Package 'nadaverse'

November 9, 2026

Type Package

Title Browse Microdata Catalogs Using 'NADA' REST API

Version 0.2.0

Date 2026-11-09

Description Provides a unified, programmatic interface for searching, browsing, and retrieving metadata from data repositories that use the National Data Archive ('NADA') software, such as the World Bank, 'FAO', the International Household Survey Network ('IHSN'), 'UNHCR', the World Health Organization, and the statistics offices of Kenya and Zambia. Functions discover available data collections, country codes, and access types, search by keyword, country, and year, and retrieve study metadata, file lists, and variable-level data dictionaries.

License MIT + file LICENSE

Encoding UTF-8

Imports cli, httr2

URL https://guturago.github.io/nadaverse/, https://github.com/guturago/nadaverse

BugReports https://github.com/guturago/nadaverse/issues

Suggests httptest2, knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Gutama Girja Urago [aut, cre, cph] (ORCID: <https://orcid.org/0000-0001-5588-2301>)

Maintainer Gutama Girja Urago <girjagutama@gmail.com>

Repository CRAN

Date/Publication 2026-11-09 18:50:02 UTC

Contents

catalogs 2
data_files 4
search_catalog 6
set_user_agent 9

Index **10**

`catalogs` *Small Helper Functions for Data Catalog Access*

Description

A suite of small **helper functions** designed to interact with and retrieve essential metadata from various international organization data repositories (catalogs). These functions standardize the process of obtaining lists of available **data access codes**, **collections**, **country codes**, and **latest entries** from specified sources.

Usage

```
catalogs(show = TRUE)

access_codes(catalog)

collections(catalog)

country_codes(catalog)

latest_entries(catalog, limit = NULL)

metadata(catalog, id)
```

Arguments

`show`	Logical. If `TRUE` (default), prints the supported catalogs to the console. The catalog table is returned invisibly either way.
`catalog`	A required character string specifying the name of the data catalog (e.g., `"fao"`, `"wb"`) from which to retrieve metadata.
`limit`	A positive integer number, applicable only to `latest_entries`, indicating the **maximum number of results** (data entries) to return. If `NULL` or omitted, a default limit set by the API will be used.
`id`	A required **study identifier**. Accepts either the numeric **Study ID** (integer, e.g., `101`) or the character **Study ID Number** (string, e.g., `"ALB_2012_LSMS_v01_M_v01_A_PUF"`). These values are typically returned in the search results from `search_catalog()`, `latest_entries()` or `data_files()`.

Details

All functions require a valid `catalog` code (see the *Supported Catalogs* section) and return their results in a standardized format. An unknown code stops with an error that lists the supported catalogs.

Each request waits up to 30 seconds and is attempted up to three times before it fails. If a catalog replies with something other than JSON, for example a block page for automated clients, the function stops with a clear message. Use `set_user_agent` to change the user agent sent with requests.

Catalogs differ in what they publish. For example, `ihsn` has no collections, so `collections("ihsn")` warns that it returned 0 collections and gives an empty result.

Value

- `catalogs`: Invisibly, a data frame with one row per supported catalog and the columns `catalog`, `name`, `home` and `api`.
- `access_codes`: A data frame of the data access types, with columns such as `id`, `type` and `title`.
- `collections`: A data frame of collections, with columns such as `id`, `repositoryid` and `title`.
- `country_codes`: A data frame of country codes, with columns such as `countryid`, `name` and `iso`.
- `latest_entries`: A data frame of the most recently added studies, with columns such as `id`, `idno`, `title`, `nation`, `created`, `changed` and `url`.
- `metadata`: A `list` with the study description. Its main parts are `doc_desc` (about the documentation) and `study_desc` (about the study itself).

Column names come from each catalog's API, so they can differ slightly between catalogs. If a catalog returns no records, a warning is issued and the result is empty. Network failures, HTTP errors and non-JSON replies raise an error.

Supported Catalogs

The `catalog` argument must be one of the following short codes (case-insensitive) corresponding to the respective microdata repository. **The list is sorted alphabetically by code.**

- `"df"`: **Data First** (https://www.datafirst.uct.ac.za)
- `"erf"`: **Economic Research Forum** (https://erfdataportal.com)
- `"fao"`: **Food and Agriculture Organization** (https://microdata.fao.org)
- `"ihsn"`: **International Household Survey Network** (https://catalog.ihsn.org)
- `"kenya"`: **Kenya National Bureau of Statistics** (https://statistics.knbs.or.ke/nada/index.php)
- `"unhcr"`: **United Nations High Commissioner for Refugees** (https://microdata.unhcr.org)
- `"wb"`: **The World Bank** (https://microdata.worldbank.org)
- `"who"`: **World Health Organization Health Information Systems** (https://apps.who.int/healthinfo/systems/surveydata
- `"zambia"`: **Zambia Statistics Agency** (https://nada.zamstats.gov.zm)

