

# SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

## Eyasi Plateau Paleontological Expedition, Laetoli, Tanzania, fossil specimen database 1998–2005

Denné Reed <sup>1\*</sup>, Terry Harrison <sup>2</sup> & Amandus Kwekason<sup>3</sup>

The Eyasi Plateau Paleontological Expedition (EPPE) Laetoli specimen database contains 13716 records of plant and animal fossils (ca. 28248 specimens) collected by EPPE field teams working at Laetoli, Tanzania between 1998 and 2005. This dataset is a digital version of the original hard-copy specimen catalog, and it documents the discovery, stratigraphic provenience and taxonomic diversity of Plio-Pleistocene fauna and flora in northern Tanzania between 4.4 Ma and >200 ka. Laetoli is renowned for the discovery of important hominin fossils, including the lectotype for *Australopithecus afarensis*, one of our early hominin ancestors, the first record of *Paranthropus aethiopicus* outside Kenya-Ethiopia, and an early record of our own species *Homo sapiens*. This database is one of the few publicly available palaeoanthropological fossil datasets and serves as an example for expanding open access to primary fossil occurrence data in palaeoanthropology. The taxonomic identifications appearing in this dataset are the original field identifications and are provisional. Any taxonomic analysis employing this dataset should refer to updated taxonomic identifications published by specialists.

### Background & Summary

Laetoli in northern Tanzania is one of the most important palaeontological and palaeoanthropological sites in eastern Africa. The site is located on the Eyasi Plateau, an uplifted fault block at the southern end of the eastern branch of the East African Rift Valley, bordered to the south by Lake Eyasi (Fig. 1). To the east is the Ngorongoro Volcanic Highland Complex, comprising a series of Pliocene and Pleistocene volcanoes. These are the source of the primary air fall tuffs and reworked tuffaceous sediments in which the fossils are buried and preserved<sup>1</sup> (Fig. 2). Tephra throughout the stratigraphic sequence are amenable to radiometric dating and this has allowed an excellent geochronological framework to be established<sup>2</sup>. Fossils were first discovered in the area in the 1930s (when the site was referred to as Garusi), but the significance of Laetoli for human evolution was not fully appreciated until the late 1970s when Mary Leakey began her field research in the area<sup>3</sup>. The current phase of geological and palaeontological fieldwork, co-directed by TH and AK, was initiated in 1998 and continues to the present-day<sup>4,5</sup>. The database comprises data derived from the first seven years (8 field seasons) of renewed paleontological collections (1998–2005).

Laetoli is perhaps best known for the discovery of important fossil remains and ichnological traces of early hominins. The site has yielded hominins from three stratigraphic units: Upper Laetolil Beds, Upper Ndolanya Beds, and Upper Ngaloba Beds (Fig. 2). The specimens of *Australopithecus afarensis* from the Upper Laetolil Beds (3.85–3.6 Ma) are relatively few, but nevertheless represent one of the largest and geologically oldest assemblages, including the lectotype of the species<sup>6</sup>. The Upper Laetolil Beds also preserve several trails of footprints of *Au. afarensis* that provide direct evidence of early hominin bipedal behaviour<sup>3,7–9</sup>. Hominins from the Upper Ndolanya Beds (2.66 Ma) include the first specimen of *Paranthropus aethiopicus* recovered from outside the Turkana Basin of northern Kenya and Ethiopia, and one of the oldest securely dated specimens attributable to this species<sup>6</sup>. Finally, a partial cranium of an archaic *Homo sapiens* has been recovered from the Late Pleistocene Upper Ngaloba Beds (>200 ka) associated with Middle Stone Age artefacts<sup>10,11</sup>. In addition to the hominins, a rich record of the fossil remains of animals and plants from Laetoli offers important insights into the faunal and floral diversity of Africa during the Pliocene, and provides a well-dated reference for comparisons with other Plio-Pleistocene faunas from Africa and Eurasia.

<sup>1</sup>Department of Anthropology, University of Texas at Austin, Austin, USA. <sup>2</sup>Center for the Study of Human Origins, Department of Anthropology, New York University, New York, USA. <sup>3</sup>National Museum of Tanzania, Dar es Salaam, Tanzania. \*email: [reedd@austin.utexas.edu](mailto:reedd@austin.utexas.edu)



