
What is Open Science? Why Should You Care?

— Mette Peters —

Discussion

- Science has a problem
- What if we change the scientific paradigm to make research
 - Available
 - Accessible
 - Transparent
 - Inclusive
 - Democratic

We have a problem

Replication
Crisis

Misinformation

Fraud

New
Therapeutics

Slow
Publication
Process

Publish or
Perish

Waste

Paywalls

A bit about me

- Senior Advisor at the NIA/NIH in the Division of Neuroscience
- PhD in cell and molecular biology
- Worked in data science related positions in academia and pharma
- Passionate about Open Science

Disclaimer

The opinions expressed here are the presenter's own and do not reflect the view of the National Institute on Aging, the National Institutes of Health, the Department of Health and Human Services, or the United States government.

What is Open Science

- The maverick lone scientist is a myth - science is like building a house. New discoveries are build on previous work
- Open Science is a movement to make scientific research available, accessible, transparent, inclusive, and democratic
 - Includes knowledge, protocols, licenses, samples, data, code, hardware

UNESCO Open Science

Open science is a set of principles and practices that aim to make scientific research from all fields accessible to everyone for the benefits of scientists and society as a whole. Open science is about making sure not only that scientific knowledge is accessible but also that the production of that knowledge itself is inclusive, equitable and sustainable.

- UNESCO member states are asked to
 - Promote a shared understanding of open science and set out diverse paths to achieving it
 - Develop an enabling policy environment for open science
 - Invest in infrastructure and activities that contribute to open science
 - Invest in training, education, digital literacy and capacity-building to support open science
 - Foster a culture of open science and align incentives to support it
 - Promote innovative approaches for open science at all stages of the scientific process
 - Encourage international and multi-stakeholder cooperation in the context of open science to reduce gaps in technology and knowledge

The Year of Open Science!

JANUARY 11, 2023

FACT SHEET: Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research



OSTP ► NEWS & UPDATES ► PRESS RELEASES

OSTP launches Year of Open Science to advance national open science policies across the federal government in 2023

Today, the White House Office of Science and Technology Policy (OSTP) announced new actions to advance open and equitable research, including new grant funding, improvements in research infrastructure, broadened research participation for emerging scholars, and expanded opportunities for public engagement. OSTP is also launching the *Year of Open Science*, featuring actions across the federal government throughout 2023 to advance national open science policy, provide access to the results of the nation's taxpayer-supported research, accelerate discovery and innovation, promote public trust, and drive more equitable outcomes.

<https://www.whitehouse.gov/ostp/news-updates/2023/01/11/fact-sheet-biden-harris-administration-announces-new-actions-to-advance-open-and-equitable-research/>

Open Science aims to

Making sure that scientific knowledge is accessible

Making sure that the production of scientific knowledge is inclusive

Facilitate and reward team science

Open the process to knowledge creation to people outside the research community



Scholarly publishing

A \$19 billion industry

Article

Global evidence of rapid urban growth in flood zones since 1985

<https://doi.org/10.1038/s41586-023-06468-9>

Received: 17 March 2022

Accepted: 21 July 2023

Jun Rentschler^{1,2*}, Paolo Avner¹, Mattia Marconcini^{2,3}, Rui Su¹, Emanuele Strano³,
Michalis Vousdoukas⁴ & Stéphane Hallegatte¹

Disaster losses are increasing and evidence is mounting that climate change is driving

Global evidence of rapid urban growth in flood zones since 1985"

You are viewing an article preview. The following formats are available to purchase through ReadCube:

\$9.99 [Rent for 48 hours](#)
Printing and saving restrictions apply

\$24.99 [Buy Cloud Access](#)
Printing and saving restrictions apply

\$39.95 [Buy PDF](#)

Slow
Publication
Process

Paywall

Some publishing solutions

- Publications supported by NIH funding
 - Must be made available on PubMed Central one year from official date of publication
- Open access manuscripts
 - Nature Journal = \$11,600
- Public Library of Science (PLOS) journals
 - Non-profit open access journals
 - Much lower publication fee
- Preprints
 - bioRxiv
 - medRxiv

Some unintended problems - predatory journals

- Open access journals taking advantage of publishing for a fee
- No editorial board
- Sham peer review
- Misinformation
- Pseudoscience
- Fraud
- Somewhat easy to detect if you get a request to publish
- Much harder as a reader
- There is a list - Predatory Reports

How do we solve that?

