

Serving data from the SCAR Southern Ocean Observing System (SOOS) using the SeaDataNet infrastructure

T. F. de Bruin

NIOZ Royal Netherlands Institute for Sea Research, P.O. Box 59, 1790 AB, Den Burg, The Netherlands

Netherlands National Oceanographic Data Committee (NL-NODC) p/a N. Kinneging, Rijkswaterstaat Waterdienst, Zuiderwagenplein 2, P.O. Box 17, 8200 AA LELYSTAD, The Netherlands

Received: 25 January 2010 – Revised: 1 March 2010 – Accepted: 1 June 2010 – Published: 27 September 2010

Abstract. The Scientific Committee on Antarctic Research (SCAR) and the Scientific Committee on Oceanic Research (SCOR) jointly intend to build a Southern Ocean Observing System (SOOS). This paper addresses the required data flow infrastructure.

SOOS will use a system of systems approach, using existing observation programmes and projects. Data should be submitted to professional data centres. The problem arises how to link all these data centres and get a central overview of the SOOS data as well as direct access to the data.

The Netherlands National Oceanographic Data Committee (NL-NODC) has successfully built a national distributed oceanographic data access infrastructure, adopting and implementing technology developed by the European SeaDataNet project.

The Dutch system has been operational since early 2009.

The conclusion is that the SeaDataNet technology can be used to build an operational, distributed data delivery infrastructure, featuring all elements required by the Southern Ocean Observing System (SOOS).

1 Introduction

Nowadays, all big projects in observational science recognize the importance of data preservation and the data management required to achieve that. Data resulting from a project may be seen as “the most important single outcome of (a) programme” (International Council for Science, 2004) and need special care to be preserved for future (re)use.

The Scientific Committee on Antarctic Research (SCAR) is a highly respected and influential organization. SCAR’s mission is “the initiation, promotion and co-ordination of scientific research in Antarctica”¹. Established during the International Geophysical Year 1957–1958 (IGY) by the International Council for Science (ICSU) and starting with 12 member countries, SCAR has now grown to 31 member countries and 9 Union members. SCAR also provides international, independent scientific advice to the Antarctic Treaty system¹.

In certain ways, the Scientific Committee on Oceanic Research (SCOR) can be considered to be SCAR’s oceanographic counterpart. Also established by ICSU during IGY, SCOR focuses on “promoting international cooperation in planning and conducting oceanographic research, and solving methodological and conceptual problems that hinder research”². Currently, thirty-six nations have formed national SCOR committees and are active in SCOR.

SCAR and SCOR are jointly planning to build a Southern Ocean Observing System (SOOS). Recognizing that “the Southern Ocean is an integral and key component of the global climate system”³ and also that “the region is vast, remote and logistically difficult to access and thus is one of the least sampled regions on Earth”³ the SCAR/SCOR Expert Group on Oceanography of the Southern Ocean is “designing an observing system that encompasses physical, biogeochemical and ecological processes”³.

Some key elements are:

¹<http://www.scar.org/about/introduction/>

²<http://www.scor-int.org/>

³<http://www.clivar.org/organization/southern/expertgroup/SOOS.htm>



Correspondence to: T. F. de Bruin
(taco.de.bruin@nioz.nl)

The design, choice of parameters to be measured and deployment of the observing system will be guided by science and policy questions which need to be addressed.

The research conducted with the SOOS system will be multi-disciplinary and inter-disciplinary.

The SOOS system will be circumpolar and will thus involve many nations.

Instrumentation will be deployed in remote locations, where, until now, measurements have hardly been possible to make (e.g. under sea-ice).

Very importantly, it will use a system of systems approach, combining the efforts of many existing programmes and coordinating the development of new programmes and projects.

Special attention has been given to data preservation and data management right from the start of the planning phase of SOOS. In the words of the SOOS Strategy Discussion Paper: "A Data Management System will be an integral part of SOOS, to ensure data is easily accessible and of the highest possible quality, but like the observations themselves the data system will rely heavily on what already exists."⁴

The question is whether there are there any existing elements which may form the SOOS Data Management System?

2 Existing elements

There are many professional, discipline-based data centres. More than 64 nations have a National Oceanographic Data Centre (NODC). ICSU has established 4 World Data Centres (WDC) for marine and oceanographic data. Many of the major oceanographic projects have set up some sort of a data unit or committee, to handle and preserve the data resulting from that project, at least during the lifetime of the project.

