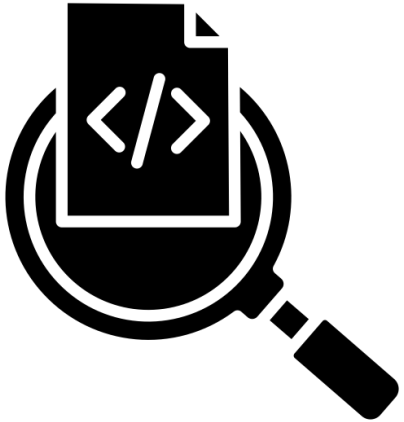
Code peer review

towards more reproducible and reliable research



Stefan Vriend & Joey Burant
Netherlands Institute of Ecology (NIOO-KNAW)

Open Science Community-Wageningen (OSC-W)
Lunch seminar, 23 November 2023

NIOO NETHERLANDS INSTITUTE OF ECOLOGY (NIOO-KNAW)

OSCW

Access today's workshop materials on the
Open Science Framework

https://osf.io/5ykhq

Who are we?

Stefan Vriend

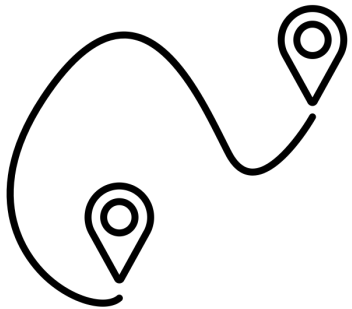
Coordinator Veluwe, LTER-LIFE
Manager & Developer, SPI-Birds
@StefanJGVriend

Joey Burant

Tenure-track Researcher,
Department of Animal Ecology
@jbburant

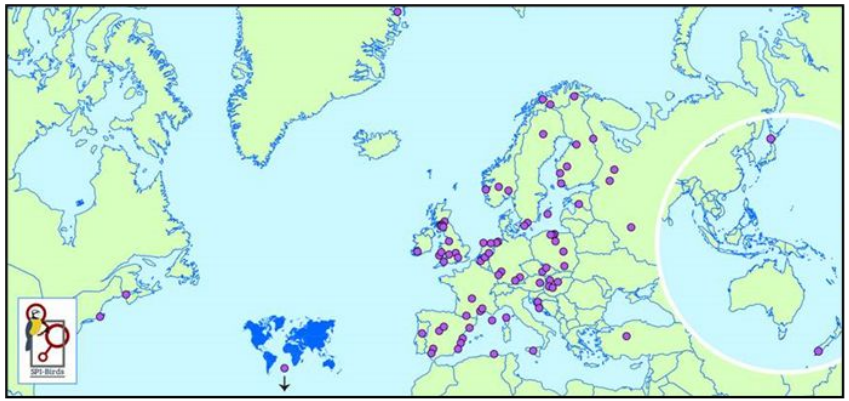
Rough schedule

- 12:00 - 12:12 — Introduction + the basics of code peer review
- 12:12 - 12:15 — Explanation of interactive session + lingering questions
- 12:15 - 12:50 — Break-out session
- 12:50 - 13:00 — Discussion + concluding remarks