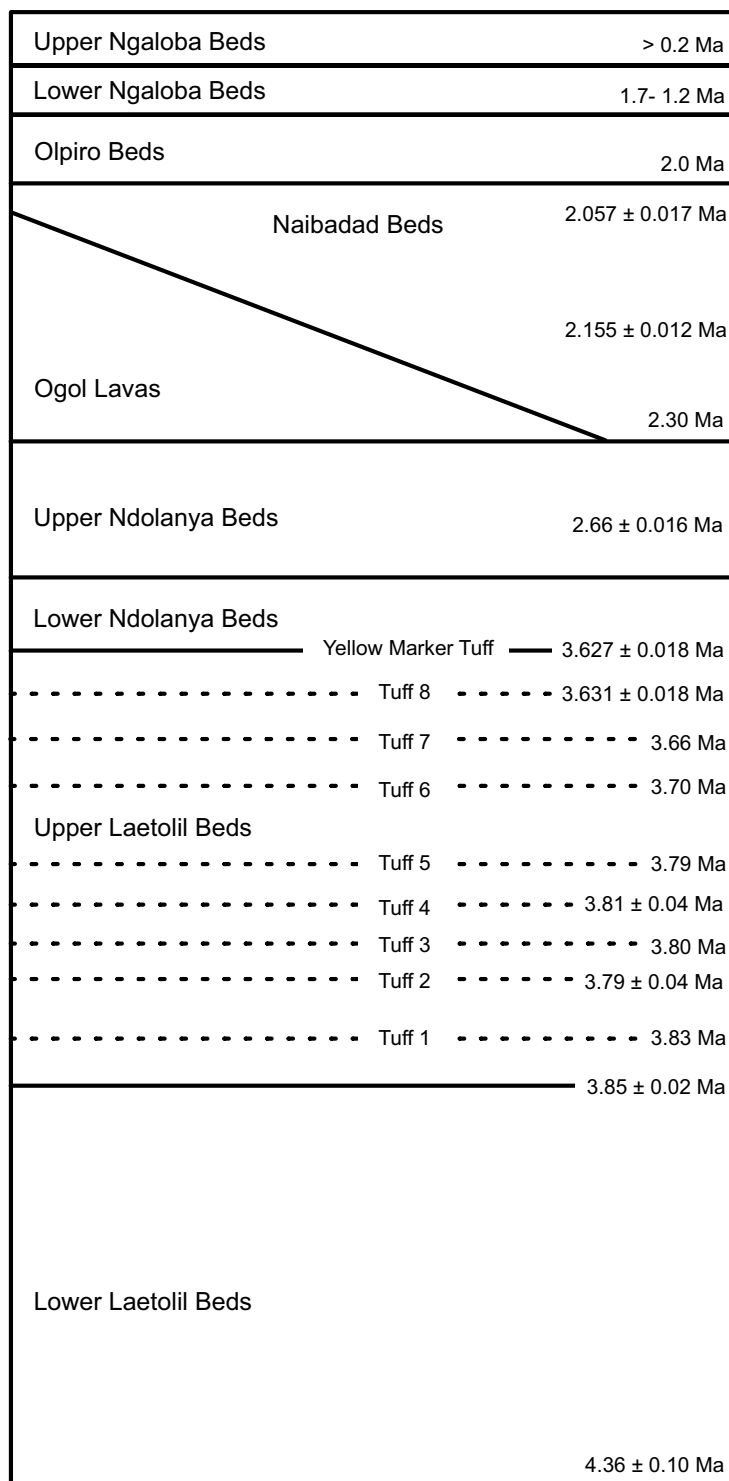
**Fig. 1** Map showing the location of Laetoli relative to other paleoanthropology fossil sites.

The database represents an important addition to the resources currently available for researchers investigating human evolution and vertebrate palaeontology in Africa. Field data of this kind, which provide crucial documentation about the nature and history of fossil collections, is rarely available to other researchers, and in the past essential contextual data about historical collections have been lost. For example, in 1938–1939 Ludwig and Margrethe Kohl-Larsen made one of the most important collections of fossil vertebrates from Laetoli<sup>12–14</sup>, which is housed in the Museum für Naturkunde in Berlin and the University of Tübingen. However, no comprehensive documentation of the collecting localities and stratigraphic provenance of the specimens was made at the time of discovery (or at least none that survives to the present-day), so crucial information about the context is largely unknowable and this greatly lessens the value and significance of the Kohl-Larsen collections. Making the Laetoli field data available ensures that future researchers have access to the history and contextual information relating to the discovery of individual specimens. As such, the database becomes an important historical resource. The database also provides important information, especially when combined with corresponding data from other palaeontological sites in Africa, which can be used for analyses of palaeoecology, palaeobiogeography, taphonomy, biochronology, and macroevolutionary patterns of speciation and extinction. Finally, the database can be used in broader-scale analyses of the impact of regional and global climate change on biotas during the Pliocene.

## Methods

The data were collected and processed following the steps outlined in Fig. 3, from field collection and hard-copy documentation, to digitization, alignment/import, cleaning/harmonization, and metadata annotation.

**Field collection and documentation.** The original fossils were recovered from the Eyasi Plateau, an uplifted fault block on the northwest margin of Lake Eyasi, located in the Ngorongoro Conservation Area, Tanzania (3.25°S, 35.10°E). Most of the fossils were recovered from the Laetoli area, but smaller collections have been recovered from Kakesio and Esere-Noiti. The project area covers approximately 400 square km. Fossils were primarily recovered from the surface of exposed outcrops after they have eroded out of the sediments. Partially exposed fossils *in situ* were excavated. No systematic screening for microinvertebrates and microvertebrates was undertaken, although dry screening methods were employed to recover associated remains and at localities where fossil hominins were recovered. The collection protocol stipulates collecting all vertebrate fossils that were anatomically identifiable (with the exception of rib fragments and limb bone shaft fragments that did not retain at



