# MaRESS: Mapping Research in Earth System Sciences

A Modular Web Application for Literature Analysis and Geographic Data Mapping in Earth System Scienes

## Benjamin Schmidt<sup>1</sup>, Marco Otto<sup>1</sup>

<sup>1</sup>Chair of Climatology, Institute of Ecology, Technische Universität Berlin Rothenburgstraße 12, 12165 Berlin, Germany

benjamin.schmidt@tu-berlin.de, marco.otto@tu-berlin.de

#### Project Overview

In the process of conducting systematic literature reviews in Earth System Sciences, researchers often face challenges in efficiently geolocating study sites. While study sites are frequently mentioned in the text, they are rarely included in machine-readable formats or metadata. This gap hinders the ability to perform comprehensive spatial analyses and identify research trends.

MaRESS (Mapping Research in Earth System Sciences) is a modular web application designed to address these critical challenges in Earth System Sciences research data management and analysis.

#### Core Objectives:

- Automated extraction and mapping of research data from scientific literature
- Identification of geographic knowledge voids, thematic research gaps and citation network
- Extension of bibliographic metadata with spatial analysis capabilities
- AI-assisted categorization and semantic mapping of scientific publications

Target Application: Initial deployment will focuses on the expansive test dataset on High Andean Wetlands research, supporting comprehensive literature review and ecosystem service analysis

#### Modular System Design

**MaRESS** is structured in frontend/backend architecture, allowing usage as a standalone web application or integration into existing research data infrastructures.

Each part is separated technically and functionally into four main modules (Figure 2):

#### Module 1: Geographic Mapping

- Interactive clustered visualization using OpenLayers for Point data
- Coordinate extraction from research papers
- Spatial clustering analysis of study locations

#### Module 2: Semantic Mapping

- Zotero API integration for bibliographic management
- Citation/Author network analysis using graph algorithms
- Keyword extraction and clustering

#### Module 3: Research Data Mapping

- Integration with PANGAEA, Zenodo APIs
- DataCite DOI metadata extraction
- Automated dataset-publication linkage

#### Module 4: AI-Assistant

• Third-party API usage for LLM document categorization

# Implementation Workflow

#### Data Processing Pipeline:

#### 1. Authentication:

Current: Email registration with Bearer Token
Planned: OAuth2 integration with ORCiD user accounts

#### 2. Metadata Retrieval:

Automated download of bibliographic records and PDFs via Zotero  $Web\ API$ 

- 3. **NLP Analysis:** spaCy pipeline processes full-text content for:
- Geographic entity recognition (GPE, LOC tags)
- Coordinate pattern extraction using regex
- Sentence-level geolocation context analysis
- 4. **Spatial Analysis:** Algorithms implemented for:
- Coordinate validation and projection
- Clustering algorithms (DBSCAN) for study site grouping
- 5. **Visualization:** Real-time rendering of:
  - Interactive maps with study location markers
- Citation network graphs with node clustering
- Metadata dashboards with filtering options

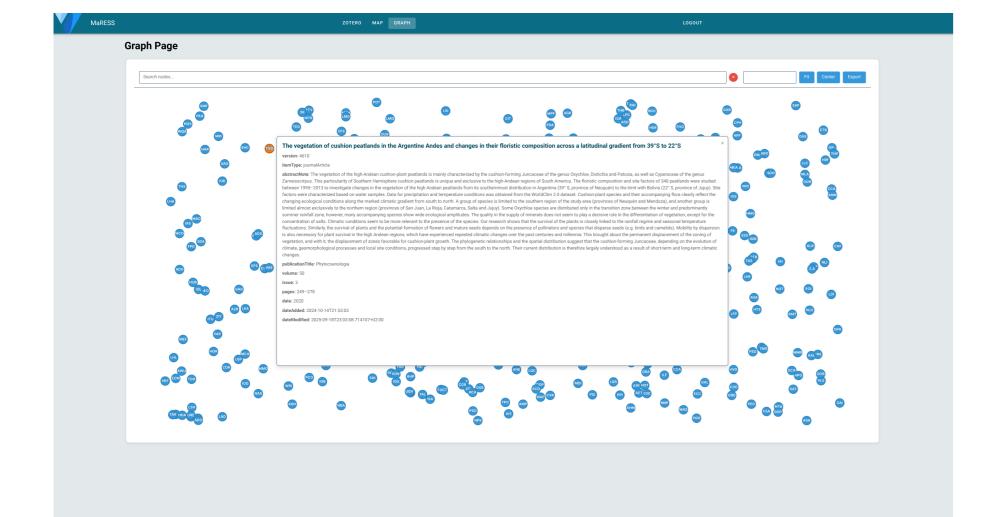


Figure 1: Interactive citation network and keyword visualization using Cytoscape.js with an overlay of metadata information of selected paper

