

# Reproducibilidad

Miriam Lerma

Mayo 2021

# Intro

- Reproducibilidad
- Publicaciones reproducibles
- Manuscritos reproducibles

## Ustedes

- Quieren crear proyectos reproducibles
- Mejorar su seguimiento de proyectos
- Mejorar el ciclo de trabajo en su laboratorio

## Preguntas

Responder en el chat 

- Han querido replicar análisis de algún artículo ya publicado y se preguntaron como lo hicieron?
- Les ha pasado que a seguir los pasos de alguna tesis o artículo se dieron cuenta que no todos los métodos estaban descritos?

# Créditos & Recursos

## Lecturas

-  [The turing way](#)
-  [Articulos reproducibles](#)

## Videos en español

- Daniela Ballari  [ReproHack](#)
- Francisco Rodríguez-Sánchez  [ReproHack](#)

## En inglés

-  [RLadiesJohannesburg](#)
-  [Matt Dray](#)
-  [Experimenting with reproducibility](#)
-  [r3 course](#)

## Imagenes

- Portada [Unsplash by Joel Filip](#)
- Imagenes de The turing test creadas por Scriberia para la comunidad The Turing Way y se usa bajo una licencia CC-BY.

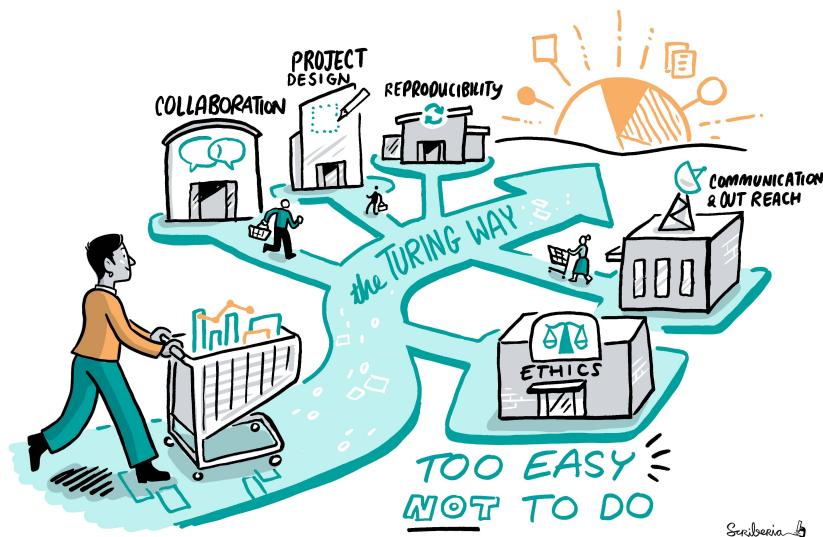
# 1. Reproducibilidad

# 1.1. Reproducibilidad

**Reproducibilidad computacional** es que otra persona pueda ejecutar el código que hemos creado y tenga los mismos resultados

Esto incluye a:

- Científicos de otras instituciones
- Colaboradores
- Nosotros mismos



# 1.2. Reproducibilidad

La reproducibilidad tiene varias dimensiones, de acuerdo a

- si usamos los mismos datos o datos diferentes,
- si usamos el mismo código o lo adaptamos.

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

- **Reproducible:** mismos datos deben dar el mismo resultado.
- **Replicable:** mismos análisis pero con diferentes datos dan resultados cualitativamente similares.
- **Robusto:** los mismos datos pero diferente análisis, dan resultados similares.
- **Generalizable:** diferentes datos y análisis para entender que los resultados no son dependientes de ese set de datos en particular y de ese análisis en particular.

Fuente: [The turing way](#)

# 1.3. Reproducibilidad

Recordemos que:

El método científico esta sustentado en dos pilares: **reproducibilidad** y refutabilidad

La reproducibilidad implica que se pueda repetir el experimento, en cualquier lugar y persona para verificar que los resultados obtenidos sean verídicos.