**Fig. 2** Summary of stratigraphic units and associated radiometric dates at Laetoli.  $^{40}\text{Ar}/^{39}\text{Ar}$  dates are given with standard errors.

least a portion of one articular surface). Bone fragments that were not anatomically identifiable but preserved traces of taphonomic interest (such as carnivore bite marks, rodent gnawing, cut marks, insect damage, and root etching) were also collected. Isolated fragments of tortoise shells and ostrich eggshell, terrestrial gastropods, insects and insect traces, and macrobotanical fossils were not collected systematically, but representative specimens were collected at each locality as reference specimens. Collecting events occurred at 60 designated localities and sublocalities within specific stratigraphic units in those localities<sup>15</sup>. Fossils were cataloged the same day they were discovered and field numbers were inscribed on the fossils with permanent ink. Collection details (date of collection, locality, stratigraphic unit, anatomical element, taxonomic identification, other remarks) for each



**Fig. 3** Data collection and processing workflow.

Year	Record Count (modifications)
1998	1691 (1693 – 2 duplicate)
1999	377
2000	4377 (4374 – 2 duplicates + 5 splits)
2001	1576 (1575 + 1 split)
2003	2540
2004	1749 (1754 – 5 duplicates)
2005	1406 (1407 – 1 duplicates)
Total	13716 (13720 – 10 duplicates + 6 splits)

**Table 1.** Number of records imported by year.

fossil were written onto collection cards that remain with the fossil and details were also written into a hard-copy collection catalog. Preliminary taxonomic identifications included in the catalog are based on expertise and literature sources. All specimens were accessioned into the collections of the National Museums of Tanzania (NMT), Dar es Salaam.

**Digitization.** Original records of the collected materials were written into a hard-copy catalog that is kept at the National Museums of Tanzania (NMT), Dar es Salaam, Tanzania. Entries from the paper catalog were digitized into seven spreadsheet files, one for each field campaign in 1998, 1999, 2000, 2001, 2003, 2004, 2005. In all there were eight field campaigns, with two separate campaigns in 2000 (one in January–February and another in August). Data from the spreadsheets were imported into the Paleo Core data repository (<http://paleocore.org>), aligned to standard fields and harmonized to established vocabularies and formats as described below. In total 13720 records were read from the spreadsheets, of these 10 were deleted as duplicates and 6 records were added as the result of splitting bulk records, resulting in the final count of 13716 data records (Table 1).

**Alignment and import.** Data from the digitized spreadsheets were mapped to a set of verbatim fields that record the original values from the spreadsheets in the Paleo Core database. The data in the verbatim fields were then processed and used to populate the cleaned fields in the database (Table 2). Two (2) of the spreadsheet columns, “Tray” and “Published” contained no data and were dropped.

	Spreadsheet Column	Verbatim Field	Cleaned Field
1	Specimen Number	verbatim_specimen_number	catalog_number
2	Date Discovered	verbatim_date_discovered	date_discovered
3	Storage	verbatim_storage	institution_code
4	Tray	N/A	N/A
5	Locality	verbatim_locality	locality
6	Horizon	verbatim_horizon	bed
7	Element	verbatim_element	part_of_organism
8	Kingdom	verbatim_kingdom	kingdom
9	Phylum/Subphylum	verbatim_phylum_subphylum	phylum, subphylum
10	Class	verbatim_class	class
11	Order	verbatim_order	order
12	Family	verbatim_family	family
13	Tribe	verbatim_tribe	tribe
14	Genus	verbatim_genus	genus
15	Species	verbatim_species	specific_epithet
16	Other	verbatim_other	taxon_remarks
17	Comments	verbatim_comments	remarks
18	Published	N/A	N/A
19	Problems	verbatim_problems	problem_remark

**Table 2.** Mapping between the 19 columns in the spreadsheet files and the corresponding verbatim fields and cleaned fields in the dataset. Two spreadsheet columns, Tray and Published, contained no data and were dropped.

Standard	Namespace URI	Abbreviation
Darwin Core	<a href="http://rs.tdwg.org/dwc/terms/">http://rs.tdwg.org/dwc/terms/</a>	dwc
ABCD-EFG	<a href="http://www.synthesys.info/ABCDEFG/1.0">http://www.synthesys.info/ABCDEFG/1.0</a>	abc

**Table 3.** Standards, namespaces and abbreviations.

**Cleaning and harmonization.** For each of the 17 columns imported into Paleo Core the data were cleaned, and where appropriate, harmonized to a data encoding scheme and structured vocabulary. Details on this process are described in detail below for each field. Alignment and harmonization were automated in Python, so that every step in the process is documented in code and reproducible from the original Excel files.

### Data Records

**Data files.** This dataset comprises a single data file in comma delimited format (.csv). The first row is a header of column names matching standard terms. The dataset is available for download from the Paleo Core data repository at: <https://paleocore.org/projects/epee/> and the figshare data repository<sup>16</sup> at: <https://doi.org/10.6084/m9.figshare.8847935.v2>.

