

Preparation for Querying & Accessing Scholarly Literature Metadata Workshop at the Texas Conference on Digital Libraries

Clarke Iakovakis

May 21, 2019

1 Introduction and schedule

Hello everyone. I'm looking forward to our session at [TCDL](#) on Tuesday on "Querying & Accessing Scholarly Literature metadata: Using `rcrossref`, `rorcid`, and `roadoi`." I am Clarke Iakovakis, the Scholarly Services Librarian at Oklahoma State University. The workshop materials are hosted [on my GitHub](#). I'll also be posting them to [my figshare](#) when they're finished.

I would ask that you complete the following if you have time before Tuesday's session. A number of people have signed up for this workshop, and it will be difficult to do troubleshooting before and during. Please [contact me](#) if you have problems with any of this. I plan on getting to the conference center by 7AM, so we can talk then as well. You will be receiving another set of documents and code for working with the code packages during the session. This document is only to help you get set up.

We have 4 hours to get through a lot of complex material. We will take 5-10 minute breaks at the end of every hour to try and catch our breath. Below is a tentative schedule.

8:00-8:50: Getting used to R and the R Studio environment

- What is R? Why R? Assigning values, evaluating expressions, calling functions, data types,

9:00-9:50: Working with data in R

- Data frames, lists, data exploration, installing packages

10:00-10:50: `rorcid`

- Finding ORCID iDs by name and affiliation, understanding what is returned in the API call, getting biographical information, getting researcher works, writing to disk

11:00-12:00: `rcrossref`

- Getting publications by journal, understanding what is returned in the API call, getting article metadata, getting BibTex/RIS files, writing to disk

2 Download R & R Studio

To download R, go to <https://www.r-project.org/>. Click on CRAN (Comprehensive R Archive Network) under Download, and scroll down to your country. Select the download link corresponding to the city that is geographically closest to you.

Go to <https://www.rstudio.com/products/RStudio/#Desktop> to download the RStudio desktop software. RStudio is a user interface for working with R. It is called an Integrated Development Environment (IDE) and acts as a sort of wrapper around the R language. You can use R without RStudio, but it's much more limiting. RStudio makes it easier to import datasets, create and write scripts, and has an autocomplete activated for functions and variables you've already assigned. RStudio makes using R much more effective, and is also free and open source.

3 Install R packages

Open R Studio and go to **File > Open File**. Navigate to the directory for the folder from which you opened this document, and open `setup_code`.

Click on the first line, `install.packages("rorcid")`. Then click the **Run** button in the upper right corner of the window.

It will print some lines of code and make take a few seconds as it installs the package to your R library. If successful, it will tell you the packaged was successfully unpacked. Now run the next line, `library(rorcid)`. As long as it doesn't say "there is no package called 'rorcid'", you should be good.

Do the same thing with the remaining packages.

4 Set up rorcid

If you haven't done so already, create an ORCID account at <https://orcid.org/signin>. If you have an ORCID but can't remember it, search for your name at <https://orcid.org>. If you try to sign in with an email address already associated with an ORCID, you'll be prompted to sign into the existing record. If you try to register with a different address, when you enter your name you'll be asked to review existing records with that name and verify that none of them belong to you—[see more on duplicate ORCID records](#). Make sure you have verified your email address.

Next, you need to authenticate with an ORCID API Key. According to the [ORCID API tutorial](#), anyone can receive a key to access the public API.

Run the line `orcid_auth()` from the `setup_code` script. You should see a message stating: `no ORCID token found; attempting OAuth authentication` and a window will open in

your default internet browser. Log-in to your ORCID account. You will be asked to give `rorcid` authorization to access your ORCID Record for the purposes of getting your ORCID iD. Click “Authorize.”

If successful, the browser window will state: “Authentication complete. Please close this page and return to R.” Return to R Studio and you should see in your R console the word **Bearer**, followed by a long string of letters and numbers. These letters and numbers are your API key. At this point, this should be cached locally in your working directory.

Highlight and copy the API key (the letters and numbers only—exclude the word “Bearer” and the space). Paste it into the quotation marks on the `setup_code` script in the line `ORCID_TOKEN="copy and paste your token here"`. Copy this line to the clipboard.

Click on the line that says `usethis::edit_r_envIRON()` and click the Run button. A new window will open in R Studio. Paste the line into that window. Leave the tab open.

Navigate back to the `setup_code` and run the next `orcid_auth()` line. It should print the token to your R console.

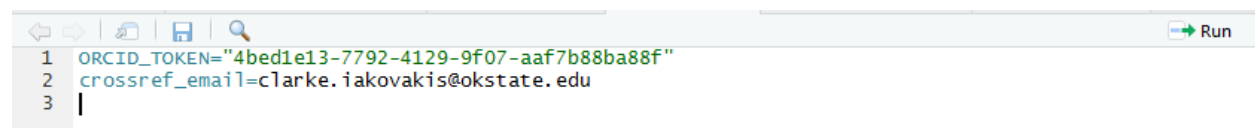
5 Set up `rcrossref`

In the `setup_code` script, replace your email in `crossref_email=name@example.com`. Copy this line to the clipboard.

Go back to the window that opened when you ran `edit_r_envIRON()` and paste in the `crossref_email`.

Then press enter to create a new line below `crossref_email=name@example.com`, and leave it blank.

Your R environment should look like this



```
1 ORCID_TOKEN="4bed1e13-7792-4129-9f07-aaf7b88ba88f"  
2 crossref_email=clarke.iakovakis@okstate.edu  
3 |
```

Click **File** > **Save**. Then in the R Studio navigation menu, click **Session** > **Restart R**.

6 Help with R and R Studio

- [Try R](#) is a browser-based interactive tutorial developed by Code School.
- `swirl` is a package you can install in R to learn about R and data science interactively. Just type `install.packages("swirl")` into your R console, load the package by typing `library("swirl")`, and then type `swirl()`. Read more at <http://swirlstats.com/>.

- [R For Data Science](#) by Garrett Golemund & Hadley Wickham [the best book out there for learning R]
- [Introduction to R for Libraries](#) by Clarke Iakovakis [webinar recording, slides, handouts]
- [R cheat sheets \(all\)](#)
 - [base R cheat sheet](#)
 - [purrr cheat sheet](#)
 - [data transformation cheat sheet](#)
 - [data import cheat sheet](#)