

rOpenSci Community Call - Reproducible Research with R

This doc - http://bit.ly/commcall_reproresearch - is a place for collaborative note taking and adding your questions during the call. This call will be recorded and posted along with any other resources within two business days at <https://ropensci.org/commcalls>.

Information on today's topic, with speaker bios:

<https://ropensci.org/blog/2019/07/11/commcall-jul2019/>

Agenda

1. Stefanie Butland, rOpenSci Community Manager - welcome (3 min)
2. Ben Marwick on research compendia (10 min)
3. Karthik Ram on his holepunch package (10 min)
4. Anna Krystalli on ReproHacks (10 min)
5. Q & A (moderated by Karthik Ram, 20 min)

Speakers' slides

Links will be added here and on the [archive page for the Call](#)

Ben Marwick's [slides](#)

Karthik Ram's [slides](#)

Anna Krystalli's [slides](#)

In attendance (please add yourself here):

- *Name, Affiliation*
- Stefanie Butland, rOpenSci
- Ben Marwick, U Washington
- Karthik Ram, Berkeley Institute for Data Science and rOpenSci
- Anna Krystalli, U Sheffield and rOpenSci Software Peer Review
- Ju-Chi Yu, The University of Texas at Dallas
- Florencia D'Andrea, National Agricultural Technology Institute of Argentina.
- Will Landau, Eli Lilly and Company
- Noushin Nabavi, Digital Platforms and Data Division, Ministry of Citizens' Services (BC Gov), Victoria, British Columbia

- Gail Clement, Caltech Library and Author Carpentry
- Tom Morrell, Caltech Library and Author Carpentry
- Kari L. Jordan, The Carpentries
- Julie Galloway, Ministry of Environment (BC Gov) + University of Victoria (British Columbia)
- Harrison Dekker, University of Rhode Island Libraries
- Evan Simpson, The British Council, Shanghai, PRC
- Marcus Beck, Southern California Coastal Water Research Project
- Scott Chamberlain, ropensci
- Amanda Devine, Smithsonian Institution, Washington DC
- Julie Hawkins and Stephanie Yurchak, BC Stats (BC Gov)
- Hugo Gruson, University of Montpellier
- Genevieve Perkins, Ministry of Environment (BC Gov), Canada.
- Liliana Paloma Rojas Saunero, Erasmus MC, Rotterdam - Netherlands
- Kathleen Wendt, Colorado State University (Fort Collins, CO, USA)
- Marion Louveaux, University of Heidelberg (Germany)
- Clemens Schmid, University of Bern
- Mike Trizna, Smithsonian Institution, Washington, DC
- Kara Woo, Sage Bionetworks
- Hannah Frick, Mango
- Pierre Montagano, Code Ocean
- Mauro Lepore, 2 Degrees Investing Initiative
- Pierrette Lo, Oregon Health & Science University
- Rhiannon Cameron, Unaffiliated (Recent University of Victoria Graduate)
- Holly Deniston-sheets, Data Analysis and Tactical Operations Center ([DATOC](#))
- Donna Wrublewski, Caltech
- Jesse Sadler, Loyola Marymount University
- Marieke Jones, University of Virginia, RLadies Charlottesville
- Kelly Hondula, University of Maryland & SESYNC
- Peter Higgins, University of Michigan
- Andy Teucher, Ministry of Environment (BC Gov), Canada
- Mara Alexeev, At home with baby soon to be at Memorial Sloan Kettering in NYC
- Nasim Taba, BC Stats (BC Gov), Canada
- Dimitri Bourilkov, University of Florida
- Sina Rüeger, Swiss Federal Institute of Technology in Lausanne, Switzerland
- Greg Botwin, Cedars-Sinai Medical Center
- Rachael Blake, SESYNC, University of Maryland
- Steven Hodge, University of Massachusetts Medical School, Worcester MA
- Daniela Gawahns, Leiden University, Netherlands
- Antoine Thibaud, Uruguay
- John Blischak, University of Chicago
- Neal Fultz - DeclareDesign.org
- Roxana Noelia Villafañe, National University of Northeast - CONICET (Argentina)

