a mix of research grants but some Institutional support would likely be necessary.

4. **Decide that the OIWG has performed its mission** and Institutional support is no longer necessary. This "null role", opting to provide minimal support for an Institution "commons" of data informatics expertise is an option to consider, but it is becoming more and more clear that, with the increasing focus on the importance of proper data handling in society, this approach might be a mistake.

An analysis of the current OIWG efforts in terms of strengths, weaknesses, opportunities, and threats is provided separately in order to assist WHOI staff in forming Institution strategy and making a decision on the next steps forward.

Brief history of the OIWG

In 2008 the WHOI Directorate created a small working group with internal and external participants to examine the needs of WHOI's scientific community in the area of "Ocean Informatics". This was done partly in response to several recommendations of the Data Management and Visualization Subgroup in the 2004 WHOI Access to the Sea report, at addressing the changing needs for more data-centric and collaborative approaches to ocean science. The OIWG started with a series of meetings, activities and proposal efforts organized by myself, Peter Fox (Rensselaer Polytechnic Institute and Adjunct AOPE), Art Gaylord, and Jennifer Schopf (then at NSF Office of Cyber-infrastructure, now at IEEE Computer Society). Further information on the background and strategy of the OIWG is available at: http://www.whoi.edu/DoR/special-projects/ocean-informatics-working-group.

With the support of external grants that derived from OIWG efforts (detailed below), an inter-departmental informatics team is currently working at Caryn House (thanks to the AOPE department). Staff include myself (WHOI/CIS), Joe Futrelle (WHOI/AOPE) and Massimo DiStefano (RPI/TWC) at full-time and Stace Beaulieu (WHOI/BIO) at half-time. We also have part-time participation of a number of staff in other departments and summer students that work in this space off and on.

Summary of OIWG and other Ocean Informatics project accomplishments

The following accomplishments were or are being derived from the projects listed in Appendix A.

- *Established key informatics-related partnerships* with RPI's Tetherless World Constellation (TWC), USGS Woods Hole, NOAA NEFSC, and Lamont Doherty Earth Observatory (in the R2R project). TWC provides the in-kind support of faculty, students, methodologies, software tools, contacts in other organizations and expertise in the realms of computer science and geo-informatics. Many of our accomplishments have been made possible through a strong, mutually beneficial, collaborative relationship with Peter Fox at the Tetherless World Constellation at Rensellaer Polytechnic Institute.

- *Awarded new projects supported by federal agencies and foundations* by establishing strong scientist/informaticist partnering relationships with WHOI scientific staff members with science goals as foci. Our largest projects include the development of improved ocean science research outcomes through an informatics-focused Gordon and Betty Moore Foundation (GBMF) award that has allowed WHOI to hire it's first full-time "ocean informaticist"; in addition, we have obtained two NSF-funded informatics-related grants: (Integrated Ecosystem Assessment Interoperability Initiative (ECOOP) Project funded through NSF/OCI, and Rolling Deck To Repository Program (R2R) funded through NSF/OCE.

- *Produced software tools for the Ocean Sciences community and individual investigators at WHOI.* Tools, at various levels of completion and described in Appendix B. Examples include a modular, web-based underwater image browsing tool that is customizable for different instrument systems, a modularly-designed underwater image annotation tool, the WHOI-developed "Local Data Resolver" approach to making laboratory-scope data more available to research collaborators; facilitating design sessions during development of the NSF Biological and Chemical Oceanography Data Management Office (BCO-DMO) semantically-enable faceted advanced search capability; developed the Imaging Flow CytoBot (IFCB) image browser; Rolling Deck to Repository Shipboard Event Logger; a prototype project vocabulary mapping tool; the physical oceanography data catalog (which we hope to be able to customize for other WHOI departments who are interested in it); a "reproducible science" environment based on the IPython Notebook to facilitate NMFS scientists' ecosystem work – also plans for use of this tool by WHOI scientists; and the intense informatics-related work we have done with BCO-DMO (with Cyndy Chandler's crucial leadership and involvement of BCO-DMO staff) which has led to the development of the "Ocean Data Ontology" by BCO-DMO staff and collaborators, applicable to cataloging and describing a large percentage of oceanographic datasets.

- *Led and participated in community workshops and conferences.* Cyndy Chandler, Heidi Sosik, Peter Fox, Janet Fredericks, I, and other WHOI staff have all been active

and in some cases helped to lead national and inter-national workshops and conference sessions with foci on scientific data and informatics related topics. In the process, we have explored the "milieu" of initiatives such as EarthCube, Research Data Alliance, DataONE, Federation of Earth Science Information Partners and attempted to evaluate the role, value, and funding potential these might have for WHOI scientists in meeting the demands of their data-related challenges. WHOI's presence in these geo-informatics forums has been of a high-quality and a solid reputation has been built.

