

1st Workshop on Paleoecological Databases in South America

Claudio Latorre^{1,2}, P.I. Moreno^{2,3} and E.C. Grimm⁴

Olmué, Chile, 12-16 June 2013

Creating a state-of-the-art database for paleoecological data is an expensive and time-consuming endeavor that requires long-term funding. Currently, no institution in South America is likely to have access to the necessary funding for such an effort. An existing open-access platform, however, now available through the Neotoma Paleoecological Database (www.neotomadb.org) offers an extremely valuable infrastructure that scientists can use throughout the developing world. The use of this common platform would maximize compatibility among the regions and the versatility of this global open access archive of paleoenvironmental data.

Forty-one scientists from seven South American countries gathered in Olmué (located north of Santiago in the foothills of La Campana National Park) for an intense two-day meeting on the policy and practices involved in creating a public-access database for South American paleoecologists. The workshop was followed by a three-day course for young scientists taught by Eric Grimm on how to access and use the Neotoma database, and on how to use the newest versions of the Tilia (pollen analysis; Grimm 1991) and Bacon (age modeling; Blaauw and Christen 2011) applications.

Why Neotoma?

Neotoma is a large, multi-proxy database that includes datasets spanning back to the Pliocene. It is a client-server database housed at the Center for Environmental Informatics at Pennsylvania State University, USA, and is comprised of many virtual databases representing different data types and geographic regions. It arose from a partnership between IT domain scientists and developers in an attempt to answer large-scale questions about paleoenvironmental evolution including climate, fauna and flora.

Workshop participants included archeologists, paleoecologists, zooarcheologists, paleontologists, diatomists, and experts in other paleo fields, who addressed the following questions: (1) How to involve South American scientists in a ground-up effort to build and to contribute to a database? (2) Where can funding be sought to assemble legacy data? (3) Who owns the intellectual rights to the contributions made? (4) What data should be publicly available or restricted (e.g. locations of geographically sensitive sites such as unprotected archeological sites)? (5) Who could act as data stewards for the different South American data types (vertebrates, pollen, rodent middens, zooarcheofaunas, etc.)? (6) Should it be possible to upload unpublished datasets? (7) Should it be possible to upload data to Neotoma for ongoing research projects to take advantage of database analysis and visualization tools but

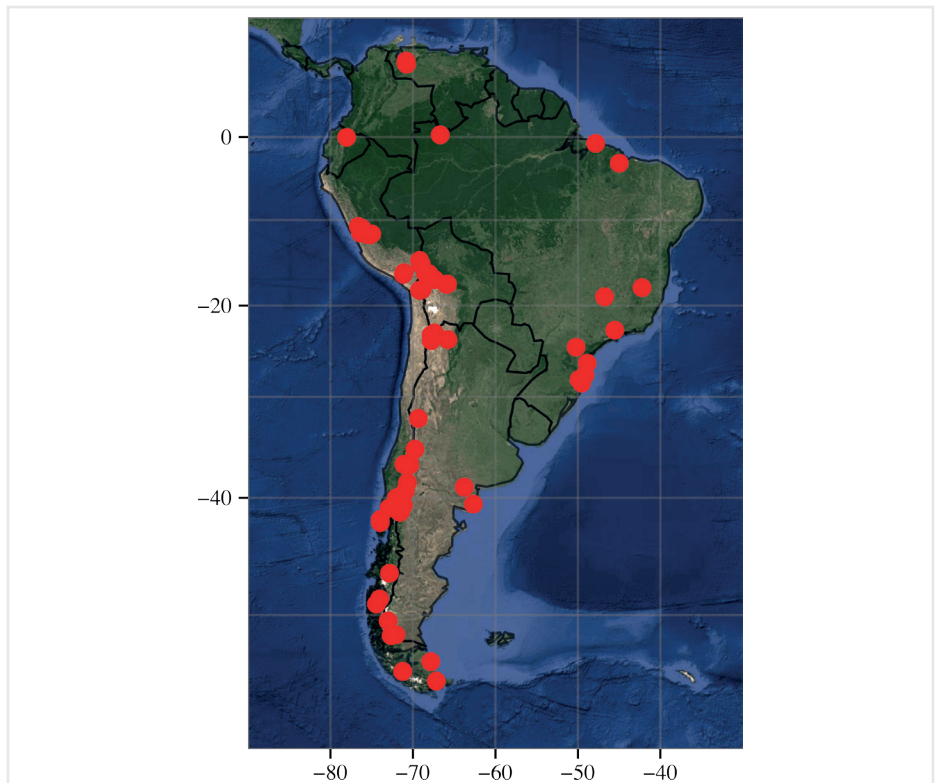


Figure 1: Current South American paleoecological sites in Neotoma. The vast majority of these are part of the constituent Latin American Pollen database. Future collaborations could add much needed faunal data as well as other databases (for figure source code see <http://pastebin.com/fgQSVTKb>; image credit: Simon Goring)

have it embargoed for public release until after publication?

Workshop attendees were divided into breakout groups for further discussion. While recognizing the importance of these large public-access databases, several key concerns need to be addressed, such as data propriety rights. In particular, when dataset authors should be invited to be co-authors on data-synthesis papers and how should the use of data be adequately cited, especially for previously unpublished data. A proposed solution would be to make datasets themselves citable, with for example, a DOI number. This solution could stimulate the publication of legacy data, probably one of the more cost-intensive efforts of creating such a database for South America.

Also, there is the attractive possibility of creating “embargoed” contributions, i.e. datasets uploaded to Neotoma but not publicly available for a determined amount of time. This would allow time to take advantage of the database’s infrastructure and ample time to publish an article before releasing data. Other important aspects discussed included how to create “fuzzy” coordinates for geographically sensitive sites, in which only the data contributors and relevant land management agencies

know the actual locations (e.g. to protect sites from looting).

Clearly much work lies ahead, but researchers were enthusiastic about the possibility of generating a public database which could address and generate new and exciting research questions, particularly those associated with broad-scale climate and land-use change that could contribute to published data and incorporate the vast quantity of legacy data generated in South America for over a century.

AFFILIATIONS

¹Department of Ecology, Pontifical Catholic University of Chile, Santiago, Chile

²Institute of Ecology and Biodiversity, University of Chile, Santiago, Chile

³Department of Ecological Sciences, University of Chile, Santiago, Chile

⁴Illinois State Museum, Springfield, USA

CONTACT

Claudio Latorre: clatorre@bio.puc.cl

REFERENCES

- Blaauw M, Christen JA (2011) *Bayesian Anal* 6: 457-474
 Grimm EC (1991) *Tilia and Tiliagraph*. Illinois State Museum, Springfield