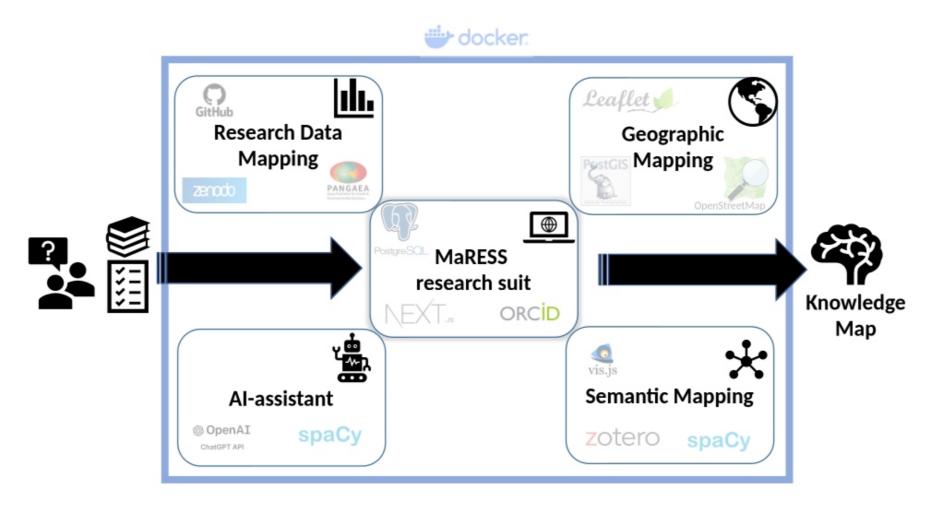


Figure 2: MaRESS modular architecture integrating four core modules

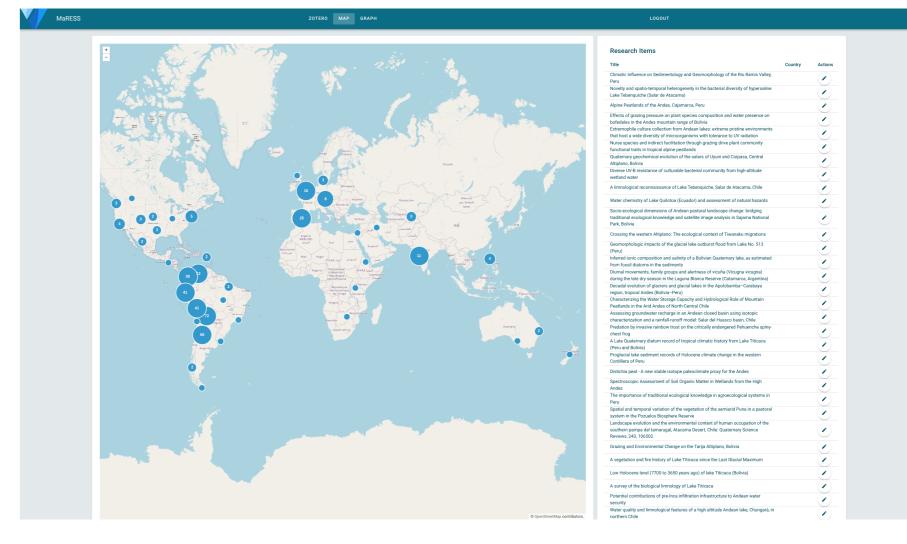


Figure 4: Interactive map view displaying clustered research study locations extracted from Zotero library via Zotero Web API integration

## Technical Architecture

## Backend Infrastructure:

- Python-based with asynchronous request handling
- Relational database with geoinformational extension for spatial data
- NLP model pipeline for Named and Custom Entity recognition
- Zotero Web API, DataCite DOI resolution, Zenodo integration

## Frontend Components:

- Modern, reactive JS/TS-based Browser application
- Interactive geographic visualization with OpenStreetMap and clustering capabilities
- Dynamic data tables for metadata filtering and analysis
- Cytoscape.js for keyword and author network visualization