In general, these data centres are much better equipped and resourced to preserve oceanographic data than individual scientists or research institutes are.

The logical choice for building the SOOS Data Management System would be to use the NODCs as "existing elements" and have all data submitted to and preserved by professional data centres.

Problem solved?

Not really. Even though all these NODCs and WDCs are cooperating on a global level within the International Oceanographic Data and Information Exchange Committee (IODE) of the Intergovernmental Oceanographic Commission (IOC) and on various regional levels, there was no modern, Internet-based technical infrastructure for data exchange until very recently.

All this changed with the start of the European SeaDataNet project⁵. SeaDataNet is a consortium of NODCs from 35

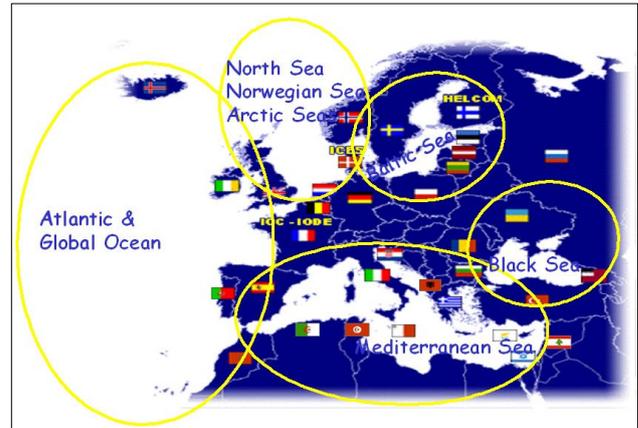


Fig. 1. Participating countries and international organizations in SeaDataNet. The yellow ellipses indicate regional areas of interest.

coastal nations in and around Europe (Fig. 1) and includes international organizations such as IODE, the International Council for Exploration of the Sea (ICES) and the Joint Research Centre (JRC) as well as several university groups.

A total of 49 partners are cooperating towards a common goal of building a Pan-European distributed data access infrastructure.

If it is achievable and works, the SeaDataNet infrastructure is a prime candidate for the SOOS Data Management System.

3 Oceanographic data management in the Netherlands: a miniature SeaDataNet

In 1997, The Netherlands established the Netherlands National Oceanographic Data Committee (NL-NODC)⁶. The NL-NODC is a committee, with broad representation from all major data collecting institutes and organizations in The Netherlands, as opposed to the situation in many other countries which have one single, oceanographic data centre. The activities of the NL-NODC are based on a Memorandum of Understanding, signed by the directors of the participating institutes.

Currently, the following seven institutes are participating in the NL-NODC:

Rijkswaterstaat – Directorate General for Public Works and Water Management, NIOZ Royal Netherlands Institute for Sea Research (also representing the academic community), TNO B&O – Built Environment and Geosciences (Geological Survey of The Netherlands), Hydrographic Service of the Royal Netherlands Navy, Centre for Estuarine and Marine Ecology (CEME), Wageningen IMARES, Deltares.

⁴SOOS Strategy Discussion Paper (not available any more)

⁵<http://www.seadatanet.org>

⁶<http://www.nodc.nl>

The Marine Information Services (MARIS), a private company, advises the NL-NODC, while the Royal Netherlands Meteorological Institute (KNMI) used to be an NODC partner until the end of 2008.

These seven organizations form a very interesting mix of two government agencies, two research institutes and three privatized, former government agencies. Where the government agencies have specific tasks, usually limited to territory and Exclusive Economic Zone (EEZ) of the Kingdom of The Netherlands, Netherlands Antilles and Aruba, the other NODC Partners may operate all over the world.

It is estimated that these organizations together manage at least 90% of all marine and oceanographic data collected by organizations in The Netherlands.

Within the NL-NODC, many different datatypes are handled, coming from all disciplines: Coastal research, bathymetry, geology, biology, ecology, physical oceanography, marine meteorology, etc.

