PIDapalooza 2016 Session Description

Title: Persistent Identifiers and Linked Ocean Data

Authors: Cynthia L. Chandler[1], Adam Shepherd[1], Robert A. Arko[2] and Matthew B. Jones[3]

1. Woods Hole Oceanographic Institution, Woods Hole, MA USA
2. Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY USA
3. National Center for Ecological Analysis and Synthesis (NCEAS), University of California Santa Barbara, CA USA

Marine ecosystem science requires access to a wide variety of data. Research topics include investigation of complex food webs, sustainable fisheries science, coastal and marine ecosystem studies that contribute to better informed management decisions, and climate change vulnerability and potential mitigation. Such research requires access to data from the natural, social and economic sciences and often large distributed teams of investigators with expertise in a broad range of disciplines. As data are made more freely available and discoverable on the Web, it is likely that data are being used and re-used far from their point of origin. That distance defined in time, space and research domain increases the need for robust metadata (e.g. structured documentation) to provide proper context for the data.

In this session three speakers, Chandler, Shepherd and Arko, will talk about work being done at several complementary marine research data repositories in the USA. Over the past few years, a variety of different Persistent Identifiers (PIDs) have been added to the metadata catalogs they support. Use of Open Researcher and Contributor ID (ORCiD) and Digital Object Identifiers (DOIs) have shown great potential to disambiguate persons and research objects respectively. We have made extensive use of DOIs for a variety of research objects including published datasets and documents, but also for repositories (e.g. re3data.org codes), expeditions (e.g. research cruises), and research networks (e.g. IRIS sensor networks for seismological data). In addition Open Funder Registry (FundRef) codes are used to identify funding sources, Global Research Identifier Database (GRID) identifiers help to disambiguate institutional affiliations, and International Geo Sample Numbers (IGSNs) help to track the provenance of physical samples. The PIDs are integrated into the metadata records of the repository catalogs, published out as Linked Open Data and enhanced with semantic markup to provide additional, critical context for improved discovery of resources of interest.

In this session we will review progress made thus far, reflecting on what has worked or failed, and thoughts on logical next steps. Use of PIDs has resulted in the ability to validate metadata records by comparing the contents with trusted, authoritative sources. The PIDs also serve as reliable facts in Linked Open Data resources that can be connected with related research results, e.g. published data sets with literature publications.