

Estrategias de difusión de la actividad investigadora: el poder de la **ciencia abierta**

Wenceslao Arroyo Machado

Departamento de Información y Comunicación

[versión 2]

Contenidos

1. Problemas de la ciencia abierta
2. Nuevas posibilidades
3. La tripleta ganadora
4. Buenas prácticas

Objetivos

1. Conocer los problemas y desafíos de los datos abiertos
2. Aprender nuevas prácticas para compartir datos y código



1.

Problemas de la ciencia abierta

Consideraciones previas

Contenidos

1. **Problemas de la ciencia abierta**
2. Nuevas posibilidades
3. La tripleta ganadora
4. Buenas prácticas

Problemas de la ciencia abierta



open science

Son 6 los principios de la Ciencia Abierta:

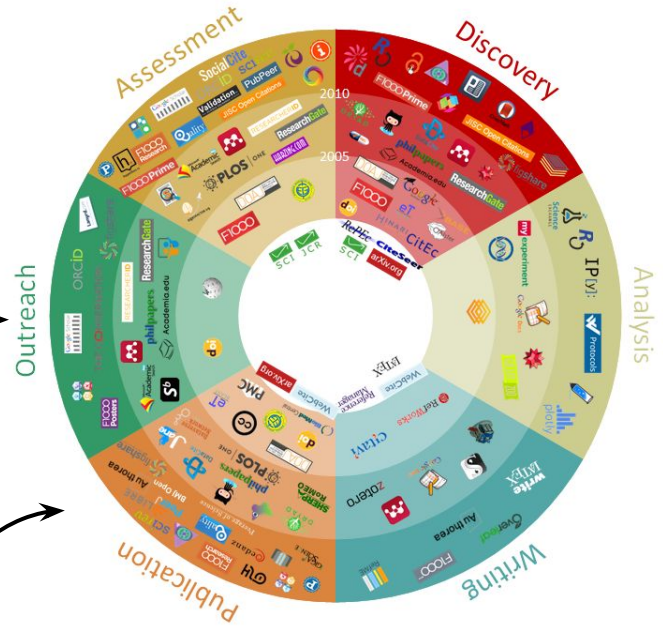
1. acceso abierto
2. **código abierto**
3. **datos abiertos**
4. metodologías abiertas
5. recursos educativos abiertos
6. revisión por pares abierta

Problemas de la ciencia abierta

Una parte fundamental del proceso de investigación está en **compartir**, no solo los resultados de investigación sino también los datos y procesos involucrados en ello.

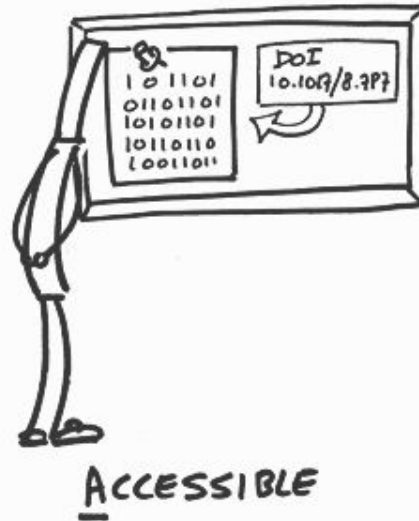
ANTES DE LLEGAR AQUÍ HAY QUE GARANTIZAR LA PUBLICACIÓN

HAY UNA AMPLIA VARIEDAD DE HERRAMIENTAS



Problemas de la ciencia abierta

FAIR DATA PRINCIPLES



Problemas de la ciencia abierta

Datos
Código

Disponibilidad
Usar repositorios abiertos, con un identificador asociado y licencia que garantice su reutilización

Zenodo, Digibug
DOI, ORCID, ROR
Creative Commons

Datos
Código

Formato
Emplear formatos abiertos y estandarizados

CSV, TSV, TXT

Datos
Código

Documentación
Detallar y explicar qué se incluye exactamente y los procesos

README.md
wikis

Datos
Código

Limpieza
Simplificar y eliminar cualquier tipo de ruido

Código

Reproducibilidad
Asegurar su correcta reproducción

Problemas de la ciencia abierta

De todos ellos los principales desafíos son

Disponibilidad



Uf. Tenemos problemas para encontrar ese sitio.

No podemos conectar al servidor en www.datosdeinvestigacion.com.

Si esa dirección es correcta, aquí hay otras tres cosas que puede probar:

- Vuelva a intentarlo más tarde.
- Compruebe su conexión de red.
- Si está conectado a través de un cortafuegos, compruebe que Firefox tiene permiso para acceder a la web.

Reintentar

Los materiales tienen que estar disponibles 24/7

Reproducibilidad

```
> analisisDatos(datos = 'carpeta')
ERROR FATAL
Es necesario un paquete que ya no existe
La versión instalada del software no es la requerida
Warning message:
In analisisDatos(datos = "carpeta") : Falta un archivo
> |
```

Los análisis tienen que ser reproducibles sin errores



Problemas de la ciencia abierta | Disponibilidad

Aunque es una práctica cada vez más extendida y requerida, de manera general todavía no cuenta con la atención y reconocimiento que debe.

COMMENT | 04 June 2019 | Correction [05 June 2019](#)

Make scientific data FAIR

All disciplines should follow the geosciences and demand best practice for publishing and sharing data, argue Shelley Stall and colleagues.

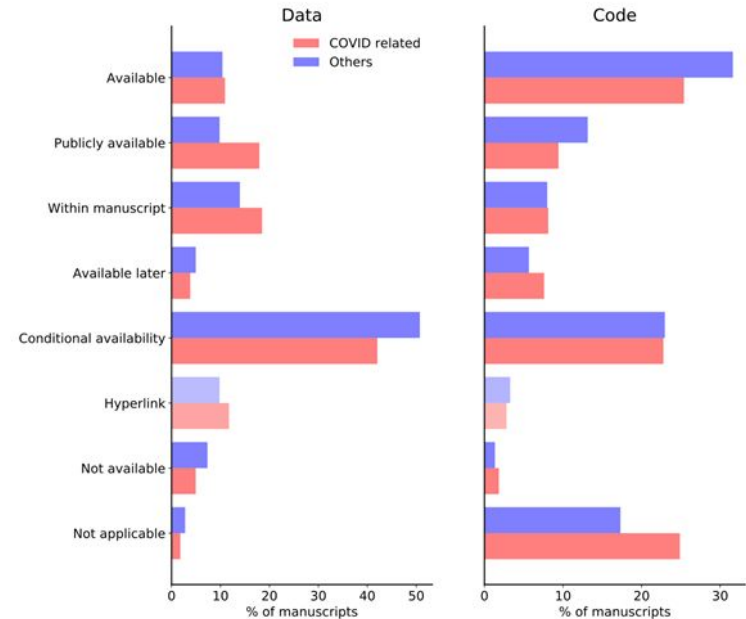
[Shelley Stall](#) , [Lynn Yarmey](#), [Joel Cutcher-Gershenfeld](#), [Brooks Hanson](#), [Kerstin Lehnert](#), [Brian Nosek](#), [Mark Parsons](#), [Erin Robinson](#) & [Lesley Wyborn](#)

Stall, S., Yarmey, L., Cutcher-Gershenfeld, J., Hanson, B., Lehnert, K., Nosek, B., Parsons, M., Robinson, E., & Wyborn, L. (2019). Make scientific data FAIR. *Nature*, 570(7759), 27–29. <https://doi.org/10.1038/d41586-019-01720-7>

Problemas de la ciencia abierta | Disponibilidad

La pandemia de COVID-19 ha evidenciado la falta de transparencia y compromiso en este sentido.