- Sam Albers, Digital Platforms and Data Division, Ministry of Citizens' Services (BC Gov), Victoria, British Columbia
- Christopher Callaghan, Naval Postgraduate School
- Kate Nyhan, Yale
- Michelle Koo, Museum of Vertebrate Zoology, UC Berkeley
- Helen Miller, Fred Hutch
- Jeroen Ooms, UC Berkeley
- Martin Monkman, BC Stats, Province of British Columbia, Victoria, British Columbia
- Ryan Ames, FoodCorps
- Mladen Cucak, Maynooth University, Ireland
- Kim Cressman, Grand Bay National Estuarine Research Reserve
- Jeff Newmiller, DNV GL
- Liying Wang, University of Washington
- Corinne Bassin, NOAA
- Christine Stawitz, NOAA
- Jennifer Huck, University of Virginia Library
- Bianca Peterson, North-West University, South Africa
- Yanina Bellini SAibene, Instituto Nacional de Tecnología Agropecuaria (INTA), Argentina
- Matt Mulvahill, Charter Communications
- Vijay Barve, Florida Museum
- Adam Batten, Veterans Health Administration, Seattle
- Tomomi Suwa, Field Museum, IL
- Andrea Sánchez-Tapia, Rio de Janeiro Botanical Garden
- Ju Kim, Fred Hutch
- Micaela Chan, UT Dallas

Questions for speakers (add yours here as you think of them):

Please include your name with your question. The Moderator will call on you to ask your question, unless you note you prefer to have the moderator ask on your behalf

- Gail Clement, Caltech: do you have licensing recommendations around making the objects in the compendium, and the compendium as a whole, reuseable and able-to-be-modified with confidence and rights-clarity? The Research Data Alliance Legal IG is looking at the Research Compendium model for legal interoperability and not sure what licensing scheme you feel is appropriate for heterogeneous content inside the compendium as a whole...and also whether you feel a single 'omnibus' license could be attached for the compendium as a whole. If the latter is the case, what features would the 'omnibus license' carry?

- Stodden advice: license each component of the compendia with a relevant open license, eg MIT or GPL for code; CC-BY for narrative, figures; CC0 for data
- Stodden citation: Stodden, Victoria, Enabling Reproducible Research: Open Licensing for Scientific Innovation (March 3, 2009). International Journal of Communications Law and Policy, Forthcoming. Available at SSRN: <https://ssrn.com/abstract=1362040>
- Thank you!!! :-) RDA-CODATA Legal Interop is also supporting this approach so now needs to outreach to repositories and other platforms to accommodate licensing fields with multiple rights statements: JISC is demoing this in their repo. #progress!
- Peter Higgins, UMich: Is there a screencast(s) for best practices in using rrtools / holepunch for newbies, i.e. a mock analysis of the iris or mtcars or gapminder dataset?
 - <https://github.com/annakrystalli/rrtools-repro-research> 😍
 - BM: here are slides I've used in a few workshops that give a hands-on walk-through of getting started with rrtools:
<https://benmarwick.github.io/Marwick-CSSS-May-2019-Reproducibility/Marwick-CSSS-May-2019-Reproducibility-Workshop.html#1>
- Sam Albers, bcgov: Could Ben/Karthik/Anna comment on when they think introducing this compendium work flow during someones learning cycle with R should be? Does anyone have any experience teaching this to beginners?
- Daniela Gawehtns, Leiden Uni: What if Binder/mybinder.org runs out of funds?
 - For setting up your own binderhub, have a look at the materials from the Turing Way “Build a BinderHub” workshop:
<https://github.com/alan-turing-institute/the-turing-way/blob/master/workshops/build-a-binderhub/workshop-presentations/zero-to-binderhub.md>
 - Also resources to automatically launch a binderhub on Azure:
<https://github.com/alan-turing-institute/binderhub-deploy>
- To Anna: Have you seen the [unconf toolbox](#) (in active development by Angela Li and Sam Tyner, funded by R Consortium)? It might make sense for ReproHacks to be listed/connected. Mentored by Brooke Anderson (rOpenSci editor), I am planning a pilot “collab” using the toolbox at Colorado State University in April 2020. Angela is planning a sister event in Chicago in March 2020. (From Kathleen Wendt)
 - No I wasn't aware of this! Looks great, will make sure to look into it properly 🙏

