MANAGING MARINE POLLUTION DATA AND INFORMATION

Kent H. Hughes

National Oceanographic Data Center

ABSTRACT

An Ocean Pollution Data and Information Network (OPDIN), composed of a central office in Washington, D.C. and five regional offices in key coastal locations, has been proposed to ensure timely and useful availability of marine pollution data and information. The National Oceanographic Data Center recently convened the Workshop on Marine Pollution Information Management as a first step in defining system requirements and to initialize OPDIN design. Important features of the Network include regional access (for input and distribution of data and information), a system of cost recovery, and a means of facilitating data synthesis. The use of existing resources to the maximum extent is emphasized. The publication, Coastal Ocean Pollution Assessment (COPAS) News, is now available for data and information dissemination.

BACKGROUND

In late 1978, the Interagency Committee on Ocean Pollution Research and Development and Monitoring tasked its subordinate group, the Subcommittee on Ocean Pollution Data Collection, Storage, and Distribution, to explore problems associated with the handling of marine pollution data and information. The Subcommittee was chaired by Dr. H. William Menard, Director of the U.S. Geological Survey, U.S. Department of the Interior. The Subcommittee recognized the acute need for adequate data and information services to support ocean pollution research, development and monitoring and examined the current state of such services within the Federal structure. The report of the Subcommittee reemphasized the mandate of the National Ocean Pollution Planning Act of 1978 (PL95-273), that proposals should use existing facilities to the greatest extent possible. Recommendations were focused largely on the needs of secondary data users, those whose use of environmental data and information is subsequent to the use for which the data were originally intended, or collected.

Invited experts in scientific data and information management and marine pollution data collection and use gathered January 13-15, 1981, in Frederick, Maryland, to participate in a Workshop on Marine Pollution Information Management. The Workshop was sponsored by the Environmental Data and Information Service (EDIS) of the National Oceanic and Atmospheric Administration (NOAA) and organized by the National Oceanographic Data Center (NODC), the EDIS center that has been delegated responsibility for implementing Section 8 (Information Dissemination) of PL95-273. NODC convened the Workshop to enable representatives of academia, private industry, public interest groups, and other government bodies and agencies to review and comment on proposals for fulfilling the mandate of the Act.

Section 8 of the Act presents its charge in one sentence of bold generality: "The Administrator (of NOAA) shall ensure that results, findings, and information regarding ocean pollution research and development and monitoring programs conducted or sponsored by the Federal Government be disseminated in a timely manner, and in useful forms, to relevant departments, agencies, and instrumentalities of the Federal Government, and to other persons having an interest in ocean pollution research and development and monitoring." Central to the Workshop discussions were the questions...What is timely? and What is useful?...in regard to marine pollution data and information products and services.

A point of common reference was the proposed Ocean Pollution Data and Information Network (OPDIN) as presented in the first five-year Federal Plan for Ocean Pollution Research Development and Monitoring. OPDIN is an institutional arrangement using largely existing data processing facilities organized to expedite the orderly flow of marine pollution data and information (Figure 1). The OPDIN will be initially composed of the Central Coordination and Referral Office (CCRO--an office within the National Oceanographic Data Center) and Regional Coordination and Referral Offices (RCRO--the NODC Liaison Officers in Anchorage, Seattle, La Jolla, Miami, and Woods Hole).
The Workshop was designed such that each participant was assigned to one of three panels: Digital Data, Data Synthesis, and Information Management. The three panels met separately, but each discussed all three of these broad subject areas. Background information for the panels' deliberations was provided by issue papers developed and presented by panel leadership. This Workshop format allowed the discussion groups to be kept to workable size and allowed problems to be viewed from different professional perspectives.

The main conclusions and recommendations of each of the panels' deliberations are briefly summarized:

**GENERAL RECOMMENDATIONS**

- The Ocean Pollution Data and Information Network (OPDIN) was endorsed as a mechanism to facilitate the timely and useful acquisition and dissemination of marine pollution data and information.
- A reasonable system of cost recovery and international exchange should be built into Network operations.

**Digital Data**

- Regional (or problem area where appropriate) approaches to data accession and access were strongly emphasized (Figure 1). Establishment of Regional Coordination and Referral Offices (RCROs) was recommended.
- Reluctance to submit data to data centers can be reduced by stressing cooperation. Support to data collectors should not be overlooked: use the carrot before the stick.
- The Ocean Pollution Data and Information Network should emphasize the use of available data management tools including standard formats, codes, quality assurance procedures, and output products.
- Remote access to the archive should be investigated.
- Complete documentation must be included with data sets submitted for archival. Quality assurance statements must accompany data and information.