SPI-Birds — data and code standards for bird studies

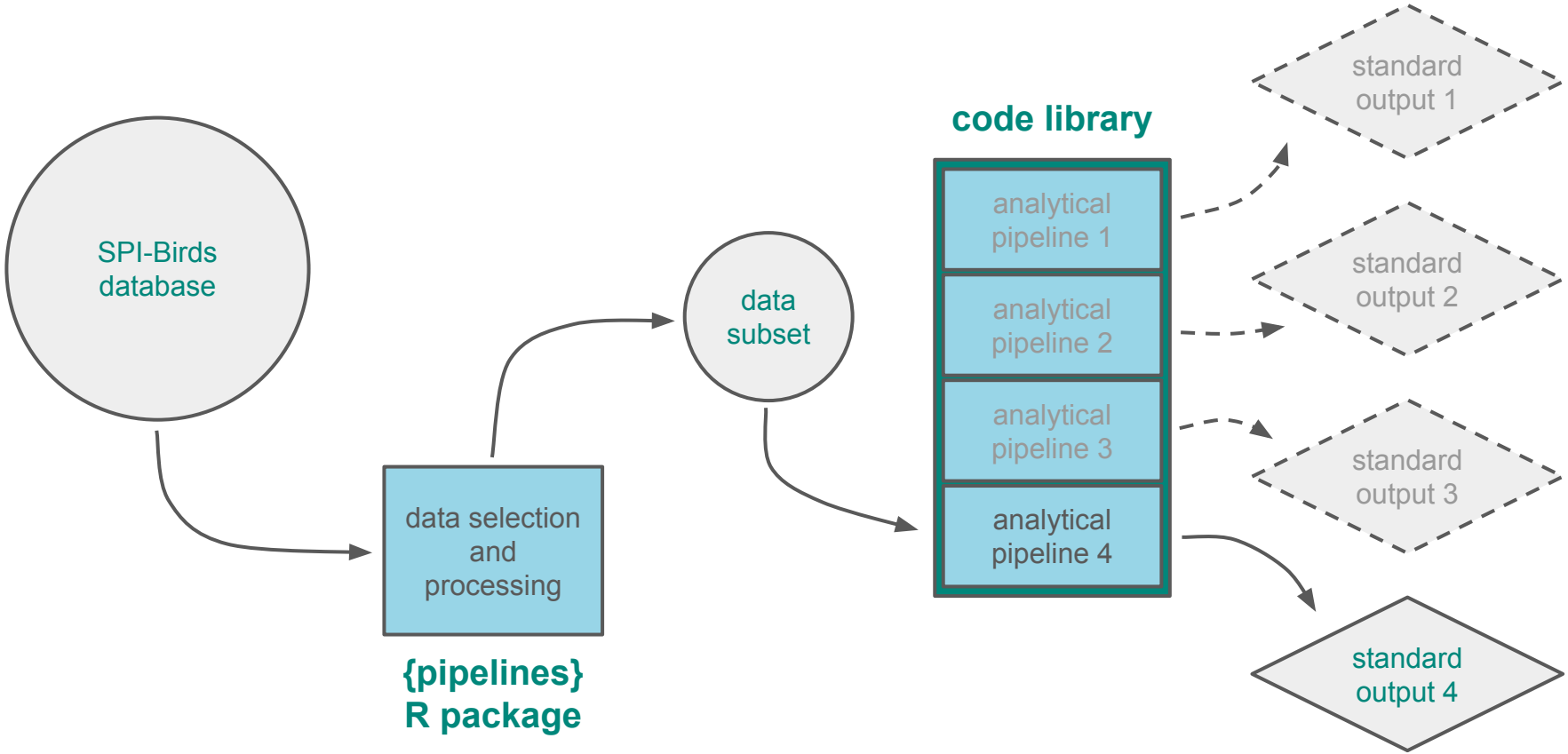
- Community-led metadata repository and data standard for studies of of individually-marked breeding birds
- Aims to promote increased data re-use and synthesis
- **>200** populations, **117** studies, **34** species, **25** countries



