

Herramientas prácticas para investigación reproducible

CICLO: I- 2023

SIGLAS: B-676 / SP-8225

CREDITOS: 4

CUPO: 12 de pregrado y 5 de posgrado

ESTUDIANTES: pregrado y posgrado

HORARIO: 4 horas por semana (jueves 13-16:50)

MODALIDAD: presencial

REQUISITOS: B-0304 (Ecología General) y B-0305 (Laboratorio de Ecología General)

PROFESOR:

Marcelo Araya Salas; correo: marcelo.araya@ucr.ac.cr

Página web: <http://marceloarayasalas.weebly.com>

Oficina: #3 Centro de Investigación en Neurociencias

MEDIACION VIRTUAL:

Se utilizará la plataforma Mediación Virtual para que las y los estudiantes tengan acceso a todos los elementos del curso y para la entrega de tareas correspondientes.

HORAS DE ATENCIÓN: L: 11-12:00 / J: 11-12:00 (o a convenir con el profesor)

DESCRIPCIÓN DEL CURSO:

Vivimos en un periodo único en la historia de la producción científica en el que se generan datos y publicaciones a una tasa nunca antes vista. Esta mega producción se debe tanto al crecimiento en el número de científicas y científicos a nivel mundial, como a los avances tecnológicos que permiten tomar datos de formas cada vez mas eficientes. Sin embargo, una grave desventaja de este crecimiento es que ha generado incentivos profesionales que en ocasiones priorizan el volumen científico sobre su calidad. Uno de los efectos mas permisivos es la **poca reproducibilidad, producto de la escasa documentación de los métodos y análisis utilizados, así como de la falta de acceso a los datos generados**. La biología no ha escapado de esta llamada crisis de reproducibilidad científica, que afecta la credibilidad de nuestra comunidad. Por suerte, en respuesta a esta crisis se ha desarrollado una multitud de herramientas, en la mayoría de los casos libremente disponibles, para aumentar la transparencia y accesibilidad de datos que respaldan las conclusiones de los estudios científicos. El poder hacer llegar estas herramientas a la comunidad científica y fomentar su uso de forma temprana en la formación de los investigadores resulta fundamental para solventar la crisis de reproducibilidad. Estas herramientas giran alrededor de **tres ejes principales: 1) repositorios en línea para compartir datos y productos de la investigación, 2) herramientas digitales de libre acceso y 3) la implementación de buenas prácticas de manejo y análisis de datos**. Estos tres ejes representan los módulos en que se organiza el curso. Este curso pretende profundizar en el uso de herramientas (en su mayoría digitales) así como de buenas prácticas a lo largo de cada uno de las etapas comunes al desarrollo de proyectos de investigación en biología. El curso provee beneficios tanto para la comunidad científica, al promover altos estándares de calidad, como

beneficios profesionales para las y los estudiantes, al prepararles para un mercado laboral que empieza a privilegiar la capacidad de desarrollar ciencia abierta, colaborativa y reproducible y en el cual utilizar eficientemente herramientas digitales es una capacidad deseada.

OBJETIVO GENERAL:

Familiarizar a las y los estudiantes con herramientas digitales para el manejo de datos y buenas prácticas para garantizar transparencia, accesibilidad de datos y reproducibilidad en la investigación científica biológica.

OBJETIVOS ESPECIFICOS:

- Dar a conocer la problemática actual sobre reproducibilidad científica y sus implicaciones para la producción de conocimiento y la legitimidad de este en la sociedad
- Brindar a los estudiantes herramientas computacionales de software libre que faciliten la documentación y accesibilidad de la investigación.
- Identificar procesos en las diferentes etapas de la investigación científica en las que se pueden adoptar prácticas que facilitan la reproducibilidad
- Proveer a los estudiantes con experiencia en la aplicación de las herramientas brindadas por medio de prácticas y proyectos individuales.

