

Package ‘datadict’

September 17, 2026

Title Validate Data Against a 'data-dict.yaml' Data Dictionary

Version 0.1.0

Description A data dictionary describes a collection of related tables: their columns, types, constraints, relationships, and glossary. Validate a dataset against its dictionary and read the findings as an HTML report. Reference: <<https://data-dict.tidyverse.org>>.

License MIT + file LICENSE

URL <https://data-dict.tidyverse.org>,
<https://github.com/tidyverse/data-dict>

BugReports <https://github.com/tidyverse/data-dict/issues>

Depends R (>= 4.1)

Imports cli (>= 3.2.0), processx, tools, utils

Suggests nanoparquet, testthat (>= 3.0.0), withr

Encoding UTF-8

Config/testthat/edition 3

Config/roxygen2/markdown TRUE

Config/roxygen2/version 8.1.0

NeedsCompilation no

Author Gábor Csárdi [aut, cre] (ORCID:
<<https://orcid.org/0000-0001-7098-9676>>),
Hadley Wickham [aut] (ORCID: <<https://orcid.org/0000-0003-4757-117X>>),
Posit Software, PBC [cph, fnd] (ROR: <<https://ror.org/03wc8by49>>)

Maintainer Gábor Csárdi <gabor@posit.co>

Repository CRAN

Date/Publication 2026-09-17 11:20:02 UTC

Contents

dd_install	2
dd_path	3
dd_run	3
dd_validate_data	4
Index	5

dd_install	<i>Download the data-dict binary</i>
------------	--------------------------------------

Description

Downloads the release archive for this platform, checks it against the published SHA-256, and unpacks the binary into tools::R_user_dir("datadict", "cache").

Usage

```
dd_install(version = "latest", force = FALSE, quiet = FALSE, dest = NULL)
```

Arguments

- | | |
|---------|---|
| version | Release to install, e.g. "0.0.1" (a leading v is also accepted). Defaults to the latest release. |
| force | Whether to download again when a binary is already installed. |
| quiet | Whether to suppress the download progress bar and messages. |
| dest | The destination path for the downloaded binary. Defaults to machines canonical data directory, see tools::R_user_dir() . Warning: if you change the default then datadict will not be able to find it automatically, so we strongly suggest leaving it as the default. |

Value

The path to the installed binary, invisibly.

Examples

```
dd_install()
```

dd_path	<i>Locate the data-dict binary</i>
---------	------------------------------------

Description

Searched in order: the DATA_DICT environment variable, the copy installed by `dd_install()` in the package's cache directory, then the PATH.

Usage

```
dd_path(check = TRUE)
```

Arguments

check	Whether to throw an error when no binary is found. With FALSE, return "" instead.
-------	---

Value

Path to the binary, or "" if it was not found and check is FALSE.

Examples

```
dd_path(check = FALSE)
```

dd_run	<i>Run the data-dict binary</i>
--------	---------------------------------

Description

Both output streams are captured, interleaved as the binary wrote them. With echo = TRUE they are also printed as they arrive, so a long run shows its progress.

Usage

```
dd_run(args, echo = FALSE, ...)
```

Arguments

args	Character vector of command line arguments.
echo	Whether to print the binary's output while it runs, on top of capturing it.
...	Passed on to <code>processx::run()</code> , e.g. timeout or wd.

Details

A non-zero exit status is an error, carrying the captured output as the message.

Value

A list with the exit status and the captured output, a character vector of lines, invisibly.

Examples

```
dd_run("--version")
```

dd_validate_data	<i>Validate a dataset's values against its data dictionary</i>
------------------	--

Description

Runs `data-dict render-report`, which validates the data and writes the findings as a self-contained HTML report, and opens it in a browser. The data files themselves come from each table's source: `parquet:` entry in the dictionary, resolved relative to the dictionary's directory.

Usage

```
dd_validate_data(
  dict = ".",
  table = NULL,
  html = tempfile(fileext = ".html"),
  browse = interactive()
)
```

Arguments

<code>dict</code>	A <code>data-dict.yaml</code> file, or a directory holding one.
<code>table</code>	Validate only this table, instead of every table in the dictionary.
<code>html</code>	Where to write the report.
<code>browse</code>	Whether to open the report with <code>utils::browseURL()</code> .

Details

A dataset that fails validation is not an R error: the report is the point, so check `status` to see whether the run passed.

Value

A list with the `html` path and the CLI's exit status, where 0 means the dataset validated, invisibly.

Examples

```
# a directory holding a data-dict.yaml and the parquet files it points at
dd_validate_data("inst/data")
```

Index

`dd_install`, [2](#)
`dd_install()`, [3](#)
`dd_path`, [3](#)
`dd_run`, [3](#)
`dd_validate_data`, [4](#)

`processx::run()`, [3](#)

`tools::R_user_dir()`, [2](#)

`utils::browseURL()`, [4](#)