The collection of this dataset was carried out under the auspices of the Cultural Evolution of Neolithic Europe project (EUROEVOL) led by Professor Stephen Shennan, UCL. The dataset represents one of the largest collections of archaeobotanical data for the Neolithic of Europe (Figure 1), comprising c.8300 records for c.1500 different species, genera and families and representing over a million identified items. This is one of three datasets resulting from the EUROEVOL project; the other two comprising the core spatial and temporal structure of the project, including all radiocarbon dates (EUROEVOL Dataset 1) and zooarchaeological data (EUROEVOL Dataset 2) – http://discovery.ucl.ac.uk/1469811/.

Keywords: EUROEVOL; Neolithic Europe; archaeobotanical data; preservation conditions; recovery methods

Funding statement: The data collection was funded by the European Research Council by an Advanced Grant (# 249390) to Stephen Shennan for the EUROEVOL Project

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.
(2) Methods

The majority of data in this dataset was obtained directly from source publications, which included several PhD theses and unpublished reports. Qualitative and quantitative details for all identified taxa are included in the dataset, together with information on sample provenances, recovery methods, and preservation status. The dataset utilises the same recording system as the zooarchaeological and radiocarbon datasets, as can be seen in the full published MySQL database found at [http://discovery.ucl.ac.uk/1469811/](http://discovery.ucl.ac.uk/1469811/).

Steps

Archaeobotanical data collection involved two main approaches: 1) locating site monographs and journal articles online or in libraries; and 2) liaising with regional specialists in the UK and mainland Europe to request unpublished reports and less easily accessible published reports. Once a report had been obtained a hardcopy and digital copy were made and kept on file for future reference at the Institute of Archaeology, UCL. Where necessary reports were translated using Google translate and other online translation programmes. Any relevant site data, e.g., stratigraphic and contextual information, radiocarbon dates, etc., were also archived with the archaeobotanical report. If the site did not already exist in our database its precise location was identified and recorded in decimal degrees. Each site was then assigned a unique SiteID and details of excavation, sampling strategy and recovery methods were recorded. The archaeobotanical data for each site were initially entered on a separate excel spreadsheet; sample-by-sample information for each taxon (either count or presence data) was recorded, as were context descriptions, volumes sampled and any other relevant details given in the original report. All taxa and individual plant parts listed by the original analyst were included and recorded as a seven-character code (TaxonCode) denoting genus (the first four characters) and species (three characters) affiliations. Each record was assigned a level identification (LevelOfIdentification) based on the degree of accuracy cited in the original report and any criteria used by the analyst to distinguish between taxa were noted in the spreadsheet. Total counts (i.e., of identified items per taxon) and ubiquity (i.e., percentage of sample units where specific taxa have been observed) were calculated at the level of the cultural unit (e.g., LBK, Michelsberg, Rössen, etc.) and thus represented aggregated data for each site phase. Each cultural unit was assigned a unique PhaseCode and the aggregated datasets were then entered in the database. All archaeobotanical data are identifiable either at the site level (based on SiteID), or at the level of cultural unit or phase (based on PhaseCode), and can
therefore be linked to other associated datasets, e.g., radiocarbon dates and faunal data. Each of the archaeobotanical tables is published alongside the associated spatial and temporal datasets and zooarchaeological data at http://discovery.ucl.ac.uk/1469811/.

Quality Control
We have adopted a fully inclusive approach to the data collection and have entered data from all archaeobotanical reports irrespective of whether or not they pre-dated the adoption of the current standard methods of sampling, recovery and recording. Similarly, we have not made any judgements about the accuracy of the identifications of taxa (i.e., based on the skill levels of the original analysts) and therefore have not excluded any from the dataset. We have included the full range of preservation types (e.g., charring, waterlogging, mineralisation, desiccation, impressions) and methods of recovery (e.g., flotation, dry-sieving, hand-picking). All records have been checked and standardised wherever possible, so that consistency across dataset is guaranteed, e.g., synonym species names, such as Alliaria petiolata and Alliaria officinalis have been recorded under a single taxonomic name to avoid duplication, in this instance Alliaria petiolata.

