Journal of Computer Applications in Archaeology

Odyssey: A Spatial Data Infrastructure for Archaeology

CASE STUDY

]u[ubiquity press

RAFAEL SÁ
LUÍS JORGE GONÇALVES
JORGE MEDINA
ANTÓNIO NEVES
FERNANDO MARSH
MOHAMMED AL-RAWI
DANIEL CANEDO
RITA DIAS
TIAGO PEREIRO
JOÃO HIPÓLITO
ALBERTO LAGO DA SILVA
JOÃO FONTE
LUÍS GONÇALVES SECO
MARTA VÁZQUEZ
JOSE MOREIRA

*Author affiliations can be found in the back matter of this article

ABSTRACT

Geospatial data acquisition methods like airborne LiDAR allow for obtaining large volumes of data, such as aerial and satellite imagery, which are increasingly being used in archaeology. As in other subjects, the ability to produce raw datasets far exceeds the capacity of domain experts to process and analyze them, but recent developments in image processing, Geographic Information Systems (GIS), Machine Learning (ML) and related technologies enable the transformation of large volumes of data into useful information. However, these technologies are challenging to use and not designed to interact with each other. Hence, tools are needed to efficiently manage, share, document, and reuse archaeological data.

This article presents the Odyssey SDI platform, a spatial data infrastructure for annotating, validating, and visualizing data about archaeological sites. This platform is built upon GeoNode, and special-purpose modules were developed for dealing with archaeological information. The main contribution is the integration of remote sensing, GIS features and ML algorithms in a single framework.

CORRESPONDING AUTHOR:
Rafael Sá
IEETA, LASI, University of Aveiro, Portugal
rafael.sa@ua.pt

KEYWORDS:
Archaeology Information System; Remote Sensing; Spatial Data Infrastructure; Web Geographical Information System

TO CITE THIS ARTICLE:
Sá, R, Gonçalves, LJ, Medina, J, Neves, A, Marsh, F, Al-Rawi, M, Canedo, D, Dias, R, Pereiro, T, Hipólito, J, da Silva, AL, Fonte, J, Seco, LG, Vázquez, M and Moreira, J. 2024. Odyssey: A Spatial Data Infrastructure for Archaeology. *Journal of Computer Applications in Archaeology*, 7(1): 225–236. DOI: https://doi.org/10.5334/jcaa.115

1. INTRODUCTION

Archaeology has benefited from the utilization of modern tools like satellite imagery and aerial Light Detection and Ranging (LiDAR), enabling the acquisition of extensive data that can be employed for remote sensing of archaeological sites. LiDAR is a laser system that allows the creation of point cloud datasets of landscapes and surface features by sending pulses of energy to the Earth's surface and collecting them through a sensor (Parcak, 2017). From the point clouds, derived products can be obtained, such as Digital Terrain Models (DTM) to represent the bare Earth surface, removing all natural and non-natural features other than the ground (Zhou, 2017), helping to detect archaeological findings that might have been covered by vegetation. Visualization techniques can also be applied to these images to enhance or improve the visibility of patterns or anomalies in the landscape (Lasaponara & Masini, 2012), including traditional techniques such as hill-shaded maps (Luo et al., 2019), or more complex techniques such as the Local Relief Model (LRM) (Hesse, 2010), sky-view factor (Kokalj et al., 2011) or openness (Doneus, 2013).

The data can be integrated into a Geographic Information System (GIS) to be analyzed in conjunction with other sources of archaeological information available. However, as McCoy (2017) points out, the sheer size and complexity of the collected data far exceeds the ability of domain experts using today's software packages to process and analyze them. Recent developments in image processing, GIS, Machine Learning (ML), and related technologies enable the transformation of this large volume of data into useful information. The use of remote sensing with GIS allows the identification of areas with high potential as archaeological sites using predictive models (Nsanziyera et al., 2018; Parcak, 2017) and other techniques such as shape detection (Di Iorio et al., 2015; Di Iorio et al., 2008). Yet, these technologies are challenging to use and are not designed to interact with each other. The use of GIS is mainly for visualization and interaction with the data, and no platform covers the entire process, spanning from data acquisition, transformation, and management to automatic data analysis using machine learning, in an integrated and standardized way.

The project *Odyssey – Platform for Automated Sensing in Archaeology* aims to develop a Spatial Data Infrastructure (SDI) for storage, processing, distribution, access, and enhanced utilization and preservation of geospatial resources, such as geospatial data and services (Hu & Li, 2017). This SDI uses several sources of georreferenced data and applies image processing and artificial intelligence techniques to detect archaeological sites remotely. Georreferenced data is loaded and consolidated into a platform allowing the annotation of archaeological sites and artefacts and combining data from different sources for interactive data visualization. It also makes the interaction with third-party tools for data analysis and ML easier. It is prepared to support ground-truthing activities to discard potential false positives, validate the identified sites and contribute to the continuous improvement of the ML algorithms.

This paper presents the requirements, design, architecture and features of the Odyssey platform. This includes the terminology and data model for annotating archaeological sites and the design and the architecture of the system implemented on the top of GeoNode for data management, visualization and validation. The use of Open Geospatial Consortium (OGC) standards promotes the integration of spatial data from a wide range of sources and interoperability with other systems for data analytics and ML. The platform enables archaeological surveys to be targeted towards areas with greater archaeological potential, resulting in a more efficient management of human and financial resources. It also supports validation activities during fieldwork, which contributes to improving data annotation processes. To showcase the capabilities and key features of the platform, a case study is presented that focuses on the remote detection of burial mounds using LiDAR data captured using a Unmanned Aerial Vehicle (UAV).

The Odyssey platform is designed to improve the traditional process of archaeological prospection. Currently, information is collected through bibliographic research, databases, and aerial photographs, which must be manually consolidated by an expert and subjected to a visual inspection in the field to produce the report with the results. With the platform, there is a greater emphasis on georeferenced data, such as data obtained through LiDAR. All data are accessible in a single platform that allows an integrated view of the territory and its archaeological sites, even by non-expert users. It also allows for preparing data for the automatic detection of archaeological features using ML algorithms, integrating the results in a database and validating them, which is a novelty in information systems for archaeology.

This paper is structured as follows: Section 2 presents the design, architecture and key technical details of the Odyssey platform implementation. Section 3 presents a case study on the remote detection of burial mounds, including acquiring LiDAR data using drones, importing and annotating data, interacting with third-party tools to detect burial mounds, and validating the results automatically. Section 4 highlights the contributions of this work and the relevance of the platform, and presents the results of tests performed. Lastly, Section 5 presents the conclusions.

2. DESIGN AND ARCHITECTURE

The goal was to create an SDI, ensuring the use of standards for interoperability and giving emphasis to the use of metadata and sharing of geospatial data (Tripathi et al., 2020), and including a web GIS for the storage, manipulation, analysis and presentation of geospatial data. Similar to other SDIs and web GISs with applications in archaeology (Bernardoni et al., 2017; Gallo & Roberto, 2012; Matsui et al., 2012; McCool, 2014; Serlorenzi et al., 2021), this platform has functionalities such as uploading geospatial and non-geospatial data, searching information through metadata or geographical extent, visualizing different layers on maps (including layers produced by third-party applications), downloading data, and so on. The SDIs and GISs just cited give particular importance to standards for the interoperability of geospatial content, especially the OGC standards, and in some cases (Gallo & Roberto, 2012), to the INSPIRE directive, which aims to ensure the standardization of metadata of spatial datasets and services in the European Union.

