
Article:
Pagé-Perron, Émilie and Richards, Julian D (2026) Advanced digital workflows for archaeology. The Archaeologist. pp. 4-5. ISSN: 2754-0448

Advanced Digital Workflows for Archaeology

Émilie Page-Perron,
Research Associate, ADS

Julian Richards MCIfA (73),
Director, ADS and HSDS

The Archaeology Data Service (ADS) is working with European partners to investigate and develop best practice for digital workflows, from field recording to archiving, for multiple data types, including context sheets, fieldwork reports, images and IIIF, sound data, 3D data, BIM, and geospatial data. The research is funded by the European Commission as part of ATRIUM (Advancing fronTier Research In the arts and hUManities) (https://atrium-research.eu/). The project is coordinated by DARIAH (the digital research infrastructure for the Arts and Humanities) but also includes CLARIN (covering language and linguistics), and ARIADNE (the research infrastructure for Archaeology), as well as individual archaeological partners, including the Swedish National Data Service, the Cyprus Institute, and ARUP and ARUB, the archaeological research centres in the Czech Republic. The archaeological bodies, including ADS, are providing test datasets, but ATRIUM also relies on computer scientists to enhance metadata and investigate novel ways of visualising the datasets. The outputs will be demonstrated using the infrastructures of the ARIADNE portal (https://portal.ariadne-infrastructure.eu/) and the ADS data catalogue (https://archaeologydataservice.ac.uk/data-catalogue/).

The workflows and demonstrators are split into five strands, according to data type. We start with textual data, and go back to the beginning with hand-written proforma recording sheets. At ADS we hold lots of scanned context sheets but generally have very little metadata about their contents. Extracting information is a challenging task, involving handwriting recognition software, and segmentation to recognise the different parts of the form. Once the text has been transcribed, we can identify keywords based on controlled vocabularies, covering the subject terms, period names, and placenames. We are also working with colleagues at the Universities of Sheffield and South Wales to refine AI techniques to extract rich metadata from unpublished fieldwork reports, as well as from the large digitised journal back runs we hold in the ADS Library. The latter is more challenging as the articles are more heterogeneous than the OASIS fieldwork reports, particularly as we go back into the nineteenth century. At the moment, we often only have the authors' names and title for these articles, so our aim is to add enhanced subject metadata to the Library catalogue (https://archaeologydataservice.ac.uk/library), making it far easier to identify relevant articles.

The second strand is focused on image data. Our Czech colleagues are applying visual recognition software to images and metadata from the UK Portable Antiquities Scheme to train their software to recognise artefacts from photos. The ADS is also experimenting with IIIF, the international image interoperability framework. Using the IIIF Mirador viewer, we can compare images drawn from different repositories. For example, researchers will be able to place Bronze Age rock art from Sweden side-by-side with rock art images from Northern England archived by ADS. In another case study, we are rendering images from the Corpus

of Anglo-Saxon stone sculpture in IIIF, so that the motifs can be compared with those from manuscript art held by European galleries.

In the third strand, we move from 2D to 3D. In these workflows, the team is looking at 3D models and Building Information Modelling (BIM) datasets and enhancing the ways such data can be visualised in ADS and ARIADNE using online viewers embedded in web browsers.

The collaboration with CLARIN means that the fourth strand focuses on sound data. We are investigating the use of oral recording on site, with archaeologists speaking into microphones and reading context sheets, including in realistic, noisy field environments. Transcription software produces text from sound, and the texts can then be fed back into the workflow developed in Strand 1, allowing the extraction of keywords for indexing.

The fifth and final strand is looking at geospatial data where we are examining how place can be used to combine archaeological and historical information. At ADS, we will be taking a case study of a dataset from one of the cemeteries excavated in advance of HS2, the St.James Burial Grounds at Euston, London.

ATRIUM is now halfway through a four-year project. Work on many of the workflows is well advanced, and in the second half of the project, we are now focusing on the demonstrators. We aim that by year four many of these enhancements will have been realised for users of the ADS and ARIADNE catalogues.

These reusable workflows will not only improve the findability and interoperability of archaeological data across ADS and the ARIADNE portal but also support broader public engagement by making rich, well-documented digital resources easier to explore and reuse.