HUNGARIAN ARCHAEOLOGY
E-JOURNAL • 2025/3 – AUTUMN www.hungarianarchaeology.hu

COMMUNITY-BASED ARTEFACT AND DATA COLLECTION IN THE JÁSZSÁG REGION

‘Data-scratching’ by the Jász Kakasok and its scientific use

András Zoltán Gulyás[1] **– Viktor Hegyesi**[2]
Hungarian Archaeology Vol. 14 (2025) Issue 3, pp. 49–53.

The community of volunteer civilians supporting the work of the Jász Museum, known as the Jász Kakasok *(‘Jász Roosters’), has discovered and documented more than 7,000 finds during instrument-assisted surveys in recent years. These artefacts have already been incorporated into the central database of the museum. This unprecedented amount of data opens new dimensions in archaeological research; at the same time, it has become necessary to ensure that the data collected by volunteers are available in a reliable form enabling scientific interpretation. This need is met by the system we have named FindVault, a community archaeology platform that organises the data recorded in the field into a unified framework and integrates them into professional research databases using automated methods. Community archaeology thus gains a new toolkit: the enthusiastic work of volunteers is complemented by high-precision surveying instruments, QR-code tagging, and a smartphone application. In the following, the operation of the system, the steps of field data collection and processing, are presented, together with the ways in which GIS analyses, statistical procedures, and machine-learning methods are integrated into the workflow. A further key advantage of the system is that all recorded finds can be retrieved at any time within the museum, and their identification and inventorying can be completed quickly using a handheld barcode reader.*

Keywords: community archaeology, mobile application, server-side (backend) application, metal detecting, QR code

THE FINDVAULT SYSTEM

The purpose of FindVault is to support field data collection using modern tools and to integrate the data into the research process. Its central components are a smartphone-based application and a server-side backend service. Volunteers record their finds via the mobile application, which stores 12 attributes for each find (e.g., date, coordinates, unique code, finder, GPS accuracy, material, type, period, photographs, etc).

Finds are labelled in the field with a unique QR code; once scanned, the data are automatically written into a GeoPackage (GPKG) file. This eliminates the need for manual field notes, and coordinates are uniformly recorded in the Hungarian EOV (*Egységes Országos Vetület,* Hungary’s own projection system) projection.

The system is primarily online – each client connects to the central server through a dedicated VPN connection – but it also supports offline recording: in the absence of mobile data, the application stores data locally and synchronises automatically once connectivity is restored.

On the server side, a multi-layered system provides data management, consisting of an SQL-based database and WMS and PostgreSQL servers. A backend service receives uploads from the mobile clients; here users’ data appear as layers along with their associated maps. The map service provides access to essential museum basemaps (e.g., elevation, EOTR sheets [ie. sheets created in EOV projection], custom UAS rasters, manuscript maps), so the user does not need to download or store the entire map library, only what is required for the designated survey area.

A PostGIS database stores the archaeological and other vector datasets (e.g., the official site-registry layers), which can also be accessed through the field application by on-site archaeologists. Together, these

[1] Jász Museum; e-mail: gulyas.andras.zoltan@jaszmuzeum.hu
[2] Jász Museum; e-mail: hegyesi.viktor@jaszmuzeum.hu

modules ensure that the finds recorded on mobile devices enter the museum's database smoothly and securely. The Jász Museum and FindVault Ltd jointly presented the application within the frame of a workshop in December 2025.

FIELD DATA COLLECTION: ACCURACY AND AUTOMATION

Each code used in the application incorporates both the finder's name and the object's serial number, linking the two immediately. In the field, it is sufficient to record the QR labels with the device camera and select a few basic attributes (type, material, culture) from predefined dropdown menus; all other data (date, GPS coordinates, photo, etc.) are recorded automatically in the user's personal GeoPackage file. This process is fast and error-free, eliminating inaccuracies associated with handwritten field notes.

In addition to the smartphone's built-in GPS module (which yields an accuracy of approx. 5 m), the application can triangulate using GSM towers, a more battery-efficient solution. It can also connect to external RTK (Real-Time Kinematic) GPS modules via Bluetooth, enabling positioning accuracy of 2–3 cm. The resulting coordinate data provided by volunteers therefore meet field-survey precision requirements.

