

In. Albritton, K D ; Bradley, R S. & Pedersen, T F (eds)
 Paleoclimate, Global Change and the Future 220 pp.

Appendix B - The PAGES Data System

Sprunger, Berlin
 2003

C. M. Eakin

NOAA/National Geophysical Data Center, World Data Center for Paleoclimatology, 325 Broadway E/GC,
 DSRC 1B139, Boulder, CO 80305-3328, U.S.A

175-120

M. Diepenbroek

Alfred Wegener Inst. for Polar & Marine Research, Am Alten Hafen 26, DE-27568 Bremerhaven, Germany

M. Hoepffner

CNES, MEDIAS- FRANCE, BPI 2102 18 avenue E. Belin, FR-31401 Toulouse-Cedex 4, France

B.1 The PAGES data system and its components

When PAGES first began, centralized management of paleoclimatic data rested in the World Data Center (WDC) for Paleoclimatology (then called WDC-A for Paleoclimatology) and in disparate but burgeoning data management efforts. By the second PAGES Data Meeting in 1998 (Anderson and Webb, 1998), additional major data management efforts were taking shape. The PAGES Data Guide highlighted three main components with important roles in data management:

- Scientists, who generate, publish and contribute the paleoclimatic data
- Data cooperatives, project-level data management, thematic and regional archives, and
- The World Data Center for Paleoclimatology (and regional mirrors) that provide the long-term archive and access to the data.

major archive for paleoceanographic data and for major European paleoclimatic projects. The MEDIAS-France group provides archive and access to a variety of European paleoclimatic data efforts and plays a crucial role in developing data management and scientific infrastructure in Africa. Many other centers are developing data sharing for a wide range of paleoclimatic data. The diversity of the PAGES Data System has increased dramatically, and with it so has the challenge. All components of the PAGES Data System work to a common goal: to support the data management needs of the International Geosphere-Biosphere Programme (IGBP) core project on Past Global Changes (PAGES) and its parent body, the International Council for Science (ICSU).

B.2 The World Data Center for Paleoclimatology

The World Data Center for Paleoclimatology in Boulder, Colorado (WDC-Paleoclimatology, <http://www.ngdc.noaa.gov/paleo>) is dedicated to providing the paleoclimatic data and information needed to understand and model interannual to centennial scale environmental variability. Focused on the principle of science-driven data management, WDC-Paleoclimatology relies upon the expertise of scientists at universities and institutions around the world to produce and share scientific data and information. Committed to the ICSU (International Council for Science) principle of full and open exchange of data and information for scientific and educational purposes, WDC-Paleoclimatology works to both enhance the compilation of paleoclimatic data within its holdings and to improve the accessibility and usability of these data. All data holdings are freely available via the Internet.

WDC-Paleoclimatology is the international component of the National Oceanic and Atmos-

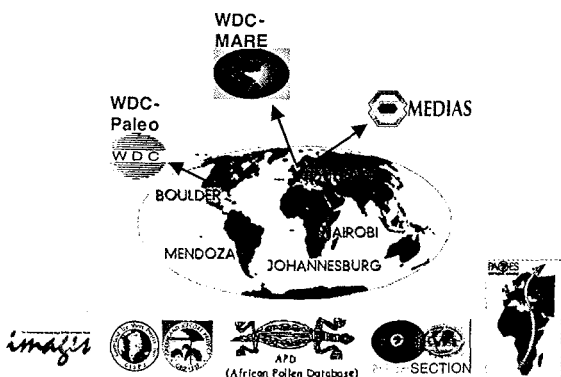


Fig. B.1. Many of the data centers and data activities that are part of the PAGES Data System.

Today, additional major data management efforts play essential roles in archiving and distributing paleoclimatic data (Figure B.1). The German PANGAEA group, host to the World Data Center for Marine Environmental Sciences serves as a

pheric Administration (NOAA) Paleoclimatology Program and is located at NOAA's National Geophysical Data Center in Boulder, Colorado. The NOAA Paleoclimatology Program and the WDC-Paleoclimatology has worked for over a decade to support the data management needs ICSU, IGBP-PAGES, and the U.S. Global Change Research Program (USGCRP). All three organizations have established similar policies promoting the free and open exchange of scientific data.

