A NEW POSSIBILITY TO WORK WITH OPEN GEOSPATIAL DATA: WHEROBOTS CLOUD, WHEROBOTSDB AND WHEROBOTS AI

Ana Cornelia BADEA, Prof.PhD.Eng., Faculty of Geodesy, Technical University of Civil Engineering Bucharest, Romania, ana.badea@utcb.ro
Gheorghe BADEA, Prof.PhD.Eng., Faculty of Geodesy, Technical University of Civil Engineering Bucharest, Romania, gheorghe.badea@utcb.ro

Abstract *In this article we describe a new possibility to access open geospatial data in the cloud, Wherobots. One of the advantages is that the data is not stored on the user's own computer, nor is the workflow performed on the user's own device. AI technology is also integrated into this solution. By accessing the Wherobots Cloud, WherobotsDB and Wherobots AI can be utilized. The Wherobots programming languages are Python, Scala, Java and WherobotsDB supports the SparkSQL query optimizer. There is also a map matching feature applicable when the GNSS device provides GPS coordinates that may be inaccurate and do not necessarily align with geospatial features in the field (e.g. road networks).*

Keywords: *open data; Wherobots; AI; cloud; geospatial; spatial intelligence cloud; GPS; GIS; GNSS*

1. Introduction

Wherobots Cloud is a spatial intelligence cloud platform designed to make it significantly easier and more efficient to work with geospatial data at a planetary scale, bridging the gap between complex geospatial data and actionable insights.

The study started from general to particular, with a survey of articles on the “spatial intelligence cloud” and related topics in WoS, gathering 318 results. (figure 1)

Spatial Data Science (SDS) is a subset of Data Science that focuses on the special characteristics of spatial data, using modeling to know where and why things happen.[1]

The adaptable use of external visual-spatial representations is a crucial aspect of spatial intelligence. Medical professionals, including surgeons and radiologists, utilize imaging technologies such as ultrasound and MRI. Interactive computer visualizations also play a significant role in medical education, particularly in anatomy. [2]

The spatial intelligence cloud represents a paradigm shift in geospatial analysis, wherein the scalability and computational power of cloud computing are harnessed to manage, analyze, and derive insights from vast amounts of location-based data. This concept integrates core geospatial functionalities with advanced artificial intelligence and machine learning within cloud environments.

Several use cases in which the spatial intelligence cloud can help to make more informed decisions based on location data are the following:

- Smart Cities - Cloud-based spatial intelligence platforms can be used to monitor and manage urban infrastructure, traffic flow, and public safety. Real-time data from sensors and cameras can be analyzed to optimize city services.
- Precision Agriculture - Satellite imagery and sensor data can be analyzed to optimize crop yields and resource management. Cloud-based spatial intelligence platforms can provide farmers with real-time insights into crop health and soil conditions.

- Disaster Management - Spatial data can be used to assess damage, plan evacuation routes, and coordinate relief efforts. Cloud-based platforms enable rapid access to critical information during emergencies.

The relationship between a spatial intelligence cloud and Geographic Information Systems (GIS) is one of evolution and enhancement. The spatial intelligence cloud builds upon and extends the capabilities of traditional GIS, because GIS provides the foundational tools and techniques for working with spatial data.

Traditional GIS is primarily focused on the capture, storage, analysis, and display of geospatial data. It relies heavily on desktop software and local data storage and analysis often involves manual processes and specialized GIS expertise. It has limitations, like the struggling with processing and analyzing very large datasets.

Spatial intelligence cloud represents the next generation of GIS, leveraging the power of cloud computing, enabling scalable processing, storage, and analysis of massive geospatial datasets, facilitating data sharing and collaboration. It integrates AI and machine learning to automate tasks and extract deeper insights, automating tasks like image analysis, feature extraction, and predictive modeling.

Cloud platforms facilitate the integration of GIS with other data sources and applications, such as IoT devices and big data analytics and can be accessed from any device with an internet connection, improving accessibility and collaboration.