**Field definitions.** The data fields (columns) included in this dataset are of two types: verbatim fields and cleaned fields. The verbatim fields contain the uncleaned, digitized data copied from the spreadsheets. They provide a record of the digitized version of the paper specimen catalog. The cleaned fields contain data cleaned and harmonized from the verbatim data and mapped to one of the standards listed in Table 3. Details about the processing are described in the sections describing each field, and in the Python code for the import script.

All field names are presented in snake\_case (i.e. in lower-case and words are joined by underscores) to standardize their presentation and promote readability.

**Verbatim fields.** The spreadsheets contained 19 columns of data. Seventeen (17) columns were copied, unmodified, into corresponding “verbatim” fields in the final dataset (Table 2). Dates in the spreadsheets were a mixture of string values and integer date formats, which were parsed accordingly and converted to a standardized format in the database, all other fields were copied, unmodified, as text. Two (2) of the spreadsheet columns, “Tray” and “Published” contained no data and were dropped.

**Cleaned fields.** In addition to the verbatim fields the dataset includes the fields listed in Online-only Table 1. The data from the verbatim fields were cleaned, validated and used to populate these fields. Unless otherwise indicated, each field is defined according to the Darwin Core standard<sup>17</sup>. One field, verbatim\_element, does not conform well to an existing Darwin Core term and was aligned to the term PartOfOrganism from the ABCD-EFG data standard. Two of the taxonomic fields, verbatim\_phylum\_subphylum and verbatim\_tribe also are not represented by Darwin Core terms. The column for verbatim\_phylum\_subphylum was divided into dwc:phylum and subphylum, the latter is not standard but its definition is similar to that of other taxonomic fields in Darwin Core.



Format	Example 1	Example 2
EP <item_number> [item_part]/<yy>	EP 001/98	EP 1345a/05

**Table 4.** Formatting and examples of entries in the catalog\_number field.

Similarly, *verbatim\_tribe* was cleaned and transferred to *tribe*. The *tribe* field is particularly significant in the analysis of Plio-Pleistocene bovid faunas where many fossils cannot be identified to genus level but can be identified to ecologically informative tribe level designations. A complete listing of all the cleaned fields and their definitions is provided in Online-only Table 1. Additional comments about each field are provided in the subsequent sections.

*Catalog number.* Field name – *catalog\_number*.

Definition – *dwc:catalogNumber*.

Catalog Number is the primary key (pk) for the EPPE Laetoli Database and as such is unique for all records, but not guaranteed globally unique outside this dataset. Catalog numbers match the values written on the fossil specimens and take the form shown in Table 4, where EP stands for Eyasi Plateau, <item\_number> indicates a three- or four-digit unique (to this dataset) integer, [a-z] indicates an optional, lower-case, lettered part of a specimen (Example 2) and yy indicates a two-digit year. Item numbers less than 100 have leading zeros to the hundreds place, as shown in Example 1, though values may expand to the thousands as indicated in Example 2. There is always a single space between EP and the item number. Leading zeros were retained for consistency with published specimen numbers.

*Institution code.* Field name – *institution\_code*.

Definition – *dwc:institutionCode*.

All values for this field are the string, ‘NMT’ for the National Museums of Tanzania.

*Collection code.* Field name – *institution\_code*.

Definition – *dwc:collectionCode*.

All values for this field are the string ‘EP’ for Eyasi Plateau. This collection code distinguishes these collections from others made at Laetoli, e.g. earlier Leakey collections under code ‘LAET’ or ‘LIT’.

*Occurrence remarks.* Field name – *occurrence\_remarks*.

Definition – *dwc:occurrenceRemarks*.

This field is copied from *verbatim\_comments*. No additional processing was applied to the *verbatim* data.

*Event date.* Field name – *event\_date*.

Definition – *dwc:eventDate*.

The values for event date are derived from *verbatim\_date\_discovered*, which in the spreadsheet files were a mix of dates in number format where the value corresponds to the number of days (or fractions thereof) since a designated start date (which is stored in the workbook), or the entries were string values in the format dd/mm/yy, where dd corresponds to the day, mm to the month and yy to the year, all as two digits. Dates were converted to Python date format using the *xlrd* Python library depending on the data type of the cell in the spreadsheet (date type vs string type). Each record was validated for a value that fell in the interval between 1998 and 2005 inclusive.

*Basis of record.* Field name – *basis\_of\_record*.

Definition – *dwc:basisOfRecord*.

This is a Darwin Core term used to indicate the type of data record. For this dataset all records have the value ‘FossilSpecimen’, which is from the recommended Darwin Core type vocabulary.

*Part of organism.* Field name – *part\_of\_organism*.

Definition – *abc:PartOfOrganism*.

The *part\_of\_organism* field records free text description of the fossil elements preserved for each specimen. As there is no Darwin Core term to handle free text anatomical element descriptions, this field definition is drawn from the Access to Biological Collections Data standard (ABCD) Extended for Geology (EFG). Values in the description field were pared down to 2293 unique descriptions from the 3976 unique entries in *verbatim\_element*. The reduction was accomplished by standardizing or expanding abbreviations for anatomical elements, parts and sides.

*Organism quantity.* Field name – *organism\_quantity*.

Definition – *dwc:organismQuantity*.

