

ORCID US Community Call - Collaboration Data Visualization

- **When:** Thurs. May 4 @ 1-2:30pm Eastern/ 12-1:30pm Central/ 11am-12:30pm Mountain/ 10-11:30am Pacific
- **Recording:**
 - YouTube: <https://youtu.be/kslWPY8SKw4>
 - Vimeo: <https://vimeo.com/824795011>
- **Slides:**  2023-05-04 Community Call - Collaboration Data Visualization
- **Helpful resources:**
 - Tools:
 - RStudio: <https://posit.co/downloads/>
 - Tableau Public: <https://public.tableau.com/app/discover>
 - Open Refine: <https://openrefine.org/>
 - Documentation: <https://github.com/lyrasis/ORCID-Data-Visualization>
 - R script:
https://github.com/lyrasis/ORCID-Data-Visualization/blob/main/Rorcid_Crossref_Authors.R
 - Dashboard documentation:
<https://github.com/lyrasis/ORCID-Data-Visualization/blob/main/tableau-visualization/customizing-dashboard.md>
 - Data Visualization documentation page:
<https://orcidus.lyrasis.org/data-visualization/>
- ORCID US Community Code of Conduct: <https://orcidus.lyrasis.org/events/>

Attendees: (~63 attendees)

- Sheila Rabun (Program Leader for PID Communities, Lyrasis)
- Paolo P. Gujilde (ORCID US Community Specialist, Lyrasis)
- Olivia Given Castello (Temple University)
- Maria Mercedes Tuya (Institute for Advanced Study, Princeton)
- Negeen Aghassibake (University of Washington)
- Hanna Schmillen (Ohio University)
- Ellen Cline (Elon University)
- Drew Wallsh (SUNY System Administration)
- Yingting Zhang (Rutgers University)
- Wendy Shook (Brandeis University)
- Sherry Lake (University of Virginia)
- Angelica Chire (CSU, Chancellor's Office)
- Sharon Shafer (UCLA)
- Matt Carson (Northwestern)
- Cyril Oberlander (Cal Poly Humboldt)
- Reid Boehm (Purdue University)
- Sara Scheib (University of Iowa)
- Colleen Lyon (UT Austin)
- Doug Skillins (Harvard Medical School)
- Belinda Orland (AHA)

- Aaron Retteen (Texas A&M School of Law)
- Natalie Meyers(University of Notre Dame)
- Greta Heng (San Diego State University)
- Angela Clark (University of Miami, Rosenstiel School)
- Lisa Chinn (University of Nebraska Medical Center)
- Dana Peterman (UCLA)
- Margaret Henderson (San Diego State University)
- Eugene Skibinsky (University of Southern California)
- Sandy De Groote (University of Illinois)
- Karen Vaughan (Old Dominion University)

Agenda:

- Welcome - Brief overview of the ORCID Data Visualization project w/ example of final product (Sheila, 10 min.)
- How to create a collaboration map for your own institution:
 - Gathering the data via the R script (Olivia, 25 min.)
 - Download R Studio
 - Download R script and open it in R Studio
 - Gather organization identifiers needed to fill in variables within the script
 - Run the script - tips and tricks, such as using open refine to clean org names
 - Get the final CSV file
 - (Optional/if time allows) Do any further data cleaning you want in OpenRefine or Excel
- Questions about the script? (5 min.)
 - Visualizing the data via Tableau Public (Negeen, 25 min.)
 - Get a Tableau Public account
 - Follow documentation to load CSV file into Tableau (copy template dashboard, load data, etc.)
 - Customization and branding
- Questions about Tableau/the visualization? (5 min.)
- Q&A and discussion (20 min.)

Notes:

Example visualization:

<https://public.tableau.com/app/profile/olivia6835/viz/ORCIDCollaborationsTempleUniversityandHospitalsince2022/Fullvisualization>

The first step is to run the R script, which gathers the following data within a specified time frame:

- Finds all of the ORCID iDs for people who have your organization listed as their current place of employment in their ORCID record
- From those ORCID iDs, collects all of the works that have Crossref DOIs

- Queries the Crossref API to get co-author information from the DOI metadata

If you are not comfortable with R, or if you run into any issues, contact orcidus@lyrasis.org for support.

