

# TETR ARCHS

**Transforming Data Reuse in Archaeology**

## **D5.1 Legacy Data Incorporation**

**15 May 2023**

### **Authors**

Benjamin Štular, ZRC SAZU

James Taylor, University of York

Edisa Lozić, ZRC SAZU

Nicolò Dell'Unto, Lund University

Anna Simandiraki-Grimshaw, MOLA

## Document History

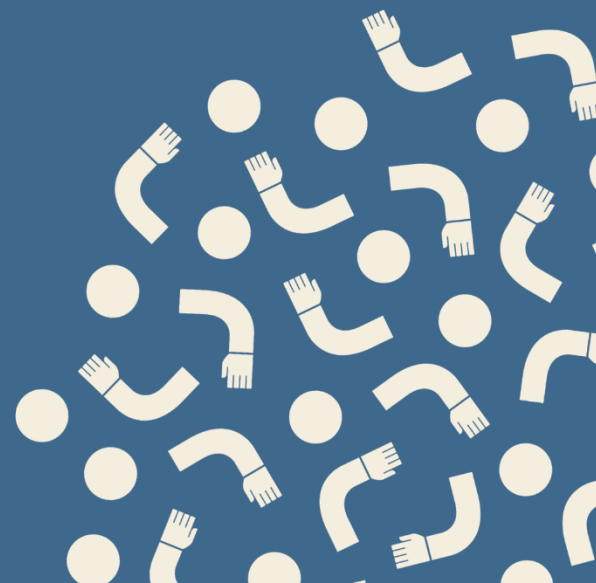
Date	Version	Author	Comments
15/5/23	0.8	Benjamin Štular James Taylor Edisa Lozić Nicolò Dell'Unto, Anna Simandiraki- Grimshaw	Full draft
7/12/23	0.9	Holly Wright	Quality Control Check
12/12/23	1.0	Nicolò Dell'Unto	Final Check

# Table of Contents

Document History .....	2
Table of Contents .....	3
Summary .....	4
Introduction .....	5
Current case studies and experiments .....	6
Reuse of archaeological field archives.....	6
Reuse of Existing Archaeological Site Photographs and Illustrations.....	9
Iterative Capture of New Data.....	13
Case study and experiment in preparation .....	15
Conclusion .....	18

## Summary

This Deliverable 5.1, titled "*Legacy Data Incorporation*," focuses on the publication of legacy data scenarios and associated project planning documentation for application in D5.2. The project aims to incorporate legacy data into archaeological fieldwork and investigations using digital documentation and advanced recording technologies. It provides an overview of the experiments planned for the first year of the project, made up of three experiments, two focusing on archaeological fieldwork in Sweden and Greece and one at MOLA in the United Kingdom. The experiments aim to explore the reuse of digital documentation and archaeological records, as well as the incorporation of legacy data in field investigations and museum presentations. This deliverable also previews several experiments being developed to explore the incorporation and utilisation of legacy data in archaeological fieldwork. In particular, the application of LiDAR (Light Detection and Ranging) techniques in an Iron Age settlement excavation project located in western Slovenia. Overall, these experiments will contribute to understanding how legacy data can be effectively incorporated and reused within archaeological fieldwork, documentation, and engagement with diverse user communities within the TETRARCHS project.



## Introduction

This deliverable reports the activities associated with *T5.1: Incorporation of Legacy Data* within WP5: *Data Experimentation*. The purpose of T5.1 was to “identify sets of legacy data associated with appropriate landscapes, sites, and artefacts available to Project Partners across Europe”. Task 5.1 co-designed legacy data experiments for trialling in Task 5.2 *Experiment Design Implementation*. This task was led by TETRARCHS partner Lund University with support from the University of York and ZRC SAZU.

This deliverable focuses on legacy data scenarios, and associated project planning documentation to incorporate legacy data into archaeological fieldwork and investigations using digital documentation and advanced recording technologies. The deliverable provides an overview of the experiments planned for the first year of the project. It mentions three separate experiments, two focusing on archaeological fieldwork in Sweden and Greece and one at MOLA in the United Kingdom. The experiments aim to explore the reuse of digital documentation and archaeological records, as well as the incorporation of legacy data into field investigations and museum presentations.

The first experiment, conducted in collaboration with Blekinge Museum in Sweden, focuses on the reuse of digital documentation and archived archaeological records from previous excavation campaigns. The experiment aims to investigate how field archaeologists reuse digital documentation in the field and how the museum then reuses the archived data to engage a wider audience. It also explores the creation of stories with archaeological information and the use of keywords and metadata to describe and annotate the data.