WDC-Paleoclimatology serves as the home to many large archives of paleoclimatic data, including the International Tree-Ring Databank, the Global Pollen Database, Global Database of Borehole Temperatures and Climate Reconstructions, the International Ice Core Data Cooperative and data for the PAGES project on Annual Records of Tropical Systems (ARTS). In addition it holds paleoclimatic data from most every discipline, including climate forcings, corals and sclerosponges, paleoceanography, paleolimnology, pollen and plant macrofossils, fauna, insects, model data, and reconstructions. It mirrors of data from the Paleoclimate Modelling Intercomparison Project and provides visualization tools for model data and a variety of paleoclimatic reconstructions. WDC-Paleoclimatology distributes free software and tools produced by many researchers. Major efforts are placed on tools for data browse and visualization (e.g., WebMapper) and providing information products such as the "Paleo Perspectives".

B.3 The World Data Center for Marine Environmental Sciences / PANGAEA

The World Data Center for Marine Environmental Sciences (WDC-MARE, <http://www.pangaea.de>) is aimed at collecting, scrutinizing, and disseminating data related to global change in the fields of environmental oceanography, marine geology, paleoceanography, and marine biology. WDC-MARE uses the scientific information system PANGAEA (Network for Geosciences and Environmental Data) as its operating platform.

Essential services supplied by WDC-MARE / PANGAEA are project data management (e.g. for the PAGES project IMAGES, the International Marine Global Change Study), data publication, and the distribution of visualization and analysis software (freeware products). Organization of data management includes quality control and publication of data and the dissemination of metadata according to international standards. Data managers are responsible for acquisition and mainte-

nance of data. The data model used reflects the information processing steps in the earth science fields and can handle any related analytical data. A relational database management system (RDBMS) is used for information storage. Users access data from the database via web-based clients, including a simple search engine (PangaVista) and a comprehensive data-mining tool (ART). With its comprehensive graphical user interfaces and the built in functionality for import, export, and maintenance of information PANGAEA is a highly efficient system for scientific data management and data publication.

WDC-MARE / PANGAEA is operated as a permanent facility by the Centre for Marine Environmental Sciences at the Bremen University (MARUM) and the Alfred Wegener Institute for Polar and Marine Research (AWI) in Bremerhaven, Germany.

B.4 MEDIAS-France

Located in Toulouse, MEDIAS-France (<http://medias.obs-mip.fr:8000/>) is a non-profit public corporation that works to develop cooperative research projects, set up permanent observation systems, build up data banks, develop models, and train and provide exchanges for students and researchers. MEDIAS-France has built databases in a wide range of scientific disciplines relating to climatic changes in the global environment, primarily paleoclimatology, hydrology, atmospheric chemistry in tropical region, meteorology (with rainfall forecasts covering North Africa) and oceanography (both dynamic and biological aspects). MEDIAS-France is also skilled in drawing up catalogues, developing integrated or specialized databases and educational products, and distributing its products. In addition MEDIAS-FRANCE provides training services to the scientific community, especially international institutional development such as: the SysTEM for Analysis, Research and Training (START), ACMAD, the Sahara-Sahel Observatory, ENRICH, IGBP/DIS, etc. An important effort in developing START activities is directed at the Mediterranean and Africa, where Planning Committees have early been established. MEDIAS and START have evolved in close partnership, jointly sponsoring several activities. ENRICH promotes collaboration in Western Europe, encourages the endogenous research capabilities in developing countries, including in Africa and the Mediterranean Basin, and promotes support for relevant research initiatives in the countries of Central and

Eastern Europe and the New Independent States of the former Soviet Union (NIS).

MEDIAS-France develops and supports several major paleoclimatic database activities including the European and African Pollen Databases, Forest Modelling Assessment and Tree Rings (Format), and the European Diatom Database (EDDI, coordinated by Dr. S. Juggins, University of Newcastle). It distributes paleoclimatic data analysis software and hosts one of the mirror sites for WDC-Paleoclimatology.

B.5 Mirror sites / World Data Center partners

While the major centers provide ready access to data and information, they can only work as well as the users' connection via the Internet. This is one of the reasons that WDC-Paleoclimatology has worked with partners in various countries to establish mirror sites (Figure B.2). These distributed archives hold complete copies of the web and ftp holdings of the WDC-Paleoclimatology, providing regional access points to the data. This means that data can be made more accessible, benefiting scientists in the region of the data's origin. Bringing the data close to home encourages scientists to contribute their data – increasing participation in PAGES-organized regional to global efforts to understand our climate system. Currently four mirror sites are operating, with more under consideration.