SPI-Birds

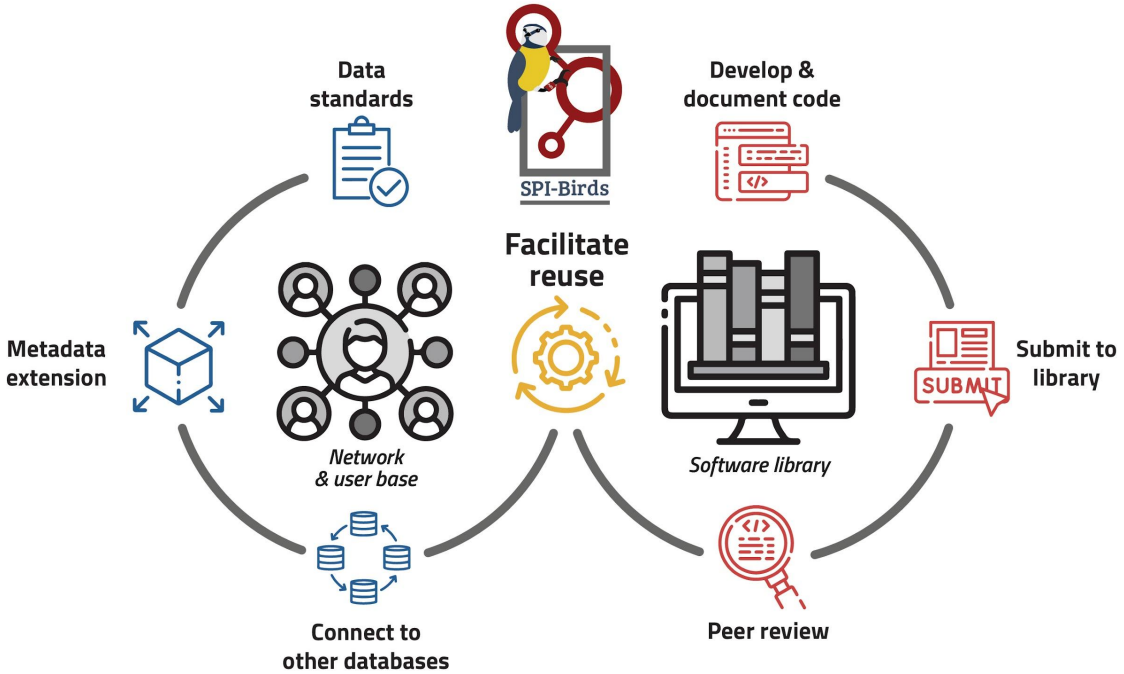
https://spibirds.org

Why are we interested in code peer review?



CoreBirds — Connecting Open Research outputs in the Ecology of Birds

"...create a library of data processing and analytical codes, and implement a peer review process... ensuring the codes are themselves standardised and applicable to any SPI-Birds dataset."



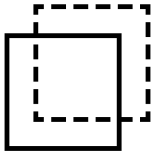
https://doi.org/10.53962/m95v-7drc

Why should you be interested in code peer review?

Minimise potential conceptual, programmatic, and syntactic **errors**

Improve **reliability** and **quality** of your analysis and manuscript

Contribute to **transparent** and **reproducible** science

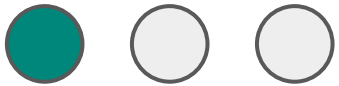


Support and learn from colleagues; contribute to a culture of collaboration

Ivimey-Cook et al. (2023), *J. Evol. Biol.*

Sharing code for transparent and reproducible science

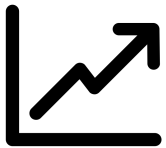
< 1 in 3 articles are accompanied by code to reproduce the outputs



~75% of ecology journals have policies encouraging/mandating code sharing

Community service; the target audience finds value in seeing code

Boosts impact; code sharing increases citation rates



Culina et al. (2020), PLoS Biol.; Maitner et al. (2023), *pre-print*

Sharing code for transparent and reproducible science

< 1 in 3 articles are accompanied by code to reproduce the outputs

~75% of ecology journals have policies encouraging/mandating code sharing

the most useful code is *correct*, well-documented, and openly shared!