**Data Synthesis**

- Synthesis is a process involving the selection, acquisition, merging, manipulation, and interpretation of available relevant data and information. At least three levels, or categories of synthesis can be identified: 1) data and information summaries; 2) information products (e.g., maps or charts); and 3) analyses requiring scientific expertise.
- Users should be able to interact with OPDIN functions at whatever level or activity appropriate.
- Many syntheses will be referred to external sources for action. Lists of synthesizers should be maintained by the Central Coordination and Referral Office (CCRO). Basic Ordering Agreements, or other prearranged contractual arrangements should be explored for facilitating selection of vendors.

**Information Management**

- OPDIN should publicize its activities as widely as possible.
- Special emphasis should be placed by OPDIN on making relevant gray literature available at the local level--perhaps through a remote access bibliographic search capability, or through special collections of literature held locally.
- Insofar as possible, data and information products should be prepared in an anticipatory mode, e.g., special studies of high priority pollutants should be commissioned. Information should be prepared describing ongoing or planned data collection projects.
The Ocean Pollution Data and Information Network as originally proposed by the Interagency Committee on Ocean Pollution Research, Development and Monitoring, is intended to supplement existing agency information and data services, not replace them. Project data processing is still a responsibility of project management.

FEATURES OF OPDIN

- OPDIN should sponsor in-depth analyses of critical pollution problems using existing data and information resources. These reports should examine particular pollution-related problems, either a specific pollutant or a class of pollutants (e.g., PCBs), pollutant sources (e.g., municipal sewage outfalls), or pollution impact problems (e.g., altered abundance of certain species).

- A system of reasonable cost recovery must be built into the system:
  - a) Recovery is expected of all or some percentage of the cost of products and services.
  - b) A fee structure should be established and adjusted as products and services become available.
  - c) Special arrangements should be made with users who are also data contributors via cooperative agreements or other means.
  - d) Special charges should be established for guaranteed rapid delivery.

- International data exchange should be encouraged per NOAA guidelines with those countries where data exchange agreements exist, or where programs are relevant.

Digital Data

Central, automated digital data systems for the archival of environmental data are well developed in several Federal agencies, e.g., NOAA's EDIS, EPA's STORET and USGS' WATSTORE. In 1977 the Chairman of the Council on Environmental Quality, pursuant to the President's 1977 environmental message, established an interagency task force on environmental data and monitoring. A subgroup of that body concluded that: most users require digital data from a single medium and are satisfied by single systems; that existing Federal data systems operate successfully in meeting the needs of single Government agencies and single discipline users; but that cross-disciplinary analyses are becoming increasingly important for decisionmaking. It was concluded that mechanisms for providing data support and analyses for cross-disciplinary purposes in a timely fashion are presently inadequate.

The Workshop on Marine Pollution Information Management recognized that environmental impact analysis depends heavily on the collection, archiving, analysis, and dissemination of digital data on marine pollutants and related parameters. These analyses are usually produced using data and information, which were collected for other purposes which then become available for secondary uses. A major concern was how best to obtain data from the data collector, recognizing the reluctance on the part of some data collectors for releasing data to data centers. In general, it was agreed that data collectors, who are generally single-discipline scientists, have different interests and points of view from data centers. Data centers tend to serve primarily the needs of the secondary data user community. A prevailing viewpoint is that the threshold of reluctance toward data submission could be reduced by emphasizing products and services (e.g., data entry and/or reformatting assistance) offered by data centers to the data collector.

Many users of marine environmental data comment on difficulties encountered in submitting or receiving oceanographic data. Many of these problems can be attributed to the distances between the data collectors, usually located in Washington, D.C., and the customer (submitter or requester). It was recommended that Regional Coordination and Referral Offices (RCROs) be established to better span the gap between suppliers and users of data. The proposed RCRO-CCRO Network will be in a better position, to deal with local, or regional marine pollution problems.

It is recommended that OPDIN (CCRO and RCROs) stress the development of and availability of data management tools and procedures to facilitate interaction among data collectors, processors, archivists and users. Standard formats, data exchange procedures, adequate documentation, biological and chemical nomenclatural codes, quality assurance and quality control procedures, and data product and analysis techniques all fit this category.

It is essential that inventories of important pollution related parameters, not held by central data centers, be maintained by the Central Coordination and Referral Office (CCRO). Relevant non-digitized data sets should be identified for conversion to the archives of the central data centers. It was further recognized that because of the diverse nature of marine pollution data, OPDIN must limit its scope to those data types which are truly marine pollution related in order that a significant contribution can be made with existing resources.

