CATALOGING OF NODC ARCHIVES
ON THE NASA OCEAN DATA SYSTEM

Steven L. Patterson
and
Michael T. Simmons

National Oceanographic Data Center
2001 Wisconsin Avenue, N.W.
Washington, D.C. 20235

ABSTRACT

The efficient management of rapidly increasing quantities of satellite-derived oceanographic data and related in-situ data poses a major challenge for the immediate future. In an effort to respond to this challenge the NASA Ocean Data System (NODS) was developed at the Jet Propulsion Laboratory. NODS is a computer-based on-line data information system which provides a data catalog, sample browse files, and data acquisition capabilities. The Global On-Line Data (GOLD) catalog of NODS provides a user with the ability to identify data sets which meet his specified search criteria. NODS is being linked via a telecommunications network with various remote archives including the National Oceanographic Data Center (NODC). In order to evaluate the GOLD catalog's utility for in-situ data, NODC elected to initially load information about its archive of drifting buoy data. This effort identified some problems which led to recommendations to improve the GOLD catalog design.

INTRODUCTION

The efficient management of rapidly increasing quantities of satellite-derived oceanographic data and related in-situ data poses a major challenge for the immediate future. In the past, too little consideration has been given to the process of adequately documenting, inventorying, and archiving data sets in a manner that will make them readily accessible and useful to secondary users. Ocean researchers frequently encounter difficulty in determining what relevant historical data are available, and then are obliged to expend considerable time and resources to obtain the data. Once the data are assembled, the researcher may find that documentation on calibration, processing techniques, and quality assessment is either incomplete or absent.

Recent advances in the data-collecting capability of satellites have now made it possible for the oceanographic community to plan and conduct research on a global scale.^{1,2} Such efforts, however, require the cooperation of many scientists and government agencies. This mutual interdependence has the effect of making all of the involved parties secondary data users, and high-

lights the need for efficient and effective data management policies and procedures. In an effort to respond to this challenge the Jet Propulsion Laboratory (JPL) in Pasadena, California, initiated the development of the NASA Ocean Data System (NODS).

DESCRIPTION OF NODS
AND THE GOLD CATALOG

NODS is a computer-based on-line information system which provides a data catalog, sample browse files, bibliographical references, and data extraction capabilities. The Global On-Line Data (GOLD) catalog of NODS^{3,4} provides a user with the ability to identify data sets within the catalog which meet his needs. From his remote terminal a user can log onto the GOLD catalog and specify selection criteria based on platform type, measurement, location, or time. For any data set that meets his criteria, the user may obtain detailed information about the data including where the data are archived and their accessibility, the identification of the principal investigator or other cognizant individual, a description of the platform and sensor with which the data were collected, the spatial and temporal coverage of the data set, a description of the data processing steps employed, an assessment of the quality of the measured parameter including the temporal and/or spatial resolution of the measurements, a description of any available data products, information about browse files, and a list of bibliographic references.

The GOLD catalog was originally designed to accommodate information about the satellite data sets resident at JPL and Scripps Institution of Oceanography. These included data obtained by SEASAT, TIROS-N, NOAA6, NOAA7, NOAA8, and NIMBUS-7. In order to incorporate other oceanographic data sets, including non-satellite data, NODS is being implemented as a distributed archive system. JPL is being linked via the Space Physics Analysis Network (SPAN) with various remote archives including the University of Miami, the University of Rhode Island, the University of Delaware, the University of Alaska, the National Snow and Ice Data Center in Boulder, Colorado and the National Oceanographic Data Center in Washington, D.C. A current limitation of this system is that it only links VAX computers manu-

factured by Digital Equipment Corporation (DEC). However, since most oceanographic research facilities have VAX computers, this was not deemed to be a serious restriction.

