A New Online Database (http://anthropologicaldata.free.fr) and a Short Reflection About the Productive Use of Compiling Internet Data

JEAN-LUC VOISIN
UMR 7194 - CNRS & USM 103, Muséum National d'Histoire Naturelle, IPH, 1 rue René Panhard, 75013 Paris, FRANCE; and, ADES Anthropologie Bioculturelle, Droit, Ethnique et Santé, UMR 7268 - CNRS / Université Aix Marseille / EFS, CS80011, Boulevard Pierre Dramard, 13344 Marseille Cedex 15, FRANCE; jeanlucvoisin2004@yahoo.fr

SILVANA CONDEMI
ADES Anthropologie Bioculturelle, Droit, Ethnique et Santé, UMR 7268 - CNRS / Université Aix Marseille / EFS, CS80011, Boulevard Pierre Dramard, 13344 Marseille Cedex 15, FRANCE; silvana.condemi@univmed.fr

MILFORD H. WOLPOFF
Department of Anthropology, University of Michigan, Ann Arbor, MI 48109, USA; wolpoff@umich.edu

DAVID W. FRAYER
Department of Anthropology, University of Kansas, Lawrence, KS 66045, USA; frayer@ku.edu

ABSTRACT

This note provides information on a new, online dental database and concludes with some observations about the use of the Internet and the growth of available online databases. First, we describe our web site, which allows downloading, free of charge, of dental metric data from apes to recent modern humans. Currently, the sample consists of length and breadth data for orangutans, gorillas, chimpanzees, a large sample of European Upper Paleolithic and Mesolithic, Neolithic and Hungarian Medieval humans, and late Upper Paleolithic specimens from North Africa. We plan to add more data on australopithecines to Neandertals as samples are accumulated and edited. Two types of data sets are presented—*Excel* files for mesial-distal and labial (buccal)-lingual data are arranged 'tooth by tooth' for each specimen with references. The webpage also lists *Excel* files for the specimens by species (e.g., orangutans) or sites (e.g., Abri Pataud, Vlasac) for the various grouped samples. In the first type of *Excel* files we plan to add data on tooth wear for each specimen. For the second, only length and breadth measurements are provided.

Sharing data is the basis of scientific research. The web is the most powerful tool to achieve this goal, since it permits anyone to access a nearly unlimited range of data. We encourage scholars to use and contribute to our website.

> "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing." (US National Science Foundation, 2012)

While working on Neanderthal dental variation (Condemi et al. 2010), we realized that most analytical studies do not give raw data and much time is required to collect them. One of us (Voisin) had the idea to create a web site (http://anthropologicaldata.free.fr) providing free access to dental metric data of extant great apes and hominins. The website includes *Excel* files in two formats, by individuals and by sample. For individual data, at present more than 6,000 permanent teeth from more than 400 individuals in the Upper Paleolithic and Mesolithic are available. For apes, currently there are 1000+ *Pongo* teeth from more than 50 individuals, more than 900 chimpanzee (*Pan troglodytes*) teeth from more than 100 individuals, and nearly 145 *Gorilla* teeth from more than 10 individuals. For great apes, all dental data derive from the Ph.D. dissertation of Paul Mahler (1973). We strongly recommend that users of his files check his dissertation (available on line at http://search.proquest.com/pqdtft/docview/302652203/fulltextPDF/139D035B8B341C676D0/6?accountid=14556) for his measurement techniques, especially for the canine and P_3. For now, most human data are Upper Paleolithic, Mesolithic, Neolithic, and Medieval teeth with some Middle and Lower Paleolithic specimens. We have added, however, the complete Krapina dental sample compiled and recently updated by Wolpoff (Radovčić et al. in prep.; Wolpoff 1979). For the most part, these come from original measurements collected by Wolpoff (1999) on a NAS/IREX and NSF funds

PaleoAnthropology 2012: 241–244. ISSN 1545-0031
doi:10.4207/PA.2012.ART76

or Frayer (1978, 1984, 2004), collected with NSF funds in 1973–74 and 1984–86, and NAS/IREX grants in 1981, 1983 and 1992. More great ape, Neolithic, and Medieval Europeans will be added in this format and we plan to upload data from North African late Paleolithic sites. We intend to eventually include data from the australopithecines, other fossil groups, and a worldwide sample of modern humans. We also hope to make available human and ape deciduous tooth measurements. Most measurements are standard length (mesial-distal) and breadth (labial [buccal]-lingual) dimensions (Wolpoff 1971). In some limited cases, crown height and length and breadth dimensions at the cervix are included, based only on published data.