Author(s)

Gutama Girja Urago

See Also

The main search function: search_catalog; study files and variables: data_files; the user agent setting: set_user_agent

Examples

```
# --- Examples for Supported Catalogs ---

# 1. Data First (df): Get available access codes.
df_codes <- access_codes("df")

# 2. Economic Research Forum (erf): Get latest data entries (limited to 5).
erf_latest <- latest_entries("erf", limit = 5)

# 3. Food and Agriculture Organization (fao): Get available collections.
fao_collections <- collections("fao")

# 4. International Household Survey Network (ihsn): Get supported country codes.
ihsn_countries <- country_codes("ihsn")

# 5. Kenya National Bureau of Statistics (kenya): Get available collections.
kenya_collections <- collections("kenya")

# 6. United Nations High Commissioner for Refugees (unhcr): Get available collections.
unhcr_collections <- collections("unhcr")

# 7. The World Bank (wb): Get supported country codes.
wb_countries <- country_codes("wb")

# 8. World Health Organization HIS (who): Get available access codes.
who_codes <- access_codes("who")

# 9. Zambia Statistics Agency (zambia): Get latest entries (limit 10).
zambia_latest <- latest_entries("zambia", limit = 10)

# Example for the metadata function (requires a study ID)
wb_study_metadata <- metadata("wb", id = 8098)
str(wb_study_metadata)
```

---

data_files *Get Study Data Files List and Data Dictionary*

Description

Retrieves information about the files included in a study, or the detailed **data dictionary** (variables) for the entire study or a specific data file.

Usage

```
data_files(catalog, id)

data_dictionary(catalog, id, file_id = NULL)
```

Arguments

catalog	A required character string specifying the name of the data catalog (e.g., "wb", "fao"). Valid codes can be found in the documentation for catalogs().
id	A required study identifier. Accepts either the numeric **Study ID** (integer, e.g., 101) or the character **Study ID Number** (string, e.g., "ALB_2012_LSMS_v01_M_v01_A_PUF"). These values are typically returned by search_catalog() and latest_entries() (columns id and idno).
file_id	An optional character identifier, applicable only to data_dictionary(). This is the ID of a specific data file within the study, typically found in the file_id column returned by data_files(). If NULL (default), data_dictionary() attempts to fetch variables for the entire study.

Details

data_files() returns the list of files available for a study, along with metadata like file name, size, and ID.

data_dictionary() retrieves the variable-level metadata, including variable names, labels, and definitions. If file_id is provided, it retrieves the dictionary for that specific file; otherwise, it attempts to fetch the dictionary for the entire study. The function automatically detects whether the provided study identifier (id) is numeric or character.

Value

The return value depends on the function called:

- data_files(): A **data frame** with one row per data file. Typical columns include file_id, file_name, description, case_count and var_count.
- data_dictionary(): A **data frame** with one row per variable. Typical columns include vid, name and labl (the variable label).

If the study has no files or variables, an empty result is returned without a warning. Column names come from each catalog's API and can differ between catalogs.

Author(s)

Gutama Girja Urago

See Also

search_catalog, latest_entries

Examples

```
# Example 1: Get the list of files for a World Bank study (using idno)
study_idno <- "ALB_2012_LSMS_v01_M_v01_A_PUF"
files_wb <- data_files(catalog = "wb", id = study_idno)
print(files_wb)

# Example 2: Get the data dictionary for the entire study (using idno)
dictionary_all <- data_dictionary(catalog = "wb", id = study_idno)
head(dictionary_all)

# Example 3: Get the data dictionary for a specific file
# First, retrieve the files to find a file_id
file_id_to_use <- files_wb$file_id[1] # Use the ID of the first file
dictionary_file <- data_dictionary(
  catalog = "wb",
  id = study_idno,
  file_id = file_id_to_use
)
head(dictionary_file)
```

`search_catalog` *Search Catalogs*

Description

Performs a comprehensive search in the specified catalog's API endpoint, utilizing a full range of available searching, filtering, and sorting parameters.

Usage

```
search_catalog(
  catalog,
  keyword = NULL,
  from = NULL,
  to = NULL,
  country = NULL,
  inc_iso = NULL,
  collection = NULL,
  created = NULL,
  dtype = NULL,
  sort_by = NULL,
  sort_order = NULL,
  ps = NULL,
  page = NULL,
  rows = TRUE
)
```

