

Package: kwb.geoportal (via r-universe)

June 14, 2026

Title R Package for Getting Spatial Data from Berlin Geoportal

Version 0.0.0.9000

Description R Package for getting spatial data from Berlin Geoportal
(<https://gdi.berlin.de/geonetwork/srv/ger/catalog.search#/search>).

License MIT + file LICENSE

URL <https://github.com/KWB-R/kwb.geoportal>

BugReports <https://github.com/KWB-R/kwb.geoportal/issues>

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports dplyr, purrr, tibble, xml2

Suggests covr, DT, knitr, rmarkdown

VignetteBuilder knitr

Config/pak/sysreqs libxml2-dev

Repository <https://kwb-r.r-universe.dev>

Date/Publication 2025-10-30 09:59:15 UTC

RemoteUrl <https://github.com/KWB-R/kwb.geoportal>

RemoteRef HEAD

RemoteSha 43e42fa79de6bb4ca48052ca8515f3320647a049

Contents

parse_gn_link	2
read_metadata	3
read_metadata_all	4

Index	6
--------------	----------

`parse_gn_link`*Parse a single GeoNetwork link element*

Description

GeoNetwork often encodes links as a single string separated by |, e.g.: |Darstellungsdienst (WMS)|https://...|OGC:WM
This helper splits such a string into named columns and pads missing parts up to 6 elements.

Usage

```
parse_gn_link(x)
```

Arguments

x Character string as found inside a <link> XML node.

Details

The order used here is:

1. link name
2. link description
3. link URL
4. link protocol (e.g. "OGC:WMS")
5. MIME type
6. order

Note: In many Berlin GDI records the **first** field (link name) is empty, and the actual meaningful text is in the **second** field (description).

Value

A one-row tibble with columns:

- link_name
- link_desc
- link_url
- link_protocol
- link_mime
- link_order

Examples

```
parse_gn_link("|Darstellungsdienst (WMS)|https://example.org/wms?|OGC:WMS|||")
```

read_metadata	<i>Read GeoNetwork service metadata (one row per service)</i>
---------------	---

Description

This function reads a GeoNetwork XML search response (as delivered by <https://gdi.berlin.de/geonetwork/srv/ger/>) and converts it into a tidy tibble with **one row per metadata record**. All `<link>` elements of a record are kept together in a **list column** called `links`, where each entry is itself a tibble created by [parse_gn_link\(\)](#).

Usage

```
read_metadata(path_xml)
```

Arguments

<code>path_xml</code>	Path or URL to the GeoNetwork XML document. This can be a local file (e.g. "geoportal_metadaten.xml") or a remote URL such as "https://gdi.berlin.de/geonetwork/srv/ger/geoportal_metadaten.xml"
-----------------------	--

Details

This structure is convenient when you want to keep the dataset-level information (title, abstract, uuid, ...) together, but still be able to inspect or unnest all service/download/view links later on.

The function assumes a GeoNetwork-style XML with `<metadata>` elements and the namespace `geonet:` available for the info block. It is tailored to the GDI Berlin instance but should work for other similar GeoNetwork responses that use the same link encoding (| separated).

If a record has **no** `<link>` elements, the `links` column will contain a single-row tibble with all NA values. This preserves the 1:1 alignment between records and rows.

Value

A tibble with one row per metadata record and the columns:

geonet_uuid UUID from `<geonet:info><uuid>` (character).

geonet_id Internal GeoNetwork id from `<geonet:info><id>` (character).

title Dataset/service title.

abstract Dataset/service abstract/description.

serviceType Service type, if present (e.g. WMS).

types Semicolon-separated `<type>` elements.

source_logo Logo path, if present.

links List column; each element is a tibble with the parsed links.

Examples

```
## Not run:
df <- read_geonetwork_services(
  "https://gdi.berlin.de/geonetwork/srv/ger/q?facet.q=type/service&resultType=details&sortBy=changeDate&from=1&to=100"
)
dplyr::glimpse(df)
df$links[[1]]

## End(Not run)
```

read_metadata_all	<i>Read all GeoNetwork / Geoportal metadata in chunks</i>
-------------------	---

Description

This function uses [read_metadata\(\)](#) repeatedly to fetch **all** available metadata records from a GeoNetwork endpoint that supports from / to pagination (like the GDI Berlin instance).

Usage

```
read_metadata_all(
  base_url =
    "https://gdi.berlin.de/geonetwork/srv/ger/q?facet.q=type/service&resultType=details&sortBy=changeDate&from=1&to=100"
  chunk_size = 100
)
```

Arguments

base_url	Base GeoNetwork query URL without from and to parameters. Must return an XML with a <summary count="..."> node. Defaults to the GDI Berlin service search.
chunk_size	Number of records per request. Default: 100.

Details

It first downloads the initial XML, reads the <summary count="..."> attribute to know how many records exist, then iterates in chunks (default: 100) until all records are read.

Value

A tibble with one row per metadata record, identical in structure to the return value of [read_metadata\(\)](#), but for **all** pages.

See Also

[read_metadata\(\)](#)

Examples

```
## Not run:  
all_md <- read_metadata_all()  
nrow(all_md)  
  
## End(Not run)
```

Index

`parse_gn_link`, [2](#)
`parse_gn_link()`, [3](#)

`read_metadata`, [3](#)
`read_metadata()`, [4](#)
`read_metadata_all`, [4](#)