Climate and Environmental Database Systems

edited by

Micheal Lautenschlager
DKRZ German Climate Computing Centre
Hamburg

and

Manfred Reinke
Alfred-Wegener-Institute for
Polar and Marine Research
Bremerhaven

14

SEPAN - SEDIMENT AND PALEOCLIMATE DATA NETWORK
THE INFORMATION SYSTEM FOR THE PALEOSCIENCES

Michael Diepenbroek
Hannes Grobe
Manfred Reinke[1]

ABSTRACT

In order to fulfill the requirements of the paleoscience community, an information system has been developed and implemented which is able to archive paleoclimatic data of any kind, together with the related metainformation. The system provides standardized import and export routines, easy access with uniform retrieval functions and tools for graphical presentation. The network is designed as a client/server system providing access through the Internet.

THE PAST IS THE KEY TO THE FUTURE

Goals of Global Change Research

In our efforts to understand the natural variability of the Earth, most recent research concentrates on the climate system under the influence of anthropogenic changes. In predicting global change, we are faced with questions which can only be answered through a better knowledge of the past. For this purpose we need to know the history of climate and the related changes in the entire Earth system, and we have to understand the large-scale processes and responses of the climate system to forcing factors including our more recent human impact.

Paleoclimate data play a key role by extending the baseline of environmental and climatic observations far back into the geological past. The library of natural climatic archives is our best means of determining how the climate system operated under boundary conditions substantially different from today and to improve our under-

1. Alfred Wegener Institute for Polar and Marine Research, PO-Box 120161,
D-27515 Bremerhaven, Germany
E-mail: sepan@awi-bremerhaven.de

standing to predict future climate development and environmental changes in models. In order to validate their effectiveness, models must be capable of accurately reproducing conditions known to have occurred in the past.

Data concerning previous environmental conditions are available from instrumental records and documentary history, as well as from the different types of the worlds natural archives for paleoclimate. These include continental ice, marine, lakustrine and terrestrial sediments, cave speleothems, tree rings or corals. The archives provide records of different time spans from years to millions of years and resolutions between month and kilo years. Over 50 years of studies of the physical, chemical and biological parameters have clearly demonstrated that these archives contain a wealth of information about past changes in the Earth's environment.

A number of parameters are significant for models of climatic processes and forecasts of possible future climatic changes. These include past temperature variations, the composition and circulation of the atmosphere and the oceans, changes in the flux of materials transported by wind and water to the oceans, variations in biological productivity and the reactions of biological systems to environmental changes. The only way to obtain a continous and detailed record of these processes is to collect the proxy data and to store them in a consistent format. This data collection might be the source for future interpretations of global processes.

The Earth's Library

Ocean sediments provide us with a very long record in time, reaching back about 140 million years. Since the invention of piston coring by Kullenberg, geologists working on sediments from the marine environment have archived a tremendous amount of sampling material in their repositories, together forming a valuable and comprehensive library of the earth's geological history. The material, mainly taken during the last 4 decades, consists of sediment cores recovered with different technologies, such as piston, gravity or box coring for sampling of the uppermost layers (< 50 m) of the ocean floor. Also conventional drilling is being undertaken in the international ocean drilling projects (DSDP: Deep Sea Drilling Project, ODP: Ocean Drilling Program) to recover long sediment sequences of up to several hundred of meters.

The samples, taken from these sediment cores are analyzed to help reconstruct the earth's history. The development of a technique to measure the isotopic composition of fossil plankton shells and to use these data for the reconstruction of the ice age climate (Emilian 1955) founded the field of paleoceanography. This fast growing young branch of marine geology tries to extract paleoclimatic information stored in the sediments. Interpretation of these data contribute to the knowledge about the most important parts of the climatic system, which are the hydrosphere and the kryosphere. Results and data are generally published in the Journal 'Paleoceanography', the first issue being printed in 1986. Today 'Paleoceanography' is one of the first journals to require that authors store their data in established data bases (Webb et al. 1994).

The Data Management Initiative

Analytical data from sediments are complex. With the introduction of new and more efficient methods the number of parameters and the amount of data has increased by an order of magnitude during the last decade. To effectively use this scientific resource in the future we need an information system which guarantees a consistent storage and provides easy access for the scientific community.