Therefore, the spatial intelligence cloud is not a replacement for GIS, but rather an evolution that expands its capabilities and applications.

The trends can be identified in figure 2. The domains in which articles were published in WoS were identified using [3]. (figure 3)

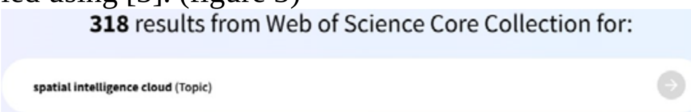

Figure 1 – Gathering results using the topic “spatial intelligence cloud”

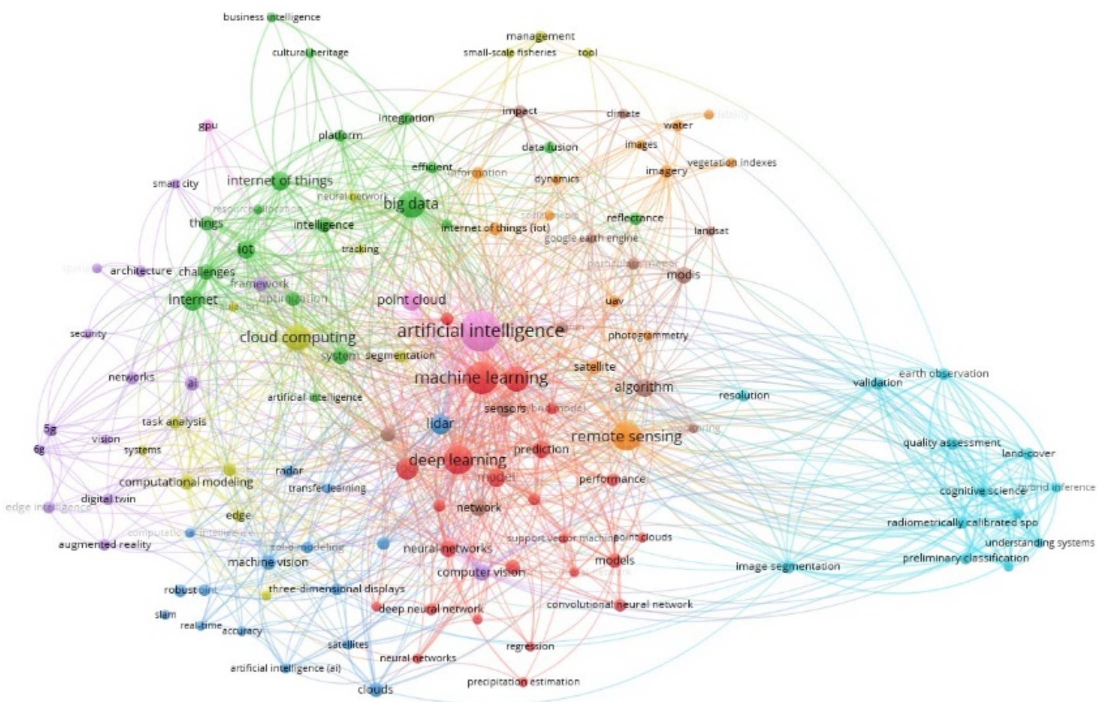

Figure 2 - Identifying trends and thematic coverage

2. Wherobots Cloud, WherobotsDB, Wherobots AI, Wherobots Spatial Catalog

WherobotsDB is a cloud-native analytics engine optimized for geospatial workloads, having the possibility to automate spatial data processing and to extract insights from aerial imagery. It is using Job Runs to schedule and automate data processing workloads. To start using WherobotsDB it is needed to launch a SedonaContext in Wherobots cloud.

WherobotsDB empowers users to work with vector and raster data through a range of SQL APIs. It aligns with the SQL/MM Part3 Spatial standard for vector data and provides familiar Apache Sedona public APIs. To ease the learning curve for Sedona users, the documentation maintains the "Sedona" nomenclature. Raster data is used for terrain, elevation, and satellite imagery and Havasu supports to store and query raster data.

