

# Retrieve clinical trial information

Ralf Herold

2021-11-20

## Get started

### Attach package ctrdata

```
library(ctrdata)
citation("ctrdata")
```

Remember to respect the registers' terms and conditions (see `ctrOpenSearchPagesInBrowser(copyright = TRUE)`). Please cite this package in any publication as follows: Ralf Herold (2021). `ctrdata`: Retrieve and Analyze Clinical Trials in Public Registers. R package version 1.8.0. <https://cran.r-project.org/package=ctrdata>

### Open register's advanced search page in browser

These functions open the browser, where the user can start searching for trials of interest.

```
# Please review and respect register copyrights:
ctrOpenSearchPagesInBrowser(
  copyright = TRUE
)
# Open browser with example search:
ctrOpenSearchPagesInBrowser(
  url = "cancer&age=under-18",
  register = "EUCTR"
)
```

### Adjust search parameters and execute search in browser

Refine the search until the trials of interest are listed in the browser. The total number of trials that can be retrieved with package `ctrdata` is intentionally limited to queries with at most 10000 result records.

### Copy address from browser address bar to clipboard

Use functions or keyboard shortcuts according to the operating system.

## Get address from clipboard

The next steps are executed in the R environment:

```
q <- ctrGetQueryUrl()
# Found search query from EUCTR.

q
#                                     query-term  query-register
# 1 query=cancer&age=under-18&status=completed&phase=phase-one      EUCTR

# To check, a browser with this query
# is opened with this command
ctrOpenSearchPagesInBrowser(
  url = q
)
```

## Retrieve protocol-related information, transform, save to database, check

```
# Connect to a database and chose a table / collection
db <- nodbi::src_sqlite(
  dbname = "sqlite_file.sql",
  collection = "test"
)

# Count number of trial records
ctrLoadQueryIntoDb(
  queryterm = q,
  only.count = TRUE,
  con = db
)$n
# * Found search query from EUCTR: query=cancer&age=under-18&status=completed&phase=phase-one
# (1/3) Checking trials in EUCTR:
# Retrieved overview, multiple records of 71 trial(s) from 4 page(s) to be downloaded
# [1] 71

# Retrieve records, download into database
ctrLoadQueryIntoDb(
  queryterm = q,
  con = db
)
# * Found search query from EUCTR: query=cancer&age=under-18&status=completed&phase=phase-one
# (1/3) Checking trials in EUCTR:
# Retrieved overview, multiple records of 71 trial(s) from 4 page(s) to be downloaded
# Checking helper binaries: done
# Downloading trials (4 pages in parallel)...
# Note: register server cannot compress data, transfer takes longer, about 0.3s per trial
# Pages: 4 done, 0 ongoing
# (2/3) Converting to JSON, 279 records converted
# (3/3) Importing JSON records into database...
# = Imported or updated 279 records on 71 trial(s)
# No history found in expected format.
```

```
# * Updated history ("meta-info" in "test")

# Show which queries have been downloaded into database
dbQueryHistory(con = db)
#      query-timestamp query-register query-records
# 1 2021-11-18 22:27:12      EUCTR      279
#
#      query-term
# 1 query=cancer&age=under-18&status=completed&phase=phase-one
```

With a file-base SQLite database, this takes about 3 minutes for 1000 records, with most of the time needed for internet-retrieval with is slow from this register. Speed is higher with other registers, with using MongoDB and with memory-based SQLite.

## Repeat and update a previous query

```
ctrLoadQueryIntoDb(
  querytoupdate = "last",
  con = db
)
```

Instead of “last”, an integer number can be specified for `querytoupdate` that corresponds to the number when using `dbQueryHistory()`.

Depending on the register, an update (differential update) is possible or the original query is executed fully again.