The second experiment focuses on the reuse of existing archaeological site photographs and illustrations. It involves enhancing the existing metadata of these records to make them reusable for different stakeholders. The experiment will select sites from MOLA's archives and collaborate with user audiences in workshops to enrich the metadata with relevant terms and tags. The goal is to make the records more accessible for various purposes and to test their reusability.

The third experiment centres around the Toumba Serron Research Project in Greece. It involves the deployment of image-based modelling techniques and a digital workflow at an excavation site. The experiment aims to assess if the current graphic representation and metadata used for 3D models are sufficient to meet research goals and user needs. It also involves stakeholder evaluation and mapping to facilitate input from different communities and foster the co-creation of archaeological knowledge.

In preparation for the second year of the project, several experiments are being developed to explore the incorporation and utilisation of legacy data in archaeological fieldwork. One such experiment focuses on the application of LiDAR (Light Detection and Ranging) techniques in an Iron Age settlement excavation project located in western Slovenia. The experiment involves the reprocessing of airborne LiDAR data collected nationally in 2015 specifically for archaeology, along with the integration of aerial photographs, historical maps, geological and soil maps, and spatial records from the Arkas 2.0 sites and monuments database (<http://arkas.zrc-sazu.si/>).

Overall, these experiments will contribute to the understanding of how legacy data can be effectively incorporated and reused in archaeological fieldwork, documentation, and engagement with diverse user communities.

## Current case studies and experiments

This section will provide a detailed overview of the fieldwork experiments planned during the project's first year, including a description of the user groups, people involved, and tools employed.

### Reuse of archaeological field archives

#### Site/scenarios description

The first experiment will be conducted in collaboration with the Blekinge Museum, Sweden (TETRARCHs Cooperating Partner) and will take place between May and October 2023. It will (1) explore how field archaeologists reuse digital documentation in the field to support their practice and (2) investigate how the digital archaeological records archived during the field investigation are reused by the museum to address a wider audience.

#### Legacy data scenarios

The legacy data included in this experiment are: (1) 3D models created using Image Based Modelling and laser scanning, representing 3D graphical documentation of trenches, contexts, and artefacts, and (2) shape files representing context and artefact positions created using total station or RTK GPS.

#### Overview of the experiment

The excavation at Västra Vång will be carried out using 3D recording technology. A team of archaeologists with different skills will use a combination of advanced methods to record the information gathered during the survey. The data will be input into the Archaeological Interactive Reports (AIR) system (<https://omeka.ht.lu.se/s/reports/page/home>), and by the end of the field season, the excavation team will input all the information retrieved (including the 3D models). The AIR system allows data input during fieldwork to be linked directly to the digital report (and can accommodate different versions). In addition, using AIR during the experiment will allow us to track how different users reuse and combine the same dataset to reach different audiences. This experiment focuses on creating a narrative with a given set of archaeological information by expressing/annotating that information using keywords/metadata. This type of work is useful for collecting potential metadata sets and gap analysis (how storytellers annotate digital resources concerning the stories/interpretations they tell and how this relates to existing standards).

