
Open Science in Bioinformatics

Caleb Kibet



What to expect...

- What is Open Science
- Why Open Science
- Open Science Taxonomy
- Open Science Tools
- Open Science in Bioinformatics



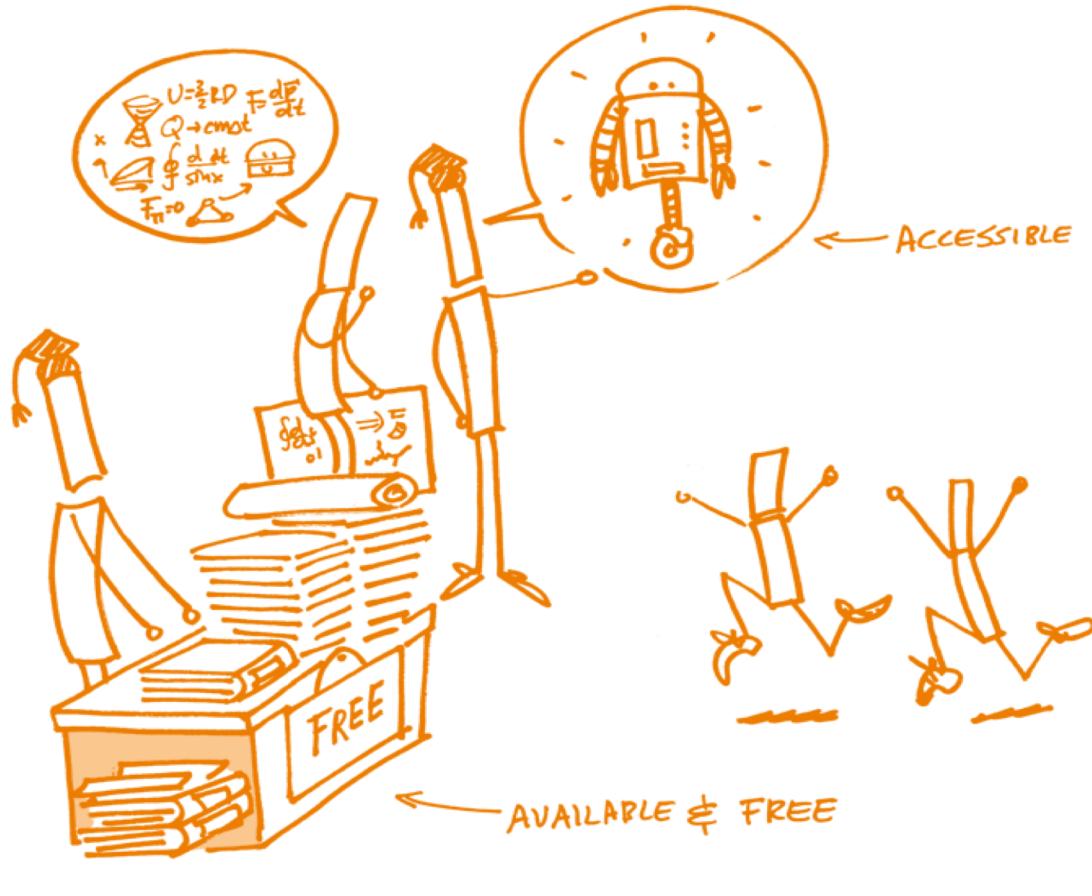
What is Open Science ?

Open Science is the practice of **science** in such a way that others can **collaborate** and **contribute**, where research data, lab notes and other research processes are **freely available**, under terms that enable **reuse**, **redistribution** and **reproduction** of the research and its underlying data and methods.

(<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>)



4 FUNDAMENTAL RULES OF OPEN SCIENCE



Why Open Science ?

"When all researchers are aware of Open Science, and are trained, supported and guided at all career stages to practice Open Science, the potential is there to **fundamentally change the way** research is performed and disseminated, **fostering a scientific ecosystem** in which research gains increased visibility, is shared more efficiently, and is performed with enhanced research integrity."