The recovery and interpretation of paleoclimate related proxy data has traditionally been done through individual or single laboratory efforts, often examining regional records that cover a limited temporal domain. The emergence of an integrated Earth system science call for a much fuller knowledge of the past, in both space and time, and for data sets that are drawn as composites from different efforts and techniques. Through the organization of coordinated national and international scientific efforts, PAGES (Past Global Changes), a core project of the IGBP (International Geosphere-Biosphere Program), seeks to obtain and interpret a variety of paleoclimatic records and to provide the data essential for the validation of predictive climatic models. PAGES seeks the integration and comparison of ice, ocean and terrestrial paleorecords and encourages the creation of consistent analytical and database methodologies within the paleosciences (IGBP 1990).

In 1993, scientists from various German research instituts contributing to the marine part of the IGBP/PAGES, initiated a project in response to these needs, which has been to design and implement a system to fulfill the requirements above. Based on the discussion and recommendations of this group, the information system SEPAN (Sediment and Paleoclimate Data Network) was developed at the Alfred Wegener Institute for Polar and Marine Research (AWI), financed by the German Ministry of Education, Science, Research and Technology (BMBF). The goal of the three year project (1994-1996) was the implementation of an information system holding paleoclimate data of any kind in a consistent form, and to make this data available to the scientific community in an easily accessible form. The goals were to provide: 1. an overview of sampling material with related metainformation, 2. storage of analytical data in a consistent form with easy access, 3. tools for import/export, graphical presentation and complex retrievals. In the first implementation step the system should store data from research on marine sediments only but it should be able to store paleodata of any kind.

THE INFORMATION SYSTEM SEPAN

General Aspects

The most important generic aspects of an information system are the quality and availability of the data as well as a high adaptability and effective usage of the system by the user. Data quality can be described in terms of validity of methods and precision and objectivity of measurements (Gilb 1988). It is not essential to have only data sets of excellent quality, however it is important that the quality can be estimated. The completeness of the metainformation is crucial for the understanding of the analytical data. Legal aspects also have to be taken into account (Epstein 1988). The analytical method and the reference where the data are published for the first time should be included in the metadata. The user of a specific data set must be able to verify the data by reading the reference and so decide about the quality and usefulness of the data.

A general quality check is supplemented by an evolving system of generic and parameter specific validation routines during the import of data. To improve the data consistency, data sets can be stored in different levels of their processing (primary data, e.g. counts; secondary data, e.g. percentage; tertiary data, e.g. paleotemperature). The archiving of the raw data allows recalibration or new interpretations of data sets. The definition of new parameters requires a given standard unit and a defined range (minimum, maximum) of the data, thus data have to be imported in the same units, the range can be used by the validation routines to find outlyers. These routines are a further step in improving the data quality and consistency in general.

Whenever dealing with the publication and archiving of data, the copyright has to be considered (Diepenbroek and Reinke 1995). If an information system also stores unpublished data, it is crucial for the acceptance and the trust in the database that the data are protected by a hierarchical system, which can be organized and controlled by the user. The owner of a specific data set is the data producer (principal investigator) who must be able to either give copyrights to individual users/groups or open data sets system wide. When using foreign data, a given reference has to be cited. For unpublished data the principal investigator has to be asked.

Data Model

The great variety of parameters, methods, calibrations, and interpretations in the field of paleoenvironmental reconstruction, as well as the modification of established methods or the invention of new interpretations are major obstacles for the integrative use of data sets. The challenge to manage these heterogenic and dynamic data has to be solved through a flexible and simple relational data model (Fig. 1).

The structure of SEPAN is shown as a graphic on the opening screen of the client software. The graphic is the entrance to all levels, tables and tools just by selecting the required field. The structure reflects the standard processing steps for paleoclimatic data. Lists including standardized metainformation are connected to the main data fields (e.g. gear, ship, method). Different institutes/projects (PROJECT), working in the field of paleoclimatology carry out expeditions (CRUISE) to take samples. During a cruise at a number of sites (STATION) different samples may be taken or certain equipment is used (CORE). At distinct points/intervals the sample (which is a sequence of sampling material, e.g. a core from ice, coral or sediment) is sub-sampled for different investigations (PROCESSING). Down to this level, all data are considered to be metadata. From each sample one or more 'real' analytical data will be produced which can be found on the DATA level. From this field, data can be exported as a table or plotted (Fig. 2).

