Week 9 – Obtaining fossil occurrence data

Some databases with fossil occurrences

1. Paleobiology Database (http://paleobiodb.org/)
   FossilWorks interface to Paleobiology Database (http://fossilworks.org)
   occurrence data for all taxonomic groups
2. NOW Database (http://www.helsinki.fi/science/now/)
   occurrence data for mammals, primarily from the Neogene of Eurasia
3. NeoMap (http://www.ucmp.berkeley.edu/neomap/)
   occurrence data for mammals from the Miocene and Quaternary of North America (compilation of the FaunMap database of the Quaternary and the MioMap database of the Miocene).
5. Global Biodiversity Information Facility GBIF (http://www.gbif.org)
   Mostly modern occurrence records, but also fossil data (almost exclusively from Paleobiology Database)

**NOTE 1 data quality:** occurrence data in any of these databases cannot be expected to be complete. The databases are compilations from many sources, each of which had its own unique purpose. The occurrences in the database may therefore not even be a random sampling of real occurrences; they may be skewed to certain taxa, certain time periods, or certain geographical areas. The data from these sources is suitable for our class projects, but examine them very carefully if you choose to publish on them or use them for serious research. In examining data for serious uses, it is often wise to:

1. check whether all the taxa in your group are represented in the data
2. check whether the occurrences reported for taxa adequately sample the known geographic or temporal range
3. spot check age and location of individual occurrences
4. look carefully for gaps in time, space, or taxonomy

**Note 2 data citation:** Despite shortcomings of data, a lot of work went into making the compilations of occurrences found in these databases. A single occurrence record involved field work, taxonomic work, publication, assembling publications, and entering into the database, probably totalling 10 to 100 person hours per occurrence. Give the people who did this work as much credit as possible. Each database has a suggested form of citation to the database itself, and most of these contain data that credit the original source of the occurrence and the person who entered it into the database. Give as much credit as possible because researchers are frequently evaluated on the impact of their work, most notably by the number of citations their work has received.
Obtaining occurrence data

Occurrences are records of the presence of a taxon at a particular time and place. The temporal duration of a taxon is known from the oldest and youngest occurrence, and the geographic range is known from the geographic spread of the occurrences. A typical occurrence record has fields for the taxonomy, the site name, the geological age, and the location.

Obtain fossil occurrence data for the group you want to work on. We’ll use these data in the next assignment. Make sure your data have the following fields: taxonomic identification, geological age in millions of years (or some other absolute time). You will want at least 100 records, probably no more than 5000 records. Once you have downloaded the data, organize them in Excel by getting rid of columns or rows that you don’t need, giving the columns sensible, one-word names (e.g., “Species”, “Locality”, “Longitude”). Save and import into MySQL.

Note that if you use Paleobiology Database, it has four tabs in the data download section, each with options. Under Collection Fields you’ll want to choose latitude/longitude (decimal), 10 my bin and midpoint age, under Occurrence Fields you might want Species Name and, perhaps, Enterer (since the data in PBDB were contributed by voluntary effort and you will want to acknowledge the people who entered it).

Reading