## Experiencias con la gestión y publicación de datos científicos



Marcela Alfaro Córdoba, Ph.D. Profesora Invitada, Escuela de Estadística Universidad de Costa Rica

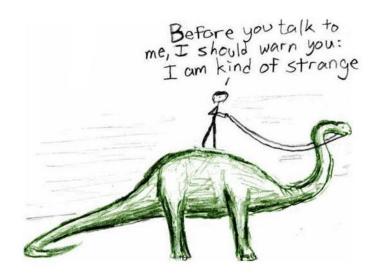
### Puede encontrar la presentación aquí:

https://malfaro2.github.io/RDA/Datos Inv Abiertos Revistas.pdf

### Cuéntenme de Uds

- Afiliación institucional.
- Área de especialización de su revista.
- ¿Qué opina de la Ciencia Abierta?
- ¿Qué opina de los datos abiertos?

Por favor responda en el link que será circulado en el chat.

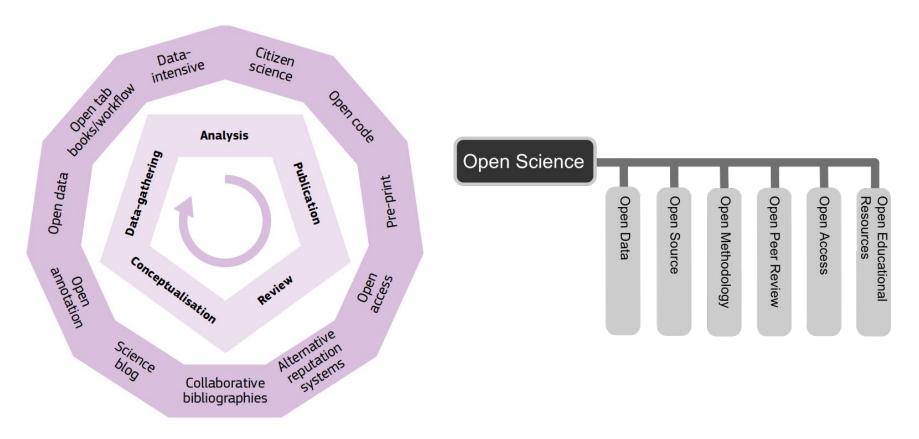


Fuente: https://xkcd.com/15/

## ¿Quién soy yo?

- Profesora e investigadora de la Escuela de Estadística, del Posgrado en Estadística y del CIMPA de la Universidad de Costa Rica. Además, colaboro con investigadoras y docentes del CICIMA, CIBCM, CIGEFI, y del Posgrado en Computación. Más información: <a href="https://malfaro2.github.io/">https://malfaro2.github.io/</a>
- Fundadora del capítulo <u>R-ladies en San José</u> y de <u>ConectaR2019</u>, además de ser la anfitriona para la 16va Plenaria de RDA, que se llevará a cabo en Costa Rica en noviembre de este año. Co-directora de <u>CODATA-RDA Schools of Data Science</u>.

CODATA: Committee on Data of the International Council for Science y RDA: Research Data Alliance, son organizaciones internacionales preocupadas por el estado y el futuro de los datos, las habilidades para manejarlos y los esfuerzos de estandarización que aseguren un uso eficiente.



Fuente: European Commission Public Consultation Science 2.0: Science in Transition



EDITORIAL · 04 FEBRUARY 2020

## Calling all coronavirus researchers: keep sharing, stay open

As the new coronavirus continues its deadly spread, researchers must ensure that their work on this outbreak is shared rapidly and openly.



















A ministry of health official visits a family in Guatemala City to help control the mosquitoes that transmit Zika virus.

CONRED GUATEMALA (CC BY-NC-ND 2.0)

A plea for open science on Zika

By Gretchen Vogel | Feb. 10, 2016, 1:00 PM

# Recomendaciones generales para personas investigadoras:

- Hacer un plan de manejo de datos de investigación a la hora de planear el artículo. <a href="https://dmptool.org/">https://dmptool.org/</a>
- 2. Buscar herramientas que se acoplen a nuestro contexto (repositorios por área de conocimiento, protecciones legales, estabilidad tecnológica y política).
- 3. Responder la pregunta más importante en este caso: ¿existe una persona "dueña" de los datos de investigación que estoy utilizando?

## Recomendaciones generales para personas encargadas de las revistas:

- 1. Tener una propuesta de plan de manejo de datos, en el idioma de la población meta (español en este caso).
- 2. Recomendar herramientas que se acoplen al contexto (repositorios por área de conocimiento, protecciones legales, estabilidad tecnológica y política).
- 3. Hacer parte del protocolo las preguntas más importantes en este caso: ¿existe una persona "dueña" de los datos de investigación que se están utilizando? ¿qué es considerado como datos en la disciplina?