User beware! We add some cautionary notes about using dental data, whether presented here or in any other data sets, especially but not only when there are comparisons of apes and some australopithecines and humans. Wolpoff (1971), White (1977), Frayer (1978), and others define how MD and BL (or LL) are taken. But these are often not homologous between authors, even sometimes for the same author. There are many examples, but really just two main sources of problems with homology. The first is in (or between) samples of teeth measured *in situ* and measurements taken on isolated teeth. The problem here could come for any teeth, but especially for asymmetric ones such as upper molars, lower P_3, and canines of both jaws. The second comes from the fundamentally different shapes of the lower P_3 and canines of both jaws, comparing apes and some australopithecines with other hominids. These dimensions are poorly described and, while some authors (cf. White 1977) are quite clear about how the measurements can be defined to allow homology, most authors are not. An excellent example is in the argument about the identification and shape of a molar from Bed I, Olduvai. Leakey, Von Koenigswald, Robinson, and Dahlberg could not decide whether it was an upper or lower deciduous M2, or a permanent molar tooth, longer than broad or broader than long. This went on in *Nature* for several years and was never resolved. Comparing data sets invariably has homology issues like these—they are rarely brought to light.

Our web site is made up of four pages: (i) a "Home," (ii) "Tables and Bibliography," (iii) "About Us," and (iv) "Links." The first page introduces Internet users and the "About Us" identifies the organizers of the web site. The "Tables and Bibliography" and the "Links" pages are the web site's core. The former allows downloading of two kinds of data sets and the latter provides links to other databases.

The first database available in the "Tables and Bibliography" page is dental metric data in *Excel* 2003 format. These tables are arranged by country and each line corresponds to a single tooth with the name of each column identified in the listing. One column specifies the source of each data point and whether it was taken on the original specimen or a cast (more than 99% of our data are currently taken on originals, the only exceptions are destroyed originals). For each tooth there are, at least, mesial-distal and buccal (labial)-lingual crown dimensions. In the future other information will be included such as root dimensions, degree of wear, pathology, non-metric traits, photographs, and CT scans. A column with a primary bibliography for each specimen is listed. The bibliography includes (i) initial publications about each specimen such as description of the site and description of the teeth, and (ii) a bibliography of other data sources, if different from the main work. If two or more data provided by different authors are listed, all the data are put online and the user must choose which data are more appropriate for their problem. Individuals are separated in the files by two solid black lines.

Another database in PDF format corresponds to the complete bibliography cited in the database files. For some references, the journal name is written in blue, which allows the user to download the reference free of charge. The web site is currently in HTML, which provides only static pages and allows only downloading the entire tables without making a choice about the data before downloading. Eventually we will upgrade the web site by using PHP programming and MySQL tables.

The second type of files are grouped, length and breadth data of humans from Krapina, and the Upper Paleolithic to Medieval Hungarians. These are the same raw data files used for the individual file types discussed above, but are presented in a different *Excel* format. In some cases, in place of individual data for left and right sides, averages of the two sides are given or only the left when the right is missing or only the right when the left is absent. Because the data files are not converted to the more extensive data files listed above, we are able to provide a few more samples, including deciduous teeth from the Upper Paleolithic, Mesolithic, Neolithic, and the Hungarian Medieval sites at Zalavar. We also have uploaded permanent teeth from Afalou-bou-Rhummel (Algeria) and Taforalt (Morocco), all measured by Frayer. All of Mahler's files for *Gorilla, Pan*, and *Pongo* are included here too.

The web site (http://anthropologicaldata.free.fr) is completely free with no required login or password. However, when using data from this web site, we ask the user to reference the database as Voisin, Condemi, Wolpoff, and Frayer 2012 (or the most recent year the dataset is updated), http://anthropologicaldata.free.fr and send the leader of the project (Voisin: Jeanlucvoisin2004@yahoo.fr) a PDF of your published paper.

In the past few years, online databases have rapidly increased, especially with respect to mouse data (Schofield et al. 2009), and ancient DNA and the 1000 Genomes project (anon 2011; Kirsanow and Burger 2012; Meyer et al. 2012). Recently, *PaleoAnthropology* announced a data bank based on the archaeology of Montet-White (Hovers and Rosenberg 2012; Montet-White 2012), but there are only a few cases where dental and other osteological data for human fossils have been posted, such as Brown: http://www-personal.une.edu.au/~pbrown3/palaeo.html. Despite a conference aimed at making paleontological data available online (Delson et al. 2007), no outcome for basic data has emerged. Similarly, during the 2011 meeting of the American Association of Physical Anthropologists (Reed and Campisano

2011), a session was devoted to web databases and numerous options were discussed, but no implementation of free databases materialized. Today, any search with different sites, like *Google* or *Bing*, will supply many results, but not the database one is seeking.