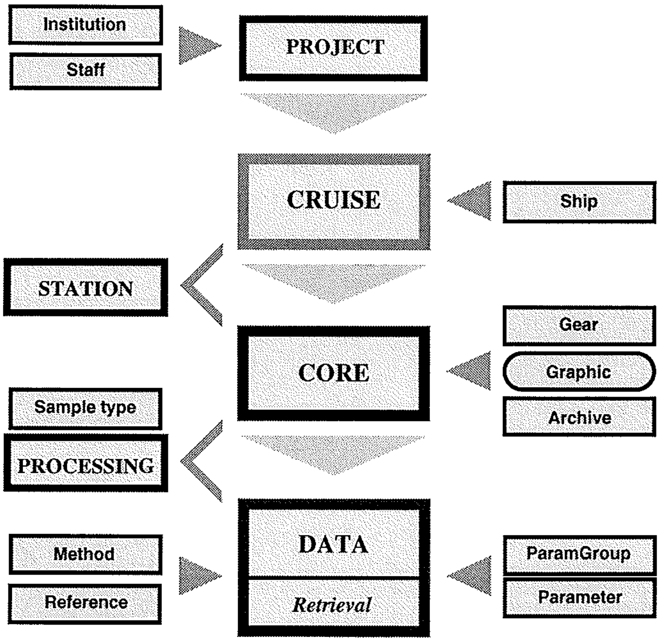

Figure 1: Data model of the Sediment and Paleoclimate Data Network (SEPAN). The hierarchy follows the typical route used to process samples and to produce analytical data. The essential part of the model is the combination of DATA, PARAMETER and METHOD which allows the definition of an unlimited number of parameters.

The parameters are gathered into parameter groups for a better overview. Data types are defined as primary data (e.g. weights, counts), secondary data (e.g. %, m/sec) and tertiary data as they are produced using specific algorithms from the secondary data (e.g. paleotemperatures). Calibration algorithms are stored as method information. The combination of the DATA, PARAMETER and METHOD fields is the essential part of the model, which allows the definition of new, unique parameters by the user at any time. Data types can be numeric, string or graphic.

The Network

In a first step, the main instituts in Germany working in the field of paleoceanography should be connected to the network. SEPAN uses client/server technology through the different local networks and the Internet (Fig. 3).

