

## **AWG, Oxford 22<sup>nd</sup> April 2023: Data and Standardisation**

Minutes, Ruth Pelling

The spring 2023 AWG meeting was hosted by Oxford University Archaeology Department. The meeting was timed to follow the second seminar of the Rewilding Project (see below) in which concepts of data publishing, sharing and metadata publishing were discussed in relation to the development of an Oasis + module.

The meeting was a hybrid event. A total of 26 people attended in-person event, while 9 attended on-line.

Books from the libraries of two much valued and missed late colleagues were available to buy with voluntary contributions paid to their chosen charities: donations for Lisa Lodwick's books went to <https://coppafeel.org/> and those of Lisa Moffett went to the AEA where they will provide funding for student travel bursaries. Many thanks to Kelly Reed, Gill Campbell and Don O'Meara for organising.

Thank you to Amy Bogaard for kindly hosting us, and for monitoring online activity and questions. And thank you to speakers, all those who helped with practicalities on the day and to those who contributed to the discussion.

### **Welcome by Amy Bogaard, Professor in Prehistory**

#### **ANNOUNCEMENTS**

**[OASIS & Research Framework Reporting Walkthroughs | Eventbrite](#)** : Next events Thursday 27<sup>th</sup>, East Midlands, Friday April 28<sup>th</sup> (South Yorkshire) and Wednesday May 3<sup>rd</sup> (North East). Each walkthrough is c.45 minutes and designed to demonstrate the links between OASIS and the Research Frameworks. Each session is the same, but with a regional example.

***OASIS V Specialist training***: A training event for specialist metadata entry into OASIS V will be made available to AWG members later in the year – date tbc

**[Free Online Training! | Open Phytoliths \(open-phytoliths.netlify.app\)](#)** : The International Committee of Open Phytolith Science (ICOPS) is providing a series in Open Research Skills relevant. Relevant to archaeobotanists as well as phytolith specialists.

28th April 2023 - GitHub collaborative workflows

26th May 2023 - Standard vocabularies and ontologies

30th June 2023 - FAIR data

Past meetings have been on Open access publishing, Using repositories and GitHub basics.

All workshops will be on a Friday at 14:00 to 16:00 GMT. Recording of previous meetings will can be found on the ICOPS You Tube Channel

<https://www.youtube.com/@icopsopenphytoliths1370/videos>

Slides are on zenodo:

- open access - <https://doi.org/10.5281/zenodo.7669467>
- using repositories - <https://doi.org/10.5281/zenodo.7717229>

**AEA Spring Conference – [News \(envarch.net\)](#)** : The spring conference will be an online event on 13<sup>th</sup> May. It will be followed by R Training, free to AEA members, or £20 to non-members on 14<sup>th</sup> and 15<sup>th</sup> May.

**Next AWG meeting:** Sheffield, on the subject of isotope analysis. Details tbc. General support for a weekday event.

**Open Reference Collection Days at Fort Cumberland:** Open days are fully back on following the pandemic. They are held on the second Wednesday of each month. Please email Ruth Pelling or Gill Campbell to book in.

## PRESENTATIONS

**Anwen Cooper and Tina Roushannafas, Oxford Archaeology: *Rewilding Later Prehistory Project***  
<https://rewilding.oxfordarchaeology.com/>

Anwen presented an introduction to the project and the seminar held the previous day, including the work with ADS to develop an Oasis + module. An OASIS + module already exist for geophysics, and modules in human remains and scientific dating are in production. Specialist input into development of modules for archaeobotany and zooarchaeology is vital to create a fit for purpose product. The modules will ensure usable and searchable metadata which will help resolve current difficulties of finding data. More automated data uploads would be possible with better mapping of data codes.

Specialist working groups are being created to shape the details of the proposed OASIS+ modules for collecting/signposting palaeoenvironmental (meta)data. Interested AWG members are invited to participate. The initial remit for this working group will be to design broad templates for the OASIS+ plant macrofossil and vertebrate animal remains modules, which colleagues will refine, test, and give feedback on (at, and following on from, the next 'Rewilding' workshop in December 2023). The project has budgeted fully for up to 26 specialists to contribute to this process – your input would be greatly appreciated; it is extremely important to the team that the specialist community are an integral part of this initiative from the outset.

