INTERNATIONAL COUNCIL FOR
THE EXPLORATION OF THE SEA

C.M.1980/D:12
Statistics Committee
ref: Hydrography Cttee

Introductory paper to the Joint Session of the Hydrography and Statistics Committees

by
ir.P.J.F.Geerders
Netherlands Centre for Oceanographic Data,
c/o KNMI, Wilhelminalaan 10,
3730 AE, De Bilt, the Netherlands.

ABSTRACT

Today mankind is facing an increasing amount of numbers. This amount will soon be overwhelming if we don't take some precautions.
Causes of this process can be found in the technical developments regarding data acquisition as well as data processing. These have caused a clear shift in scientific thinking.
Also marine science suffers from this development. More and more instruments and methods produce incredible amounts of numbers and we start wondering where to go and what to do with all this data. Does data really contribute to knowledge and understanding?
New developments in technology lead on one hand to higher amounts of data, but give on the other hand a possibility for "smart" data acquisition. Also the amounts of data that can be stored increase rapidly. Our historical files expand but how much should we really keep and for how long?
The situation asks for an increasing awareness of the problems of today and tomorrow. International arrangements should be made in order to develop practical and scientifically justifiable solutions. These solutions should regard:
- the acquisitions,
- the processing,
- the storage,
- the exchange and
- the presentation of data.

Since the problems are not limited to one type of science, a multi-disciplinary, international approach might be needed to solve these problems.

1. History

A few hundred years ago, science had a much more philosophical character than today. Scientists were those who had time (and money) to walk and look around in Nature, to think about observations and to perform simple experiments.
Also in marine science this attitude was present. A few privileged did some research and/or compiled knowledge and experience e.g. from sailors and fishermen. This knowledge sometimes dates from long time ago. Already the Egyptians, the Phoenicians and the Greek drew maps with indications of depths and currents. At that time in the Pacific empirical knowledge about the interference patterns of waves, refracted by islands and reefs, was used for navigational purposes.
Information was available only in a limited amount and from a limited number of people. Some of them became known by their data-products (Mercator, Halley).
However, as the exploitation of the sea increased, the demand for information increased too. This information was almost purely used for practical purposes, e.g.:
- where will I find a certain fish,
- how can I navigate best to get as quickly as possible from A to B.

Thus, a separation developed between the operational and the purely scientific use of information. Still this gap exists between information that is used to support a certain task, and information that is used to enrich Man's knowledge of Nature.

2. Present situation

With the advent of mechanisation and automatisation, our ability to "measure" has

/2

enormously increased. Since Nature contains many variables - and it seems that we discover every day more of them - the process of doing measurements becomes more and more a daily routine.
The process of measurements can be simplified as follows:

Nature $\overset{1}{\longleftrightarrow}$ measurements $\overset{2}{\rightarrow}$ interpretation $\overset{3}{\rightarrow}$ knowledge (any feedback loops neglected)

Today we have more and better means to do measurements. This has led to higher data rates. Compare:
- a simple bucket temperature measurement, about: .1 bit/sec;
- continous shipboard temperature measurements, about: 100 bits/sec;
- thermal scanner in a satellite, about: 10.000 bits/sec.

This same trend is present in physical, chemical and geophysical oceanography. In biology a slightly different situation has developed whereby the advent of more and flexible platforms to observe and catch species has led to higher data rates.
Measured data is present all over the world in many different formats, even the same data. In many cases it is unsufficiently known that certain data is available and the data-holders are not always prepared to exchange data with others.
The step towards the interpretation of the data has consequently been complicated to a large extent. Of course, computers - digital and/or analog - give us some possibilities to limit the high data rates, to "play" with the data, to archive and present the data in some way, but this is not very satisfactory. Several attempts have been made to find long-term solutions in hardware and software, but for each, the increasing amounts of data have caught up with them.
Statistical analysis of data leads many times to complicated schemes where the relation between the original data and the final results is a bit vague. Compression of data volumes can be easily done, but what data can be disposed of and what data is really important? For archival and exchange of data many different formats have been developed. Mostly they are very specific for a discipline or even an organisation while implementation of software on different computersystems gives problems. Many approaches have been made towards an optimum way to present data but no panacea exists as far as I am aware of.
Since the interpretations of data should eventually lead to an increase of Man's Knowledge the human factor is a key factor in the whole process. The human brain has impressive capabilities to handle data but they must be presented to him in an optimum way.
It should be clear that the same remarks are valid for the products of science - reports, articles and books. Worldwide activities are in progress to improve the access to literature but it seems sometimes that the systems that exist are not used with the intensity that could be explected. Is it perhaps that computer and datacommunication specialists develop systems that they like but that do not meet the real users demands? In this field copying machines and suchlike facilities do not contribute very much to a solution. Maybe we should only copy something if we consider it also worthwhile to copy it in writing, if necessary.

3. Future

The development of technology alone will not bring us a solution to our problems. It will only lead to still higher datarates, faster and smarter computers, more complicated software and CRTs wherever you look. The fundamental problems could be handled with (a combination of) a few approaches:
- the key role of the human factor should be brought back into the discussions on this subject. In close relation between data experts and psychologists/physiologists ways should be sought to optimise the use of the ability of Man to transfer numbers into knowledge. Key fields are here: data presentation, supervised dataprocessing and user-friendly databases.
- the increase in knowledge should be used as soon as possible to limit the data rates from our measurements. Technology certainly gives us ways to do this but today in many cases lowering of the datarate from one sensor leads to the inclusion of more sensors, by which the gain is annihilated again.
- the likely limitation of financial means for all of us, should be regarded as a strong impetus to strive for an intensification of exchange of data. Administrations should be convinced that the "struggle for money" must not lead to embargos on data. National and international cooperation of scientists is the only worthy answer to the financial problems.
- in the field of long term data storage we should be careful. Again technology gives us the possibilities but at the expense of what? Criteria should be developed to select

/3

data for long term storage. Is it likely that we will really use today's data in the year 2080? It might be worthwhile to seek for "black holes" for data, as they exist in the universe for matter and energy.
- since the fields of datacollection, dataprocessing and data-interpretation have developed into specialised areas, it might be useful to consider them as separate fields with separate personnel. It seems very difficult to keep up with the developments in all these fields and still run your primary job.
- one of the big problems is to know where certain data is available. Reference systems should be developed to enable scientists to access certain types of data. The IOC Roscop forms might be a help in setting up these systems and make these accessible to the scientific community.
- IOC's national oceanographic data centres should focus upon the following tasks:
  . provide references to data and information,
  . supply and reformat data upon request,
  . give advice on data matters,
  . perform trend studies.

4. Option

The data explosion is certainly not unique for marine science. All sciences and everybody is confronted with this problem. Several national and international activities are in progress or planned to cope with this explosion but these activities are mostly focused upon one or a few aspects or disciplines. They also suffer from geographic limitations (local formats or systems, national interests).
It might be of great value to approach the problem of the data explosion with an international, probably intergovernmental organisation, e.g. on UNESCO level, to maintain coordination, give advice and stimulate where necessary. Only such an organisation is able to find more universal solutions to the problems. Such solutions will definitely have a longer life than specific, national or regional solutions.

-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-o-