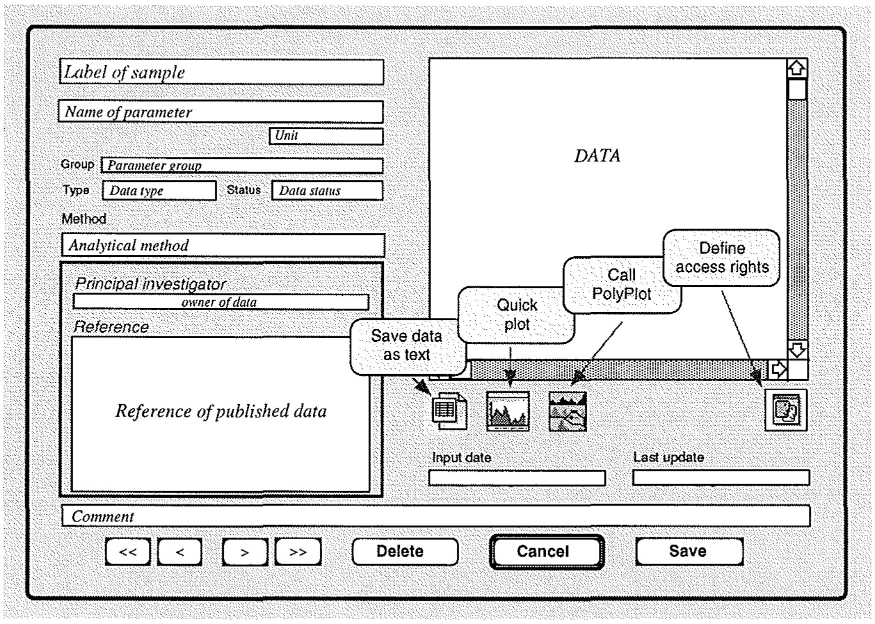

Figure 2: The DATA field of SEPAN. Related to the analytical data are metadata about the method used, the principal investigator and the reference where the data are published. Detailed information is given with the parameters name. Buttons allow an export of the data as text or graphic and the definition of access rights to other users.

The main server, located in the computer center of the AWI, is connected through the Internet with the different external instituts (remote site). The clients, which are the personal computers of the scientists, are connected through the local network with the sub server of the related institute. To increase access speed all metainformation is mirrored on each local sub server (50 000 sampling sites, as imported by 1996, need about 100 MB hard disk capacity). The mutual update of newly imported metainformation is made in the background through the network. The update is based on optimistic strategies, thus avoiding the problems connected with the handling of complex data dictionaries. With the development of reliable high speed electronic networks this feature may become obsolete. The first members of the network being connected in April 1996 are the Institutes in Kiel and Bremen. Further remote installations are planned for other institutions in Germany and Europe (Tab. 1)

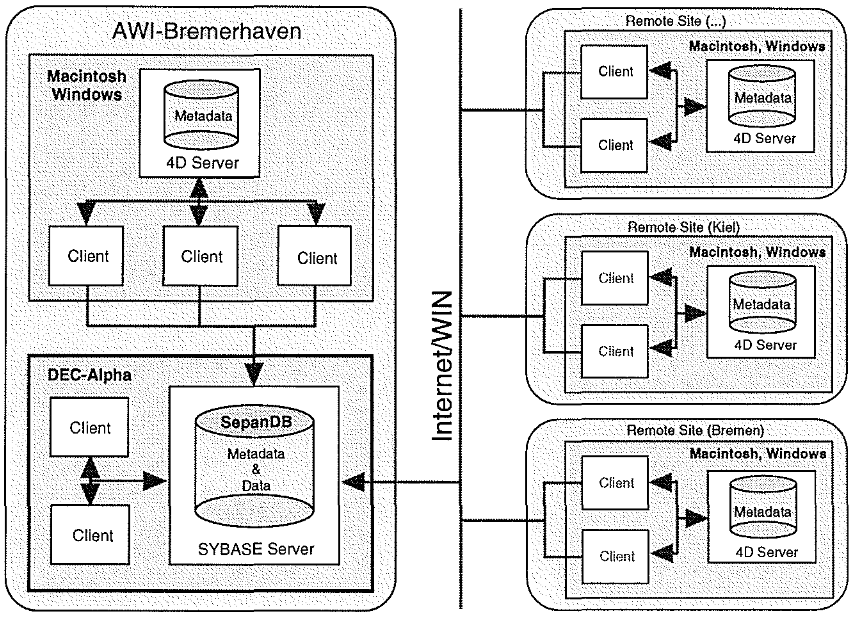

Figure 3: The Network concept of SEPAN. The client/server system consists of a main server in the data center (AWI) and a sub server in each institute connected to the network via Internet. The client computers of the scientists are connected through the local network with the subserver. Updates are made in the background.

The main server, located in the computer center of the AWI, is connected through the Internet with the different external instituts (remote site). The clients, which are the personal computers of the scientists, are connected through the local network with the sub server of the related institute. To increase access speed all metainformation is mirrored on each local sub server (50 000 sampling sites, as imported by 1996, need about 100 MB hard disk capacity). The mutual update of newly imported metainformation is made in the background through the network. The update is based on optimistic strategies, thus avoiding the problems connected with the handling of complex data dictionaries. With the development of reliable high speed electronic networks this feature may become obsolete. The first members of the network being connected in April 1996 are the Institutes in Kiel and Bremen. Further remote installations are planned for other institutions in Germany and Europe (Tab. 1).

