W. Henry Gilbert & Joshua P. Carlson - Data Models

DATA MODELS AND GLOBAL DATA INTEGRATION IN PALEOANTHROPOLOGY: A PLEA FOR SPECIMEN-BASED DATA COLLECTION AND MANAGEMENT

December 2010

In book: Pleistocene Databases: Acquisition, Storing, SharingEdition: 4Chapter: DATA MODELS AND GLOBAL DATA INTEGRATION IN PALEOANTHROPOLOGY: A PLEA FOR SPECIMEN-BASED DATA COLLECTION AND MANAGEMENTPublisher: Wissenschaftliche Schriften des Neanderthal MuseumsEditors: Roberto Macchiarelli & Gerd-C. Weniger

William GilbertWilliam GilbertJoshua CarlsonJoshua Carlson

# Specimen table schema

• **Project**: Unique identifier of study area or project.

• **Specimen type**: Specimens will generally be of 6 types: Geology, vertebrate paleontology, archaeology, non-vertebrate paleontology, paleobotany, and microfossils. It is common, but not necessary, for projects to devise unique number schemes or abbreviations to distinguish among the types readily.

• **Specimen number**: Alphanumeric code that uniquely identifies specimen. Often, a specimen number contains metadata that allows for quick detection of the area of origin or current repository (e.g. KNM ER-1470). Avoiding characters that have a function in common query and markup languages (like quotes, percent signs, parentheses, semicolons, backslashes, etc.) is strongly advised. Although not absolutely necessary, a good rule of thumb is to use characters that would work in a MS-DOS filename. Care should be taken that no more than one individual is represented by each specimen number.

• **Locality**: Alphanumeric code that uniquely identifies locality.

• **Elements preserved**: List of skeletal elements preserved. Abbreviations are not recommended due to their incompatibility with keyword searches. Many projects use abbreviations, but there are not currently universal standards.

• **Geological formation**: Formal name of geological formation. Care should be taken to utilize the published geological formation name with established priority.

• **Geological member**: Formal name of geological member. Care should be taken to utilize the published geological member name with established priority.

• **Stratigraphic horizon**: Description of geological unit containing the specimen. Ge­nerally this description provides stratigraphic information beyond the geological member, and care should be taken to provide as much detail as possible.

• **In situ? (Y/N)**: Was the specimen recovered in situ? This is always a discrete yes or no answer.

• **Estimated upper age limit (ma)**: Estimated specimen younger age limit in millions of years.

• **Upper limit age basis**: Basis for upper (younger) age limit estimate.

• **Estimated lower age limit (ma)**: Estimated specimen older age limit in millions of years.

• **Lower limit age basis**: Basis for lower (older) age limit estimate.

• **Sediment or matrix adhering?**: Indication from adhering matrix or sediment that the specimen was extracted from an in situ position. This specifically does not apply to adherent recent residue.

• **Repository**: Institution formally recognized as owner of the specimen.

• **Has a replica been made? (Y/N)**: Indication of whether a cast has been made of the specimen.

• **Curatorial problem? (Y/Never/Corrected)**: Indication of occurrence of curatorial problem. See below for examples of curatorial problems. If a curatorial problem existed, but was later corrected, indication of this should be made such that the database becomes an archive of curatorial work with specimens through museum time.

• **Curatorial problem description**: Description of curatorial problem. Curatorial prob­lems include lost specimens, corrected specimen numbers, specimens where MNI is found to be greater than one, specimens that have illegible numbers, speci­mens that have been loaned for which return is overdue, etc. Curatorial problem descriptions may indicate that the problem was corrected, but recorded problems should never be deleted and should be archived in curatorial notes upon correc­tion of the curatorial problem.

• **Curatorial notes**: Curatorial notes include descriptions of corrected problems, loans, movements, damage, and other phenomena affecting the physical disposi­tion of the specimen.

• **Taxonomic problem? (Y/N)**: Indication of occurrence of a taxonomic problem. Taxonomic problems generally result from discrepant identifications among specialists or from the inability of a collection manager to confidently identify a specimen to the level of precision perceived possible.

• **Taxonomic identifier’s names and dates**: Names of identifiers and dates for identifi­cation of each taxonomic revision that triggers a change to the catalog.

• **Taxonomic notes**: Notes on taxonomic identification and revision. These are recor­ded by both specialists and catalog managers, and care should be taken by the catalog manager to acquire any notes from specialists.

• **Taxon**: It is advisable to establish a standardized taxonomic lookup table using a singular source to insure consistency in application of higher taxonomic nomen­clature to identified specimens. It is useful to separate prefixes like cf., aff., and sp. from the Linnaean nomen to facilitate efficient data management: Class prefix, Class; Order prefix, Order; Suborder prefix, Suborder; Infraorder prefix, Infraorder; Superfamily prefix, Superfamily; Family prefix, Family; Subfamily prefix, Subfa­mily; Tribe prefix, Tribe; Genus prefix, Genus; Species prefix, Species; Subspecies prefix, Subspecies.

• **Tool type/description**: Type of artifact collected.

• **Technique/industry**: Name of industry or technique of manufacture. This field can be more verbose than ‘Tool type.’

• **Raw material**: Raw material of collected artifact.

• **Artifact identifier’s name and dates**: Names of identifiers and dates of identifica­tion of each archaeology revision that triggers a change to the catalog.