The NL-NODC has a fully distributed structure, analogous to the structure of both SeaDataNet and SOOS, albeit at a much smaller, national scale. In fact, from a data management point of view, one can consider the NL-NODC to be a miniature SeaDataNet or a miniature SOOS. The NL-NODC thus forms an ideal test bed to implement and test the SeaDataNet data access technology and to evaluate the feasibility of this technology for SOOS.

4 The NODC-i project

Prior to 2005, the NL-NODC was in a situation very similar to SOOS now, concerning data access and data exchange. The formal structure for data exchange was in principle well organized, with a Memorandum of Understanding and a National Oceanographic Data Policy underpinning the willingness to provide access to data and to exchange data.

However, the technical infrastructure for data exchange was completely lacking. Organizations were using different relational data base management systems (RDBMS) and stored the data in different formats. Some of these data bases were online accessible, while others were not connected to the Internet.

As a result, the user had no overview of data availability and, if she or he could get access to the data, had to perform a whole range of format conversions.

To remedy this situation, the NL-NODC started the NODC-i project. The goal of this project was to build a "National infrastructure for access to oceanographic and marine data and information". The project ran from February 2005 until December 2008.

SeaDataNet system

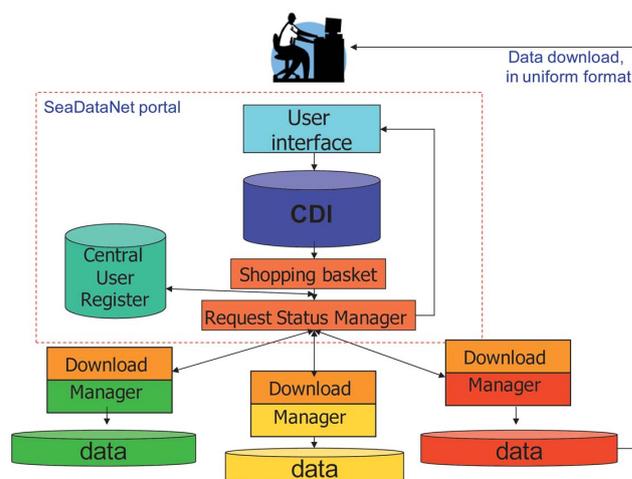


Fig. 2. NODC-i/SeaDataNet data flow. See text for complete description.

The project had two main objectives:

1. To build a central index of all data in the partner databases, using a common, standard description of each individual data point.
2. To provide transparent access via Internet to the partner databases and deliver the data in standard formats.

These objectives are, not coincidentally, identical to the objectives of SeaDataNet, which started at about the same time. Right from the start of the NODC-i project, it was decided to adopt and implement SeaDataNet technology, thus providing an ideal test bed for SeaDataNet, while contributing to an European scale standardization at the same time.

The project was concluded successfully, with all objectives achieved. The national oceanographic data access infrastructure has been operational in The Netherlands since early 2009.

5 Data flow within the NODC-i/SeaDataNet infrastructure

The infrastructure is a one-stop, single sign-on access system to a distributed system of data sources. The data flow, from the user's perspective, is shown in Fig. 2.

The user interacts with the system through a dedicated user interface. Using this interface, the user will get an overview of the entire data holdings of all participating data centres. This is achieved by way of a central data description register, which provides a centralized overview of data availability. The register contains data descriptions in the so-called Common Data Index (CDI) format. The CDI format is a fine-grained index to individual data measurements (such as

a Conductivity-Temperature-Depth (CTD) cast or a moored instrument record), with a “common” or standard description (location, time, parameter, availability, owner, etc.) of those individual measurements. Each of the participating data centres has contributed the CDI descriptions of its data holdings to the central register.

Having gained this data availability overview, the user can now order data with a shopping basket mechanism. Before transferring this order to the distributed data centres, the portal software checks whether the user is allowed to access that data. This is done by a central user register. Each user has to register once in order to get a user account. In this process, one or multiple roles (e.g. “Public”, “Academic”, “Commercial”, “National and local government”, “Pan-National government”) are assigned to the user. Each data set may have data access restrictions, which may also be dependant on the assigned role of the user. With the help of a business matrix, the portal software determines whether access restrictions apply for that user and that data.

The Request Status Manager (RSM) handles all incoming data requests and interacts with the Download Managers. The RSM splits incoming requests into parts which are tailor-made for each data centre. This way, a data centre will only receive data requests for data it actually holds. The data requests are transferred to the Download Manager. A Download Manager is customized software on top of an RDBMS, which can communicate with the local database and retrieve data from that database.