It is assumed that the data is not public, so the platform must manage users and their access to information. Currently, it supports Portuguese and English languages. Initially, a spatial data infrastructure and a WebGIS application were designed to manage information on archaeological sites. Still, the platform is prepared to host other applications, namely a mobile application to support ground-truthing activities in the field and the ability to access data in locations where there is no mobile network connection. Sobotkova et al. (2021) address this issue in their work, particularly in the amount of information that needs to be stored offline. Furthermore, precautions should be taken so that, when synchronizing with the database, no data is lost due to changes that may have occurred.

The platform is implemented using GeoNode, a web framework that allows the development and implementation of SDIs. Buonanno et al. (2019) use GeoNode as the basis for implementing a DInSAR-related SDI, and conclude that it is an excellent framework for implementing an SDI that can be adapted to the project requirements, considering one of its greatest strengths to be the use of OGC standards.

GeoNode supports the uploading of vector and raster data, including shapefiles and GeoTIFFs, and handles metadata management, automatically exposing it to the catalog service for search and discovery capabilities. The users responsible for the data can assign permissions, defining which users or groups can view, edit and download the data and metadata. Once the data is uploaded, different layers, including external ones, can be combined and viewed on thematic maps. All data are accessible through OGC standards such as Web Map Service (WMS), Web Feature Service (WFS) and Catalog Services for the Web (CSW) (Corti et al., 2019).

2.1. ARCHAEOLOGICAL INFORMATION MODULE

GeoNode is designed to be extended and customized, so new Django applications can be created to implement the project requirements that GeoNode does not have. As the platform must allow the insertion of archaeological information, a Django application was created to handle and integrate this data with GeoNode. Django is a framework for the development of Web applications.

During the project, it was necessary to agree on the terminology to store this information to avoid misunderstandings and ambiguity. For a better organization of the information, it was decided to distinguish between archaeological sites and archaeological occurrences. An archaeological site is defined as the area of archaeological intervention or area of interest delimited by a surrounding polygon. The surrounding polygon can be updated over time. The occurrences are the features of archaeological interest in a given site, There may be several occurrences for each site and each occurrence is delimited by a bounding polygon. For visualization and reporting purposes, both the sites and the archaeological occurrences can be depicted as points, calculated using the centroid of the polygons. Figure 1 presents the domain model using the notation presented in Teorey et al., 2006, describing the main entities of the archaeological module and how they relate to each other. The entities are represented by rectangles that include their name and attributes, and the relationships between the entities are represented by the lines that connect them. Relationships of 1 to 1, or 0..1 if not mandatory, indicate that there is a relationship of at most one instance between the entities. For instance, a site has only a site status, which is “Not verified” or “Verified”. One-to-many relationships, with many being represented by 1..* or 0..* if not mandatory, indicate that one entity can relate to multiple instances of the other entity. For instance, a site can have one or more occurrences, and an occurrence is related to a single site. The many-to-many relationship indicates that one entity can be related to several instances of the other entity and vice-versa. For instance, a document can be linked to several occurrences and an occurrence can be linked to several documents.

Archaeological sites and occurrences represent the main entities for archaeological information. They can be represented geographically by a point or polygon and their attributes are defined based on the fields of the site records used in Portugal. Archaeological sites must have at least one occurrence and it is possible to insert information, namely the national site code, name, country, parish, and status, which can be “Verified” or “Not Verified”. Similarly, it is also possible to insert information about the occurrence, in particular its name, acronym, toponym, altitude, owner, and status. The attribute status of an occurrence can hold two additional values relative to the status of a site. These values are “Verified – False Positive” and “Verified – True Positive”, differentiating the occurrences entered manually from those detected by the algorithms. Metrics can also be

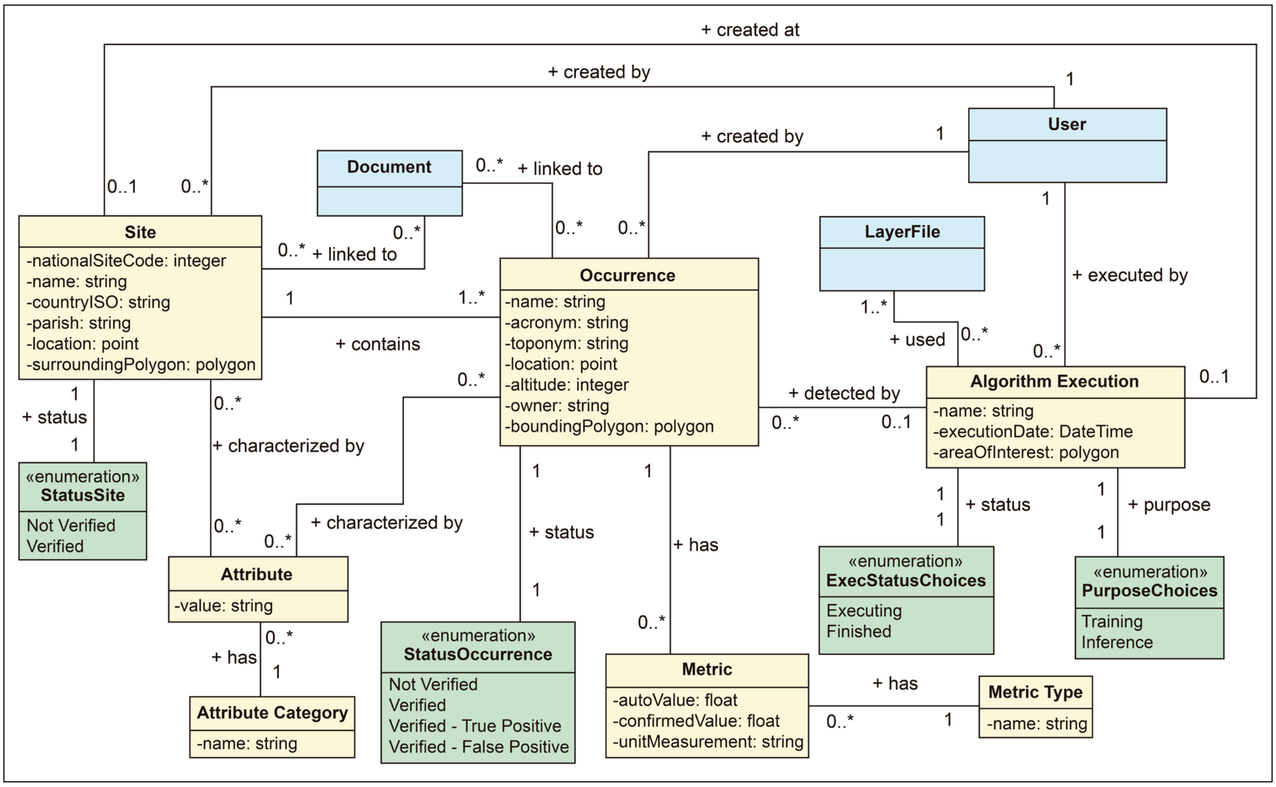


Figure 1 Domain model of the archaeological information module. The yellow entities represent the models that have been created from scratch, the blue entities represent the models already existing in GeoNode, and the green enumerations represent the possible options for the attribute.

associated with occurrences, such as area, diameter, length, and width, to name a few.