• **Archaeology notes**: Notes on artifact identification and analysis. These are recor­ded by both specialists and catalog managers, and care should be taken by the collection manager to acquire any notes from specialists.

• **Formal excavation**: Identifier of formal excavation from which specimen was deri­ved.

• **Collection procedure**: Specific procedure used to obtain specimen.

• **Reason for collection**: Reason for collecting specimen.

• **Collector‘s name**: The collector is the person who found or encountered the spe­cimen. While under some circumstances it may not be correct to assign discovery to a single person, as, for example, in excavations, it is not advisable to compro­mise this field by including more than one person’s name. In our experience, it is either a single person or simply ‘group.’

• **Collection date**: Date of collection of specimen.

• **Latitude**: Latitude in a standardized format (for example, decimal degrees).

• **Longitude**: Longitude in a standardized format (for example, decimal degrees).

• **Geographic imagery reference**: Reference to satellite or air photo imagery source that records geographic origin of the specimen.

• **Specimen geographic location notes**: Notes on geographic origin of specimen.

• **Geography reference provider**: Name of remote geographic information provider. For example, if a DGPS service is used, the name of the service provider and the correction signal type and frequency should be entered. If standard GPS is used, it should be noted here.

• **Geography reference date**: Date of geographic reference acquisition.

• **Geography reference type**: GPS, DGPS, total station, georeferenced image, etc.

• **GPS unit (if applicable)**: Brand and model of GPS unit.

• **Geo-reference notes**: Notes on method used for georeferencing, including infor­mation on grid or alternative coordinate system.

• **Elevation**: Elevation in meters.

• **Elevation type**: Basis of elevation. Examples include GPS, altimeter, and topogra­phic map.

• **Elevation provider**: Specific information on elevation provider. Examples include cartographer or DGPS service provider.

• **Elevation date**: Date elevation was recorded.

# Locality table schema

• **Project**: Unique identifier of study area or project.

• **Locality ID**: Alphanumeric code that uniquely identifies locality. Avoid characters that have a function in common query and markup languages (like quotes, per­cent signs, parentheses, semicolons, backslashes, etc.). Although not absolutely necessary, a good rule of thumb is to use characters that would work in a MS-DOS filename.

• **Locality common name**: Name used to refer to locality. Often this is the local name of the vicinity of the locality.

• **Geological member**: Formal name of geological member to which the locality be­longs. Care should be taken to utilize the published geological formation name with established priority. Only one member should be entered. If a locality exposes more than one formal member, the less common member(s) should be discussed in Stratigraphic interval.

• **Geological formation**: Formal name of geological formation. Care should be taken to utilize the published geological formation name with established priority (Sal­vador 1994)

• **Stratigraphic interval**: Detailed description of stratigraphic interval represented at the locality. Care should be taken to avoid redundancy between stratigraphic interval description and locality description, although some overlap is inevitable. This field has no length limitations and should include as much detail as possible.

• **Estimated upper age limit (ma)**: Estimated locality younger age limit in millions of years.

• **Upper limit age basis**: Basis for upper (younger) age limit estimate.

• **Estimated lower age limit (ma)**: Estimated locality older age limit in millions of years.

• **Lower limit age basis**: Basis for lower (older) age limit estimate.

• **Locality description and boundaries**: Description of locality using landscape fea­tures. Care should be taken to avoid botanical features and to minimize redundan­cy between locality description and stratigraphic interval.

• **Locality dimensions (N/S)**: Approximate north to south locality dimensions in me­ters.

• **Locality dimensions (E/W)**: Approximate east to west locality dimensions in meters.

• **Uncollected taxa present**: List of taxa represented among uncollected fossils. Taxo­nomic nomenclature should follow the taxonomic code adopted for the database.

• **Archaeological evidence**: Detailed description of archaeological evidence associa­ted with the site.

• **Macrobotanical evidence**: Detailed description of non-microscopic botanical evi­dence associated with the site.

• **Locality discoverer**: Person or persons responsible for discovering locality. Un­like specimens, because it is potentially unrealistic to establish the first specimen found, localities may list the names of more than one discoverer.

• **Locality discovery date**: Date of locality discovery.

• **Locality recorder**: Person recording locality information.

• **Locality recording date**: Date of locality information recording.

• **Latitude**: Latitude of approximate center of locality.

• **Longitude**: Longitude of approximate center of locality.

• **Geographic imagery reference**: Reference to satellite or air photo imagery source that records geographic position of the locality.

• **Geography reference provider**: Name of remote geographic information provider. For example, if a DGPS service is used, the name of the service provider and the correction signal type and frequency should be entered. If standard GPS is used, it should be noted here.

• **Geography reference date**: Date of geographic reference acquisition.

• **Geography reference type**: GPS, DGPS, georeferenced image, etc.

• **GPS unit (if applicable)**: Brand and model of GPS unit.

• **Geo-reference notes**: Notes on method used for georeferencing. Excavation grids, total station systems, and other sub-locality geo-spatial systems should be menti­oned here.

• **Elevation**: Elevation in meters.

• **Elevation type**: Basis of elevation. Examples include GPS, altimeter, and topographic map.

• **Elevation provider**: Specific information about elevation information. Examples in­clude topographic map cartographer and DGPS service provider.

• **Elevation date**: Date elevation was recorded.

• **Field notes**: Notes on locality taken in field.

• **Repository notes**: Notes on locality derived from museum work.

• **Data notes**: Notes about locality data handling.