

# Package ‘scimetr’

May 9, 2026

**Type** Package

**Title** Analysis of Scientific Publication Data with R

**Version** 1.2.0

**Date** 2026-03-01

**Description** Tools for quantitative research in scientometrics and bibliometrics.  
This package provides routines for importing bibliographic data from Clarivate Web of Science (<<https://www.webofscience.com/wos/>>) and performing bibliometric analysis.

**Depends** R (>= 4.1.0), graphics

**Imports** dplyr, tidyr, stringr, stringi, scales, ggplot2, rlang,  
openxlsx

**Suggests** DT, rmarkdown, knitr

**License** GPL (>= 2)

**URL** <https://rubenfcasal.github.io/scimetr/>,  
<https://github.com/rubenfcasal/scimetr/>

**BugReports** <https://github.com/rubenfcasal/scimetr/issues/>

**Encoding** UTF-8

**VignetteBuilder** knitr

**LazyData** true

**RoxygenNote** 7.3.2

**NeedsCompilation** no

**Author** Ruben Fernandez-Casal [aut, cre] (ORCID:  
<<https://orcid.org/0000-0002-5785-3739>>),  
Borja Lafuente-Rego [aut] (ORCID:  
<<https://orcid.org/0000-0002-2443-3084>>),  
Maria Jose Lombardia [aut] (ORCID:  
<<https://orcid.org/0000-0001-9452-9818>>),  
Javier Tarrío-Saavedra [aut] (ORCID:  
<<https://orcid.org/0000-0002-9584-127X>>),  
Julian Costa-Bouzas [aut] (ORCID:

<<https://orcid.org/0000-0001-9760-9581>>),

Yesica Fernandez-Ramos [ctb],

Guillermo Lopez-Taboada [ctb]

**Maintainer** Ruben Fernandez-Casal <rubenfcasal@gmail.com>

**Repository** CRAN

**Date/Publication** 2026-03-16 20:00:02 UTC

## Contents

scimetr-package . . . . .	2
add_jcr . . . . .	3
dbjcr . . . . .	4
db_bib . . . . .	4
db_jcr . . . . .	6
get_id_docs . . . . .	7
get_jcr . . . . .	8
import_wos . . . . .	9
summary.wos.db . . . . .	10
summary_year . . . . .	11
wosdf . . . . .	13

**Index** **16**

---

scimetr-package	<i>scimetr: Analysis of Scientific Publication Data with R</i>
-----------------	--

---

## Description

Tools for quantitative research in scientometrics and bibliometrics. This package provides routines for importing bibliographic data from Clarivate Web of Science (<https://www.webofscience.com/wos/>) and performing bibliometric analysis. For more information visit <https://rubenfcasal.github.io/scimetr/articles/scimetr.html>.

## Author(s)

**Maintainer:** Ruben Fernandez-Casal <rubenfcasal@gmail.com> ([ORCID](#))

Authors:

- Borja Lafuente-Rego <borjalrego@gmail.com> ([ORCID](#))
- Maria Jose Lombardia <maria.jose.lombardia@udc.es> ([ORCID](#))
- Javier Tarrio-Saavedra <javier.tarrio@udc.es> ([ORCID](#))
- Julian Costa-Bouzas <julian.costa.bouzas@udc.es> ([ORCID](#))

Other contributors:

- Yesica Fernandez-Ramos <yesica.ramos@udc.es> [contributor]
- Guillermo Lopez-Taboada <guillermo.lopez.taboada@udc.es> [contributor]

**See Also**

Useful links:

- <https://rubenfcasal.github.io/scimetr/>
- <https://github.com/rubenfcasal/scimetr/>
- Report bugs at <https://github.com/rubenfcasal/scimetr/issues/>

---

 add\_jcr

*Add JCR data to a bibliographic database*


---

**Description**

Extends the bibliographic database by adding JCR metrics to sources, per year and WoS category.