For more information visit the [project webpage](#).

Tina discussed practicalities of collating data from old reports and highlighted the following points:

- Tabulated data is always helpful no matter how summary it is
- Shorthand in tables is not useful – additional observations are better made in a separate column
- Access to original Excel sheets speeds up data entry
- Phasing is often missing, incomplete, or changed at late stages of projects when limited time is available for specialists to revise their data/reports

- Images/illustrations are useful
- Data cleaning is a necessary part of data collection

Comments/resources:

Concern around data ownership for specialists working on commercial projects was mentioned – needs to be considered in terms of the specialist uploading site data.

The following paper is of relevance

Wright, H & Richards J D (2018) Reflections on Collaborative Archaeology and Large-Scale Online Research Infrastructures, *Journal of Field Archaeology*, 43:sup1, S60-S67, <https://doi.org/10.1080/00934690.2018.1511960>

Historic England have created the ADAPT toolkit to assist in the compiling of data for archive. <https://historicengland.org.uk/research/methods/archaeology/archaeological-archives/adapt-toolkit/>

Digital copies of data should be included in project archives as well as any paper copies.

**Elizabeth Stroud, Oxford University: *Seeing the fields through the weeds: introducing the WeedEco R package for comparing past and present arable farming systems using functional weed ecology***

Liz presented a summary of a forthcoming paper to be submitted to *Vegetation History and Archaeobotany* (open access). She included a theoretical background to the use of functional traits of arable weeds (physical attributes which reflect a species ecological potential in relation to environmental variables such as soil productivity and disturbance). The combination of traits, as expressed by combinations of species, are used to model arable regimes by comparing ancient floras with floras from present-day farming regimes which differ primarily in terms of soil productivity and/or disturbance. The functional traits for nearly 1000 arable weed species, and the present-day models developed from data collected across Europe and North Africa over a period of nearly 30 years have been made available.

Analysis of archaeological arable weed data relies on identification of seeds to species or genus where species variability is minimal (and possible species limited to few). Statistical analysis is conducted in RStudio using a package developed by the team. Members of the AWG (Ruth Pelling, Michael Wallace, Anne de Vareilles, Meriel McClutchie) had all trialled the method, following instructions given in the forthcoming paper, using archaeobotanical data from Saxon Stafford. All found the instructions sufficiently detailed and easy to follow.

Discussion points:

Identification level of wild (weed) taxa in archaeobotanical assemblages is important to enable full ecological analysis.

Analytical methods always need to be appropriate to the assemblage. The Functional Ecology analysis requires a minimum number of wild taxa identified to species level (or genus level if all possible species have very similar functional attributes), and sufficient numbers of samples. It is important to recognise the presence of dung, turf or other sources of wild taxa.

The functional attribute database behind the statistical model will be updated as more taxa are included, raising the importance of live databases.

A similar paper is being developed using Glynis Jones' crop processing data. Some members of the audience felt they might not have time to apply the ecology package, but were very interested in the crop processing application. Ideally both would be used in combination.

**Rachel Ballantyne, Oxford University: *Waterlogged plant remains from the EXPLO project***  
<https://exploproject.org/>

The EXPLO (Exploring the dynamics and causes of prehistoric land use change in the cradle of European farming) project is an international project exploring Neolithic-Bronze Age lakeshore settlements in the 'lake district' of the southern Balkans (Greece, Albania, North Macedonia). Rachel presented a background to the project and an introduction to the waterlogged plant remains so far being recovered, particularly from a pile dwelling site (Lin 3) on the edge of lake Ohrid, Albania which she is working on (the other major sites being Dispilio in Greece and Ploča in N. Macedonia). A substantial, centuries long, time period is represented by up to 1.8m depth of organic deposits. Rachel included sampling and sub-sampling methodology, taxa range, preservation, and quantification.