CONTENIDO Y CRONOGRAMA:

Semana	Tema
1 (30 de marzo)	Introducción al curso Lectura del programa Reproducibilidad y otras propiedades deseables en investigación. ¿Por qué importa la reproducibilidad? Crisis de reproducibilidad: estado de la cuestión Lectura: <i>Baker, M. 2016. Reproducibility crisis.</i>
Módulo 1: repositorios y bases de datos	
2 (06 de abril)	Introducción a ciencia abierta Contribución a repositorios de datos. Principios y prácticas FAIR- Convenciones de nomenclatura. Respaldos. Documentación y metadatos. Selección de repositorios y ejemplos. Lectura: <i>Gallagher et al. 2020. "Open Science Principles for Accelerating Trait-Based Science Across the Tree of Life."</i>
3	Uso de repositorios de datos en investigación Manejo y estructura de bases de datos. Datos compartibles. Formatos de archivo, "Non-data content", README, tidy data.

Semana	Tema
(13 de abril)	Open Science Framework: crear repositorios Open Science Framework: colaborar y compartir datos Lectura: <i>Foster & Deardorff. 2017. Open science framework (OSF).</i>
5	Pre-registro de proyectos de investigación
(20 de abril)	Compartir publicaciones, datos, recursos educativos, métodos, software, y revisión abiertas. Licencias. Selección de artículo para reproducción. Idea general del proyecto final. Lecturas: 1) <i>Wilkinson. 2016. The FAIR Guiding Principles for scientific data management and stewardship.</i> 2) <i>Parker et al. 2016. Promoting transparency in evolutionary biology and ecology.</i>
Módulo 2: Herramientas digitales	
6	Repaso de R
(27 de abril)	Control de versiones y colaboración usando git Control de versiones y colaboración usando github y github desktop Avance de proyecto final I Lectura: <i>Bryan. 2018. Excuse Me, Do You Have a Moment to Talk about Version Control?</i>
07	Compartiendo datos y código en github Colaborar usando git y github Github y Open Science Framework Lectura: <i>Ram. 2013. Git Can Facilitate Greater Reproducibility and Increased Transparency in Science</i>
(04 de mayo)	
8	Programación como herramienta indispensable para la reproducibilidad (con énfasis en R). Ventajas de automatización vs “a mano” Ambientes reproducibles con packrat y holepunch Reportes dinámicos en Rmarkdown Lectura: <i>Marwick et al. 2018. Packaging Data Analytical Work Reproducibly Using r (and Friends)</i>
(11 de mayo)	
9	Buenas prácticas en la escritura y documentación de código. Generalidad. Claridad vs eficiencia. Gráficos dinámicos en R Compartiendo código en Rpubs Lectura: <i>Vuorre et al. 2020. Sharing and Organizing Research Products as r Packages.</i>
(18 de mayo)	
10	Organización de datos y código de proyectos de investigación Compendios de investigación Avance de proyecto final II

Semana	Tema
(25 de mayo)	Lectura: <i>Schwab et al. 2000. Making Scientific Computations Reproducible.</i>
Módulo 3: Buenas practicas para el manejo y análisis de datos	
11 (01 de junio)	Antes de tomar los datos: formatos estándar en la toma de datos Precisión taxonómica. Bases de datos en línea. Taxize. Darwin Core y GBIF Lectura: Vink. 2012. Taxonomy and Irreproducible Biological Science
12 (08 de junio)	Buenas prácticas de análisis estadísticos Simulaciones como herramienta fundamental para el aprendizaje estadística P-hacking. Trampas estadísticas y cómo detectarlas. Avance de proyecto final I Lectura: <i>Gigerenzer. 2018. Statistical rituals: The replication delusion and how we got there.</i>
13 (15 de junio)	Replicación y pseudoreplicación. Valores de P, Tamaño de efecto. Supuestos. Contrastes apropiados e inapropiados. ¿Por qué considerar estadística Bayesiana? Causación y modelos gráficos Lectura: <i>Arif & MacNeil. 2022. Predictive models aren't for causal inference.</i>
14 (22 de junio)	Predicciones. Consideraciones de diseño. Datos ciegos. Redacción para ciencia reproducible. Métodos y Resultados. Citar correctamente. Presentación de artículos replicados I
15 (29 de junio)	Del capítulo de tesis a la publicación científica. Checklist de reproducibilidad. Preprints. Revisión por pares. Cartas de presentación. Tiempos y expectativas. Retracción. Presentación de artículos replicados II
16 (06 de julio)	Presentación de artículos replicados III
17 (13 de julio)	Entrega de proyectos finales
18 (20 de julio)	Ampliación

