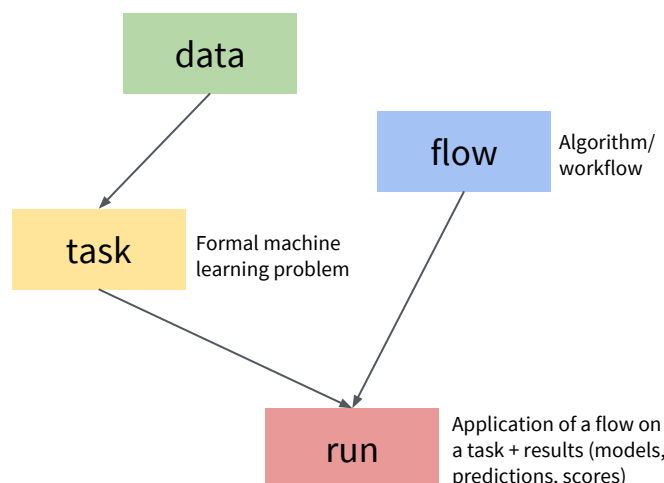


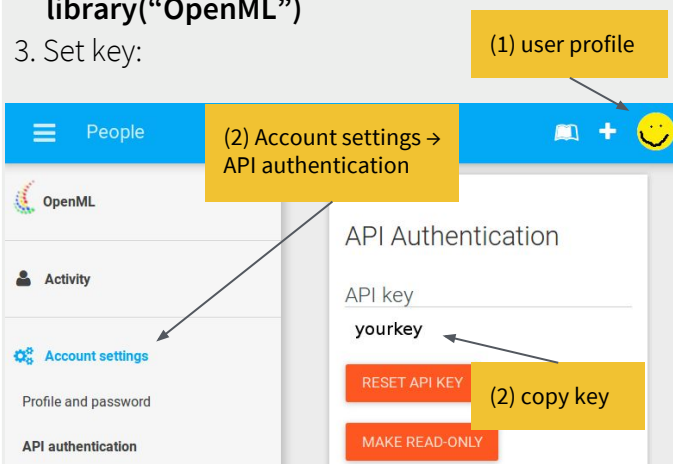
R Cheat Sheet

OpenML Objects



Setup

1. Create account on OpenML.org
2. Install and load R package
`install.packages("OpenML")`
`library("OpenML")`
3. Set key:



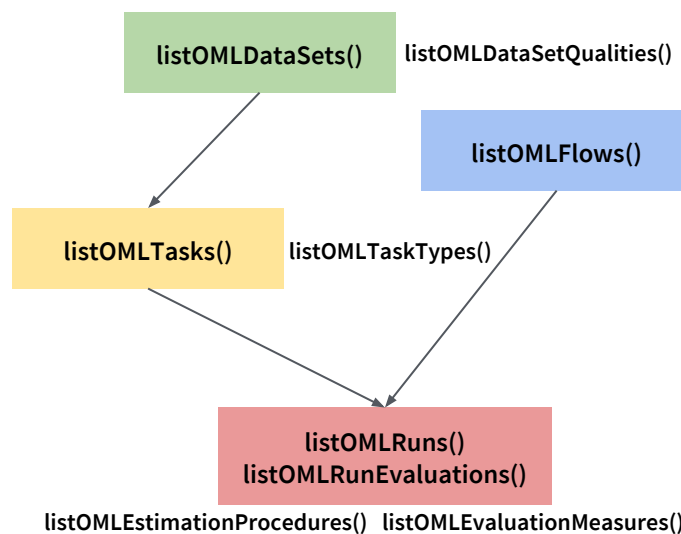
`saveOMLConfig(apikey = "yourkey")`

Getting/setting setup options

`getOMLConfig()`
`setOMLConfig()`
`saveOMLConfig(..., overwrite=TRUE)`
Find your config-file: `path.expand("~/openml/config")`

Listing, Download, Running and Uploading

Listing - Result is always a data frame with info



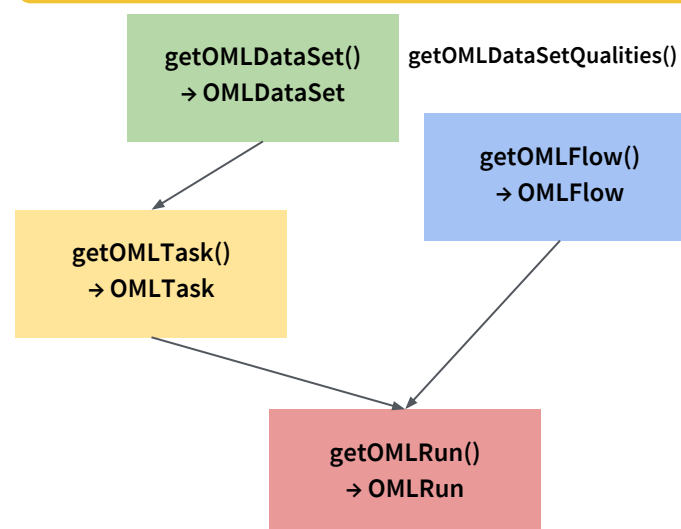
Running models on tasks

Example:

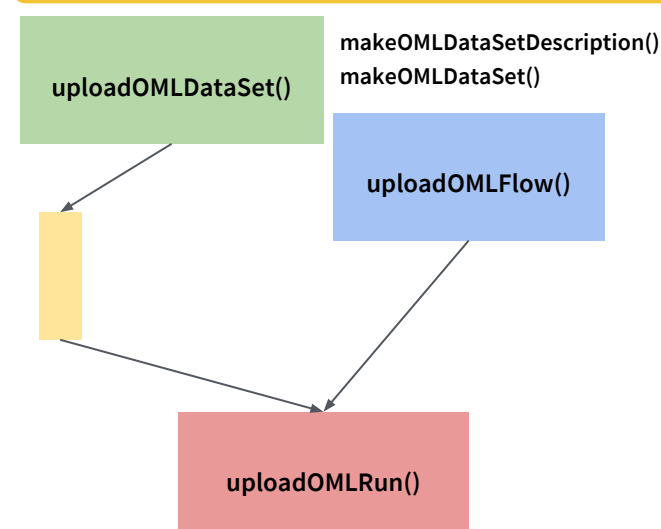
```
task = getOMLTask(task.id = 59)
lrn = mlr::makeLearner("classif.rpart")
run.mlr = runTaskMlr(task, lrn)
```

For more complex analyses see also [mlr cheatsheet](#)

Downloading

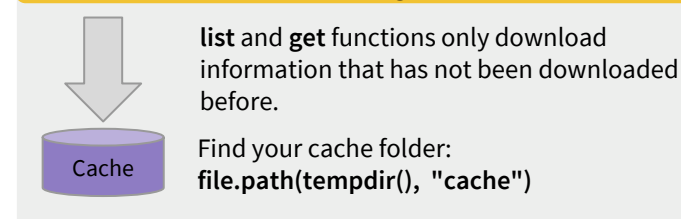


Uploading



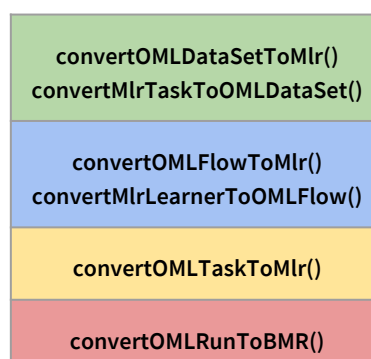
Other features

Caching

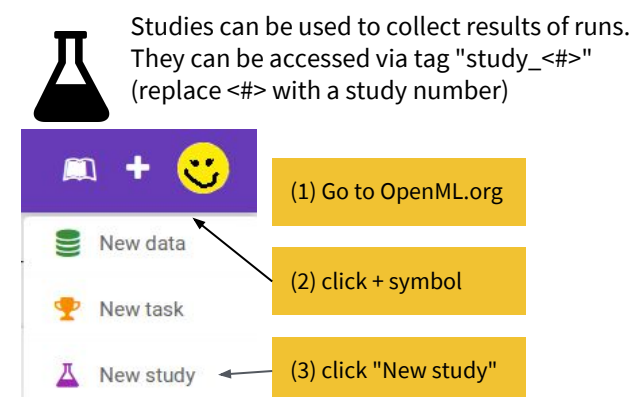


Conversion OpenML (OML) -- mlr (Mlr)

OpenML objects can be converted to mlr objects and the other way around



OpenML Studies



List all results of a study, e.g. study 34:
`listOMLRunEvaluations(tag = "study_34")`

Use cases

Data owner

TODO: Upload a data set and create a task so that others can build models.

1. Clean data set
2. `makeOMLDataSetDescription()`
3. `makeOMLDataSet()`
4. `uploadOMLDataSet()`
5. Wait for data set to be active on OpenML.org
6. Create task on OpenML.org: go to your data set page
→ DEFINE A NEW TASK

In case users raise issues, start with step 1.

Machine learner

TODO: Solve a task as well as possible.

1. `getOMLTask()`
2. `mlr::makeLearner()`
3. `runTaskMlr()`
4. `uploadOMLRun()` (this automatically also uploads the flow if not yet available)
5. Check performance of run on website and compare to other runs, if improvement possible/needed, return to step 2.

Benchmark

TODO: Compare the performance of different algorithms (learners) on different tasks.

1. `taskinfos = listOMLTasks(<characteristics>)`
2. `tasks = lapply(taskinfos$task.id, getOMLTask)`
3. `mylrns = lapply(<mylearners>, mlr::makeLearner)`
4. `grid = expand.grid(task.id = tasks$task.id, lrn.ind = seq_along(mylrns))`
5. `myruns = lapply(seq_row(grid), function(i) { task = getOMLTask(grid$task.id[i]) ind = grid$lrn.ind[i] runTaskMlr(task, lrn.list[[ind]]) })`
6. `lapply(myruns, uploadOMLRun, tags = "<mytag>")`
7. View results on OpenML.org or get results with `evals = listOMLRunEvaluations(tag = "<mytag>")`