Community service; the target audience finds value in seeing code

Boosts impact; code sharing increases citation rates

Culina et al. (2020), PLoS Biol.; Maitner et al. (2023), *pre-print*

Implementing code review in the scientific workflow

Ensuring code matches the reported methods is imperative to evaluate whether the code does what is/was intended

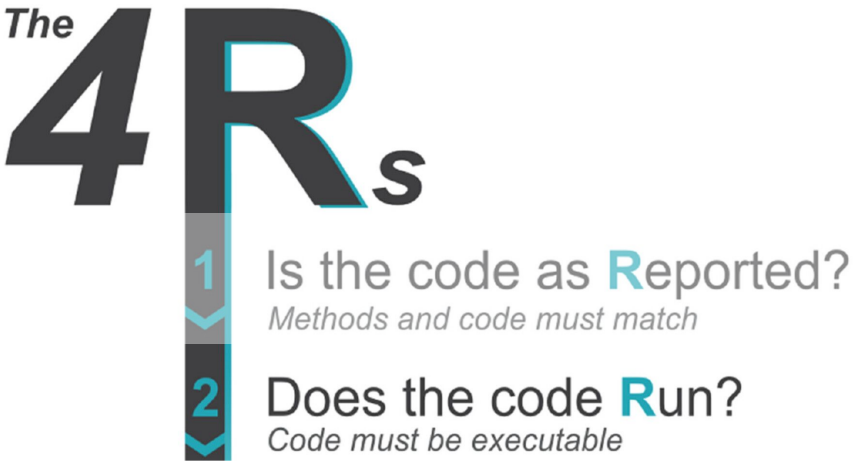
Common issues

- data processing
- model specification
- packages used (versions)

Ivimey-Cook et al. (2023), *J. Evol. Biol.*

Implementing code review in the scientific workflow

Clean code should run without error, and where warnings are expected they should be clearly annotated within the script



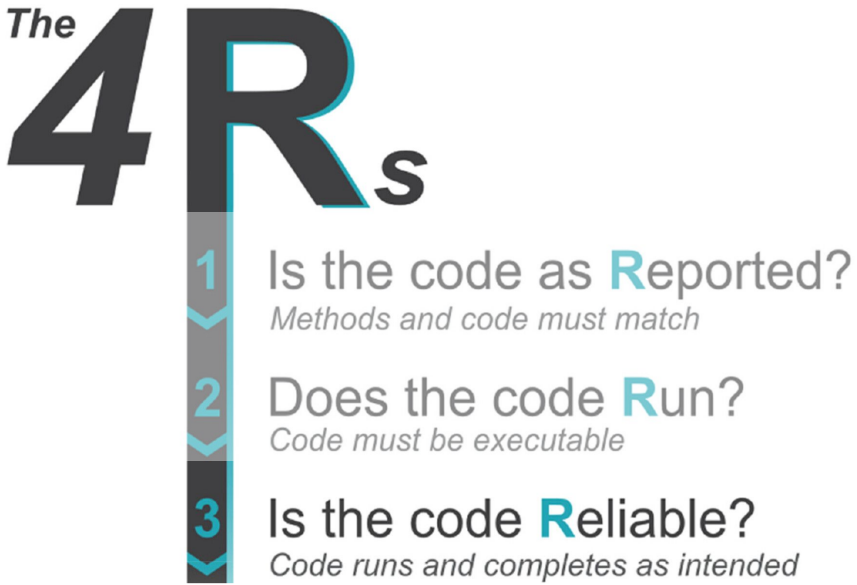
Common issues

- data availability
- missing packages/functions
- simulations
- run-time

Ivimey-Cook et al. (2023), *J. Evol. Biol.*

Implementing code review in the scientific workflow

Reliable code and data processing builds in quality checks (unit tests, QA/QC) throughout, ensuring that the intermediate and final outputs are not just similar, but identical



Common issues

- hard-coding / indexing
- copy-paste-replace approaches
- lack of unit tests
- missing steps

Ivimey-Cook et al. (2023), *J. Evol. Biol.*

Implementing code review in the scientific workflow

The code should (near-)identically produce all outputs as reported in the paper's analysis and results sections, including the figures, tables, and other supporting details

Common issues

- no specified random seed
- imprecision in reporting
- qualitative but not quantitative similarity