Los esfuerzos para compartir datos y códigos no han cambiado y siguen siendo minoría.



Larregue, J., Vincent-Lamarre, P., Lebaron, F., & Larivière, V. (2020). COVID-19: Where is the data? *Impact of Social Sciences*. <https://blogs.lse.ac.uk/impactofsocialsciences/2020/11/30/covid-19-where-is-the-data/>

Problemas de la ciencia abierta | Disponibilidad

Compartir los datos implica no solo hacerlos **públicamente** disponibles sino seguir unas buenas prácticas que garanticen su disponibilidad y correcta reutilización.

A diferencia de los manuscritos, las normas para publicar los datos son generales y fácilmente aplicables.

Data availability

Please provide a Data Availability statement in the Methods section under "Data Availability"; detailed guidance can be found in our [data availability and data citations policy](#). Certain data types must be deposited in an appropriate public structured data depository (details are available [here](#)), and the accession number(s) provided in the manuscript. Full access is required at publication. Should full access to data be required for peer review, authors must provide it.

We encourage provision of other source data in unstructured public depositories such as [Dryad](#) or [figshare](#), or as supplementary information. To maximize data reuse, we encourage publication of detailed descriptions of datasets in [Scientific Data](#).

Fuente: <https://www.nature.com/nature/for-authors/initial-submission>

Problemas de la ciencia abierta | Disponibilidad

A la hora de compartir datos no se pueden obviar **cuestiones legales** ni tampoco que estas sean una excusa para evitar hacerlo.

En caso de que sea imposible compartirlos, ni siquiera de forma agregada o filtrada, asegúrate de justificarlo.

 Developer Platform



Redistribution of Twitter content

If you need to share Twitter content you obtained via the Twitter APIs with another party, the best way to do so is by sharing Tweet IDs, Direct Message IDs, and/or User IDs, which the end user of the content can then rehydrate (i.e. request the full Tweet, user, or Direct Message content) using the Twitter APIs. This helps ensure that end users of Twitter content always get the most current information directly from us.

We permit limited redistribution of hydrated Twitter content via non-automated means. If you choose to share hydrated Twitter content with another party in this way, you may only share up to 50,000 hydrated public Tweet Objects and/or User Objects per recipient, per day, and should not make this data publicly available (for example, as an attachment to a blog post or in a public Github repository).

Fuente:

<https://developer.twitter.com/en/developer-terms/more-on-restricted-use-cases>

Problemas de la ciencia abierta | Reproducibilidad

No es suficiente con compartir los datos en bruto y el *script* en el que se ha realizado el análisis.

Es necesario garantizar que estos sean fácilmente **comprensibles** y que el proceso al completo pueda **reproducirse** sin problema.



Fuente: <https://hackernoon.com/its-time-we-code-in-english-e02df6b62ecc>

Problemas de la ciencia abierta | Reproducibilidad

SOLO UNA PEQUEÑA PARTE DE LOS RESULTADOS SON REPRODUCIBLES



PNAS PNAS PNAS

An empirical analysis of journal policy effectiveness for computational reproducibility

Victoria Stodden^{a,1}, Jennifer Seiler^b, and Zhaokun Ma^b

^aSchool of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL 61820; and ^bDepartment of Statistics, Columbia University, New York, NY 10027

Edited by David B. Allison, Indiana University Bloomington, Bloomington, IN, and accepted by Editorial Board Member Susan T. Fiske January 9, 2018 (received for review July 11, 2017)

A key component of scientific communication is sufficient information for other researchers in the field to reproduce published findings. For computational and data-enabled research, this has often been interpreted to mean making available the raw data from which results were generated, the computer code that generated the findings, and any additional information needed such as workflows and input parameters. Many journals are revising author guidelines to include data and code availability. This work evaluates the effectiveness of journal policy that requires the data and code necessary for reproducibility be made available postpublication by the authors upon request. We assess the effectiveness of such a policy by (i) requesting data and code from authors and (ii) attempting replication of the published findings. We chose a random sample of 204 scientific papers published in the journal *Science* after the implementation of their policy in February 2011. We found that we were able to obtain artifacts from 44% of our sample and were able to reproduce the findings for 26%. We find this policy—author remission of data and code postpublication upon request—an improvement over no policy, but currently insufficient for reproducibility.

reproducible research | data access | code access | reproducibility policy | open science

computational reproducibility of published results. We use a survey instrument to test the availability of data and code for articles published in *Science* in 2011–2012. We then use the scientific communication standards from the 2012 Institute for Computational and Experimental Research in Mathematics (ICERM) workshop report to evaluate the reproducibility of articles for which artifacts were made available (11). We then assess the impact of the policy change directly, by examining articles published in *Science* in 2009–2010 and comparing artifact ability to our postpolicy sample from 2011–2012. Finally, we discuss possible improvements to journal policies for enabling reproducible computational research in light of our results.

Results

We emailed corresponding authors in our sample to request the data and code associated with their articles and attempted to replicate the findings from a randomly chosen subset of the articles for which we received artifacts. We estimate the artifact recovery rate to be 44% with a 95% bootstrap confidence interval of the proportion [0.36, 0.50], and we estimate the replication rate to be 26% with a 95% bootstrap confidence interval [0.20, 0.32].



Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *Proceedings of the National Academy of Sciences*, 115(11), 2584–2589. <https://doi.org/10.1073/pnas.1708290115>



2.

Nuevas posibilidades

Preparación de los materiales

Contenidos

1. Problemas de la ciencia abierta
- 2. Nuevas posibilidades**
3. La tripleta ganadora
4. Buenas prácticas

Nuevas posibilidades

Existen una amplia variedad de posibilidades para llevar a cabo la difusión de estos materiales.

En lo relativo al código, el punto en común en buena parte está en el uso de *notebooks*.

communications physics

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [communications physics](#) > [comment](#) > [article](#)

[Comment](#) | [Open Access](#) | [Published: 19 August 2020](#)

Creating an executable paper is a journey through Open Science

[Jana Lasser](#) 

[Communications Physics](#) **3**, Article number: 143 (2020) | [Cite this article](#)

6234 Accesses | **3** Citations | **74** Altmetric | [Metrics](#)

Executable papers take transparency and openness in research communication one step further. In this comment, an early career researcher reports her experience of creating an executable paper as a journey through Open Science.

Lasser, J. (2020). Creating an executable paper is a journey through Open Science. *Communications Physics*, 3(1), 143. <https://doi.org/10.1038/s42005-020-00403-4>

Nuevas posibilidades | Wikis

Tanto para generar una documentación extensa como para incluir materiales complementarios, las **wikis** tienen mucha utilidad y están presentes en muchas plataformas.

