

WETSCAPES2.0 Dataset Explorer

v1.0

The WETSCAPES 2.0 project produces a diverse range of datasets originating from field measurements, laboratory analyses, environmental monitoring, and molecular biology workflows. These datasets are stored and published through the project's CKAN data infrastructure.

As the number of datasets grows, efficiently locating and exploring relevant data becomes increasingly important for project partners. To address this need, a web-based CKAN Dataset Exploration and Filtering Application was developed.

The application provides a map interface that allows researchers to visualize dataset locations, inspect metadata, and dynamically filter datasets. The goal of the application is to simplify dataset discovery and improve accessibility of project data for all WETSCAPES 2.0 partners.

1. Application Architecture

The application integrates several components of the WETSCAPES data infrastructure. The architecture combines a web-based visualization interface with existing project data services and GIS tools. The main components include:

CKAN API

Dataset metadata and resource information are retrieved from the WETSCAPES CKAN instance using the CKAN API. The application queries CKAN dynamically and extracts relevant metadata such as: dataset title; authors and maintainers; temporal coverage; spatial coordinates; and additional metadata fields. This allows the map to always reflect the latest datasets available in CKAN.

Web Application (JavaScript)

The interactive application is implemented in JavaScript and runs directly in the web browser. It processes the metadata retrieved from CKAN, converts it into geospatial points, and manages all filtering logic. Key responsibilities of the application include: retrieving and parsing CKAN metadata; managing dataset filtering; handling map interaction and popups; and dynamically updating the displayed dataset points

Map Visualization (Leaflet)

The map interface is implemented using the Leaflet JavaScript mapping library. Leaflet provides the interactive map environment used to visualize dataset locations. It enables: interactive markers; popup windows displaying dataset metadata; dynamic updating of visible points; and user-defined layer management.

Predefined Layers (QGIS / QField)

In addition to CKAN datasets, the application can also display predefined spatial layers created within the WETSCAPES GIS workflow. These layers originate from QGIS projects and QField field mapping workflows. They represent spatial datasets such as: field sites; monitoring locations; near-site parking; and parking-to-sensor paths. These layers are maintained externally in the GIS environment and integrated into the web application as reference layers.

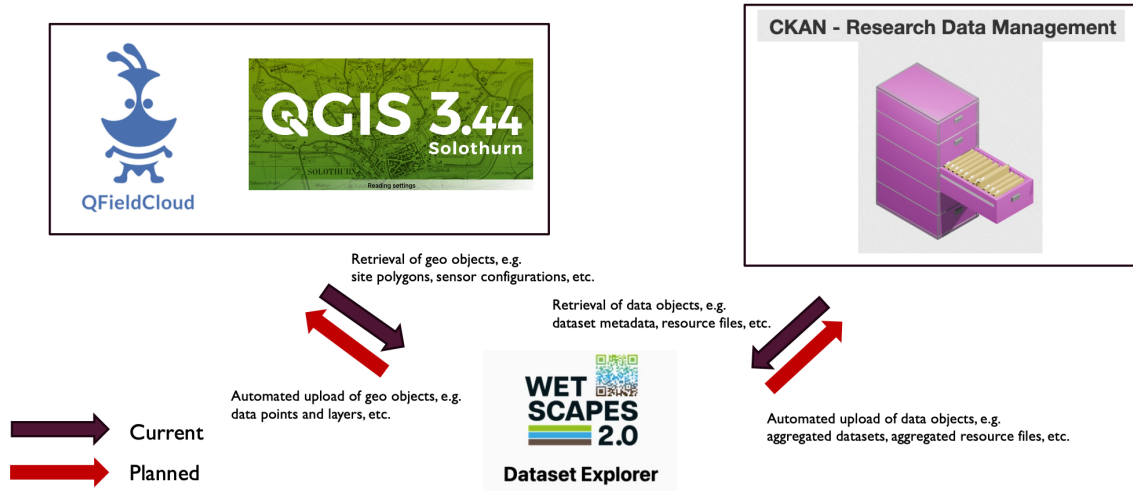


Figure 1 - Application Architecture

2. Application Overview

The application provides an interactive environment for exploring datasets stored in CKAN. When datasets are loaded, the application retrieves the metadata and displays the corresponding dataset locations as points on the map. Each **point** represents a dataset resource file measurement, associated with geographic coordinates. Users can interact with the map by clicking on dataset points to view detailed metadata information.

