Building the icus.data data package

Alan Engel

March 27, 2024
File version: 0.02

The **icus.data** package is intended to be compatible with the rOpenSci Project[2] and relies mainly on the DataPackageR[3] package. I don't use RStudio but instead use R-Gui with a crude project control setup, and Cntl-R execution of code from separately edited script files.
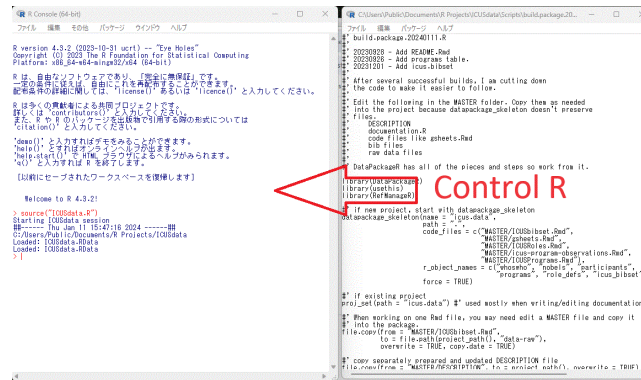

Figure 1: R-Gui setup

DataPackageR has all of the pieces and is compliant with rOpenSci. DataPackageR appears to be optimized for researchers who need to bang out factory quantities of datasets. It is less optimized for independent and unfunded researchers who are working through data objects over a period of months. After much trial and error, I arrived at the workflow below. The code below is NOT intended to be run as a standalone script, but instead run in R-console step by step.

1 Start with the libraries

```
library(DataPackageR)
library(usethis)
library(RefManageR)
```

2 Create or open data package skeleton

2.1 If new, create.

The first command is **RefManageR::datapackage_skeleton**. This wipes out
(initializes) several files so I maintain a MASTER set of components and copy
from it. The listing below gives the code files in RMarkdown plus a list of the
datasets to be included in the package. The Rmd code files are available on the
Github repository for this package. **DataPackageR::package_build** executes
the code files in the sequence listed so I find it convenient to put the code I am
trying to insert in to the project at the top of the list. That way I hit the crash
points early in the build.

Each of these Rmd code files contains code zero or more datasets. **icus-
program-observations.Rmd** is there so I can insert a separately prepared
pdf vignette into the documentation.

```
DataPackageR :: datapackage_skeleton (name = "icus.data",
path = ".",
code_files = c("MASTER/ICUSbib.Rmd",
              "MASTER/gsheets.Rmd",
              "MASTER/ICUSRoles.Rmd",
              "MASTER/vignettes/build-package.Rmd",
              "MASTER/vignettes/search-examples.Rmd",
              "MASTER/vignettes/icus-program-observations.Rmd",
              "MASTER/ICUSPrograms.Rmd"),
r_object_names = c("whoswho", "nobels",
                  "participants", "programs",
                  "role_defs", "icusbib",
                  "viaf"),
force = TRUE)
```

Among several files and folders, *DataPackageR::datapackage_skeleton* cre-
ates a **documentation.R** that contains descriptions for the datasets. This
content is displayed when queried from the R command line, for example **?viaf**.
DataPackageR::datapackage_skeleton overwrites any existing skeleton, so
if you require multiple rebuilds like I do, maintaining this file separately and
copying it into the project is much more convenient.

2.2 If existing, open.

If working on an existing project, such as when working on documentation, the
next line connects to it.

```
usethis :: proj_set (path = "icus.data")
```

When working on one Rmd file, I usually edit the MASTER file and copy it
into the package.
Example:

```
file.copy(from = "MASTER/ICUSbib.Rmd",
          to = file.path(project_path(), "data-raw"),
          overwrite = TRUE, copy.date = TRUE)
```

3 Insert separately maintained code and data files.

I also maintain the DESCRIPTION file separately..

```
file.copy(from = "MASTER/DESCRIPTION",
          to = project_path(), overwrite = TRUE)
```

Several raw data files are provided in **inst/extdata** for researchers who may want to use them. These are copied by this code.

```
file.copy(from = "MASTER/data/ICUSRoles.xlsx",
    to = file.path(project_path(), "inst/extdata"),
    overwrite = TRUE, copy.date = TRUE)
file.copy(from = "MASTER/data/ICUSPrograms.xlsx",
    to = file.path(project_path(), "inst/extdata"),
    overwrite = TRUE, copy.date = TRUE)
#' This master set is in a different R project.
file.copy(from = "[local-path]/ICUSbiblio/bibset-raw",
    to = file.path(project_path(), "inst/extdata"),
    overwrite = TRUE, copy.date = TRUE, recursive = TRUE)
```

