Satellite Big Data for Archaeological Discovery: A Digital Earth Perspective

Yaroslav Vasyunin*[1] | Ceslav Cieslar[1] | Cristhian Jesus Neyra Kunkel[1,2]
[1] Paititi Research, London, UK
[2] National University of Engineering (UNI), Lima, Peru
yaroslav@paititi.info, ceslav@paititi.info, cristhian@paititi.info

Abstract

The advent of modern big data platforms has revolutionized archaeology, enabling the discovery of new sites on a planetary scale. This article overviews how these platforms, within the Digital Earth framework, process and analyze vast amounts of remote sensing data. We explore the capabilities of major publicly accessible platforms in integrating diverse geospatial datasets and employing advanced analytical tools. Through case studies, we illustrate the transformative impact of these technologies on archaeological research, highlighting notable discoveries and the role of collaboration and crowdsourcing. Despite challenges such as data quality and computational costs, future advancements promise to enhance our ability to uncover and understand historical heritage. This work underscores the significant potential of big data platforms in advancing archaeological discoveries and enriching our knowledge of the past through the Digital Earth initiative.

Keywords

Big Data, Remote Sensing, Digital Earth, Archaeology, Archaeological Sensitivity

Introduction

The identification of new archaeological sites using geospatial data and tools, particularly airborne and satellite Remote Sensing (RS), has a century-long history of successful applications (Luo et al., 2019). Detailed remote sensing data allow the identification of subtle features and patterns left by human terrain transformations that are only visible from above. Additionally, Geographic Information Systems (GIS) have enabled researchers to create archaeological sensitivity maps, which estimate the probability of encountering archaeological sites within landscapes (Caracausi et al., 2018; Howey & Burg, 2017). However, despite these clear benefits, Howey et al. (2020) noted that such techniques have been applied mostly to high-resolution research at local scales, often neglecting broader extents.

In the past decades, the amount, variety, and speed of RS data acquisition, particularly from publicly available programs like Landsat and Copernicus, have dramatically increased (Deng et al., 2019), leading to the problem of RS Big Data (RSBD). As defined by Herndon et al. (2023), RSBD is geospatial data or derived datasets collected via satellite or airborne sensors at an increasingly

* *Corresponding author

larger scale (spatial, temporal, spectral), facilitating the use of new computationally-intensive analytical techniques.

Various solutions for dealing with RSBD were developed (Loukili et al., 2022). These solutions, also known as Geospatial Big Data Platforms (GBDP), could provide archaeologists with new tools to handle spatial datasets and expand their research across cross-border landscapes at multiple scales, even globally.

However, available RSBD combined with computing resources form only the basic infrastructure. To enable easy and efficient data discovery, access, processing, and analysis, user-oriented software and services are also necessary. To navigate the complexities of technological advancements in the geospatial domain, we adopt the Digital Earth (DE) concept, popularized by Al Gore in 1998. This concept integrates various systems, tools, and datasets into a multi-resolution 3D representation of the planet, making it accessible to a wide range of users, from experts to children (Gore, 1998).

In this work, we explore publicly accessible Geospatial Big Data Platforms based on satellite remote sensing data archives that lower the entry barrier for archaeologists, enabling them to access, analyze, and visualize Remote Sensing Big Data for site discovery and research, and discuss their potential development within the context of Digital Earth.

Discussion

Considering the invariant characteristics of DE, formulated by Annoni et al. (2023), as a guiding star, the importance of accessibility and collaborative nature, support for education and knowledge sharing, help us practically narrow down the selection of GBDPs only to those platforms that are aligned with DE. Thus, we define DE-aligned Big Data Platform (DE-GBDP) as a cloud-based system that offers an extensive archive of geospatial data, featuring an accessible interface for users to load, analyze, and visualize this data, and are designed to support a wide range of applications at global scale, and available for free use at least for private, non-commercial and academic purposes. Only a few existing GBDPs satisfy this definition: Google Earth Engine (Gorelick et al., 2017), Sentinel Hub (*Sentinel Hub*, 2023), and, arguably, Microsoft Planetary Computer (Microsoft Open Source et al., 2022). The latter is de-facto a catalog of big geospatial data with a standardized API to access it, but mainly devoted to geospatial experts rather than less experienced users. Altough these platforms are not technologically-neutral, as required by DE characteristics (Annoni et al., 2023), they are the most mature developments of what is currently available to a broad user.

