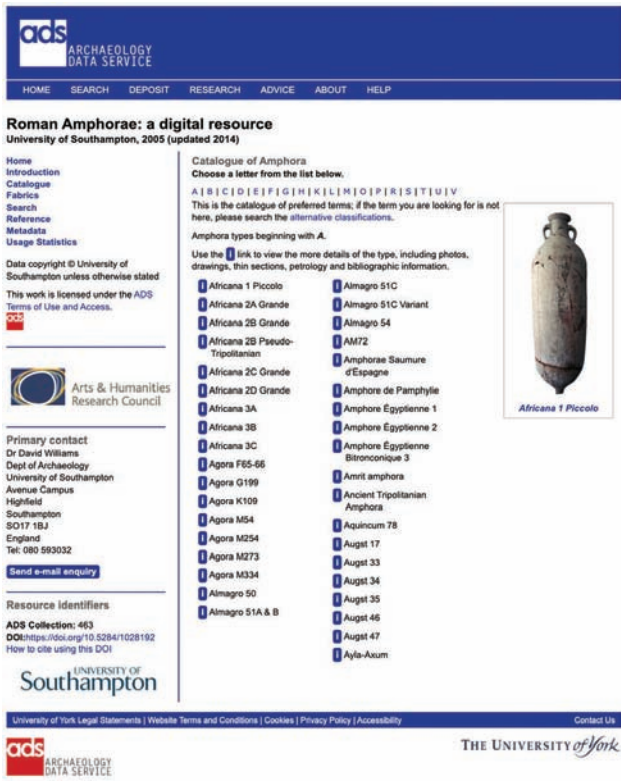
DIGITAL ARCHIVES, OPEN DATA AND UNEXPECTED LEGACIES

Holly Wright, Archaeology Data Service

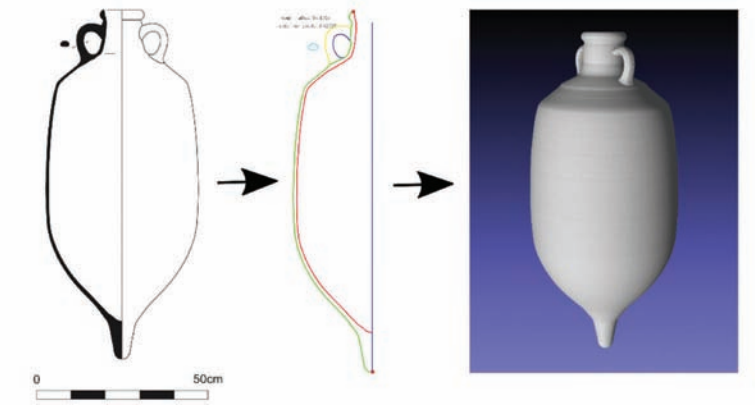
At the Archaeology Data Service, we often discuss how and why the data we hold might be used, and how we might optimise it for the many communities we work to serve. This has become more urgent as we try to align our archiving processes with the FAIR Principles: the best-practice guidance promoting that digital data should be made Findable, Accessible, Interoperable and Re-usable (Wilkinson et al 2016). While we try to understand and predict how our data might be re-used, there are times when the advent of new technologies creates uses for our open data that neither we at the Archaeology Data Service nor the data creator could have imagined.



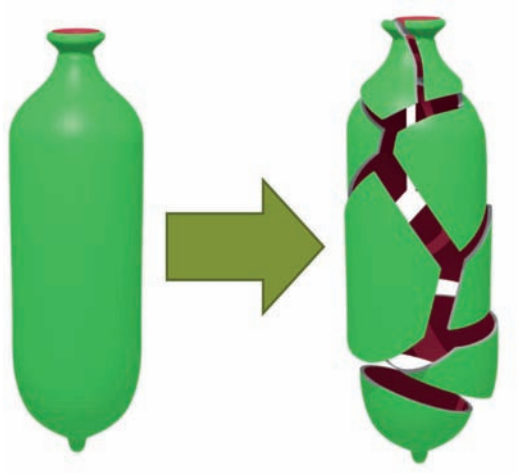
The popular archive: Roman Amphorae: a digital resource*, first deposited in 2005. Credit: Archaeology Data Service*

One of our most popular archives is *Roman amphorae: a digital resource* (University of Southampton 2014), with around 25,000 pageviews every month, year after year. One of the main reasons for its popularity is that it is a comparative collection, rather than the data from a particular archaeological project. The archive includes a searchable interface with the significant characteristics for a wide range of Roman amphorae types and sub-types, including details about everything from spatial and temporal distribution to typical contents and distinctive features, but it also includes photographs of typical examples, petrology and most importantly for this discussion, profile drawings for each amphorae type and sub-type.

The popularity of this archive already illustrates the importance of open, digital comparative collections for archaeology to speed the identification of archaeological artefacts, but the fact that it is openly licensed for re-use means it is incorporated into other research in ways that paper-based and proprietary comparative data cannot. For example, it is also currently in use by the *Production and distribution of food during the Roman Empire: economic and political dynamics* project ('EPNet' n.d.). As the Archaeology Data Service also actively participates in research, we agreed to be partners in a recent project funded by the European Commission called ArchAIDE (Anichini et al 2020). ArchAIDE set out to design an app



The conversion process carried out by the CNR-ISTI Visual Computing Lab. Standardised profile drawings of archaeological pottery were converted into vector files in SVG format, from which 3D scale models were derived. Credit: Archaeology Data Service



To address the paucity of training data available, 3D models were broken into 'virtual sherds' and used to train the shape-based algorithm. Credit: Archaeology Data Service

3D printed Roman amphorae from the ArchAIDE archive. Credit: Holly Wright

Screenshot detail of the ArchAIDE archive, showing the interactive model for Roman amphorae type African 2D Grande in 3DHOP, along with the 2D and 3D models available for download, and the links back to the full description within Roman amphorae: a digital resource. *Credit: Archaeology Data Service*

'we agreed to be partners in a recent project called ArchAIDE ... (which) set out to design an app to speed up the time-consuming process of pottery identification.'

to speed up the time-consuming process of pottery identification, and set out to use Artificial Intelligence in the form of automated image recognition to make it easier to match sherds with their correct type.