Constraints
The level of recording in the archaeobotanical reports varied, e.g., at some sites absolute counts were given for all taxa, whereas at others there were presence data only and no quantitative records (e.g., in instances where plant impressions are identified). As noted above, data from all reports have been included and this has permitted presence/absence analysis of as full a range of taxa as possible. Our records include taxa preserved under different conditions and the disparity in the range and type of remains represented by each imposes a potential constraint when using the dataset. In comparison to charring, for example, waterlogging results in less taphonomic bias and thus a far greater diversity of taxa is preserved which is more likely to comprise the full spectrum of species originally used; under all other preservation conditions the large seeded crops and wild edible fruits/nuts are resistant to decay but fragile taxa rarely survive, hence the dataset is biased in favour of the more robust plant parts [7]. Allowances should be made accordingly when making comparisons between sets of data from different sites. Similarly, account should be taken of the fact that because the dataset includes some sites where recovery has involved the use of large mesh sizes (for flotation or dry sieving) whereby any small taxa are lost, there is also likely to be a bias in favour of large taxa. For reference, there are notes on recovery methods and mesh sizes in the AbotSites table. Sampling strategies also differed greatly between sites and unfortunately it was rare for relevant details to be given in the original reports, but if available these are described in the AbotSites and AbotPhases tables.

(3) Dataset description

Object name
AbotSites – two files providing the data (EUROEVOL-13-07-2015-ABotSites.csv) and field type definitions (ABotSites_fields.csv) for all sites with associated archaeobotanical data, recovery methods and mesh sizes. The SiteID links to the CommonSites table described in the EUROEVOL Dataset 1: Sites, Phases and Radiocarbon Data.

AbotPhases – two files providing the data (EUROEVOL09-07-201516-34_ABotPhases.csv) and field type definitions (ABotPhases_fields.csv) for all PhaseCodes with associated archaeobotanical data, context type descriptions and numbers of samples.

AbotTaxaList – two files providing the data (EUROEVOL09-07-201516-34_ABotTaxaList.csv) and field type definitions (ABotTaxaList_fields.csv) for the full taxonomic description in relation to the unique Taxoncode of all taxa represented in the database

AbotSamples – two files providing the data (EUROEVOL09-07-201516-34_ABotSamples.csv) and field type definitions (ABotSamples_fields.csv) for each taxon assigned by PhaseCode together with quantification, plant parts identified and preservation status.

Data type
Primary and secondary.

Format names and versions .csv, SQL

Creation dates
Some records were created in 2001–2004 as part of the AHRB funded ‘The origin and spread of plant economies in the Near East and Europe’ project, however, the majority of records, and the current SQL database were created in 2010–2015.

Dataset Creators
The primary researcher responsible for the data collection (both in the previous project and the EUROEVOL project) was Sue Colledge. Meriel McClatchie (School of Archaeology, University College Dublin) collected all the Irish and British data for the EUROEVOL project.

Language
English.

License
The open license under which the data has been deposited (e.g. CC0).

Repository location
The full relational database is available as a SQL dump file and the individual tables (CommonSites, CommonPhases and C14Samples) are available as .csv files at http://discovery.ucl.ac.uk/1469811/.

Publication date

(4) Reuse potential
The dataset represents one of the largest collections of archaeobotanical data for the Neolithic of Europe (c.8.300 records for c.1500 different species, genera and families, representing over a million identified items) and as such it has great analytical potential for future researchers. The EUROEVOL data are particularly reusable because
the sample sizes are so large, permitting robust comparative analysis between sites and regions, and across time. Furthermore, all data are fully georeferenced, offering considerable mapping potential and most importantly, the data are linked to associated faunal and radiocarbon data from the same sites. Hence, there is considerable scope for further palaeoecological and palaeoeconomic analyses that incorporate both the plant and animal bone data. This dataset will prove most useful for archaeobotanists; it may however also be of benefit to geographers and palaeoecologists interested in past species distribution. The dataset will provide a useful training resource for student archaeobotanists interested in developing quantification techniques and statistical analyses of processed data.

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

Acknowledgements
Several researchers from across Europe have been extremely supportive of the project in allowing access to libraries and providing reports (e.g., including unpublished theses) as well as bibliographies with lists of useful references. I am particularly grateful to the following colleagues: Amy Bogaard, Otto Brinkkemper, Stefanie Jacomet, Philippe Marinval, Charlene Murphy, Welmoed Out, David Robinson, Aurélie Salavert, Hans-Peter Stika, Jean Denis-Vigne and Julian Wiethold.

References

How to cite this article: Colledge, S. 2016 The Cultural Evolution of Neolithic Europe. EUROEVOL Dataset 3: Archaeobotanical Data. Journal of Open Archaeology Data, 5: e1, DOI: http://dx.doi.org/10.5334/joad.42
Published: 27 January 2016
Copyright: © 2016 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.