### Plan de manejo de datos

### Recopilación de datos

- ¿Qué datos recopilará o creará?
- ¿Cómo se recopilarán o crearán los datos?

### Documentación y metadatos

¿Qué documentación y metadatos acompañarán los datos?

### Ética y cumplimiento legal

- ¿Cómo manejará cualquier problema ético?
- ¿Cómo manejará los problemas de derechos de autor y propiedad intelectual (IP / IPR)?

## Plan de manejo de datos

### Almacenamiento y respaldo

- ¿Cómo se almacenarán y respaldarán los datos durante la investigación?
- ¿Cómo administrará el acceso y la seguridad?

### Selección y preservación

- ¿Qué datos son de valor a largo plazo y deben ser retenidos, compartidos y / o preservados?
- ¿Cuál es el plan de preservación a largo plazo para el conjunto de datos?

## Plan de manejo de datos

### **Compartir datos**

- ¿Cómo va a compartir los datos?
- ¿Se requieren restricciones para compartir datos?

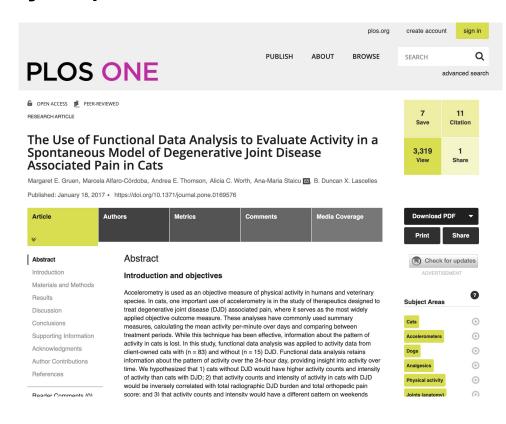
### Responsabilidades y recursos

- ¿Quién será responsable de la gestión de datos?
- ¿Qué recursos necesitará para entregar su plan?

## ¿Cómo hago para hacerlo en la práctica?

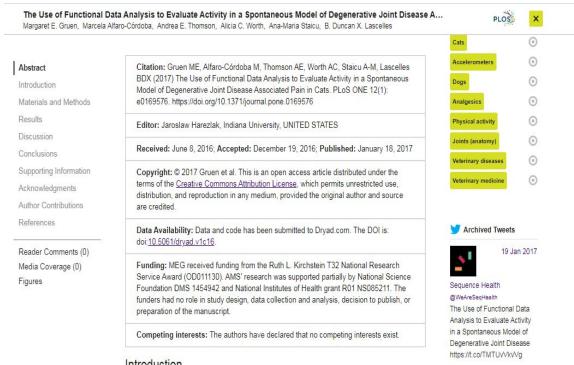
Hay muchos caminos, y en parte depende de las particularidades del área del conocimiento, pero les voy a dar dos ejemplos completos de dos publicaciones científicas.

### Ejemplo 1: Artículo de veterinaria en PLOS One



- Artículo:
  - https://journals.plos.org/plosone/article?id =10.1371/journal.pone.0169576
- Datos y código:
  - https://datadryad.org/stash/dataset/doi:10 .5061/dryad.v1c16

### Ejemplo 1: Artículo de veterinaria en PLOS One



Introduction

### Ejemplo 1: Artículo de veterinaria en PLOS One



Data from: The use of functional data analysis to evaluate activity in a spontaneous model of degenerative joint disease associated pain in cats

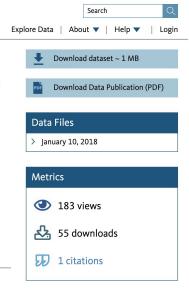
Gruen, Margaret E., North Carolina State University
Alfaro-Córdoba, Marcela, North Carolina State University
Thomson, Andrea E., North Carolina State University
Worth, Alicia
Staicu, Ana-Maria, North Carolina State University
Lascelles, B. Duncan X., North Carolina State University, University of North Carolina
Worth, Alicia C., North Carolina State University, University of North Carolina
Worth, Alicia C., North Carolina State University
Publication date: January 10, 2018
Publisher: Dryad
https://doi.org/10.5061/dryad.v1c16

#### Citation

Gruen, Margaret E. et al. (2018), Data from: The use of functional data analysis to evaluate activity in a spontaneous model of degenerative joint disease associated pain in cats, Dryad, Dataset, https://doi.org/10.5061/dryad.v1c16

#### Abstract

Accelerometry is used as an objective measure of physical activity in humans and veterinary species. In cats, one important use of accelerometry is in the study of therapeutics designed to treat degenerative joint disease (DJD) associated pain where it serves as the most widely applied objective outcome measure. These analyses have commonly