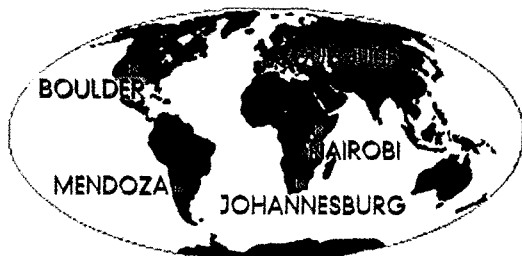


Fig. B.2. Map showing the locations of the World Data Center for Paleoclimatology, Boulder and its data mirrors around the world.

The idea of mirror sites has taken hold within the ICSU structure. The World Data Centers Panel has decided to begin new approaches to expand global data sharing. One of these, World Data Center *Partners*, entails the collaboration of a World Data Center with partner organizations in developing countries. The approach is intended to encourage data exchange without requiring the infrastructure commitments entailed in establishing new WDCs. While details of this program are

still under development, the activities of the PAGES Data System and the WDC mirror sites established the model for this new ICSU program.

B.6 Data cooperatives and project level data management

A key step between the scientists producing data and centralized data management are the data management efforts focused on disciplinary or regional programs. As each paleoclimatic data stream has individual characteristics and idiosyncrasies, data cooperatives have provided the expertise needed to develop proper protocols for handling data and metadata. These collaborations between expert university scientists, disciplinary data managers, and data management centers establish data management protocols, resolve taxonomic, methodological, dating, and data quality problems. These provide the procedures that scientists and data managers use for managing the data in the future. The role of data cooperatives and the activities of many of these were described in *The PAGES Data Guide* (Anderson and Webb 1998). Some of these include large international programs such as the International Ice Core Data Cooperative (for GRIP and GISP2 data), and data management groups for international programs such as the Paleoclimate Modelling Intercomparison Project (PMIP) and the International Marine Past Global Changes Study (IMAGES). Other data cooperatives pull together large but less structured groups of scientists and data such as the International Tree-Ring Data Bank and the Global Pollen Database. Once the data coop has developed appropriate data management procedures, the data are contributed to WDC holdings for permanent archive and access.

B.7 Users of paleoclimatic data

While the primary users of paleoclimatic data are still members of the paleoclimatic research community, a much broader suite of users is now realizing the value of pre-instrumental records of climate. The most active of these have been the modern climate research and prediction communities. However, in recent years a growing number of policy makers, planners, and resource managers are using paleoclimatic data in planning and decision support. Additionally, paleoclimatic data now are accessed by a broad array of educators, students, the media, and the general public. While the research community mainly comes to the PAGES Data System to access data, these newer users

typically seek information products that make past climate changes more understandable. In all, users in 147 countries have accessed data from the WDC-Paleoclimatology holdings (Figure B.3, as of December 2001).

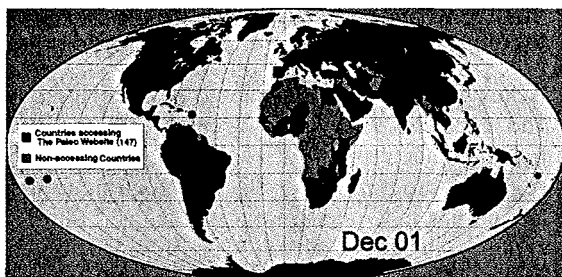


Fig. B.3. Countries that have downloaded paleoclimatic data from the World Data Center for Paleoclimatology, Boulder.

B.8 Data access and information tools

Archived data are valuable only if they are accessible and used. For this reason, members of the PAGES Data System work to make sure that their data holdings are readily accessible. Text based search engines across the PAGES Data System provide the user with the ability to define a variety of search parameters to find desired data. These include searches by contributor, proxy, variables or expedition. Advanced text-based search tools such as PANGAEA's Advanced Retrieval Tool enables the user to retrieve and download data using user-defined configurations.

In addition to text-based information search and delivery tools, new geospatial tools have proven themselves as valuable ways to access the data. The WDC-Paleoclimatology's WebMapper (Figure B.4), a web-based browse, visualization and access tool and PANGAEA's PangaVista (Figure B.5) search engine are powerful new ways to search and access data from PAGES Data System archives.