Add your notes and resources here:

- [Research Compendium resources](#) curated by Daniel Nüst, Ben Marwick, Carl Boettiger

- From John Blischak: R has tons of great workflow packages
<https://jdblischak.github.io/ctv-project-workflows/>
- The [holepunch](#) package
- [A few slides from today's community call on the holepunch section.](#)
- Longer talk on research compendia and binder: <http://inundata.org/talks/rstd19/#/>
- [ReproHacks](#)
 - Newsletter: <https://tinyletter.com/rephack-hq>
- AuthorCarpentry *Reproducible Reporting and Research Compendium* short courses at Force11 Scholarly Communication Institute (Daniel Nust remotes in for part of the discussion on the Compendium)
 - 2018 version: <https://authorcarpentry.github.io/FSCI-2018/>
 - 2019 version (to be offered next week!):
https://docs.google.com/document/d/1KF2i1L2Dx-gPoac4GHaLa0TwrmZ4PxGD_TzFhY-SwvE0/edit#heading=h.clIn6ws8o5du;
<https://authorcarpentry.github.io/FSCI-2019/>
 -
- [ReScienc C](#): An online journal that publishes replication studies
- [Scigen.report](#): webportal for logging the status of reproducibility efforts
- CODESH: a tool for reproducibility/sharing of general command line (shell) work:
<https://sourceforge.net/projects/codesh/>
- [unconf toolbox](#) in active development by Angela Li and Sam Tyner (funded by R Consortium)
- We are planning a local ReproHack in the Netherlands (Paloma Rojas Saunero was also on the call and is on the team, and obviously Anna and Florencia are helping with the tech/ templates). Reach out if you are planning your own, we should share experiences (promotion, obstacles, funding possibilities, set-up -- all very hands-on)! Contact: gawehnsd@liacs.leidenuniv.nl
 - 
- A question was asked about teaching a compendium workflow approach to beginners. Although not specific to R, Project TIER has been doing good work in this area for a number of years and has training materials that have been used in undergraduate curricula. TIER has a social science focus, but their approach is general enough for use with most research projects involving code, data, and documentation. In the past, they've also offered teaching fellowships to train interested instructors.
<https://www.projecttier.org/>

What would you like to hear about on a future rOpenSci Community Call?

Past Calls: <https://ropensci.org/commcalls/>

Next Call: Reproducible workflows at scale with drake, with drake's developer Will Landau.
Tuesday, September 24th at 9am Pacific.

Other ideas we're considering

- Geospatial analysis in R with Antarctic & Southern Ocean research as example
- Learning from other orgs' Equity, Inclusion, and Accessibility Roadmaps
- Testing
- Finding and reading source code
- Make the most of GitHub for package development and R ecosystem knowledge
- Live code review for advanced developers
- Maintaining a package
- rOpenSci tools to access academic literature
- Text processing & analysis featuring rOpenSci tools

Tell us here what would you like to hear about on a future rOpenSci Community Call. Add your name, if you don't mind.

- Testing and training data considerations for machine learning (e.g. for tree-based modeling) - Noushin Nabavi
- Promoting R and reproducible research in social and behavioral sciences (resistant SPSS users) (Kathleen Wendt, Colorado State University)
- Workflowr as a workflow tool?
- Best practices for administering a (remote) R based analysis team (Greg Botwin, CSMC)
- A detailed tutorial about how to make your paper reproducible
- Tips on licensing open source materials and code: A licensing 101 discussion if you will, maybe with material that people on the call can take back to their groups/ institutes/ friends and colleagues?
- Accessing data via an API (e.g. the spocc package). Interested in hearing about coding best practices for data users as well as package developers, reproducibility issues, and any insights ROpenSci can share based on their experience developing (and using) R packages for this purpose. (Harrison Dekker, URI)