(iv) photographic methods of recording for taxonomic, behavioral, and morphological purposes and long-term environmental changes with suggestions on documentation and retrieval of both still and cine film

(v) basic statistical techniques and parameters for use in plankton studies, and the

(vi) suitability of freeze-drying for biological archives.

While the Working Group is sympathetic with regards to the needs for the examination of certain of the above-listed topics, others seem to have been adequately covered for the present in recent publications. In either case, the Working Group 23 feels that these subjects are outside their terms of reference. It does, however, suggest that SCOR examines the need for forming working groups to study these subjects, in particular nos. ii, iii, iv and vi. Working Group 23 does recommend that any future working groups issuing manuals on these topics consider the use of the loose-leaf style for their publications.

3.5 Working Group 23 expressed the hope that on the basis of the information obtained through the activities of this group and the need for additional experience, UNESCO would organize courses and programmes in plankton laboratory methods.

3.6 Dr. A. Fehlmann reported that he had arranged for the compilation of more than 3,000 references in the field of fixation and preservation of biological material including methods applied in industry. The Working Group took note of this information and expressed the hope that the Smithsonian Oceanographic Sorting Center would, in due time, make these references available to interested marine biologists.

ANNEX VIII

REPORT OF SCOR WORKING GROUP 32
BIOLOGICAL DATA INVENTORIES

Proposal for a Second-Level Inventory System to Facilitate Dissemination and Exchange of Biological Oceanographic Data

1. Introduction

The objectives of any system to facilitate the dissemination and exchange of marine biological data are:

(i) To provide Institutes and individual workers with any available information which can supplement their own observations.

(ii) To facilitate the compilation of comparative data from different regions.

(iii) To provide individual workers with information about possible sources of material for taxonomic and other studies.

(iv) To assist in local, national and international programme planning by making it possible to identify gaps in existing information and to provide a convenient source of information on the form and status of current work in a given area.

It is generally recognized that the traditional means of information exchange, by reference to and exchange of publications, no longer provides an adequate means of obtaining data. There is a growing need for improved handling and accessibility of biological data. The number of large international projects, FAO projects and other investigations in biological oceanography is growing. The amount of data being collected by continuous measuring and sampling devices is rapidly increasing. There is a growing need for long-term monitoring of marine communities for pollution, conservation and management studies and in relation to these there is a growing interest in eco-system modelling.

In many instances it is not practical to meet the needs for data management with respect to these developments through the medium of publication and it seems necessary to look for an alternative. Despite repeated and continuing efforts to standardize methods, most biological data still require detailed information to be available on the methods of collection and analysis. The interpretation of the data is often influenced by these considerations. Therefore, the submission of raw data to an agency and subsequent retrieval based on a centralized data bank cannot be regarded as a satisfactory basis for data dissemination in many fields of marine biology.

What is required is a method of establishing communication between a user and those agencies and institutions which hold data that might be of interest to the user. The centralized storage of inventories of the data holdings of institutions supported by a retrieval system adequate to meet the needs of the majority of users would appear to be, at present, the most generally applicable means of information exchange.

It seems reasonable that the institutional basis for any system should be the national and international oceanographic data centres. It would appear that the referral concept is playing an important and increasing role in the activities of these centres. An interdisciplinary (including biology and pollution) data inventory system based on the ROSCOP reporting form is being developed which is designed to provide an immediate report that observations have been made. In most disciplines the data reported by means of the ROSCOP system should be immediately available for exchange. A large proportion of biological observations, however, take the form of samples of organisms which have to be sorted and counted or at least require fairly extensive treatment in the laboratory subsequent to their collection. The time lag between collection and the production of exchangeable data, other than the material itself, is commonly several months and may extend to several years.

In addition, the considerable diversity of biological data and the associated comparability problems means that most users require, at least for a first look, developed statements of the results of observations rather than simple listings of raw data.

Taking all these factors into consideration it is believed that there is a need to supplement the initial reports that observations have been made, the ROSCOP system, by a second-level inventory system the objective of which is to provide a medium for the reporting and centralized banking of information about the structure and format of the results of biological observations.

It is visualised that the main banking agency for the second-level inventory will be WDC-A, Oceanography. Reports of data and requests for information will probably be channeled through national oceanographic data centres or designated national agencies. These may wish to provide limited banking facilities.

2. The Proposed Inventory System

2.1 Content and Structure

The basic aim of any data inventory is to establish communication between a user and all the holders of all the data which might be of value to the user.

The inventory must contain information required for search and select processes in response to an inquiry by a user. This information must satisfy the question WHO did WHAT and WHERE and WHEN was it done?