In the DE-GBDPs, petabytes of analysis-ready data are co-located with a high-performance computation services, accessible through the internet Internet via application programming interfaces (APIs) of various complexity and enable rapid prototyping and visualization of results. The simplest APIs are interactive web maps, while others use popular programming languages (e.g., Python). Since these platforms are by definition open to a wide range of people, a community of users has grown around each of them, ready to share and jointly create new algorithms for data analysis and visualization, and to solve technical difficulties together. A massive archive of satellite images, as well as products derived from them, can be combined with information uploaded by users. For most common tasks, there are already developed algorithms that users can utilize. Extensive documentation, tutorials, and a supportive community are available for both platforms.

Google Earth Engine and Sentinel-Hub provide robust visualization tools, enabling users to create interactive maps, graphs, and dashboards. These visualizations help in interpreting analysis results and communicating findings effectively.

Both platforms offer APIs that facilitate integration with other tools and services. This interoperability allows users to incorporate RS data into broader analytical workflows, combining it with other datasets and analytical tools.

As DE-GBDPs lack 3D visualization capabilities, required by DE definition, the information derived from them can be integrated with 3D visualization tools, such as Google Earth (*Earth Versions*, 2023) or ArcGIS Earth (*ArcGIS Earth App*, 2023).

Both platforms leverage cloud computing to handle large-scale data processing efficiently. This scalability is essential for analyzing global datasets and running complex algorithms without the limitations of local computing resources. For example, H. Orengo & Petrie (2017) analyzed a series of more than 1,700 Landsat images within Google Earth Engine to map relict channels and meander scars, demonstrating the complexity of Holocene fluvial history in relation to human settlement. Kalafatić et al. (2020) explored a complex network of densely populated settlements with the analysis of aerial and satellite imagery and geomagnetic survey, where Sentinel Hub was one of the tools.

Both platforms provide comprehensive suites of advanced analytical tools for RS data analysis, including machine learning algorithms, pattern recognition, and anomaly detection. These tools are essential for extracting information from vast amounts of satellite imagery and geospatial data. Thus, Google Earth Engine have been already used by archaeologist in the recent years, as summarized by Herndon et al. (2023) and Agapiou (2017). Herndon et al. describes three ways of the applications of Google Earth Engine in archaeology:

- Archaeological site, feature, and artifact identification
- Cultural heritage site assessment
- Environmental characterization and reconstruction

With respect to Sentinel Hub, there is no much literature available on its use in archaeology. We can speculate that it is due to the fact that Google Earth Engine emerged six years earlier than Sentinel Hub and was initially devoted to the research community, while Sentinel Hub was positioned as a commercial service. Despite this fact, we claim that Sentinel Hub can be used as efficiently as Google Earth Engine in this application.

Orengo et al. (2020) illustrated the potential of machine learning-based classification of multisensor, multitemporal satellite data, implemented through Google Earth Engine, for the remote detection and mapping of archaeological mounded settlements in arid environments, which allowed the examination of a very-large-scale study area of ca. 36,000 km^2. By leveraging such advanced analytical tools available on Google Earth Engine and Sentinel-Hub, researchers can efficiently identify and analyze potential archaeological sites, thereby enhancing the understanding of historical human activities and cultural heritage. However, the use of machine learning to identify sites and features has primarily focused on monumental structures and large sites easily identifiable in topographic relief models (Davis, 2021).

While earth terrain is constantly changing, it does so at a much slower rate than other features on the earth's surface. In this regard, the identification of landform features suspected to be associated with archaeological sites has been proven successful, especially in mountain environments that tend to constrain movements (Visentin et al., 2016). Caracausi et al. (2018) suggests that at high altitudes. due to the same reason, paths and locations suitable for temporary camps did not change significantly during the centuries. Another example is the attempt undertaken by Cieslar & Vasyunin (2023) to discover new potential sites in the poorly studied Peruvian Amazon basin. They apply Google Earth Engine to find relationship between thousands of small modern settlements in the Andes and terrain morphometry, and use the gained information to create an archaeological sensitivity map covering over 3000 km^2.

Challenges

Using big data platforms for archaeological research poses challenges such as data quality, computational costs, and the need for expertise in data analysis. Data quality issues, like sensor errors and atmospheric conditions, require rigorous preprocessing and validation to ensure accuracy. Computational costs can be significant due to the large volumes of data and the intensive processing power needed, necessitating efficient resource management and potential reliance on cloud computing services. Additionally, the complexity of data analysis demands expertise in both archaeology and advanced data analytics, highlighting the need for interdisciplinary collaboration and specialized training to fully harness the capabilities of big data platforms.

