Overview
Context
This dataset was collected in the scope of the EUROEVOL project, representing the largest repository of faunal data from Neolithic Europe at the time of publishing. The time frame of the Neolithic in this part of the world broadly encompasses 8000–4000 BP, and is characterised by the spread of domestic plants and animals from the Near East alongside the development of novel ceramic and lithic technologies via two routes of dispersal; a more maritime one, linking the Levant with the Aegean coast of Turkey and Greece into the western Mediterranean, and a more continental one linking central and northwest Anatolia with the more eastern part of Bulgaria and into continental central Europe [1–4]. This spread of early agro-pastoral lifeways also correlates with fundamental changes in past human demography, ecology and social organization [5–6].

In keeping with this, we found correlations between the population patterns and changing economic patterns [8], as well as with investment in conspicuous monument construction and in the incidence of evidence for violence, which appears to be associated with societies exceeding their limits. We have also shown that the cultural transmission processes that produce distinctive patterns of similarity and difference in the archaeological record have recognisable signatures that can be identified from the archaeological material [9–11]. In addition, we have assessed the relationship between different dating approaches for the European Neolithic and demonstrated the underlying shape of the intensity of European Neolithic cultures through time [12].

Spatial coverage
Central and northwest Europe
Description: Poland, Germany, Austria, Switzerland, France, Czech republic, Denmark, Sweden, Belgium, Liechtenstein, Luxembourg, Netherlands and Britain and Ireland.
Northern boundary: + 64.622N
Southern boundary: + 42.618N
Eastern boundary: + 23.963E
Western boundary: – 10.457E

Temporal coverage
8000 BP – 4000 BP
Whilst the majority of data falls within this time range, some sites may have associated radiocarbon data that exceeds these boundaries.
Methods

The majority of data in this dataset was obtained directly from source publications, as well as Masters theses, PhD theses and occasional unpublished reports. There are five main components of the faunal data, including the NISP counts of all taxa found, summary age data, summary sex data, instances of pathology and biometric data. The dataset utilises the same recording system as the archaeobotanical [14] and radiocarbon [13] datasets, as evidenced in the full published MySQL database found at http://discovery.ucl.ac.uk/1469811/. The MySQL dump file, which contains the full relational database for the EUROEVOL project can be downloaded and opened in any SQL interface, or alternatively, imported into MS Access.

Steps

The data collection process involved three primary approaches: 1) contacting specialists in the field of archaeozoology to request specific site reports and publications; 2) Accessing libraries across the UK to access site monographs and journal articles; 3) visiting regional museums and university departments across Europe to access published and unpublished reports on faunal data from site excavations. Once a report had been obtained, both a

Figure 1: Map of northwestern Europe showing sample locations.
hardcopy and digital copy were made and these are kept on file at the Institute of Archaeology, UCL. Where necessary reports were translated, often using Google translate and other online translation programmes. Any additional data from the site e.g. stratigraphic information, radiocarbon dates etc. were also archived with the faunal report. If the site did not already exist in our database, its precise location was identified and recorded in decimal degrees. The site was then assigned a unique SiteID and details of the excavation and sampling strategy were recorded. An excel spreadsheet was made for each site, with the different faunal data components being assigned a separate worksheet, including all raw data relating to NISP, sex ratios, age trends, biometrics, pathologies and body part representation. At this stage of the data collection, the original format of the report was maintained. For example, if the analyst had recorded all bones at the context level, this information was retained in the excel spreadsheet. For the sake of consistency in the database all taxa were recorded as a seven-character code (TaxonCode) denoting genus (the first four characters) and species (the first three characters) affiliations. Once all the raw data had been copied, it was then aggregated at the level of the cultural unit, for example LBK, Michelsberg, Cerny, Chasséen etc. Each cultural unit was assigned a unique PhaseCode and these aggregated datasets were then recorded in the database. All faunal data could therefore be identified at either the site level (based on SiteID), or at the phase level (based on PhaseCode), and linked to other associated datasets, for example radiocarbon dates and archaeobotanical data. Each of the faunal tables are published alongside the associated spatial and temporal datasets and archaeobotanical data at [13] & [14].

Quality Control
We have adopted a fully inclusive approach to the data collection, including all faunal data, irrespective of the date of publication or original analyst. The only exclusion criteria at the data input stage relates to the biometrics. Due to variation in the measuring protocol of different analysts, we have only included measurements that follow the von den Driesch [15] standard. This provides the largest possible sample size of broadly comparable biometrics. Similarly, the raw counts of demographic data e.g. age and sex consisted of several different quantification techniques. Hence, we have summarised this information in the database, recording simply the age/sex trend e.g. ‘majority male’, majority female’ etc. Additional comments on the demographic trends are recorded in the relative notes fields. All records have been checked to ensure that they are standardised wherever possible. For example, synonym species names, such as Pagophilus groenlandicus and Phoca groenlandica have been recorded under a single taxonomic name to avoid duplication, in this instance Phoca groenlandica.