The methods follow those established by extensive experimental analyses for Alpine pile-dwelling settlements (e.g. Steiner et al. 2017 <https://doi.org/10.1016/j.revpalbo.2017.05.013>; Antolín et al. 2017 <https://doi.org/10.1016/j.jasrep.2017.02.008>). Samples are being processed by hand using wash over onto a stack of sieves. They are kept wet. A grid method is used to subsample resulting wash over to ensure it is fully mixed and randomised. Substantial charred material is present as well as waterlogged and preservation is beautiful due to still clear water and consistent water level. Rachel compared it to some of the Scottish crannog sites in terms of preservation conditions.

Fragmented items, such as acorn pericarp are quantified by count if they are larger than an arbitrary mm<sup>2</sup> area in size (separated by mesh size).

The archaeobotanical remains fall into the categories of crop plants and weeds, scrubby vegetation/hedgerow, and woodland, giving a much broader landscape or vegetation community representation than would be expected from vegetation in the immediate catchment.

#### Discussion

Quantification – currently all items and taxa are identified and quantified to MNI. Fragmented items are counted by fragment over an arbitrary size that varies according to the original plant structure (e.g. 25mm<sup>2</sup> for acorn and hazelnut pericarp, but 12mm<sup>2</sup> for Pistacia pericarp and legume pod fragments).

Contrasts made with Must Farm – which reflects a much shorter time period and more partial preservation due to the unstable water table (species-rich fen vegetation with many taxa, but few delicate anatomical parts). Must Farm deposits also reflected a long period of natural environment prior to and after the building of the pile dwelling.

Each lake for EXPLO is producing botanical remains of slightly different character suggesting varied subsistence regimes involving farmed and wild resources.

Determining provenance of WL material, for example, arable weeds mixed with naturally deposited flora from disturbed ground: In pilot work an initial crop processing analysis of ‘waterlogged midden’ samples from one of the EXPLO sites, Ploca, was made to help assess whether the potential weeds actually were consistent with the associated crop chaff. Normally, they were. Unusually for WL remains, perhaps, the assemblages are very much skewed towards arable harvesting.

Interesting to consider what weeds may be missing from charred assemblages.

### **Ruth Pelling, Historic England: *Update on ArboDat and taxa lists***

As noted during the Rewilding seminar, while ArboDat might not be used widely by commercial archaeobotanists, the database principles behind it are worth considering and it can be a useful source of information for identification groups.

ArboDat is managed by Felix Bittmann at The Lower Saxony Institute for Historical Coastal Research (NIhK - Das Niedersächsische Institut für historische Küstenforschung), Wilhelmshaven, Germany. It is free to use under license, and enquiries can be made to Ruth Pelling at HE. For further information see <https://nihk.de/forschung/aktuelle-projekte/arbodat-ape>

An updated version of ArboDat is in development – will make updates to repositories (especially Pangaea) more straight forward. HE has offered support with the development of the English Language version.

#### ***Pangaea***

For long term data storage ArboDat data can be uploaded to Pangaea, a German data repository, for free using a web app “APE – Arbodat Pangaea Export”. Particularly useful for single sites/projects <https://nihk-server.de/ape/>

<https://www.pangaea.de/?q=arbodat>

Data is freely downloadable for use in statistical analysis using R or statistics software, but not directly into an ArboDat database.

**GFZ ArboDat 1016 Datacentre:** Alternatively, Angela Kreuz and uploaded complete research ArboDat databases to GFZ. <https://dataservices.gfz-potsdam.de/portal/>

Data is searchable and downloadable in a format that can be entered directly into another copy of ArboDat. But, currently access to data is restricted by approval.

**OASIS V** includes an ArboDat field and data repository location, so if data is put into ADS, Pangaea or GFZ Data Services, it can be findable in searches.