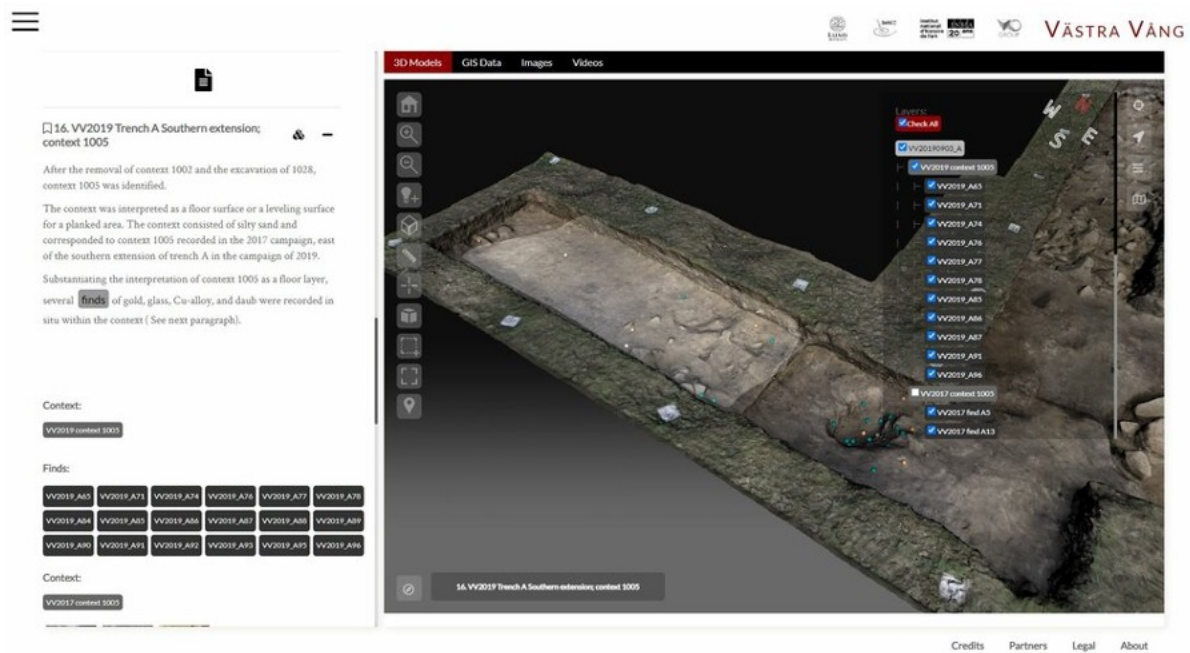


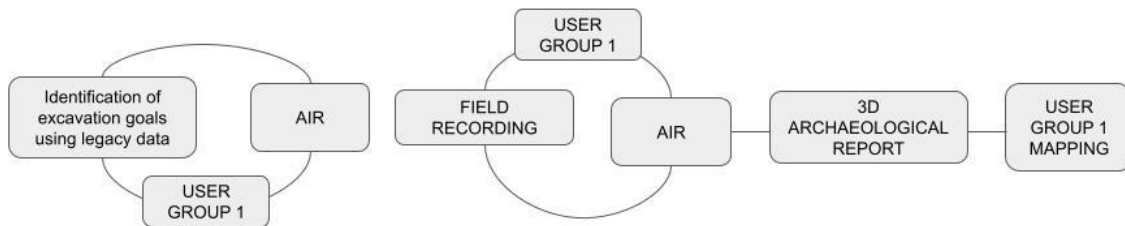
Figure 1: The interface of the 3D online web archive, known as Archaeological Interactive Reports (AIR), which has been used in previous years to store, organise and create narratives from the archaeological records collected during the field campaigns at Västra Vång. The same system will be used during upcoming field experiments.

To use this material to support the analysis carried out in WP3, we will ask museum staff and field archaeologists to interact with images and 3D models accompanied by narratives. We will ask them to describe each paragraph using a set of keywords and a short explanation of why such a set of keywords was used in relation to the data.

## Methodology

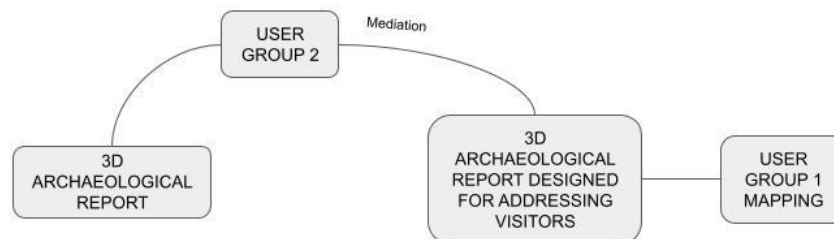
The experiment consists of two distinct steps. The first step aims to investigate how archaeologists use or incorporate legacy data previously stored in AIR to formulate the excavation strategy for upcoming work. Prior to the excavation, AIR will be used to review the extensive dataset to identify the objectives of the excavation campaign. The entire meeting will be recorded and analysed to determine the focus of the discussion and how the information was used or reused. During the excavation the archaeological data will be recorded, processed, archived, and used for discussion. The discussions will be recorded daily, and the videos will be used to observe how and when archaeologists reuse data to support different interpretations and hypotheses. The results of this work will provide indications of the real affordances of the system in supporting the 'dynamic' construction of different narratives.

## STEP 1



The second step of the experiment will focus on reusing the archaeological records available through AIR to generate new narratives to address different museum visitors. With the help of communicators and archaeologists from Blekinge Museum, we will use AIR to co-produce stories that directly address museum visitors and that can be used by the museum to support various activities. Once the story is published through AIR, we will ask the museum staff to describe the archaeological records used in the stories using keywords.

