

# Package: readbulk (via r-universe)

September 6, 2024

**Type** Package

**Version** 1.1.4

**Title** Read and Combine Multiple Data Files

**Description** Combine multiple data files from a common directory. The data files will be read into R and bound together, creating a single large data.frame. A general function is provided along with a specific function for data that was collected using the open-source experiment builder 'OpenSesame'  
<<https://osdoc.cogsci.nl/>>.

**Maintainer** Pascal J. Kieslich <pascal.kieslich@gmail.com>

**URL** <https://github.com/pascalkieslich/readbulk>

**BugReports** <https://github.com/pascalkieslich/readbulk/issues>

**License** GPL-3

**Depends** R (>= 3.1.0)

**Imports** utils, plyr

**Encoding** UTF-8

**LazyData** TRUE

**RoxygenNote** 7.2.3

**Repository** <https://pascalkieslich.r-universe.dev>

**RemoteUrl** <https://github.com/pascalkieslich/readbulk>

**RemoteRef** HEAD

**RemoteSha** 4f044a018b8bca260ee65fb8370a4e9c41a81cdd

## Contents

read_bulk . . . . .	2
read_opensesame . . . . .	4

<b>Index</b>	<b>6</b>
--------------	----------

---

`read_bulk`*Read and combine multiple data files*

---

### Description

Read and combine multiple data files. The files will be merged into one [data.frame](#).

### Usage

```
read_bulk(  
  directory = ".",  
  subdirectories = FALSE,  
  name_contains = NULL,  
  name_filter = NULL,  
  extension = NULL,  
  data = NULL,  
  verbose = TRUE,  
  fun = utils::read.csv,  
  ...  
)
```

### Arguments

<code>directory</code>	a character string. Name of the folder where the raw data are stored. If it does not contain an absolute path, the file name is relative to the current working directory. Defaults to current working directory.
<code>subdirectories</code>	logical indicating whether the directory contains subdirectories. If FALSE (the default), it is assumed that all raw data files are directly included in the directory. If TRUE, it is assumed that the raw data files are stored in folders within the directory. Alternatively, a vector of folder names that contain the raw data.
<code>name_contains</code>	an optional character string. If specified, only files whose name contains this string will be merged.
<code>name_filter</code>	an optional regular expression. If specified, only files whose name matches this regular expression will be merged.
<code>extension</code>	an optional character string. If specified, only files ending with the specified extension will be merged.
<code>data</code>	A <code>data.frame</code> to which the new data will be added. This is optional, and an empty <code>data.frame</code> is used if none is provided.
<code>verbose</code>	logical indicating whether function should report its progress.
<code>fun</code>	the function used for reading the individual files. By default, this is <a href="#">read.csv</a> . Can be any data import function as long as it takes the file name as first argument.
<code>...</code>	additional arguments passed on to <code>fun</code> .

## Details

read\_bulk provides a wrapper around a specific data import function ([read.csv](#) by default) to load the individual data files. After loading, the different data files are merged using [rbind.fill](#). This function can deal with varying column names across files, and still places data into the appropriate columns. If a column is not present in a specific file, it will be filled with NA.

## Value

A [data.frame](#) containing the merged data.

One column in the data.frame (File) contains the name of the raw data file. If the `subdirectories` option is set, an additional column (Subdirectory) with the name of the subdirectory is added.

## See Also

[read.table](#) for reading individual data files.

[rbind.fill](#) is responsible for merging files.

[write.table](#) for data export.

## Examples

```
## Not run:
# Merge all files in the main folder "raw_data"
# (which is in the current working directory)
raw_data <- read_bulk(directory = "raw_data")

# Merge files with file extension ".csv"
raw_data <- read_bulk(directory = "raw_data",
  extension = ".csv")

# Merge all files stored in separate folders
# within the folder "raw_data"
raw_data <- read_bulk(directory = "raw_data",
  subdirectories = TRUE)

# Merge all raw data stored in the folders "Session1"
# and "Session2" within the folder "raw_data"
raw_data <- read_bulk(directory = "raw_data",
  subdirectories = c("Session1", "Session2"))

# Merge tab separated data files and prevent
# character vectors from being converted to factors
raw_data <- read_bulk(directory = "raw_data",
  fun=read.delim, stringsAsFactors=FALSE)

## End(Not run)
```

---

read_opensesame	<i>Read and combine raw OpenSesame data</i>
-----------------	---

---

## Description

Read and combine multiple raw data files that were collected with OpenSesame (Mathot, Schreij, & Theeuwes, 2012). The files will be merged into one `data.frame`.

## Usage

```
read_opensesame(  
  directory = ".",  
  subdirectories = FALSE,  
  extension = NULL,  
  data = NULL,  
  verbose = TRUE  
)
```

## Arguments

<code>directory</code>	a character string. Name of the folder where the raw data are stored. If it does not contain an absolute path, the file name is relative to the current working directory. Defaults to current working directory.
<code>subdirectories</code>	logical indicating whether the directory contains subdirectories. If FALSE (the default), it is assumed that all raw data files are directly included in the directory. If TRUE, it is assumed that the raw data files are stored in folders within the directory. Alternatively, a vector of folder names that contain the raw data.
<code>extension</code>	an optional character string. If specified, only files ending with the specified extension will be merged.
<code>data</code>	A <code>data.frame</code> to which the new data will be added. This is optional, and an empty <code>data.frame</code> is used if none is provided.
<code>verbose</code>	logical indicating whether function should report its progress.

## Details

OpenSesame generally produces an output `.csv` file for each participant in the experiment. This is handy during data collection, but for the analysis it is often useful to combine many such files into a single `data.frame`. This is the single task of the `read_opensesame` function, which loads all files from a given directory and attempts to combine them into a `data.frame`.

`read_opensesame` provides a wrapper around `read_bulk` to load the raw data files. After loading, the different data files are merged using `rbind.fill`. This function can deal with varying column names across files, and still places data into the appropriate columns. If a column is not present in a specific file, it will be filled with NA.

**Value**

A [data.frame](#) containing the merged raw data.

One column in the data.frame (File) contains the name of the raw data file. If the subdirectories option is set, an additional column (Subdirectory) with the name of the subdirectory is added.

**References**

Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314-324.

**See Also**

[read\\_bulk](#) for reading and combining multiple data files that have other file formats.

**Examples**

```
## Not run:
# Read single raw data file from OpenSesame
raw_data <- utils::read.csv("raw_data/subject-1.csv",encoding = "UTF-8")

# Merge all files in the main folder "raw_data"
# (which is in the current working directory)
raw_data <- read_opensesame(directory = "raw_data")

# Merge files with file extension ".csv"
raw_data <- read_opensesame(directory = "raw_data",
  extension = ".csv")

# Merge all files stored in separate folders
# within the folder "raw_data"
raw_data <- read_opensesame(directory = "raw_data",
  subdirectories = TRUE)

# Merge all raw data stored in the folders "Session1"
# and "Session2" within the folder "raw_data"
raw_data <- read_opensesame(directory = "raw_data",
  subdirectories = c("Session1","Session2"))

# Export merged data to a file using write.table
write.table(raw_data, file = "raw_data.csv",
  sep="," , row.names = FALSE)

## End(Not run)
```

# Index

`data.frame`, 2–5

`rbind.fill`, 3, 4

`read.csv`, 2, 3

`read.table`, 3

`read_bulk`, 2, 4, 5

`read_opensesame`, 4

`write.table`, 3