## Retrieve results

For EUCTR, result-related trial information has to be requested to be retrieved, because it will take longer to download and store. For CTGOV, any results are always included in the retrieval.

```
ctrLoadQueryIntoDb(
  queryterm = q,
  euctrresults = TRUE,
  con = db
)
# [...]
# * Retrieving results if available from EUCTR for 71 trials:
# (1/4) Downloading results (max. 10 trials in parallel):
# 71 downloaded, extracting x x x . . . . . PDF . . . PDF . . PDF . . . PDF . x PDF . x x x . PDF . x x
# (2/3) Converting to JSON, 53 records converted
# (3/4) Importing JSON into database...
# (4/4) Results history: not retrieved (euctrresultshistory = FALSE)
# = Imported or updated results for 53 trials
# * Updated history ("meta-info" in "test")
```

The download or presence of results is not recorded in `dbQueryHistory()` because the availability of results increases over time.

## Add information from another register

The same database and table / collection can be used to store (and analyse) trial information from different registers. At the moment, `ctrdata` only supports the two registers <https://ClinicalTrials.Gov/> and <https://ClinicalTrialsRegister.EU/>. Example:

```
ctrLoadQueryIntoDb(  
  queryterm = "https://clinicaltrials.gov/ct2/results?cond=neuroblastoma&recrs=e&age=0&intr=Drug",  
  con = db  
)  
  
# * Found search query from CTGOV: cond=neuroblastoma&recrs=e&age=0&intr=Drug  
# (1/3) Checking trials in CTGOV:  
# Retrieved overview, records of 196 trial(s) are to be downloaded  
# Checking helper binaries: done  
# Downloading: 1.3 MB  
# (2/3) Converting to JSON, 196 records converted  
# (3/3) Importing JSON records into database...  
# = Imported or updated 195 trial(s)  
# * Updated history ("meta-info" in "test")
```

## Add personal annotations

When downloading trial information, the user can specify an annotation to all records that are downloaded. By default, annotations are accumulated if trial records are loaded again or updated; alternatively, annotations can be replaced.

Annotations are useful for analyses, for example to specially identify subsets of records in the database.

```
ctrLoadQueryIntoDb(  
  queryterm = "https://clinicaltrials.gov/ct2/results?cond=neuroblastoma&recrs=e&age=0&intr=Drug&cntry=DE",  
  annotation.text = "site_DE ",  
  annotation.mode = "append",  
  con = db  
)  
  
# * Found search query from CTGOV: cond=neuroblastoma&recrs=e&age=0&intr=Drug&cntry=DE  
# (1/3) Checking trials in CTGOV:  
# Retrieved overview, records of 11 trial(s) are to be downloaded  
# Checking helper binaries: done  
# Downloading: 69 kB  
# (2/3) Converting to JSON, 11 records converted  
# (3/3) Importing JSON records into database...  
# = Imported or updated 11 trial(s)  
# = Annotated retrieved records (11 records)  
# * Updated history ("meta-info" in "test")
```

## Find synonyms of active substance names

Not all registers automatically expand search terms to include alternative terms, such as codes and other names of active substances. To obtain a character vector of synonyms for any active substance name, use:

```
ctrFindActiveSubstanceSynonyms(
  activesubstance = "imatinib"
)
# [1] "imatinib" "gleevec" "sti 571" "glivec" "CGP 57148" "st1571"
```

These names can then be used in queries in any register.

## Using a MongoDB database

This example works with a free service here. Note that the user name and password need to be encoded. The format of the connection string is documented at <https://docs.mongodb.com/manual/reference/connection-string/>.

```
# Specify base uri for remote mongodb server,
# as part of the encoded connection string
db <- nodbi::src_mongo(
  # Note: this provides read-only access
  url = "mongodb+srv://DWbJ7Wh:bdTHh5cS@cluster0-b9wpw.mongodb.net",
  db = "dbperm",
  collection = "dbperm")

# Since the above access is read-only,
# just obtain fields of interest:
dbGetFieldsIntoDf(
  fields = c("a2_eudract_number",
             "e71_human_pharmacology_phase_i"),
  con = db)

#           _id a2_eudract_number e71_human_pharmacology_phase_i
# 1 2010-024264-18-3RD      2010-024264-18                TRUE
# 2 2010-024264-18-AT      2010-024264-18                TRUE
# 3 2010-024264-18-DE      2010-024264-18                TRUE
# 4 2010-024264-18-GB      2010-024264-18                TRUE
# 5 2010-024264-18-IT      2010-024264-18                TRUE
# 6 2010-024264-18-NL      2010-024264-18                TRUE
```