GeoLabel: WebGIS Tool for Geospatial Data Annotation

Teja Kattenborn[1], Janusch Jehle[1], Jan Herrmann, Anna Göritz[1]
[1]Department for Sensor-based Geoinformatics, University of Freiburg, Germany.
Date of submission: 21.06.2024

Abstract

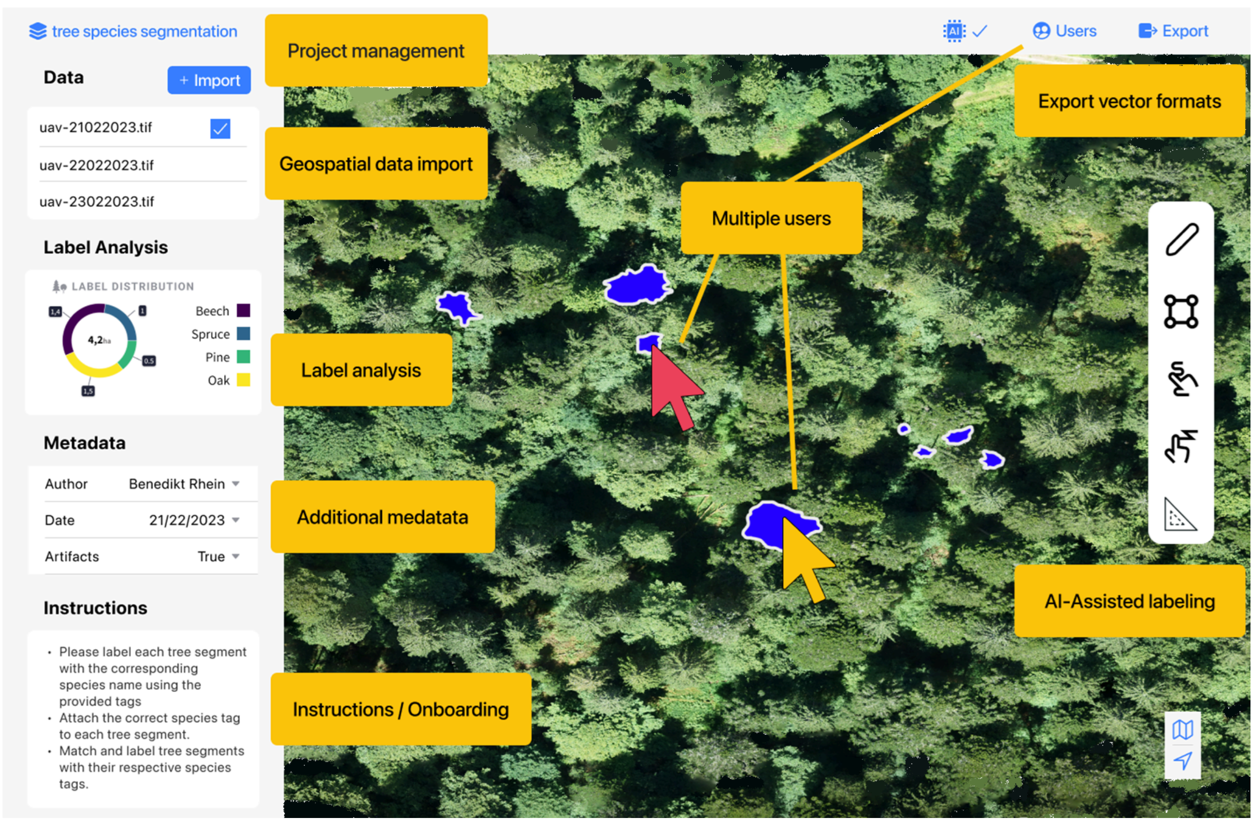
In Earth System Science (ESS), the generation and curation of geospatial data labels are crucial for training AI models. However, existing annotation tools are inadequate for handling large-scale geospatial data due to incompatibility with common raster formats like Zarr and GeoTIFF, lack of user-friendly and collaborative features, and limited export capabilities for AI-ready datasets. GeoLabel will develop a WebGIS-based annotation tool that supports georeferenced labeling, enhancing the efficiency and accuracy of AI model training. GeoLabel will provide high accessibility, project and user management, standardized metadata, crowd-sourcing compatibility, geospatial data import, annotation tools, AI-assisted labeling, label analysis, and export capabilities. The prospective users include researchers, data curators, citizen scientists, and public authorities, among others. This pilot will significantly advance data analytics, research data management, and the broader NFDI4Earth efforts by facilitating efficient and standardized geospatial data labeling workflows.