The number of identified fossil specimens (NISP) included with each record. The values in this field derive from the item counts appearing in parentheses in *verbatim\_element*, e.g. “Proximal Ulnae (3)”. The anatomical description is preserved in *part\_of\_organism* and the number of specimens appearing in parentheses is recorded in *organism\_quantity* as an integer value, without parentheses.

*Organism quantity type.* Field name – *organism\_quantity\_type*.

Definition – *dwc:organismQuantityType*.

Format	Example 1	Example 2
<place> <number> [number range]	Laetoli 1	Kakesio 2–4
<place> <number> <sublocality>	Laetoli 10 East	Laetoli 13 “Snake Gully”
<place>	Engesha	Garusi Southwest

**Table 5.** Formats and examples for entries in the locality field.

This field is set to ‘NISP’ for all records, indicating the number of identified specimens as the quantity expressed in Organismal Quantity.

*Country.* Field name – country.

Definition – dwc:country.

All values for this field are the string, ‘Tanzania’.

*Locality.* Field name – locality.

Definition – dwc:locality.

A locality is the place within the Laetoli project area from where a specimen was recovered. Harrison and Kweka<sup>15</sup> list 60 fossil localities at Laetoli, which they group into three major areas: Laetoli, Kakesio, and Noiti-Esere (Fig. 1). Three localities (Ndoroto, Olaltanaudo, Oleisusu) fall outside these areas. The 219 unique entries in *verbatim\_locality* were cleaned to match one of the 65 locality terms in the Laetoli locality vocabulary. The additional 5 entries in the vocabulary correspond to conflated values such as ‘Kakesio 1–6’ which indicate that the fossils came from one of the 6 Kakesio localities but it is unclear which one. Locality names follow one of the formats shown in Table 5.

Locality place and number are always separated by a single space, and ranges are indicated by a single n-dash with no spaces around it. All entries were stripped of leading and trailing whitespace.

*Bed.* Field name – bed.

Definition – dwc:bed.

Each fossil derives from a stratigraphic unit or range of units, which comprise the geological context for the fossil. Most collections were surface finds and their provenience is inferred from the surrounding sediments, adhering matrix, and preservation. Geological samples from volcanic tuffs were submitted for radiometric analysis to determine geochronological age controls for the site<sup>3</sup>. The original 196 unique entries in *verbatim\_horizon* were pared down to 34 unique entries in the cleaned, bed field and the values correspond to the units shown in Fig. 2.

*Minimum/maximum chronometric age.* Field names – *minimum\_chronometric\_age*, *maximum\_chronometric\_age*.

Definitions – dwc:minimumChronometricAge, dwc:maximumChronometricAge.

The age fields are drawn from the Chronometric Age extension to Darwin Core. The minimum and maximum are based on the absolute age estimates for each unit as illustrated in Fig. 2. The dates provided are the best estimates (central tendency) for the minimum and maximum age respectively. Most ages were determined using <sup>40</sup>Ar/<sup>39</sup>Ar radiometric dating, with Bayesian interpolation for beds between dated tuffs<sup>2,18</sup>. Dates for the Upper Ngaloba are based on Amino Acid Racemization while dates for the Lower Ngaloba Beds are based on biochronology. Each fossil was ascribed dates based on the bed or interval of beds from which it was recovered.

*Maximum/maximum chronometric age reference system.* Field names – *minimum\_chronometric\_age\_reference\_system*, *maximum\_chronometric\_age\_reference\_system*.

Definitions – dwc:minimumChronometricAgeReferenceSystem, dwc:maximumChronometricAgeReferenceSystem.

All dates for minimum and maximum chronometric age are provided in Ma (megaannum, i.e. millions of years).

*Chronometric age uncertainty in years.* Field name – *chronometric\_age\_uncertainty\_in\_years*.

Definition – dwc:chronometricAgeUncertaintyInYears.

While the max and min ages are provided in Ma, the uncertainty is provided in years to conform with the standard. In cases where uncertainty differed between minimum and maximum ages, or where uncertainty is known only for one of the values, the greatest uncertainty is reported. For example, fossils found in the Laetoli Beds, Upper Unit, Between Tuffs 7 – 8, are bracketed between a maximum age of 3.66 Ma and a minimum age of 3.631 ± 0.018 Ma. The maximum age is interpolated and the minimum age is based on argon-argon analysis and has a standard error of 0.018 Ma. The reported uncertainty for this example is thus 18000 years.

*Taxonomic fields.* Field names – kingdom, phylum, subphylum, class, order, family, tribe, genus, *specific\_epithet*.

Definitions – dwc:kingdom, dwc:phylum, dwc:class, dwc:order, dwc:family, dwc:genus, dwc:specific\_epithet.

The taxonomic fields, kingdom, phylum, class etc., record preliminary identifications for each specimen at the designated taxonomic rank and are derived from the **verbatim** taxonomic fields. Spelling and typographical errors were corrected, but no attempt was made to update taxonomic assignments. Values from the

verbatim\_phylum\_subphylum field were split into the cleaned fields phylum and subphylum accordingly. There are no standard terms in Darwin Core or ABCD-EFG for the taxon ranks subphylum and tribe. These ranks are included because of their significance in paleoanthropological faunal analysis. Their definitions are as for the other taxon fields.