We do not consider our site the only option to share metric data online, but once someone opts to share data, it is relatively easy to create "links" pages in each online database providing numerous web addresses to other important web sites. As each online database creates its own link page, they will not provide identical addresses and a great diversity of links would be thus available. In other words, a "dead end" web site, as most online databases currently look should be avoided. It would be worthwhile if major journals, like *American Journal of Physical Anthropology, Journal of Human Evolution, Bulletins et Mémoires de la Société d'Anthropologie de Paris* provided some 'e-space' each year with links to main databases as well as in the paper and in the online issues, as the PaleoAnthropology Society web site already does.

In this regard, an Internet portal of databases, such as http://paleoanthroportal.org/, is very useful for a variety of purposes, but includes no links, which provide large samples of metric data for teeth. For example, *Nespos* (Neanderthal Studies Professional Online System) can be subscribed to for a minimal enrollment fee and lists a number of fossil specimens, including high quality photographs and site data, but not metrics for most specimens. Similarly, the *Digital Archive of Fossil Hominids* (University of Vienna) has images and other data, but no metric data and access is not free.

Another question is the citation of online databases in article bibliographies, which enhances and advertises their scientific visibility. As presented in the 2011 American Association of Physical Anthropologists Minneapolis meeting, web databases are less cited than traditional journal articles, although each online database could be considered as a citable electronic publication (Reed and Campisano 2011). This difference is mainly due to the fact that web publications are currently not widely accepted in many journals and because many scholars are reluctant to cite them in the first place. With publications going electronic and online pre-publications like *arXiv* (Cornell University 2012; Eisen 2012; Hawks 2012a,b), sooner rather than later, web citations will become commonplace.

In a field with a tradition of withholding basic data on human fossils, it is time for human paleontologists to follow the lead of the modern genetic researchers (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) and ancient (http://www.eva.mpg.de/neandertal/data.html) DNA researchers who routinely put their data online for others to use. The benefits of this are a deeper understanding of the evolutionary processes leading to the significance of prehistoric genetic relationships to Europeans (Hawks 2012c; Keller et al. 2012). Similar openness is practiced by some australopithecine paleontologists, such as Berger (Balter 2011), who shared casts of *A. sediba* with anyone interested, before their publication in *Science.* But this is the exception and still in 2012, it is difficult to find basic metric dental measurements for comparisons.

As the US National Science Foundation contends, sharing data is the basis of scientific research. The web is the most powerful tool to achieve this goal as it permits anyone to access an early unlimited range of data. However, scholars must be willing to contribute their basic data to the web and others need to know where to find these data. Main sites have to be referenced in different media and all databases must have *links* pages. Webmasters need to send the URL of their new online databases to most, if not all, other online databases, as we have done for our web site http://anthropologicaldata.free.fr. Today, this web site is referenced on the PaleoAnthropology Society web site (http://paleoanthro.org), Antrocom (http://www.antrocom.org), *Associazione Antropologica Italiana* (http://aai.unipr.it/cgi-bin/home.pl), Dental Anthropology Association (http://anthropology.osu.edu/DAA/index.htm), and soon on INEE (*Institut National Ecologie Environnement*, which houses CNRS databases). Our new website is a small step toward the future, when we hope colleagues will routinely contribute data to it to be placed online so that others can use it in their research.

ACKNOWLEDGEMENTS

We thank all the curators of human skeletal collections from European institutions who made their specimens available for Frayer's calipers. We thank Dr Jakov Radovčić and the late Dr I. Crnolatac, of the Geološko-Paleontološki muzej in Zagreb, for their cooperation, help, and the hospitality extended to Wolpoff in studying the Krapina dental remains. Our friend, Paul Mahler, who died too young (1984), was a tireless worker and always open to sharing his metric data on ape dentitions. He never saw the possibilities of posting his data on the Internet, but would have jumped at the chance. We also thank Luca Bondioli (Rome) for help in transforming the Frayer's datasets from *Statview* to *Excel*.

REFERENCES

Anon. 2012. 1000 genomes http://www.1000genomes.org/

Cornell University. 2012. http://arxiv.org

Balter, M. 2011. Paleoanthropologist now rides high on a new fossil tide. *Science* 333, 1373–1375.