METODOLOGÍA Y ACTIVIDADES PARA CUMPLIR CON LOS OBJETIVOS:

Este curso combina presentaciones del profesor, discusiones con los participantes y actividades prácticas. Cada semana el profesor presentará el tema y su contexto y se llevarán a cabo una o más actividades prácticas para ilustrarlo. Las actividades generalmente son simultáneas a la presentación. En todas las sesiones, la resolución de las actividades será evaluada como parte del curso. Se promueve el diálogo horizontal y la colaboración entre participantes, pero en la mayoría de los casos, cada estudiante debe entregar su propia solución de las actividades, a más tardar una semana después de que fueran asignadas.

Cada semana también se discutirán artículos científicos y artículos tipo opinión. Una o un estudiante estará a cargo de liderar cada discusión. La responsabilidad de liderar la discusión será asignada de manera voluntaria o aleatoria. La discusión de artículos también será tomada en cuenta en la evaluación.

Las y los estudiantes llevarán a cabo dos proyectos independientes. El primero es la reproducción total o parcial de los análisis de un artículo científico de su interés. El o la estudiante debe escoger un artículo con un nivel de complejidad adecuado con ayuda del profesor. Además debe redactar sus conclusiones sobre su experiencia, indicando las dificultades que tuvo al reproducir el estudio y los aspectos del estudio que facilitaron su reproducción. Las y los estudiantes también llevarán a cabo una breve presentación sobre el artículo seleccionado y su experiencia reproduciéndolo.

El segundo proyecto individual consiste en aplicar las herramientas de reproducibilidad aprendidas a una parte de su investigación planeada o en curso, o a un pequeño proyecto de investigación planteado para el curso y con datos pre-existentes o simulados. Para este proyecto individual las y los estudiantes deben elaborar una estrategia o plan de reproducibilidad indicando las hipótesis y predicciones de manera gráfica (si aplica), justificando el diseño experimental o de muestreo y aportando información suficiente para acceder, comprender los datos. Además, el proyecto individual incluirá elaborar código apropiadamente anotado para el manejo y análisis de al menos una parte de los datos y un resumen de métodos (del análisis), redactado al estilo de un artículo científico.

Uso de Mediación Virtual

Se utilizará la plataforma virtual de la UCR (sistema de mediación virtual) para colocar todo el material de apoyo para el curso, así como información importante para las y los estudiantes. Las y los estudiantes entregarán algunas tareas y su proyecto final por medio de la plataforma. Es indispensable que las y los estudiantes revisen el entorno 2-3 veces por semana.

EVALUACIÓN DEL CURSO:

Ejercicios/Tareas	30%
Discusión de artículos	10%
Reproducción de artículo	20%
Presentación de artículo	10%
Proyecto final	30%

Tareas (30%): Durante el semestre se dejarán 3 ejercicios de tarea (10% cada tarea), que le permitirá al estudiante poner en práctica lo aprendido en clase. Las tareas consistirán en ejercicios relacionados a : (i) manipulación de bases de datos y uso de repositorios de datos en línea, (ii) creación de reportes dinámicos, y (iii) pruebas prácticas de conceptos desarrollados en clase, . Las tareas se asignarán con una semana de antelación, y deberán ser entregadas en las siguiente fecha.

Discusión de artículos: Cada semana se discutirán artículos científicos y artículos tipo opinión. Una o un estudiante estará a cargo de liderar cada discusión (“presentación de artículo”, 10%). La responsabilidad de liderar la discusión será asignada de manera voluntaria o aleatoria. También se evaluará la participación durante las discusiones de artículos presentados por otros estudiantes (“discusión de artículos”, 10%).

