

O2A specification for PANGAEA sources (.sdi.pangaea.lst)

- Preface
- Version 0.1
 - Requirements
 - Notes
 - Query Syntax/Formatting
 - Query Examples
 - File Examples

Preface

The [Spatial Data Infrastructure](#) can be fed with data published on [PANGAEA](#) to have it published as (part of) an OGC Web Service (OWS). How to specify which data our SDI should integrate/harvest from PANGAEA can be specified with `.sdi.pangaea.lst` files. They can hold both specific [PANGAEA DOIs](#) and [PANGAEA Search Queries](#). The latter will result in all resulting DOIs to be integrated/harvested into the SDI. All of an DOI's content which is available as PANGAEA tab files will get integrated/harvested.

Version 0.1

Requirements

- plain text file
- file extension: `.sdi.pangaea.lst`
- encoding: UTF-8
- one PANGAEA URL (query or direct DOI) per line
 - entries get separated by [newline](#)
 - no other separators allowed!
 - see section on [Query Syntax/Formatting](#) 

Notes

- readability features for humans
 - empty lines are fine
 - lines beginning with a hashtag (#) can be used for comments
 - no comments in lines with URLs
- a file's basename (e.g. `foobar` from `foobar.sdi.pangaea.lst`) can be used for convenient structuring but will not have any impact on SDI /OWS issues.
 - [Exchange Folder Specification](#) might apply

Query Syntax/Formatting

Advanced Search

The used syntax formatting can be explored in the [PANGAEA WIKI: Advanced Search](#) webpage. But it must be used as decoded link address without '+' signs.

- **do copy** filter from PANGAEA Search Bar content (and combine it with <https://www.pangaea.de/?q=>)
- **do not copy** filter from browser location bar (URL)

The screenshot shows the PANGAEA search interface. The URL in the address bar is highlighted with a yellow box and contains the query: `device:"Buoy" AND location:Arctic Ocean AND (project:AWI_Sealce OR project:AWI_BioOce)`. A yellow arrow points from the search bar to the search results. The search results page displays 89 datasets found for the query "device:'Buoy' AND location:Arctic Ocean AND (project:AWI_Sealce OR project:AWI_BioOce)". The first result is a dataset by Nicolaus et al. (2021) titled "Snow height on sea ice, meteorological conditions and drift of sea ice from autonomous Snow Buoys during MOSAiC 2019/20". A yellow checkmark is placed next to the search bar.

Query Examples

- valid** usage of Advanced Search Syntax:
 - Decoded URL: [https://www.pangaea.de/?q=device:'Buoy' AND location:Arctic Ocean AND \(project:AWI_Sealce OR project:AWI_BioOce\)](https://www.pangaea.de/?q=device:'Buoy' AND location:Arctic Ocean AND (project:AWI_Sealce OR project:AWI_BioOce))
- invalid** usage of Advanced Search Syntax:
 - Encoded URL: [https://www.pangaea.de/?q=device%3A%22Buoy%22+AND+location%3AArctic+Ocean+AND+\(project%3AAWI_Sealce+OR+project%3AAWI_BioOce\)](https://www.pangaea.de/?q=device%3A%22Buoy%22+AND+location%3AArctic+Ocean+AND+(project%3AAWI_Sealce+OR+project%3AAWI_BioOce))
- invalid**, Facet Filtering Syntax:
 - Decoded URL: [https://www.pangaea.de/?f.method\[\]='Buoy'&f.location\[\]='Arctic Ocean'&f.project\[\]='AWI_Sealce'&f.project\[\]='AWI_BioOce'](https://www.pangaea.de/?f.method[]='Buoy'&f.location[]='Arctic Ocean'&f.project[]='AWI_Sealce'&f.project[]='AWI_BioOce')
 - Encoded URL: https://www.pangaea.de/?f.method%5B%5D='Buoy'&f.location%5B%5D='Arctic+Ocean'&f.project%5B%5D='AWI_Sealce'&f.project%5B%5D='AWI_BioOce'
 - Facet Filtering Syntax is **NOT** supported!

The screenshot shows the PANGAEA search interface with an invalid facet filtering query. The URL in the address bar is highlighted with a yellow box and contains the query: `project:dam`. A yellow circle with a slash is placed over the facet filtering section. The search results page displays 79 datasets found for the query "project:dam". The results list includes various datasets from the MERIAN cruise, such as bathymetry raw data from RV MARIA S. MERIAN during cruise MSM55, MSM96, and MSM97. A yellow checkmark is placed next to the search bar.

File Examples

Valid File

valid_example.sdi.pangaea.lst

```
# single DOI(s)
https://doi.pangaea.de/10.1594/PANGAEA.123456
https://doi.pangaea.de/10.1594/PANGAEA.234567

# Filtered by project label 'DAM' as decoded URL
https://www.pangaea.de/?q=project:label:DAM

# Filtered by method and project as decoded URL
https://www.pangaea.de/?q=method:"Multicorer with television" AND (project:Hausgarten OR project:HERMES OR
project:FRAM)
```

Invalid File

invalid_example.sdi.pangaea.lst

```
# single DOI with invalid comment behind entry
https://doi.pangaea.de/10.1594/PANGAEA.123456 # my favourite DOI!

# single DOI with invalid address (missing https prefix)
doi.pangaea.de/10.1594/PANGAEA.234567

# Filtered by project label 'DAM' ...
# A) using Advanced Search Syntax but as encoded URL
https://www.pangaea.de/?q=project%3Alabel%3ADAM

# B) URL copied from address bar after using facet filtering
https://www.pangaea.de/?f.project[]=%DAM

# URL copied from address bar after using facet filtering -- works in browsers but not for harvesting
https://www.pangaea.de/?q=method:"Multicorer+with+television"&f.project[]=%Hausgarten&f.project[]=%HERMES&f.
project[]=%FRAM
```