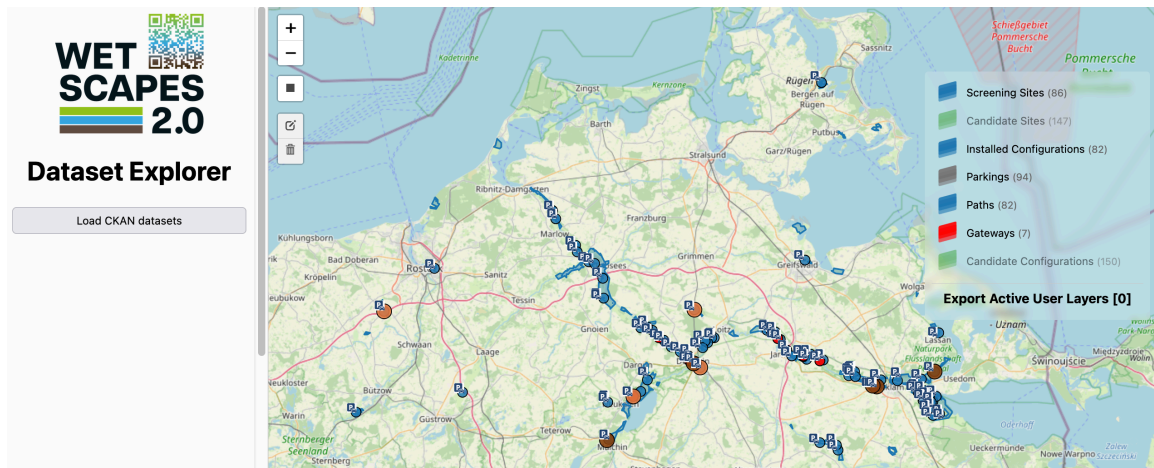


Figure 2 - Main Application Interface

3. Dataset Visualization

Dataset locations are displayed as markers on the map. Clicking on a marker opens a popup containing metadata extracted from CKAN. The popup may include information such as: **dataset name**; **dataset description**; **responsible persons**; **temporal coverage**; and **links to dataset resources**. The visualization dynamically updates whenever filters are applied or removed.

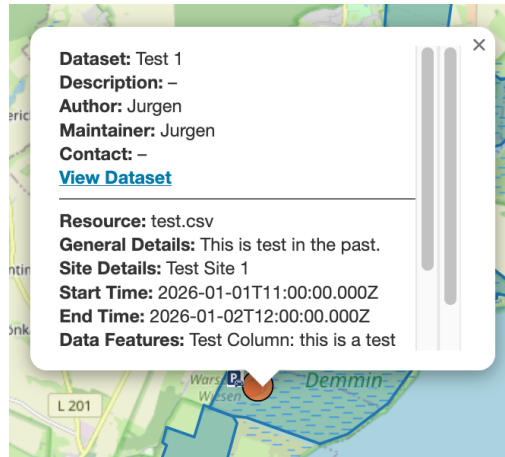


Figure 3 – Dataset Metadata Popup

4. Temporal Highlighting of Datasets

Dataset resource measurements, whose temporal coverage extends into the future are visually highlighted on the map. If a dataset has a start date that occurs today or in the future, it is displayed using a **darker brown marker colour**. This visual distinction allows users to quickly identify: **concluded datasets**; **ongoing datasets**; **planned data collection**; **datasets that will become available in the future**. This feature provides an intuitive overview of upcoming or currently active datasets within the project.