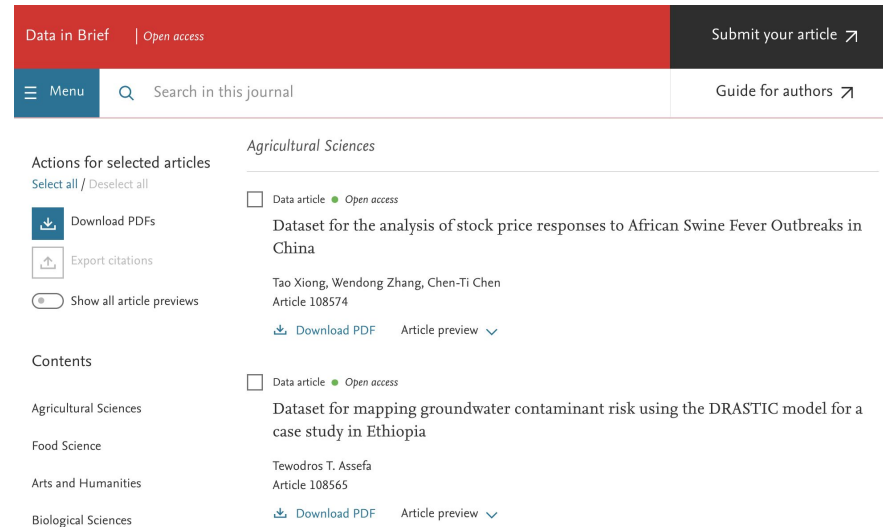
API	Tipo	Acceso	Documentación
Altmetric Details Page API	Agregador altmétrico	Público	DOC
Altmetric Explorer API	Agregador altmétrico	Privado	DOC
AMiner	Base de datos/agregador bibliográfico	Público	DOC
arXiv	Repositorio	Público	DOC
BASE	Base de datos/agregador bibliográfico	Público	DOC
CIRC	Índice	Privado	DOC
COCI	Índice	Público	DOC
CORE	Base de datos/agregador bibliográfico	Público	DOC , Swagger
Crossref	Base de datos/agregador bibliográfico	Público	DOC , Swagger , Simple Query
Crossref Event Data	Agregador altmétrico	Público	DOC
DataCite	Base de datos/agregador bibliográfico	Público	DOC
Dialnet CRIS	Base de datos/agregador bibliográfico	Privado	Swagger
Dimensions Metrics	Índice	Público	DOC
Dimensions Search Language	Base de datos/agregador bibliográfico	Privado	DOC
DOAJ	Base de datos/agregador bibliográfico	Público	DOC

Fuente: https://github.com/Wences91/Fuentes_de_informacion/wiki/APIs

Nuevas posibilidades | Paper de datos

Si los datos son lo suficientemente relevantes, puedes publicarlos en un **paper de datos**.

Es un pequeño paper en el que describes completamente los datos, ya alojados en algún repositorio. Con ello generas un recurso documentado y de mayor alcance.



The screenshot shows the 'Data in Brief' section of the ScienceDirect journal website. The header includes 'Data in Brief | Open access', a search bar, and a 'Submit your article' button. Below the header, there are navigation options like 'Menu' and 'Guide for authors'. The main content area is titled 'Agricultural Sciences' and lists two data articles. Each article entry includes a checkbox for 'Data article' (with an 'Open access' indicator), the article title, authors, and options to 'Download PDF' or view an 'Article preview'.

Actions for selected articles
Select all / Deselect all

- Download PDFs
- Export citations
- Show all article previews

Contents

- Agricultural Sciences
- Food Science
- Arts and Humanities
- Biological Sciences

Agricultural Sciences

- Data article Open access
Dataset for the analysis of stock price responses to African Swine Fever Outbreaks in China
Tao Xiong, Wendong Zhang, Chen-Ti Chen
Article 108574
[Download PDF](#) [Article preview](#)
- Data article Open access
Dataset for mapping groundwater contaminant risk using the DRASTIC model for a case study in Ethiopia
Tewodros T. Assefa
Article 108565
[Download PDF](#) [Article preview](#)

Fuente: <https://www.sciencedirect.com/journal/data-in-brief/>

Nuevas posibilidades | Notebooks

Aunque las notebooks nacieron en los 80s, su uso se ha popularizado con el auge de la ciencia de datos.

Integran en una misma interfaz **procesador de textos** (habitualmente en Markdown) y **programación**.

ALGUNAS PERMITEN LA INTERACCIÓN

The Lorenz Differential Equations

Before we start, we import some preliminary libraries. We will also import (below) the accompanying `lorenz.py` file, which contains the actual solver and plotting routine.

```
[1]: %matplotlib inline
      from ipywidgets import interactive, fixed
```

We explore the Lorenz system of differential equations:

$$\begin{aligned}\dot{x} &= \sigma(y - x) \\ \dot{y} &= \rho x - y - xz \\ \dot{z} &= -\beta z + xy\end{aligned}$$

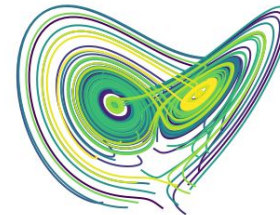
Let's change (σ, β, ρ) with ipywidgets and examine the trajectories.

```
[2]: from lorenz import solve_lorenz
      w=interactive(solve_lorenz, sigma=(0.0,50.0), rho=(0.0,50.0))
      w
```

sigma

beta

rho



TAMBIÉN SE MUESTRAN LOS GRÁFICOS

Nuevas posibilidades | Notebooks

Destaca sobre todo **Jupyter**, un proyecto nacido en 2014 para facilitar el desarrollo de software de código abierto y la programación interactiva.

Ofrece soporte para aproximadamente **40 lenguajes**, entre los que destacan Python, R y Julia.



Nuevas posibilidades | Notebooks

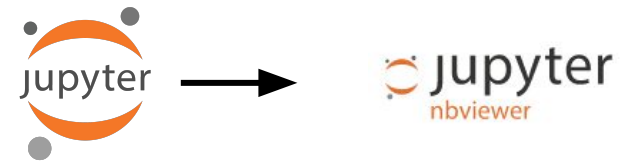
Las notebooks pueden compartirse en la red por medio de diferentes plataformas que permiten almacenarlas y/o visualizarlas correctamente.

No ofrece interacción pero hace posible su lectura.

Repositorio de notebooks



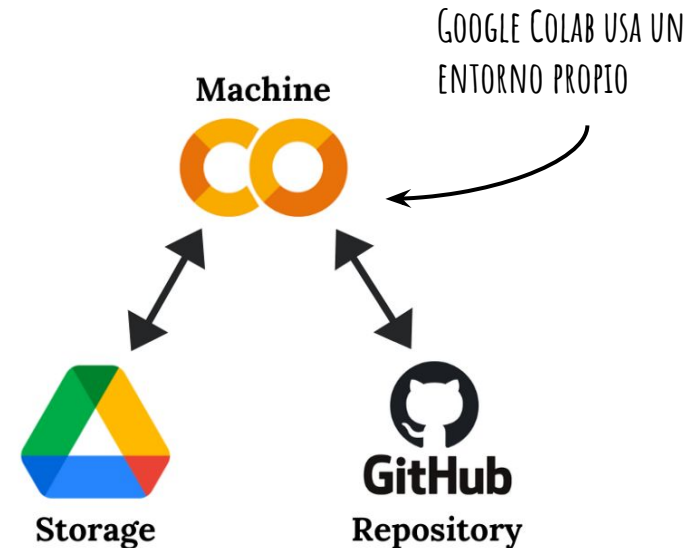
Visualizador de notebooks



Nuevas posibilidades | Entornos colaborativos

Es posible además trabajar directamente con código de manera **colaborativa** haciendo uso de notebooks.