[Open Science Skills Working Group Report](#) (2017)



Benefits of Open Access for you and society



Source: Australian Open Access Support Group





eLIFE

ABOUT LABS COMMUNITY [SUBMIT MY RESEARCH](#)

LOG IN/REGISTER

MENU HOME MAGAZINE



BIOCHEMISTRY AND CHEMICAL BIOLOGY



Point of View: How open science helps researchers succeed



Erin C McKiernan , Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, Brian A Nosek, Karthik Ram [see all »](#)

National Autonomous University of Mexico, Mexico; National Institutes of Health, United States; University of California, Davis, United States; Laura and John Arnold Foundation, United States; BioMed Central, United Kingdom; CrossRef, United Kingdom; University of Texas at Austin, United States; Center for Open Science, United States; University of California, Berkeley, United States [see all »](#)

FEATURE ARTICLE Jul 7, 2016



@Calkibet

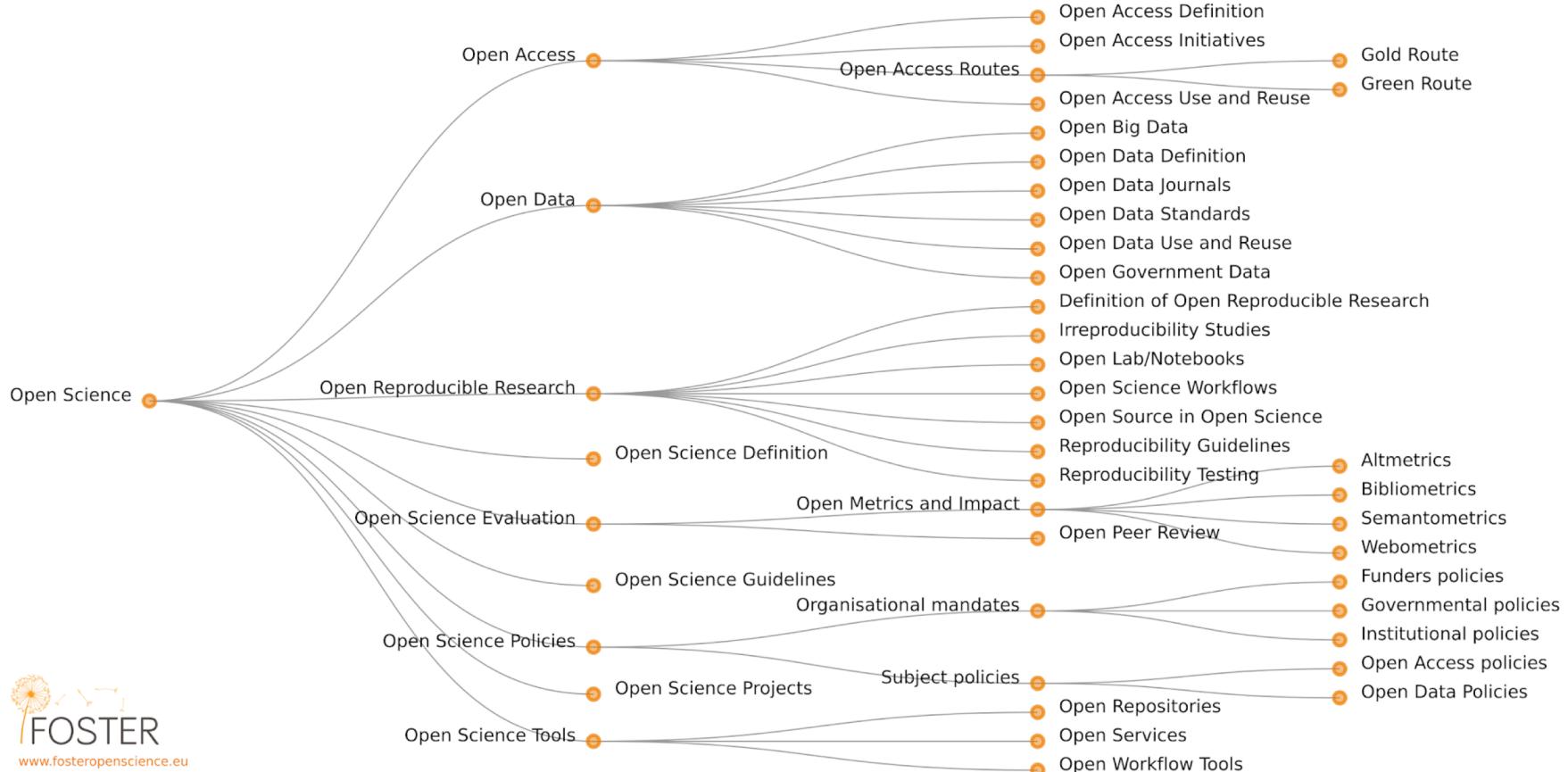
Open Science is good for the researcher...