An ArboDat (Taxa Transfer) Excel recording sheet is available for archaeobotanical data recording for anyone who would like one. Email [Ruth.Pelling@HistoricEngland.org.uk](mailto:Ruth.Pelling@HistoricEngland.org.uk)

#### ***Nomenclature:***

Mark McKerracher has kindly shared a list of botanical names following Stace 2010 which he compiled as part of the FeedSax Project. Not a complete list but does include all taxa encountered during the project.

Mark also created an Excel macro which converts synonyms to the Stace 2010 format.

Volunteers are sought to add to the list and to build the synonym list to enable much quicker conversion of synonyms (could be used to convert Flora Europea nomenclature outputs from ArboDat to Stace 2010 for example). Email Ruth if you would like volunteer.

### **Michael Wallace, Hedland Archaeology: *The A14 Project in Numbers***

Michael has been coordinating the synthesis for the large A14 infrastructure project which inevitably brings challenges when compiling data created by different specialists. This useful presentation highlights some of the inconsistencies that we can think about in terms of how we record archaeobotanical data and the creation of large project wide databases. Some differences were the result of different sites/landscape units and preservation of material, whereas others might be due to different experiences or expectations of individual specialists.

The synthesis is at different levels: site, project, regional and national. The project numbers presented were:

397,353 plant items

12,885 individual taxonomic entries

197,266 plant parts identified by estimate, eg scale of abundance, eg for fragmented items, or very large assemblages

15.8% of cereal grain entries which mention fragments in notes

214 taxonomically implausible quantifications, eg *Anthemis cotula* recorded as “chaff” – wrong category selected in error (to think about in future database design)

nine archaeobotanists worked on the project, each with different landscape blocks

7% to 25% - proportions of cereal grains given a cf. (before genus or species) by specialist

1% to 36% - proportions of cereal chaff given a cf.

69% of cereal grains were identified with multiple genus eg *Cerealia* indet or *Hordeum/Triticum*

87% of cereal grains were identified to single species or an either/or species

32.1% of charred weed seeds were identified to species level (including either/or), of which

25% were identified to single species

Data entry errors are inevitable and to be expected – amendments take time

#### Discussion

Identification level and use of cf. were discussed. Partly an issue of experience (from eager early career specialists wanting to identify everything to increasing caution amongst more experienced specialists). Can experiment in groups to find common agreement, but generally this is difficult.

Positive identification should be to the lowest level that is definite (better to be cautious but correct).

But, time/costs can restrict full identification/counting – if not counting in full, always state in methodology

Use of cf, – agreement that this is an important inclusion but should be used with caution to prevent later researches upgrading an identification inappropriately. Most contributors use cf. when they are pretty sure of an identification but can't be 100% because they may be lacking reference material to confirm, or spot some identification characteristics but not all. “The use of cf. (*confer -compare*) means uncertainty because of:

- a) Poor or fragmentary preservation
- b) Inadequate reference material
- c) Ill-defined morphology (Dickson 1970, 250)

cf. is also sometimes used on ecological grounds (only species known to grow in temperate climate).

Dickson, C 1970 The study of plant macrofossils in British Quaternary deposits, *in* D Walker and R G West *Studies in the Vegetational History of the British Isles*. Cambridge. Cambridge University Press 233-254

## DISCUSSION

The afternoon was spent discussing some of the issues raised. Useful resources:

### Integrated Archaeobotanical Research Project – on line tutorial

<https://sites.google.com/sheffield.ac.uk/archaeobotany/home>

There is a list of resources on the AWG webpage – it does need tidying up and updating

<https://historicengland.org.uk/research/current/heritage-science/archaeobotanical-work-group/>

### Quantification. Well established criteria following Hillman et al 1996, Charles et al 2009).

#### What to count?

- a feature that survives well archaeologically (to avoid under-representation),
- easily identifiable (to ensure reliable identification)
- unique (to avoid double counting).

#### For cereals:

the embryo end of the grain (not the embryo itself)\*

the base of the glume (for glume wheats, the most durable chaff part), where one spikelet fork = two glume bases

the rachis node (for free threshing cereals, more durable than glumes bases)

the culm node (the most durable part of the straw).