Google Colab es una de las herramientas más populares. Permite integrar Google Drive y GitHub.



Fuente: <https://medium.com/analytics-vidhya/how-to-use-google-colab-with-github-via-google-drive-68efb23a42d>

Nuevas posibilidades | Entornos colaborativos

Aunque se trata de una opción gratuita, ofrece **recursos limitados**.

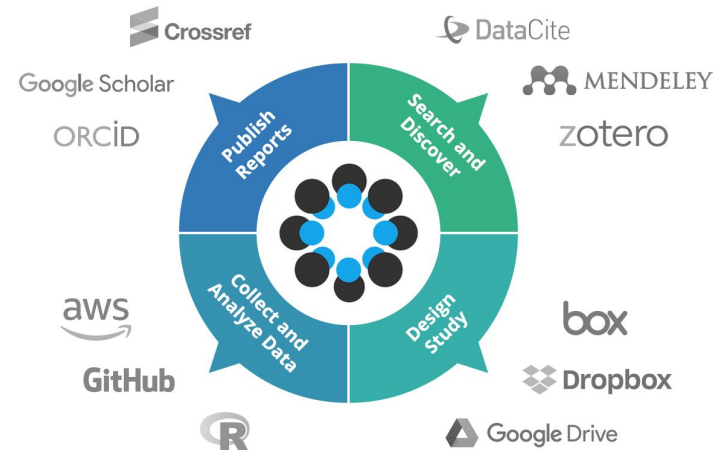


Existe una versión de pago así como una amplia variedad de alternativas.



Nuevas posibilidades | Entornos colaborativos

Otra posibilidad se encuentran en herramientas como OSF.io, que ofrece un completo entorno para la investigación, pudiendo integrar numerosas herramientas y repositorios.



Nuevas posibilidades | Reproducibilidad

Existen aplicaciones que permiten replicar de manera interactiva y online experimentos empleando las mismas condiciones que en el original para garantizar su reproducibilidad.



Nuevas posibilidades | Reproducibilidad

PAPER

Scientometrics (2021) 126:1725–1743
<https://doi.org/10.1007/s11192-020-03814-w>

Exploring WorldCat identities as an altmetric information source: a library catalog analysis experiment in the field of Scientometrics

Daniel Torres-Salinas¹ · Wenceslao Arroyo-Machado¹ · Mike Thelwall²

Received: 28 July 2020 / Accepted: 18 November 2020 / Published online: 3 January 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

Abstract

Assessing the impact of scholarly books is a difficult research evaluation problem. Library Catalog Analysis facilitates the quantitative study, at different levels, of the impact and diffusion of academic books based on data about their availability in libraries. The WorldCat global catalog collates data on library holdings, offering a range of tools including the novel WorldCat Identities. This is based on author profiles and provides indicators relating to the availability of their books in library catalogs. Here, we investigate this new tool to identify its strengths and weaknesses based on a sample of Bibliometrics and Scientometrics authors. We review the problems that this entails and compare Library Catalog Analysis indicators with Google Scholar and Web of Science citations. The results show that WorldCat Identities can be a useful tool for book impact assessment but the value of its data is undermined by the provision of massive collections of ebooks to academic libraries.

Torres-Salinas, D., Arroyo-Machado, W., & Thelwall, M. (2021). Exploring WorldCat identities as an altmetric information source: A library catalog analysis experiment in the field of Scientometrics. *Scientometrics*, 126(2), 1725-1743.
<https://doi.org/10.1007/s11192-020-03814-w>

CÓDIGO Y DATOS

File	Initial commit	3 years ago
data	Add files via upload	2 years ago
.gitignore	Initial commit	3 years ago
LICENSE	Initial commit	3 years ago
README.md	Update README.md	2 years ago
download_names.py	Add files via upload	3 years ago
download_uris.py	Add files via upload	3 years ago
environment.yml	Update environment.yml	3 years ago
index.ipynb	Add files via upload	3 years ago

REPRODUCCIÓN

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import worldcatidentities
from ipywidgets import interact, fixed, widgets

