

# Package ‘metajam’

August 16, 2024

**Type** Package

**Title** Easily Download Data and Metadata from 'DataONE'

**Version** 0.3.1

**Date** 2024-08-06

**Maintainer** Julien Brun <julien.brun@alumni.duke.edu>

**Description** A set of tools to foster the development of reproducible analytical workflow by simplifying the download of data and metadata from 'DataONE' (<<https://www.dataone.org>>) and easily importing this information into R.

**License** Apache License (== 2.0)

**Encoding** UTF-8

**Language** en-US

**RoxygenNote** 7.3.2

**SystemRequirements** Mac OSX: redland (>= 1.0.14) ; Linux: librdf0 (>= 1.0.14), librdf0-dev (>= 1.0.14)

**URL** <https://nceas.github.io/metajam/>, <https://github.com/NCEAS/metajam>

**BugReports** <https://github.com/NCEAS/metajam/issues>

**Depends** R (>= 3.6.0)

**Imports** dataone, dplyr, EML, emld, lubridate, purrr, readr, stats, stringr, tibble, tidy, XML

**Suggests** knitr, rmarkdown, testthat

**VignetteBuilder** knitr

**NeedsCompilation** no

**Author** Julien Brun [cre, aut] (<<https://orcid.org/0000-0002-7751-6238>>),  
Irene Steves [aut] (<<https://orcid.org/0000-0002-5511-9717>>,  
<https://github.com/isteves>),  
Mitchell Maier [aut] (<<https://orcid.org/0000-0001-6955-0535>>),  
Kristen Peach [aut] (<<https://orcid.org/0000-0002-6388-0901>>),  
Nicholas Lyon [aut] (<<https://orcid.org/0000-0003-3905-1078>>,  
<https://njlyon0.github.io/>),

Nathan Hwangbo [ctb] (<<https://orcid.org/0000-0002-3777-0821>>),  
 Derek Strong [ctb] (<<https://orcid.org/0000-0002-0347-8574>>),  
 Colin Smith [ctb] (<<https://orcid.org/0000-0003-2261-9931>>),  
 Regents of the University of California [cph]

**Repository** CRAN

**Date/Publication** 2024-08-16 17:50:02 UTC

## Contents

check_version . . . . .	2
download_d1_data . . . . .	3
download_d1_data_pkg . . . . .	4
read_d1_files . . . . .	4
tabularize_eml . . . . .	5

**Index** **6**

---

check_version	<i>Check PID version</i>
---------------	--------------------------

---

## Description

This function takes an identifier and checks to see if it has been obsoleted.

## Usage

```
check_version(pid, formatType = NULL)
```

## Arguments

pid	(character) The persistent identifier of a data, metadata, or resource map object on a DataONE member node.
formatType	(character) Optional. The format type to return (one of 'data', 'metadata', or 'resource').

## Value

(data.frame) A data frame of object version PIDs and related information.

## Examples

```
## Not run:
# Most data URLs and identifiers work
check_version("https://cn.dataone.org/cn/v2/resolve/urn:uuid:a2834e3e-f453-4c2b-8343-99477662b570")
check_version("doi:10.18739/A2ZF6M")

# Specify a formatType (data, metadata, or resource)
check_version("doi:10.18739/A2ZF6M", formatType = "metadata")
```

```
# Returns a warning if the identifier has been obsoleted
check_version("doi:10.18739/A2HF7Z", formatType = "metadata")

# Returns an error if no matching identifiers are found
check_version("a_test_pid")

# Returns a warning if several identifiers are returned
check_version("10.18739/A2057CR99")

## End(Not run)
```

---

download_d1_data	<i>Download data and metadata from DataONE</i>
------------------	--

---

## Description

Downloads a data object from DataONE along with metadata.

## Usage

```
download_d1_data(data_url, path)
```

## Arguments

data_url	(character) An identifier or URL for a DataONE object to download.
path	(character) Path to a directory to download data to.

## Value

(character) Path where data is downloaded to.

## See Also

[read\_d1\_files()] [download\_d1\_data\_pkg()]

## Examples

```
## Not run:
download_d1_data("urn:uuid:a2834e3e-f453-4c2b-8343-99477662b570", path = file.path("."))
download_d1_data(
  "https://cn.dataone.org/cn/v2/resolve/urn:uuid:a2834e3e-f453-4c2b-8343-99477662b570",
  path = file.path(".")
)

## End(Not run)
```

---

download\_d1\_data\_pkg    *Download all data and metadata of a data package from DataONE*

---

**Description**

Downloads all the data objects of a data package from DataONE along with metadata.

**Usage**

```
download_d1_data_pkg(meta_obj, path)
```

**Arguments**

meta\_obj            (character) A DOI or metadata object PID for a DataONE package to download.  
path                (character) Path to a directory to download data to.

**Value**

(list) Paths where data are downloaded to.

**See Also**

[read\_d1\_files()] [download\_d1\_data()]

**Examples**

```
## Not run:  
download_d1_data_pkg("doi:10.18739/A2028W", ".")  
download_d1_data_pkg("https://doi.org/10.18739/A2028W", ".")  
  
## End(Not run)
```

---

read\_d1\_files            *Read data and metadata based on 'download\_d1\_data()' file structure*

---

**Description**

Reads data along with metadata into your R environment based on [download\_d1\_data()] file structure.

**Usage**

```
read_d1_files(folder_path, fnc = "read_csv", ...)
```

**Arguments**

folder\_path (character) Path to a directory where data and metadata are located.  
 fnc (character) Function to be used to read the data (default is [readr::read\_csv()]).  
 ... Parameters to pass into the function specified in 'fnc'.

**Value**

(list) Named list containing data and metadata as data frames.

**See Also**

[download\_d1\_data()] [download\_d1\_data\_pkg()]

**Examples**

```
data_folder <- system.file(file.path("extdata", "test_data"), package = "metajam")
soil_moist_data <- read_d1_files(data_folder)

# You can specify the function you would like to use to read the file and pass parameters
soil_moist_data_skipped <- read_d1_files(data_folder, "read.csv",
                                       skip = 8, stringsAsFactors = FALSE)
```

---

tabularize_eml	<i>Get tabular metadata</i>
----------------	-----------------------------

---

**Description**

This function takes a path to an EML (.xml) metadata file and returns a data frame.

**Usage**

```
tabularize_eml(eml, full = FALSE)
```

**Arguments**

eml An emld class object, the path to an EML (.xml) metadata file, or a raw EML object.  
 full (logical) Returns the most commonly used metadata fields by default. If full = TRUE is specified, the full set of metadata fields are returned.

**Value**

(data.frame) A data frame of selected EML values.

**Examples**

```
eml <- system.file("extdata", "test_data", "SoilMois2012_2017__full_metadata.xml",
                  package = "metajam")
tabularize_eml(eml)
```

# Index

`check_version`, [2](#)

`download_d1_data`, [3](#)

`download_d1_data_pkg`, [4](#)

`read_d1_files`, [4](#)

`tabularize_eml`, [5](#)