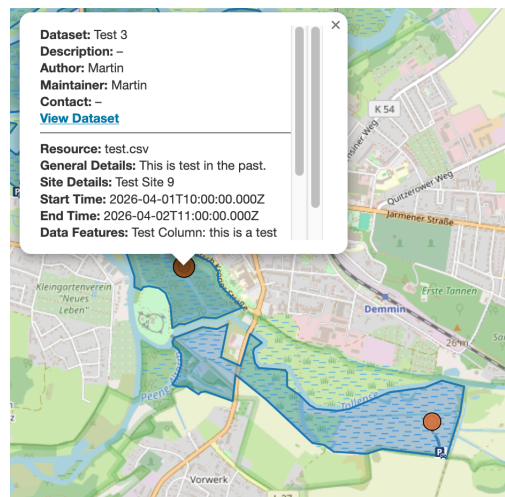


Figure 4 – Dataset Temporal Highlighting

5. Filtering Functionality

A key feature of the application is its ability to dynamically filter datasets based on metadata attributes. The filtering system allows users to narrow down the displayed dataset points according to several criteria. Multiple filters can be applied simultaneously, enabling flexible and targeted dataset exploration. The currently implemented filters include: **Site name filter**; **Person filter**; **Temporal filter**; and **Bounding-box filter**.

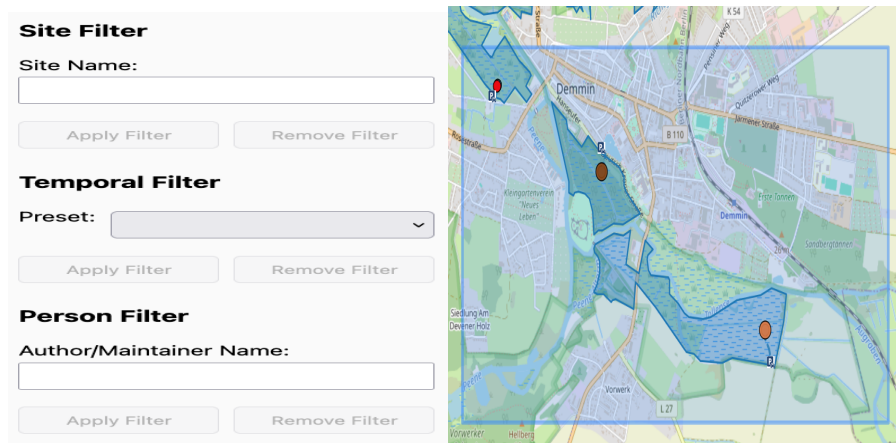


Figure 5 – (left) Site, Person, and Temporal filtes; (right) Bounding-box filter

5.1 Site Name Filter

The **Site Name Filter** allows users to search for datasets associated with specific field sites. Users may enter a full or partial site name. The application then filters the dataset points and displays only those whose site name contains the entered text. Features include: **case-insensitive search**; **partial text matching**; **dynamic update of map visualization**; and **ability to remove the filter independently**.

5.2 Person Filter

The **Person Filter** allows users to locate datasets associated with specific researchers. The filter searches the following CKAN fields: **dataset author**; and **dataset maintainer**. Users can enter a name or part of a name, and the application will display datasets linked to that person. This feature helps researchers quickly identify datasets created or maintained by specific project members.

5.3 Temporal Filter

The **Temporal Filter** allows users to restrict datasets based on their temporal coverage. Several predefined temporal filters are available: **past 1 week**; **past 1 month**; **next 1 week**; and **next 1 month**. In addition, users can define a **custom temporal range** by specifying the following options: **an after date** and **before date**. If only one boundary is provided, the application automatically extends the filter range to the project's temporal boundaries: project start date: **1 April 2025**; project end date: **31 December 2028**. Datasets are included if their temporal coverage **overlaps with the selected time interval**, ensuring that partially overlapping datasets are not excluded.

