Report

Advisory Board of the European Pollen Database

21-22 September 1990

Niedersächsisches Institut für historische Küstenforschung

Wilhelmshaven, Germany

Participants:

B. Ammann, Bern, Switzerland (1)
J. L. de Beaulieu, Marseille, France (1#)
K.-E. Behre, Wilhelmshaven, Germany (2)
B. E. Berglund, Lund, Sweden (2*)
H. J. B. Birks, Bergen, Norway (2)
E. Bozilova, Sofia, Bulgaria (2)
M.-J. Gaillard-Lemdahl, Lund, Sweden (2#)
E. C. Grimm, Springfield, Illinois, USA (4)
J. Guiot, Marseille, France (3)
B. Huntley, Durham, England (1#)
G. L. Jacobson, Jr., Orono, Maine, USA (2)
C. R. Janssen, Utrecht, Netherlands (2)
A. Pons, Marseille, France (1*)
M. Ralska-Jasiewiczowa, Cracow, Poland (2)
W. A. Watts, Dublin, Ireland (1*)

(1) Executive Committee member
(2) Present member of the Advisory Board
(3) Database coordinator at Marseille
(4) North American Pollen Database representative
* Until 22.9.90
\# As of 22.9.90

Ammann and Gaillard-Lemdahl proposed an agenda, chairpersons, and secretary for the workshop. Three formal sessions were held, Friday evening, Saturday morning, and Saturday afternoon. Chairpersons for the sessions were Watts, Janssen, and Jacobson. Minutes were taken by Grimm. A number of themes were discussed during the meeting, some discussed in more than one session. The following report is abstracted from the minutes and is organized by session and theme. In some cases, the summary of the complete discussion of a theme is consolidated, although the actual discussion was dispersed throughout a session or over more than one session.

Session 1. 21 September 1990, 18.00 hrs, W. A. Watts, Chairperson

Apologies for absence were received from M. Follieri, M. Kabailiene, and J. C. Ritchie.

Pons gave a report of the EPOCH program, which is focussed on global change over the past 30,000 years. One aspect of the EPOCH program is the development of a European pollen database, to be centered at Marseille, for which funding of 105,000 ecu has been obtained for the term July 1990 - July 1993. A postdoctoral candidate will be engaged for 30 months, beginning in January 1991. The position is open, and suggestions for qualified candidates were solicited.

The Marseille group has also obtained approximately 250,000 Ffrs for computer equipment, office equipment, etc. from regional government authorities. Guiot reported on facilities available for the data centre. Office space has been acquired in a research centre in the city of Arle, which is approximately 80 km (about 1 hr) from Marseille. The complex has a central room for computers and meetings and quiet offices for about 5 people. Funds are also available to purchase an 80486 microcomputer, two smaller computers, and a printer.

Grimm reported on the development of the parallel North American pollen database. Funding of $52,500 has been obtained for the first year from the (U.S.) National Geophysical Data Center, an agency in the National Oceanic and Atmospheric Administration (NOAA). NOAA is supporting the development of palaeoclimate databases, including several for palaeoclimate proxy data (fossil pollen, tree rings, ice cores, marine data, etc.). The pollen data centre will be the Illinois State Museum in Springfield, Illinois. The project began 15 August 1990. A computer programmer has been hired to develop database software, and considerable progress has already been made. The objective of the first year is to incorporate the COHMAP database into a relational database.

Gaillard-Lemdahl reported the results of a poll sent to members of the INQUA Eurosiberian and Mediterranean Subcommissions. Of 120 questionnaires sent, 68 were returned. The results were:

91% would provide raw pollen data
57% want to use the database
22% preferred a central database only
76% preferred regional databases
69% preferred a relational database
19% had experience with database software

Concerning data accessibility:

28% thought all data should be public domain
18% thought data should be available to contributors only
63% thought data should be available to noncontributors if contributors were contacted

Session 2. 22 September 1990, 9.00 hrs, C. R. Janssen, Chairperson

Other databases.

Huntley reported on a grant he has received from NERC (Great Britain) for a compilation of European late-glacial pollen data. This database will be developed in full collaboration with the Marseille group and will be a "root"

delivering data to the central database.

Huntley also discussed the origin of the Huntley and Birks data compilation used for the European pollen maps they published. That database consisted mainly of percentages read off published pollen diagrams at certain time slices and contained few original count data. Moreover that database, developed during the 1970's, used homegrown software and was tied to a mainframe computer and to a particular FORTRAN compiler. As a result access to the data was difficult. He emphasized that microcomputers and commercial software now facilitate development of pollen databases having wide availability.

Ammann reported on the organization of the Alpine database. A workshop was held last spring and considered topics of why a database is needed, what is a database, critical scientific questions (re. geographic domain, sites, data), software, and database accessibility and users. It was agreed that, at a minimum, every contributor to the Alpine database will have access to the database. Meetings will be held periodically to resolve questions of taxonomy, etc.

