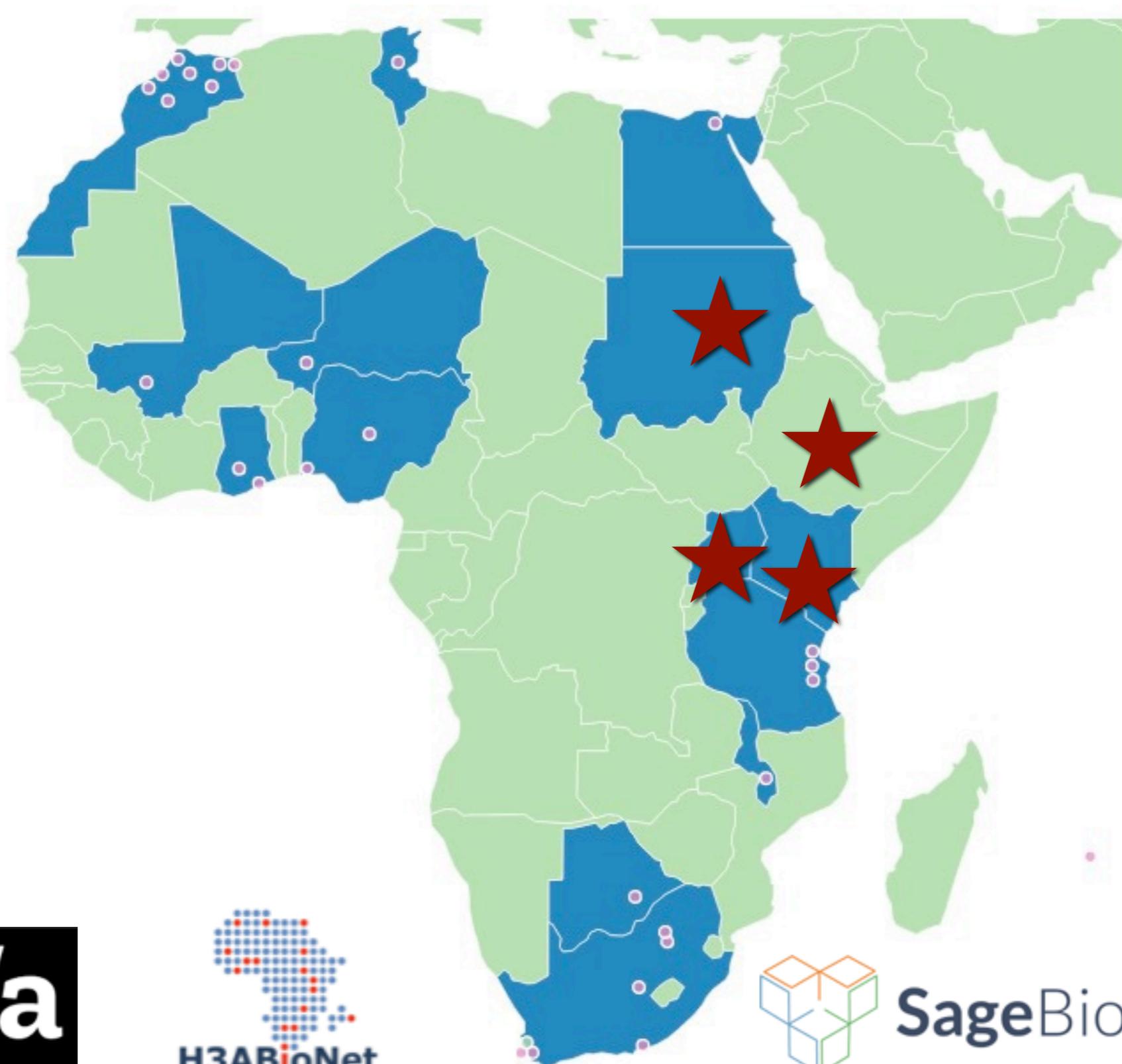


East-Africa Open Science tour

November 12-23, 2018



<https://mozillascience.github.io/EastAfricaWOW/>



Schedule

Notes

Help

Participants

Venues

East Africa Open Science Tour 2018

Uganda, Ethiopia, Sudan and Kenya



mozilla
Science Lab



SageBionetworks



H3ABioNet

Pan African Bioinformatics Network for H3Africa

Mozilla Science group and H3ABioNet in collaboration with SageBionetworks are running a series of 2-days Open Science trainings in East Africa. These 2 days trainings aim to raise awareness, increase skills around open research and reproducibility in Science, tools and platforms, and ultimately help scientist and researchers overcome the fears and challenges.