The R script is available on GitHub: <https://github.com/lyrasis/ORCID-Data-Visualization>

- 1) Gather a listing of organization names and identifiers (Ringgold ID, GRID ID, and ROR) and email domains for the institutions you want to include in your search
- 2) How to find these identifiers:
 - a) Go to <https://ror.org> and search for your organization - this allows you to find the ROR ID and GRID ID
 - b) To find Ringgold ID - you will either need to create a guest account on the Ringgold website, or email orcidus@lyrasis.org and we will help you find the Ringgold IDs that you need
- 3) This list and identifiers will allow you to make a query that will bring back the results that you want
- 4) To get the script - download the entire repository by going to the Code button, download zip file, unzip, and you will see the R script. Also note - there will be an empty "data" folder - that is where the output files from the script will show up
- 5) Download R studio (free software)
- 6) Open the script in R studio
- 7) There are many detailed comments in the script that give you instructions on what to do (code comments start with # to indicate that it is a comment, not a command)
 - a) To uncomment so that you can run a line, remove the #
- 8) There are a few things you will need to do the first time, and once done you will not need to do them again:
 - a) Get your Public API client id and client secret (<https://info.orcid.org/documentation/features/public-api/>)
 - b) Install the packages in R on your computer
 - c) Get your ORCID access token - store it in R (there are detailed instructions in the code comments in the script). Will also need to enter an email address for the Crossref API
 - d) Then restart R
- 9) Plug in the values for start year, institution name, email domain, identifiers, location, and keyword
 - a) If your institution has common words in the name, you may want to use the entire organization name as the keyword, for example the keyword "New" would not be helpful for an org such as "The New School" because multiple organizations have the word "New" in the name
- 10) You can choose to do a simple query (choice 1) or if you want to include multiple identifiers/org names, do choice 2 - if you use choice 2, you need to edit the query a bit - copy the lines and edit to fill in additional variables for the different org names/identifiers.

- 11) The script has various sections - there are opportunities for you to export the data so far, and write it back in later without having to run the whole script again...
- 12) There is a point in the script to see what the different organization names are that have been returned from the script - you can view the org names and make edits to them - alot of times there will be typos, extra spaces, and other variations so you can limit the search further to only include the org names that you want...
- 13) Once you have the CSV output, you may want to clean the org names and city names in Open Refine <https://openrefine.org/>

Once you have the CSV file from the script:

- Tableau Desktop is paid (lets you save files locally).
- Tableau Public is free, but everything that you save is saved to the Tableau Public server.
 - You can publish a dashboard in Tableau Public and hide it from searching
 - Some customization options are limited, so if you want to add anything to the dashboard or use it more robustly, we recommend using Tableau Desktop
- To get started, you need a Tableau Public account
- Follow the instructions here:
<https://github.com/lyrasis/ORCID-Data-Visualization/blob/main/tableau-visualization/customizing-dashboard.md>
 - The Tableau template is only able to recognize the data as it comes out of the script - do not change any of the field names or variable names in the CSV output file, otherwise the Tableau template will not recognize it
 - Make sure your CSV file is named "orcid-data"
 - Tableau offers worksheets (for individual single visualizations), dashboards (consist of multiple visualizations), and stories (can contain both visualizations and dashboards)
 - Customizations will be done in worksheets (can't really make any changes in a story)
 - Important to note the data that data was pulled - the script pulls data from Jan. 1 of the year entered in the script, to the present data that the script was run, so it's important to note the date that the script was run...
 - To add your organization's logo - we will add these instructions to the dashboard customization instructions
- Once the visualization dashboard is published on your profile:
 - You can show sheets as tabs
 - You can show the visualization on your profile - makes it visible and findable, otherwise you can only share by sharing the link
 - At the bottom, you can edit details, such as the title, or additional information about where your users can reach out for help, etc.
 - Save changes
- Instructions also include accessibility resources - Tableau has improved accessibility over the years

- Geocoding may have errors or ambiguous data - it is possible to clean up the geographic info in Open Refine in the CSV, or using tools within Tableau (info included in documentation)

Additional notes:

- The map and dashboard can be used as a tool to show the importance of ORCID records having complete information, the importance of ORCID member organizations adding works and other information (like employment) to peoples' ORCID records
- Ideally, information will get added to ORCID records by organizations via the ORCID member API

Questions:

Q: Sheila and Paolo are you able to assist institutions outside the US?