Jacobson, who is chairman of NOAA's Palaeoclimate Advisory Committee, described the philosophy behind databases in the United States. There it is considered the professional obligation of scientists to contribute their data to a database, particularly when the investigator receives public support. The basic philosophy is that if the public pays for the data, the public has access to them. Federal agencies in the U.S., such as the National Science Foundation, are now requiring that data acquired with federal monies be made public.

Justification for financial support.

Watts emphasized that because of the current interest in global climate change, possibilities now exist for funding owing to the usefulness of pollen data for palaeoclimate studies. The need exists for both pollen stratigraphic and pollen surface data. Regardless of the financial justification of the database, however, it will be highly useful for more local studies and for studies of human impact. Behre and several other participants emphasized that palaeoclimate studies must not neglect human impact.

Data quality.

Behre also argued that the database must, from the very beginning, insist on high standards of taxonomy, correctness of radiocarbon chronologies, etc. All participants concurred in the ensuing discussion.

Jacobson maintained that data must be of high quality and that once the data are entered into the database, the data originator should be given the opportunity to verify that the data and taxonomy are correct. Once the data are in the database and verified, it is up to the user to use the data appropriately. Thus, use of the data is analogous to the use of any other kind of published data.

Huntley stressed that data quality, including taxonomic considerations, is the responsibility of the data originator, not the central database. The database must rely on the expertise of the local scientists for questions of taxonomy, hardwater effects on radiocarbon dates, etc. A hierarchical structure of taxonomy in the database will accommodate the finest taxonomic precision possible provided by the data originators, while permitting grouping at higher taxonomic levels for more general research objectives. Huntley further maintained that the central database cannot make judgments about data quality beyond some minimum qualifications, i.e. the data cannot be graded. However, the database must include the relevant data necessary for database users to make judgments concerning data quality or suitability for given research aims.

General discussion among the participants reached general agreement that the database must include the finest taxonomic precision possible and that the taxonomy in the database must be hierarchical. Furthermore, the central database must work closely with local individuals or groups to resolve questions of taxonomy.

Database structure: centralized or regional.

Several participants argued the merits of regional databases, particularly regarding the issues of synonymy and taxonomy, evaluation of radiocarbon dates, language, and motivation. It is clear that a centralized database cannot resolve all issues of synonymy, taxonomy, and chronology, and, in some fashion, must work with regional palynologists.

Watts noted that the issue of funding was basic. Marseille had acquired funding and facilities for a database and must produce a functioning database if funding is to be maintained in the future. The subject of regional databases must inevitably face the critical questions of from where will the funding come, who will organize them, how will compatibility be maintained between the databases, and when will they be developed. The reality was that Marseille had funds now for development of a centralized database.

In addition, the organization of the Alpine regional database is also underway. Discussion among the participants led to the conclusion that the formal development of regional databases for certain geographic or political regions was not necessarily inconsistent with the development of a centralized database. In fact, because the centralized database must rely on local expertise for technical issues, the timely development of given regional databases in close collaboration with the centralized database, would be meritorious and could expedite the flow of verified data into the central database. Berglund and Watts emphasized that the system should not be rigid, but flexible, suiting those desiring local databases as well as those wishing to contribute data directly to the central database. Pons diagrammatically proposed a system featuring a centralized database that would work closely with any regional databases that developed as well as with individual palynologists in other regions. He also argued that development of certain regional databases was not necessarily inconsistent with the idea of a centralized database. He envisioned that in some areas, such as the Alps, groups of palynologists might wish to develop a regional database in concert with the central database, whereas in other regions individual researchers might prefer to work directly with the central database.

Ralska-Jasiewiczowa noted that although palynologists in eastern countries might prefer local databases, the reality was that they had no access to funds, and therefore must work with the centralized database.

Consensus was reached that the choices were not of either a centralized database or of regional databases, but that a system accommodating both could emerge. Regional databases should work closely with the central database in order to maintain compatibility.

Communication.

Excellent communication must be maintained if the database is to succeed. Jacobson noted that the report of the Frostavallen meeting was a brief summary of the workshop and did not contain details of all that was discussed there. As a result some misunderstandings ensued. Huntley suggested that the central database should distribute a newsletter to palynologists throughout Europe to inform of progress of the database, data currently available in the database, software development, research utilization of the database, etc.

Session 3. 22 September 1990, 14.00 hrs, G. L. Jacobson, Jr, Chairperson

Protocols.

Birks outlined a set of proposed protocols for the database, which then guided the discussion. The protocols, as modified by the participants, are the appendix to this report. The protocols establish credibility and integrity of the database and set standards for acceptable use of the database.

Computer software and database design

Grimm outlined the philosophy of relational database design and presented a preliminary set of database tables for the North American pollen database. Several queries, suggestions, and comments about the table design and content were made, with particular reference to needs in Europe. Guiot informally proposed that a meeting on these and related technical matters be arranged in Marseille in January 1991.

Need for computer hardware/software.