5.4 Nested Filtering

The filtering system supports **nested filtering**, meaning that multiple filters can be applied simultaneously. Users may combine filters in any order, for example: Site name + temporal filter, Person + site name, Person + temporal filter, Site + person + temporal filter. Each filter progressively reduces the set of displayed dataset points. Importantly, individual filters can be removed without resetting the entire filtering configuration. This enables users to explore datasets iteratively and refine their search interactively.

6. User-defined Layers

The application allows users to create **custom layers** from the currently visible dataset points. A typical workflow is: 1) **Apply** one or more **filters**; 2) **Identify** relevant dataset **points**; and 3) **Save them** as a new map layer. These layers allow users to group datasets according to thematic criteria or research interests. User-defined layers can then be displayed independently on the map.

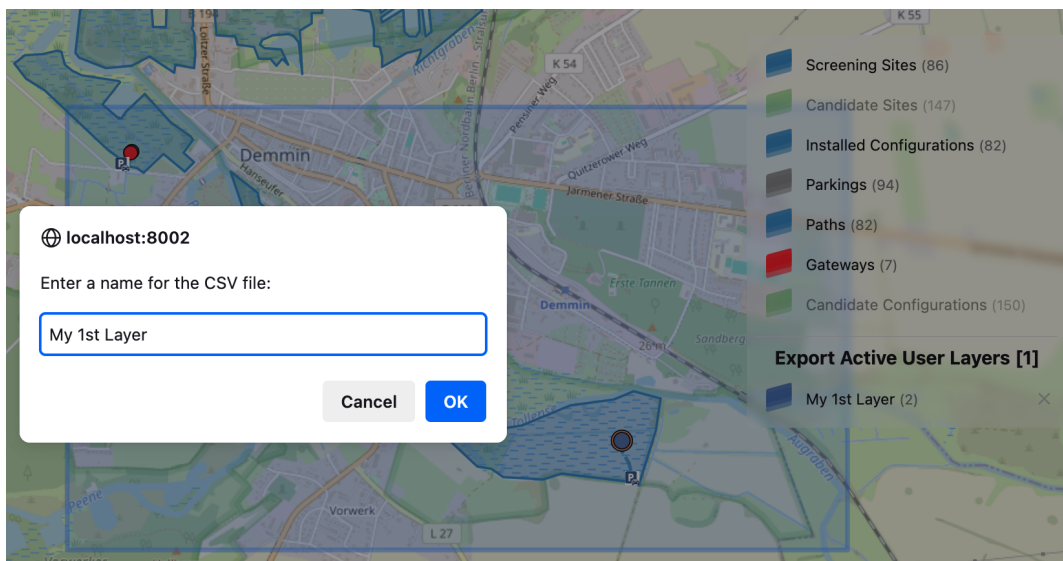


Figure 7 – Creating a User-Defined Layer

7. Benefits for the WETSCAPES 2.0 Project

The CKAN Dataset Exploration Application provides several advantages for project partners:

- improved accessibility of project datasets
- simplified dataset discovery
- visual overview of spatial dataset distribution
- interactive exploration of temporal dataset coverage
- efficient filtering of datasets based on research needs

The application therefore contributes to better data management and improved collaboration within the WETSCAPES 2.0 project.

8. Future Development

Several future developments may further enhance the capabilities of the application. Potential directions include:

Automated GIS Integration

Future work may explore automated workflows for transferring spatial data layers from **QGIS projects into the web application environment**. This would simplify the process of publishing new spatial layers.

Automated Dataset Publication

Another possible extension is the development of automated pipelines that upload **aggregated datasets directly to CKAN**. Such workflows could streamline the publication of processed datasets generated within the project.

Additional Metadata Filters

Additional filters could be implemented to support dataset exploration based on: data type; research domain; measurement method; and/or project work package