Biblography files for the documentation go in the same folder as the Rmd files.

```
file.copy(from = "MASTER/gsheets.bib",
    to = file.path(project_path(), "data-raw"),
    overwrite = TRUE)
file.copy(from = "MASTER/kondo197808diet.bib",
    to = file.path(project_path(), "data-raw"),
    overwrite = TRUE)
file.copy(from = "MASTER/icus.programs.bib",
    to = file.path(project_path(), "data-raw"),
    overwrite = TRUE)


#' write citations for R packages
knitr::write_bib(c("DataPackageR", "googlesheets4",
    "data.table", "readxl", "usethis", "RefManageR", "R.rsp"),
                 file.path(project_path(), "data-raw", "R.bib"),
                 width = 60)
```

4 Now build the package

```
#' don't post this in NEWS.md yet
options("DataPackageR_interact" = FALSE)
DataPackageR::package_build(packageName = project_path(), install = FALSE)
```

Building may require several iterations. The Rmd files are executed in the order listed in the **DataPackageR::datapackage_skeleton** call and **DataPackageR::package_build** stops at the first error. If I am adding a new Rmd file, I place it at the top of the list so the build catches any errors early.

5 Copy in separately maintain documentation files

5.1 documentation.R

package_build completes and displays messages about updating the documentation.R file in the source data-raw subdirectory. This is a major task and this file is overwritten on every build so I maintain a copy in MASTER and edit it separately.

```
file.copy(from = "MASTER/documentation.R",
    to = file.path(project_path(), "data-raw"), overwrite = TRUE)
```

5.2 Separately prepared vignettes

The next section of code is a work-around to place separately prepared pdf files in the vignettes.

5.2.1 Remove interfering html files that were created during the build.

package_build generates html files for each Rmd file, but these interfere with the listing of pdf files in the vignettes index. Remove them.

```
file.remove(file.path(project_path(), "inst", "doc",
                "icus-program-observations.html"))
file.remove(file.path(project_path(), "inst", "doc",
                "build-package.html"))
```

5.2.2 Copy vignettes

The next steps are based on the vignette "Static PDF and HTML vignettes" in the R.rsp package[1]. Copy pdf vignettes and corresponding pdf.asis files into vignettes folder.

```
file.copy(from = "MASTER/icus-program-observations.pdf.asis",
        to = file.path(project_path(), "vignettes"),
        overwrite = TRUE)
file.copy(from = "MASTER/icus-program-observations.pdf",
        to = file.path(project_path(), "vignettes"),
        overwrite = TRUE)
file.copy(from = "MASTER/build-package.pdf.asis",
        to = file.path(project_path(), "vignettes"),
        overwrite = TRUE)
file.copy(from = "MASTER/build-package.pdf",
        to = file.path(project_path(), "vignettes"),
        overwrite = TRUE)
```

Then copy the correponding Rmd and pdf files into the **inst/doc** folder.

```
file.copy(from = file.path(project_path(),
    "vignettes",
```

```
      "icus−program−observations.Rmd"),
      to = file.path(project_path(), "inst", "doc"),
      overwrite = TRUE)
file.copy(from = file.path(project_path(),
    "vignettes",
    "icus−program−observations.pdf"),
    to = file.path(project_path(), "inst", "doc"),
    overwrite = TRUE)
file.copy(from = file.path(project_path(),
    "vignettes",
    "build−package.Rmd"),
    to = file.path(project_path(), "inst", "doc"),
    overwrite = TRUE)
file.copy(from = file.path(project_path(),
    "vignettes",
    "build−package.pdf"),
    to = file.path(project_path(), "inst", "doc"),
    overwrite = TRUE)
```

5.2.3   Copy and render master README.Rmd file.

Finally, copy the master README.Rmd file into the project, render it, and rebuild the documentation.

```
file.copy(from = "MASTER/README.Rmd",
          to = project_path(),
          overwrite = TRUE)
file.copy(from = "MASTER/bibliography.bib",
          to = project_path(),
          overwrite = TRUE)

rmarkdown::render(file.path(project_path(),
                    "README.Rmd"))

DataPackageR::document(project_path())
```

6   Install and admire your work.

The following are some commands to install the new package and look at it.

```
detach(package:icus.data, unload = TRUE, character.only = TRUE)

install.packages(project_path(), repos = NULL,
    type = "source")

install.packages(file.path(getwd(),
    "icus.data_0.6.0.tar.gz"), repos = NULL)

(.packages())
library(icus.data)
```

```
icus.data::programs
?programs
```

References

[1]   Henrik Bengtsson. *R.rsp: Dynamic Generation of Scientific Reports*. R package version 0.45.0. 2022. URL: https://github.com/HenrikBengtsson/R.rsp.

[2]   Carl Boettiger et al. "Building Software, Building Community: Lessons from the rOpenSci Project". In: *Journal of Open Research Software* 3.1 (2015). DOI: 10.5334/jors.bu.

[3]   Greg Finak. *DataPackageR: Construct Reproducible Analytic Data Sets as R Packages*. R package version 0.15.8.9000. 2024. URL: https://docs.ropensci.org/DataPackageR/.