The participants emphasized the special need of the countries of Eastern Europe for the acquisition of computer hardware and software. Many labs have no access to microcomputers, while others have only limited access. A number of suggestions were made for financial support for computer acquisition by Eastern European labs and will be pursued. Minimum equipment necessary are an IBM compatible microcomputer with an 80286 processor and 40 MB harddisk. A math coprocessor and colour VGA monitor are also desirable. A printer is also necessary. A provisional list of labs requiring equipment was drawn up.

Executive Committee and Advisory Board members.

A. Pons resigned from the Executive Committee and suggested J. L. de Beaulieu as a replacement. W. A. Watts also resigned and suggested B. Huntley as a replacement. Both resignations and replacements were accepted. B. E. Berglund resigned from the Advisory Board and suggested M.-J. Gaillard-Lemdahl as a replacement, which was accepted. Current membership is now:

Executive Committee

B. Ammann
J. L. de Beaulieu
B. Huntley

Advisory Board

K.-E. Behre
M.-J. Gaillard-Lemdahl
H. J. B. Birks
E. Bozilova
M. Follieri
G. L. Jacobson Jr.
C. R. Janssen
M. Kabailiene
M. Ralska-Jasiewiczowa
J. C. Ritchie

The responsibilities of the Executive Committee were agreed to be finding further financial support for the Database and for overseeing the development of the Database project. The advisory Board is responsible for protocols, assisting with taxonomic decisions, and arbitration in any disputes between contributors, users, or coordinators.

Appendix

Protocols for the European Pollen Database

These protocols are provisional and probably incomplete. These were the protocols actually discussed by the participants.

A. Data

1. Data must consist of the original counts, not percentages or digitized data.

2. Database must contain the original taxonomic identifications, with exceptions of exact nomenclatural synonymy. Taxa will not be lumped into higher taxonomic groups in the database.

3. Data will be flagged as public or restricted. All data will be available in the database. In other words, the central database will distribute all data, not just the public data. Thus, restricted data can be viewed by a user, but cannot be used except as provided below.

4. Public domain data are available for all uses.

5. Restricted data are available only by permission of the data originator. Appropriate and ethical use of restricted data is the responsibility of the data user.

   [An issue not completely resolved was whether restricted data would be available for synoptic studies without permission from the data originator. Synoptic studies are those involving many sites over a large geographic area (i.e. hundreds of sites on a continental scale) and in which actual data values are never plotted. An argument was made that for synoptic studies it is impractical for investigators to contact every data originator. Another possible solution is that a third category of data availability be made, that is data restricted except for synoptic studies. This category might considerably reduce the number of permissions a synoptic investigator would be required to obtain. This point should be discussed among palynologists in Europe, and the Advisory Board will further consider this critical point before establishing the protocol regarding the requirement for obtaining permission from data originators for data use for synoptic studies.]

B. Contributors

1. Can declare data public domain or restricted.

2. Can ask to verify that data in the database are correct. As a matter of general policy, the central database should routinely return to the data originator a hardcopy printout of the data as they are entered in the database for optional verification by the originator.

3. Can have access to all public domain data.

4. Can have access to database software and the database itself for use on his/her own computer.

5. Should receive a periodic newsletter or report concerning the database.

6. Can ask at any time that his/her data be withdrawn from the database or that the status (public domain or restricted) be changed.

7. In the case of a dispute regarding inappropriate use of restricted data, the Advisory Board will serve as arbitrator.

C. Users

1. Must ask permission from the data originator for use of restricted data (with the possible exception of synoptic studies as discussed above).

2. Should as a matter of courtesy, inform data originators of the uses being made of their data.

3. If the contributor wishes, should show the contributor results of analyses and manuscripts for publication for critical comment.

4. Should cite in any publication using data from the database contributors' original publications describing their data.

5. Should send contributors reprints of publications that use their data.

6. Should acknowledge contributors for use of unpublished data and for any technical advice they may have provided.

7. No user can pass data on to another party. All users must obtain data from the central database.

8. Normal ethics apply to co-authorship of publications. The contributor should be invited to be a co-author if a user makes significant use of a single contributors data, or a single contributor's data comprises a substantial portion of a larger dataset analyzed, or a contributor makes a significant contribution to the analysis of the data or to the interpretation of the results. This guideline applies to public domain as well as to restricted data.

9. The data are available only to not-for-profit organizations and research. Use of the data by for-profit (commercial) organizations, even for legitimate uses, can be only with the written consent of the Advisory Board, who will determine or negotiate the payment of any fee required.

4. Coordinators

1. Should prepare a periodic newsletter or report about every six months for contributors, users, the Advisory Board, etc.

2. For use of the data, must abide by the same protocols that apply to all other users.

3. Should make taxonomic decisions in close cooperation with the data originator or relevant regional group.

4. Should assemble a mailing list of all Quaternary palynologists in Europe and should contact all palynologists notifying them of the opportunity to contribute to and participate in the database development. In addition should announce the development of the database in appropriate newsletters and publications.

5. Should incorporate all data in the database, subject to certain minimum requirements, without assignment of quality.

6. Should organize workshops on computing and should work to facilitate acquisitions of computing equipment and software by laboratories not having access to such equipment.

7. Should send agreed protocols to all potential contributors and users.