The GOLD catalog software has been installed at several remote archives and efforts are underway at these sites to load information about their data holdings into the catalog. There are two major benefits to distributing the catalog. First, increasing the number and diversity of data sets listed will make the catalog a much more valuable resource for the research community. Second, the effort to incorporate diverse types of oceanographic data reveals inadequacies in the present software design which, when modified, will improve the system functionality and performance. Therefore, the present GOLD catalog is considered to be a prototype system whose design will be upgraded as deficiencies are identified and corrected with the aid of the ocean research community.

DRIFTING BUOY DATA AT THE NATIONAL OCEANOGRAPHIC DATA CENTER

The National Oceanographic Data Center (NODC) archives many diverse types of physical, chemical, and biological data from the deep ocean and the shallow coastal zone and estuaries.⁵ Since in-situ data at the sea surface provide "ground truth" for parameters that are remotely-sensed by satellites, NODC has elected to initially load information about its archive of surface drifting buoy data into the GOLD catalog. At this writing NODC's archive of drifting buoy data consists of over 600 trajectories (Table 1) and is growing steadily because of NOAA's ongoing TOGA (Tropical Ocean Global Atmosphere) Project. The data span the period from 1975 to the present and come from virtually all parts of the world ocean. Most of the drifting buoys were tracked by satellites. Their trajectories provide information about ocean surface currents. Although the other parameters measured by the drifting buoys vary because of differing project objectives, the following parameters are represented in the archive: sea surface temperature, sea level pressure, air temperature, wind speed, wind direction, and subsurface temperature profile.

ASSESSMENT OF THE GOLD CATALOG

Not surprisingly, when NODC attempted to load information about drifting buoy data into the GOLD catalog, some problems emerged. These were reported at a workshop convened at NODC on 6-7 May, 1986. The workshop was arranged by JPL to provide a forum in which representatives from the various sites attempting to load the catalog could relate their experiences and suggest improvements to be implemented. The major deficiencies reported at the workshop are summarized below.

The first problem relates to searches of the catalog based on location and time. Within the GOLD catalog, information about the spatial coverage of each data set is stored and searched separately

from the information about its temporal duration. Data from moving platforms, such as drifting buoys and satellites, are functions of both space and time. In such cases, searches of the catalog based on location and time could incorrectly identify data sets as having satisfied the search criteria.

The second problem involves the definition of a "data set". The GOLD catalog considers each platform-measurement combination to be a separate data set. This definition is important because 1) each data set is accompanied by extensive documentation, and 2) the result of any search of the catalog is an identification of the data sets which meet the search criteria. For non-satellite data such as those archived at NODC, this definition of a data set poses some difficulties. Consider a drifting buoy that provides a time-series of position, sea surface temperature, and sea level pressure. The catalog considers this to be three data sets. Consequently, the drifting buoy archive alone would soon contribute several thousand data sets to the catalog with much redundant documentation. Similar arguments apply to the other data types archived by NODC. It was concluded by the workshop that for the purposes of the GOLD catalog, the entire drifting buoy archive should be considered a single data set. Hence, the workshop recommended that the definition of a data set allow for multiple platforms with multiple sensors.

Another issue involves the control of acronyms. Each platform type, sensor name, parameter name, etc., is designated by an acronym up to eight characters in length. Although a list of acceptable acronyms resides within the GOLD catalog software, new acronyms may be added with relatively little central control. Without such control it is possible that several acronyms with the same or very similar meaning might be introduced. Potentially this is a serious problem, since getting the desired result from a search requires a judicious specification of search criteria.

Each archive site has exclusive responsibility and capability for maintaining the catalog information about its own holdings. As each site updates its own copy of the GOLD catalog, a record of these transactions (called an audit trail) is automatically compiled. It is planned that JPL will periodically collect these audit trails from the various sites and prepare a revised copy of GOLD for distribution. In this way, all of the participating archives will have a periodically updated copy of the comprehensive catalog. This would allow a user to log onto his favorite node (archive) and be able to perform quick interactive searches of the holdings of all participating archives. However, in order to be sure that he is obtaining the most recent information from a particular archive, he would be obliged to log onto that archive's local copy of the GOLD catalog. While this seems to be a reasonable plan for maintaining uniformity of the catalog throughout the system, at the time of the May workshop JPL had not yet designed the software required to use the audit trails to update the catalog.