**Scientific name.** Field name – scientific\_name.

Definition – dwc:scientificName.

This field records full name at the lowest level taxonomic designation available for the specimen. It is derived from the **cleaned** taxonomic fields (unlike the taxonomic fields which derive directly from the verbatim taxonomic fields). For genus and above it provides a single word and for species the complete species name (binomen). An important note about this field is that it does not include designations of uncertainty, for example *Australopithecus cf. afarensis* will have scientific\_name *Australopithecus afarensis* and identification\_qualifier cf. *afarensis*, as per the recommended best practice under Darwin Core. Users are encouraged to consult the Darwin Core documentation for dwc:scientificName and dwc:identificationQualifier for definitions and details.

**Identification qualifier.** Field name – identification\_qualifier.

Definition – dwc:identificationQualifier.

This field is derived from the verbatim taxonomic fields. The verbatim taxonomic fields were searched for incidents of ‘?, cf., aff, indet., sp., nov.’ and where found the taxonomic field and identification\_qualifier field were updated accordingly. The import script includes detailed notes on the regular expression searches and parsing used to update this field. Generally, the identification qualifiers are taken to follow common convention for open nomenclatures *sensu* Sigovini *et al.*<sup>19</sup>. Users are referred to the Darwin Core documentation on dwc:scientificName and dwc:identificationQualifier to understand their definition and use.

**Taxon rank.** Field name – taxon\_rank.

Definition – dwc:taxonRank.

This field was updated by analysing the taxon fields to identify the most precise rank for which there is a value. All ranks are presented in lower case.

**Taxon remarks.** Field name – taxon\_remarks.

Definition – dwc:taxonRemarks.

The values in this field were updated for cases where taxonomic changes were made such that taxonomic fields differ from verbatim taxon fields.

**Problem remarks.** Field name – problem\_remarks.

Definition – Remarks describing potential or known problems with a record.

This field derives from verbatim\_problems and provide free text remarks about possible or known problems associated with a data record. Twenty records have entries for verbatim\_problems, two of these referring to duplicate catalog numbers were removed. The remaining 18 problem remarks indicate potential missing or problematic fossil specimens. There is no standard term in Darwin Core or ABCD-EFG that matches this field. It could be accommodated in remarks, but we preferred to keep problem remarks separate to facilitate search.

## Technical Validation

Automated validation was conducted for catalog\_number, event\_date, locality, and the taxonomic fields.

**Catalog number.** All catalog\_number strings were validated for consistency of formatting and uniqueness.

Entries in the catalog\_number field were copied from verbatim\_specimen\_number and validated against the following Python regular expression:

```
EP \d{3,4} [a-zA-Z]?/[09][01234589]$
```

Leading zeros were retained for consistency with published specimen numbers.

**Splits.** Five items for bulk collections contained multiple or mixed taxa and had to be split into parts. The splits added 6 new items to the catalog (Table 6).

**Corrections.** Six (6) specimens raised validation errors (Table 7). Three (3) of these specimens had an entry in catalog\_number that did not match the formatting regular expression. The first, with verbatim specimen number EP 120A + B/98 is an associated set of left and right upper teeth and skull fragments. These were published<sup>20</sup> as parts A and B but exact designation for all elements was not given so the catalog number is merged in the official version of the catalog to simply EP 120/98. The other formatting errors are typos in the year suffix, which were corrected. Another three (3) were duplicate catalog numbers resulting from digitization errors. These were corrected as shown in Table 7.

**Deletions.** Ten (10) records were deleted from the dataset. Seven (7) records were duplicate entries with identical data (Table 8).

Two (2) pairs of duplicate records were the result of emended taxonomic identifications (Table 9, items 1–2). The amended data was retained, the earlier version deleted, and a note of the change was added to the taxonomic remarks field. One (1) duplicate stems from an incorrect entry, which was deleted (Table 9, item 3).



	Catalog Number	Locality	Description	Taxon	Updated to
1	EP 1280/01	Laetoli 9 South	shells (2)	Achatina zanzibarica	1280a/01
			shells (5)	Pseudoglossula cf. gibbonsi	1280b/01
2	EP 3129/00	Laetoli 10	shells (8)	Achatina zanzibarica	EP 3129a/00
			shell (1)	Pseudoglossula cf. gibbonsi	EP 3129b/00
3	EP 1181/00	Laetoli 8	shell (1)	Achatina zanzibarica	EP 1181a/00
				Burtoa nilotica	EP 1181b/00
4	EP 3635/00	Kakesio 5	shell (1)	Achatina zanzibarica	EP 3635a/00
			shells (8)	Limicolaria martensiana	EP 3635b/00
5	EP 1177/00	Laetoli 8	nuchal bone	Stigmochelys brachygularis	EP 1177a/00
			xiphialastron	Stigmochelys brachygularis	EP 1177b/00
			fragment of carapace	Stigmochelys brachygularis	EP 1177c/00

