



research data publication services

Datasets available through DOI records are downloaded together with existing metadata of DOI registration agency and data publisher.

input data

locally held unpublished data

Data that exist solely on a local computer can be uploaded and processed. Resulting metadata image may be published without the dataset itself (but what's the use of a dataset that is not published?)

input data

data available through URLs

Single files (incl. file archives) downloadable from a URL may be included into the dataset metadata processing.

input data

extensible set of metadata entity types

Adjusted and generalized metadata scheme of BonaRes is adopted to represent various metadata entities.

Pointer to specific location in source data where particular instance of metadata entity (i.e. the metadata element) is found, is stored and processed with each of the elements.

metadata scheme

how about other languages?

Using a third party APIs to translate default english keywords for metadata entities in various languages will be tested

Using synonyms and generalisation of keywords by AGROVOC thesaurus will be tested.

term dictionaries

graphical user interface for data and metadata recognition and manipulation

Interactive streamlit based web application for editing metadata elements and their mapping to source files. Provides functionalities to incorporate data from local computer, web service or URLs.

web application

API for metadata structure image composition

Set of methods that provide access to core functionalities of metadata retrieving, file parsing and metadata entities recognition and adjustment.

Datasets and metadata files may be provided for incorporation to dataset structure. All files are analyzed according to their nature and coherent data structures are captured. Tree-structure of building elements is constructed for the whole dataset (ideally) up to the level of a table header and values recognition.

Generated metadata elements structure may be edited to add, remove, change locators and other properties (e.g. entity type). The metadata structure image may be stored to permanent storage to become findable in the query procedures.

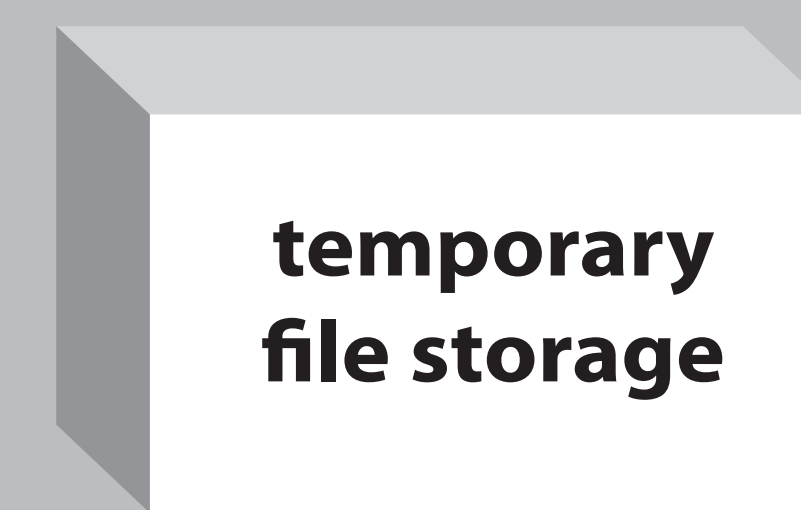
API

a software package for (semi-) automated metadata management and publication

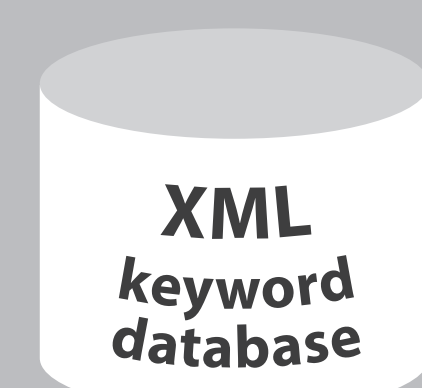
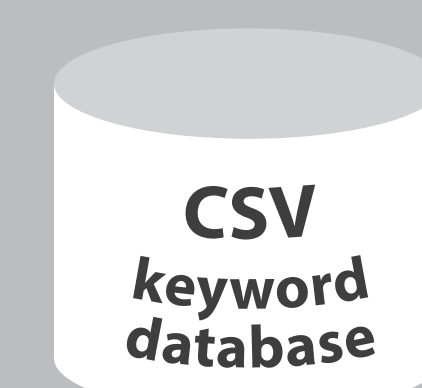
SoilPulse

The package core consists of basic methods to establish dataset file structure, recognize and validate file types and request metadata from provided sources.

Generally applicable (re) patterns specific for particular entity types are defined in file database distributed with the package.



Implementation of various file and other storage types (e.g. MySQL) is possible through adding a new module. Besides defining specialized procedures for different data types each module also includes search keywords definition file database.



package core

API for datasets searching

Set of methods that provide access to core functionalities for querying records in the central database. Structured information about datasets with metadata image records matching desired criteria can be retrieved and manipulated.

API

graphical user interface for datasets searching

Interactive streamlit based web application that provides user friendly access to functionalities of the SoilPulse API. Properties of datasets with metadata image records matching desired criteria are presented in a comprehensive way.

web application

metadata structure templates

Specific combination of measured phenomena may be needed for a particular research task (e.g. calibration procedures of Erosion-3D model). 'Metadata template' can be used within the metadata image generation phase to match if the dataset fulfills expected contents or to specify desired combination of contained data when gathering resources for that particular research task in a search procedure.

extension

We aim to develop a python library capable of handling metadata from existing data sets of various types and origin.

The package shell contain tools for:

- metadata retrieval and recognition (if already existing)
- metadata creation (by parsing the data and recognizing metadata elements)
- representing found elements by a common general structured metadata scheme (i.e. dataset metadata image)
- storing the resource's metadata image
- providing tools to query the storage to find available datasets fitting particular conditions
- matching existing metadata images against task-specific metadata templates

In the pilot phase we focus on handling textual, numerical and tabular data. Geographical data, images and other more complicated data structures should come next.

Jan Devátý

Jonas Lenz

Conrad Jackisch

jan.devaty@fsv.cvut.cz

jonas.lenz@iproconsult.com

conrad.jackisch@tbt.tu-freiberg.de

to get more about the idea in progress see also



EGU24-19497 | ECS | PICO | SSS2.2

SoilPulse – Towards FAIR soil process data!

Jonas Lenz, Jan Devátý, and Conrad Jackisch

Tue, 16 Apr, 16:36–16:38 (CEST) | PICO spot 2 | PICO2.9