Ivimey-Cook et al. (2023), *J. Evol. Biol.*

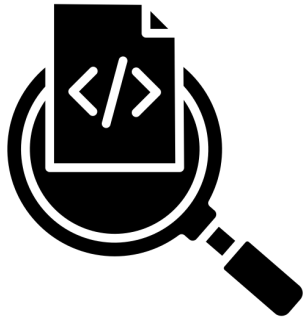
Other considerations

- The 4 Rs are broad and cover only the barest of minimums
- What are some other things we might check for? What makes code (re-)usable?
  - think about the parallels between data and code sharing/re-use

Suggestions

- metadata (!!!!)
- consistent style (there are many different code style guides; pick one!)
- documentation
- efficiency
- succinctness (functional programming, iteration, ~~copy-paste-replace~~)

Today's interactive session



- Working by yourself or in a pair, you have ~35 minutes to:
  1. select a paper and corresponding code to review
  2. read the methods briefly, with a particular eye for the data selection and analytical steps
  3. download the data and code
  4. open the script in R/RStudio (or elsewhere) — you need to be able to run the code!
  5. work through our suggested 14-step code review checklist
  6. brainstorm other checks you think are relevant

- All materials (manuscript, data, and code) are available on the OSF project, as well as from the original sources

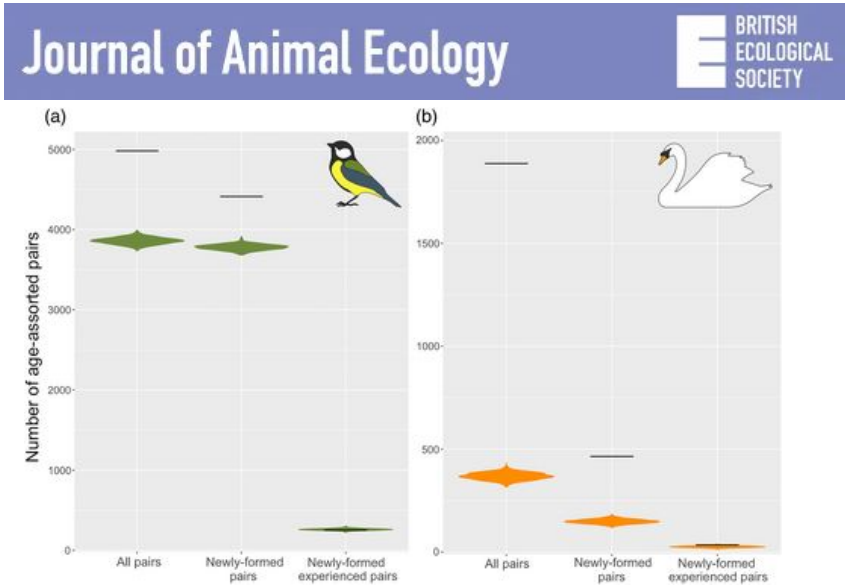
OSF

Access today's workshop materials on the
Open Science Framework

https://osf.io/5ykhq

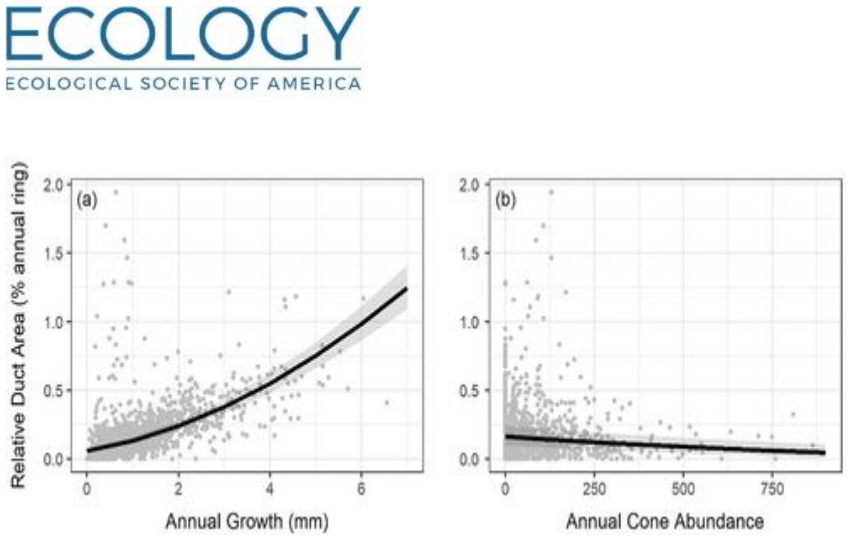
Today's interactive session

Option 1: assortative mating in bird populations with contrasting life-history strategies