1. **Increased quality of research** from reduced errors and fraud due to wider scrutiny and evaluation brought about by transparency
2. **Increased opportunities** for both local and global participation in research
3. **Faster transfer of knowledge** required to solve problems
4. **Fosters innovation** which produces new products and services
5. **Improves productivity and research output** due to reduced duplication
6. Promote awareness among citizens which **improves willingness in participation** in experiments and data collection.



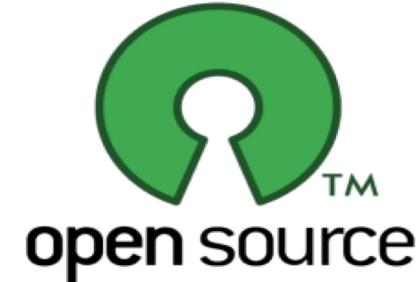
Open Science Taxonomy

Open Science Taxonomy





open peer review



Your Open Data should adhere to...

FAIR Principles

Make your data:

- **Findable**
- **Accessible**
- **Interoperable**
- **Reusable**



- Findable**
- Descriptive metadata
 - Persistent Identifiers

Accessible

- Determining what to share
- Participant consent and risk management
- Access status

Interoperable

- XML standards
- Data Documentation Initiative
- CDISC

Reusable

- Rights and licence models
- Permitted and non-permitted use

<http://datafairport.org/>



The monster of Paywalls and Impact Factors...

In 2010, Elsevier's reported a 36% profit margin – higher than Apple, Google, or Amazon posted that year



The long read

Is the staggeringly profitable business of scientific publishing bad for science?

It is an industry like no other, with profit margins to rival Google - and it was created by one of Britain's most notorious tycoons: Robert Maxwell. By Stephen Buranyi

A thumbnail image of a Guardian newspaper article. The title is "Is the staggeringly profitable business of scientific publishing bad for science?". Below the title is a short blurb: "It is an industry like no other, with profit margins to rival Google - and it was created by one of Britain's most notorious tycoons: Robert Maxwell. By Stephen Buranyi". The background of the thumbnail features abstract, colorful shapes resembling a microscopic view of cells or data.

@Calkibet

“When a measure becomes the target, it ceases to be a good measure”

--Goodhart's Law



The Irony...



Purchase PDF

Export ▾



ELSEVIER

Journal of Business Research

Volume 88, July 2018, Pages 428-436



Open Science now: A systematic literature review for an integrated definition

Ruben Vicente-Saez , Clara Martinez-Fuentes

[Show more](#)

<https://doi.org/10.1016/j.jbusres.2017.12.043>

[Get rights and content](#)

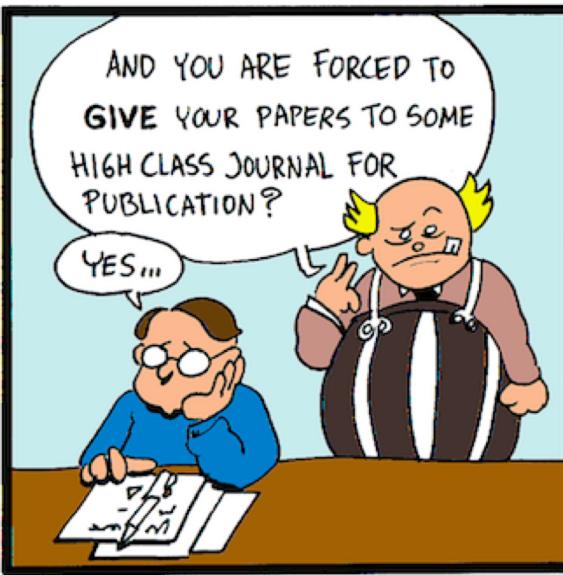
“Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks”



@Calkibet

The Paper is the advertisement...