- *Education and outreach, including WHOI's first summer student fellow in Ocean Informatics.* Eric Rozell's research during summer 2009, working directly with a number of WHOI's ocean scientists and RPI/TWC staff, resulted in the development of S2S (surface-to-seafloor) - a powerful software package/framework that has become a cornerstone for several graduate and undergraduate projects at RPI/TWC. His AGU poster on this topic got the top student prize in the Earth and Space Science Informatics area. Projects in the OIWG "portfolio" have supported other summer guest students in computer-science-related disciplines. We are hopeful that as the Ocean Informatics program grows that more Ocean Informatics summer fellows will be attracted to WHOI's student programs.

- *Provided education and outreach to scientific and technical staff* regarding data management plans and tools for publishing and citing data. This work has been done in collaboration with the WHOI Data Library and Archives.

For a full list of projects and their status, please see Appendix A.

In addition to learning OI related skills and techniques, WHOI staff have had the willingness to try new ways of doing their day-to-day work and shared their opinions on what does and does not work in this first phase of our collaborative efforts. We apologize ahead of time for the people who worked on these projects who we mean to thank but may mistakenly forget to list here: Dicky Allison, Susan Avery, Stace Beaulieu, Massimo DiStefano (TWC), Peter Fox (TWC), Janet Fredericks, Joe Futrelle, David Glover, Scott Gallager, Bob Groman, Jon Hare (NMFS), Alison Kline, Fran Lightsom (USGS), Andrew Maffei, Deborah McGuiness (TWC), Chris Mentzel (GBMF), Carolina Nobre, Lisa Raymond, Cindy Sellers, Heidi Sosik, Adam Shepherd, Laura Stolp, Patrick West (TWC), Norman Vine, Peter Wiebe, Amber York, Stephan Zednick (TWC).

Appendix A – Table of WHOI Informatics Related Projects

The following table presents the informatics-related projects that the Ocean Informatics group and other WHOI staff have been facilitating or actively working on since 2009. Project name, project goals pertinent to WHOI's ocean informatics work, and current status are given.

Project Name and Goals pertinent to WHOI Ocean Informatics	Status
BCO-DMO Ocean Data Ontology (ODO) and advanced search development *(BCODMO_ODO)* - Supported BCO-DMO in their collaborative work with RPI-TWC to create a high-quality ontology as a centerpiece of an informatics infrastructure for oceanographic data that contains concepts and relationships that pertain to a broad range of repositories that contain oceanographic data. - BCO-DMO will offer the Ocean Data Ontology (ODO) to the broader oceanographic community for adoption as a community tool for documenting relationships between oceanographic data held in repositories. - BCO-DMO and R2R are pioneering a new era of oceanographic data access via the development of an "open linked-data network" of oceanographic data using Linked Open Data (LOD).	Funded, by NSF OCE BIO, semantically enabled faceted search has been deployed.
Ocean Informatics Working Group planning *(WHOI_OIWG):* **-** Explore ocean informatics strategies for WHOI researchers and the institution as a whole. - Educate science staff on what Ocean Informatics is - Establish funded informatics–related projects at WHOI that can be used to further demonstrate informatics advantages.	Funded by WHOI internal, "phase 1" complete.
WHOI/RPI-TWC Collaboration *(WHOI_TWC)*: - Maintain close relationship with Peter Fox and his students as WHOI's primary computer/earth science informatics collaborator. - Collaborate on funding opportunities of mutual interest to both WHOI and TWC	MOU signed. Proposal ideas are under active consideration.
Underwater Ocean Imaging Informatics Program *(GBMF_OII)*: - Develop informatics infrastructure components that improve scientific outcomes for projects that employ underwater ocean imaging systems; - Research and develop an approach for effective and productive partnering between science and technical staff initially in Sosik, Gallager, and Singh laboratories and staff informaticists. - Create infrastructure components that track the provenance of data products so that algorithms that were used to process individual images can be documented.	Funded by GBMF, some useful tools already in hand