After having retrieved the data, the Download Manager will convert the data into one of two standard output formats. For profile data this is a very well described ASCII format (called Ocean Data View ASCII or ODV-ASCII), while for gridded data this will be in the CF variant of NetCDF (not yet implemented). The Download Manager will signal the RSM that the data are ready for downloading.

The final step of the data flow will be for the RSM to inform the user that the data are ready for downloading and can be downloaded at any convenient time for the user.

6 Experiences with the operational NODC-*i*/SeaDataNet system

The NODC-*i* system has been fully operational since early 2009.

One of the biggest advantages of this approach is that each participating data centre can continue to use whatever RDBMS it is used to work with and can continue to operate in the way it is used to operate. The NODC-*i*/SeaDataNet infrastructure of Download Managers, RSM and CDI-register simply adds an additional or intermediate layer on top of the existing local or national infrastructure.

The system has proven to be very reliable and robust. It is technically capable of handling all kind of situations,

ranging from fully automated data extraction to manual data extraction.

The system is also well designed: only minor adjustments and improvements have been necessary with each new release of the Download Manager.

Most importantly: people are starting to use it, without experiencing any problems.

The system is open to everybody. Everybody can access the CDI database and get an overview of data availability. To be able to actually download data, people have to register first. Even though people are starting to use it, there are currently only 29 registered users in The Netherlands and some 300 in all of Europe.

This clearly shows the paramount importance of a professional, well-structured and well-organized Education, Outreach and Communication (EOC) campaign, to advertize the benefits of the new infrastructure: “If nobody knows about it, nobody will use it”.

Such an EOC campaign ought to be part of every major project in observational science.

7 Conclusions

The NODC-*i*/SeaDataNet distributed data access infrastructure has been operational in The Netherlands since early 2009.

It is important to notice that this does not place an additional burden on scientists or science projects. They can continue to submit the data to the data centres, the way they are used to do. Even data centres can continue to do what they do best, in the way they are used to. All that is required is to install an additional layer to provide the functionality to get an overview of and access to data via the SeaDataNet infrastructure.

In fact, the fully automated data delivery reduces the workload on the data centres.

The conclusion is that the SeaDataNet technology can be used to build a working, distributed data delivery infrastructure, featuring all elements required by the Southern Ocean Observing System (SOOS).

The big challenge will be to couple the (European) SeaDataNet infrastructure to the emerging parallel Australian and American systems and prevent these developments from diverging. Here lies an important role for the International Oceanographic Data and Information Exchange Committee (IODE) with its Ocean Data Portal project.

Edited by: G. M. R. Manzella and S. Nativi

Reviewed by: two anonymous referees

Appendix A

List of acronyms

ASCII	American Standard Code for Information Interchange
CDI	Common Data Index
CEME	Centre for Estuarine and Marine Ecology
CF	Climate and Forecast (conventions)
CTD	Conductivity-Temperature-Depth
EEZ	Exclusive Economic Zone
EOC	Education, Outreach and Communication
ICES	International Council for Exploration of the Sea
ICSU	International Council for Science
IGY	International Geophysical Year 1957–1958
IOC	Intergovernmental Oceanographic Commission
IODE	International Oceanographic Data and Information Exchange Committee
JRC	Joint Research Centre
KNMI	Royal Netherlands Meteorological Institute
MARIS	Marine Information Services
NIOZ	Royal Netherlands Institute for Sea Research
NetCDF	Network Common Data Form
NL-NODC	Netherlands National Oceanographic Data Committee
NODC	National Oceanographic Data Centre
ODV-ASCII	Ocean Data View ASCII
RDBMS	Relational Data Base Management System
RSM	Request Status Manager
SCAR	Scientific Committee on Antarctic Research
SCOR	Scientific Committee on Oceanic Research
SOOS	Southern Ocean Observing System
TNO	B&O Built Environment and Geosciences (Geological Survey of The Netherlands)
WDC	World Data Centre

References

- (1) International Council for Science: A Framework for the International Polar Year 2007–2008, Produced by the ICSU IPY 2007–2008 Planning Group, ISBN 0-930357-61-2, 2004.