Map matching function is very useful and it is appliable when GNSS device provides GPS coordinates that might be imprecise and don't necessarily align with geospatial elements from the field (eg. road networks). This discrepancy can be solved using map matching, aligning the potentially inaccurate GPS points to their correct road segments, in order to ensure the accuracy of navigation routes and improving the navigation.

Wherobots AI is specifically designed to bring artificial intelligence capabilities to geospatial data, particularly imagery.

On December 17, 2024, in San Francisco, Wherobots, the Spatial Intelligence Cloud powered by the original Apache Sedona team, announced the general availability of Raster Inference for WherobotsAI.

The workflow and applicability can be seen in figure 3.

Figure 3 – Wherobots approach [4]

This approach provides SQL and Python developers with instant analytical access to satellite and drone imagery, offering pay-as-you-go flexibility and custom model integration to reduce development time and cost.

It leverages cloud computing to provide scalable and on-demand access to processing and analysis capabilities, eliminating the need for users to manage complex infrastructure.

Tutorials are available that greatly help users and guide them step by step in raster and vector analysis. [5]

Figure 4 highlights the types of data and their relationship to the traditional and cloud-optimized data formats. WherobotsAI Raster Inference simplifies and reduces the cost of building satellite and drone image analysis solutions.

It offers instant, infrastructure-free inference pipelines, flexible model integration, a high-performance spatial data environment, automated workflows, and enhanced model portability via the MLM STAC standard. [6] Example notebooks facilitate quick understanding.

WherobotsAI provides a comprehensive toolkit for geospatial analysis, pointed out in figure 5.

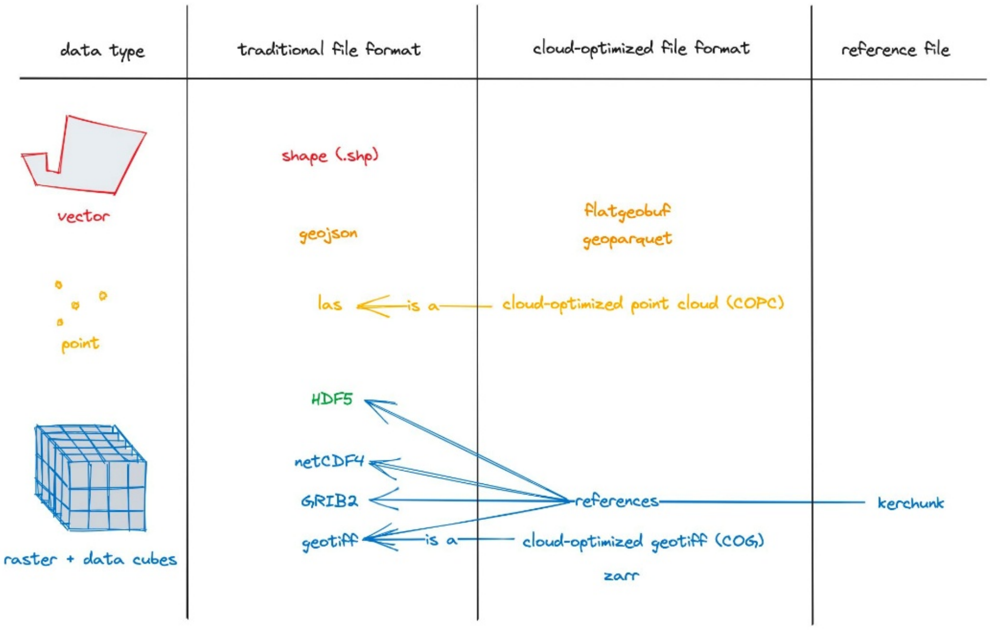

Figure 4 – The main approach in data type and file format [7]