## STEP 2



The information collected across the two steps will be used to inform tasks in WP3.

### Personnel

Nicolò Dell'Unto (Lund University)  
 Paola Derudas (Lund University)  
 Åsa Berggren (Lund University)  
 Mikael Fauvelle (Lund University)  
 Mikael Henriksson (Blekinge Museum)  
 Andreas Svensson (SydSvenskarkeologi)

### Users

Field archaeologists with strong experience in producing advanced digital documentation  
 Field archaeologists with extensive experience in the use of digital data  
 Field archaeologists with no experience of producing or using digital data  
 Museum staff specialising in communication and archaeological data management  
 Archaeologists from the commercial sector  
 Archaeologists in the academic sector



## Reuse of Existing Archaeological Site Photographs and Illustrations

### Site/scenarios description

The second experiment is being conducted by MOLA London and is taking place between May and Autumn/Winter 2023. It sets out to (1) explore which types of legacy photographic materials are reused and generate user metadata; (2) investigate the ways in which these legacy materials and their combinations can aid or elicit narrative creation; (3) test how these metadata can be processed and inform the creation of controlled vocabularies and ontologies.

### Legacy data scenarios

The legacy data included in this experiment are (1) born-digital photographs from a particular excavation conducted by MOLA (Bloomberg site); (2) the existing metadata (from a variety of sources) which are attached to the digital photographs, but which are collated from more than one source of information.

### Overview of the experiment

This scenario will act as a first pilot test of the working relationship and methodology used by MOLA, WP2 and WP3 collaborators and external audiences, drawing on photographic records linked to at least one of MOLA's previous excavations. The aim is to enhance these records to be more usable by different stakeholders (see below under Possible Users) by working directly with those users to tag the records with more emotive, values-driven, descriptive, and evocative metadata. This work is important to exploring not only the creative/storytelling potential of particular legacy data (primary and secondary) over others, but also in mapping how different users employ different strategies in reusing these data and creating their own (tertiary) metadata.