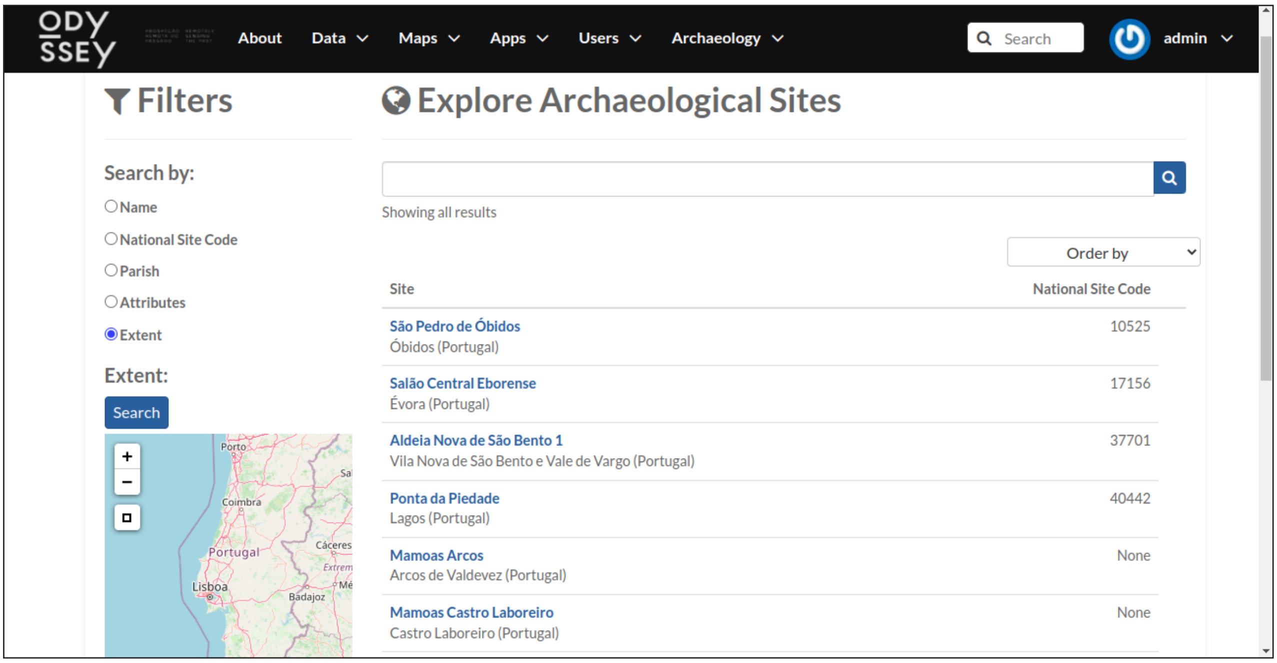
A variety of attributes can be associated with both sites and occurrences. The categories and values for these attributes were adapted from the Endovélico thesaurus, the Portuguese archaeological information and management system, and can be changed as needed. The uploaded files, whether spatial data, text files, photos, videos or others, can be linked to the sites and occurrences, thus making all available information about a particular site and its occurrences centrally available. Also, the trace of the execution of the algorithms to detect archaeological sites is saved, and it is possible to check whether a given execution has finished or is still being executed and its purpose.

During the implementation of the data models, an effort was made to use the models already existing in GeoNode. For the users, documents, and layer files, the existing models were used in order to take advantage of the existing functionality, while the remaining models were created from scratch. The spatial data regarding sites and occurrences were stored using the WGS84 - EPSG:4326 reference system to be consistent with the spatial reference system that GeoNode uses, despite the project was using ETRS89/Portugal TM06 - EPSG:3763.

To import the occurrences that have been manually annotated using external tools such as QGIS, a feature has been implemented that allows to load CSV files in a predefined format. The first column of the file, “WKT”, should contain the polygon(s) delimiting the occurrences in Well-Known Text (WKT) format, an OGC standard for the representation of geographical data, such as points, lines and polygons. The polygons must be of MultiPolygon type, even if there is only one occurrence and, therefore, only one polygon. The spatial reference system should be the one used in the Odyssey project, and then the necessary transformations are made to store it in the database. The second column, “Id”, is the type of the occurrence(s) of the respective MultiPolygon, so that is automatically assigned.

To help the user interact with this information, a search tool was developed that allows obtaining a list of archaeological sites, occurrences and algorithm executions, with the possibility to fine-tune the search by text or by selecting a geographical area of interest (Figure 2).

When searching for archaeological sites by text, it is possible to filter by name, national site code, parish, and other attributes. For the search by archaeological occurrences, it is possible to filter by name, archaeological site, owner and altitude range. As for searching algorithm executions, it is possible to search by the name of execution or by the geographic extent of the area of interest. For the search by geographical extent, the area of interest is defined through zoom-in and zoom-out operations on the map’s visible area. For an improved presentation, it is possible to sort the search results, for example, by most recent or least recent, by name in ascending or descending order, and others.



Search by:

○ Name
○ National Site Code
○ Parish
○ Attributes
● Extent

Extent:

Showing all results

Site	National Site Code
São Pedro de Óbidos Óbidos (Portugal)	10525
Salão Central Eborense Évora (Portugal)	17156
Aldeia Nova de São Bento 1 Vila Nova de São Bento e Vale de Vargo (Portugal)	37701
Ponta da Piedade Lagos (Portugal)	40442
Mamoas Arcos Arcos de Valdevez (Portugal)	None
Mamoas Castro Laboreiro Castro Laboreiro (Portugal)	None

Figure 2 Interface for archaeological sites search. The left sidebar shows the textual filters and the map for the geographical search. The textual search bar, the search results and sorting options are displayed on the right side.

2.2. AUTOMATIC DETECTION OF ARCHAEOLOGICAL SITES

A key feature of the platform is the possibility to execute algorithms that allow the identification of sites with a high potential of having archaeological occurrences. These algorithms (Canedo et al., 2023), developed by members of the Odyssey project, use the information uploaded to the platform and can be executed for training purposes or to infer information. The results obtained are also stored and published on the platform.

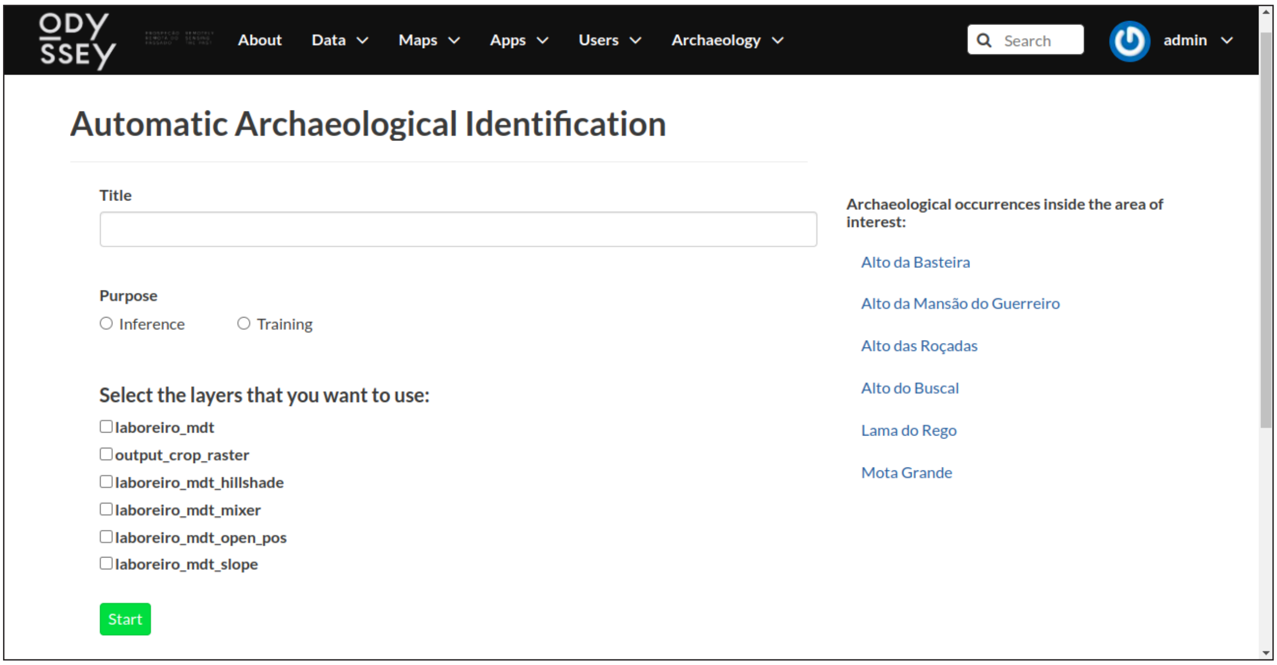
To execute the algorithms, an area of interest has to be selected first, which must intersect at least one layer and, if it is for training the algorithm, contain at least one verified occurrence with a defined polygon and occurrence type. It is essential to have the polygon defined, as the point does not define the geographical limits of the occurrence and, therefore, cannot be used by the algorithms. The need to have the occurrence type assigned is because this is how the algorithms distinguish between occurrences. Finally, occurrences with the status "Not Verified" and "Verified – False Positive" are not considered to not mislead the algorithms. If all conditions are met, it is possible to select the layers that intersect the area of interest and that are intended to be used in the execution (Figure 3). The ability to choose the layers is useful, as results may differ depending on the processing the layer has undergone.