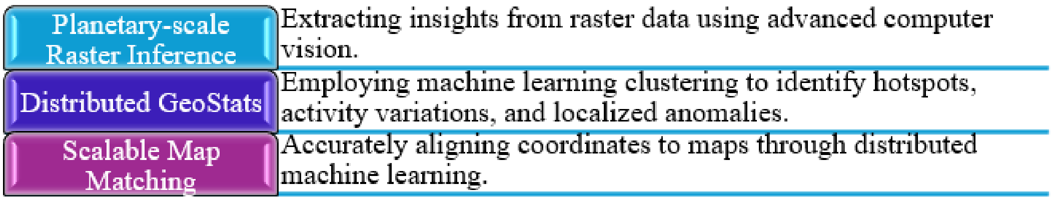

Figure 5 – WherobotsAI Toolkit

Wherobots Spatial Catalog is a tool to find and manage spatial datasets. (figure 6) Wherobots integrates with and utilizes open data, recognizing its importance in geospatial analysis.

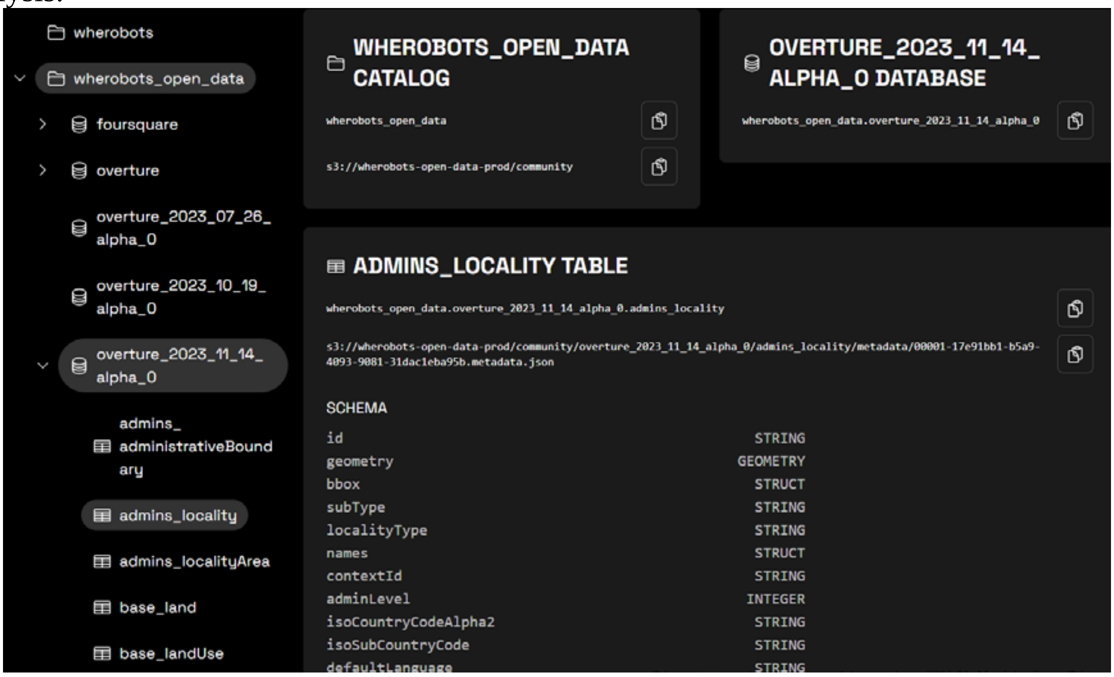

Figure 6 - Retriving open data using Wherobots platform

Wherobots collects and maintains open datasets from various sources. These datasets are then cleaned and transformed into their "Havasu" format, designed to optimize them for efficient analytics within the Wherobots Cloud environment, particularly with WherobotsDB. Wherobots provides these open datasets to its users, having the possibility to leverage publicly available geospatial data to enhance their analyses within the Wherobots platform. (figure 7)