Each find can be retrieved through the QR code: when a code is entered or scanned with a QR reader, the system extracts the associated record and images from the volunteers' GeoPackage files. This significantly accelerates post-fieldwork processing and ensures that metadata (timestamp, finder, description, photograph, coordinates) are always available.

DATA PROCESSING AND STRUCTURING

FindVault's analytical workflow consists of several automated steps. Each process is coded in Python. PyQGIS is especially advantageous, providing direct access to GIS operations while allowing data handling and automated reporting within a unified environment. Thus, large quantities of archaeological data can be processed in a single coherent workflow—from field recording to scientific analysis.

The server stores all volunteer field projects and related files. The system ensures that researchers always work with the most up-to-date field database, automatically updating earlier versions and organising the data: for example, unifying names of collectors, image fields and spatial points. All volunteer data are combined into a single merged GeoPackage file, preparing the dataset for further analysis.

The next step is decoding: the processing module loads a 'period-order' virtual layer and links the numerical codes to historical periods or cultural designations (e.g., code '1' = Roman period; '2' = medieval period). All coded fields in the data are then replaced with their corresponding text labels.

The outcome is a GeoPackage file containing the complete volunteer database with updated period/culture attributes, along with a separate line layer storing the recorded survey routes. These files can now be used directly for scientific analysis.

Spatial context is then added to the dataset. Using the QGIS Python API and spatial intersection operations, each find is assigned to its administrative unit and known archaeological site(s). If a point falls within one or more polygons (e.g., municipal boundary, archaeological site), the corresponding attributes (settlement name, site ID) are written into the record. This creates 'belonging-to' fields for each find, supporting accurate spatial interpretation and contributing valuable information for potential site-revision work.

The processing chain ensures that the field data are transformed into a free-to-research format suitable for scientific processing, ready for use in standard GIS software such as QGIS.

REPORT GENERATION, DAILY LOGS, AND WEATHER INTEGRATION

From the consolidated and processed database, automated daily reports are generated, i.e. the database is synchronised, merged, analysed each day, and a report is created on the results. The system produces a daily work log: a statistical summary and a document detailing the previous day's field-survey results, in PDF, DOCX, XLSX and GPKG formats.

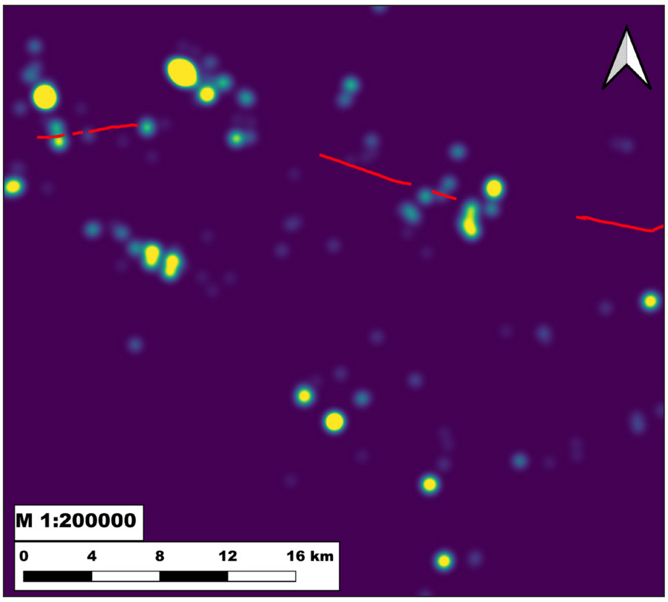

Fig. 1. KDE-based hotspot map depicting the Jászság region based on point data from 1142 Sarmatian era metal and pottery finds collected by the Jász Roosters; georeferenced RGBA raster in EOV projection. The high-intensity spots support the estimation of presumed find concentrations and the extent/activity zones of the sites. The red line marks the explored section of the Kis-árok, which is part of the Csörsz-árok earthwork system stretching between the Danube and the Tisza (considered one of the elements of the Sarmatian-era Limes Sarmatiae)

The report includes the duration and participants of the survey, the distribution of surveyed administrative units and archaeological sites, and the quantity and distribution of finds by historical or archaeological period, type, and material. Each find's photograph appears on a separate page with its corresponding metadata. The PDF field log thus provides an accessible, clear documentation for researchers. Reports are created nightly and can be emailed or uploaded to cloud storage automatically.

