

# Workgroup Report: External Database Linking

Dave Anderson, Tony Barnosky, Brian Bills, Angela Bruch, Rob Guralnick, Jenn Marlon, Richard Pelletier, Noel Heim

## Priorities

1. Establish capabilities to exchange catalog level metadata among databases.
  - a. Content: Darwin Core (would provide immediate integration to Paleoportal)
  - b. Suggested format: NASA Directory Interchange Format (DIF)
  - c. Suggested implementation: Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)
2. Develop a detailed plan for managing changes to data within the database.
  - a. What is the process for handling requests to investigate questionable data and update them if necessary?
  - b. Will mechanisms exist for propagating updates upstream to consuming databases (i.e. error discovered at source) as well as downstream to the source database (i.e. error discovered by consumer)?
  - c. What is the Neotoma policy concerning synchronization with databases that have overlapping contents?
3. Determine a strategy for enabling the micro-attribution of datasets associated with publication records in the major bibliographic databases.

Example scenario:

- a. Paper published with primary references
- b. Web of Science harvests and creates citation record
- c. WoS record supplemented with additional data references (may be numerous)
  - i. Do the bibliographic databases have a mechanism to allow this?
  - ii. Can we automate?
- d. Provide use statistics back to cited dataset providers

## Discussion

- What role does Neotoma wish to play in the broader ecosystem of databases across which higher-level aggregate tools and services will be developed?
- The keys to enabling query across paleo-databases are resolution of disparate taxonomy, chronology, and geography.
  - Who will do the translating?
- Rob may have funding to assist Neotoma with development of standards-compliant metadata interoperability services.
- It is important to consider the needs of two very different audiences of data providers and consumers: those of individuals and those of other database (and collectives of databases)
- Investigate linkages to modern flora/fauna distribution databases

## Databases to Consider Linking

ROCEEH/ROAD

- Archaeological focus; 393 sites (24.09.2010)
- Content: archaeology, palaeontology, palaeobotany, palaeoclimate, palaeoenvironment

- Needs
  - Fossil plant and animal -> want to search Neotoma with their tools.
  - Get data- harvest/ periodical update or web source
- Want to contribute new data (field + published datasets)
- Major use: climate and environmental reconstructions
- Time: 20ka to 3 Ma
- Extent: Africa, Eurasia
- ROCEEH - <http://www.roceeh.net/>
- ROAD - [http://www.roceeh.uni-tuebingen.de/roadweb/smarty\\_switch\\_menu.php?menutype=main](http://www.roceeh.uni-tuebingen.de/roadweb/smarty_switch_menu.php?menutype=main)
  - User id – public; Password – roceeh
- Research products e.g., maps: <http://www.roceeh.net/index.php?id=120&L=0>

#### Neomap

- Fossil mammal occurrence data from 30 million years ago through the Holocene. Mostly United States
- <http://www.ucmp.berkeley.edu/neomap/>
- >30,000 occurrences of fossil mammals plus or minus
  - Faunmap ~15,000 taxon occurrences (?? check with Russ)
  - Miomap ~15,000 taxon occurrences

#### Neogene of the World

- Fossil mammal occurrence data primarily from the Neogene (emphasis on Pliocene and Miocene) but with some coverage of Oligocene and Pleistocene as well. Primarily Eurasia, but expanding to cover other continents as well.
- <http://www.helsinki.fi/science/now/>
- > 50,000 occurrences

#### NOAA World Data Center for Paleoclimatology

- Provides data and information scientists need to understand natural climate variability and future climate change.
- Operate the World Data Center for Paleoclimatology which distributes data contributed by scientists around the world.
- <http://www.ncdc.noaa.gov/paleo/ftp-search.html>

#### Paleobiology Database

- All fossil occurrences through time.
- <http://paleodb.org/cgi-bin/bridge.pl>
- > 874,000 occurrences

#### Paleontology Portal

- Links museum records of paleontological collections
- <http://www.paleoportal.org/>
- many hundreds of thousands of records

#### VertNet

- Data publishing network that provides access to species occurrence records from a single portal that mines through museum and observational data from a variety of sources.
- <http://vertnet.org/index.php>
- 84.3 million records