**Usage**

```
add_jcr(db, jcrdb)
```

```
## S3 method for class 'wos.jcr'
plot(
  x,
  filter,
  plot = TRUE,
  all = FALSE,
  ask = plot && all && interactive(),
  ...
)
```

**Arguments**

db	a bibliographic database (a <code>wos.db</code> object; typically the output of the function <code>db_bib()</code> ).
jcrdb	a JCR database (a <code>jcr.db</code> object; typically the output of the function <code>db_jcr()</code> ).
x	a <code>wos.jcr</code> object.
filter	vector of document identifiers (usually a result of <code>get_id_docs()</code> ).
plot	logical; if TRUE (default), the plots are drawn, otherwise only the ggplot2 object is (invisibly) returned.
all	logical; if TRUE, function <code>plot.wos.db()</code> is called (additional parameters ... are passed to this function), otherwise only a JCR metrics plot is generated.
ask	logical; if TRUE, the user is asked before each plot (see <code>par(ask=.)</code> ).
...	further arguments passed to or from other methods.

**Value**

An S3 object of class `wos.jcr`. A `wos.db` object with additional components `JCRSour` and `JCRCatSour`.

**See Also**

[db\\_bib](#), [db\\_jcr](#).

---

dbjcr

*Mathematics UDC 2018-2023 bibliographic database*

---

**Description**

Bibliographic database with JCR metrics (a [wos.jcr](#) S3 object) corresponding to a WoS search by the Affiliation field of *Universidade da Coruña (UDC)* in the research area "Mathematics" during the years 2018–2023 (generated from [wosdf](#) data set, using the functions [db\\_bib](#), [db\\_jcr](#) and [add\\_jcr](#)).

**Usage**

```
dbjcr
```

**Format**

An object of class `wos.jcr` (inherits from `wos.db`) of length 22.

---

db\_bib

*Bibliographic data base creation*

---

**Description**

`db_bib.wos.data()` converts a data.frame with WoS variables (as returned by [import\\_wos](#); see [wosdf](#)) into a relational database (a list of data.frames).

**Usage**

```
db_bib(data, ...)

## S3 method for class 'wos.data'
db_bib(data, label = "", progress = interactive(), verbose = FALSE, ...)

## S3 method for class 'wos.db'
print(x, ...)

## S3 method for class 'wos.db'
plot(
  x,
  filter,
  which = 1:3,
  plot = TRUE,
  warning = FALSE,
  ask = plot && (length(which) > 1) && interactive(),
  ...
)
```

**Arguments**

data	a data.frame with bibliographic data.
...	further arguments passed to or from other methods.
label	character string describing the data.
progress	logical; if TRUE the progress is printed.
verbose	logical; if TRUE additional information is printed.
x	an object used to select a method.
filter	vector of document identifiers (usually a result of <a href="#">get_id_docs</a> ).
which	if a subset of the plots is required, specify a subset of the numbers 1:3.
plot	logical; if TRUE (default), the plots are drawn, otherwise only the list of <code>ggplot2</code> objects is (invisibly) returned.
warning	logical; if FALSE (default), warnings are ignored.
ask	logical; if TRUE, the user is asked before each plot (see <a href="#">par(ask=.)</a> ).

**Value**

An S3 object of class `wos.db`. A list with the following components:

- Docs: document-level records (year, type, citations, identifiers).
- Authors and AutDoc: author dictionary and document-author links.
- OI and OIDoc: ORCID identifiers and their linkage to documents.
- RI and RIDoc: ResearcherID identifiers and their linkage to documents.
- Affiliations and AffDoc: affiliation dictionary and document links.
- Addresses and AddAutDoc: address and document-author-address links.
- Sources: document sources (journals, books, proceedings, ...).
- Categories, Areas, CatSour and AreaSour: thematic classifications and their linkage to sources.
- WSIndex and SourWSI: Web of Science indexes and their linkage to sources.
- Label and date: metadata that helps tracking the dataset identity over time.

This object has specialized [print](#) and [plot](#) methods.