November 12-13 UVRI: Entebbe, Uganda

November 15-16 AASTU: Addis Ababa, Ethiopia

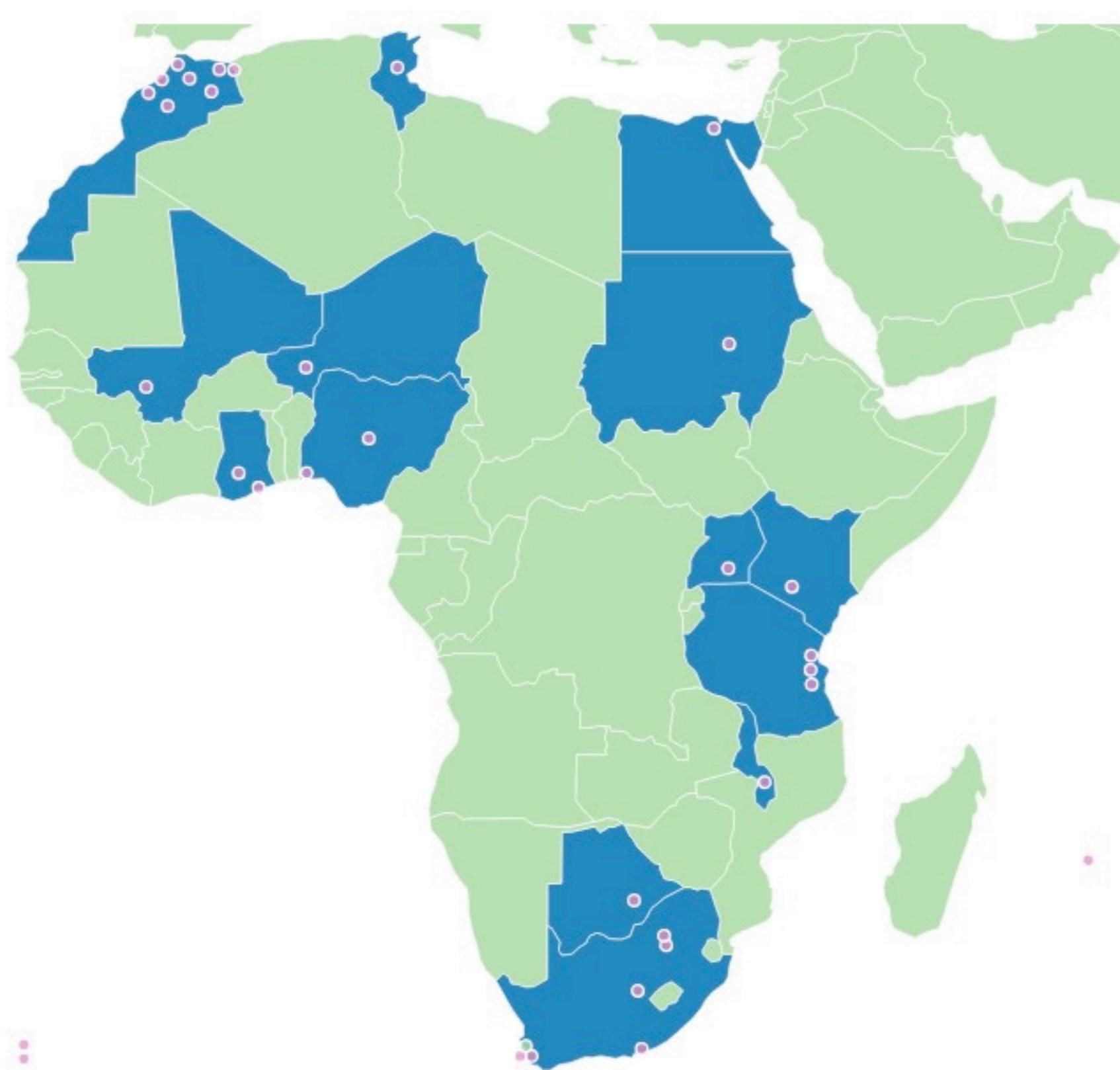
November 18-19 UofK: Khartoum, Sudan

November 22-23 ICIPE: Nairobi, Kenya



H3ABioNet

Pan African Bioinformatics Network for H3Africa



Opening up the research life cycle

Amel Ghouila, PhD
amel.ghouila@pasteur.tn



This presentation is licensed
[Creative commons](#) Attribution 4.0 International



You are free to:

Share: Copy and redistribute the material in any format

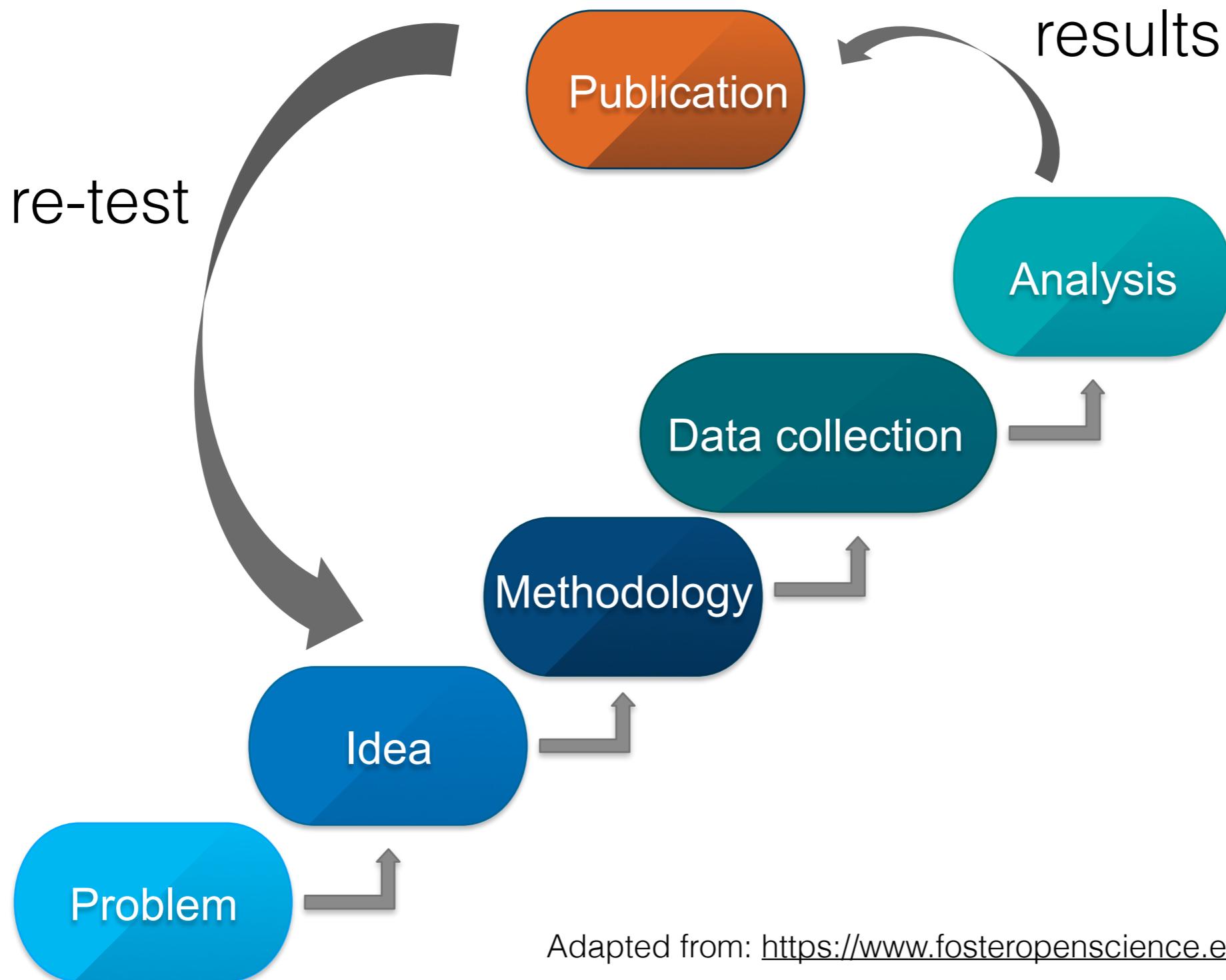
Adapt: remix, transform, and build upon the material for any purpose

Under the following condition:

Attribute: cite this presentation

More info: <https://creativecommons.org/licenses/>

Research life cycle



Adapted from: <https://www.fosteropenscience.eu/node/2323>

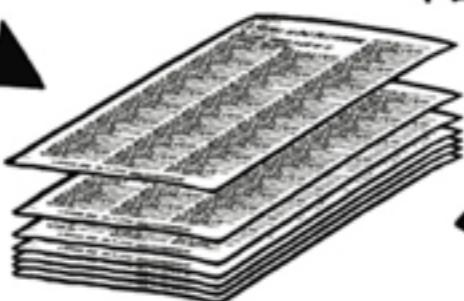
HOW MUCH SCIENCE IS THERE?

SCIENTIFIC PUBLISHING HAS BEEN ACCELERATING—A NEW PAPER IS NOW PUBLISHED ROUGHLY EVERY 20 SECONDS. LET'S IMAGINE A BIBLIOGRAPHY LISTING *EVERY* SCHOLARLY PAPER EVER WRITTEN. HOW LONG WOULD IT BE?