Proyectos: Las y los estudiantes llevarán a cabo dos proyectos independientes. El primero es la reproducción total o parcial de los análisis de un artículo científico de su interés (reproducción de artículo, 20%). El o la estudiante debe escoger un artículo con un nivel de complejidad adecuado con ayuda de los profesores. Además debe redactar sus conclusiones sobre su experiencia, indicando las dificultades que tuvo al reproducir el estudio y los aspectos del estudio que facilitaron su reproducción. Las y los estudiantes también llevarán a cabo una breve presentación sobre el artículo seleccionado y experiencia reproduciéndolo. El segundo proyecto individual consiste en aplicar las herramientas de reproducibilidad aprendidas a una parte de su investigación planeada o en curso, o a un pequeño proyecto de investigación planteado para el curso y con datos pre-existentes o simulados (“proyecto final”, 30%). Para este proyecto individual las y los estudiantes deben elaborar una estrategia o plan de reproducibilidad indicando las hipótesis y predicciones de manera gráfica (si aplica), justificando el diseño experimental o de muestreo y aportando información suficiente para acceder y comprender los datos. Además, el proyecto individual incluirá elaborar código apropiadamente anotado para el manejo y análisis de al menos una parte de los datos y un resumen de métodos (del análisis), redactado al estilo de un artículo científico.

BIBLIOGRAFÍA

- Alston, Jesse, and Jessica Rick. 2020. “A Beginner’s Guide to Conducting Reproducible Research.” <https://doi.org/10.32942/osf.io/h5r6n>.
- Arif, S., & MacNeil, M. A. (2022). Predictive models aren't for causal inference. *Ecology Letters*, 25(8), 1741-1745. <https://doi.org/10.1111/ele.14033>
- Auker, Linda A, and Erika L Barthelmess. 2019. “Teaching r in the Undergraduate Ecology Classroom: Approaches, Lessons Learned, and Recommendations.” *bioRxiv* 11: 666768. <https://doi.org/10.1101/666768>.
- Baker, M. (2016). Reproducibility crisis. *Nature*, 533(26), 353-66. <https://doi.org/10.1038/533452a>
- Bryan, Jennifer. 2018. “Excuse Me, Do You Have a Moment to Talk about Version Control?” *American Statistician* 72: 20–27. <https://doi.org/10.1080/00031305.2017.1399928>.
- Carey, By Maureen A, Kevin L Steiner, and William A Petri. 2020. “10 Simple Rules for Reading a Scientific Paper,” 1–8. <https://doi.org/10.1371/journal.pcbi.1008032>.
- Culina, Antica, and Simon Evans. 2019. “Low Availability of Code in Ecology : Call for Urgent Action.” *Open*

- Science Conference*, 1–9. <https://doi.org/10.1371/journal.pbio.3000763>.
- Foster, E. D., & Deardorff, A. (2017). Open science framework (OSF). *Journal of the Medical Library Association: JMLA*, 105(2), 203. <https://doi.org/10.5195/jmla.2017.88>
- Gallagher, Rachael V., Daniel S. Falster, Brian S. Maitner, Roberto Salguero-Gómez, Vigdis Vandvik, William D. Pearse, Florian D. Schneider, et al. 2020. “Open Science Principles for Accelerating Trait-Based Science Across the Tree of Life.” *Nature Ecology & Evolution*, February. <https://doi.org/10.1038/s41559-020-1109-6>.
- Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, 1(2), 198–218. <https://doi.org/10.1177/2515245918771329>
- Heesen, Remco. 2018. “Why the Reward Structure of Science Makes Reproducibility Problems Inevitable.” *Journal of Philosophy* 115: 661–74. <https://doi.org/10.5840/jphil20181151239>.
- Ince, Darrel C., Leslie Hatton, and John Graham-Cumming. 2012. “The Case for Open Computer Programs.” *Nature* 482: 485–88. <https://doi.org/10.1038/nature10836>.
- Janz, Nicole. 2016. “Bringing the Gold Standard into the Classroom: Replication in University Teaching.” *International Studies Perspectives* 17: 392–407. <https://doi.org/10.1111/insp.12104>.
- Marwick, Ben, Carl Boettiger, and Lincoln Mullen. 2018. “Packaging Data Analytical Work Reproducibly Using r (and Friends).” *American Statistician* 72: 80–88. <https://doi.org/10.1080/00031305.2017.1375986>.
- McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315372495>
- Mislan, K. A. S., Jeffrey M. Heer, and Ethan P. White. 2016. “Elevating the Status of Code in Ecology.” *Trends in Ecology and Evolution* 31: 4–7. <https://doi.org/10.1016/j.tree.2015.11.006>.
- Parker, T. H., Nakagawa, S., Gurevitch, J., & IIEE (Improving Inference in Evolutionary Biology and Ecology) workshop participants. (2016). Promoting transparency in evolutionary biology and ecology. *Ecology Letters*, 19(7), 726–728. <http://dx.doi.org/10.1111%2Fele.12610>
- Piwowar, Heather. 2013. “Value All Research Products.” *Nature* 493: 159.
- Prlić, Andreas, and James B Procter. 2012. “Ten Simple Rules for the Open Development of Scientific Software.” *PLOS Computational Biology* 8 (December): e1002802. <https://doi.org/10.1371/journal.pcbi.1002802>.
- Ram, Karthik. 2013. “Git Can Facilitate Greater Reproducibility and Increased Transparency in Science.” *Source Code for Biology and Medicine* 8: 1–8. <https://doi.org/10.1186/1751-0473-8-7>.
- Romero, Felipe. 2019. “Philosophy of Science and the Replicability Crisis.” *Philosophy Compass* 14: 1–14. <https://doi.org/10.1111/phc3.12633>.
- Schwab, M., N. Karrenbach, and J. Claerbout. 2000. “Making Scientific Computations Reproducible.” *Computing in Science & Engineering* 2: 61–67. <https://doi.org/10.1109/5992.881708>.
- Vink, Cor J., Pierre Paquin, and Robert H. Cruickshank. 2012. “Taxonomy and Irreproducible Biological Science.” *BioScience* 62: 451–52. <https://doi.org/10.1525/bio.2012.62.5.3>.
- Vuorre, Matti, and Matthew J. C. Crump. 2020. “Sharing and Organizing Research Products as r Packages.” *Behavior Research Methods*, 1–11. <https://doi.org/10.3758/s13428-020-01436-x>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1–9.