**See Also**

[wosdf](#), [import\\_wos](#).  
[summary.wos.db](#).  
[plot.summary.wos.db](#), [plot.summary.year.wos](#)

**Examples**

```
db <- db_bib(wosdf)
print(db)
summary(db)
```

db\_jcr

*Import JCR data from Web of Science (WoS)***Description**

Reads JCR data from excel files downloaded from WoS and generates a relational database (a list of data.frames). It is assumed that the file name format is *JCR\_[WE]\_[PY]*, where *[WE]* is the WoS index (SCIE, SSCI, ...) and *[PY]* the JCR year.

**Usage**

```
db_jcr(path = ".", files = dir(path, pattern = "*.xlsx"), verbose = TRUE)
```

**Arguments**

path	character; path to the directory containing the files. Defaults to the working directory.
files	character vector with the file names. Defaults to filenames in directory path with extension <i>.xlsx</i> .
verbose	logical; indicating whether the name of the file being processed is printed. Defaults to TRUE.

**Details**

A subscription to Web of Science is required to download JCR data. For successful execution, Excel files must retain the original header fields from Clarivate exports: Title20, ISO\_ABBREV, TITLE, ISSN, EISSN, ISSUES/YEAR, COUNTRY, LANGUAGE, 1ST\_YR\_PUB, categories, TOT\_CITES, CITES\_JCR\_YR, CITES\_JCR\_YR1, CITES\_JCR\_YR2, IF\_NUMERATOR, CITES\_JCR\_YR3, CITES\_JCR\_YR4, CITES\_JCR\_YR5, 5YR\_IF\_NUMERATOR, ITEMS\_JCR\_YR, ITEMS\_JCR\_YR1, ITEMS\_JCR\_YR2, IF\_DENOMINATOR, ITEMS\_JCR\_YR3, ITEMS\_JCR\_YR4, ITEMS\_JCR\_YR5, 5YR\_IF\_DENOMINATOR, IMPACT\_FACTOR, IMMEDIACY\_INDEX, CITED\_HALF\_LIFE, 5YR\_IMPACT\_FACTOR, EIGENFACTOR, NORM\_EIGENFACTOR, ARTL\_INFLUENCE, FREQUENCY, PUBCODE, PUBLISHER\_NAME, CATEGORY\_CODE, CATEGORY\_DESCRIPTION, CATEGORY\_RANKING, QUARTILE\_RANK, JIF\_PERCENTILE.

**Value**

An S3 object of `class` `jcr.db`. A list with components: Sources, Categories, JCRSour and JCRCatSour.

**See Also**

[db\\_bib](#), [add\\_jcr](#).

---

get_id_docs	<i>Get table identifiers</i>
-------------	------------------------------

---

## Description

Retrieve internal identifiers (entity keys) for the corresponding relational table (authors, sources, categories, areas, ...). Any field in the target table can be used as a condition, and multiple conditions are combined with & (only IDs for which all conditions evaluate to TRUE are returned). *Tidy evaluation* can be used to construct the logical expressions, see [dplyr::filter\(\)](#).

## Usage

```
get_id_docs(  
  db,  
  ...,  
  id_sources,  
  id_authors,  
  id_addresses,  
  id_oi,  
  id_ri,  
  id_affiliations  
)  
  
get_id_authors(db, ...)  
  
get_id_addresses(db, ...)  
  
get_id_areas(db, ...)  
  
get_id_categories(db, ...)  
  
get_id_wsi(db, ...)  
  
get_id_sources(db, ..., id_categories, id_areas, id_wsi)  
  
get_id_oi(db, ...)  
  
get_id_ri(db, ...)  
  
get_id_affiliations(db, ...)
```

## Arguments

db	Object of <code>class</code> <code>wos.db</code> , as returned by <code>db_bib</code> .
...	Logical predicates. Multiple conditions are combined with & (see <a href="#">filter</a> ).
id_sources	optional; sources identifiers.

id_authors	optional; author identifiers (values of db\$Authors\$ids).
id_addresses	optional; addresses identifiers.
id_oi	optional; ORCID identifiers or codes (values of db\$OI\$OI).
id_ri	optional; RI identifiers or codes (values of db\$RI\$RI).
id_affiliations	optional; affiliations identifiers or names (values of db\$Affiliations\$C3).
id_categories	optional; categories identifiers or names (values of db\$Categories\$WC)
id_areas	optional; research area identifiers or names (values of db\$Areas\$SC)
id_wsi	optional; WoS Index identifiers or names (values of db\$WSI\$WE)

**Value**

An integer vector of identifiers. For instance, `get_id_docs` returns the document identifiers (values of the `db$Docs$idd` variable) corresponding to identifiers of authors, categories, areas, addresses and/or sources. Logical expressions defined in terms of the variables in `db$Docs` can also be used as arguments.