## Compliance & Standards:

- Improves FAIR data principles (Findability, Accessibility)
- Open-source licensing (MIT/Apache 2.0)
- RESTful API design with OpenAPI 3.1 specification

## Deployment:

Docker containerization ensuring reproducible environments and crossplatform compatibility

Self-hosted or server deployment options

		ZOTERO M	AP GRAPH				.ogout					
1.7												
Libra	<u> </u>											
Items	LIBRARY											
	Title	Latitude	Logitude	Abstract	Authors	Journal	Date	ISSN	Catalog	Archive	Series	DOI URL
	Climatic Influence on Sedimentology and Geomorphology of the Rio Ramis Valley, Peru	-6.8700	-6.8700	Fluctuations in regional precipitation and in base-level (the level of Lake Titicaca) trigger changes in fluvial erosion and deposition in the largest tributary valley of Lake Titicaca, the Rio Ramis, Peru. Relationships between fluvial sedimentation and		Journal of Sedimentary Research	2005					Ø
•	Novelty and spatio-temporal heterogeneity in the bacterial diversity of hypersaline Lake Tebenquiche (Salar de Atacama)	-23.3836	-23.3836	Lake Tebenquiche is one of the largest saline water bodies in the Salar de Atacama at 2,500 m above sea level in northeastern Chile. Bacteria inhabiting there have to deal with extreme changes in salinity, temperature and UV dose (i.e., high		Extremophiles: life under extreme conditions	2008					
•	Alpine Peatlands of the Andes, Cajamarca, Peru	5.6556	5.6556	An ecological analysis of wetlands in the high mountain jalca above 3700 m elevation in the Andes near Cajamarca, Peru, indicated that most wetlands are groundwatersupported peat- accumulating fens. The floristic composition of fen		Arctic, Antarctic, and Alpine Research	2010					
	Effects of grazing pressure on plant species composition and water presence on bofedales in the Andes mountain range of Bolivia	5.6556	5.6556	Bofedales are high-Andean peatland plant communities with high capacity for water retention, which are regarded as oases of biodiversity. These areas have great social and economic value for livestock grazing, which plays an important role in their		Mires and Peat	August 13, 2018		DOLorg (CSL JSON)			Ø
	Extremophile culture collection from Andean lakes: extreme pristine environments that host a wide diversity of microorganisms with tolerance to UV radiation	38.5188	38.5188	A total of 88 bacterial strains were isolated from six Andean lakes situated at altitudes ranging from 3,400 to 4,600 m above sea level: L. Aparejos (4,200 m), L. Negra (4,400 m), L. Verde (4,460 m), L. Azul (4,400 m), L. Villama (4,600 m), and Salina		Microbial ecology	2009					
	Nurse species and indirect facilitation through grazing drive plant community functional traits in tropical alpine peatlands	-34.4407	-34.4407	Facilitation among plants mediated by grazers occurs when an unpalatable plant extends its protection against grazing to another plant. This type of indirect facilitation impacts species coexistence and ecosystem functioning in a large array of		Ecology and evolution	2017					
	Quaternary geochemical evolution of the salars of Uyuni and Coipasa, Central Altiplano, Bolivia	-20.4628	-20.4628	The central trough of the Bolivian Altiplano is occupied by two wide salt crusts: the salar of Uyuni, which is probably the largest salt pan in the world (10,000 km2) and the salar of Coipasa (2,500 km2). Both crusts are essentially made of porous halite		Chemical Geology	1991					Ø
	Diverse UV-B resistance of culturable bacterial community from high-altitude wetland water	40.5182	40.5182	Isolation of most ultraviolet B (UV-B)-resistant culturable bacteria that occur in the habitat of Laguna Azul, a high-altitude wetland [4554 m above sea level (asi)] from the Northwestern Argentinean Andes, was carried out by culture-based methods.		Current microbiology	2006					
	A limnological reconnaissance of Lake Tebenquiche, Salar de Atacama, Chile	-31.7613	-31.7613	A number of expeditions to the area of Mar de Atacarna, Chile, 68° 15°W, 20° 30°S, have involved studies of the biological and chinical features of Lake Tebenquiche, situated in the interior of the salar. Chemically, Tebenquiche is hypersaline, with		Hydrobiologia	1991					
	Water chemistry of Lake Quilotoa (Ecuador) and assessment of natural hazards	-0.8716	-0.8716	A geochemical survey carried out in November 1993 revealed that Lake Quilotoa was composed by a thin (saim14 m) oxic epillimnion overlying a \sim200 m-thick anoxic hypolimnion. Dissolved CO2 concentrations reached 1000 mg/kg in the lower		Journal of Volcanology and Geothermal Research	2000					Ø
	Socio-ecological dimensions of Andean pastoral landscape change: bridging traditional ecological knowledge and satellite image analysis in Sajama National Park, Bolivia	-18.1206	-18.1206	n the Andes, indigenous pastoral communities are confronting new challenges in managing mountain peatland pastures, locally called bofedales. Assessing land cover change using satellite images, vegetation survey, and local knowledge (i.e.,		Regional Environmental Change	2019					
	Crossing the western Altiplano: The ecological context of Tiwanaku migrations	38.5188	38.5188	Mobility and migration are critical processes that influence cultural and socio-economic development, and have lasting effects on demographic and ecological arrangements. However, these are ephemeral behaviors which are difficult to reconstruct		Journal of Archaeological Science	2020					
	Geomorphologic impacts of the glacial lake outburst flood from Lake No. 513 (Peru)	-41.7994	-41.7994	This article deals with the 2010 glacial lake outburst flood (GLOF) which happened in the Chucchu n' Valley in the Cordillera Blanca (Peru). The volume of the ice and nock fall from Mt. Huaica n was estimated to be about 500,000 m3 and the		Environmental Earth Sciences	2015					
•	Inferred ionic composition and salinity of a Bolivian Quaternary lake, as estimated from fossil diatoms in the sediments	-22.2063	-22.2063	Multiple Linear Regression (MLR) was used to estimate past salinity and ionic composition in sediments of a paleolake in the sounthermost Bolivian Altiplano. Modern diatom assemblages were sampled from the subsurface sediments of 14 saline		Hydrobiologia	3/1991		DOI.org (Crossref)			Ø
	Diurnal movements, family groups and alertness of vicuña (Vicugna vicugna) during the late dry season in the Laguna Blanca Reserve (Catamarca, Argentina)	-34.9965	-34.9965	A study of daily movementsand aspects of social behavior of vicufias (Ficugna vicugna) in Laguna Blanca Reserve (Catamarca, Argentina) was carried out duringthe last month of the dry season (November). Mean composition of a vicu-a		Small Ruminant Research	6/1992		DOI.org (Crossref)			Ø