As proof-of-concept, ArchAIDE decided to use Terra Sigillata, Majolica of Montelupo and Roman amphorae to test whether image recognition algorithms could be used to identify sherds based on their shape or decoration. It was important not only to create training data so the algorithms could identify the correct types, but also to provide comparative data within the app, so that archaeologists could confirm whether or not they agreed with the identification, before saving the identification in a digital assemblage. For Terra Sigillata and Majolica of Montelupo, this meant expensive and time-consuming photo campaigns were undertaken within archives in Italy, Spain and Germany, but for Roman amphorae, the *Roman amphorae: a digital resource* archive could simply be re-used within the app.

As the project progressed, however, the real power of *Roman amphorae: a digital resource* became apparent. First, the ArchAIDE partners at the CNR-ISTI Visual Computing Lab in Italy were able to take all the digitised profile drawings and use them to create 3D models of every amphorae type, using an automated workflow. The visual computing lab is also responsible for creating an interactive viewer called 3DHOP, which allows user interaction with the models, including movement, measurement, sectioning in three dimensions, and a variety of other features. It was thought that this could be incorporated into the comparative data for the app, but it turned out to be even more critical to the success of the project.

The training of algorithms for automated image recognition typically relies on thousands or even millions of examples in order to learn what to recognise. This was simply impossible for archaeological pottery, and ArchAIDE partners were fortunate if they were able to find 10 examples for each type. The accuracy of the decoration-based image recognition was quite good, but the ArchAIDE partners at the Deep Learning Lab at Tel Aviv University who were training the algorithm struggled with the accuracy of the shape-based image recognition with so few examples to properly train the data, and the outlook for this part of the project was grim.

'we hope it represents an unexpected legacy that can be an exemplar for the future.'

Partners then hit upon the idea of creating training data by taking the 3D models of the Roman amphorae, breaking them into 'virtual sherds' and using them to train the algorithm... and... it worked! The accuracy was still not as high as for decoration, but was good enough to show the proof-of-concept could be successful, representing a way to create training data for many applications across the Humanities where there is a scarcity of well-identified types.

As the project came to a close, one of the tasks for the Archaeology Data Service was to archive the outputs of the ArchAIDE project (ArchAIDE consortium 2019). This included the 2D and 3D models created from every profile drawing from *Roman amphorae: a digital resource*, and it became obvious that this would be a useful addition to an already popular archive. While the two archives had to remain distinct, and all the 2D and 3D models accessible directly from the ArchAIDE archive, after discussion with the original depositors it was agreed the two archives should be linked. Now users of *Roman amphorae: a digital resource* can see the ArchAIDE logo and a link to the 3D model(s) for that amphorae type. In turn, each of the pages with the 3D models has links back to the full descriptive record within *Roman amphorae: a digital resource*. Information can be accessed directly through one archive to the other (and vice versa). Within the ArchAIDE archive, users can interact with the models in 3DHOP and they are easily downloaded for re-use, including 3D printing.

This represents a fascinating example of data re-use that could never have been imagined when *Roman amphorae: a digital resource* was first deposited with the Archaeology Data Service in 2005. It has resulted in the enrichment of the original archive, making it an even more useful resource going forward. None of this would have been possible, however, if the data had not been made FAIR and open, and we hope it represents an unexpected legacy that can be an exemplar for the future.

Anichini, F, Banterle, F, Buxeda i Garrigós, J, Callieri, M, Dershowitz, N, Dubbini, N, Lucendo Diaz, D et al 2020 Developing the ArchAIDE application: a digital workflow for identifying, organising and sharing archaeological pottery using automated image recognition. *Internet Archaeology*, 52. https://doi.org/10.11141/ia.52.7.

ArchAIDE consortium, 2019 ARCHAIDE portal for publications and outputs [data-set]. *York: Archaeology Data Service [distributor]*. https://doi.org/10.5284/1050896.

'EPNet.' n.d. Accessed 7 April 2020. http://www.roman-ep.net/wb/home/.

University of Southampton, 2014 Roman amphorae: a digital resource [data-set]. *York: Archaeology Data Service [distributor]*. https://doi.org/10.5284/1028192.

Wilkinson, M D, Dumontier, M, Jsbrand Jan Aalbersberg, I, Appleton, G, Axton, M, Baak, A, Blomberg, N et al, 2016 The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific Data* 3 (March), 160018.

Holly Wright

Holly is International Projects Manager for the Archaeology Data Service at the University of York. Her work has included ADS participation in the SEADDA, ArchAIDE, NEARCH, E-RIHS, ARIADNE, ARIADNEplus, LoCloud, CARARE and SENESCHAL projects. Her research focuses on archaeological field drawing and the Semantic Web, including specialisation in visual documentation and communication, archaeological data management, the use of Web standards and the Semantic Web in archaeology.