ECO-OP: Employing Cyber Infrastructure Data Technologies to Facilitate Integrated Ecosystem Assessments for Climate Impacts in NE & CA LME's (#3 & #7) ***(NEFSC_ECOOP)*** - Provide NOAA NMFS researchers with the ability to collaborate and appropriately automate the creation of ecosystem indicators derived from observed data (ocean health indicators are related) - Create infrastructure components that track the provenance of indicators and the ability to publish the interactively executable code and parameters that were used to create these indicators. - Develop an ontology that describes concepts and relationships for the progression and provenance of observed data to indicators.	Funded by NSF OCI, underway. NEFSC partners very pleased with direction this is headed. Journal article in prep.
Rolling Deck to Repository (R2R) Repository ***(NSF_R2R)*** - Strong collaboration with LDEO, SIO, and FSU (LDEO lead) informatics efforts. - Establish central online repository for shipboard underway data collected during all UNOLS research cruises, with aim of improving consistency and quality of these data. - Establish and implement framework to assess the quality of underway-shipboard instrumentation collected on UNOLS vessels and measure quality and consistency over time. - Support the concept of the research fleet being an integrated "virtual/mobile observatory" in the future thru the linked-data concept. - international collaboration with EU and AU on ocean data interoperability	Funded by NSF OCE, already in use. See http:/rvdata.us/ for research cruise data inventories, etc.
R2R Eventlogger Development ***(R2R_EVENTL)*** - Provide scientists with an electronic science event logger available for use on research cruises on UNOLS vessels - Establish a common shipboard webserver on all UNOLS vessels that can be used as a "commons" for R2R quality-assessment software that can be run on ships. - Incorporate the ability to map cruise-specific terminology for instruments, people, cruise-ids, etc. to ocean community scale controlled vocabularies.	Funded by NSF OCE, Version 1.0 of EL deployed on many UNOLS cruises.
R2R Ship Operator Cruise Planning Website ***(R2R_CRUISEP)*** - Port the well-respected WHOI Cruise Planning software to a software platform that UNOLS can offer to other ship operators. - Employ controlled vocabularies and standard webservices to describe and document cruise metadata so that it can be more easily discovered and evaluated for use.	Funded by NSF OCE, Complete and well received. Tech transfer to UNOLS soon.

R2R CTD Data Quality Assessment Software *(R2R_CTDQA)* - Develop software that assess and report on the quality of CTD data collected on UNOLS research vessels and archived at NODC by the R2R project. - Deploy the CTD quality assessment software on the R2R shipboard server in the future	Funded,by NSF OCE. Carolina Nobre (PO) doing this work.
R2R Ocean Data Interoperability Program *(R2R_ODIP)* - Collaborate with EU colleagues on the mapping of community vocabularies (i.e., British Oceanographic Data Center (BODC) vocabularies, employed throughout the EU) to other oceanographic datasets so that cross-repository searches related to oceanographic instrumentation and science parameters can be performed.	Funded by NSF SAVI program. Funds available for students as part of this.
PO Department-level Data Catalog *(OIWG_PODC)* - Build a science-department level data catalog that WHOI Physical Oceanography department scientists, graduate students, and administrative assistants can use to provide minimal dataset-level metadata about datasets that have been collected over time. - Incorporate the BCO-DMO Ocean Data Ontology and its related controlled vocabularies, as part of the PO Data Catalog. - Encourage PO department researchers to refer to the PO Data Catalog as part of their NSF Data Management Plans. - Assuming value of website proves worthwhile, make the data catalog website available to other departments for their own use.	Funded internally, still needs work to reach production level for PO.
RPI/TWC Semantics Methodology with Use Case Development *(WHOI_USECASE)* - Support the development of use cases for informatics-related needs for researchers at WHOI, USGS, NMFS, and others. In software and systems engineering, a use case defines interactions between a role (or "actor") and a system, to achieve a goal. Initial use cases include the development of activity diagrams and concept maps. - Encourage the use of RPI/TWCs Semantics Methodology for those projects that will benefit from it.	Informal, ongoing as needed by science projects that can benefit from it.
USGS Coastal and Marine Spatial Planning (CMSP)-related RPI/TWC Semantic Methodology *(USGS_USECASE)* - Provide regular workshops to the 3 USGS Coastal Marine Centers (Woods Hole, MA., St. Petersburg, FL., Santa Cruz, CA.) to support their ongoing use and interest in the RPI/TWC semantic methodology. - Develop controlled vocabulary and ontology elements that support research in the area of Coastal Marine Spatial Planning (CMSP); Underwater Ocean Imaging; Satellite Imagery; etc.	Funded via USGS/WHOI Cooperative, 4 workshops so far, more planned for winter/spring 2013