**Table 6.** Specimens that were split into new records. Five (5) items were split resulting in 6 new records.

	Catalog Number	Locality	Description	Taxon	Updated to
1	EP 120A + B/98	Laetoli 10 East	assoc. teeth and cranial fragments	Rhinocerotidae	EP 120/98 and comment added.
2	EP 507/07	Laetoli 9 South	scute	Chelonia	EP 507/05
3	EP 756/06	Laetoli 5	lower incisor	Madoqua	EP 756/05
4	EP 1075/03	Laetoli 10 West	proximal phalanx	Rodentia	same
	EP 1075/03	Laetoli 7	distal metapodials	Serengetilagus	EP 1975/03
5	EP 348/04	Laetoli 8	proximal ulnae	Serengetilagus	same
	EP 348/04	Laetoli 8	distal radius	Serengetilagus	EP 349/04
6	EP 2188/99	Laetoli 7	distal humerus	Bovidae	EP 2188/03
	EP 2188/99	Laetoli 7 East	vertebral centrum	Bovidae	EP 2188/00

**Table 7.** Corrected catalog numbers.

	Duplicate Catalog Numbers (duplicate deleted)
1	EP 1582b/00
2	EP 1144/04
3	EP 1173/04
4	EP 1400/04
5	EP 1403/04
6	EP 1542/04
7	EP 515/05

**Table 8.** Deleted duplicate records, one copy of each was deleted.

	Catalog Number	Locality	Description	Taxon	Updated to
1	EP 001/98	Kakesio 1	vertebra	Crocodylidae	deleted
	EP 001/98	Kakesio 1	cervical vertebra	Rhinocerotidae	retained with comments added
2	EP 1477b/00	Laetoli 7 East	metatarsal	Suidae	deleted
	EP 1477b/00	Laetoli 7 East	right metatarsal iii	Felidae	retained with comments added
3	EP 1052/98	Laetoli 9 South	calcanei	Serengetilagus	—
	EP 1052/98	Laetoli 9 South	middle phalanx	Aves	deleted

**Table 9.** Deleted and corrected duplicate records reflecting amended taxonomic identifications.

**Date discovered.** Discovery dates were automatically validated for formatting and appropriate time intervals, i.e. all dates were tested to fall in the range between 1 Jan 1998 and 31 Dec 2005. Dates were converted to python date objects in the Paleo Core database, and exported as text strings in ISO 8601:2004(E) format.

**Locality.** Locality values were validated against the Laetoli locality vocabulary listed in Online-only Table 2. The Verbatim Locality column indicates the corresponding entries in the verbatim\_locality field that were harmonized to a common locality name. For example ‘Kakesio #1’ and ‘Kakesio #1’ were both harmonized to the value of ‘Kakesio 1’. Some verbatim entries are identical except for trailing whitespaces.

**Taxonomic fields.** Entries in the taxonomic fields were validated for spelling and taxonomic placement (in the appropriate hierarchy) against the iDigBio taxonomic backbone at genus level and above. Taxonomic validation was used to verify taxa, their spellings and their appropriate ranks. Three genera were not found in the iDigBio taxonomic backbone; these were confirmed against the original print catalog and retained.

### Usage Notes

These data represent the primary digital version of the specimen catalog and are suitable for analyses on differences between localities and comparisons with other fossil catalogs from Laetoli and other sites.

The taxonomic identifications and item descriptions are provisional and should be used with care. Taxonomic identifications are provided primarily to assist in data discovery and to assist taxonomic specialists and systematists in identifying specimens in their area of expertise or that might be relevant for subsequent analysis. Quantitative analyses making use of taxonomic information at higher levels (e.g. Tribe and above) may be appropriate but lower-level, fine-grained taxonomic analysis using these data is probably not appropriate without consulting published analyses of the fossils conducted by specialists<sup>5</sup>.

Similarly, the descriptions of anatomical elements preserved is meant as a guide to data discovery and further analysis. Quantitative analysis of element preservation or taphonomic processes based on these data is not recommended without further reference to published analyses of the specimens or confirmation with the physical specimens.

The digital version of this dataset is public and distributed under a Creative Commons CC-0 license<sup>16</sup>. The physical fossils specimens are under the jurisdiction of the National Museums of Tanzania and the Antiquities Division of the Ministry of Natural Resources and Tourism.

### Code availability

Data were imported into the Paleo Core data repository using Python (version 3.6) standard libraries (re, datetime, pytz), the xlr library (version 1.20) for reading Excel files, and the database API included with the Django web framework (version 1.11.20). All of the original source code used to process the data are freely and publicly available through the Paleo Core github repository at: [https://github.com/paleocore/paleocore110/blob/master/epe/import\\_1998\\_2005.py](https://github.com/paleocore/paleocore110/blob/master/epe/import_1998_2005.py).

