

The Paleointegration Project in GEON (The Geosciences Network)

Allister Rees (University of Arizona)
arees@email.arizona.edu

The Paleointegration Project (PIP) within GEON (<http://www.geon.org/>) is facilitating interoperability between global-scale fossil and sedimentary rock databases, enabling a greater understanding of the life, geography and climate of our planet throughout the Phanerozoic.

Background: In the first phase of PIP, I developed five interoperable databases of sedimentary rock types, plants, and dinosaurs, training and working with University of Arizona IT students. These comprise some 175,000 occurrence records from 55,000 localities worldwide. The Paleogeographic Atlas Project (PGAP, University of Chicago) compiled most of the sedimentary and floral data (Permian through Cenozoic). Dinosaur data were derived from “The Dinosauria” encyclopedia, in collaboration with David Weishampel (Johns Hopkins University). Judy Parrish (University of Idaho) provided her dataset of Phanerozoic oil source rocks for integration within the system. The databases are text- as well as map-searchable, through the use of age and geography ontologies I developed, linked to GIS mapping tools.

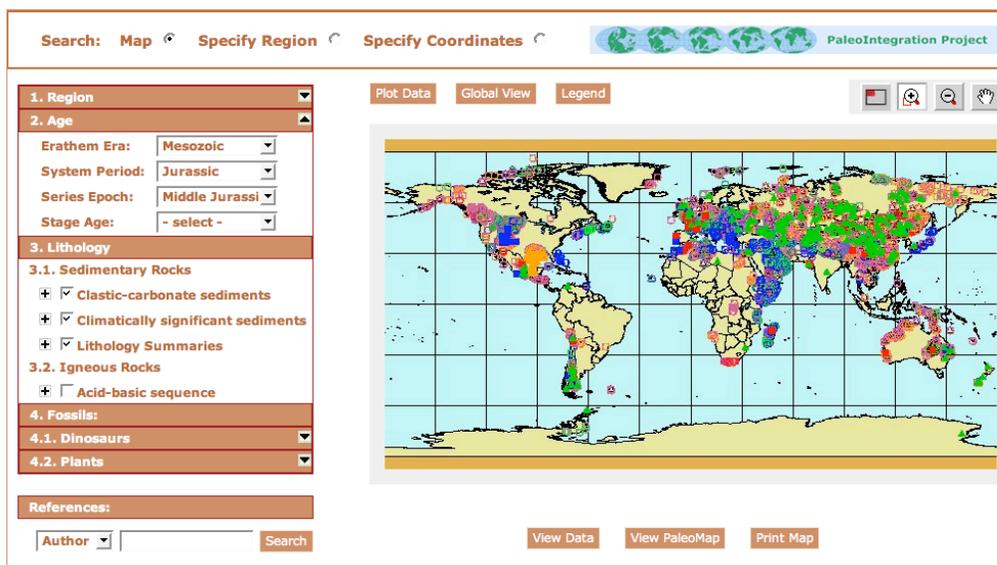


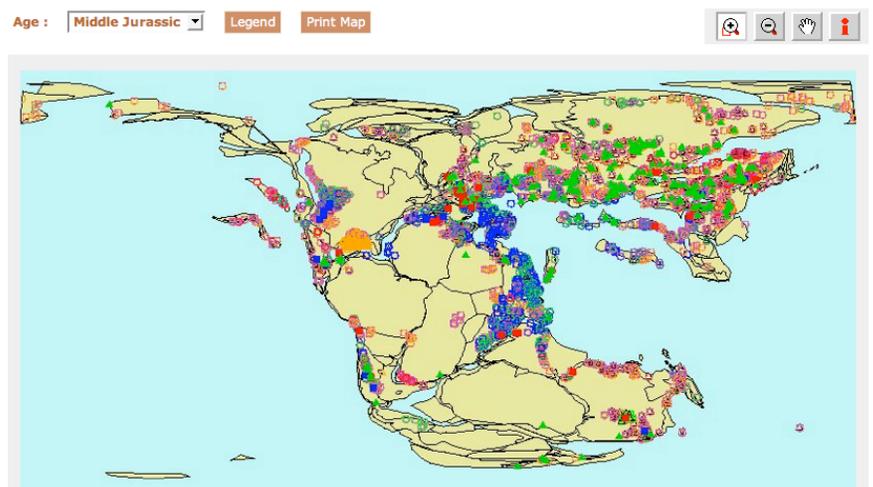
Figure 1. The PIP user interface, showing the search options on the left and mapped results on the right. Smaller regions can be selected from the map. Other search options include: specify ocean, continent or country; enter bounding coordinates; select geologic age (era through stage); select various rock and/or fossil types; search references. Other features include: view data on paleomaps; view data tables (also downloadable as text files); print maps.

Results are viewable on present-day maps, as well as paleo- maps courtesy of Chris Scotese (<http://www.scotese.com/>). Results can also be downloaded for further detailed analyses.

Figure 2. The same data from the example in Figure 1, plotted on a Middle Jurassic paleomap.