Drupal Website Hosting and Support *(OIWG_DRUPAL)* - Host websites for PO Data Catalog, Cruise Planning, and a variety of other researcher and department level websites (PO Promotion, Community Vocabulary Server, workshop websites, program office websites and databases) - Provide researchers and technical staff that wish to develop Drupal sites with a "commons" server.	Ongoing at an "as-time-available" basis. Still determining a "sustainability solution"
Interridge Drupal-7 "Hotvents" Database *(INTERRIDGE_HOTVENT)* - Serve and maintain data previously held in a researcher's Excel spreadsheet more effectively via a content management system - Serve data for machine-machine interactions via a SPARQL endpoint. - Iteratively integrate linked-data capabilities for "columns" as standards and capabilities permit	Originally funded by US Interridge Office. Semantically-enabled website successfully transferred to China office.
NDSF Data Management Subcommittee *(DESSC_DATAM)* - Document existing NDSF vehicle data management products and practices - Identify data management / informatics synergies in existing programs that might be applied to NDSF needs - Report back to DESSC with recommendations of directions for NDSF to explore for improving its data management.	Looking for additional funding. Being done informally at present.
NSF EarthCube initiative *(OIWG_EARTHC)* - Represent WHOI and its informatics efforts in the NSF Earthcube initiative - Remain aware of possible collaborative opportunities, being careful to focus only on initiatives synergistic with WHOI's Ocean Informatics programs – improving ocean science outcomes – grounding informatics to existing needs and not doing informatics for informatics sake. Sometimes simpler "better data management" is the best informatics!!	Background task Maffei, Chandler, Fredericks. Needs direction from WHOI science staff on appropriate level of effort
International Research Data Alliance Initiative *(OIWG_RDA)* - Represent WHOI and its informatics efforts in the international RDA Effort at bi-annual meetings - Co-chair meetings and support creation of the "Marine Data Harmonization Working Group" (led by Helen Glaves, British Geographic Survey) - Play co-leadership role in the RDA "Community and Engagement Working Group". - Stay in contact with US Lead (Fran Berman) at RPI/CS as part of the WHOI/RPI MOU.	Background task for Maffei who co-chairs one of the working groups, needs direction from WHOI science staff on appropriate level of effort

WHOI Data Library & Archive Informatics *(OIWG_DLA)* - collaborate and participate in efforts such as the SCOR/IODE/WHOI-MBL Library Working Group on Data Publication - work with Lisa Raymond to help define a structure for support of emerging informatics services for the WHOI DLA	Lisa Raymond, Cyndy Chandler, Andy Maffei
Information Modeling Services for Science Projects *(DLA_INFMOD)* - work closely with library as it provides information modeling services in support of the development of informatics use cases for Woods Hole Scientific Community science projects	Services being provided to WHOI 'omics researchers and USGS
Marine Metadata Initiative *(NSF_MMI)* - Maffei currently sits on executive committee and attends bi-monthly steering committee meetings. - helping to guide future home for MMI. OOI is currently supporting this effort.	Originally funded by NSF. All expended. complete, v. small effort at present.
Database and Data Access for Microbial Metabolism Products Development *(GBMF_MMPD)* - Provide high-level design/informatics planning support for Liz Kujawinski in WHOI MGG Department	Funded by GBMF. Informatics use case being developed for project
NMFS Use-case Workshop facilitation *(NMFS_USECASE)* - design use-case workshops similar to those offered to USGS (USGS_USECASE)	Pre-proposal stage
Next Collaborative Informatics Proposal? *(OIWG_PROPOSALS)* - Identify scale of next proposals we should submit and whom we should submit them to.	Pre-proposal stage.

Appendix B – Emerging OI Toolbox for Project and Laboratory Level Informatics Partnering

Many people involved in WHOI's Ocean Informatics work have been working on tools and resources that are growing into a toolbox of software, methodologies, and approaches that can hopefully be applied at broader scopes. This is a peek at what some of these tools are:

- **Ocean Data Ontology** – WHOI/NSFs BCO-DMO project has worked closely with the staff at RPI/TWC over the past 3+ years to develop a high quality ontology representing concepts that are important for oceanographic data. The Ocean Data Ontology (ODO) will be announced at the December 2013 AGU meeting. The BCO-DMO advanced search capability can now demonstrate to WHOI researchers the advantages of mapping to community vocabularies and employing ontologies. We plan to use the BCO-DMO example as a way of explaining these advantages to WHOI researchers.

- **RPI/TWC Use-case Development Methodology** – Researchers and technical staff at WHOI and partnering organizations involved in several of the OI projects described above have been trained by TWC and WHOI staff in the use-case centered methodology taught as a core of the RPI/TWC graduate program. It has proven itself to be very effective in identifying and solving real problems of scientists and provides a common language and set of tools for these staff to employ in the informatics arena. New projects are begun by customizing the methodology to meet the unique needs of a project/science laboratory, keeping its core principles intact.