**Desafortunadamente**, por muchos años se ha perdido el acceso a datos y métodos  
**Afortunadamente**, estamos en una época de cambio de paradigmas



# 1.4. Credibilidad

Al hacer nuestro trabajo reproducible.  
Podemos:

- Evitar desastres al permitir detección de errores
- Escribir mas fluido
- Asegurar la continuidad
- Mejorar la credibilidad/transparencia

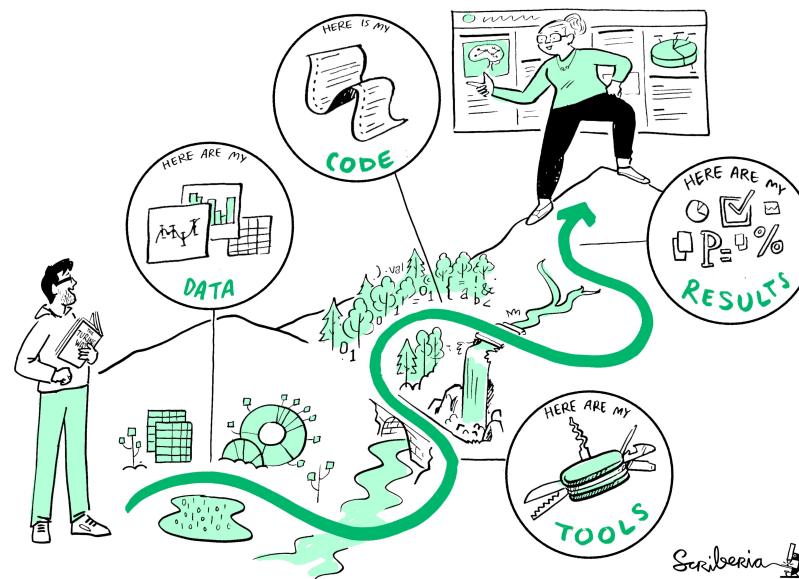


Por ejemplo, evitando casos como: **revistas retiran artículos sobre COVID debido a falta de integridad en los datos**

# 1.5. Recomendaciones

Para crear un proyecto reproducible hay que:

- Agregar comentarios y escribir código en forma ordenada.  
Ver [programacion literaria, un concepto de 1984](#)
  - Usar código legible
  - Usar directorios relativos
  - Usar control de versiones
- Esto evita redundancias: no intentes lo mismo muchas veces

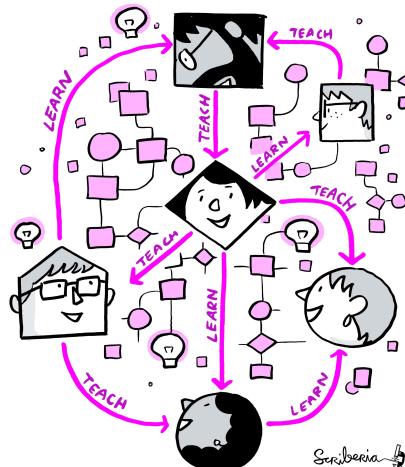


# 1.6. Supervisores

Como **supervisor** al tener proyectos reproducibles, se puede:

- Evaluar como se procesaron los datos
- Detectar errores
- Facilitar la escritura y edición
- Facilitar el ciclo de investigación

Tener un flujo de trabajo reproducible en un laboratorio ahorra tiempo para los investigadores, estudiantes y colaboradores. Aunque hay que invertir tiempo en generar proyectos reproducibles, a la larga esto mejora la productividad