Tests of the client/server connections through the Internet have shown that this system would presumably run in Europe with only one main server, but response times for a connected client/server system in the United States for example are too slow. If scientists in the US want to use the system, as already announced, a mirror site, could be set up at, for example, the NGDC/WDC-A (National Geophysical Data Center/World Data Center A for Paleoclimatology) in Boulder, Colorado or in one of the main core repositories.

Hardware and Software

The main database server is a DEC Alpha 8200 (4 processor, 1 GB internal memory, 50 GB hard disc capacity) running SYBASE Version 11 under DEC/UNIX as the database software. The client software for access to the server was written in 4th Dimension (4D of ACI). 4D allows the design of graphical user interfaces and compilation of the front end software for all processors recently found in PC's (68k, RISC, 486, Pentium). Thus client software is available for Macintosh and Windows systems as well.

The client software was modularized into a database front-end together with attached tools individually developed for the processing of specific data sets. The modularization and the open environment facilitates future adaptation of the system. The entry requirements for the handling of the software are low because the functionality is uniform for all tables and tools.

System requirements for running SEPAN at an institute connected to the network are a fast Macintosh or Windows computer with at least 40 MB of RAM which is to be used as the sub server system only. The server has to be connected via the local network to the clients and the local network needs access to the Internet. Additionaly, a number of licenses for the 4D sub server software is needed depending on the number of clients.

Tools

The import of metadata is organized through well defined form tables, therefor raw data files have to be edited by the data curator for the import. Metadata form tables are defined for the import of references, cruises, stations/cores, archive information and the processing/sampling of cores; for the import of data the name or ID of the relevant parameter is used in the header of the input matrix. The related metainformation (method, owner, comments) have to be defined prior to the import and are also updated during the import procedure.

The retrieval tool for finding and extracting data from the system is uniformly designed for all levels and allows complex combinable search criteria relevant to the desired data. Data can be exported as spreadsheets or plotted with one of the graphic tools. The spreadsheets can be sorted and configured individually. In general the functionality of the database frontend comprises import and export of data in common exchange formats (Text, NetCDF, NASA-DIF for the metainformation).

The main retrieval function for extracting multidimensional data sets is found at the bottom of the DATA field (Fig. 1). The retrieval button starts the spreadsheet tool, a routine which allows the retrieval, display and export of any combination of analytical data, cores or ages (e.g. time slices). The matrix can be configured to be equivalent to the export functionality on the upper levels, including columns of metadata information (latitude, longitude etc.) and any combination of parameters (up to 100 columns). The analytical values are plotted versus depth. Multiple data sets can either be displayed with identical parameters and locations in one column, or the data can be split by data sets and location into separate columns, thus allowing the comparison of data sets from different investigators or multiple versions for one data set. The data can be downloaded into a textfile or plotted graphically with PolyPlot. If the geographical information is included in the matrix, values can also be displayed in a map using MacMap or data can be exported in formats for appropriate contouring tools.