Future Directions

Future developments in GBDPs will likely emphasize open ecosystems, facilitating the sharing and integration of diverse datasets from various sources. This approach ensures the viability and evolvability of the DE concept by allowing continuous updates and improvements. Open ecosystems will enable archaeologists to access a broader range of data, including high-resolution satellite imagery, LiDAR scans, and hyperspectral data, thereby increasing the chances of discovering previously unknown archaeological sites. Collaboration between governments, research institutions, and private companies will further enhance data availability and quality, fostering a more comprehensive understanding of the earth's historical and cultural heritage.

Addressing the challenge of modeling archaeological sensitivity maps without relying on high-resolution data requires a multi-faceted approach that leverages diverse data sources, advanced analytical techniques, and cutting-edge technologies.

A user-centered and transparent methodology will be paramount in the future of remote sensing big data platforms. These platforms will need to prioritize usability, ensuring that archaeologists and other researchers can easily navigate and analyze complex datasets. User-friendly interfaces, intuitive data visualization tools, and transparent algorithms will enable non-experts to harness the power of big data for archaeological research. Transparency in data processing and analysis will also enhance the credibility and reproducibility of research findings, fostering greater trust and collaboration within the scientific community.

The strategy for developing digital public goods will drive the creation of open-access databases and tools that promote a sustainable and fair society. In archaeology, this could manifest as publicly available repositories of satellite imagery and other remote sensing data, as well as open-source software for data analysis. Such resources will democratize access to advanced technologies, enabling researchers from around the world to contribute to and benefit from archaeological discoveries. This democratization will not only enhance research capabilities but also ensure that the benefits of technological advancements are equitably distributed.

An interoperability framework will be crucial for connecting and utilizing digital technologies and resources related to the earth's surface and subsurface. Future RS platforms will need to adhere to standardized protocols and formats, allowing seamless integration of diverse datasets. This interoperability will enable researchers to combine data from multiple sources, such as ground-penetrating radar, aerial surveys, and satellite imagery, to gain a more comprehensive understanding of archaeological sites. Enhanced interoperability will also facilitate the integration of data from different time periods, providing insights into the temporal dynamics of historical and cultural landscapes.

The integration of advanced technologies such as artificial intelligence (AI) and quantum computing holds significant potential for enhancing the discovery process in archaeology.

Artificial Intelligence algorithms can analyze vast amounts of RS data more efficiently than traditional methods. These algorithms can identify patterns and anomalies that may indicate the presence of archaeological sites, automating the initial stages of site detection. AI can also assist in predictive modeling, helping researchers prioritize areas for further investigation based on historical and environmental factors.

Quantum Computing has the potential to revolutionize data processing by performing complex calculations at unprecedented speeds. In the context of remote sensing and archaeology, quantum computers could rapidly analyze massive datasets, identifying subtle correlations and patterns that classical computers might miss. This capability could significantly accelerate the discovery of new sites and enhance our understanding of known ones.

Conclusions

Big data platforms are revolutionizing the field of archaeology, providing unprecedented opportunities to uncover and understand our shared human history. By leveraging the power of these platforms, archaeologists are making groundbreaking discoveries, pushing the boundaries of our knowledge about past societies and their interactions with the environment. As technology continues to evolve and our ability to harness big data grows, we can expect even more astonishing discoveries in the years to come, further enriching our understanding of the past and shaping our interpretations of the present.

Future developments in remote sensing big data platforms, guided by the invariant characteristics of the Digital Earth concept, will profoundly impact archaeological research. The adoption of open ecosystems, user-centered methodologies, digital public goods, and interoperability frameworks will enhance data accessibility, usability, and integration. The incorporation of advanced technologies such as AI and quantum computing will further streamline the discovery process, enabling researchers to uncover and analyze archaeological sites with greater efficiency and accuracy. These advancements will not only advance our knowledge of the past but also promote a more inclusive and equitable approach to archaeological research, aligning with the broader goals of the Digital Earth initiative.

References

Agapiou, A. (2017). Remote sensing heritage in a petabyte-scale: Satellite data and heritage Earth Engine© applications. *International Journal of Digital Earth*, *10*(1), 85–102. https://doi.org/10.1080/17538947.2016.1250829

Annoni, A., Nativi, S., Çöltekin, A., Desha, C., Eremchenko, E., Gevaert, C. M., Giuliani, G., Chen, M., Perez-Mora, L., Strobl, J., & Tumampos, S. (2023). Digital Earth: Yesterday, today, and tomorrow. *International Journal of Digital Earth*, *16*(1), 1022–1072. https://doi.org/10.1080/17538947.2023.2187467

ArcGIS Earth App. (2023). [Software]. ESRI. https://www.esri.com/en-us/arcgis/products/arcgis-earth/overview