Received: 12 July 2019; Accepted: 31 October 2019;

Published online: 03 December 2019

### References

- Ditchfield, P. & Harrison, T. Sedimentology, Lithostratigraphy and Depositional History of the Laetoli Area. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 1: Geology, Geochronology, Paleocology and Paleoenvironment* (ed. Harrison, T.) 47–76 (Springer Netherlands, 2011).
- Deino, A. L. <sup>40</sup>Ar/<sup>39</sup>Ar Dating of Laetoli, Tanzania. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 1: Geology, Geochronology, Paleocology and Paleoenvironment* (ed. Harrison, T.) 77–97 (Springer Netherlands, 2011).
- Leakey, M. D. & Harris, J. M. *Laetoli, a Pliocene site in northern Tanzania* (1987).
- Harrison, T. (Ed). *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 1: Geology, Geochronology, Paleocology and Paleoenvironment*. (Springer, Dordrecht, 2011).
- Harrison, T. (Ed). *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 2: Fossil Hominins and the Associated Fauna*. (Springer, Dordrecht, 2011).
- Harrison, T. Hominins from the Upper Laetoli and Upper Ndolanya Beds, Laetoli. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 2: Fossil Hominins and the Associated Fauna* (ed. Harrison, T.) 141–188 (Springer Netherlands, 2011).
- Leakey, M. D. & Hay, R. L. Pliocene footprints in the Laetoli Beds at Laetoli, northern Tanzania. *Nature* **278**, 317–323 (1979).
- Hay, R. & Leakey, M. The fossil footprints of Laetoli. *Sci. Am.* **246**, 50–57 (1982).
- Masao, F. T. *et al.* New footprints from Laetoli (Tanzania) provide evidence for marked body size variation in early hominins. *Elife* **5**, 1–29 (2016).
- Day, M. H., Leakey, M. D. & Magori, C. A new hominid fossil skull (LH 18) from the Ngoloba Beds, Laetoli, northern Tanzania. *Nature* **284**, 55–56 (1980).
- Magori, C. C. & Day, M. H. Laetoli Hominid 18: an early Homo sapiens skull. *J. Hum. Evol.* **12**, 747–753 (1983).
- Dietrich, W. O. Die säugetierpaläontologischen Ergebnisse der Kohl-Larsen'schen Expedition 1937–1939 im nördlichen Deutsch-Ostafrika. *Neues Jahrb. Mineral. Geol. Palaeontol. Abh. Abt. B*, **8**, 217–223 (1941).
- Dietrich, W. O. Ältestquartäre Säugetiere aus der südlichen Serengeti, Deutsch-Ostafrika. *Palaeontographica* **94A**, 43–133 (1942).
- Kohl-Larson, L. *Auf den Spuren des Vormenschen Forschungen, Fahrten und Erlebnisse in Deutsch-Ostafrika*. (Strecker und Schröder, 1943).
- Harrison, T. & Kweka, A. Paleontological Localities on the Eyasi Plateau, Including Laetoli. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 1: Geology, Geochronology, Paleocology and Paleoenvironment* (ed. Harrison, T.) 17–45 (Springer Netherlands, 2011).
- Reed, D., Harrison, T. & Kwekason, A. Eyasi Plateau Paleontological Expedition, Laetoli (Tanzania) fossil specimen database 1998–2005. *figshare*, <https://doi.org/10.6084/m9.figshare.8847935.v2> (2019).
- Wieczorek, J. *et al.* Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS One* **7**, 1–8 (2012).
- Mollel, G. F., Swisher, C. C., Feigenson, M. D. & Carr, M. J. Petrology, Geochemistry and Age of Satiman, Lemagurut and Oldeani: Sources of the Volcanic Deposits of the Laetoli Area. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 1: Geology, Geochronology, Paleocology and Paleoenvironment* (ed. Harrison, T.) 99–119 (Springer Netherlands, 2011).
- Sigovini, M., Keppel, E. & Tagliapietra, D. Open Nomenclature in the biodiversity era. *Methods Ecol. Evol.* **7**, 1217–1225 (2016).
- Hernesniemi, E., Giaourtsakis, I. X., Evans, A. R. & Fortelius, M. Rhinocerotidae. In *Paleontology and Geology of Laetoli: Human Evolution in Context: Volume 2: Fossil Hominins and the Associated Fauna* (ed. Harrison, T.) 275–294 (Springer Netherlands, 2011).

### Acknowledgements

The authors would like to gratefully acknowledge the National Museums of Tanzania in Dar es Salaam, Tanzania Commission for Science and Technology (COSTECH), Division of Antiquities, and Ngorongoro Conservation Area Authority for permission to conduct research in Tanzania. Funding for Laetoli field work was provided by NSF (Grants BCS-0309513, BCS-0216683 and BSC 1350023), the National Geographic Society, the Leakey

Foundation, and NYU University Research Challenge Fund grants. Funding for Paleo Core was provided by NSF (Grant BCS-1244735). We also wish to thank the editor and two anonymous reviewers for their helpful comments and suggestions.

### Author contributions

Denné Reed generated the original idea for a data publication, implemented the data transfer to the Paleo Core data repository, wrote the Python code, and conducted the data alignment, harmonization and validation. Amandus Kwekason and Terry Harrison co-directed the field work to collect the original fossils, supervised the original digitization of the data and validated the data against the fossil collections. Denné Reed and Terry Harrison contributed to the writing and revision of the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to D.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019