The layers and occurrences are fed to the algorithm in a predefined format. Occurrences are grouped according to their type, with the polygons of each type being grouped in a MultiPolygon in WKT format. The layers are converted to binary format in base64 encoding. The response with the algorithm's detections follows the same format as the occurrences sent. If the results contain new detections, then a new archaeological site is created where the identified occurrences are inserted. By default, the status of the occurrences detected by the algorithms is "Not Verified", as they need to be verified in the field. After being verified, the status should be changed to "Verified – True Positive" or "Verified – False Positive", depending on whether it was, in fact, an archaeological occurrence or not. Using these can be useful for calculating metrics of the algorithms, such as their accuracy.

2.3. PLATFORM ARCHITECTURE

The platform architecture is derived from GeoNode, adhering to standard web GIS architectures and service catalogue conventions. A web GIS typically follows a three-tier client-server architecture, consisting of the data layer, middle layer that includes the map server, and presentation layer (Rodríguez Luaces et al., 2004; Tiwari & Jain, 2013). The service catalogue, on the other hand, relies on a Service-Oriented Architecture (SOA), as it is a distributed, loosely coupled architecture which follows the publish-find-bind pattern (Agrawal & Gupta, 2017; Hu & Li, 2017; Vaccari et al., 2009).

GeoNode combines the components needed: the spatial database, map server, cache server, catalogue service, and user interface. The spatial database is PostgreSQL with PostGIS, the map server is GeoServer, which already includes GeoWebCache as a caching server, and the service catalogue is pycsw. For the user interface, GeoNode uses Django templates, together with JavaScript libraries such as Leaflet and OpenLayers (Corti et al., 2019).



Automatic Archaeological Identification

Title

Purpose
○ Inference ○ Training

Select the layers that you want to use:
☐ laboreiro_mdt
☐ output_crop_raster
☐ laboreiro_mdt_hillshade
☐ laboreiro_mdt_mixer
☐ laboreiro_mdt_open_pos
☐ laboreiro_mdt_slope

Archaeological occurrences inside the area of interest:
Alto da Basteira
Alto da Mansão do Guerreiro
Alto das Roçadas
Alto do Buscal
Lama do Rego
Mota Grande

Figure 3 Interface to fill in the information and select the layers to be used to start the algorithm execution. This interface is presented after selecting an area of geographic interest and presents the user with the occurrences that are within that area.

Figure 4 presents an overview of the platform architecture. The platform receives the data acquired using an UAV, including DTMs in GeoTIFF format derived through the LiDAR point clouds (see arrow ①), and archaeological annotations in shapefile format ②. The annotations delimit the region of archaeological interest in the GeoTIFF files and are carried out by domain experts. These data are integrated into the platform ③④ and can be used for visualization and edition ④⑤, or to feed the algorithms for the detection of archaeological sites automatically ④⑥, among other applications. This allows the archaeologists to directly analyze the results and check if ground-truthing is necessary. In these cases, the fieldwork can be focused on the areas with greater archaeological potential.

The algorithms are integrated with the platform in such a way as to be independent of whether they are hosted on the same server or not. The algorithms are made available through a web service with an endpoint prepared to receive the information needed for them to execute in JSON objects, using the predefined format. The endpoint is responsible for receiving and processing the information, feeding the algorithm, and returning the results obtained if any ⑥.

To support the analysis of archaeological information, the use of base layers is necessary. These base layers can be uploaded as external services ⑦, having the advantage of being easier to keep the information up to date from official sources. The base layers for the Odyssey project include the Official Administrative Map of Portugal, the Land Use and Land Cover Map and aerial orthophotos, all made available by DGT (Directorate General for Territorial Development), Geological maps from LNEG (National Energy and Geology Laboratory), and others.

The Django application created to handle the archaeological information was included in GeoNode ⑧. However, after creating the data models for the archaeological sites and occurrences, it was necessary to find a solution to publish these models as GeoNode layers to take advantage of the existing functionalities, such as metadata handling and viewing features in the maps.

The solution is to publish the sites and occurrences as layers in GeoServer ⑨ and then use a command that GeoNode provides to synchronize its layers with the GeoServer layers ⑩. However, publishing the sites and occurrences as layers directly from the database tables is not possible, as each object can be represented by both point and polygon, and GeoServer assumes only one geometric representation. GeoServer addresses this issue and provides several alternatives for publishing the data correctly. The solution used is the creation of views, as it is a less invasive solution that does not require data restructuring and is versatile. Views also allow the selection of data with specific properties that are not restricted to a single data table. This is relevant, as the attributes of sites and occurrences are stored in different tables, and also because it allows publishing the information separately according to its status. Thus, a total of eight views were created for the occurrences and four views for the sites. Each view presents information from a specific status and with only one of the geometry types, so the total of twelve views created consists of the possible combinations between the available status and geometry types. This way, the user can choose to view, for example, only the sites or occurrences that have already been verified in the field, without seeing at the same time those that have not been, and also choose what kind of geometric representation to see ⑪ (Figure 5).

