Member Node Description: PISCO

Version 1.2 07/03/15 Kristen Milligan, Michael Frenock

General

Name of resource:	Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO)
URL(s):	data.piscoweb.org
Institutional affiliation(s):	University of California at Santa Barbara (UCSB), University of California at Santa Cruz (UCSC), Stanford University at Hopkins Marine Station (HMS), Oregon State University (OSU)
Primary geographic location:	Corvallis, Oregon, USA
Project Director & contact info:	Kristen Milligan Department of Zoology Cordley Hall 3029 Oregon State University Corvallis, OR 97331-2914 Phone: 541-737-8862 Email: milligak@science.oregonstate.edu
Technical Contact & contact info:	Michael Frenock Department of Zoology Cordley Hall 3029 Oregon State University Corvallis, OR 97331-2914 Email: frenockm@science.oregonstate.edu
Age of resource:	1999
Funding support:	PISCO was established in 1999 with major grants from The David and Lucile Packard Foundation. Beginning in 2005, the program was funded by 5-year collaborative grants from The David and Lucile Packard Foundation and the Gordon and Betty Moore Foundation. Starting in 2010, major operational support is from the David and Lucile Packard Foundation. This core support plus additional funding from diverse public (local, state, and federal including National Science Foundation) and private sources make the unique, multi-institutional partnership possible.
Proposed Unique Identifier:	urn:node:PISCO

Content

Content description/collection policy (1 paragraph, domain and spatial/temporal coverage, uniqueness of content, exclusions, as applicable):

PISCO's research extends more than 1,200 miles (2,000 kilometers) along the west coast of North America**.** Most research is conducted within 10 miles (16 km) of shore and focuses on three important components of the marine ecosystem: kelp forests, rocky shores, and coastal currents. PISCO researchers collect biological, chemical, and physical data about ocean ecosystems in the nearshore portions of the California Current Large Marine Ecosystem.

Types of data (complex objects, text, image, video, audio, other):
Data is stored as ASCII text, either space or comma delimited. Metadata is stored using the Ecological Metadata Language (EML) specification implemented by the Knowledge Network for Biodiversity (KNB).

Data and metadata availability (rights, licensing, restrictions):

DATA SHARING POLICY

GENERAL POLICIES

1. Datasets will be uploaded to the data catalog for availability within PISCO within one year of collection.
2. Full documentation (metadata) will be developed for each dataset and will be publicly available within one year of collection.
3. All data will be available to the public via the data catalog within two years of collection, or at the time of publication of the main findings of the project, whichever comes first.

EXCEPTIONS AND USE POLICIES

There will be instances where sensitive information cannot be shared prior to publication or where data summaries rather than raw datasets are more appropriate for sharing. PISCO has a formal internal process for reviewing such exceptions. Use of the data and how to appropriately acknowledge PISCO is specified in the Usage Rights section of the metadata for each dataset.

Option for embargo (yes/no, duration):
No.

Size of holdings (number and size of datasets, mean and median granules (files) per dataset):
As of June, 2015, PISCO has 21,000+ datasets in the data catalog. Physical Oceanography datasets typically contain one data file per dataset. Biological datasets typically contain 2 to 5 files per dataset. New datasets continue to be added.

Please describe recent usage statistics, if known, including information on annual data product downloads, annual number of users, annual number of data products used in publications:

usage (2014):

- 41331 PISCO metacat reads by 847 different IP addresses in 43 different countries

Publications:

- As of June 2015: 448 contributions published in peer reviewed literature (since 1999).

User interactions

How does a user contribute data? (what can be deposited, how are data prepared, are specific software required, documentation/support available)
Data collections to be loaded into the PISCO data catalog are produced by the PISCO consortium. Prior to uploading data to the PISCO data catalog, each lab quality-controls the data according to agreed-upon methods which are documented in the metadata. Data is typically loaded into the catalog using the catalog software api via java or Matlab programs written by PISCO. Data may also be loaded via desktop tools.

How does a user acquire / access data?

Data is accessed and downloaded via the following PISCO web-based data portal applications:

Main data portal (product or phrase based):
http://data.piscoweb.org/DataCatalogAccess/DataCatalogAccess.html.

Data access by site:
http://data.piscoweb.org/PiscoSiteDataAccess/PiscoSiteDataAccess.html

Interactive map and graphing tool for biodiversity data:
http://data.piscoweb.org/marine1/intertidalmap.html

Interactive map and graphing tool for sea star wasting disease extent and severity:
http://data.piscoweb.org/marine1/seastardisease.html

What user support services are available (both for depositing and accessing/using data)?

PISCO's web-based applications contain 'How to use' links on each page. The main data portal also has a folder containing answers to frequently asked questions about the application and data catalog. Questions may be asked or issues reported at any time via email at webcontact@piscoweb.org.

How does the resource curate data at the time of deposit?

Data is curated at the time of deposit by the person uploading it.

Technical characteristics and policies

Software platform description, incl. data search and access API(s):

The PISCO data catalog uses the Metacat metadata and data management server from the Knowledge Network for Biodiversity (see http://knb.ecoinformatics.org/knb/docs/) on a Linux server running Apache Server and Apache Tomcat. The data access applications are written in java using Google Web Toolkit and SmartGWT. The biodiversity mapping and graphing application is written in javascript using jquery and accessing a web service written in jruby.

Service reliability (including recent uptime statistics, frequency of hardware refresh, if known):

As of July 2015, the servers hosting the data catalog databases and software have currently been up for 17 days with two unanticipated downtimes in the last few years which were resolved within an hour.

Preservation reliability (including replication/backup, integrity checks, format migration, disaster planning):

The servers hosting the data catalog databases and software are currently administered by the College of Science Information Network team (COSINe) at Oregon State University. Policies and procedures can be accessed here: http://my.science.oregonstate.edu/policies. Full backups of the servers are done monthly with incremental daily backups using bacula.

A local replicate of the data catalog is replicated to as data is added to the catalog. The data is also replicated to the Knowledge Network for Biodiversity (KNB) Metacat instance.

User authentication technology (incl. level of create/modify/delete access by users):

Users are currently authenticated via LDAP through the Metacat api. Most PISCO datasets can be read by public. Only specific groups of PISCO personnel have create/modify/delete access.

Data identifier system and data citation policy, if available:

Document Object Identifiers via EZID; shoulder: doi:10.6085/AA/

Data Citation Policy:

We ask that you contact PISCO prior to using this information for any purpose.
We make this request to:

1. Reduce redundancy; we may be currently working on projects that involve this information.
2. We would like to be informed of and involved in projects developed using this information.

Please cite PISCO in all publications containing these data. The citation should take the form: "This study utilized data collected by the Partnership for Interdisciplinary Studies of Coastal Oceans." Funders of the data collection, if listed in the metadata, must also be included in your citation.

Metadata standards (including provenance):
Ecological Metadata Language (EML). See http://knb.ecoinformatics.org/software/eml/

Capacity/services to DataONE

At what functional tier will you initially be operating? (see http://bit.ly/MNFactSheet for definitions)

- ☐ **Tier 1: Read only, public content**
- ☐ **Tier 2: Read only with access control**
- ☑ **Tier 3: Read/write using client tools**
- ☐ **Tier 4: Able to operate as a replication target**

If you can host data from other member nodes, what storage capacity is available?
Possibly; depends on future funding

Can you provide computing capacity to the broader network? If so, please describe.
Possibly; depends on future funding

Other Services

What other services or resources (such as expertise, software development capacity, educational/training resources, or software tools) can be provided of benefit to the broader network?
Unknown at this time.