I. Introduction

AI methods revolutionize geospatial data analytics, but data sparsity limits their full potential (Camps-Valls et al. 2021). A critical challenge is generating and curating geospatial data labels for AI models, such as those for object detection, semantic segmentation, instance segmentation or dense regression (Kattenborn et al. 2021). **Current annotation tools** (e.g. CVAT, Labelme, MONAI, Label Studio) **are inadequate for geospatial data**, as they do not support geospatial data formats like Zarr or GeoTIFF and struggle with large datasets. These tools commonly create labels for small, non-geocoded subsets, limiting flexibility and reusability for tasks like changing model tensor sizes, combining data modalities, or assessing spatial context (Kattenborn et al. 2022). Consequently, researchers in Earth System Science (ESS) often rely on general GIS software (ArcGIS, QGIS, Google Earth Engine) that lacks optimization for collaborative and efficient data curation and labeling. The NFDI4Earth pilot GeoComBi offers a web-based geospatial context for image labeling but has limitations. As is part of the BIIGLE ecosystem, customization is very limited, it requires existing WMS services, does not support direct import of custom data (e.g., Zarr), and does not allow exporting annotations in common geospatial formats (e.g., [Shapefile](), [GeoPackage]()) or as AI-ready datasets. In Summary: Existing tools suffer from user unfriendliness, low accessibility, limited collaboration functionality, lack of standards, compatibility issues with AI frameworks, and non-reproducibility due to the absence of geospatial context. **Our vision is to streamline data curation and labeling for the ESS community** by providing an integrated tool that supports georeferenced labeling.

II. Pilot description

Our proposed solution is a WebGIS-based annotation tool designed to streamline and enhance the process of georeferenced labeling for geospatial data. This tool can be accessed from common browsers and **hosted locally** or **remotely**.



GeoLabel will include the following features: **High accessibility,** allowing access to all functionality from a browser without local software installation and independence of licenses or operating systems; **project management**, enabling the creation and management of multiple projects; **User management and crowd-sourcing functionalities,** with different roles and privileges for flagging, editing, deleting, creating or comparing annotations; **metadata standardization**, which generates project metadata and documentation, including machine-readable annotations and descriptions; **Geospatial data import**, enabling the integration of geospatial raster data ([Zarr](), [GeoTIFF]()) and interactive visualization on a map; **annotation tools**, for identifying instances or entities (e.g., points, bounding boxes, segments); **AI-assisted labeling**, like Geo-SAM and pre-calculated feature maps to facilitate efficient label generation; **label analysis**, to analyze label distribution and statistics; and **export capabilities**, allowing the export of labels in georeferenced vector formats with direct compatibility with state-of-the-art ML frameworks like TorchGeo. *GeoLabel* is designed to be user-friendly and adhere to FAIR standards, ensuring annotated data is findable, accessible, interoperable, and reusable.