The system also supports manual filtering by any attribute (period, settlement, site, object type, material, finder, etc.), producing summary reports using the same format.

A built-in module integrates meteorological data into each daily report. It downloads and decodes NOAA METAR aviation weather telegrams, extracts key parameters (temperature, wind speed, visibility, precipitation, cloud cover, etc.) and presents them in Hungarian. Expected weather conditions during fieldwork (e.g., 'rainy', 'overcast') are included in the narrative section of the report. These meteorological datasets are saved as text files and become part of the daily report package.

Through the reporting module, researchers receive a structured documentation of each day's activities: automated statistical summaries, photographs, metadata and weather observations.

Application of machine-learning and statistical methods (Meyer et al. 2018)

FindVault is not only a data-collection and management tool; it also supports advanced analytical methods by leveraging the expanding dataset. The processing module includes algorithms that help identify patterns and deviations in volunteer-generated data.

One such method is Fisher's exact test, which examines whether the proportion of specific object types found by a volunteer (e.g. coins) significantly differs from the average of the surrounding area. This is useful if a volunteer discovers an unusually large number of a specific find type.

The Isolation Forest algorithm (Guyot et al. 2019), an unsupervised anomaly-detection method, identifies days or users with unusually high or low numbers of finds. These tools help filter noise (e.g., erroneous entries) while highlighting rare and archaeologically significant events (such as indicators of hoards).

Additional spatial-analytical tools are also available. A module under development conducts point-density analysis (hotspot analysis) and cluster detection using DBSCAN (Díaz-Rodríguez & Fábregas-Valcarce 2022). DBSCAN identifies areas where finds cluster at high density—potential new sites (*Fig. 1*).

A Kernel Density Estimation (KDE) module (Beardah & Baxter 1996) produces continuous distribution surfaces, visualising high-intensity concentrations of finds. The software generates multi-page PDF reports and georeferenced raster outputs representing clusters and hotspots.

These statistical and machine-learning approaches provide effective tools for detecting archaeological patterns and anomalies within the dataset.