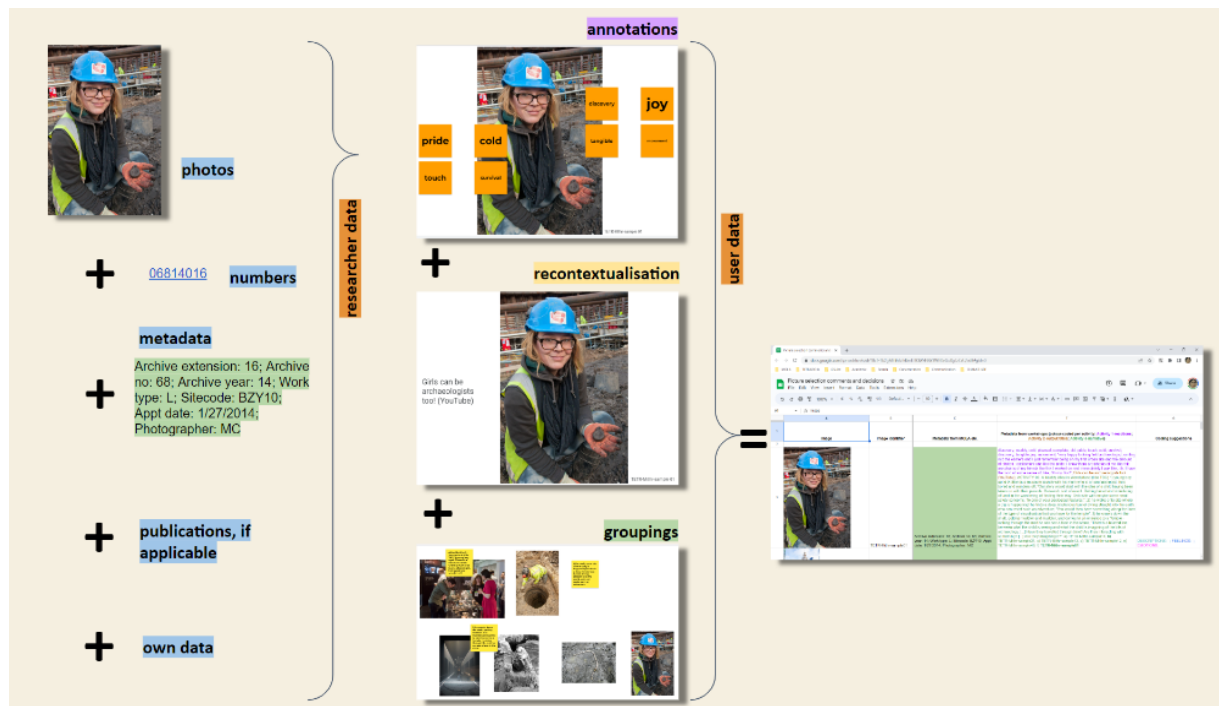


Figure 2: Visualisation of the first half of the photographic experiment, showing (from left to right) a) the process of collating the 'original' primary (photographs) and secondary (metadata) materials, b) examples of the categories of user-generated additional (tertiary) metadata, resulting from three specific tasks, c) an example of the collation of all materials (primary, secondary and tertiary) derived from a single photograph, prior to coding and other processing in the context of WP3.

## Methodology

Our methodology runs as follows:

1. We selected one existing MOLA site (per the criteria listed below).
2. We will work with small groups of user audiences (per Possible Users listed below) in short workshops to attempt to enhance the site's existing photographic metadata with richer terms and tags (based on the interests, values and needs of the different users themselves). We have done this digitally thus far, but may also do this by hand, or by uploading a small selection to Madoc or another appropriate platform to enable crowdsourcing of the work.
3. We will mine the captions associated with published site records available from the ADS archive of some of these same photos/illustrations, looking for descriptive terms that could be used to enhance the metadata.
4. We will work to integrate the outputs of steps #2 and #3 above into Omeka-S so that they can then be searched by other users for their own creative purposes.
5. We will then host at least one workshop where these other users test how reusable the enhanced records are now.

This case study required us to select one or more high-potential sites from MOLA's archives.

**What are the criteria for selecting the site(s) to use in this case study?** The site(s) must have:

1. Existing photos or illustrations which include a variety of finds and features, including studio-based artefact shots, typical on-site unit and feature shots, shots of team members (whether full body shots or partial shots of hands or heads, etc.).
2. Photos or illustrations of sufficient quality to be archived (e.g., shots should be in focus, and of high enough resolution to allow reproduction in different forums). The content of the shots themselves do not need to have 'wow-factor', but the quality of the reproduction itself needs to be good or excellent.
3. A local community from whom we could seek out representatives to participate in co-development and evaluation activities. In other words, we should prioritise sites which have a local audience currently living or working on top of or next to them.
4. Records deposited with ADS to allow us to mine captions from published imagery for relevant storytelling content.
5. Interesting potential for storytelling. In other words, a more vertically or horizontally complex site (covering different time periods or large stretches of the landscape) is preferable to a site where stories might be limited to a handful of speculations about a limited number of finds or features. We may choose a site that has already been the subject of considerable storytelling efforts (e.g., already been picked up by the press or featured in magazines or on TV) or we may wish to choose site(s) that have seen no attention, as advised by the MOLA graphics and comms teams.

## **Process**

The experiment will use legacy photos and associated legacy metadata (as above). A series of workshops will be run, the first being a pre-test by MOLA colleagues. The participants for each workshop will be chosen according to different criteria (as explained above). A selection of photographs will be made available to different groups of participants (specialists, creatives, memory institutions, local communities, and wider publics - see below). Although the content (questions/tasks) of the workshops will be adjusted for the needs of each audience, the basic structure remains the same, for consistency (and after internal project consultation): three progressively more detailed tasks designed to elicit creative responses and metadata. The tasks will involve annotation, re-contextualisation, and grouping. These tasks will create further metadata, which are then collated and evaluated, before they are utilised and processed as part of WP3.