Table 1. NODC Surface Drifting Buoy Data
(as of 20 May 1986)

<u>Project</u>	<u>Institution</u>	<u>Investigator</u>	<u>Trajectories</u>	<u>Region</u>	<u>Period</u>
Outer Continental Shelf Environmental Assessment Project (OCSEAP)	Univ. of Washington	P. Martin	67	Beaufort Sea	1975-77
OCSEAP	Atlantic Oceanographic and Meteorological Laboratory (AOML)	D.V. Hansen	20	Gulf of Alaska	1975-78
OCSEAP	Science Applications International Corporation (SAIC)	P.R. Temple	6	Gulf of Alaska	1980-81
OCSEAP	Flow Research	Pritchard	7	Norton Sound	1981
Ocean Dumping	Texas A & M University	J.C. Mungall	1	Gulf of Mexico	1977
North Pacific Experiment (NORPAX)	AOML	D.V. Hansen	22	Equatorial Pacific	1977-78
NORPAX	Scripps Institution of Oceanography	W.C. Patzert	57	Equatorial Pacific	1979-80
Ocean Thermal Energy Conversion (OTEC)	AOML	R. Molinari	7	Equatorial Atlantic	1978-79
First GARP Global Experiment (FGGE)	Institute of Ocean Sciences	J.F. Garrett	359	Southern Hemisphere	1978-81
Marginal Ice Zone Experiment (MIZEX)	Pacific Marine Environmental Laboratory (PMEL)	C.H. Pease	21	Bering Sea	1981-83
Minerals Management Service/Santa Barbara Channel	SAIC	J. Karpen	3	Santa Barbara Channel	1984
Tropical Ocean Global Atmosphere Project (TOGA):	National Data Buoy Center (NDBC)				
S. Africa		P. Leroux	12	S. Atlantic	1984-86
U.S.A.		R. Partridge	39	Southern Hemisphere	1984-86
Australia		P. Shaw	9	S. Indian	1985-86
New Zealand		R. Munster	13	S. Pacific	1985-86
U.K.		R. Partridge	5	S. Atlantic	1985-86
Arctic Polynya Experiment (APEX)	PMEL	C.H. Pease	12	Bering Sea	1985

CONCLUSION

Considering the anticipated increase in the size and number of future data sets, the need for improved methods of data management is acute. The concept of a diversified and distributed oceanographic data archive with a common on-line

catalog, such as the prototype being developed by JPL, is a very worthwhile undertaking. Although the evaluation process has identified some problems with the system, these are being addressed. The evolution of this prototype into an effective operational system will require a continued cooperative effort.

REFERENCES

1. U.S. WOCE Science Steering Committee: World Ocean Circulation Experiment, U.S. WOCE Planning Report Number 3, January 1986, 229 pp.
2. Global Ocean Flux Study Committee: Global Ocean Flux Study, Proceedings of a workshop, September 10-14, 1984, National Academy of Sciences Woods Hole Study Center, Woods Hole, Massachusetts, National Academy Press, 360 pp.
3. Bonbright, D.I., J.W. Brown, J.E. Hilland, I.T. Hsu, J.A. Johnson, T.L. Kotlarek, R.A. Lassanyi, C.L. Miller, C.S. Morris, and F.J. Salamone: NASA Ocean Data System User Handbook, Volumes 1 and 2, JPL D-61, Jet Propulsion Laboratory, 1 February 1986, 272 pp.
4. Johnson, J.A.: Getting into PODS with GOLD, Information Systems Newsletter, NASA Office of Science and Applications, June 1985, pg. 8-9 and 12.
5. National Oceanographic Data Center: National Oceanographic Data Center Users Guide, March 1984, 250 pp.