**See Also**

[db\\_bib](#), [filter](#).

---

get_jcr	<i>Get JCR metrics</i>
---------	------------------------

---

**Description**

`get_jcr()` combines document indexes with their source JCR metrics per year.

`get_jcr_cat()` combines document indexes with their source JCR metrics per year and WoS category.

**Usage**

```
get_jcr(db, filter)
```

```
get_jcr_cat(db, filter, best = TRUE)
```

**Arguments**

db	a bibliographic database with JCR information (a <a href="#">wos.jcr</a> object; typically the output of the function <a href="#">add_jcr()</a> ).
filter	vector of document identifiers (usually a result of <a href="#">get_id_docs()</a> ).
best	logical; if TRUE (default), only the results for the WoS category with the best ranking for each document are returned.

**Value**

A data.frame with document indexes and their source JCR metrics.

**See Also**

[add\\_jcr\(\)](#), [db\\_jcr\(\)](#), [db\\_bib\(\)](#).

**Examples**

```
head(get_jcr(dbjcr))
head(get_jcr_cat(dbjcr, best = TRUE))
```

---

import_wos	<i>Import bibliographic data downloaded from Web of Science (WoS).</i>
------------	--

---

**Description**

Reads bibliography entries from UTF-8 encoded Tab-delimited files containing "Full Record and Cited References" (see [wosdf](#) and vignette [Downloading data from the Web of Science](#)).

**Usage**

```
import_wos(path = ".", pattern = "*.txt", all = TRUE, progress = NULL)
```

**Arguments**

path	character; path to the directory containing the files.
pattern	regular expression; only matching files will be loaded. Defaults to "*.txt".
all	logical; indicating whether sources without ISSN are included (TRUE) or not (FALSE) in the result. Defaults to TRUE.
progress	logical; indicating whether a progress bar should be displayed. Defaults to TRUE when R is used interactively and there is more than one file, and FALSE otherwise.

**Details**

A subscription to Web of Science is required to download bibliometric data.

**Value**

A data.frame with rows corresponding to sources and columns to WoS variables.

**See Also**

[wosdf](#), [db\\_bib\(\)](#).

summary.wos.db

*Summaries of a bibliographic database***Description**

`summary()` methods for bibliometric databases.

**Usage**

```
## S3 method for class 'wos.db'
summary(object, filter, index = c("H", "G"), nmax = 10, ...)

## S3 method for class 'summary.wos.db'
print(x, ...)

author_metrics(db, id_authors)

## S3 method for class 'summary.wos.db'
plot(
  x,
  which = 1:5,
  pie = FALSE,
  others = !pie,
  accuracy = 0.1,
  plot = TRUE,
  ask = length(which) > 1 && interactive(),
  ...
)

## S3 method for class 'wos.jcr'
summary(object, filter, all = FALSE, ...)

## S3 method for class 'summary.wos.jcr'
print(x, digits = 2, all = TRUE, ...)

## S3 method for class 'summary.wos.jcr'
plot(x, plot = TRUE, all = FALSE, ask = plot && all && interactive(), ...)
```

**Arguments**

object	an object for which a summary is desired.
filter	vector of document identifiers (optional, usually a result of <code>get_id_docs()</code> ).
index	citation indexes.
nmax	number of top levels.
...	further arguments passed to or from other methods.