In [2]: import download_names as dn
import download_uris as du
```

First download

First we tried to automate the process of retrieving data from the WorldCat Identities API. However, the data downloaded required extensive review.

GitHub:

https://github.com/Wences91/library_catalog_wi/

Binder:

https://mybinder.org/v2/gh/Wences91/library_catalog_wi/master?urlpath=tree/index.ipynb

Nuevas posibilidades | Dashboards

Los materiales complementarios, especialmente todo lo relativo a visualizaciones pueden ser transformados en **dashboards** que permitan profundizar más en los resultados y de forma interactiva.



Nuevas posibilidades | Dashboards

INTERSECTIONAL INEQUALITIES

Home

Explore ▾

Data

Methods

About

Publications

Open positions ▾

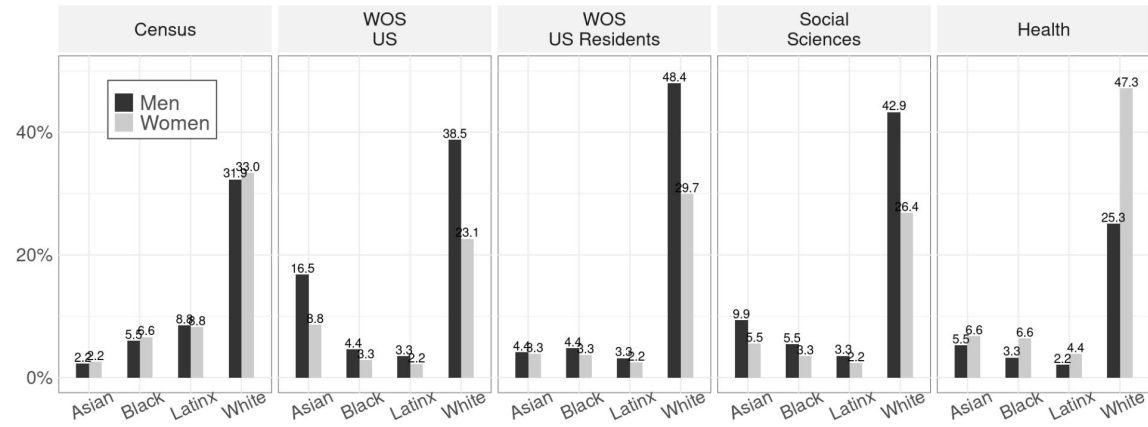
Aggregated Results

Population distribution

Disciplines, Specialities and Citations

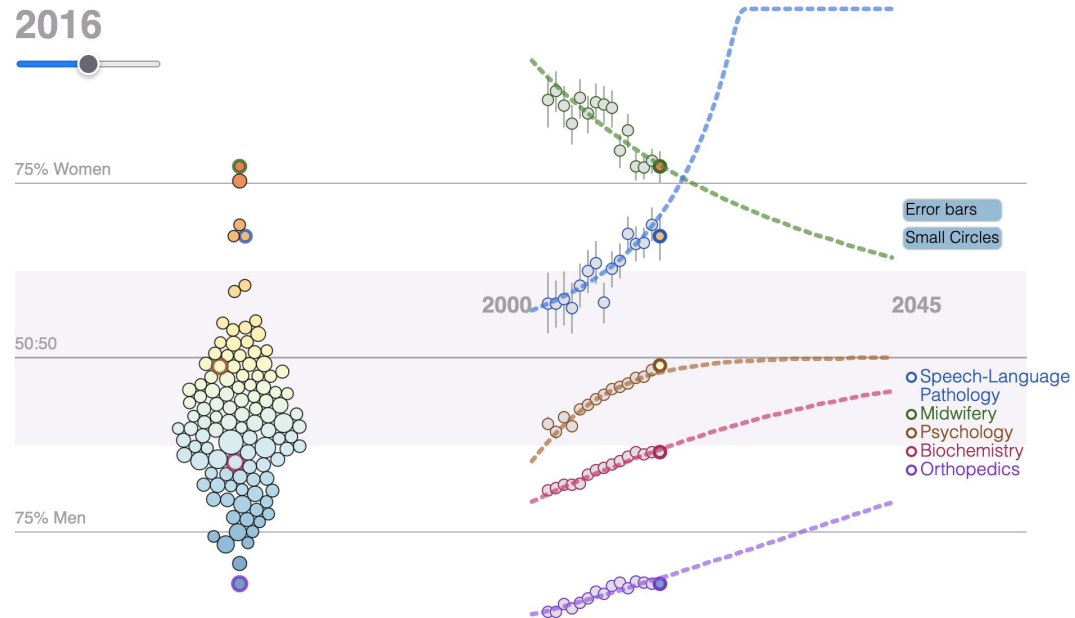
Speciality by Race and Gender

Population distribution by race & gender in US Census 2010. US first authors of Web Of Science 2008-2019: All disciplines, articles from the Health discipline, and articles from Social Sciences, Humanities and Professional disciplines



Fuente: <https://sciencebias.uni.lu/app/>

Nuevas posibilidades | Dashboards



Fuente: <https://lukeholman.github.io/genderGap/>

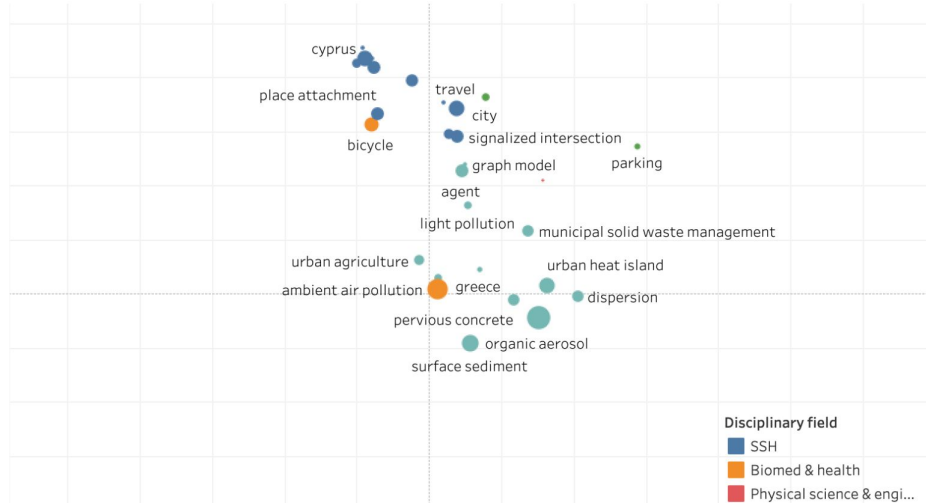
Nuevas posibilidades | Dashboards

UK Strings SDG to communities de [Ed Noyons](#)

map cluster_id1 search strings Dashboard 1 Sheet 4 Title SeedPublications Community pubs Selected communities

SDG 11 - Make cities and human settlements inclusive, safe, resilient and sustainable

Science Landscape



SDG
11

Filter clusters by number and share of publications retrieved by search strategy (seeds)

Minimum # seeds
50

and

Minimum share seeds
0,2

Sample pubs in selected clusters

pubid	
34794989	478
37075101	306
36860598	157
36797102	150
34859084	122
34722773	119
36238198	115
36221741	OR

Fuente: <https://public.tableau.com/views/UKStringsSDGtoCommunities/Dashboard1>



3.

Nuevas posibilidades

Asegurando la disponibilidad y reproducibilidad

Contenidos

1. Problemas de la ciencia abierta
2. Nuevas posibilidades
- 3. La tripleta ganadora**
4. Buenas prácticas

La tripleta ganadora | Notebooks

No es suficiente con compartir una notebook. Ciertos aspectos pueden interferir en la correcta ejecución de dichos resultados, como dependencias, versiones concretas de paquetes...

Es necesario ofrecer un entorno **concreto** para su ejecución.



La tripleta ganadora | Tripleta

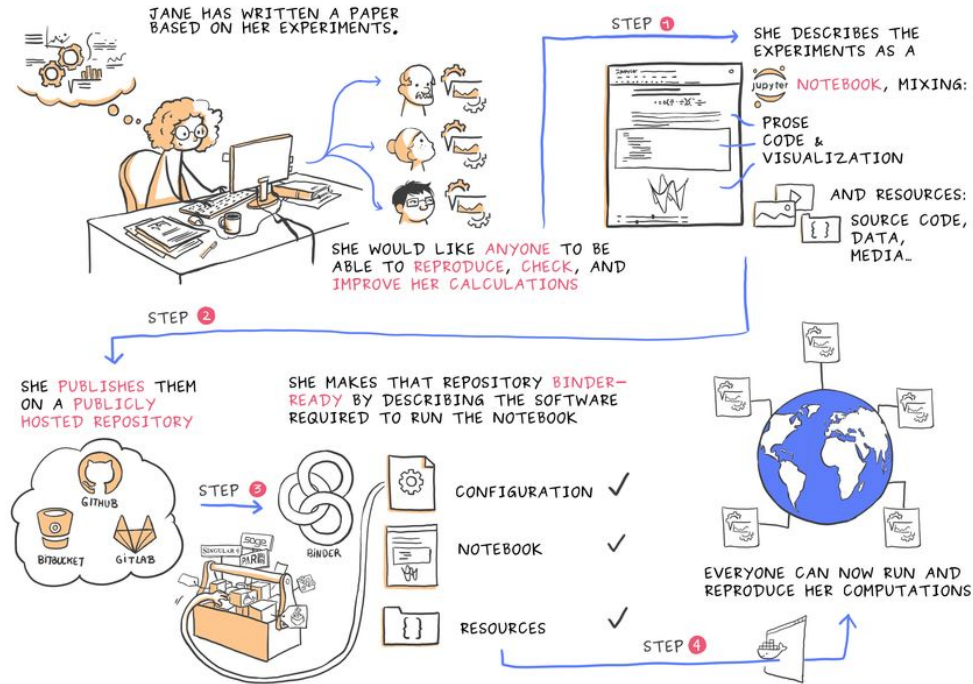
Existen herramientas que permiten partir de repositorios con código y datos y crear todo un entorno con el que ejecutarlo.

Crean **contenedores** que incluyen la configuración exacta para el correcto desarrollo de la aplicación.



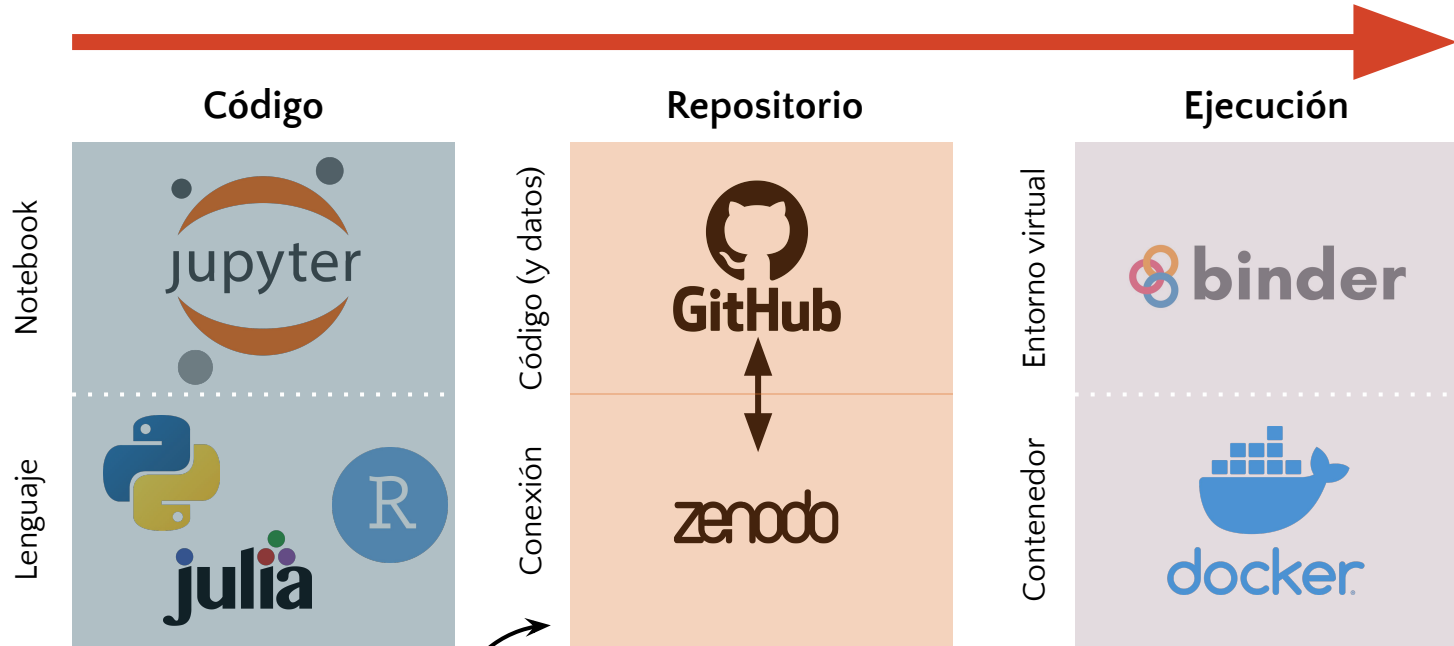
Fuente: <https://bibliometriaobarbarie.com/2019/11/22/trifuerza-open-science/>

La tripleta ganadora | Tripletta



Beg, M., Taka, J., Kluyver, T., Konovalov, A., Ragan-Kelley, M., Thiéry, N. M., & Fangohr, H. (2021). Using Jupyter for reproducible scientific workflows. *Computing in Science & Engineering*, 23(2), 36–46.
<https://doi.org/10.1109/MCSE.2021.3052101>

La tripleta ganadora | Tripleta



SE GENERA UNA VERSIÓN UNÍVOCA



4.

Buenas prácticas

Asegurando la disponibilidad y reproducibilidad

Contenidos

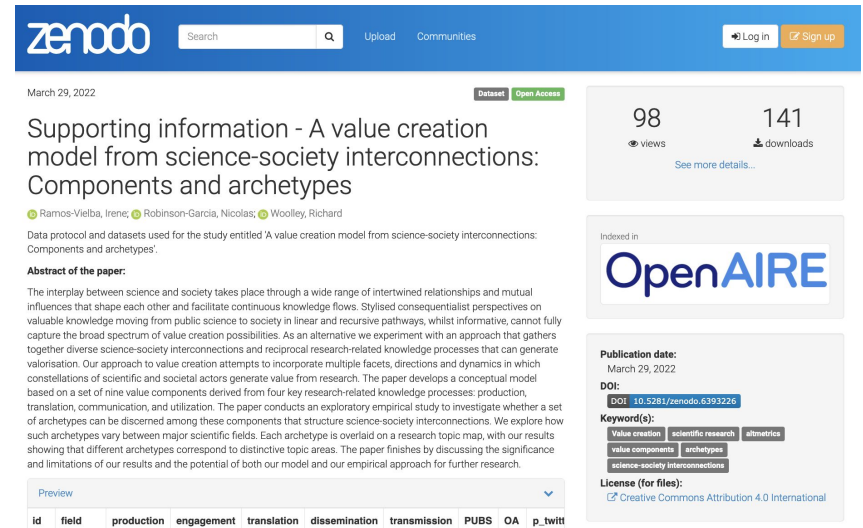
1. Problemas de la ciencia abierta
2. Nuevas posibilidades
3. La tripleta ganadora
- 4. Buenas prácticas**

Buenas prácticas | Datos

1. Usa repositorios y no apéndices

Si tienes previsto compartir algún material de investigación haz uso de repositorios en lugar de incluirlo como material complementario.

Genera recursos abiertos, identificables y citables.