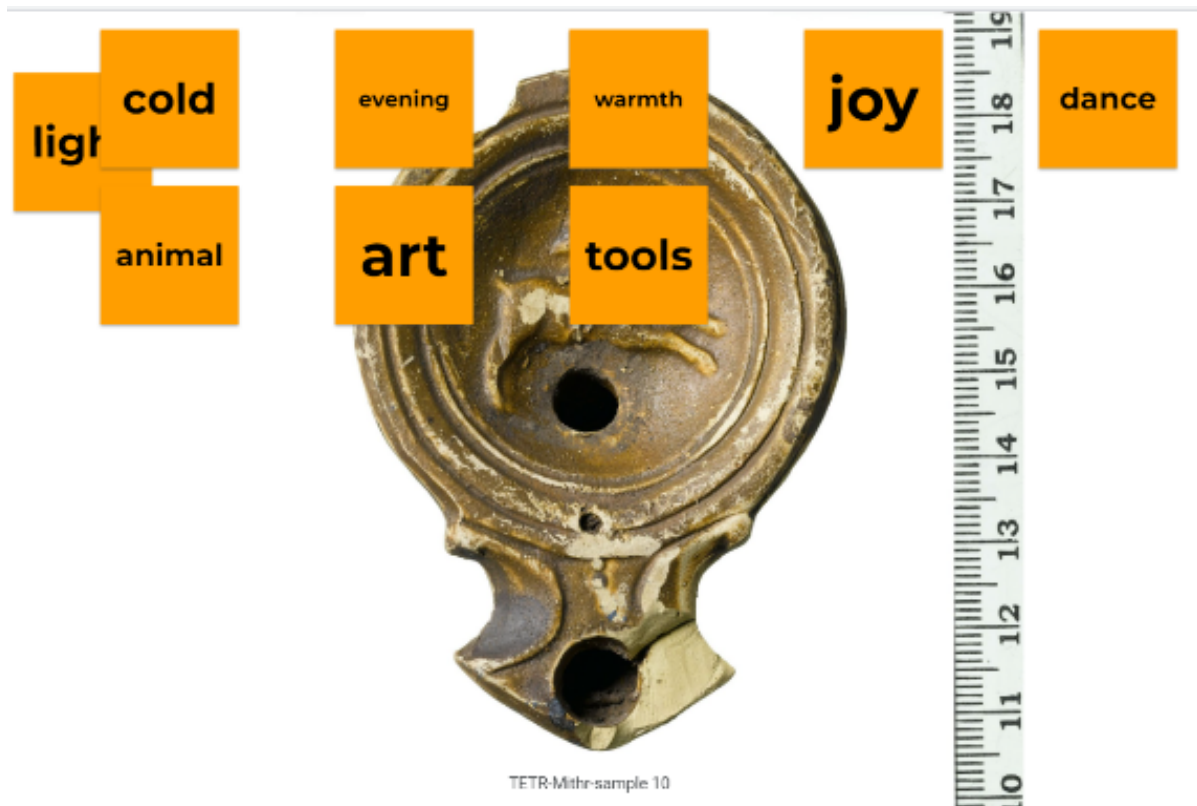


Figure 3: Example of annotation created by one of the participants during a pre-test workshop conducted by MOLA during the design of the experiment.

### Personnel

Sara Perry (MOLA)  
Anna Simandiraki-Grimshaw (MOLA)  
Andy Chopping (MOLA)  
Stu Eve (MOLA)  
Piraye Hacıgüzeller (University of Antwerp)  
Lise Foket (Ghent University)  
Aida Fadioui (University of Antwerp)

### Possible Users

As this will be a pilot test, we would like to include a variety of audiences in the work to ensure the records are enhanced in a way that represents their interests, and that we understand what supporting resources are needed to facilitate meaningful audience engagement in TETRARCHs in other case studies with different types of stakeholders.

We would like to conduct metadata enrichment workshops with the following groups:

1. MOLA specialists and archaeologists who created or use the site records [TETRARCHs audience = domain specialists].
2. MOLA communications team who use the site records to publicise archaeology to wide public audiences in different media, including press releases, on TV and social media [TETRARCHs audience = communications specialists and constituents of MOLA as a memory institution].
3. University students seeking to use the site records for educational or creative purposes [TETRARCHs audience = U Glasgow/Tessa Poller - creative practitioners and domain specialists in training].
4. Creative trainees or practitioners, who may or may not be connected to the site, but who might find artistic/creative inspiration in the sample archaeological photos, such as artists, media professionals, storytelling practitioners etc.
5. Community representatives who live directly on or next to the site(s) [TETRARCHs audience = constituents of MOLA as a memory institution], educators and other local communities/associations.

In addition, through NLP, we will aim to mine photo and illustration captions from published works associated with the site(s) for potential insights into how to enrich the metadata linked to those photos/illustrations. This will provide further understanding around how domain specialists are currently describing their records in published accounts held in the ADS's repository.

## Iterative Capture of New Data

### Site/scenarios description

This scenario will deploy image-based modelling techniques as part of a digital workflow at an excavation/survey field project centred upon a Neolithic village in northern Greece at Toumba Serron within a digital field data acquisition workflow. As part of the primary documentation of the excavation, 3D recording will be collated alongside the digitised written and graphic archive, alongside digital photograph and survey control data and spatial records in the Lund University's AIR system.