Condemi, S., Voisin, J.-L., Belmaker, M., and Moncel, M.-H. 2010. Revisiting the question of Neanderthal regional variability: a view from the Rhône Valley Corridor. *Collegium Anthropologicum* 34, 787–796.

Delson, E., Harcourt-Smith, W.E.H., Frost, S.R., and Norris, C.A. 2007. Databases, data access, and data sharing in paleoanthropology: first steps. *Evolutionary Anthropology* 16, 161–163.

Eisen, M. 2012. The glacial pace of change in scientific publishing. http://www.michaeleisen.org/blog/?p=1162

Frayer, D.W. 1978. *Evolution of the Dentition in Upper Paleolithic and Mesolithic Europe*. University of Kansas, Publications in Anthropology n°10: Lawrence, KS.

Frayer, D.W. 1984. Tooth size, oral pathology and class distinctions: evidence from the Hungarian Middle Ages.

Anthropologiai Közlemenyek 28, 47–54.

Frayer, D.W. 2004. The dental remains from Krškany and Vedrovice. *Anthropologie (Brno)* 42, 71–103.

Hawks, J. 2012a. Immediate publishing.http://johnhawks.net/weblog/topics/meta/journals/publishing-versus-nasa-eisen-2012.html

Hawks, J. 2012b. The costs of publication delays.http://johnhawks.net/weblog/topics/metascience/pickrell-preprint-2012.html

Hawks, J. 2012c. Neandertal ancestry: iced.http://johnhawks.net/weblog/reviews/neandertals/neandertal_dna/neandertal-ancestry-iced-2012.html

Hovers, E. and Rosenberg, K. 2012. The data bank: note on a new section to *Paleoanthropology*. *Paleoanthropology*, 2012, D1.

Keller, A., Graefen, A., Ball, M., Matzas, M., Boisguerin, V., Maixner, F., Leidinger, P., Backes, C., Khairat, R., Forster, M., Stade, B., Franke A., Mayer, J., Spangler, J., McLaughlin, S., Shah, M., Lee, C., Harkins, T.T., Sartori, A., Moreno-Estrada, A., Henn, B., Sikora, M., Semino, O., Chiaroni, J., Rootsi, S., Myres, N.M., Cabrera, V.M., Underhill, P.A., Bustamante, C.D., Vigl, E.E., Samadelli, M., Cipollini, G., Haas, J., Katus, H., O'Connor, B.D., Carlson, M.R.J., Meder, B., Blin, N., Meese, E., Pusch, C.M., and Zink, A. 2012. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Communications* 3, 698.

Kirasanow, K. and Burger, J. 2012. Ancient human DNA. *Annals of Anatomy* 194, 121–132.

Mahler, P. 1973. *Metric Variation in the Pongid Dentition*. Ph.D. dissertation, University of Michigan: Ann Arbor.

Meyer, M., Kircher M., Gansauge, M-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., Sudmant, P.H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R.E., Bryc, K., Briggs, A.W., Stenzel, U., Dabney, J., Shendure, J., Kitzman, J., Hammer, M.F., Shunkov, M.V., Derevianko, A.P., Patterson, N., Andrés, A.M., Eichler, E.E., Slatkin, M., Reich, D., Kelso, J., and Pääbo, S. 2012. A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338, 222–226.

Montet-White, A. 2012.The data bank: data files generated by excavations of LGM Sites from Central and Western Europe. *Paleoanthropology* 2012, D2–D8.

Radovčić, J. and M.H. Wolpoff. n.d. *An Updated Krapina Hominid Catalog*. Croatian Natural History Museum: Zagreb,

Reed, D. and Campisano, C. 2011. Session 17: Data management in the 21st Century: integrating bio- and geoinformatics in physical anthropology. *American Journal of Physical Anthropology* S52, 31.

Schofield, P.N., Bubela, T., Weaver, T., Portilla, L., Brown, S.D., Hancock, J.M., Einhorn, D., Tocchini-Valentini, G., de Angelis, M.H., and Rosenthal, N. 2009. Post-publication sharing of data and tools. *Nature* 461, 171–173.

UCSC Genome Bioinformatics 2012. http://genome.ucsc.edu/index.html

White, T.D. 1977. New fossil hominids from Laetolil, Tanzania. *American Journal of Physical Anthropology* 46,197–230.

Wolpoff, M.H., 1971. *Metric Trends in Hominid Dental Evolution*. Case Western University Press: Cleveland.

Wolpoff, M.H. 1979. The Krapina dental remains. *American Journal of Physical Anthropology* 50, 67–114.

Wolpoff, M.H. 1999. *Paleoanthropology*. McGraw-Hill: New York.