Caracausi, S., Berruti, G. L. F., Daffara, S., Bertè, D., & Rubat Borel, F. (2018). Use of a GIS predictive model for the identification of high altitude prehistoric human frequentations. Results of the Sessera valley project (Piedmont, Italy). *Quaternary International*, *490*(November 2017), 10–20. https://doi.org/10.1016/j.quaint.2018.05.038

Cieslar, C., & Vasyunin, Y. (2023). *Identification of archeological sites in the Peruvian Amazon using satellite remote sensing*. Unpublished. https://doi.org/10.13140/RG.2.2.11555.43048

Davis, D. (2021). Theoretical Repositioning of Automated Remote Sensing Archaeology: Shifting from Features to Ephemeral Landscapes. *Journal of Computer Applications in Archaeology*, *4*(1), 94. https://doi.org/10.5334/jcaa.72

Deng, X., Liu, P., Liu, X., Wang, R., Zhang, Y., He, J., & Yao, Y. (2019). Geospatial Big Data: New Paradigm of Remote Sensing Applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *12*(10), 3841–3851. https://doi.org/10.1109/JSTARS.2019.2944952

Earth Versions. (2023). [Software]. Google Earth. https://www.google.com/intl/ru/earth/about/versions/

Gore, A. (1998). The Digital Earth: Understanding our planet in the 21st Century. *Australian Surveyor*, *43*(2), 89–91. https://doi.org/10.1080/00050348.1998.10558728

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Herndon, K. E., Griffin, R., Schroder, W., Murtha, T., Golden, C., Contreras, D. A., Cherrington, E., Wang, L., Bazarsky, A., Kollias, G. V., & Alcover Firpi, O. (2023). Google Earth Engine for archaeologists: An updated look at the progress and promise of remotely sensed big data. *Journal of Archaeological Science: Reports*, *50*, 104094. https://doi.org/10.1016/j.jasrep.2023.104094

Howey, M. C. L., & Burg, M. B. (2017). Assessing the state of archaeological GIS research: Unbinding analyses of past landscapes. *Journal of Archaeological Science*, *84*, 1–9. https://doi.org/10.1016/j.jas.2017.05.002

Howey, M. C. L., Sullivan, F. B., Burg, M. B., & Palace, M. W. (2020). Remotely Sensed Big Data and Iterative Approaches to Cultural Feature Detection and Past Landscape Process Analysis. *Journal of Field Archaeology*, *45*(sup1), S27–S38.

https://doi.org/10.1080/00934690.2020.1713435

Kalafatić, H., Klindžić, R. Š., & Šiljeg, B. (2020). Being Enclosed as a Lifestyle: Complex Neolithic Settlements of Eastern Croatia Re-Evaluated through Aerial and Magnetic Survey. *Geosciences*, *10*(10), 384. https://doi.org/10.3390/geosciences10100384

Loukili, Y., Lakhrissi, Y., & Ali, S. E. B. (2022). Geospatial Big Data Platforms: A Comprehensive Review. *KN - Journal of Cartography and Geographic Information*, *72*(4), 293–308. https://doi.org/10.1007/s42489-022-00121-7

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., & 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*(March), 111280. https://doi.org/10.1016/j.rse.2019.111280

Microsoft Open Source, McFarland, M., Emanuele, R., Morris, D., & Augspurger, T. (2022). *microsoft/PlanetaryComputer: October 2022* (2022.10.28) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.7261897

Orengo, H. A., Conesa, F. C., Garcia-Molsosa, A., Lobo, A., Green, A. S., Madella, M., & Petrie, C. A. (2020). Automated detection of archaeological mounds using machine-learning classification of multisensor and multitemporal satellite data. *Proceedings of the National Academy of Sciences*, *117*(31), 18240–18250. https://doi.org/10.1073/pnas.2005583117

Orengo, H., & Petrie, C. (2017). Large-Scale, Multi-Temporal Remote Sensing of Palaeo-River Networks: A Case Study from Northwest India and its Implications for the Indus Civilisation. *Remote Sensing*, *9*(7), 735. https://doi.org/10.3390/rs9070735

Sentinel Hub. (2024). Sinergise Solutions d.o.o., a Planet Labs Company. https://www.sentinel-hub.com/

Visentin, D., Carrer, F., Fontana, F., Cavulli, F., Cesco Frare, P., Mondini, C., & Pedrotti, A. (2016). Prehistoric landscapes of the Dolomites: Survey data from the highland territory of Cadore (Belluno Dolomites, Northern Italy). *Quaternary International*, *402*, 5–14. https://doi.org/10.1016/j.quaint.2015.10.080