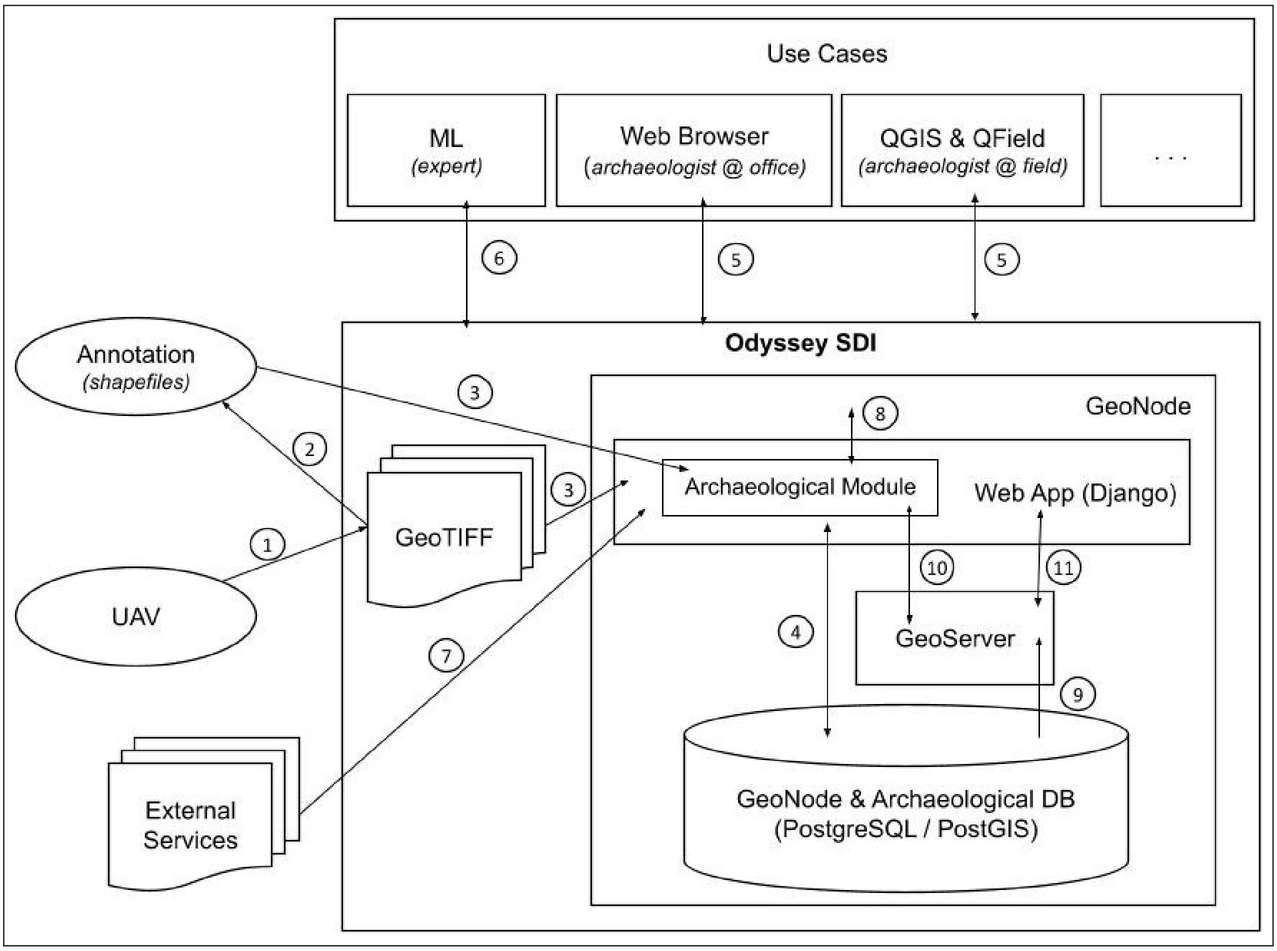
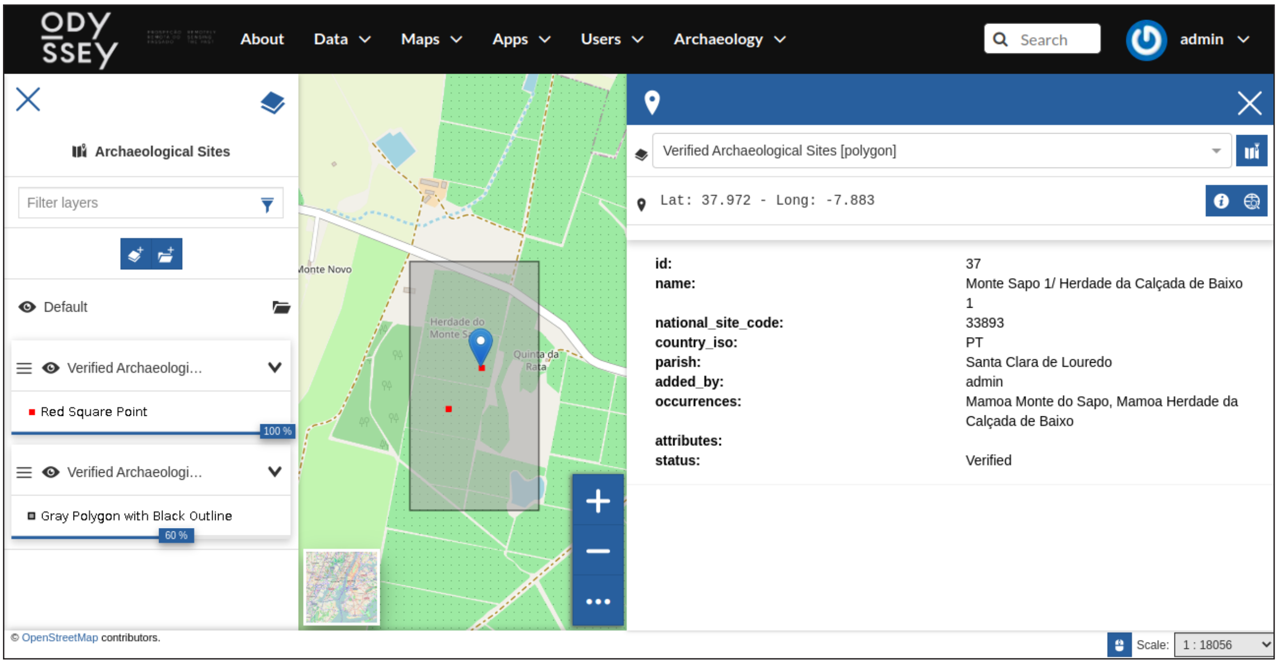


Figure 4 Architecture of the platform.



Verified Archaeological Sites [polygon]

Lat: 37.972 - Long: -7.883

id:	37
name:	Monte Sapo 1/ Herdade da Calçada de Baixo 1
national_site_code:	33893
country_iso:	PT
parish:	Santa Clara de Louredo
added_by:	admin
occurrences:	Mamoa Monte do Sapo, Mamoa Herdade da Calçada de Baixo
attributes:	
status:	Verified

Figure 5 Interface for visualization of the views as layers on a map, showing verified archaeological sites represented by polygons and verified archaeological occurrences represented by points. The right side panel displays the information provided regarding the archaeological site.

3. CASE STUDY

Portugal does not yet have a systematic LiDAR coverage for its national territory, contrary to other countries such as the United Kingdom or Spain. It is expected that in the near future there will be a greater demand for this type of data, so national coverage is anticipated. In 2018, the Intermunicipal Community of Alto Minho performed a pioneer LiDAR flight that fully covered the area of the district of Viana do Castelo, corresponding to approximately 2255 km^2, making it the first Portuguese region to have an airborne laser coverage. The generated point clouds have an average resolution of 2 points per square meter, allowing for the interpolation of DTMs with a 1 × 1 meter resolution. Additionally, several small-scale, high-resolution drone-based LiDAR surveys have been conducted at archaeological sites in Portugal, producing DTMs with a resolution of 0.2 × 0.2 meters.

Different visualization techniques were applied to the generated DTMs, such as LRM, sky-view factor, hillshade, and others. To provide archaeological data to train the ML algorithms, previously known archaeological occurrences, in this case burial mounds, were manually annotated into shapefiles. However, the platform is generic and works for the various types of archaeological occurrences. These data were used as input data to perform preliminary tests on the platform (Figure 6).