In addition the inventory must contain information to enable the data holder to identify the data set and further information for the user to assist him in selecting, from data sets listed in response to a formal WHAT, WHERE, WHEN inquiry, those which are likely to provide him with the information he needs. For this purpose the user needs to know about methods and sampling program, availability, data structure and data format.

2.1.1 Search and Select

2.1.1.1 WHO?

Each inventory entry must contain a reference to the title and full postal address of the institute or agency responsible for holding the data.

2.1.1.2 WHEN?

This should be reported by listing the dates of the first and last days on which the observations covered by the entry were taken or, in the case of monthly or yearly summaries, the time span in months and years. The inventory entry need only include month and year references.

2.1.1.3 WHERE?

This is probably the most difficult question. For the benefit of users the location of the observations should be defined as closely as possible. To allow efficient retrieval the area definition should be as simple as possible. Reporters require a versatile system which can cope with a wide variety of data acquisition systems.

The only obligatory area report system proposed for the ROSCOP inventory is by 10° squares. It is the view of WG 32 that this will not meet the requirements of a large number of users. This view is supported by the U.S. NODC from their experience with NAMDI; the FAO fisheries data centre, because most fishery data relate to areas considerably smaller than 10° square; and ICES, due to their interest in the North Sea and the Baltic Sea.

A uniform grid fine enough to satisfy the user would probably be a 1° square. From the data centres' view the search procedure may have to be stratified but is not basically different from one based on 10° squares. The reporter, however, will have difficulty in recording long-linear tracks which are likely to be involved to an increasing extent as under-way data acquisition systems continue to be developed.

Probably the most useful system from the points of view of the reporter and the user is based on a reporting system which permits differentiation between stations, lines and areas, positions being reported in terms of Lat. and Long. to the nearest minute. The position of stations may be reported individually or if a number of closely grouped stations are involved an area may be reported by means of the positions of the corner points of a delimiting polygon. A single line may be reported by the positions of start, altered course and end points. A network or set of lines may be reported as an area.

Given this reporting system the user can define his area of interest by giving the positions of the corner points of a delimiting polygon.

This system clearly presents problems for the retrieval system. The search procedure would very probably have to be stratified and the fine level search would be mathematically more complicated than for a search based on a uniform grid. The basic process involved can be reduced to one of determining whether or not two lines intersect.

WG 32 strongly recommends that this system be given serious consideration and, if it should prove unacceptable for technical reasons, at least the reporting system should be retained with the inventory containing references to 1° squares. The user would then be able to define his area of interest in terms of a rectangle of 1° squares.

2.1.1.4 WHAT?

The inventory must contain a reference to the subject of the observations. Perfect subject classifications do not exist and the number of classes that can be incorporated in a

practical report system is strictly limited. This was probably the weakest aspect of the original ROSCOP inventory and, in order to overcome some of the shortcomings, the IODE Task Team introduced additional lists of studies to supplement the subject classes.

A considerable improvement in subject definition can be achieved if the study list is used to qualify a specific subject, defining a sub-set. For example, the subject PELAGIC FISHES could be qualified by MARKING AND/OR TAGGING. And, if the report covered trials of a new fish tag the subject could be further qualified thus, PELAGIC FISHES, MARKING AND/OR TAGGING, EQUIPMENT TRIALS and STUDIES.

This system is suggested for the second-level inventory and a provisional list of study headings is incorporated in the proposed report form.

Provided reasonable AND/OR logic is applied it seems likely that retrieval will be possible on a single pass search and the system should not produce any major problems for the data centre.

2.1.2 Data Identification

The inventory must contain information to enable the data holder to identify the data set from amongst his own holdings in response to a request for a user.

The proposed report form contains entries for a form reference number, to be allocated by the reporting institute; a program name or title, if applicable; the name or title of an individual, team or department within the reporting institute; an originators' reference number allocated by the reporter, which might be a cruise number; the type of platform used for making the observations and the identification of the platform, these will commonly be an oceanographic ship and her name.

Reporters should be encouraged to complete as many of these entries as possible because it is believed that this is an area in the inventory where some redundancy will do no harm.

2.1.3 Additional Information for the User

The user should be provided with as much information as possible, within the framework of an inventory concept, to assist him in selecting data sets which are likely to contain the information he needs.

The proposed report form contains entries for the nature of the area covered by the observations, indicating whether it is an estuary or the open ocean, etc., and the IHB area code. Sections of the form cover sampling gear, sampling program, analytical methods, data structure and data format, availability and deposition of the data. Also, reporters are encouraged to supply supplementary information in the form of cruise tracks, station lists, species lists, specimen data lay-outs, etc.

2.2 Integration into the International Oceanographic Data Inventory System

The proposed biological second-level inventory can be implemented as a stand-alone system. It is believed, however, that the value both of this system and the ROSCOP inventory would be greatly enhanced if they were integrated.