<https://doi.org/10.1038/sdata.2016.18>

Wilson, Greg, D A Aruliah, C Titus Brown, Neil P Chue Hong, Matt Davis, Richard T Guy, Steven H D Haddock, et al. 2014. "Best Practices for Scientific Computing." *PLOS Biology* 12 (January): e1001745.

<https://doi.org/10.1371/journal.pbio.1001745>.

Recursos adicionales en internet

- Open Science Framework: <https://osf.io/>
- Open Science Unesco: <https://www.unesco.org/en/natural-sciences/open-science>
- Qué es ciencia abierta: https://www.fosteropenscience.eu/learning/what-is-open-science_es/
- Statistical rethinking (Richard McElreath): <https://www.youtube.com/playlist?list=PLDcUM9US4XdMROZ57-OIRtIK0aOynbgZN>



DISCRIMINACIÓN

Es un acto u omisión que afecta las oportunidades de una persona o sus derechos humanos.

SON MANIFESTACIONES DE DISCRIMINACIÓN:

- Ataques físicos
- Burlas, bromas ofensivas
- Uso de vocabulario discriminatorio
- Trato diferencial o despectivo
- Exclusión o segregación
- Desinterés o maltrato
- Negación a brindar servicios

DENUNCIA

La denuncia puede presentarse personalmente o mediante correo electrónico ante la **Comisión Institucional Contra la Discriminación (CICDI)**.

Ninguna de las personas involucradas en el proceso podrán sufrir prejuicios.

Si usted ha vivido una situación de discriminación puede acercarse a la Facultad de Ciencias para buscar apoyo.

 2511-6345  facultad.ciencias@ucr.ac.cr





Toda conducta de naturaleza sexual indeseada por quien la recibe, que provoque efectos perjudiciales en el estado general o bienestar personal.

SON MANIFESTACIONES DE HOSTIGAMIENTO SEXUAL:

- Promesa o amenaza, implícita o expresa, relacionada con favores sexuales
- Propuestas o conductas de naturaleza sexual
- Humillaciones u ofensas con palabras, gestos o imágenes
- Acercamientos o formas de contacto físico no deseados
- Intentos de comunicación ajenos a la relación profesional o académica

DENUNCIA

Las denuncias se realizan en forma verbal o escrita, ante la Comisión Institucional Contra el Hostigamiento Sexual (CICHS).

CONTACTOS

Comisión Institucional contra el Hostigamiento Sexual: 2511-4898
comision.contrahostigamiento@ucr.ac.cr
Defensoría contra el Hostigamiento Sexual: 2511-1909
defensoriahs@ucr.ac.cr