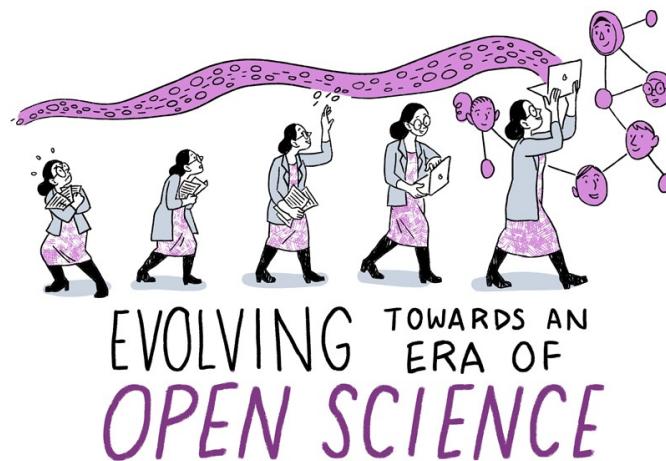


# 1.7. Tesistas

Como **tesista** entrar a un laboratorio que tenga proyectos reproducibles, permite que:

- Inician con análisis base
- Complementen, actualicen y mejoren el código
- No se pierdan en el universo de análisis y paquetes
- Tengan más confianza en los resultados

Trabajar en un laboratorio con flujos de trabajo reproducible, ahorra tiempo y permite invertir esfuerzos en mejorar el código



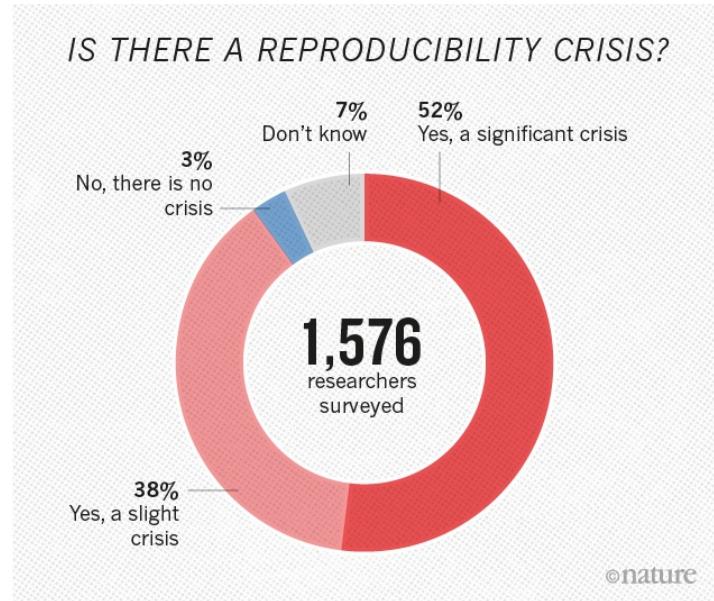
## 2. Publicaciones reproducibles

## 2.1. Publicaciones

De 1500 investigadores que intentaron reproducir experimentos publicados, 70% fallaron.

Por lo que la mayoría concluyó que existe una **crisis en la reproducibilidad**

-  Ted: crisis reproducibilidad



Fuente: **nature**

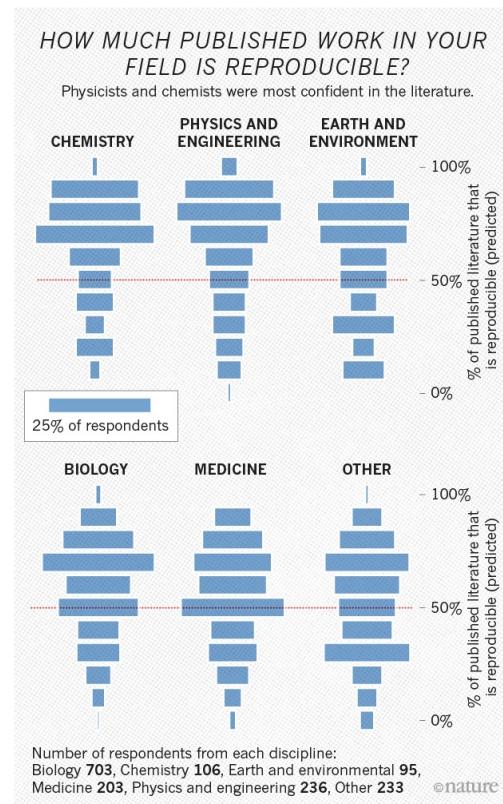
## 2.2. Disciplinas

¿Qué disciplinas son las más reproducibles?