By developing this experiment, it may be possible to understand if the graphic representation currently produced on-site is sufficiently representative to achieve the research goals set by the project, as well as further user needs from the broader network of project stakeholders. The experiment will seek to test whether additional parameters should be considered to take a broader set of (re)user needs into account when making the models. In addition, this experiment will also assess if the metadata used for describing the 3D models sufficient to support knowledge creation by groups beyond the core research team and knowledge transfer to wider audiences.

## Legacy data scenarios

The AIR system will include data from previous seasons, and the TSRP hopes to be able to incorporate relevant maps and survey data (to be digitised).

## Overview of the experiment

The experiment is still in the design phases at the time of submission of this report (to be conducted in the 2023 summer field season - June-July). However, it is anticipated that there will be two elements under consideration.

- 1) **Reflective practice and workflow development.** This may include elements of auto-ethnography considering how the workflow is implemented and output into AIR. Elements of this line of experimentation will likely follow a similar pattern to those detailed for Västra Vång (above).
- 2) **Stakeholder evaluation and user needs mapping.** The project will seek to define the broader stakeholder network and assess their potential user needs, ideally to facilitate some input by these communities into the digital workflow and data structure (co-design) and their input into the interpretation of the site (co-creation of knowledge and share value).

It is worth noting here that the Toumba Serron Research Project already includes a complementary ethnographic study which may offer a ready-made environment to conduct initial stakeholder evaluations and elicit feedback from those stakeholders on the methods deployed, as well as fostering an inclusive environment to engage in participatory approaches and the co-creation of archaeological knowledge about the site.

Through these experiments, we will identify several potential outputs:

1. A clear project-centric stakeholder map and user-needs assessment.
2. A critical reflection of the project's hybrid workflow and detailed paradata statement that will inform best practices for inclusive participatory approaches to workflow design and knowledge creation.
3. A 3D excavation sequence produced in the field and made available through a 3D web visualisation system, which could be used by the ethnographer to support her work with the community.
4. The same datasets could be used to map how information from the ongoing excavation is transferred from archaeologists working in the field to archaeologists working in the laboratory (or in different trenches). Through these experiments, it may be possible to assess the potential of this dataset to be reused for transferring knowledge.

## **Personnel**

James Taylor (University of York, UK)

Nicolò Dell'Unto and Paola Derudas (Lund University)

TSRP Team Members and Ethnographer under the co-direction of Nicolas Zorzin (NCKU, Taiwan) and Dimitra Malamidou (Ministry of Culture, Greece)

Other TSRP WP2/3 representatives (TBC)

## **Users**

User groups beyond the core TSRP/TetrARCHs team will be defined as part of the proposed stakeholder evaluation and user needs mapping exercise, so will likely be updated in detail in the final reporting on this experiment.

# **Case study and experiment in preparation**

## **Site/scenarios description**

In this scenario, archaeological airborne LiDAR techniques will be used as part of a digital workflow in an excavation/survey field project focused on an Iron Age settlement at Pivka in western Slovenia, as part of a digital workflow for field data acquisition workflow. As part of the field survey, airborne LiDAR data collected nationally in 2015 will be reprocessed specifically for archaeology. The data will be used together with aerial photographs, historical maps, geological and soil maps as well as survey control data and spatial records in the Arkas 2.0 sites and monuments database (<https://iza2.zrc-sazu.si/en/zbirka/arkas>) from ZRC SAZU.



Figure 4. “Baba” Iron Age hillfort near Pivka Slovenia. Visualisation of airborne LiDAR data after the “removal” of vegetation, symbolized in the figure by aerial photography.

Through the development of this experiment, it will be possible to understand whether LiDAR-derived data currently produced with archaeology-specific data processing is sufficiently representative to meet the research objectives set by the project, as well as further user needs from the wider network of project stakeholders. The experiment aims to test whether additional parameters should be considered when taking into account a wider range of (re) user needs when processing the data. In addition, this experiment would also assess whether the metadata used to describe LiDAR-derived visualisations is sufficient to support knowledge creation by groups outside the core research team and knowledge transfer to a wider audience.

### Legacy data scenarios

The airborne LiDAR data was acquired in 2014/15 as part of a nationwide project. The primary aim of the data collection was to strengthen flood protection, which means that the data was collected according to specifications for mapping geomorphology. In archaeology, such data is referred to as a general type, which means that its usability for archaeology must first be established.