Each item is a whole = 1

Charles, M., Crowther, A., Ertug, F., Herbig, C., Jones, G., Kutterer, J., Longford, C., Madella, M., Maier, U., Out, W., Pessin, H., Zurro, D., (2009) *Archaeobotanical Online*

*Tutorial* <http://archaeobotany.dept.shef.ac.uk/>

<https://sites.google.com/sheffield.ac.uk/archaeobotany/seeds/quantification>

Hillman G, Mason, S., de Moulins, D. and Nesbitt, M. Identification of archaeological remains of wheat: the 1992 London workshop, *Circaea* 12 (2) (p.195-210) - can be downloaded from the AEA web pages <https://envarch.net/circaea> or found on Mark Nesbitt's Research net page.

Jones, G. 1991. Numerical analysis in archaeobotany, pp. 63-80 in W. van Zeist, K. Wasylikowa and K.-E. Behre (eds.) *Progress in Old World Palaeoethnobotany*. Rotterdam: A.A. Balkema.

#### **Embryo end absent or fragmented grain**

The aim is to avoid double counting. Where most of a grain is present and it is unlikely the embryo end exists separately, the grain can be counted as one. For fragmented grain a range of approaches

were suggested. It was noted that fragmentation may be a product of pre-charring processes (eg the production of ground cereal products, malting) or as a result of post-charring preservation/taphonomy (eg very dry conditions, flotation process). Methods suggested:

Estimates of whole counts, distal ends, ventral groove (eg for waterlogged grains), number of fragments, measured volume, and estimated counts using volume. It was agreed that total weight was not useful due to variation in weight of charred material.

Method adopted would depend on the character of the assemblage in question. Always record the methodology.

Other useful references:

Antolín, F. and Buxó, R., 2011. Proposal for the systematic description and taphonomic study of carbonized cereal grain assemblages: a case study of an early Neolithic funerary context in the cave of Can Sadurní (Begues, Barcelona province, Spain). *Vegetation History and Archaeobotany*, 20, pp.53-66. <https://doi.org/10.1007/s00334-010-0255-1>

Valamoti, S.M. 2011. Ground cereal food preparations from Greece: the prehistory and modern survival of traditional Mediterranean 'fast foods'. *Archaeological and Anthropological Sciences* 3 (1): 19-39) <https://doi.org/10.1007/s12520-011-0058-z>

#### **Fragmented non-cereal remains**

**Fragmented shell/pericarp (eg *Corylus avellana*, *Quercus*):** most people counted fragments by size, easily done by separating flots into different sieve sizes. Identification may be dubious for small fragments. Some only count over a certain size (eg 2mm sieve).

Could also weigh or measure volume as an additional measure.

Given quantification is not used in statistical analysis normally, the most important thing is to give an indication of the size of the assemblage – occasional fragments discarded as food debris vs deliberate deposits or large processing features.

**For fruit stones/seeds (eg date, olive, grape):** it is usually possible to count an identifiable feature (date embryo, grape seed beak or chalaza) or estimate whole stones/seeds. Always state counting methods

See also (includes tables with scoring and counting methodology):

Antolín, F. and Jacomet, S., 2015. Wild fruit use among early farmers in the Neolithic (5400–2300 cal bc) in the north-east of the Iberian Peninsula: an intensive practice? *Vegetation History and Archaeobotany*, 24, pp.19-33. <https://doi.org/10.1007/s00334-014-0483-x>

#### **How to manage large, rich samples with little time/funding?**

Identification of wild taxa is vital if any ecological analysis is to be possible so ideally should not be compromised in favour of cereal identification.

Most preferable method is to use a riffle box and split samples – this might mean splitting different sieve sizes differently (eg ¼ >2mm, 1/6 <2mm) or counting cereal remains and weed seeds



differently according to the individual assemblage. Always be clear in methodology and make it clear if the tables contain actual counts or estimated multiplied figures.

An alternative method is to count grain and chaff but take identification less far (eg recorded hulled grains and chaff as *Triticum spelta/dicoccum* or *Triticum* sp. hulled), leaving time to fully identify and quantify weed seeds. Again, depends on the assemblage and project aims. Methods should be stated.