The YOLOv5 framework was chosen for the inference process (Canedo et al., 2023; Jocher et al., 2021). This object detection algorithm was trained with an augmented dataset. This dataset was obtained through a cropping algorithm which cropped images around the known burial mounds of Alto Minho. Afterwards, this dataset was augmented with a copy-paste algorithm, which cropped unused regions of Alto Minho to paste the original mounds, artificially increasing the background variety and the dataset size. The pasting process was not randomly done, as burial mounds were only pasted onto probable regions. This avoids pasting burial mounds onto rivers, lakes, buildings, streets, and infrastructures, just to name a few. Then, the trained YOLOv5 model was applied to the Alto Minho region to detect potential uncovered burial mounds. The generated inferences went through two post-processing validation steps. The first step validates inferences within a probable region. The second step validates inferences which have similar 3D features as the original burial mounds used in the training. These 3D features are calculated directly from the raw point cloud data. This process provided 648 inferences of potential burial mounds in the entire Alto Minho region. These were digitally validated by four archaeologists, allowing the preliminary discard of about 27.5% of the inferences (Canedo et al., 2023). The on-site ground truthing phase is underway since May 2023.

This information, together with the other data in the database, is important for planning the on-site inspections. The platform allows to overlay different visualization layers on a map, such as the locations of the burial mounds detected by visual inspection or using ML algorithms, DTMs and thematic maps, such as Land Use and Land Cover maps. It is also possible to zoom in on an area of interest or to filter the information using several criteria, for instance, to display only burial mounds with the status "Not Verified". The archaeologists can change the status of each verified occurrence to "Verified – True Positive" or "Verified – False Positive" during or after the

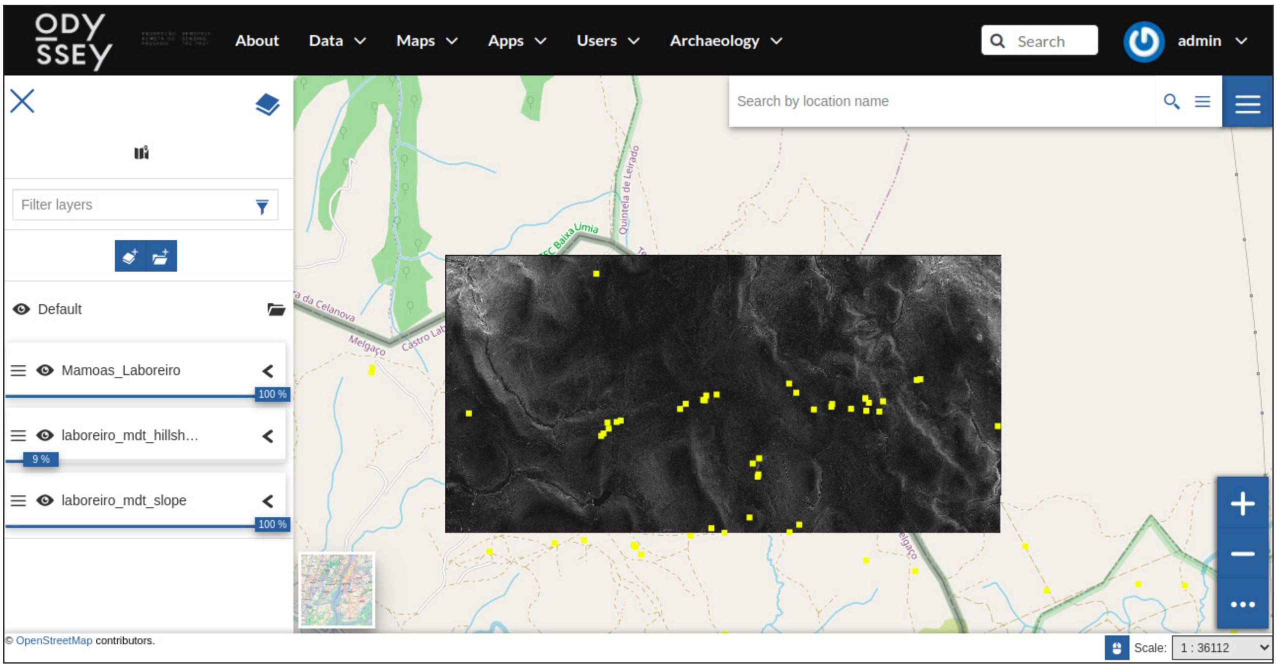


Figure 6 Interface for viewing layers on a map, showing a slope map derived from the DTM of Castro Laboreiro area with annotations, represented in yellow, of known burial mounds in the region.

on-site inspections. Not all occurrences can be validated because some are in places that are very difficult to access. The on-site inspections are an important step, as they will allow the retraining of the ML models using more reliable annotations, and it is expected to be able to gradually improve the performance of the ML algorithms used to detect burial mounds.

4. DISCUSSION

Archaeologists are gradually and sustainably adopting digital tools in their work. Currently, we are witnessing the increasing use of repositories such as The Digital Archaeological Record (tDAR) or the Archaeology Data Service, and the emergence of initiatives like the ARIADNEplus project for the integration of many archaeological data infrastructures across Europe. These resources enable searching large collections of curated data, reports, and publications, but extracting high-quality archaeological information from raw data is still a time-consuming and challenging task. However, the most recent advances in remote sensing, spatial data processing and ML can help to overcome these difficulties.

The Odyssey platform comprises a set of features, ranging from visualization, management and sharing of georreferenced and archaeological data, to metadata management and remote detection, thus offering a complete solution, while other works (Liu et al., 2021; Trepal et al., 2021) focus only on some of these functionalities. The use of OGC standards ensures high interoperability, essential for collaborative work and sharing and use of data efficiently and consistently, and avoid problems such as the lack of homogeniety mentioned by Ronzino et al. (2018). This allows for improving the efficiency and effectiveness of the processes of archaeological surveys, making them comprehensive that would otherwise be unfeasible for operational and economic reasons. With this, better quality information will be available as a result of increasing the rate of sites and their typology, correctly identified. This will also reduce the rate of sites detected only during the territorial intervention phase, facilitating their management and avoiding waste of resources. Furthermore, as the field prospection is more targeted to the areas identified as having great archaeological potential, this will reduce the time required and consequently reduce costs and improve the management of human resources.

The purpose of the machine learning algorithm employed within this platform is the identification and delineation of archaeological sites. Given the necessity to process extensive LiDAR data, the algorithm needs to be fast. Consequently, the YOLO (You Only Look Once) object detection algorithm was chosen. YOLO outperforms its predecessor algorithms in terms of processing speed and efficiency since it was designed to solve the significant processing time of region proposal-based object detection algorithms (Canedo et al., 2023). Even though the platform only uses the YOLOv5 algorithm during this beta phase, it has been developed so that the possibility of dynamically adding new algorithms can be implemented in a future version. This will give users the possibility to choose which algorithm they want to use, allowing them to compare different algorithms for the same dataset seamlessly within the platform.

A comprehensive series of tests were conducted on the platform to ensure its effectiveness and reliability. This included system tests to evaluate overall stability and performance, functional tests with use cases to ensure the platform met practical requirements, and field tests to support data validation and ground-truthing processes. This multidimensional testing approach allowed us to assess the platform's functionality in various real-world scenarios, ensuring it met the necessary standards to meet the demands of archaeological research. Furthermore, these tests provided valuable insights into the platform's impact on archaeologists' fieldwork, allowing us to analyze how it enhances their work in the field and also to identify problems and improvements for a later version.

The platform revealed strong potential to become a useful tool in the everyday work of archaeologists, not only because it allows geographic information to be uploaded with data that can later be used for multiple purposes, including the creation of maps and machine learning, but also as a database and an inventory and query tool. The platform's relevance is mainly associated with the need for archaeological study to follow technological advances in artificial intelligence and georeferencing of information collected in the field. The ability to communicate with external and mobile applications is a determining factor in the importance of continuing to develop this platform. In fact, two mobile applications are currently being developed to support archaeologists with on-site inspections, using the Odyssey platform for data access and management. Also worth mentioning is the fact that the platform was built to suit the needs of Portuguese archaeologists, which was one of the differentiating factors that stood out during the tests, considering that most of the software used in this area consists of adapting programs designed for other purposes. However, the platform is not limited to the Portuguese context and has been developed so that it is possible to configure more specific information for other countries, such as archaeological attributes and the thesaurus.