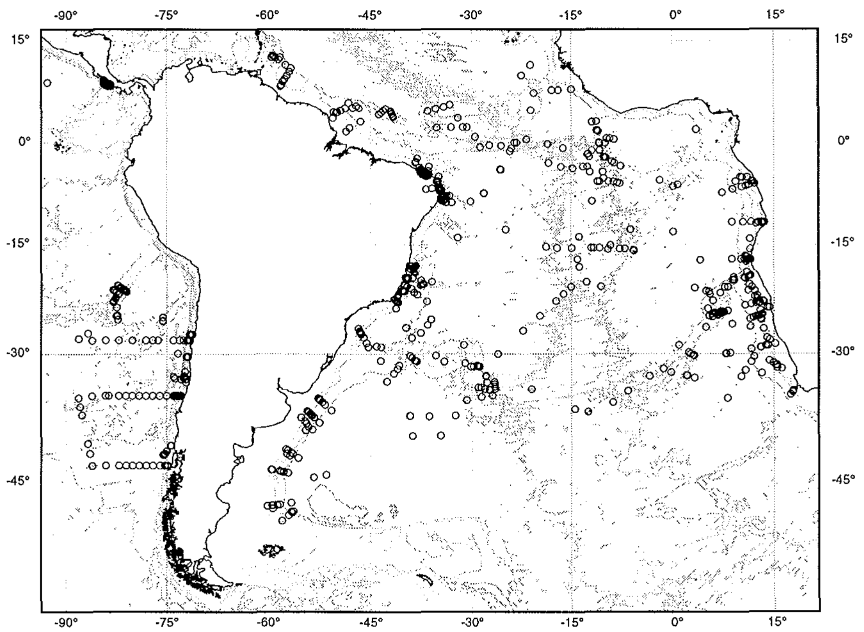

Figure 4: SEPAN tool MacMap can be called by the information system to plot metadata or analytical data in a geographical context. The graphical attributes can be changed, different projections are available. The map shows core sites in the South Atlantic (circles) taken by the Fachbereich Geowissenschaften, Bremen, during different cruises with the German research vessel METEOR.

For geographical presentation of data the SEPAN tool MacMap was developed, which is either directly connected to the database front-end or can be used as a stand-alone application (Fig. 4). MacMap can be called directly by the SEPAN user to draw sampling sites in a geographical context. Sites can be labeled with metadata (e.g. core length, water depth or core label) as well as analytical data (e.g. carbonate content, percentage of a specific mineral or microfossil). The user selects the required data set by the retrieval tools in SEPAN. After calling MacMap from SEPAN the data set is transferred to the mapping software to be plotted.

Maps can be configured with different projections, styles of map elements can be changed, individual vector data or site information can be imported. The bathymetric data source used in the maps is GEBCO (General Bathymetric Chart of the Oceans) (IOC et al. 1994). A high resolution bathymetry of the Baltic Sea, published by the Institute for Baltic Sea Research, Warnemünde, is also available. To plot the topography of continents a vectored version of the ETOPO data can be used. Maps can be exported in common graphic exchange formats or as postscript files.

Samples, taken from a natural climatic archive (e.g. a coral head), are space oriented and the distance between samples is related to time. Thus the database needs a tool for the visual inspection of the paleoclimatic data which is able to plot multiple data sets at uniform scales versus time or depth. This was realized with the SEPAN tool PolyPlot , which can also be directly called from the database frontend or used stand-alone (Fig. 5). Scales and graphic features can be modified by the user and distinct parameters can be selected from the data matrix, which was retrieved and transferred from SEPAN. PolyPlot graphs can be exported in platform specific interchange formats.

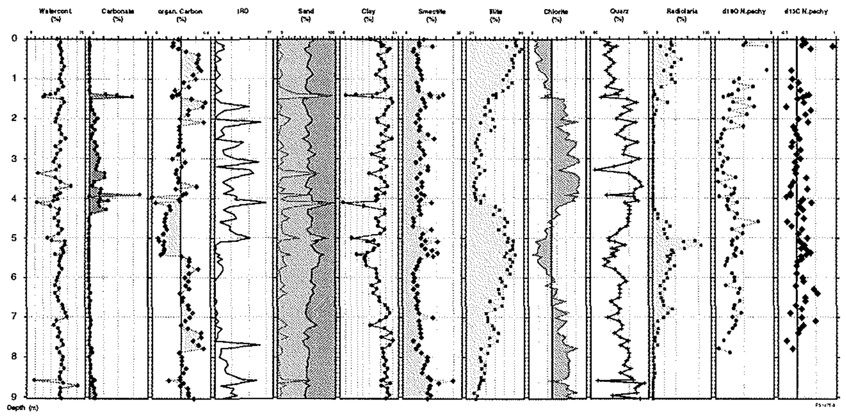

Figure 5: SEPAN tool PolyPlot was designed to plot up to 60 parameters versus depth or time. The software can be used by SEPAN after retrieval or as a stand alone application. Shown are sedimentological data of a sediment core from the Antarctic continental margin, which are used for paleoclimatological interpretations.

The methods of scientists who reconstruct oceanographic conditions of the former oceans (paleoceanography) require access to the oceanographic data of the recent oceans. The NODC (National Oceanographic Data Center, Washington) provides these data in the World Ocean Atlas 1994 (Levitus et al. 1995), (NODC 1994) as one-degree latitude-longitude mean fields for temperature, salinity, oxygen, and nutrients in standard depth levels. SEPAN allows access to these data sets from the CORE and

DATA levels to obtain the oceanographic data closest to the site of sampling.