But...



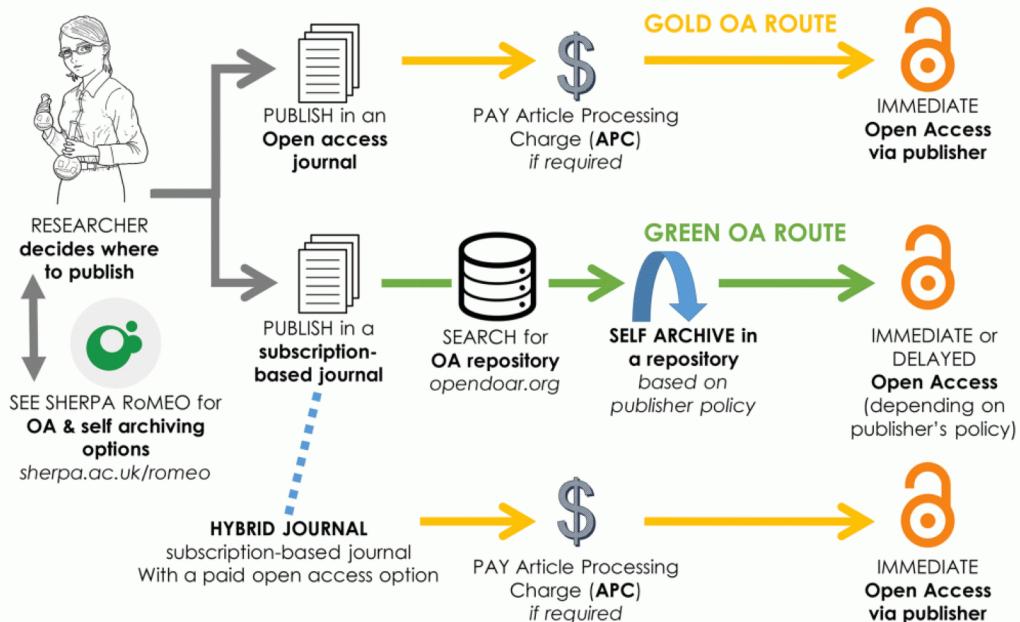
“An **article** about computational results is *advertising*, not scholarship. The actual *scholarship* is the **full software environment, code and data**, that produced the result.”

Buckheit and Donoho (1995)



What are your options?

Open Access Publishing



So, even with closed Journals, you can still be Open:

- Preprint
- Postprint



Some Great Options...

Journals with open access and open review:

- F1000Research
- The AAS Open Access Journal
- Wellcome Open Research
- eLife

F1000Research
Open for Science



AAS Open Research isn't just trying to help African researchers catch up with the West. It's trying to change the game itself.



But avoid Predatory Journals...

Credible journals



Peer review, Revisions, Rejections

Predatory journals



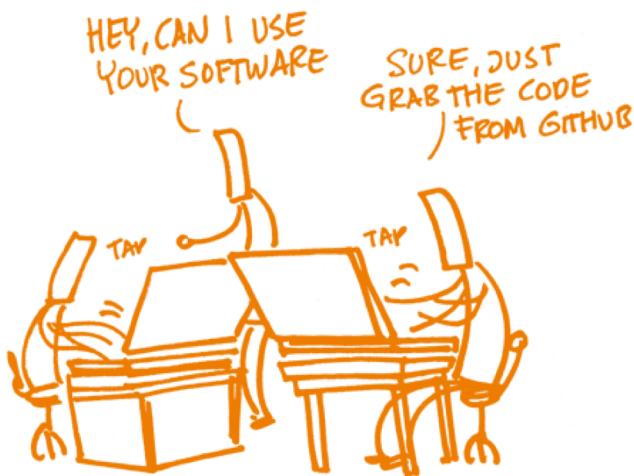
Greetings, We Adore Your Research!