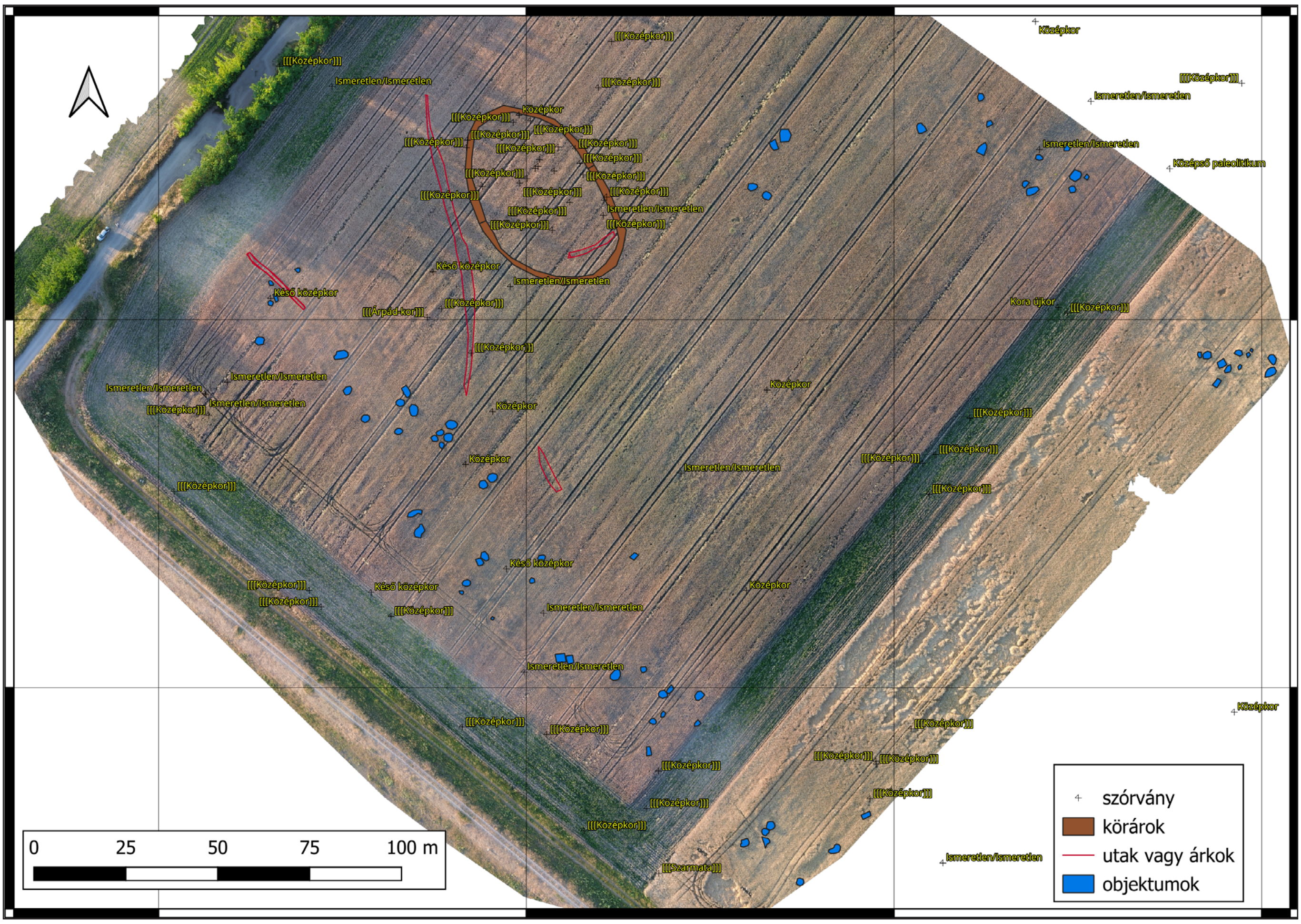

Fig. 2. Orthomosaic created from drone images taken in raking light, with spatial interpolation of the findings and vector marking of the church, circular ditch, and objects. The survey was based on a small but extremely dense concentration of human bones, medieval metal and numismatic finds; the orthophoto confirmed the presence of a medieval church, with most of the finds clustered along the circular ditch

DATA VISUALISATION AND MAPPING

In scientific analysis, visual inspection is crucial. One FindVault module creates high-resolution maps of find locations. Using PyQGIS, the mapping module loads prepared basemaps (e.g., elevation models, orthomosaics) and plots the find-points stored in the database. The resulting map is exported as a PNG image, which can be archived or shared easily. This automated map generation provides a clear spatial overview of the dataset, complementing daily or filtered reports and helping researchers recognise geographic patterns (*Fig. 2*).

THE SCIENTIFIC VALUE OF COMMUNITY-GENERATED DATA

FindVault clearly demonstrates how volunteer-collected data can be transformed into scientifically meaningful information. The system performs multiple validation and enrichment steps: erroneous files are excluded; codes are decoded; finds receive geographic interpretation (settlement, registered site where relevant); and statistical filters help ensure data quality.

Metadata (finder, coordinates, timestamp, photo) are systematically organised into output files that can be readily analysed or visualised in GIS environments. Automation minimises human error.

Overall, FindVault ensures that field-recorded data become scientifically prepared datasets. Attributes are recorded uniformly and precisely; spatial context is assigned algorithmically; and rare events are highlighted through statistical analysis. The resulting reports and maps support publication-ready outputs, making the results accessible both to the professional community and the wider public.

References

Beardah, C. C. & Baxter, M. J. (1996). *The Archaeological Use of Kernel Density Estimates*. Internet Archaeology 1. https://doi.org/10.11141/ia.1.1

Díaz-Rodríguez, M. & Fábregas-Valcarce, R. (2022). Evaluating the effectiveness of three spatial cluster analysis methods on Palaeolithic site location distributions in Galicia, NW Iberian Peninsula. *Journal of Archaeological Science: Reports* 41:103323. https://doi.org/10.1016/j.jasrep.2021.103323

Guyot, A. et al. (2019). Airborne Hyperspectral Imaging for Submerged Archaeological Mapping in Shallow Water Environments. *Remote Sensing* 11(19):2237. https://doi.org/10.3390/rs11192237

Meyer et al. (2018). Meyer, C. et al. 2018. Early Neolithic executions indicated by clustered cranial trauma in the mass grave of Halberstadt. *Nature Communications* 9:2472. https://doi.org/10.1038/s41467-018-04773-w