The SEPAN software, with its different tools will be available for Macintosh and Windows. The sub-server system in an institute connected to the network, can also be a Macintosh or Pentium computer running Windows NT, depending on the preferred systems in the institute. The SEPAN software will also be available on CD to be used as a stand alone application, providing access to the metainformation only. The CD includes manuals, references and all tools with the related data sets as described above.

The Data

The available metadata comprises related information about expeditions, sampling sites/sets and storage facilities. All scientists and instituts, related in some way to the information system are stored with full address and email. Related items, such as the names of ships, gears or sample types are defined in lists which are regularly updated. As of 1996 SEPAN can store analytical data of about 1500 parameters which are organized in 26 groups. Accessible from the core field, the core description is stored as a graphic in an interchangable format. Analytical methods can be defined with all necessary information. Reference for cruise reports or published data can be defined with the fields 'author', 'year', 'title' and 'source'. SEPAN does not have the functionality of an enhanced reference database, but all references are imported through the bibliography software EndNote Plus. Thus an EndNote Plus file is delivered together with SEPAN, allowing the user to extract references for the use in his own reference system.

Besides the sample label the most important metadata for all sampling sites are the location (latitude/longitude) and the water depth. Because the SEPAN software is based on the archiving software of the Polarstern sediment core repository, the first step was to import all metadata related to Polarstern cores and sediment samples. This set was used to test the metadata related fields of the system. In summer 1995, the storage of metadata from other core repositories started, the first being Bremen and Kiel. As a reference to the international coring projects it was also important to include the sites of DSDP and ODP.

In 1995 the NGDC/WDC-A published the prerelease of the Index to Marine Geological Samples on CD (Moore 1995). From this source the metadata of cores from six of the largest core repositories of the United States were extracted and imported into SEPAN (Tab. 1). Cores and samples from these repositories are internationally used and analytical data were also placed in SEPAN. Importing complete collections makes the import of analytical data much more efficient, because the data curator does not have to take care about any further metadata prior to the import of data of a specific core.

It is the policy of PAGES to encourage scientists to participate in international data banks. The World Data Center-A for Paleoclimatology (WDC-A), established in 1992 (Webb et al. 1994) , is located at the US. National Geophysical Data Center (NGDC) Paleoclimatology Program,Boulder, Colorado and is working with the IGBP/PAGES to coordinate a science-driven management system for the acquisition and distribution of all types of paleoenvironmental data needed by the international research community. A first step in this process is to ensure that all published data are archived in a digital format and are readily available to the community. Data, submitted to the center, are organized on a ftp-server and are distributed via magnetic tape and on-line access via Internet.

SEPAN will collect any data of a specific scientific field and store them in a consistent format. The structure and data model of SEPAN will allow the user to start comprehensive retrievals to extract data sets for specific requirements. SEPAN combines the information about sampling material with the analytical data resulting from this material and allows complex access to all combinations of metadata and analytical data and thus will provide a new scientific tool for solving questions to advance our understanding of Earth system dynamics.

Acknowledgments

The help and contributions of Rainer Sieger, Reiner Schlitzer und Uwe Siems are greatly acknowledged. Chris Dodge improved the paper through an english review. This project was financed by the German Ministry of Education, Science, Research and Technology (BMBF), Fund No: 03F0131B (Paläoklima-Datenzentrum für die marinen IGBP/PAGES-Daten). This is Publication No: 1091 of the Alfred Wegener Institute for Polar and Marine Research and Contribution No. 130 of the Sonderforschungsbereich 261 at Bremen University. Further information and a stand alone version of SEPAN on CD is available on request at sepan@awi-bremerhaven.de. PolyPlot and MacMap are available through the anonymous ftp-server of the Alfred Wegener Institute (ftp.awi-bremerhaven.de, path: /pub/geo/SEPAN).