Code is the Scholarship...

OPEN SOURCE

vs

CLOSED SOURCE



Open Science Tools

Git and GitHub

Git is a Version Control System



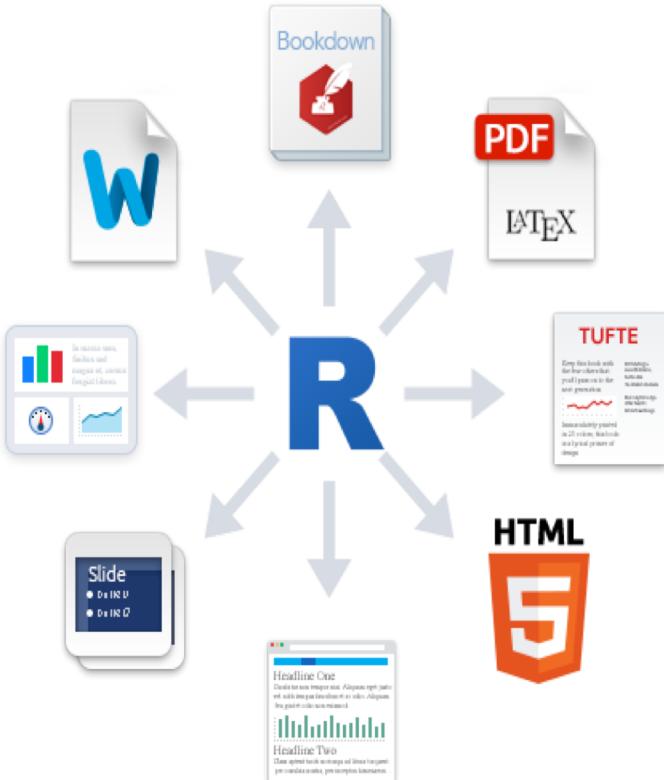
- Helps to keep track of the entire history of things that you are working on.
- Facilitates collaboration on projects

GitHub is a hosting service for Git Repositories

- Web-based service for version control and online collaboration
- The social networking site for developers
- Used to build a portfolio and get noticed by potential recruiters