Woodman et al. (2022), *J. Anim. Ecol.*

Option 2: aridity associations with ponderosa pine vital functions

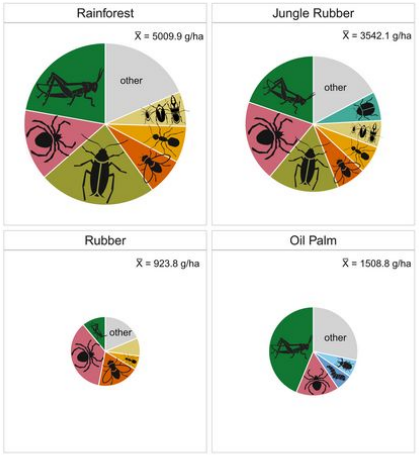


Gonzalez et al. (2023), *Ecology*

Today's interactive session

Option 3: energy fluxes and functions in canopy arthropod food webs

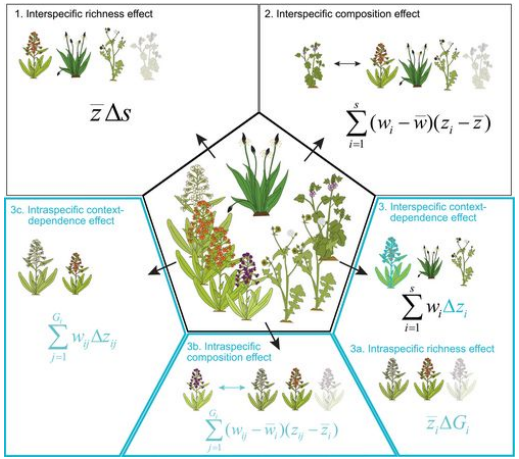
ECOLOGY LETTERS



Pollierer et al. (2023), *Ecol. Lett.*

Option 4: interspecific and intraspecific diversity effects on ecosystem functioning

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA



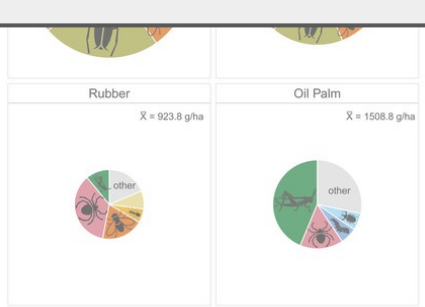
Govaert et al. (2023), *Ecology*

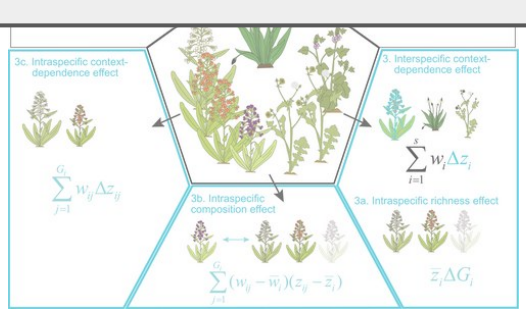
Today's interactive session

Option 3: energy fluxes and functions in canopy arthropod food webs

Option 4: interspecific and intraspecific diversity effects on ecosystem functioning

Disclaimer: we have selected these options basically at random; their inclusion here is not a judgement of their quality. We appreciate that the authors have publicly shared their data and code, making this workshop possible! Be kind.





Pollierer et al. (2023), *Ecol. Lett.*

Govaert et al. (2023), *Ecology*