The system was designed to ensure fast data retrieval, making it especially useful for extensive as well as multiple simultaneous searches (e.g. in a classroom setting). This version was launched on the UofA geongrid server (<http://www.geo.arizona.edu/~rees/>) and subsequently embedded into the GEON portal. Although I'd developed age and geography ontologies to enable seamless interoperability, I was working with static “legacy” databases

on one server. Moreover, I'd precalculated offline the paleocoordinates for all 55,000 localities in the databases. More was needed to incorporate “live” databases residing on different servers.



Current PIP Status: Close collaboration with John Alroy (The Paleobiology Database, PBDB - <http://paleodb.org/>) and Chris Scotese has resulted in significantly more data as well as tools, which are now available via the GEON portal (<http://www.geon.org/>). The PBDB is well-known as an invaluable paleontological resource, comprising some 630,000 fossil occurrence records from 69,000 collections worldwide for the entire Phanerozoic. Through the use of web services developed at the San Diego Supercomputer Center (SDSC), the PBDB server in Santa Barbara is now searchable dynamically within PIP, alongside the existing databases. In addition to the PIP features, each PBDB locality record is linked back to the original database, enabling the user to either explore and analyze the data further within that system (e.g., using the PBDB statistical tools), or to continue within the PIP interface.

Essential to this current PIP effort was a means to generate locality paleocoordinates “on the fly”, based on modern latitude and longitude as well as locality age. This has been accomplished through a module (Auto Point Tracker - APT) developed and provided to GEON by Chris Scotese, enabling dynamic calculation of locality paleocoordinates and automated plotting on his paleogeographic maps.

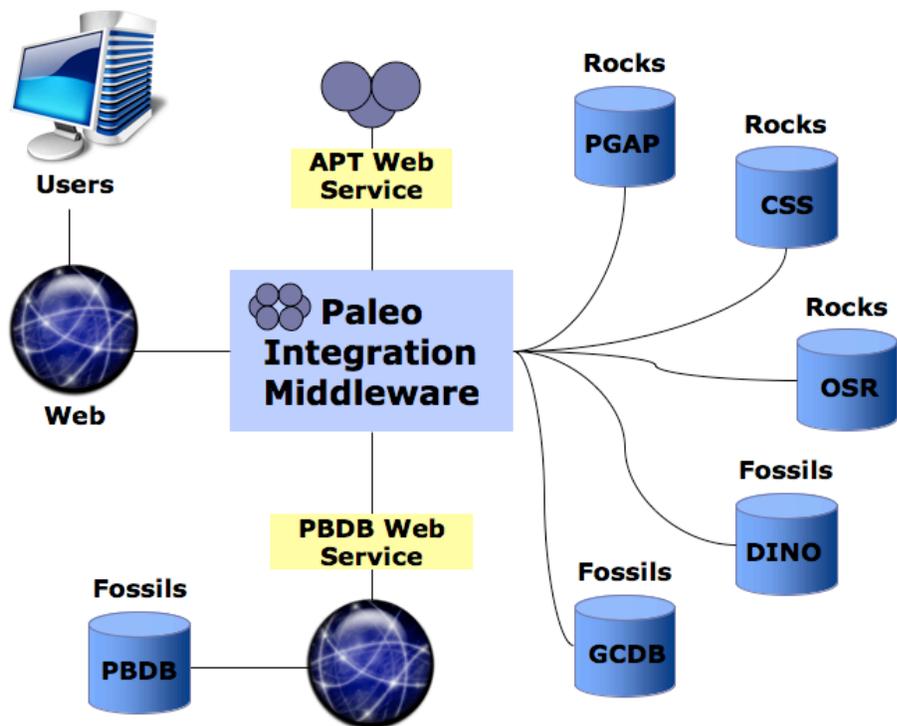


Figure 3. The architecture of PIP, showing how different databases and components interact to produce an integrated result and map. Researchers send a query for data using the PIP user interface. This query is parsed by the PIP middleware. It is then decomposed into several different queries sent to the databases using Java Databases API for the ones hosted on SDSC servers and using Web Services for querying data hosted at remote locations (e.g., the PBDB). Once all the results are accumulated, they are passed to the Auto Point Tracker Web Service for “on the fly” calculations of locality paleo coordinates so that they can be plotted on the paleogeographic maps.

The Paleointegration Project is already proving useful to researchers, teachers and students. Anyone can now access data and tools that were only available previously to specialists. I envisage continuing to develop this with the addition of new datasets, tools and services. It’s more than a paleontology and sedimentary geology project. A more complete understanding of the interactions between Earth and Life through time requires the addition of, for instance, geochemical, geophysical, igneous and metamorphic data.

The PIP should also be a useful resource for paleoclimate modelers, who can now retrieve the appropriate proxy data within seconds. I should emphasize that PIP doesn’t replace specialist expertise. It does, however, provide another means whereby researchers can develop their own scientific queries.

In summary, the PIP presently contains disparate geologic databases residing on different servers, embedded modules, web services, and GIS – structures and tools that should greatly facilitate collaborative efforts within the broader geosciences community.