The screenshot shows the Zenodo interface for a dataset. At the top, there is a search bar and navigation links for 'Upload' and 'Communities'. The dataset title is 'Supporting information - A value creation model from science-society interconnections: Components and archetypes', dated March 29, 2022. It has 98 views and 141 downloads. The authors listed are Ramos-Vielba, Irene; Robinson-Garcia, Nicolas; and Woolley, Richard. The abstract describes the interplay between science and society, focusing on knowledge flows and value creation. The dataset is part of the OpenAIRE program. The DOI is 10.5281/zenodo.6393226. The license is Creative Commons Attribution 4.0 International. A table at the bottom lists the fields: id, field, production, engagement, translation, dissemination, transmission, PUBS, OA, and p_twitter.

Ramos-Vielba, I., Robinson-Garcia, N., & Woolley, R.. (2022). Supporting information - A value creation model from science-society interconnections: Components and archetypes (Version 2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.6393226>

Buenas prácticas | Datos



March 11, 2022

Dataset Open Access

Wikipedia Knowledge Graph dataset

Arroyo-Machado, Wenceslao; Torres-Salinas, Daniel; Costas, Rodrigo

Wikipedia is the largest and most read online free encyclopedia currently existing. As such, Wikipedia offers a large amount of data on all its own contents and interactions around them, as well as different types of open data sources. This makes Wikipedia a unique data source that can be analyzed with quantitative data science techniques. However, the enormous amount of data makes it difficult to have an overview, and sometimes many of the analytical possibilities that Wikipedia offers remain unknown. In order to reduce the complexity of identifying and collecting data on Wikipedia and expanding its analytical potential, after collecting different data from various sources and processing them, we have generated a dedicated Wikipedia Knowledge Graph aimed at facilitating the analysis, contextualization of the activity and relations of Wikipedia pages, in this case limited to its English edition. We share this Knowledge Graph dataset in an open way, aiming to be useful for a wide range of researchers, such as informetricians, sociologists or data scientists.

There are a total of 9 files, all of them in tsv format, and they have been built under a relational structure. The main one that acts as the core of the dataset is the **page** file, after it there are 4 files with different entities related to the Wikipedia pages (**category**, **url**, **pub** and **page_property**) and 4 other files that act as "intermediate tables" making it possible to connect the pages both with the latter and between pages (**page_category**, **page_url**, **page_pub** and **page_link** files).

The document **Dataset_summary** includes a detailed description of the dataset.

← PEQUEÑA GUIA EN PDF

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022).
Wikipedia Knowledge Graph dataset [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.6346900>

2. Contextualiza y describe

Asegúrate de que quienes encuentren los datos sepan de dónde proceden, qué contienen y qué utilidad tienen.

Las etiquetas son igualmente importantes para su localización.

Buenas prácticas | Datos

3. No desperdicias materiales

Cualquier material generado durante la investigación puede ser reaprovechado. No desperdicias pero tampoco quieras incluir todo lo que has hecho.

1. Introduction

In the Social Sciences and the Humanities especially, of the academic book, has been bibliometry has, until quite recently, been more restricted to limited, partial applications of the lack of more ambitious initiatives and proposed set of indicators based on consulting thanks, above all, to certain technology protocol—and, especially, the Online Computer WorldCat.org, in 2006. This open-access catalogues enabling users to determine where the library count—based methodology was in Library Holdings Analysis and was one of the use of citations; it was launched two years before the Altmetric manifesto (Priem et al., 2010) was published. Since then, the framework and methods enabling researchers to analyze the impact, diffusion and use of books, have broadened substantially.