The machine retrievable sections of the proposed second-level system have been designed to be compatible with what WG 32 would like to see in the ROSCOP inventory, and it would be an advantage from the reporters point of view if the relevant sections of the report forms were made identical.

The integrated system, from the data centre's point of view, might work as follows:

ROSCOP reports are received indicating that observations have been made. After a

lapse of time, second-level reports would be received which would contain a reference to the ROSCOP report or reports up-dated by the second-level report.

If necessary the data centre could periodically (say, 4-monthly) search the ROSCOP inventory and retrieve all reports of biological observations that have not been updated and requests for second-level reports could be channelled through NODC's and DNA's to Institutes.

In response to a request for biological information from a user the second-level inventory would be searched followed by the first level inventory but, in this search, only those entries not marked as having been updated would need to be retrieved.

The user gains a virtually automatic search of both first and second-level inventories in response to a single request.

The reporter gains from the similarity of not inconsiderable sections of the two report forms.

The data centre gains from the fact that the coding and retrieval systems for the two inventories are virtually identical. The only additional requirement is a system for marking updated ROSCOP entries.

WG 32 believes that the benefits, in terms of service to users, to be gained from integration of the first- and second-level inventories are considerable, and, although the process of achieving compatibility between the two report forms might involve some delay in the implementation of ROSCOP, it is considered that the advantages will more than compensate for this.

WG 32 would also like to draw attention to the fact that the proposed biological second-level inventory system could be applied to other disciplines: there is very little in the system, apart from the subject lists, which refer specifically to biological data.

In particular, it is believed that the section on data structure and data format could be carried over unmodified to other disciplines and all that might be necessary is to extract the relevant subject list from the ROSCOP report form and modify the section on sampling gear and analysis methods.

3. The Inventory Report Form

3.1 Design and Layout

There are advantages in aiming for compatability with machine printing systems including typewriters, teletypes and line-printers. This implies a fixed line and column spacing and a separate page layout.

The advantages are:

(1) Forms may be completed in typescript.
(2) Reporters may produce complete forms on a teletype or line-printer during routine data processing procedures and the results would closely resemble the printed version.
(3) The reply to a user's request could consist of machine produced pages mixed with xerox copied pages with no marked inconsistances in style and format.
(4) Separate page layout facilitates copying on xerox and other machines.
(5) Separate page layout facilitates the publication of complete, or sections of, report forms (see, for example, the ICES publication of ROSCOP forms).
(6) Flexibility in the structure of reports.

For example, more than one data structure report could be attached to a single data identification section.

It would appear that the commonest spacings on machine printing systems are 1/6 inch line spacing and 1/10 inch column spacing. The maximum length of a line is set by teletypes at 72 characters and the maximum number of lines is set by line-printers at 60 lines (leaving 3 line margins at top and bottom). These give a printed area of 7.2 by 10 in. and a page size of about 8 by 11 in. Such a size can be copied on a wide range of document copiers and can be filed easily in normal folders.

It is probable that the form will frequently be completed in manuscript and a space of 1/5 in. by 1/3 in. (four times the printed character area) should be allowed for each manuscript character wherever possible.

A trial form was provided using a multiaccess teletype and divided into 60 line pages. Any developed version should retain the same spacings of entries but greater clarity could be introduced by varying the size and font of the printed headings. The form might be produced as a set of tear-off pads and possibly in concertina continuous form for typed entries.

4. The Inventory Request System

The system is adequately described by the proposed form and completion instructions are available separately with the exception that the final version should contain some statement on any charges involved. This will obviously depend on policy decisions by WDC-A and national agencies.

5. Implementation of the System

Resistance to the scheme is anticipated from some potential reporters who may see it as just another useless and admittedly rather complicated form to be filled in, and also from some who may be reluctant to make available to other users developed data and results over which they have expended much time and effort.

The best way of overcoming this kind of resistance is to show that the system works and is successful in meeting the needs of users.

Publicity exercises, probably at national level and mounted by national data agencies, are going to be needed to launch the scheme and see it through the early stages of development.

It seems probable that periodic publication of extracts of the inventory entries will help in publicising and selling the system.

National data agencies will also have to be responsible for seeing that all the institutes producing relevant data are asked to submit returns. The inclusion of a section on pollution observations means that a wider range of institutes (including river and harbour authorities and some public health authorities) will be involved than was envisaged in the original ROSCOP proposal.

6. Conclusion

WG 32 would like to stress their belief in the need and value of an international data referral system for biological oceanographic (including estuarine and marine pollution) data.

It is inherent in any such scheme that it can provide a really valuable service only after the passage of several years when the inventory will contain a reasonable number of entries.

The working group are forecasting that users will need the system and that their needs will be met by the system proposed in this report.