The tests demonstrated that the platform is functional and ready to be used in all stages of the remote detection of archaeological sites. The archaeologists adapted fast to the system, easily introducing it into their work routine. Overall, the platform has worked correctly and smoothly at all times. However, some areas for improvement have been detected, particularly regarding the platform's usability, since its technical features require some expertise in how it works. Usability improvements include adding more information to the help menus, for example when using automatic site identification, and refining the search engine to make search results more consistent. Although the terminology relating to archaeological sites and occurrences was defined in collaboration with the project's archaeologists, there was a lack of agreement among the wider community of archaeologists. For this reason, a future commercial version of the platform must have a feature that can handle the various interpretations made by its users, making it possible to configure its rules accordingly.

Archaeologists have also suggested another potential feature for the platform, which is the possibility of generating a PDF report with information on an archaeological site, its occurrences and all its attributes using the site record template. This feature would be an important benefit for archaeologists, since creating this record is part of the archaeologist's work routine and they currently fill out this record manually. Therefore, automating the process is an additional convenience that saves time and takes advantage of the centralization of information on the platform. These features, whose absence does not prevent the platform from being used, will make an important contribution in terms of usability to the platform's commercial success. Despite this, the platform has shown that it has the potential to become more relevant as feedback is gathered from its users, and that it has the capacity and flexibility to adapt to their needs.

5. CONCLUSION

Geospatial tools such as airborne LiDAR allow the retrieval of large volumes of data that are increasingly being used in archaeology. Recent developments in areas such as image processing, ML and GIS allow the transformation of large volumes of data into useful information. However, these technologies are complex for non-expert users and are not designed to work together.

The platform presented in this study exhibits the capacity to effectively manage and unify diverse georeferenced data alongside archaeological data. Additionally, it seamlessly incorporates machine learning algorithms to facilitate remote detection of archaeological sites. By combining and integrating these elements, the platform offers a robust solution for handling multidimensional datasets and enhancing archaeological site identification processes.

The tests conducted on the Odyssey platform showcased its functionality and relevance to archaeologists. The platform holds great potential for addressing the challenges faced by archaeologists, improving the efficiency of site detection, and enhancing data management. In the future, further usability improvements, expanded configurability, and additional features like automated report generation are planned to make the Odyssey platform an indispensable tool for archaeologists.

ADDITIONAL FILE

The additional file can be found as follows:

- Video demonstration of the platform. URL: https://vimeo.com/811442378

ACKNOWLEDGEMENTS

This work was carried out in the context of the project Odyssey – Platform for Automated Sensing in Archaeology (ALG-01-0247-FEDER-070150), made up by the consortium of ERA – Arqueologia, the University of Aveiro and the University of Maia. The project was co-financed by the European Regional Development Fund (ERDF), Operational Program Competitiveness and Internationalization.

The authors are grateful to the Intermunicipal Community of Alto Minho (CIM Alto Minho) for providing the airborne LiDAR data used in this work.

FUNDING INFORMATION

This Research was also funded by National Funds through the FCT – Foundation for Science and Technology, in the context of the project UIDB/00127/2020.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Rafael Sá orcid.org/0000-0002-7691-0110
IEETA, LASI, University of Aveiro, Portugal
Luís Jorge Gonçalves orcid.org/0000-0002-2454-127X
ESTGA, University of Aveiro, Portugal
Jorge Medina orcid.org/0000-0002-1885-7614
DGEO, University of Aveiro, Portugal
António Neves orcid.org/0000-0001-5433-6667
IEETA, LASI, DETI, University of Aveiro, Portugal

Fernando Marsh orcid.org/0000-0002-2332-8477
University of Aveiro, Portugal
Mohammed Al-Rawi orcid.org/0000-0002-0663-2485
IEETA, LASI, IT, University of Aveiro, Portugal
Daniel Canedo orcid.org/0000-0002-5184-3265
IEETA, LASI, University of Aveiro, Portugal
Rita Dias orcid.org/0000-0003-2999-3133
ERA Arqueologia, ICArEHB, Portugal
Tiago Pereiro orcid.org/0000-0003-2691-4583
ERA Arqueologia, Portugal
João Hipólito orcid.org/0009-0001-6362-4633
ERA Arqueologia, Portugal
Alberto Lago da Silva orcid.org/0000-0003-4037-3090
ERA Arqueologia, Portugal
João Fonte orcid.org/0000-0003-0367-0598
University of Maia-UMAIA, Portugal
Luís Gonçalves Seco orcid.org/0000-0002-8950-5499
University of Maia-UMAIA, Portugal
Marta Vázquez orcid.org/0000-0001-5261-4926
University of Maia-UMAIA, Portugal
Jose Moreira orcid.org/0000-0003-4633-6944
IEETA, LASI, DETI, University of Aveiro, Portugal

REFERENCES

Agrawal, S and **Gupta, RD.** 2017. Web GIS and its architecture: a review. *Arabian Journal of Geosciences*, 10(23): 518. DOI: https://doi.org/10.1007/s12517-017-3296-2

Bernardoni, S, Montanari, M and **Trojanis, R.** 2017. Open History Map. *Archeologia e Calcolatori*, 28(2): 539–548. DOI: https://doi.org/10.19282/AC.28.2.2017.44

Buonanno, S, Zeni, G, Fusco, A, Manunta, M, Marsella, M, Carrara, P and **Lanari, R.** 2019. A GeoNode-Based Platform for an Effective Exploitation of Advanced DInSAR Measurements. *Remote Sensing*, 11(18): 2133. DOI: https://doi.org/10.3390/rs11182133

Canedo, D, Fonte, J, Seco, LG, Vázquez, M, Dias, R, Pereiro, TD, Hipólito, J, Menéndez-Marsh, F, Georgieva, P and **Neves, AJR.** 2023. Uncovering Archaeological Sites in Airborne LiDAR Data With Data-Centric Artificial Intelligence. *IEEE Access*, 11: 65608–65619. DOI: https://doi.org/10.1109/ACCESS.2023.3290305

Corti, P, Bartoli, F, Fabiani, A, Giovando, C, Kralidis, AT and **Tzotsos, A.** 2019. GeoNode: an open source framework to build spatial data infrastructures. *PeerJ Preprints*, 7: 1–10. DOI: https://doi.org/10.7287/peerj.preprints.27534v1

Di Iorio, A, Biliouris, D, Guzinski, R, Hansen, LB and **Bagni, M.** 2015. Innovation technologies and applications for coastal archaeological sites FP7-ITACA. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences – ISPRS Archives*, 40(7 W3): 1367–1373. DOI: https://doi.org/10.5194/isprsarchives-XL-7-W3-1367-2015

Di Iorio, A, Bridgwood, I, Rasmussen, MS, Sorensen, MK, Carlucci, R, Bernardini, F and **Osman, A.** 2008. Automatic detection of archaeological sites using a hybrid process of Remote Sensing, Gis techniques and a shape detection algorithm. *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 1–7. DOI: https://doi.org/10.1109/ICTTA.2008.4529910

Doneus, M. 2013. Openness as Visualization Technique for Interpretative Mapping of Airborne Lidar Derived Digital Terrain Models. *Remote Sensing*, 5(12): 6427–6442. DOI: https://doi.org/10.3390/rs5126427

Gallo, P and **Roberto, V.** 2012. ANTEO: Sharing Archaeological Data on a WebGIS Platform. In: Fozzati, L and Roberto, V (eds.), *Proceedings of the 2nd workshop on the new technologies for aquileia*. CEUR Workshop Proceedings. pp. C1–C10.