1. Introducción

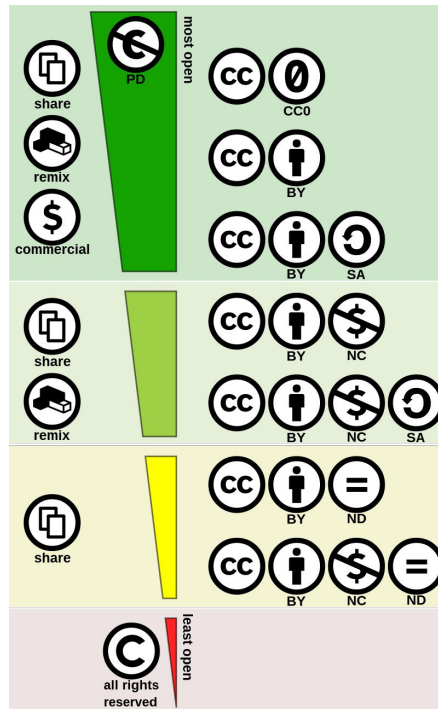
La evaluación de la actividad científica en el ámbito de las Ciencias Humanas y Sociales y, especialmente del libro académico, ha sido una de las asignaturas pendientes de la bibliometría cuyo contexto evaluativo, hasta hace bien poco, ha estado monopolizado por indicadores de citación y por las bases de datos de Thomson Reuters (Nederhof, 2006). De esta manera, pese a que casi todos los estudios manifestaban la importancia de los libros en la comunicación científica (Archambault et al., 2006; Hicks, 1999; Huang & Chang, 2008), la mayoría de las propuestas en torno a la evaluación del libro eran aplicaciones parciales de poco alcance basadas en los índices de citas tradicionales. La falta de iniciativas globales y de base de datos alternativas se vio enriquecida cuando en 2009 se propusieron una serie de indicadores basados en la consulta de *Online Public Access Catalogue (OPAC)*²⁰ gracias, sobre todo, a diversos desarrollos tecnológicos como el protocolo Z39.50 pero, especialmente, al lanzamiento en 2006 de *WorldCat.org* por parte de la Online Computer Library Center (OCLC). Este catálogo de libre acceso unificó en un buscador millones de bibliotecas de todo el mundo indicando en cuales de ellas podíamos localizar un determinado título (Nilges, 2006). La metodología basada en el conteo de bibliotecas se dio a conocer como *Library Catalog Analysis (LCA)* o *Library Holdings Analysis* y es una de las primeras alternativas evaluativas frente a las citas, lanzada dos años antes de que se promulgase el *manifesto altmetric* (Priem et al., 2010). Desde entonces se ha ampliado sustancialmente el marco y los métodos a partir de los cuales se puede analizar el impacto, difusión y uso de los libros.

Unrevised version to be published in *Evaluative informetrics – the art of metrics based research assessment. Festschrift in honour of Henk F. Moed*, edited by Cinzia Daraio and Wolfgang Glänzel

*Spanish version included just after it in page 19

Torres-Salinas, D., & Arroyo-Machado, W. (2019). Library Catalog Analysis and Library Holdings Counts: origins, methodological issues and application to the field of Informetrics. *Evaluative Informetrics – the Art of Metrics Based Research Assessment. Festschrift in Honour of Henk F. Moed*. <https://doi.org/10.5281/zenodo.3477620>

Buenas prácticas | Datos



4. Escoge la mejor licencia

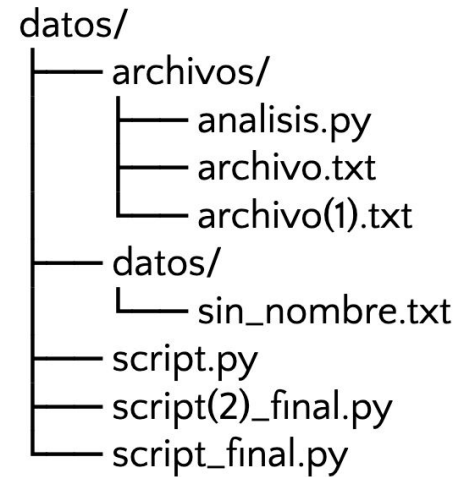
No descuides este aspecto, la elección de la licencia es fundamental para garantizar la reutilización de los datos.

No temas al Dominio Público (CC0).

Buenas prácticas | Código

5. Estructura correctamente el proyecto

Empieza desde el inicio trabajando con los entornos tradicionales y mediante scripts. Una vez estés familiarizado y tengas preparado el análisis pasa a la notebook.



Buenas prácticas | Código

Aprende git

build `passing`

Repositorio con el material para el libro *Aprende git*. Se escribió originalmente usando *Markdown* en su versión *Kramdown*, aunque finalmente hemos encontrado que funciona mejor transformándolo con *Pandoc*.

Te lo puedes [descargar](#) en cualquiera de las versiones en ePub, [comprar](#) en Amazon en formato ebook o en formato físico. Si necesitas generar otro formato, puedes inspirarte en los conversores que [hay](#) en el directorio `utils`.

Índice

1. Introducción.
2. Uso básico.
3. Resolviendo conflictos y otros problemas.
4. Flujos de trabajo y otras buenas prácticas.
5. Trabajando en GitHub.
6. Trabajando con *hooks* y otros temas de fontanería.

Fuente: <https://github.com/ll/aprende-git>

6. Usa control de cambios

Utiliza desde el inicio GitHub o herramientas relacionadas. El control de cambios mediante git ofrece muchas ventajas y te facilita la organización.

Buenas prácticas | Código

7. No empieces desde una notebook

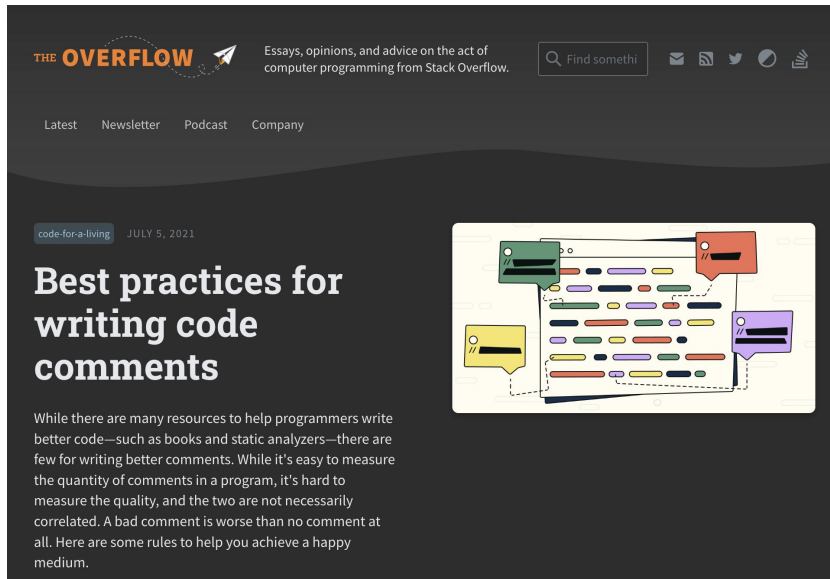
Empieza trabajando con los entornos tradicionales y mediante scripts. Una vez estés familiarizado y tengas preparado el análisis pasa a la notebook.