- **Local Data Resolver** – Joe Futrelle has developed a "middleware" package so that scientific code can be more easily written to access scientific data via URLs rather than static files system based names. The package is designed to eventually support data sitting on scientist's individual laboratory data servers, the WHOI FTP Server, Dropbox, larger science project servers, and even larger discipline-specific repositories. This is part of the "actionable URL" approach described below

- **Redmine Lightweight Project Management** – Early in the Ocean Imaging Informatics project Joe Futrelle did a survey to identify the "best" open-source project collaboration software. After some research he found Redmine. It is lightweight and has many useful features. Several projects we work on have started to use it for planning cruises and other activities – without our prompting. We think Redmine is a "keeper".

- **Science Source Code Repositories**– Using source code repositories and software versioning is a good (some say essential) software engineering practice but many of us think it is too cumbersome. Joe Futrelle and Massimo

DiStafano have been finding effective ways to entrain science and technical staff to start using it more in their day-to-day work by demonstrating its advantages to their science. More work to do here, but we are hopeful.

- **Openstack Environment for Science Lab VMs –** We have identified and customized the openstack virtual machine (VM) environment to support the practice of snapshotting the operating system, applications software, and custom code developed for individual research laboratories. Once the code has been adapted to employ web-based technologies such as those described below (actionable URLS, local data resolver, etc.) multiple copies of these VMs can sometimes be deployed so that science data can be processed in parallel. VMs can also be archived as "provenance artifacts" when documenting provenance of derived data products in a science project.

- **Actionable URL Coding** – Science projects that collect data oftentimes change the directory hierarchy, file naming conventions, and type of metadata collected from year to year, expedition to expedition. Software then must be written and re-written to accommodate these data sitting in a variety of locations. The ability to write data analysis software that accesses all the data and metadata from many years of expeditions via URLs using a well-structured URL syntax is a part of the "actionable URL" approach we have demonstrated in the GBMF-funded work. Over the next year WHOI scientists will be giving talks in an informal seminar series explaining how they can use this new programming technique to improve their lab productivity, improve science results and solve several headaches they have had to deal with in the past. *Important Note:* Data and metadata accessed by data analysis code via a URL scheme can still be resolved to a local file system, thus avoiding any performance degradation that would occur if large amounts of data had to be transported over a network.

- **Customizable Ocean Image Browser** – Once science data and metadata are available via actionable URLs, community tools can be built to act on those data and metadata. The Ocean Imaging and Informatics project has built (and then re-built in a more generic fashion) a modular image data browser that can be customized for various underwater ocean imagery datasets. Our hope is that someday data browsers for other types of data and metadata served via actionable URLs will be provided in this toolbox.

- **Customizable Ocean Image Manual Annotator -** This is a tool and related services that are created with a similar focus and approach as the above imagery data browser. We are hopeful that once data and metadata from other science-disciplines are available via the actionable URL approach that the architecture employed for the image manual annotation tool can be used to build annotators for other other source data and data products.

- **Workflow Webservices to Accelerate Ocean Image Data Analysis** – Once scientific data and metadata are available via actionable URLs and scientists have re-written their code to use them, the scientists then have the option to deploy several copies of their code on different processors on the network. The problem is then how to feed data and metadata to the many instances of the software in an organized fashion. The solution is to implement a system that you can direct – giving it a queue of data and metadata that can be parceled out to computers that are ready to take on the task. Joe Futrelle has demonstrated such an "Advanced Message Queuing Protocol (AMQP) messaging system" that does this. Once again, we plan to have ocean imaging informatics scientists in an AOPE seminar talk about how this capability has in some cases improved processing speed of their data 10-20 times. Such a capability could be developed for other science disciplines at WHOI in the future.

- **Laboratory-scale Informatics Partnering Methodology** – Another thing that Joe Futrelle and the Ocean Informatics "team" do after meetings with scientists and technical staff is to discuss what works and does not work when computer science "types" try to meet the needs of scientists – how best to form partnerships that work as well as instrument engineer – scientist partnerships have worked in the past at WHOI. The patterns for successful interactions have begun to merge these with the RPI/TWC semantic methodology to create a set of progressive benefits (many provided via the tools above) to scientists as such a partnership evolves. We are hopeful that this can be the core of a future proposal addressing the "long tail of science problem" that informaticists and computer scientists have been talking about for a long time.

- **IPython for Visualizing Provenance and Publishing Code –** the ECOOP project has identified IPython as a powerful mechanism for publishing code in addition to derived data products for a science project. Code can be published both statically and as an interactive environment. We are also looking at the possibility of using IPython to more clearly document the provenance of data products produced from observed data in other science projects such as the underwater imaging project.