- **A:** We are primarily focused on institutions within the US, but these resources are designed to be used by anyone at any organization across the globe, so please do email us at orcidus@lyrasis.org and we will do our best to support you.

Q: Would this work for a funder?

- **A:** The full script starts by finding the ORCID iDs for all employees at an organization, and then goes from there. For a funder, you may already have a list of ORCID iDs for researchers that are not necessarily employees at your organization. We would be more than happy to work with you to experiment on making adjustments to the script for this purpose (just email us)

Q: Why does one have to register to read the public API? I thought tokens were only needed for trusted (records).

- **A:** The Public API requires a general access token to query the ORCID registry, so that ORCID can have an idea of Public API usage. Meanwhile, the Member API requires a separate access token for each person/ORCID iD in order to read limited access data and/or write data to ORCID records.

Q: Would a special character in the keyword string break things? Specifically, I have an & in our name string and that some times makes scripts unhappy?

- **A:** It shouldn't, because the keywords are contained in quotes

Q: Does the initial API query only include works that are listed on ORCID profiles?

- **A:** the script starts by gathering only works that have Crossref DOIs, that are included on researchers' ORCID records. It then goes to the Crossref API to gather co-author information from the Crossref DOIs

Q: Question about time measurements - interested in changes over time? Would be really cool to see how those geographical bubbles change over time

- **A:** that is one of the top priority things for future development. Additional time considerations are not included in the script/visualization right now, simply due to time constraints. For a work-around you could just run the script for a certain time frame and then delete additional years data that you don't want. There is also a way to show time differences in Tableau with Tableau customizations.

Chat Log:

10:01:40 From Paolo Gujilde to Everyone:

Agenda and Notes:

https://docs.google.com/document/d/1jVB_ecjUlptUx5dsMUKAvwKhC_Uhr-FmiEGmsCpX82U/edit

10:04:43 From Paolo Gujilde (Lyrasis) to Everyone:

Hello, All!

10:05:21 From Paolo Gujilde (Lyrasis) to Everyone:

10:05:49 From Paolo Gujilde (Lyrasis) to Everyone:

Please feel free to add you name. Thanks!

10:11:16 From Paolo Gujilde (Lyrasis) to Everyone:

Data Visualization Documentation page: <https://orcidus.lyrasis.org/data-visualization/>

10:11:28 From Paolo Gujilde (Lyrasis) to Everyone:

GitHub page: <https://github.com/lyrasis/ORCID-Data-Visualization>

10:13:30 From Paolo Gujilde (Lyrasis) to Everyone:

email us at Orcidus@lyrasis.org

10:13:36 From Sheila Rabun (she/her) to Everyone:

Reacted to "email us at Orcidus@..." with 👍

10:13:58 From Sharon Shafer to Everyone:

Is there a readme?

10:14:13 From Sharon Shafer to Everyone:

Oh yes - I see it. cool

10:14:20 From Georgina Alsop to Everyone:

Sheila and Paolo are you able to assist institutions outside the US?

10:16:28 From Sheila Rabun (she/her) to Everyone:

Replying to "Sheila and Paolo are..."

Hi Georgina, we are primarily focused on institutions within the US, but these resources are designed to be used by anyone at any organization across the globe, so please do email us and we will do our best to support you.

10:17:11 From Georgina Alsop to Everyone:

Replying to "Sheila and Paolo are..."

Thank you!

10:17:44 From Belinda Orland to Everyone:

Replying to "Sheila and Paolo are..."

Would this work for a funder?

