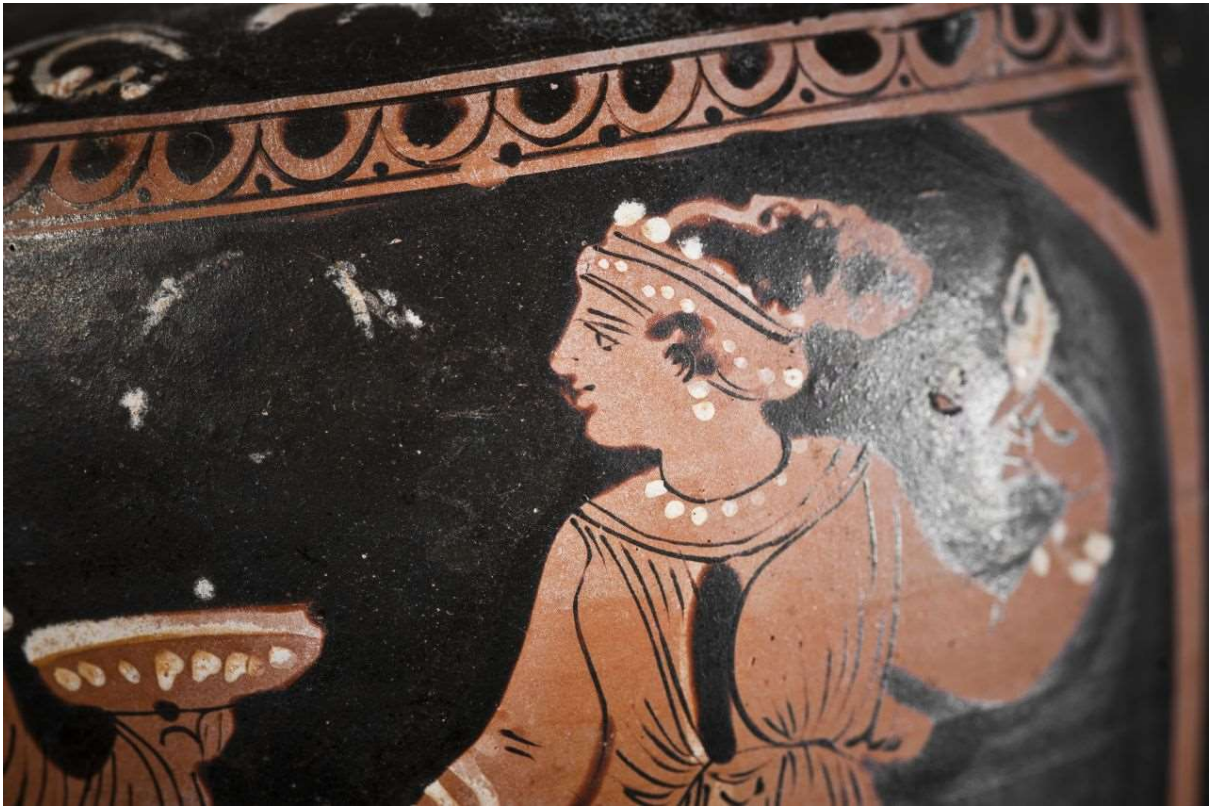


# Ancient Mediterranean Archaeology Project



Host: A.D. Trendall Research Centre for Ancient Mediterranean Studies

The Trendall Research Centre has the following objectives:

- To promote research in the general area of Ancient Mediterranean studies, particularly in the archaeology of South Italy and Sicily during the Classical period.
- To disseminate within the general community in Australia the results of the latest research in Greek and Roman art and archaeology through the sponsorship of conferences, lectures and seminars.
- To make available, at the Director's discretion, the resources of the Centre to all scholars and graduate students, whether from Australia or overseas, who wish to use the library and archive.
- To maintain and extend both the Library and the Archive (as unique research resources in Australia) through the acquisition of books and periodicals relating to Greek and Roman culture, and images of South Italian red-figure vases.

#### Background:

The Trendall Centre at La Trobe University houses a unique archive of some 40,000 images of ancient South Italian vases. These vases were made mainly in the 4th century BCE, and the imagery on them provides extraordinary detail regarding the lifestyles of the ancient Greeks who settled in South Italy, as well as the technological processes which enabled pottery production and the economic and trade environment which ensured its distribution. In collaboration with staff and students from the Dept of Computer Science and Information Technology at La Trobe, and the Beazley Archive at Oxford University, the Trendall Centre is working to produce an online database of the vase images matched with text information including details such as subject matter, creator and provenance. The matching text information has derived through OCR scanning of the publications of the archive's creator, Dale Trendall. The project involves work being done to automate the matching of images and text from different and complex sources in order to make the data accessible to 21st century users worldwide.

#### Project Description:

This project is working toward the digitisation of records made in paper-based form by the renowned researcher AD Trendall. The records have associated images of the actual artefacts which should also be logically associated.

The records are in the form of a scanned PDF file number upwards of 700 pages and the image archive is 4.3GB. We have an implementation of the extraction for you to utilised, however we are seeking a more generalised approach. Here you will design and experiment with cloud AI approaches and/or creating text extraction rule editors. The aim of the project is to perform text extraction from this and other PDF scanned archives. The extracted text will be mapped to a record format and stored in a database, along with the relevant image of the artifact (where available). In addition, a web interface will be required for researchers to search the archive and retrieve relevant information. One constraint on the search capability and perhaps that of the database, is the compatibility with Oxford University's search and data recording format.

#### Technical Skills:

- Python experience is a bonus for text extraction (easily learnt though)
- Data modelling and database operations
- Web Development