### Grass and Legume Categories/Identification

The following were not discussed in detail and can be looked at more closely at a future AWG including a microscope session. Feedback to Ruth is encouraged. A compiled document will be circulated for approval eventually.

#### Grass identification/categories

The following identification aid is really useful:

Nesbitt, M 2000 *Identification Guide for Near Eastern Grass Seeds* Routledge  
<https://doi.org/10.4324/97813154> ebook available via Taylor and Francis

Where grass seeds are not identifiable to species the following identification categories are suggested as most useful by Felix Bittmann and the European ArboDat group (see page 50 of the ArboDat manual).

The division of large and small grasses for crop processing purposes is based on the relative size of cereal grains and therefore needs to be considered on an assemblage by assemblage basis. Use of modern mesh sieves may be useful.

Identification	ArboDat code	Comments
<b>Small grasses</b>		
<i>Festuca/Lolium</i> type	FESLO	Excluding <i>Lolium temulentum</i> or <i>L. multiflorum</i>
Phleum pratense/Poa annua type	PHPPA	
<i>Poa</i> sp. <b>not annua</b>	PONAN	
Poaceae, small	POACE (=Poaceae)	Small grasses not distinguishable by species or the above identification groups
<b>Large grasses</b>		
<i>Avena/Bromus</i> sp,	AVSFB	
<i>Bromus</i> sp. (or <i>Bromus</i> typ)	BROM (BROT)	RP would tend to use this
Poaceae, large	POBFT = Poaceae Bromus/Festuca- Typ	Large grasses not identifiable as <i>Avena</i> , <i>Bromus</i> , <i>Lolium</i> , <i>Festuca</i> . Not an ideal PCode category, but unlikely to be used for anything else.
<i>Avena</i> sp.	AVEN	If no spikelet base attached. Could distinguish between large and small if there is a clear distinction but not a reliable method to distinguish wild from cultivated

## Fabaceae

The following is suggested by Felix Bittmann based on ArboDat terms.

<b>Small <i>Trifolium</i> type legumes</b>		
General group if not possible to distinguish further	LOTRI = <i>Lotus/Trifolium</i>	(includes Fabaceae- <i>Trifolium</i> type). Ruth would tend to use this
Other possible types/groups	FABTT = Fabaceae <i>Trifolium</i> Typ	
	MEDTR = <i>Medicago/Trifolium</i>	
	MEDME = <i>Medicago/Melilotus</i>	
	MEMET = <i>Melilotus/Medicago/Trifolium</i>	Or this
<b>Rounded legumes</b>		
<i>Vicia, Lathyrus, Pisum, Lens</i>	VICIE = <i>Vicieae</i>	General term, no further differentiation preferred/possible
Large <i>Vicia/Lathyrus/Pisum</i>	LAPIV = <i>Lathyrus/Pisum/Vicia</i>	
Small seeded and flattened so could include <i>Lens</i>	LENVI = <i>Lens/Vicia</i> klein	(klein = small)
Small rounded vetches, could include <i>Lathyrus</i>	FABAC=Fabaceae ( <i>Vicia</i> -Type, <i>Lens/Vicia</i> small)	HE used VICIE in Biodiversity database
Large round pea sized legumes lacking hila	PISVI = <i>Pisum/Vicia</i>	
Cultivated pea/vetch but lacking hila - can't be any other species/genus	PIVSA = <i>Pisum sativum/Vicia sativa</i>	
<i>Vicia</i> specifically so if with hila (not including <i>Lathyrus</i> )	VICGF = <i>Vicia</i> groß-früchtig	Large seeded
<i>Vicia</i> specifically so if with hila (not including <i>Lathyrus</i> )	VICKF = <i>Vicia</i> klein-früchtig	Small seeded
Large cultivated Fabaceae but not distinguishable (eg in frags)	FABAK = Fabaceae kult (kult = cultivated)	= LAPIV used in the Biodiversity database