Arguments

catalog	A required character string specifying the name of the data catalog (e.g., "fao", "wb"). Valid codes are listed by catalogs().
keyword	A character string used to search data titles, descriptions, and keywords (e.g., "lsms").
from	An integer indicating the **start year** for the data collection's coverage period (e.g., 2000).
to	An integer indicating the **end year** for the data collection's coverage period (e.g., 2010).
country	A character vector. Provide one or more **country names** or **ISO 3 codes** (case-insensitive). For valid codes, see country_codes(). Multiple values should be passed as a vector, e.g., c("afg", "Indonesia", "bra").
inc_iso	A logical value. If TRUE, the results data frame will include the ISO3 country codes; otherwise, it will contain only country names. **Default:** NULL.
collection	A character vector. Filters results by the data collection repository ID, which is returned in the repo_id column by collections(). Multiple IDs can be searched by passing a vector.
created	A character string used to filter results by the **date of creation** or update within the catalog. Use the date format YYYY/MM/DD. • Single date: "2020/04/01" (returns records created on or after this date). • Date range: "2020/04/01-2020/04/20" (returns records within the range).
dtype	A character vector. Filters results by one or more data access types. Valid values include: "open", "direct", "public", "licensed", "enclave", "remote", and "other". See access_codes() for a list of available types by catalog. Example: c("open", "licensed").
sort_by	A character string used to specify the column by which to sort the results. Valid values are: "rank", "title", "nation" (for country), or "year". Note that "country" is automatically mapped to the API field "nation".
sort_order	A character string indicating the sort direction. Must be either "asc" (ascending) or "desc" (descending).
ps	An integer indicating the number of records to display **per page** of results. **Default:** 15 records.
page	An integer specifying the **page number** of the search results to return.
rows	A logical value. If TRUE, the function returns only a data frame of the matching studies. If FALSE, it returns the full result list, which adds search information (such as found, total, limit and offset) to the rows data frame. **Default:** TRUE.

Details

This function constructs a complex API query based on the provided arguments (such as keywords, temporal range, geography, and access types) and returns the matching data entries. The function automatically handles URL encoding and JSON parsing.

All parameters correspond directly to the search options available on the NADA (National Data Archive) platform used by organizations like the World Bank and FAO.

Value

If `rows = TRUE` (default), a **data frame** where each row is a study matching the search criteria. If `rows = FALSE`, a **list** with the number of studies found (`found`), the catalog total (`total`), the page size and offset (`limit`, `offset`), and the `rows` data frame. If nothing matches, a warning is issued and `NULL` is returned. Network failures, HTTP errors and non-JSON replies raise an error.

Author(s)

Gutama Girja Urago

See Also

`access_codes`, `collections`, `country_codes`, `latest_entries`

Examples

```
# Example 1: Basic search for a keyword in the World Bank catalog
wb_search <- search_catalog(
  catalog = "wb",
  keyword = "LSMS",
  ps = 5, # 5 records per page
  page = 1
)
head(wb_search)

# Example 2: Search by country and year range
fao_search <- search_catalog(
  catalog = "fao",
  country = c("Kenya", "UGA"),
  from = 2010,
  to = 2020,
  sort_by = "year",
  sort_order = "desc"
)

# Example 3: Filter by access type and get search information
wb_info <- search_catalog(
  catalog = "wb",
  keyword = "labor",
  dtype = "public",
  rows = FALSE
)
print(wb_info$found) # Check total number of records found

# Example 4: Include ISO codes in results
ihsn_results <- search_catalog(
  catalog = "ihsn",
  inc_iso = TRUE
)
head(ihsn_results)
```

`set_user_agent` *Set the user agent used for NADA requests*

Description

Sets the `User-Agent` header sent with every request to the NADA catalogs for the rest of the R session. By default nadaverse sends a browser user agent, because some catalogs, such as UNHCR, reject requests from other clients. Use this function to send a different agent, for example one that identifies your project. If a catalog starts rejecting requests after you change the agent, call `set_user_agent(NULL)` to go back to the default.

Usage

```
set_user_agent(user_agent = NULL)
```

Arguments

`user_agent`	A single non-empty string to send as the user agent, or `NULL` to restore the default browser user agent.

Details

The setting is stored in the `nadaverse.user_agent` option.

Value

Invisibly, the previous setting: `NULL` when the default was in use. Pass it back to `set_user_agent()` to restore it.

See Also

`catalogs`

Examples

```
# Send a custom user agent for this session
old <- set_user_agent("My research project (nadaverse)")

# Restore the default browser user agent
set_user_agent(NULL)
```

Index

access_codes, *8*
access_codes *(*catalogs*)*, 2

catalogs, 2, *9*
collections, *8*
collections *(*catalogs*)*, 2
country_codes, *8*
country_codes *(*catalogs*)*, 2

data_dictionary *(*data_files*)*, 4
data_files, *4*, 4
data_files(), *2*

latest_entries, *5*, *8*
latest_entries *(*catalogs*)*, 2
latest_entries(), *2*

metadata *(*catalogs*)*, 2

search_catalog, *4, 5*, 6
search_catalog(), *2*
set_user_agent, *3, 4*, 9