The screenshot shows the top of a Medium article. The header is dark blue with the text 'towards data science' on the left, 'Sign in' and 'Get started' buttons on the right, and a profile icon. Below the header is a navigation bar with 'Follow', '598K Followers', and several category links: 'Editors' Picks', 'Features', 'Deep Dives', 'Grow', 'Contribute', and 'About'. The article title is '5 Reasons why you should Switch from Jupyter Notebook to Scripts' in large, bold black font. Below the title is the subtitle 'Using Scripts Helps me Realize the Drawbacks of Jupyter Notebook'. The author's name 'Khuyen Tran' is shown with a profile picture, along with the date 'Aug 24, 2020' and '7 min read'. At the bottom right of the article preview are social media sharing icons for Twitter, Facebook, LinkedIn, and a link icon, followed by a three-dot menu.

Fuente: <https://towardsdatascience.com/how-and-why-to-share-scientific-code-64fbd385a67>

Buenas prácticas | Código



Fuente: <https://stackoverflow.blog/2021/07/05/best-practices-for-writing-code-comments/>

8. Comenta el código

Desde el comienzo realiza anotaciones, tanto dentro como fuera del código. Puedes usar un pequeño documento a modo de guía para identificar correctamente todos los archivos y pasos.

No olvides los archivos README.

Buenas prácticas | General

9. No te compliques innecesariamente

Si ya existe un paquete que hace lo que necesitas o si un repositorio o lenguaje se acomodan perfectamente a tu problema, no hagas cambios al respecto.

Repository Name	Information on fees/costs	Size limits	Integrated with <i>Scientific Data's</i> manuscript submission system	Re3data / FAIRsharing entry
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated	Yes ✓	view FAIRsharing entry
figshare	100 GB free per <i>Scientific Data</i> manuscript. Additional fees apply for larger datasets	1 TB per dataset	Yes ✓ - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. Download instructions.	view FAIRsharing entry
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	view re3data entry
Open Science Framework	Free of charge	5 GB per file, multiple files can be uploaded	No	view FAIRsharing entry
Zenodo	Donations towards sustainability encouraged	50 GB per dataset	No	view re3data entry
Science Data Bank	Free of charge	8 GB per file, no limit to dataset size	No	view FAIRsharing entry

Fuente: <https://www.nature.com/sdata/policies/repositories>

Buenas prácticas | General

Publication date:

March 11, 2022

DOI:

DOI [10.5281/zenodo.6346900](https://doi.org/10.5281/zenodo.6346900)

Keyword(s):

[Wikipedia](#) [dataset](#) [Knowledge Graph](#) [Informetrics](#)

Related identifiers:

Compiled by

[10.5281/zenodo.6959429](https://doi.org/10.5281/zenodo.6959429) (Software)

Described by

[10.1162/qss_a_00226](https://doi.org/10.1162/qss_a_00226) (Journal article)

[10.48550/arXiv.2210.13830](https://doi.org/10.48550/arXiv.2210.13830) (Preprint)

Reviewed by

[10.5281/zenodo.6958973](https://doi.org/10.5281/zenodo.6958973) (Software)

License (for files):

[Creative Commons Zero v1.0 Universal](#)

10. Enlaza todos los recursos

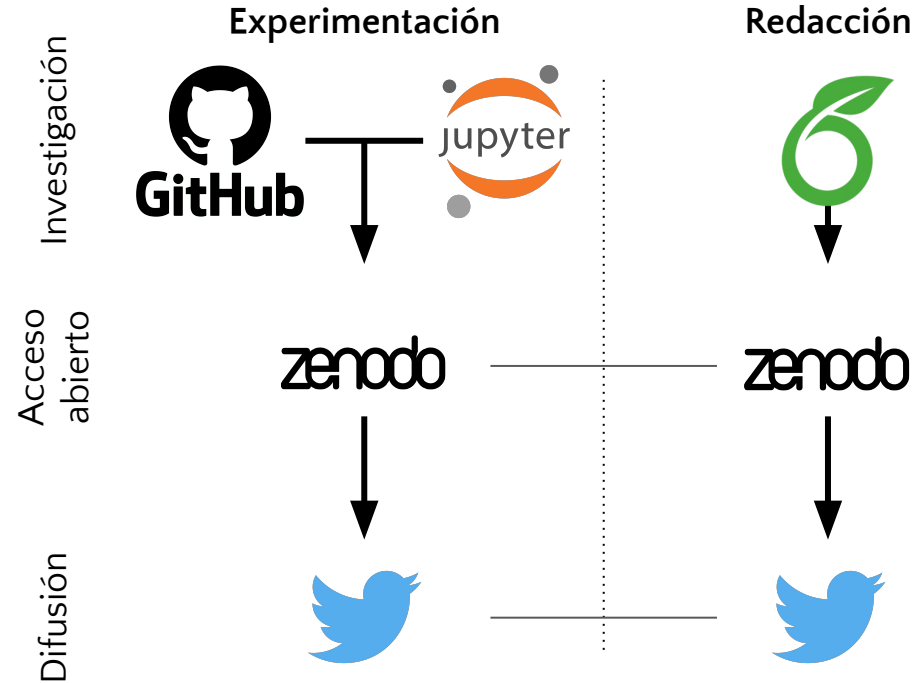
Conecta todos los resultados de investigación que estén bajo un mismo proyecto. Los repositorios permiten establecer dichas conexiones.

Arroyo-Machado, W., Torres-Salinas, D., & Costas, R. (2022).
Wikipedia Knowledge Graph dataset [Data set]. Zenodo.
<https://doi.org/10.5281/zenodo.6346900>

Buenas prácticas | General

11. Lleva a cabo un plan de difusión

Empieza trabajando con los entornos tradicionales y mediante scripts. Una vez estés familiarizado y tengas preparado el análisis pasa a la notebook.



Buenas prácticas | General



ORCID

ROR

12. Identifica todo

Haz uso en todo momento de identificadores como el DOI para los recursos, el ORCID para las personas y el ROR para instituciones.



¡Muchas gracias!

¿Alguna pregunta?

zenodo

Presentación disponible en <https://doi.org/10.5281/zenodo.7379053>