There can be accessed the following: OpenStreetMap Data - all the nodes of the OpenStreetMap Planet dataset, which includes data from European countries; Overture Maps Data (Community Edition): Buildings as millions of polygons representing any human-made structures with roofs or interior spaces; Places as millions of points representing any business or point of interest within the world; Administrative Boundaries: LineStrings representing any officially defined borders between two administrative localities; Localities: points representing countries and hierarchical subdivisions of countries; Transportation Connectors as millions of points representing points of physical connection between two locations. (figure 7)

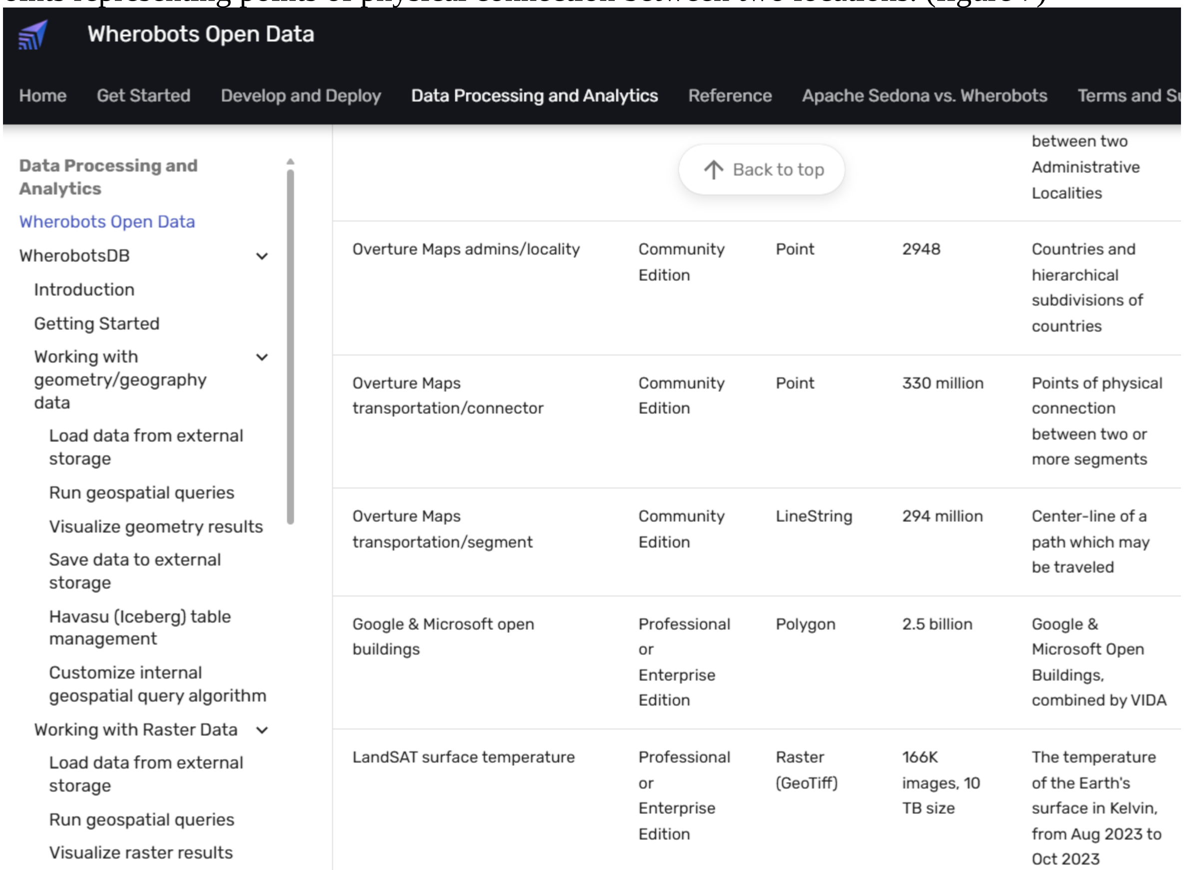

Figure 7 – Examples of open data [9]

The examples emphasized in the analysis tab include clustering DBSCAN, Getis_Ord_Gi*, K_Nearest_Neighbor_Join, Local_Outlier_Factor, Object_Detection, Raster_Classification and Raster_Segmentation. [8]

There is a study utilizing the Vehicle Energy Dataset (VED), which contains GPS tracking data from vehicles in Michigan, collected between November 2017 and November 2018. The dataset covers approximately 374,000 miles, encompassing a variety of driving conditions and vehicle types (gasoline, HEV, PHEV/EV). [10]

The workflow is highlighted in figure 8.