Hesse, R. 2010. LiDAR-derived Local Relief Models – a new tool for archaeological prospection. *Archaeological Prospection*, 17(2): 67–72. DOI: https://doi.org/10.1002/arp.374

Hu, Y and **Li, W.** 2017. Spatial data infrastructures. Wilson, JP (ed.), *The Geographic Information Science & Technology Body of Knowledge*, 2017(Q2). DOI: https://doi.org/10.22224/gistbok/2017.2.1

Jocher, G, Stoken, A, Borovec, J, NanoCode012, Chaurasia, A, TaoXie, Changyu, L, Abhiram, V, Tkianai, L, YxNONG, Hogan, A, Lorenzomammana, AlexWang1900, Hajek, J, Diaconu, L, Marc, Kwon, Y, Oleg, … Ingham, F. 2021. ultralytics/yolov5: v5.0 – YOLOv5-P6 1280 models, AWS, Supervise.ly and YouTube integrations. DOI: https://doi.org/10.5281/zenodo.4679653

Kokalj, Ž, Zakšek, K and **Oštir, K.** 2011. Application of sky-view factor for the visualisation of historic landscape features in lidar-derived relief models. *Antiquity*, 85(327): 263–273. DOI: https://doi.org/10.1017/S0003598X00067594

Lasaponara, R and **Masini, N.** 2012. Image Enhancement, Feature Extraction and Geospatial Analysis in an Archaeological Perspective. In: Lasaponara, R and Masini, N (eds.), *Satellite Remote Sensing. Remote Sensing and Digital Image Processing*, vol 16. Dordrecht: Springer. pp. 17–63. DOI: https://doi.org/10.1007/978-90-481-8801-7_2

Liu, P, Loudcher, S, Darmont, J and **Noûs, C.** 2021. ArchaeoDAL: A Data Lake for Archaeological Data Management and Analytics. *25th International Database Engineering and Applications Symposium*, 252–262. DOI: https://doi.org/10.1145/3472163.3472266

Luaces, MR, Brisaboa, NR, Paramá, JR and **Viqueira, JR.** 2004. A Generic Architecture for Geographic Information System. *Universidad Da Coruña,* (July), pp. 1–234.

Luo, L, Wang, X, Guo, H, Lasaponara, R, Zong, X, Masini, N, Wang, G, Shi, P, Khatteli, H, Chen, F, Tariq, S, Shao, J, Bachagha, N, Yang, R and **Yao, Y.** 2019. Airborne and spaceborne remote sensing for archaeological and cultural heritage applications: A review of the century (1907–2017). *Remote Sensing of Environment*, 232: 111280. DOI: https://doi.org/10.1016/j.rse.2019.111280

Matsui, Y, Gidding, A, Levy, TE, Kuester, F and **De-Fanti, TA.** 2012. ArchaeoSTOR map: publishing archaeological geodata on the web. *Proceedings of the 3rd International*

Conference on Computing for Geospatial Research and Applications – COM.Geo'12, 1–2. DOI: https://doi.org/10.1145/2345316.2345355

McCool, J-PP. 2014. PRAGIS: a test case for a webbased archaeological GIS. *Journal of Archaeological Science*, 41, 133–139. DOI: https://doi.org/10.1016/j.jas.2013.07.037

McCoy, MD. 2017. Geospatial Big Data and archaeology: Prospects and problems too great to ignore. *Journal of Archaeological Science*, 84: 74–94. DOI: https://doi.org/10.1016/j.jas.2017.06.003

Nsanziyera, A, Rhinane, H, Oujaa, A and **Mubea, K.** 2018. GIS and Remote-Sensing Application in Archaeological Site Mapping in the Awsard Area (Morocco). *Geosciences*, 8(6): 207. DOI: https://doi.org/10.3390/geosciences8060207

Parcak, SH. 2017. *GIS, Remote Sensing, and Landscape Archaeology* (Vol. 1). Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199935413.013.11

Ronzino, P, Acconcia, V and **Falcone, A.** 2018. Towards the integration of spatial data through the Italian Geoportal for Archaeological Resources. *2018 3rd Digital Heritage International Congress (DigitalHERITAGE) held jointly with 2018 24th International Conference on Virtual Systems & Multimedia (VSMM 2018)*, pp. 1–5. DOI: https://doi.org/10.1109/DigitalHeritage.2018.8810132

Serlorenzi, M, D'Andrea, A and **Montalbano, R.** 2021. SITAR: a new open-data infrastructure for a public archaeology of Rome. In: Bogdani, J, Montalbano, R, Rosati, P (eds.), *Archeofoss xiv 2020: Open software, hardware, processes, data and formats in archaeological research* pp. 108–118. Archaeopress.

Sobotkova, A, Ross, SA, Hermankova, P, Lupack, S, Nassif-Haynes, C, Ballsun-Stanton, B and **Kasimi, P.** 2021. Deploying an Offline, Multi-User, Mobile System for Digital Recording in the Perachora Peninsula, Greece. *Journal of Field Archaeology*, 46(8): 571–594. DOI: https://doi.org/10.1080/00934690.2021.1969837

Teorey, T, Lightstone, S and **Nadeau, T.** 2006. Database modeling and design: Logical design, Chapter 3 –The Unified Modeling Language (UML). Morgan Kaufmann. DOI: https://doi.org/10.1016/B978-0-12-685352-0.X5000-9

Tiwari, A and **Jain, DK.** 2013. Geospatial Framework For Dengue using Open Source Web GIS Technology. *Joint International Workshop of ISPRS WG VIII/1 and WG IV/4 on Geospatial Data for Disaster and Risk Reduction November*, 21–22.

Trepal, D, Lafreniere, D and **Stone, T.** 2021. Mapping Historical Archaeology and Industrial Heritage: The Historical Spatial Data Infrastructure. *Journal of Computer Applications in Archaeology*, 4(1): 202–213. DOI: https://doi.org/10.5334/jcaa.77

Tripathi, AK, Agrawal, S and **Gupta, RD.** 2020. Cloud enabled SDI architecture: a review. *Earth Science Informatics*, 13(2): 211–231. DOI: https://doi.org/10.1007/s12145-020-00446-9

Vaccari, L, Shvaiko, P and **Marchese, M.** 2009. A geo-service semantic integration in Spatial Data Infrastructures. *International Journal of Spatial Data Infrastructures Research*, 4: 24–51. DOI: https://doi.org/10.2902/1725-0463.2009.04.art2

Zhou, Q. 2017. Digital Elevation Model and Digital Surface Model. In: Richardson, D, Castree, N, Goodchild, M, Kobayashi, A, Liu, W and Marston, R (eds.), *International encyclopedia of geography: People, the earth, environment and technology*. John Wiley & Sons, Ltd. DOI: https://doi.org/10.1002/9781118786352.wbieg0768

TO CITE THIS ARTICLE:

Sá, R, Gonçalves, LJ, Medina, J, Neves, A, Marsh, F, Al-Rawi, M, Canedo, D, Dias, R, Pereiro, T, Hipólito, J, da Silva, AL, Fonte, J, Seco, LG, Vázquez, M and Moreira, J. 2024. Odyssey: A Spatial Data Infrastructure for Archaeology. *Journal of Computer Applications in Archaeology,* 7(1): 225–236. DOI: https://doi.org/10.5334/jcaa.115

Submitted: 14 June 2023 **Accepted:** 16 January 2024 **Published:** 09 April 2024

Journal of Computer Applications in Archaeology is a peer-reviewed open access journal published by Ubiquity Press.