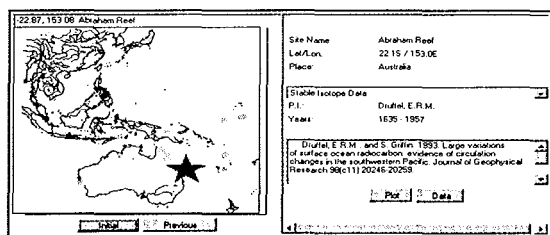


Fig. B.4. The WDC-Paleoclimatology's WebMapper

For gridded data, online plotting tools have been applied to data from paleoclimatic models, and reconstructed temperature, drought and pressure

fields to provide users with visualizations of these large and complex datasets. Many data centers are applying new tools from the realm of Geographic Information Systems (GIS). GIS systems allow users to access a variety of different data types and superimpose them as "layers" to access and analyze disparate data.

So far each of the tools described above provides access to the holdings of only one of the PAGES data centers. New tools under development will improve access by allowing users to access data from multiple data centers, from a variety of proxies, and within selected time slices. By sharing metadata among centers, the PAGES Data System will provide a paleoclimatic data "portal" to make data from multiple centers more accessible.

While data access tools such as search engines, WebMapper and portals make data more accessible; many users come to PAGES seeking interpretations of the data. By providing information products (e.g., WDC-Paleoclimatology's *A Paleo Perspective on Global Warming*), PAGES data centers have provided access to paleoclimatic data in a format that explains the importance and meaning of many of these data.

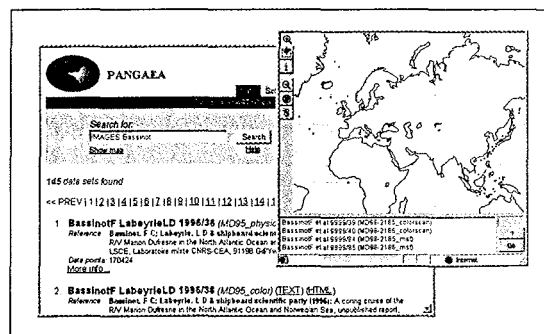


Fig. B.5. PANGAEA's PangaVista

B.9 The PAGES Data Board and data sharing

The successful development of the various data efforts within the PAGES Data System has archived more data than was possible for any one center. However, this has led to some confusion on the part of the scientific community and other users. A recent editorial in *Nature* referred to the current situation as "something of a maze of publicly supported databases" (Anonymous, 2001). To address the need for greater coordination among paleoclimatic data efforts, the PAGES Data Board was formed. The PAGES data board is an open and equal partnership between international, national and thematic groups that archive paleoenvironmental data and provide tools to

enhance their value. Its duties are to

- ensure common metadata formats
- develop a shared paleodata web portal
- promote data sharing, and
- recommend PAGES data policies

As a part of the IGBP, PAGES supports the free and open exchange of data as described by policies of IGBP and its parent organization ICSU (CODATA 2002). In particular, the PAGES Data Board has established policies that support the development and use of rules of good scientific practice (ESF, 2000), including

- making data and methods available for reproducibility of results,
- making data behind any published graphic or figure publicly available, and
- ethical use of data, including proper citation.

The data portal described above is being planned and implemented by the various PAGES data centers and its development and implementation is being facilitated through the PAGES Data Board. This includes the recent adoption of standards for archiving metadata through a standard profile and then exchanging them using extensible markup language (XML).

B. 10 Future directions

New data access tools and methodologies continue to revolutionize the options for managing paleodata. Members of the PAGES Data System are working to advance development and application

of databases, web portals and GIS technologies that will enable users to access, combine, and analyze data from a wide array of datasets across the Internet. This will allow data users to pinpoint tree-ring sites on digital elevation models, or to find lake-data from sites within drought-prone regions or time periods of interest. Implementation of standard metadata profiles will increase data sharing among data repositories, while XML and other methods to exchange metadata and data will facilitate data interchange and reuse. We plan to implement new data cooperatives, disciplinary and multidisciplinary data efforts to develop protocols for handling new proxies.

In a recent *Nature* correspondence, Alverson and Eakin (2001) pointed out that "PAGES, WDC-Paleo and the many scientists and institutions that support them have made great efforts to make data easily accessible and usable." As a community, we can all be proud that data sharing in paleoclimatology has come a long way over the last decade.

The world of data management is changing rapidly. As new developments and new ideas come forward, the PAGES Data System will continue to advance technology to manage and distribute paleoclimatic data and information. Through participation of the paleoclimatic research community, disciplinary data groups, and data management centers, the PAGES Data System will continue to make paleoclimatic data as accessible and usable as possible.