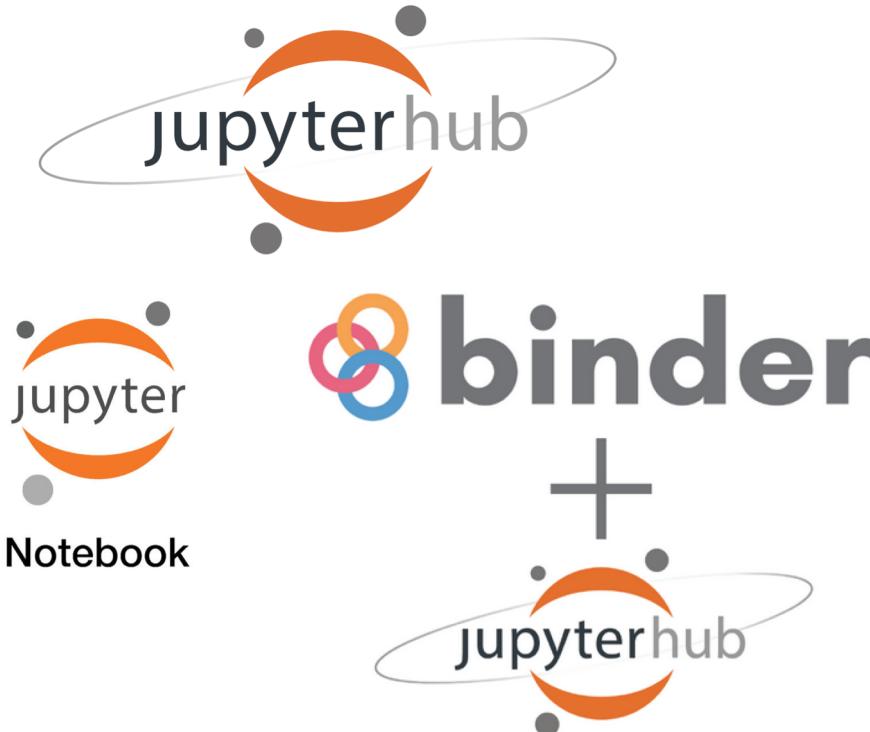


RMarkdown



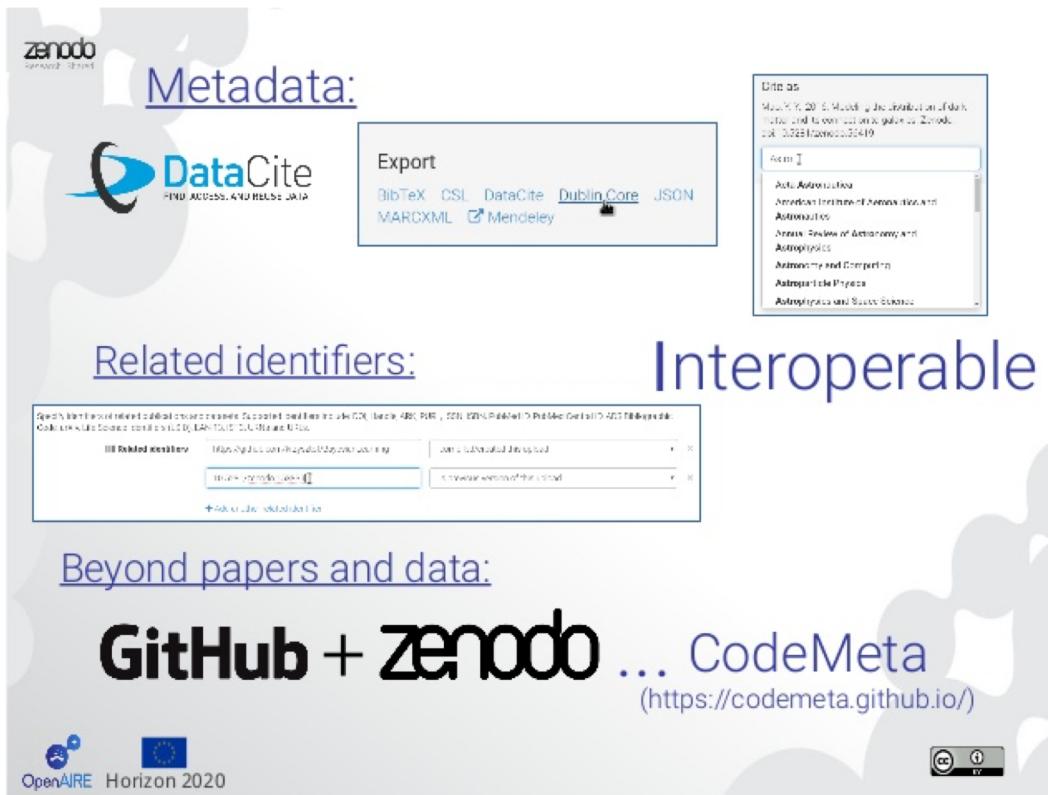
- Provides an authoring framework for data science
- It can be used to save and execute code
- As well as generate high quality and reproducible reports or presentations that can be shared with an audience
- It has built-in support for HTML, PDF, MS_Word, RTF, Github, ODT etc

Literate Programming: Jupyter Notebooks...



- Open web application for creation and sharing of documents with live code, equations, visualization
- Notebook documents are human-readable and can contain analysis description, results as well executable code that can be run to perform data analysis.

Zenodo



Data repositories

- **Zenodo** - is an open access research data repository that provides a place for researchers in any field to deposit datasets up to 50 GB. It has an integration with GitHub to make code hosted on GitHub citable.
- **Figshare** - is an online digital repository where researchers can preserve and share their research output i.e. figures, datasets, images and videos.
- **Dryad** - is a curated a general-purpose repository that makes the data undisclosed in scientific publications discoverable, freely reusable and citable
- **Dataverse** - is an open source web application to share, preserve, cite, explore and analyze research data. Dataverse repository hosts multiple dataverses