In order to support the regional structure and timely response to user requests, OPDIN should be designed and operated to offer remote and, if possible, interactive access to inventories, data bases and supportive code systems.
The Workshop on Marine Pollution Information Management posed several questions to be of concern to the operations of OPDIN:

**Data Collection:** Who is collecting digital marine pollution data? What data are currently available (which can be digitized) to assist decisionmakers and other users of digital data?

**Data Entry:** What are existing procedures, costs, and time requirements for converting data to digital form and perhaps to NODC-compatible formats?

**Data Processing:** For data that are in digital form, what kind of consistency in input formats is available among Federal agencies and other marine pollution data generators? What definitions of precision for marine pollution parameters have been specified to permit agreement among various generators and users of pollution data? Are present code systems such as the Chemical Abstract Service registry for toxic substances and the NODC taxonomic codes for marine biological studies adequate?

**Data Retrieval:** For data made available through the OPDIN, what form of cost recovery is reasonable? Should OPDIN serve as a remote, interactive source to archived marine pollution data or provide only inventories or references to data sources?

**Data Synthesis**

Except in certain cases, archived marine pollution data are not readily available in useful forms. Translation of stored data into useful forms requires synthesis. Each synthesis differs depending on the ultimate information needed by the user or requester, the availability of data (quantity and quality), and the resources available to perform the translation.

The process of synthesis implies at least three categories or levels of products or activities: (1) data and data summaries; (2) information products, such as maps or charts, and (3) analyses requiring interpretation based on scientific expertise.

In general terms, a synthesis entails six sequential steps: 1) phrasing of the initial question into discrete answerable parts; 2) development of an analytical scheme which will yield usable answers; 3) identification of input data required by the scheme; 4) acquisition of the input data; 5) operation of the analytical method to produce an analysis; and 6) interpretation of the analysis to generate an answer. Syntheses can be nested.

The OPDIN, and in particular the Central Coordination and Referral Office (CCRO), will have a significant impact on the timeliness of synthesis by reducing the time and effort required to contract with vendors of synthesis services. This can be achieved by setting up standard schedules and arranging for streamlined contracting operations. Depending upon the user, services requested of the Ocean Pollution Data and Information Network may range from generation of a bibliography, data inventory, or specific data set to generation of a complete synthesis. As a general rule, however, the OPDIN will not usually perform synthesis itself, but will restrict its activities to assisting users in arranging for syntheses and delivering data sets or prepackaged products. The major effort of each synthesis will be the responsibility of the requester. It is envisioned that OPDIN will arrange for purchase of these efforts from private industry, academia, or other government agencies.

OPDIN will provide no special services that are normally available from the private sector, nor will OPDIN knowingly provide unfair competitive opportunities.

**Information Management**

If synthesis is the process of providing timely and useful products and services, information, then, is the end product. Digital data is a subset of information. However, for the purposes of this discussion, information is the textual, or imagery counterpart of digital data.

OPDIN should be designed to address information needs of requesters whose requirements are not now being satisfied. Information products and services must be available to meet varying requirements of both scientific users in Federal agencies, and of policy-makers, non-scientific public users, and others.

Providing access to literature—both published documents and institutional reports, the so-called "gray literature"—is a major component of information management. The existing bibliographic, abstracting, and literature archival/retrieval services presently available are recognized. No attempt should be made to duplicate them. CCRO and CRRO staff should receive training in use of these services. OPDIN should particularly seek to ensure that relevant literature is available locally to users. This may be accomplished through arrangements with existing document archival services or through other means. A guide or manual describing these services may be worthwhile.

The quarterly newsletter, Coastal Ocean Pollution Assessment (COPAS) News, has great potential value as a tool for disseminating information on marine pollution. No similar periodicals produced in the United States are devoted primarily to domestic marine pollution activities.

Providing an opportunity for marine pollution information managers to meet on a
regular basis was recognized by the Workshop as a major objective that should be carried out by the CCRO. Information dissemination will be successful only if all the individuals involved—from all participating agencies—understand the capabilities of other groups and understand their own obligations. Workshops, or other forums, perhaps international in scope and including data generators, are essential to provide this interaction.

SUMMARY

The utility of pollution data and information to the descriptive or decision-making process is largely a function of timeliness, format, and completeness. Although there is no consensus on the definition of timeliness, a response to a straightforward request should take no longer than two days. Data should be processed into the system within two weeks. Formats should be simple and readily understood.

It is further recommended that users be encouraged to express their level of satisfaction with the usefulness and timeliness of products and services. A user advisory group might be formed to provide needed product feedback to OPDIN.

One of the most important products of the Network would be a User's Guide describing extent and limits of services offered and procedures to access the network. The User's Guide should be of the highest quality and priority and should receive the widest possible circulation. Periodic announcements should be made by OPDIN to keep its user community aware of current activities and available services.

REFERENCES