Show me the data

Article

Global evidence of rapid urban growth in flood zones since 1985

<https://doi.org/10.1038/s41586-023-06468-9>

Jun Rentschler^{1,2}, Paolo Avner¹, Mattia Marcon¹,
Michalis Vousdoukas³ & Stéphane Hallegatte¹

Received: 17 March 2022

Accepted: 21 July 2023

Disaster losses are increasing and evidence

Data availability

The WSF-Evo dataset is publicly available for download from

<https://geoservice.dlr.de/web/maps/eoc:wsfevolution>. We use proprietary global fluvial and pluvial flood-hazard data with the permission of Fathom Global, who provide the data for academic purposes and can be contacted at <https://www.fathom.global/contact-us/>. We use coastal flood maps developed by Vousdoukas et al.⁴⁰, which are publicly available for download from <https://doi.org/10.5281/zenodo.8057902>. Country-level summary results and subnationally and annually disaggregated results are provided in the [Supplementary Information](#) to this study.

NIH



**Taxpayer
money**

- NIH FY2023 budget is \$49 billion
- Money goes to intra- and extramural grants and contracts
- Most of that funding results in the generation of lots of data
- Should that data remain the property of the funding recipient?

Who owns taxpayer funded data?

You do

- As part of the year of Open Science NIH has issued a new Data Management and Sharing (DMS) Policy effective January 2023
- Funding recipients must
 - Prospectively plan for the managing and sharing of scientific data
 - Submit a DMS plan
 - Submit a DMS budget
 - Comply with the approved plan
- Key component of this - DMS plan must meet specifications of How data is shared

Components of the DMS plan



The
important
part

- Must describe the data types and how much
- Must describe the metadata
- Must share related tools, software, code
- Must specify what repository the data will be shared in
- Must describe how people will be able to find the data
- Must contain information about data use restrictions, and protection of privacy and rights of human research participants
- Must describe how, and who, compliance with the DMS plan will be monitored
- Yearly progress reports to the NIH must refer to progress on data sharing per DMS plan

The principle of FAIR data

It's not enough to share data. It has to be usable

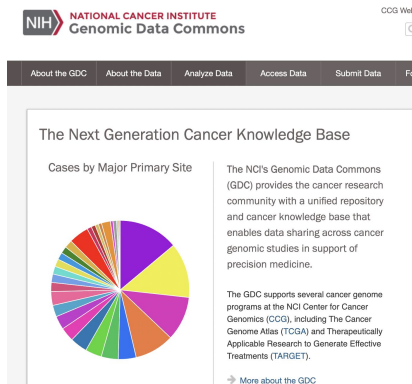
- Findable
 - You can't take advantage of Open Data if you can't find it. Data discovery is number 1
- Accessible
 - Once you find it, it should be simple to understand how to access it
- Interoperable
 - You should be able to integrate data with other data or computational workflows for processing or analysis
- Reusable
 - Data and metadata should be well described so that so that any replication of analysis or novel use can be done

Data repositories

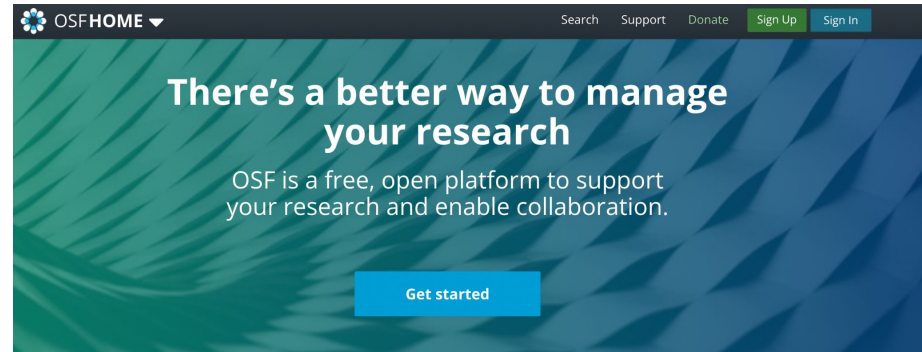
- NIH supported data repositories
- Should use on that is the most appropriate based on data type and discipline
- Also several generalist repositories

As open as possible

Access to scientific knowledge should be as open as possible, but sometimes access may need to be restricted, for example to protect human rights, confidentiality, intellectual property rights, personal information, threatened or endangered species, and sacred and secret indigenous knowledge. Open science encourages scientists to develop tools and methods for managing data so that as much data as possible can be shared, as appropriate.