Afortunadamente existen algunas iniciativas que apoyan y consultan para mejorar la reproducibilidad

- ReproHack
- rOpenScience



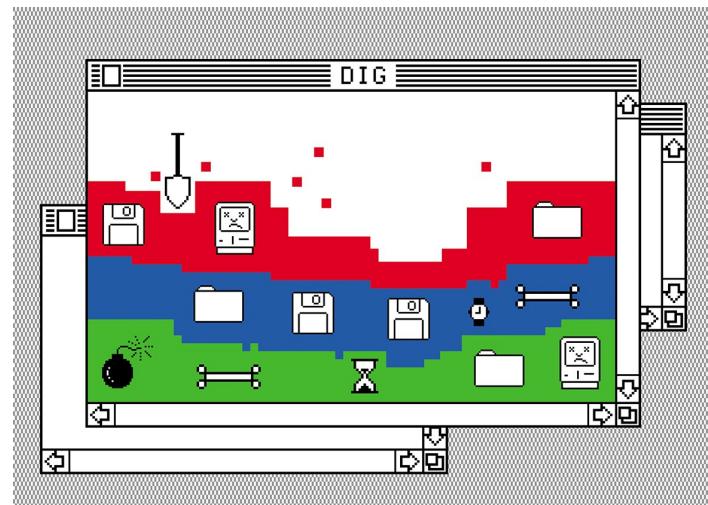
Fuente: nature

## 2.3. Código antigüo

Curiosamente, la poca reproducibilidad no solo aplicaba para resultados de otros científicos si no también incluía **sus propios resultados**.

Falta de documentación y programas obsoletos mostraron que era muy difícil reproducir incluso sus propios resultados de hace **10 años**.

- **Ten years reproducibility challenge**

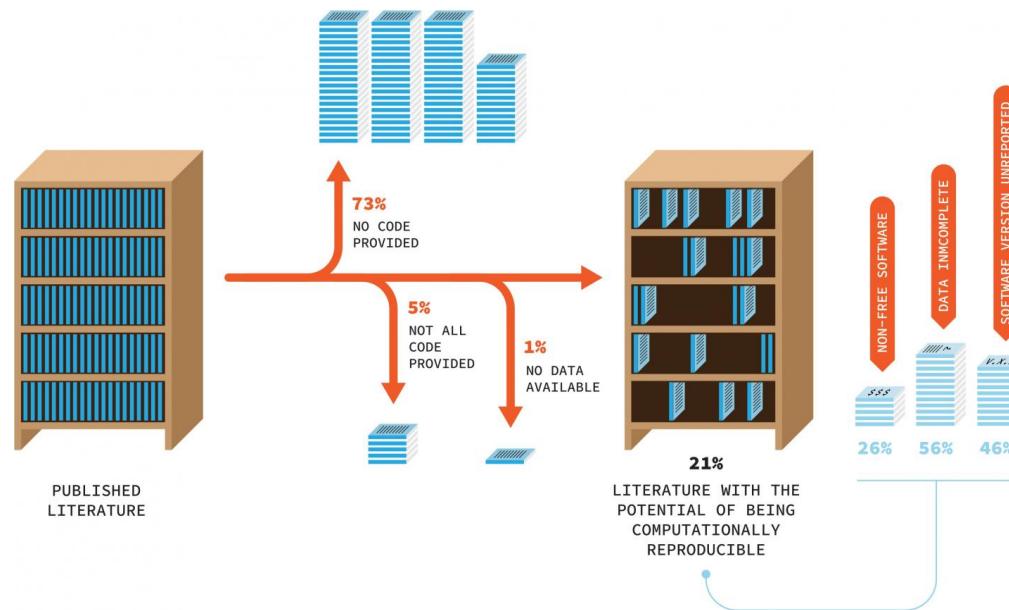


Fuente: **nature**

## 2.4. Reproducibilidad en ecología

Hay poca reproducibilidad en ecología ([Culina 2020](#))