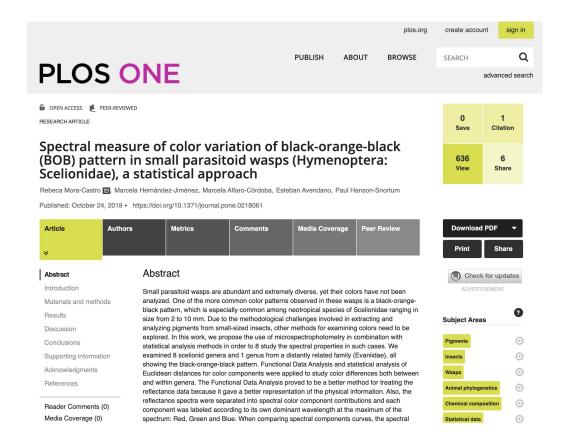


This work is licensed under a CC0 1.0
Universal (CC0 1.0) Public Domain
Dedication license.

PUBLIC
DOMAIN

- Financiamiento de un ente federal de EEUU que obliga a abrir los datos, además de las condiciones de PLOS One para la publicación.
- El costo de usar Dryad (que puede variar de \$120 hasta \$300 dependiendo del tamaño) fue cubierto por el financiamiento federal.
- Noten la licencia de dominio público en Dryad, esta la escogimos nosotros.
- Además, los datos se pueden CITAR, tienen su propio DOI.
- Desventajas:
  - Es caro.
  - No tenemos control de lo que pasará con DRYAD en el futuro.

## Ejemplo 2: Artículo de biología en PLOS One



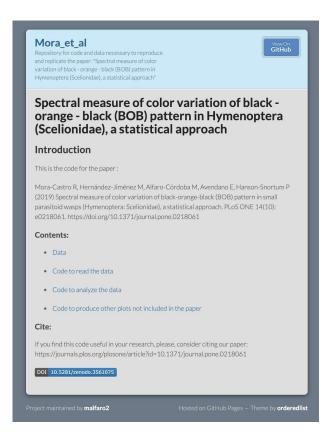
#### Artículo:

https://journals.plos.org/plosone/article/authors?id=10.1371/journal.pone.0218061

### Datos y código:

https://malfaro2.github.io/Mora et al/

## Ejemplo 2: Artículo de biología en PLOS One

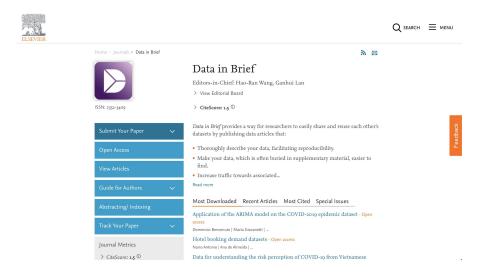


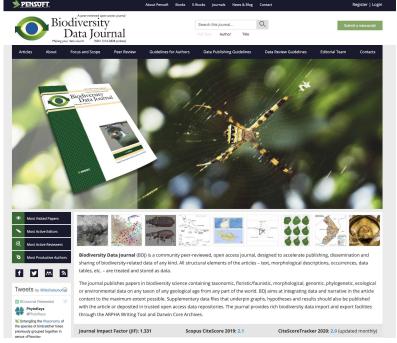
- Financiamiento viene de fondos de investigación de la UCR, que no requiere la publicación de datos. Sin embargo, era una de las condiciones de PLOS One para la publicación.
- Evitamos el costo de Dryad, creando nuestro propio repositorio usando la plataforma de GitHub (que usualmente se utiliza para código, no para datos). El uso de esta plataforma no tiene costo alguno en su versión más simple, pero tiene una curva de aprendizaje algo empinada.
- En este caso no se especifica la licencia de uso, pero si el DOI utilizando ZENODO, para que pueda citarse.

#### Desventajas:

- los datos no son fácilmente rastreables en la red, aunque si están bien definidos en el repositorio.
- No tenemos control de qué pasará con Github o con Zenodo en el futuro, aunque el segundo es mucho más confiable que el primero.

## ¿Solo PLOS One pide los datos?





## ¿Cuáles son los siguientes pasos para revistas?

- Recomendaciones de planes de manejo de datos para autores.
- Políticas de depósito de datos para acompañar los artículos.
  - Proceso de curación para los datos
  - Estándares de metadatos
  - Responsabilidad de preservación digital (incluyendo licencias y usos)
  - Privacidad y propiedad intelectual
  - Definición de objetos y formatos depositables
- El caso ideal: un repositorio de datos nacional (con la infraestructura que necesita) para datos recolectados con fondos públicos. Hay "machotes" como Dataverse: https://dataverse.org/.
  - Data Stewards o facilitadores de datos institucionales [ref].

## ¡Muchas gracias!