Figure 3: Metadata table with sorting, filtering, external links and download options

## Technical Specifications

Component	Technology Stack						
Backend API	FastAPI 0.116+ with Pydantic v2						
Database	PostgreSQL 17+ with PostGIS 3.3+						
$\mathbf{ORM}$	SQLAlchemy 2.0+ with asyncio support						
Authentication	OAuth2 + JWT tokens						
NLP Processing	spaCy 3.8+ with en_core_web_lg model						
Frontend Framework	Vue.js 3.5+ with TypeScript						
UI Components	Vuetify 3.10+ material design						
Mapping Library	OpenLayers 10.6+ with vector layers						
Graph Visualization	Cytoscape.js 3.33+						

## Current Status & Future Development

## Development Timeline:

## • July-September:

MaRESS research suite development with User handling Zotero API integration for metadata retrieval Module 1 (Geographic Mapping) development Module 2 (Semantic Mapping) development

## • October:

Algorithm accuracy evaluation and improvements
Missing feature implementation for Modules 1-2
Implementation of Module 3 (Research Data Mapping)

## • November-December:

Final testing and bug fixing
Module 4 (AI-Assistant) implementation

# Key Features Implemented:

- User authentication and management system
- PostgreSQL spatial database with coordinate storage
- Automated PDF data processing pipeline (Implementation Workflow)
- Cytoscape graph visualization (Figure 1)
- Interactive clustered map visualization (Figure 4)
- Metadata table with filtering and export options (Figure 3)

Acknowledgments: This project is supported by NFDI4Earth and carried out at the Chair of Climatology at Technische Universität Berlin. Development infrastructure provided by TU Berlin's Institute of Ecology. Source code will be made available under open-source licensing upon project completion.

Contact: https://www.tu.berlin/klima • https://gitlab.klima.tu-berlin.de/klima/maress