Technological backbone: *GeoLabel* leverages a robust stack of open-source software and modern web technologies to ensure high performance, scalability, and ease of use, deployable as a ready-to-use Docker container for consistent operation across environments. The front end, built with React, offers a dynamic user experience, utilizing Mapbox GL JS for geospatial visualization and mapbox-draw for geometry editing, with OpenStreetMap as the base map and Geotiff.js for handling Cloud Optimized GeoTIFFs (COGs). By converting GeoTIFFs or Zarr to COGs, which include built-in tiling, overviews, and compression, we can render large data formats directly in the browser without needing external web mapping services like GeoServer or MapServer, enabling highly scalable client-side rendering. For advanced geospatial data processing, we use GDAL and Python for server-side tasks, with a self-hosted local Supabase instance managing our real-time PostgreSQL database, user authentication, and storage. PostgreSQL serves as a backend database, storing georeferenced labels and metadata in a vector data format. An integrated and optional AI-assisted annotation (Geo-SAM, segment-geospatial) will significantly enhance labeling efficiency and accuracy for many tasks.

Standards and interoperability: The *GeoLabel* pilot employs ISO 19115 for metadata standards and adheres to OGC standards for geospatial data interoperability, utilizing formats such as GeoTIFF, Zarr, Cloud Optimized GeoTIFF (COGs), and widely used vector formats like GeoJSON, Shapefile, and GeoPackage. *GeoLabel* will be publicly available as open source, licensed under the MIT License, and documented on GitHub, with a Docker image available on Docker Hub.

III. Relevance for the NFDI4Earth

GeoLabel will support ESS communities across domains and data modalities, including terrestrial, marine, and planetary data from various sensors across scales. *GeoLabel* primarily targets task area 2 M2.5 of NFDI4Earth (*Advancing Tools*) and addresses different steps of the research data life cycle by facilitating Data Collection (georeferenced labeling), Data Analysis (manual and AI-assisted annotation), Data Preservation (storing annotations with standardized metadata), and Data Sharing (exporting AI-ready datasets and enabling collaboration through web-based access under FAIR principles). With regard to the thematic tracks of the NFDI4Earth 3rd pilot phase, *GeoLabel* is, hence, primarily a *Data Curation Pilot* with additional overlap to the tracks *FAIR Big Data Pilot* and *Data Publication Pilot*.

IV. Deliverables

GeoLabel Docker Container: A fully deployable Docker container of *GeoLabel*, encompassing all functionalities, ensuring ease of installation and seamless deployment across environments.
Comprehensive Documentation: Detailed documentation covering every aspect of *GeoLabel*, including step-by-step installation guides, thorough usage instructions, and troubleshooting tips to facilitate user adoption and efficient operation.
Demo Video: A high-quality demo video showcasing the practical applications and benefits of *GeoLabel*, highlighting its functionality across ESS domains and use cases.
Application case: Implement *GeoLabel* within the deadtrees.earth platform to showcase large-scale, crowd-sourced image labeling. This integration will serve as a model to attract additional application cases throughout the project's duration.
Interactive EGU 2025 Demonstration: An engaging and interactive presentation of *GeoLabel* at the EGU 2025 conference (Vienna), aimed at demonstrating its value to the geoscience community and encouraging widespread adoption.

V. Work plan & requested funding

Work package / Project month	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
WP1 - GeoLabel Front & Backend												
WP2 - User management & crowd integration												
WP3 - Export & metadata routines												
WP4 - Application cases												
WP5 - Outreach (Video, Conference)												

Requested funding: We are applying for a total of €94,500. This includes funding for 12 person-months (PM) of scientific personnel (TVL-E13; €85,296), student assistance (5 hours per week; €4,848), and travel expenses (€4,356)
In-kind contributions of the University of Freiburg: Additional student assistance 5h / week; video production and editing; travel & conference fees for a live demo at EGU 2025.

References: Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 173, 24-49. Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M. D., & Dormann, C. F. (2022). Spatially autocorrelated training and validation samples inflate performance assessment of convolutional neural networks. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 5, 100018. Camps-Valls, G., Tuia, D., Zhu, X. X., & Reichstein, M. (Eds.). (2021). Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science, and geosciences. *John Wiley & Sons*.