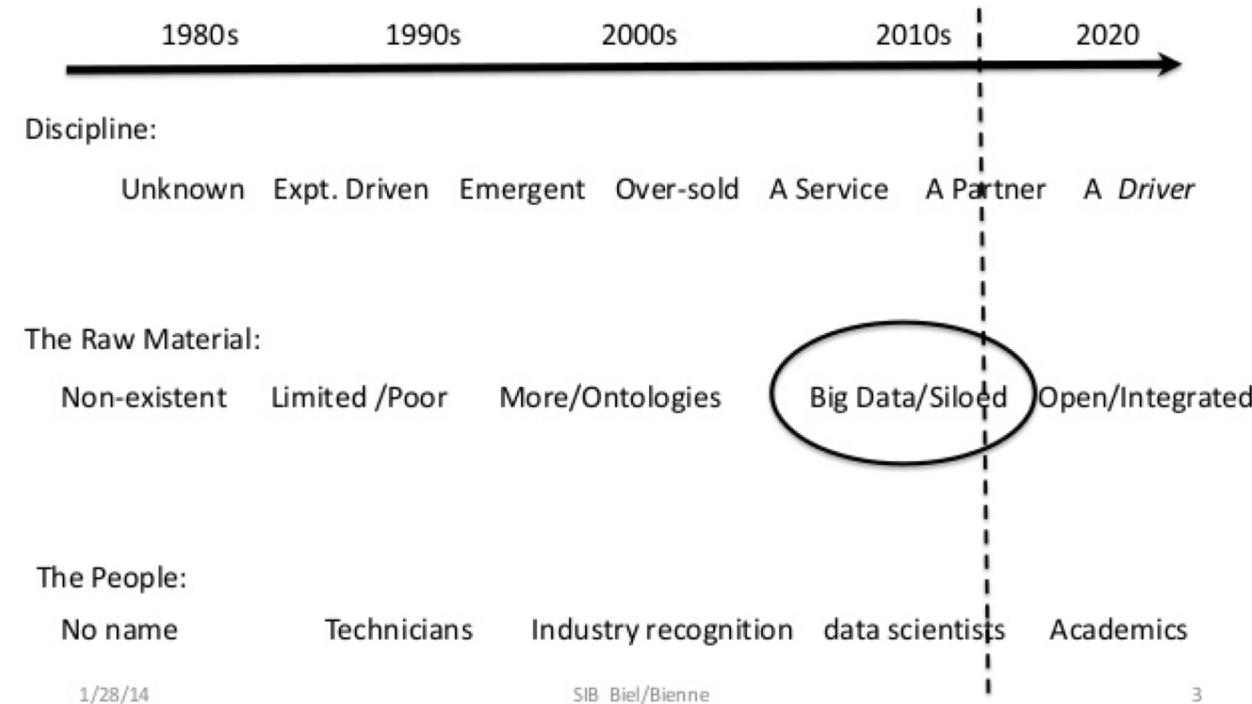
Open Science in Bioinformatics

Bioinformatics is becoming a Data Science,
therefore, Open Science Tools should be adopted.