This example and numerous examples from the literature are proving that the spatial intelligence cloud approach offers nowadays extended possibilities to access and analyze geospatial data reducing both the time and human effort. [11]

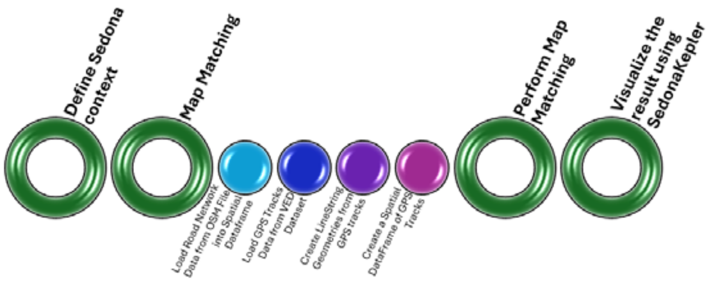

Figure 8 – Study using VED and GPS tracking data (adapted from [8], [10])

3. Conclusions

The spatial intelligence cloud combines powerful cloud computing resources with cutting-edge geospatial analytics and AI/ML capabilities, enabling organizations to gain deeper insights from location data. These capabilities are offered by vendors like Wherobots and are enabled by the cloud infrastructure of companies like Amazon, Google, and Microsoft. By providing pre-processed open data, Wherobots lowers the barrier to entry for geospatial analysis, because users can integrate these datasets into their workflows.

Here it is worth mentioning that cloud-native platforms and desktop GIS serve distinct but complementary roles. GIS software is optimal for geometry editing, static map generation, and coordinate/format transformations. However, when there is starting to experience slow rendering, long processing times, or need to connect the GIS work with data that lives in the cloud, then using cloud based solutions.

4. References

1. *https://carto.com/what-is-spatial-data-science*
2. *Hegarty,M. (2010). Chapter 7 - Components of Spatial Intelligence, Psychology of Learning and Motivation, Academic Press, Volume 52, Pages 265-297, ISSN 0079-7421, ISBN 9780123809087, https://doi.org/10.1016/S0079-7421(10)52007-3*
3. *https://www.webofscience.com/*
4. *https://wherobots.com/*
5. *https://docs.wherobots.com/latest/tutorials/wherobotsdb/set-up-wherobotsdb/*
6. *https://www.geoweeknews.com/news/introducing-wherobots-raster-inference-to-unleash-innovation-with-planetary-imagery*
7. *https://www.linkedin.com/posts/mbforr_3-things-you-should-know-about-cloud-native-activity-7305949916357361666-MFds?utm_source=share&utm_medium=member_android&rcm=ACoAAAhQTysBqH104uT4ri4xRzmSxcFzCOX95e8*
8. *https://github.com/wherobots/wherobots-examples/blob/main/python/Data/ESA_WorldCover.ipynb*
9. *https://docs.wherobots.com/latest/tutorials/opendata/introduction/#inspecting-open-data-catalogs*
10. *Geunseob (GS) Oh, David J. LeBlanc, Huei Peng. (2020). Vehicle Energy Dataset (VED), A Large-scale Dataset for Vehicle Energy Consumption Research, Published in IEEE Transactions on Intelligent Transportation Systems (T-ITS), https://github.com/gsoh/VED*
11. *Dorobanţu, G., Badea, A. C. (2024). Web GIS and Cloud Computing for Defense Decision Making, RevCAD 37/2024 http://revcad.uab.ro/upload/58_938_05_dorobantu_badea.pdf*