10:18:35 From Paolo Gujilde (Lyrasis) to Everyone:

You can also check our Agenda and Notes for links and other resources:

https://docs.google.com/document/d/1jVB_ecjUiptUx5dsMUKAvwKhC_Uhr-FmiEGmsCpX82U/e_dit

10:19:18 From Sheila Rabun (she/her) to Everyone:

Replying to "Sheila and Paolo are..."

@Belinda Orland - the full script starts by finding the ORCID iDs for all employees at an organization, and then goes from there. For a funder, you may already have a list of ORCID iDs for researchers that are not necessarily employees at your organization. We would be more than happy to work with you to experiment on making adjustments to the script for this purpose (just email us)

10:20:47 From Paolo Gujilde (Lyrasis) to Everyone:

R script:

https://github.com/lyrasis/ORCID-Data-Visualization/blob/main/Rorcid_Crossref_Authors.R

10:21:14 From Belinda Orland to Everyone:

Replying to "Sheila and Paolo are..."

Thank you, Sheila!

10:23:52 From Sherry Lake to Everyone:

Why does one have to register to read the public API? I thought tokens were only needed for trusted (records).

10:30:37 From Sheila Rabun (she/her) to Everyone:

Replying to "Why does one have to..."

Hi Sherry, my understanding is that the Public API requires a general access token to query the ORCID registry, so that ORCID can have an idea of Public API usage. Meanwhile, the Member API requires a separate access token for each person/ORCID ID in order to read limited access data and/or write data to ORCID records.

10:37:19 From Sherry Lake to Everyone:

Replying to "Why does one have to..."

So a API token is not necessarily needed to do a public query? Only "wanted" by ORCID to keep track of usage?

Maybe my real question is do we have to go through the registration process to run this code?

10:38:05 From Aaron Retteen to Everyone:

Would a special character in the keyword string break things? Specifically, I have an & in our name string and that some times makes scripts unhappy

10:43:40 From Paolo Gujilde (Lyrasis) to Everyone:

Tableau Public: <https://public.tableau.com/app/discover>

10:45:11 From Olivia Castello to Everyone:

Replying to "Why does one have to..."

Yes, you do. It's expected by the library the script uses (rorcid) to communicate with the ORCID API

10:45:56 From Olivia Castello to Everyone:

Replying to "Would a special char..."

I don't think it would. That said I don't think I have tried an & in the keyword, so it's possible.

10:47:08 From Paolo Gujilde (Lyrasis) to Everyone:

Dashboard documentation:

<https://github.com/lyrasis/ORCID-Data-Visualization/blob/main/tableau-visualization/customizing-dashboard.md>

11:05:44 From Sheila Rabun (she/her) to Everyone:

Email us for any support you may need: Orcidus@lyrasis.org

11:09:46 From Maria Aghazarian to Everyone:

Yes, definitely will try to create. there may be a bit of a learning curve for me personally though haha

11:10:08 From Sharon Shafer to Everyone:

Yes, we're going to try!

11:11:13 From Aaron Retteen to Everyone:

Thanks for thinking through all of this, putting it all together, and sharing your work for us all to benefit from! There is so much potential with the open data ecosystem

11:12:34 From Reid Boehm to Everyone:

I think so, yes. It would be interesting to look at it for the data repository (PURR). It seems like it might help to see where collaborations are happening. We'd have some gaps, but a lot of people are starting to enter them.

11:20:00 From Dan Coughlin to Everyone:

I have to leave a bit early, echoing Aaron's message, really appreciate this presentation and you putting this out there for everyone.

11:20:39 From Sharon Shafer to Everyone:

Thanks for all this wonderful work!

11:21:01 From Reid Boehm to Everyone:

Thank you all so much!

11:21:01 From Deborah Caldwell to Everyone:

Thanks so much!

11:21:08 From Olivia Castello to Everyone:

Thanks for coming everybody 😊

11:21:13 From Matt Carson to Everyone:

Thank you!

11:21:19 From Paolo Gujilde (Lyrasis) to Everyone:

Thank you, All!

11:21:40 From Paolo Gujilde (Lyrasis) to Everyone:

Thank you, Olivia and Negeen!