The screenshot shows the NIH Genomic Data Commons website. The header includes the NIH logo and the text "NATIONAL CANCER INSTITUTE Genomic Data Commons". Below the header is a navigation menu with options: "About the GDC", "About the Data", "Analyze Data", "Access Data", "Submit Data", and "For". The main content area is titled "The Next Generation Cancer Knowledge Base" and features a section "Cases by Major Primary Site" with a colorful pie chart. To the right of the chart, there is text describing the GDC's mission: "The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified repository and cancer knowledge base that enables data sharing across cancer genomic studies in support of precision medicine." Below this, it mentions that the GDC supports several cancer genome programs at the NCI Center for Cancer Genomics (CCG), including The Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research to Generate Effective Treatments (TARGET). A link "More about the GDC" is provided at the bottom.



The screenshot shows the OSF HOME website. The header includes the OSF HOME logo and a navigation menu with options: "Search", "Support", "Donate", "Sign Up", and "Sign In". The main content area features a large blue background with a white text overlay that reads: "There's a better way to manage your research". Below this, it states: "OSF is a free, open platform to support your research and enable collaboration." A prominent blue button labeled "Get started" is positioned at the bottom of the main content area.

Back to publications and data availability

- Publications should have a Data Availability statement
- Statement should point directly to the data used deposited in a legit repository
 - Manuscript supplementary materials is not a good place to discover data
 - 'Available upon request' does not qualify
- In order to replicate what was done, data should meet FAIR principles
 - Data (raw and processed), metadata, and code
- A manuscript making claims, but providing no supporting data should raise a red flag

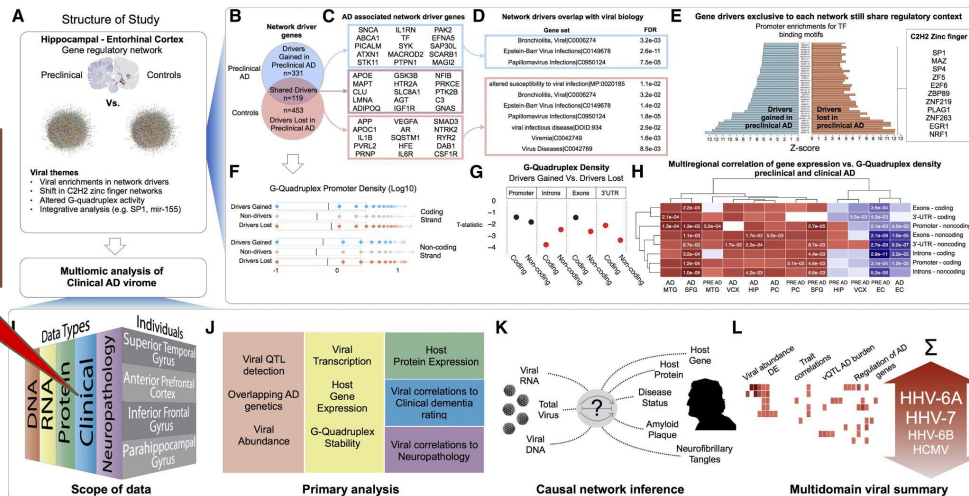
Data reuse -> novel research

It's not just about replication

Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus

Ben Readhead¹⁷ • Jean-Vianney Haure-Mirande¹⁷ • Cory C. Funk • ... Michelle E. Ehrlich • Sam Gandy¹⁷ • Joel T. Dudley^{17, 18}   • Show all authors • Show footnotes

Open Archive • Published: June 21, 2018 • DOI: <https://doi.org/10.1016/j.neuron.2018.05.023>



Existing data

National Institute on Aging

Health Information | Research & Funding | News & Events | About NIA

Home | News & Events | Could a viral illness increase chances of developing Alzheimer's or other neurodegenerative disease...

ANNOUNCEMENTS
Could a viral illness increase chances of developing Alzheimer's or other neurodegenerative disease?
 January 19, 2023
 Neuroscience | Alzheimer's Disease | Dementias

NIH biobank study suggests vaccinations against viruses may also reduce risk of neurological disorders

Changing incentives



Publish or
Perish

- NIH biosketches - a document describing your qualifications and experience for a specific role in a funding application
 - Has a contribution to science section - encouraged to add data contribution
- Data attribution is fundamental, but needs to be codified in academia and funding agencies
- Data contribution needs to become a scholarly product

Thank You!