REFERENCES

Diepenbroek M., Reinke M. Publishing scientific data - a strategy for the integration of heterogenous and dynamic data environments. IGBP Informationsbrief, 1995;19:7-9

Emiliani C. Pleistocene temperatures. J Geol 1955;63:538-578

Epstein E.F. Legal and institutional aspects of global databases. In Building Databases for Global Science, Mounsey H, Tomlinson R, ed. Taylor & Francis, 1988.

Gilb T. Principles of Software Engineering Management. Addison Wesley, Reading, Massachusetts, 1988.

IGBP. The International Geosphere-Biosphere Programme: a study of Global Change, the initial core projects. Global Change Report, 1990;12

IOC, IHO, BODC. Supporting volume to the 'GEBCO Digital Atlas' published on behalf of the Intergovernmental Oceanographic Commission and the International Hydrographic Organization as part of the General Bathymetric Chart of the Oceans (GEBCO). British Oceanographic Data Centre, Birkenhead, 1994;with CD-ROM

Levitus S., Conkright M.E., Gelfeld R.D., Boyer T. World Ocean Atlas 1994. IGBP Newsletter, 1995;20:4-6

Moore C. Index to marine and geological samples. National Geophysical Data Center, World Data Center A for Marine Geology & Geophysics, 1995;pre-release on CD-ROM

NODC. World Ocean Atlas 1994. National Oceanographic Data Center, 1994;Informal Report No.13:set of 9 CD-ROM

Webb R.S., Anderson D.M., Overpeck J.T. Archiving data at the World Data Center-A for Paleoclimatology. Paleoceanography 1994;9(3):391-939

Tables

Table 1. Institutions participating in the SEPAN-network. Shown is the number of sites of the different repositorys stored in SEPAN as of July 1996, which, in total are metadata of 28 785 locations with 45 781 gears used on 1060 cruises. (CCI= Core Curator Index; Moore, 1995)

Institution	No of sites	Source of meta information	Participation
Akademie der Wissenschaften, Universität Heidelberg			discussed
Alfred Wegener Institute for Polar and Marine Research, Bremerhaven	1870	AWI	connected
Alfred Wegener Institute for Polar and Marine Research, Potsdam	226	AWI	connected
Antarctic Research Facility, Florida State University, Tallahassee	2272	CCI	in progress
Atlantic Geoscience Center, Bedford Institute of Oceanography, Canada		AGC	
Australian Geological Survey Organisation, Canberra	51	AGSO	
Baikal Drilling Project	1		in progress
Biogeochemical Ocean Flux Study, British Oceanographic Data Center	46	BOFS	
British Antarctic Survey			discussed
Bundesanstalt für Geowissenschaften und Rohstoffe, Hannover	1361	BGR	in progress
Cape Roberts Project			in progress
Cooperation with Russian Institutes	391	AWI	
Deep Sea Drilling Project	620	DSDP	
Department of Geology, Oregon State University	2667	CCI	
Fachbereich Geowissenschaften, Universität Bremen	778	FGB	connected
Geochemical Ocean Section Studies			
Geoforschungszentrum Potsdam			discussed
Geologisch-Paläontologisches Institut, Christian-Albrechts-Universität, Kiel	2446	GIK	connected
Geologisch-Paläontologisches Institut, Universität Göttingen			discussed
Geologisch-Paläontologisches Institut, Universität Tübingen			discussed
GEOMAR-Forschungszentrum für marine Geowissenschaften, Kiel			in progress
Institute für Biogeochemie und Meereschemie, Universität Hamburg			discussed
Institut für Ostseeforschung, Warnemünde			in progress
Institut für Petrographie und Geochemie, Universität Karlsruhe	121	IPGK	
Lamont-Doherty Earth Observatory, Columbia University	7878	CCI	
Ocean Drilling Program, College Station	323	ODP	
Rosenstiel School of Marine & Atmospheric Sciences, University of Miami	1453	CCI	
Scripps Institution of Oceanography	3455	CCI	
Southampton Oceanographic Center			in progress
Woods Hole Oceanographic Institution, Department of Geology & Geophysics	2451	CCI	
World Data Center-A for Paleoclimatology, Boulder, Colorado, USA			cooperation