Constraints
Occasionally a faunal report lacked any relative counts of the species represented, and simply identified the species present. These have been included in order to permit a presence/absence analysis. Equally some analysts appear to have been more cautious in their identification of wild vs domestic taxa. For example, due to the potential misidentification of Sus scrofa domesticus as Sus scrofa ferus, or vice versa, some analysts chose to record all pig remains as Sus sp. This was equally the case with cattle (Bos taurus/Bos primigenius) and horse (Equus caballus/Equus ferus) A comparative spatio-temporal analysis of the domestic taxa should therefore work only with phases where the taxa have been recorded to species level. An additional constraint on the species identification was encountered with the sheep (Ovis aries) and goat (Capra hircus) remains. Due to the difficulties associated with distinguishing between the osteology of sheep and goat [16], the majority of ovicaprid remains were originally recorded as Ovis/Capra. It may be advisable therefore to group all sheep/goat remains in future analysis to ensure comparability between sites.

Another potential constraint on the data relates to the differential sampling strategies of the original analysts. At sites where sieving was undertaken, there is often a greater representation of fish and bird bone. Any future analyses on these species should therefore take into consideration the sampling strategy, which is noted in the FaunalPhases table.

(3) Dataset description

Object name

FaunalPhases – two files providing the data (EUROEVOL09-07-201516-34_FaunalPhases.csv) and field type definitions (FaunalPhases_fields.csv) for all phasecodes with associated faunal data, deposit type and method of recovery. The SiteID links to the CommonSites table described in the EUROEVOL Dataset 1: Sites, Phases and Radiocarbon Data.

FaunalTaxaList – two files providing the data (EUROEVOL09-07-201516-34_FaunalTaxaList.csv) and field type definitions (FaunalTaxaList_fields.csv) for the full taxonomic description in relation to the unique TaxonCode of all species represented in the database.

FaunalSpecies – two files providing the data (EUROEVOL09-07-201516-34_FaunalSpecies.csv) and field type definitions (FaunalSpecies_fields.csv) for records of all species identified, age and sex trends, associated pathologies, and body part representation.

FaunalBones – two files providing the data (EUROEVOL09-07-201516-34_FaunalBones.csv) and field type definitions (FaunalBones_fields.csv) for each bone assigned by PhaseCode and species with associated measurements.

FaunalBiometrics – two files providing the data (EUROEVOL09-07-201516-34_FaunalBiometrics.csv) and field type definitions (FaunalBiometrics_fields.csv) for all measurements (in mm) associated with each bone as identified in the FaunalBones table.

Data type
Primary and secondary data

Format names and versions
.csv, SQL
Creation dates
Some records were created in 2007-2010 as part of the AHRC funded ‘Origins and Spread of Stock-Keeping’ (OSSK) Project. However, the majority of records, and current MySQL database were created in 2010-2015.

Dataset Creators
The primary researcher responsible for the data collection was Katie Manning. Records collected as part of the OSSK project involved Barbara Stopp, and entry of the biometric data was assisted by Eva Fairnell and Rebecca Rennell.

Language
English

License
CC0

Repository location
The full relational database is available as a SQL dump file and the individual tables (e.g. FaunalPhases, FaunalTaxaList etc.) are available as .csv files at http://discovery.ucl.ac.uk/1469811/. The SQL dump file can be imported directly into any SQL-based RDMBS, such as MySQL Workbench or Sequel Pro, with all relationships maintained. Alternatively the .csv files can be imported into any other RDMBS such as MS Access.

Publication date
28/07/2015

(4) Reuse potential
This dataset comprises the largest single collation of animal bone data for the European Neolithic, and there remains considerable analytical potential for future researchers. The EUROEVOL data is particularly re-usable because the sample sizes are so large (>3 million NISP counts and >36,000 biometric measurements), permitting robust comparative analysis between sites and regions, and across time. Furthermore, all data is fully georeferenced, offering considerable spatial analytical potential. The data is linked to associated archaeobotanical and radiocarbon data from the same site offering considerable scope for further palaeoecological and palaeoeconomic analyses that incorporate both the plant and animal bone data and in depth temporal analyses. This dataset will prove most useful for archaeozoologists. However, it may also be of benefit to geographers and palaeoecologists interested in past species distribution. The data would also provide a useful training dataset for student archaeozoologists interested in developing quantification techniques and statistical analyses of processed archaeozoological data.

Competing Interests
The author declares that they have no competing interests.

Acknowledgements
A number of researchers from across Europe were extremely helpful in providing bibliographies, copies of published reports and access to libraries and museums. In particular I would like to thank the following people: Arek Marciniak, Jean-Denis Vigne, Peter Rowley-Conwy, Simon Davis, Lila Janik, Linas Daugnora, Jörg Schibler, Angéla von den Driesch, Manfred Schmitzberger, Aikaterini Glykou, Karlheinz Steppan, Roz Gillis, CyrilMarcigmy, Peter Bogucki, Finbar McCormick, Jorn Zeiler, Wil Van Neer, Els Thieren and Svenja Holtkemeier. I would also like to thank the two anonymous reviewers for their constructive comments.

References