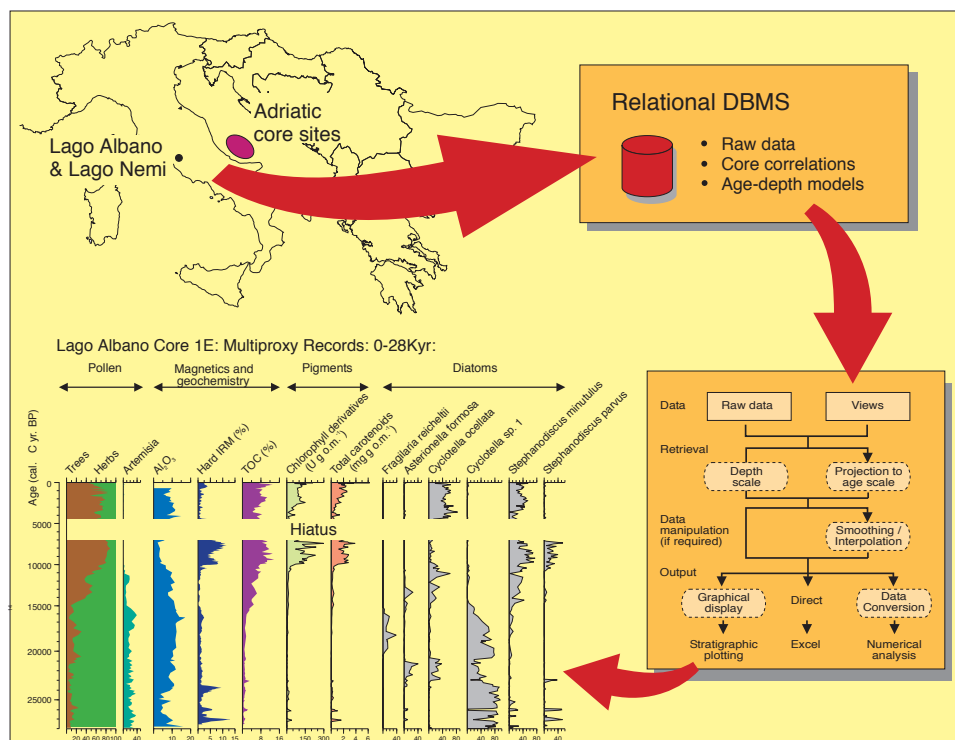


A PEP III Multi-Proxy Database for Managing and Analyzing Paleoenvironmental Data



Schematic outline of the PALICLAS database

The Pole-Equator-Pole (PEP) projects are inherently multi-site and multi-proxy, requiring the integration and analysis of different paleoenvironmental indicators recorded from regional or continent-wide site networks. Consequently, there is a need to develop an effective data management strategy that allows information from different sources to be integrated, harmonized and retrieved in a consistent format. This article describes a recent attempt to construct a multi-proxy database within the PALICLAS project, and outlines plans for its development as an integrated, scalable system for managing project data within the PEP III transect.

The PALICLAS project focused on the analysis of Italian crater lake and Adriatic sediments, with the aims to provide improved high-resolution proxy paleoclimate records for central Italy for the last 20k+ years, and to characterize the responses of both terrestrial and aquatic ecosystems during periods of rapid climatic change. Central to these aims was the need to make direct comparisons between the different biological, chemical and physical proxy records recorded in marine and lacustrine sediments. To accomplish this we constructed a PC-based data management system consisting of a relational database to store all raw data from the project, and a series of software

tools for data manipulation, visualization and analysis.

The essential features of the system are illustrated in the figure. All physical, chemical and biological information is uniquely referenced to a core and depth interval, and records may be retrieved as raw data (e.g. microfossil counts) or transformed values (e.g. percentage data) using saved queries or views. These data may be retrieved on the original depth scale, or, for between-core comparisons, transformed to the depth scale of a "master" core or to an age scale, using core-correlation and age-depth models held in the database. The resulting single- or multi-proxy records may then be output directly, plotted using built-in visualization software, or smoothed and/or interpolated in preparation for further numerical analysis.

The PALICLAS database was primarily developed as a tool for the rapid comparison and cross-correlation of different proxies within and between sites. However, as the project developed the database performed a number of other important roles. In particular, the system has facilitated:

- 1. quality control**, ensuring consistency and compatibility among data produced by different laboratories;
- 2. data sharing**, by providing a means of

rapidly distributing datasets to all participants as the project developed;

3. basic data manipulation, such as age-depth modelling and core correlation, harmonization of microfossil data recorded at different taxonomic resolutions, and calibration of the various proxy records;

4. data exploration and hypothesis generation, by allowing rapid visualization of multiple proxies on various age or depth scales;

5. hypothesis testing, by preparing the data for numerical analyses to test the significance of observed relationships; and

6. the creation of a long-term archive for all project data.

Our experience in PALICLAS has convinced us that paleoenvironmental databases should be seen not as a repository for data once a project is completed and published, but as an essential tool for data management and analysis as the project develops. In dealing with multiple proxies, from multiple sites, generated by different laboratories, PALICLAS offers a microcosm for PEP III data management. As such, we are now expanding the database structure to handle additional proxy records. Our ultimate aim is to produce an integrated database and data visualization / analysis system that can be distributed freely to laboratories and used to manage data within individual PEP III projects. By adopting a common database structure that is scalable across platforms, it is also intended that the system can be used to integrate data from these different projects as the need arises. To that end, we are collaborating closely with colleagues at the World Data Center in Boulder, to ensure compatibility with the global databases being developed there.

Further details about the PALICLAS project, including a more detailed description of the database can be found in the *Memorie dell'Istituto Italiano di Idrobiologia*, Vol. 55, 1996 (Eds. Guilizzoni and Oldfield), 357 pp, or on-line at <http://www.iii.to.cnr.it>. Details of the developing PEP III data management system can be found on the PEP III web site at <http://www.geog.ucl.ac.uk/ecrc/pep3.htm>.

STEVE JUGGINS

Department of Geography, University of Newcastle, UK

stephen.juggins@ncl.ac.uk