These data have a nominal density of 5 pnt/m<sup>2</sup> and an estimated horizontal and vertical RMSE of 0.09 m. Ground point filtering was performed with the proprietary filter gLidar and is distributed via the eVode WMS (LAS or LAZ format, ISPRS classes 0–7 are populated). The SI1 test site has 2,809,344 points with an average density of 12.28 pts/m<sup>2</sup> and the SI2 has 2,756,756 points with an average density of 12.43 pts/m<sup>2</sup>. It must be mentioned that due to the specific processing of the raw data, each point has a near-double “shadow point”. The “shadow point” is within 0.14 m horizontally and within 0.03 m vertically from the original point, but the distance and bearing to the original is not constant. This means that the functional point density is half of the above-mentioned, namely 6.14 pts/m<sup>2</sup>.



Since these data are now several years old and have not been acquired specifically for archaeology, they can in a sense be considered legacy data. However, since these data have not yet been processed specifically for archaeology and their archaeological potential is unknown prior to the experiment, they also have many similarities with archaeological data.

### **Overview of the experiment**

All remote sensing data have three main steps of data capture. The first step is sensor or raw data acquisition (e.g., the time of flight of a laser pulse). The second step is to process these data into archaeology-specific derivatives (e.g., LiDAR visualisation). The second step consists of informed but subjective discrete micro decisions by the operator, where the archaeology-specific data is created. In the third step, the archaeological interpretation takes place, which can be considered as the creation of archaeological information.

The third step includes field survey, which traditionally involves close observation and interpretation of the ground surface, direct measurement of all significant features, detailed analysis of the relationships between features, and preparation of a plan illustrating the interpretation obtained. Although access to airborne LiDAR data fundamentally redefines surveying in certain cases (e.g., dense forests), it is not a complete substitute for interpretation or field skills. There is an essential aspect of field survey that must be applied to the interpretation of LiDAR data if the latter is to retain its interpretive rigor. It is a matter of understanding the undeniably positive effect that a set of practices and sensibilities acquired during fieldwork have on archaeological interpretation. It stems from a dense, extensive, and experiential engagement with its subject landscape that few other archaeological methods can achieve.

However, archaeological practice is increasingly moving in the direction of replacing, rather than amending field survey with LiDAR and other remotely sensed data. To counter this trend, field survey needs to be redefined and fully integrated into the workflow of archaeological interpretation of digital born remote sensing data (and vice versa).

As part of the TetRARCHS project, we will conduct field experiments to qualitatively and quantitatively understand the interplay between digital remote sensing data and the practice of field survey. The goal is to understand the strengths and weaknesses of both and thus heal what is currently a deeply divided field of research.

### **Different audiences**

Archaeologists specialised in field archaeology

Archaeologists specialised in LiDAR

Archaeologists in general

General audience

story-telling specialists (not specialised in archaeology).

## Personnel

Edisa Lozić (ZRC SAZU, SI)

Benjamin Štular (ZRC SAZU, SI)

W&A (The Wolves and The Caesars: Digital Landscape Archaeology In Slovenia, Archaeological field school teaching survey, excavation, LiDAR, and photogrammetry) team members

Other ad hoc team members, representatives of different types of end-user groups

## Users

User groups beyond the core W&A/TETRARCHS team will be defined as part of the proposed stakeholder evaluation and user needs mapping exercise, so will likely be updated in detail in the final reporting on this experiment.

# Conclusion

This deliverable has outlined the range of experimental workflows developed by TETRARCHS partners within the first six months on the project. Most of these experiments will be implemented over the next six months as part of Task 5.2: *Experiment Design Implementation*. This will allow reflection and time for iterative improvements in the second year, (as appropriate) within Task 5.3: *Updates to Experiment Design Implementation*. Both the initial and updated implementation results will be reported iteratively in project Months M15, M24, M33, within Deliverable 5.2: *Experiment Trials*.

**TETRARCHs is making archaeological data accessible to a wider range of people.**

**The project explores how data from excavations and post-excavation research can be used and reused for educational, creative, and other life-enriching purposes. We work collaboratively across a variety of communities to experiment with storytelling about the past.**

-  **For heritage professionals, we create new resources for developing and measuring the effectiveness of archaeological data for storytelling.**
-  **For memory institutions like museums and cultural centres, we create reference materials to support you using these new resources for storytelling.**
-  **For creatives and local citizens, we will create a platform where you can search for and experiment with storytelling about the past.**



**Get in touch:**

**hello@tetrarchs.org**

**WWW.TETRARCHS.ORG**

TETRARCHs is funded under the CHANSE ERA-NET Co-fund programme, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no. 101004509