x	an object used to select a method.
db	Object of <code>class</code> <code>wos.db</code> , as returned by <code>db_bib</code> .
id_authors	optional; author identifiers
which	if a subset of the plots is required, specify a subset of the numbers 1:5.
pie	logical; if TRUE, pie charts are drawn.
others	logical; if FALSE, only nmax top levels categories are shown.
accuracy	numerical scalar; see <code>label_number(accuracy=.)</code> .
plot	logical; if TRUE (default), the plots are drawn, otherwise only the list of <code>ggplot2</code> objects is (invisibly) returned.
ask	logical; if TRUE, the user is asked before each plot, see <code>par(ask=.)</code> .
all	logical; if TRUE, the corresponding <code>wos.db</code> method is called (additional parameters ... are passed to it), otherwise only results of JCR metrics are returned.
digits	minimal number of significant digits, see <code>print.default</code> .

### Value

A list of summary statistics (an object of class `summary.wos.db` and/or `summary.wos.jcr`), which has specialized `print()` and `plot()` methods.

### See Also

`db_bib()`, `add_jcr()`.

### Examples

```
# Bibliographic database
db <- db_bib(wosdf)
summary(db)
# Bibliographic database with JCR metrics
summary(dbjcr)
```

---

summary\_year

*Summaries per year*

---

### Description

The `summary_year` methods return summaries by year of a bibliometric database.

**Usage**

```
summary_year(object, ...)

## S3 method for class 'wos.db'
summary_year(object, filter, ...)

## S3 method for class 'summary.year.wos'
print(x, ...)

## S3 method for class 'summary.year.wos'
plot(
  x,
  which = 1:3,
  boxplot = FALSE,
  plot = TRUE,
  ask = plot && length(which) > 1 && interactive(),
  ...
)

## S3 method for class 'wos.jcr'
summary_year(object, filter, all = FALSE, ...)

## S3 method for class 'summary.year.jcr'
print(x, digits = 3, all = TRUE, ...)

## S3 method for class 'summary.year.jcr'
plot(x, plot = TRUE, all = FALSE, ask = plot && interactive(), ...)
```

**Arguments**

object	an object for which a summary is desired.
...	further arguments passed to or from other methods.
filter	vector of document identifiers (optional, usually a result of <a href="#">get_id_docs</a> ).
x	an object used to select a method.
which	if a subset of the plots is required, specify a subset of the numbers 1:3.
boxplot	logical; if TRUE, boxplots are drawn (for plots from 2 to 3).
plot	logical; if TRUE (default), the plots are drawn, otherwise only the list of <code>ggplot2</code> objects is (invisibly) returned.
ask	logical; if TRUE, the user is asked before each plot, see <a href="#">par</a> ( <code>ask=.</code> ).
all	logical; if TRUE, the corresponding <code>wos.db</code> method is called (additional parameters ... are passed to it), otherwise only results of JCR metrics are returned.
digits	minimal number of <i>significant</i> digits, see <a href="#">print.default</a> .

**Value**

A list of summary statistics (an object of class `summary.year.wos` and/or `summary.year.jcr`), which has specialized [print\(\)](#) and [plot\(\)](#) methods.

**See Also**

[db\\_bib\(\)](#), [add\\_jcr\(\)](#).

**Examples**

```
# Bibliographic database
db <- db_bib(wosdf)
summary_year(db)
# Bibliographic database with JCR metrics
summary_year(dbjcr)
```

---

wosdf

*Mathematics UDC WoS Core Collection 2018-2023 data*

---

**Description**

The data set consists of 293 publications corresponding to a WoS search by the Affiliation field of *Universidade da Coruña (UDC)* (Affiliation: OG = Universidade da Coruna) in the research area "Mathematics" (SC = Mathematics), indexed in the Web of Science Core Collection:

- Science Citation Index Expanded (SCI-EXPANDED).
- Social Sciences Citation Index (SSCI).
- Arts & Humanities Citation Index (A&HCI).

in the years 2018-2023 (generated using the [import\\_wos](#) function).