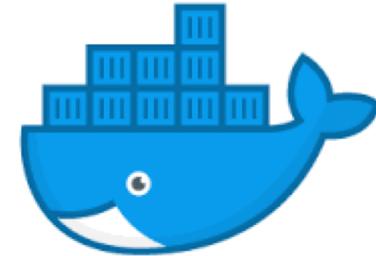


The History of Bioinformatics According to Bourne

Searls (ed) The Roots in Bioinformatics Series *PLOS Comp Biol*



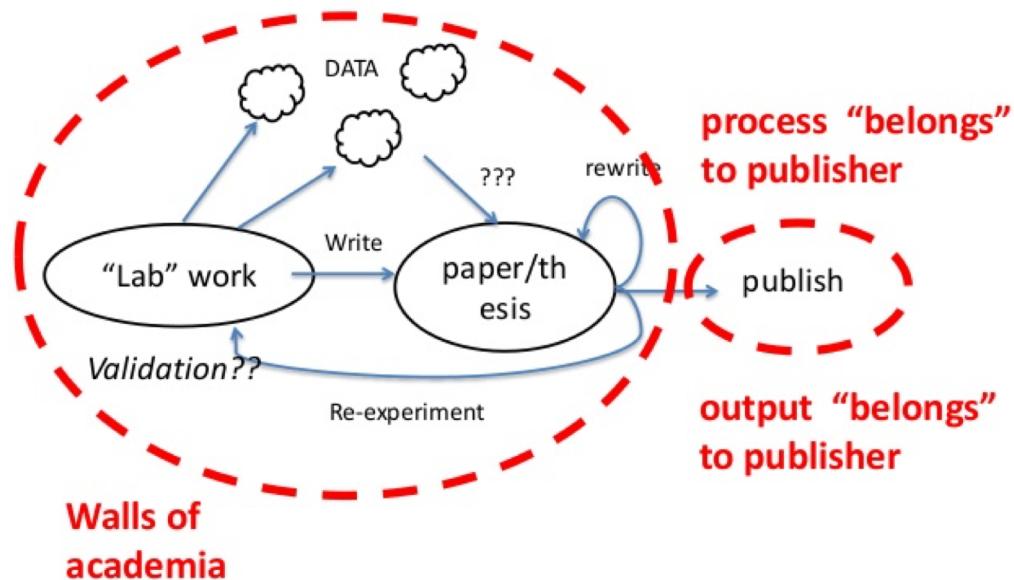
But some specific ones exist...



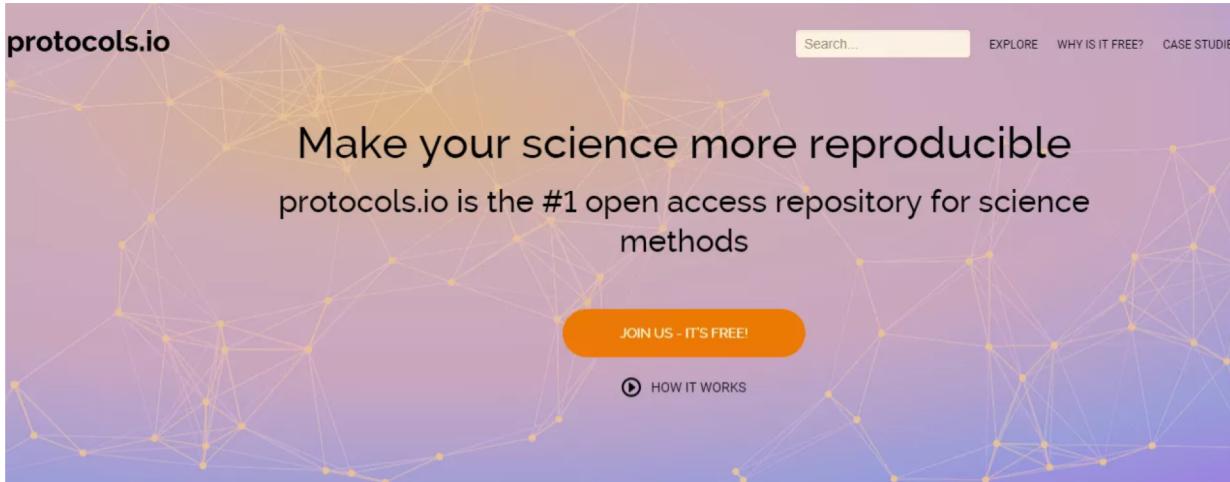
docker



Traditional Research and Publication



There is Protocol.io for experimental biologists



Easily create, edit and get credit for your methods

Protocols.io's editor gives you a structured way to create and modify your protocols in minutes. Make them from scratch or upload and convert your existing Word/PDF documents quickly and easily.



To learn more about Open Science...

- Open Science Handbook: <https://open-science-training-handbook.gitbooks.io/book/content/>
- What, exactly, is open science?: <http://openscience.org/what-exactly-is-open-science/>
- How open science helps researchers succeed:
<https://elifesciences.org/articles/16800#abstract>



My hope is that...

“Future generations [will] look on the term “open science” as a tautology – a throwback from an era before science woke up.

“Open science” will simply become known as science,

and the closed, secretive practices that define our current culture will seem as primitive to them as alchemy is to us.”

-- Brian Nosek and Chris Chambers