- **73%** de los artículos publicados (entre 2015 y 2019, n = 346) no proveen código
- **5%** no todo el código
- **1%** no proveen datos
- **21%** casi reproducible (dentro de los cuales 26% no usa software libre, 56% no provee todos los datos, y 46% no reportan la versión del software)



Fuente: [@Culina](#)

## 2.5. Barreras

## Barreras:

- Desinformación
  - Tradición
  - Falta de entrenamiento
  - Falta de reconocimiento

## Miedo a:

- Robo de ideas
  - No ser acreditado
  - Errores que lleven a humillación



 Tenemos la tecnologia, depende de nosotros implementarlo

# 3. Manuscritos reproducibles

# 3.1. Manuscritos reproducibles

Texto + Datos + Código

Pero existe un gradiente...



Fuente: Pakillo

## 3.2. Básica

Texto + Datos + Código

Almacenados y accesibles en **archivos permanentes**

- ✓ Zenodo, dryad
- ✗ Github

## 3.3. Datos

- En formato abierto:
  - ✓ csv, txt
  - ✗ excel
- Identificador permanente (DOI)
- Código de depuración (limpieza)
- README (explicando: quien, como, cuando, donde se colectaron los datos).  
Evita: **variables no localizadas**
- Licencias
  - Las más comunes son:  
**CCO** Creative Commons: no condiciones  
**CC-BY** Uso con atribución  
**MIT** para software

Leer más: [Tierney & Ram 2020](#)

## 3.4. Código

- En formato abierto:
  - ✓ Script en R o Rmd
- Identificador permanente (DOI)
- Licencias (CCO, CC-BY)
- README (incluir sessionInfo()).  
Evita problemas con actualización de **librerías**

Leer más: Eglen et al 2016

# 3.5. Ejercicios

En equipo 🧸

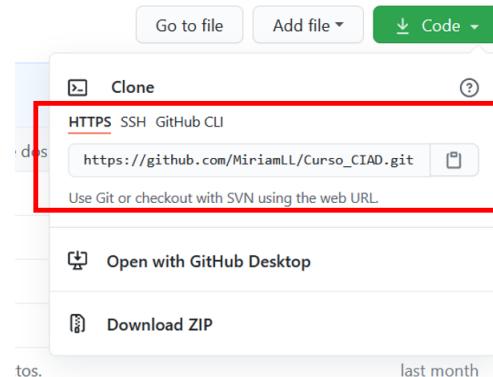
- **Replicar** análisis de otro equipo: [Equipo 1](#), [Equipo 2](#), [Equipo 3](#).
- **Reproducir** análisis para obtener resultados de pingüinos de Papúa.
- Subir nuevo repositorio a github, de preferencia incluyendo README.
- Al final reflexionar que podrían incluir en sus futuras instrucciones para facilitar reproducibilidad.

Con git

Sin git

README

Si ya tienen **git** instalado. [Clonar repositorio](#).



# 4. Paquetes

## Rmarkdown

- Todo está contenido por lo que permite tener continuidad en los análisis

## Fertile

- Revisa continuamente que el artículo sea reproducible. Por ejemplo, si usas una ruta absoluta te va a sugerir que uses una ruta relativa

## Checkpoint

- Permite regresar a algún punto de tu análisis

... hay muchos más



Fuente de la imagen [The Turing way](#)

# Contacto

## Recapitulando

- Reproducibilidad
- Publicaciones reproducibles
- Manuscritos reproducibles

Para dudas, comentarios y sugerencias:

- Escríbeme a [miriamjlerma@gmail.com](mailto:miriamjlerma@gmail.com)

Este material esta accesible y se encuentra en  
mi [github](#) y mi [página](#)

 **Volver**