**Usage**

```
wosdf
```

**Format**

A data frame with 293 rows and 48 columns:

**PT** Publication Type  
**AU** Author  
**AF** Author Full Name  
**TI** Article Title  
**SO** Source Title  
**SE** Book Series Title  
**BS** Book Series Subtitle  
**LA** Language  
**DT** Document Type  
**C1** Address  
**C3** Affiliation

**RI** Researcher Ids  
**OI** ORCID  
**NR** Cited Reference Count  
**TC** Times Cited, WoS Core  
**Z9** Times Cited, All Databases  
**U1** 180 Day Usage Count  
**U2** Since 2013 Usage Count  
**PU** Publisher  
**PI** Publisher City  
**PA** Publisher Address  
**SN** ISSN  
**EI** eISSN  
**BN** ISBN  
**J9** Journal Abbreviation  
**JI** Journal ISO Abbreviation  
**PD** Publication Date  
**PY** Publication Year  
**VL** Volume  
**IS** Issue  
**PN** Part Number  
**SU** Supplement  
**SI** Special Issue  
**MA** Meeting Abstract  
**BP** Start Page  
**EP** End Page  
**AR** Article Number  
**DI** DOI  
**D2** Book DOI  
**EA** Early Access Date  
**PG** Number of Pages  
**WC** WoS Category  
**WE** Web of Science Index  
**SC** Research Area  
**HC** Highly Cited Status  
**HP** Hot Paper Status  
**DA** Date of Export  
**UT** Unique WOS ID

**Source**

Clarivate Web of Science: <https://www.webofscience.com/wos/>.

**See Also**

[import\\_wos](#), [db\\_bib](#)

# Index

- \* **datasets**
  - dbjcr, 4
  - wosdf, 13
- add\_jcr, 3, 4, 6
- add\_jcr(), 8, 9, 11, 13
- author\_metrics (summary.wos.db), 10
- class, 3, 5–7, 11
- db\_bib, 4, 4, 6–8, 11, 15
- db\_bib(), 3, 9, 11, 13
- db\_jcr, 4, 6
- db\_jcr(), 3, 9
- dbjcr, 4
- dplyr::filter(), 7
- filter, 7, 8
- get\_id\_addresses (get\_id\_docs), 7
- get\_id\_affiliations (get\_id\_docs), 7
- get\_id\_areas (get\_id\_docs), 7
- get\_id\_authors (get\_id\_docs), 7
- get\_id\_categories (get\_id\_docs), 7
- get\_id\_docs, 5, 7, 12
- get\_id\_docs(), 3, 8, 10
- get\_id\_oi (get\_id\_docs), 7
- get\_id\_ri (get\_id\_docs), 7
- get\_id\_sources (get\_id\_docs), 7
- get\_id\_wsi (get\_id\_docs), 7
- get\_jcr, 8
- get\_jcr\_cat (get\_jcr), 8
- import\_wos, 4, 5, 9, 13, 15
- jcr.db, 3
- jcr.db-class (db\_jcr), 6
- label\_number, 11
- par, 3, 5, 11, 12
- plot, 5
- plot(), 11, 12
- plot.summary.wos.db, 5
- plot.summary.wos.db (summary.wos.db), 10
- plot.summary.wos.jcr (summary.wos.db), 10
- plot.summary.year.jcr (summary\_year), 11
- plot.summary.year.wos, 5
- plot.summary.year.wos (summary\_year), 11
- plot.wos.db (db\_bib), 4
- plot.wos.db(), 3
- plot.wos.jcr (add\_jcr), 3
- print, 5
- print(), 11, 12
- print.default, 11, 12
- print.summary.wos.db (summary.wos.db), 10
- print.summary.wos.jcr (summary.wos.db), 10
- print.summary.year.jcr (summary\_year), 11
- print.summary.year.wos (summary\_year), 11
- print.wos.db (db\_bib), 4
- scimetr (scimetr-package), 2
- scimetr-package, 2
- summary(), 10
- summary.wos.db, 5, 10
- summary.wos.jcr (summary.wos.db), 10
- summary\_year, 11
- wos.db, 3, 11, 12
- wos.db-class (db\_bib), 4
- wos.jcr, 3, 4, 8
- wos.jcr-class (add\_jcr), 3
- wosdf, 4, 5, 9, 13