

## Catch Per Unit Research Effort Supplementary Information

### Data Dictionary & Licensing Information

#### Explanation

This document describes the data files included within the supplementary information folder for the paper ‘Catch per unit research effort: sampling intensity, chronological uncertainty, and the onset of marine fishing in historic London’ (Orton, Morris, and Pipe, *Open Quaternary*). Data are derived from a snapshot of the MOLA (Museum of London Archaeology) internal database in October 2013. Unfortunately the raw data cannot be included for commercial/copyright reasons, and the files included here have been cleaned, filtered, and ‘anonymised’ (in the sense that original site codes have been replaced by arbitrary numbers) prior to upload. Data from sites not included in the study - due to location, dating, or complete lack of suitable wet-sieved zooarchaeological data - are excluded from **all** tables. Otherwise, all contexts, samples, and zooarchaeological specimens are included, whether they make it into the eventual analysis or are filtered out in the associated R code. For this reason, the numbers of records in each table may not always match those mentioned in the text.

The data files described below are licensed under a [Creative Commons Attribution 4.0 International License](#), and should be attributed to [MOLA](#).



The associated code, images, and this document itself are likewise licensed under a [Creative Commons Attribution 4.0 International License](#), but should be attributed to David Orton, James Morris, and Alan Pipe.



The *archSeries* package, an archive version of which is included here, is licensed by David Orton under the [GNU General Public License version 3](#).



#### period.csv

This file contains data on the chronological ‘periods’ (i.e. broad phases) into which each site is divided. Each row represents one period at one site.

*SITECODE* (character): unique identifier for the site from which the context derives. Original Museum of London site codes have been replaced with a numeric sequence to ‘anonymise’ the data.

*PERIOD* (character): the chronological ‘period’ (i.e. broad phase) to which each stratigraphic context belongs, within its parent site. These are the original codes as used in the site archive/publications.

*SITE\_P* (character): unique identifier for each ‘period’ within the overall dataset, created by concatenating *SITECODE* and *PERIOD* with an intervening dash.

*Start* (numeric): approximate start date for the period, in years AD. Where necessary, this is derived from a textual description such as “16th century” or “Roman, pre-Boudiccan”.

*End* (numeric): approximate end date for the period, in years AD. Where necessary, this is derived from a textual description such as “16th century” or “Roman, pre-Boudiccan”.

### **context.csv**

This file contains data on the individual stratigraphic contexts used in the study. Each row represents one stratigraphic unit, based on single-context recording.

*SITECODE* (character): unique identifier for the site from which the context derives. Original Museum of London site codes have been replaced with a numeric sequence to 'anonymise' the data.

*CONTEXT* (numeric): number given to each stratigraphic context within its parent site. These are the original codes as used in the site archive/publications.

*PERIOD* (character): the chronological 'period' (i.e. broad phase) to which each stratigraphic context belongs, within its parent site. These are the original codes as used in the site archive/publications.

*SITE\_C* (character): unique identifier for each stratigraphic context within the overall dataset, created by concatenating *SITECODE* and *CONTEXT* with an intervening dash.

*BASIC\_INT* (character): code for the interpretation of the deposit type. Links to **int\_codes.csv**.

*SITE\_P* (character): unique identifier for each 'period' within the overall dataset, created by concatenating *SITECODE* and *PERIOD* with an intervening dash.

### **sample.csv**

This file contains data on the wet-sieved 'environmental' samples from which the bones used in the study derive. Each row represents one sample.

*SITECODE* (character): unique identifier for the site from which the context derives. Original Museum of London site codes have been replaced with a numeric sequence to 'anonymise' the data.

*SITE\_S* (character): unique identifier for each wet-sieved sample within the overall dataset, created by concatenating *SITECODE* and *SAMPLE* with an intervening dash.

*SAMPLE* (integer): number given to each wet-sieved sample within its parent site. These are the original codes as used in the site archive/publications.

*SITE\_C* (character): unique identifier within the overall dataset for the stratigraphic context from which the sample derives. Created by concatenating *SITECODE* and *CONTEXT* with an intervening dash.

*WTS\_VOL* (numeric): volume of sediment from the sample that was wet-sieved and processed, in litres (nb. this may be a sub-sample).

*CONTEXT* (numeric): stratigraphic context from which the sample derives. These are the original codes as used in the site archive/publications.

## **ZOO.CSV**

This file contains the zooarchaeological data used in the study. Each row represents one or more fragments of a given taxon in a given stratigraphic context and/or wet-sieved sample.

*SITECODE* (character): unique identifier for the site from which the context derives. Original Museum of London site codes have been replaced with a numeric sequence to ‘anonymise’ the data.

*CONTEXT* (numeric): stratigraphic context from which the remains derive. These are the original codes as used in the site archive/publications.

*SAMPLE* (integer): wet-sieved sample from which the remains derive, where appropriate. These are the original codes as used in the site archive/publications.

*WEIGHT* (numeric): total mass of fragments, in grams.

*SPECIES* (character): three- or four-letter code for the taxon represented by each set of remains. Nb. this is often in fact a higher level classification rather than a species. Links to **species\_codes.csv**.

*FRAG\_COUNT* (integer): number of specimens (i.e. bone fragments) represented by each record.

*SITE\_S* (character): unique identifier within the overall dataset for the wet-sieved sample from which the remains derive. Created by concatenating *SITECODE* and *SAMPLE* with an intervening dash.

*SITE\_C* (character): unique identifier within the overall dataset for the stratigraphic context from which the remains derive. Created by concatenating *SITECODE* and *CONTEXT* with an intervening dash.

## **species\_codes.csv**

Look-up table of information on animal taxa within the database.

*SPECIES* (character): three- or four-letter taxon code. Nb. this is often in fact a higher level classification rather than a species.

*COMMON\_NAME* (character): typical common name for the taxon, in English.

*LATIN\_NAME* (character): accepted Latin name for the taxon at the point of first entry into database. May include English qualifiers e.g. “(DOMESTIC)”.

*CLASS* (character): class to which the taxon belongs (if vertebrate), in the traditional sense of “mammal”, “bird”, “fish”, “reptile”, or “amphibian” rather than strict phylogenetic terms. Invertebrates are coded simply as “inv”. Nb. added by authors; not in original database.

*FRESH\_MARINE* (character): for fish only, status as obligate marine (“marine”), obligate freshwater (“fresh”), catadromous/anadromous (“migratory”), or potentially estuarine/brackish (“fresh/marine”). Nb. added by authors; not in original database.

**int\_codes.csv**

Look-up table of information on deposit interpretations used within the dataset.

*BASIC\_INT* (character): code for each deposit-type interpretation.

*INTERP* (character): brief description of deposit type.

*INT\_GROUP* (character): higher-level grouping of deposit types. Nb. added by authors; not in original database.

*INT\_GROUP2* (character): second higher-level grouping of deposit types, as used in the analysis. Nb. added by authors; not in original database.