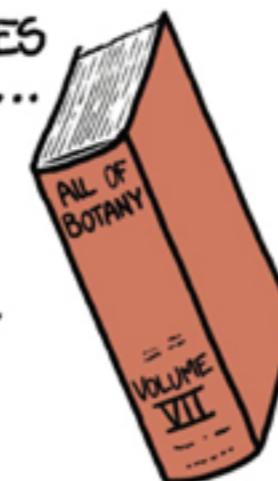
IF WE CAN FIT
140 CITATIONS
PER PAGE...



...1000 PAGES
PER BOOK...



...AND THEN WE START
STACKING BOOKS...



A LIST OF PAPERS
PUBLISHED IN 1880
WOULD FILL
100 PAGES.



BY 1920, THE LIST
WOULD BE GROWING BY
500 PAGES PER YEAR.



THE 1975 SECTION
WOULD FILL FOUR
HUGE VOLUMES.

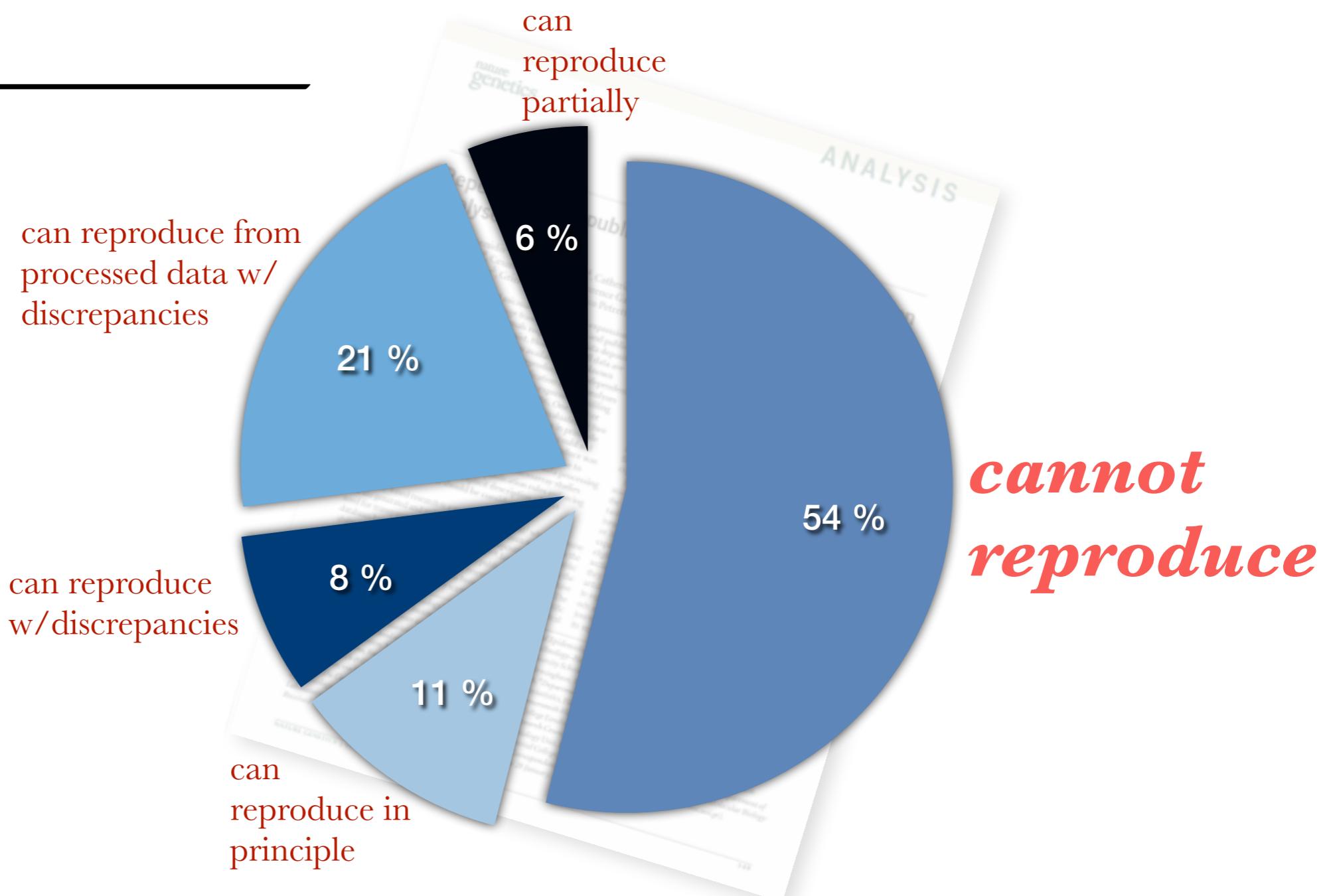


TODAY, WE'RE UP TO 15
VOLUMES PER YEAR—
A PAGE EVERY
45 MINUTES.

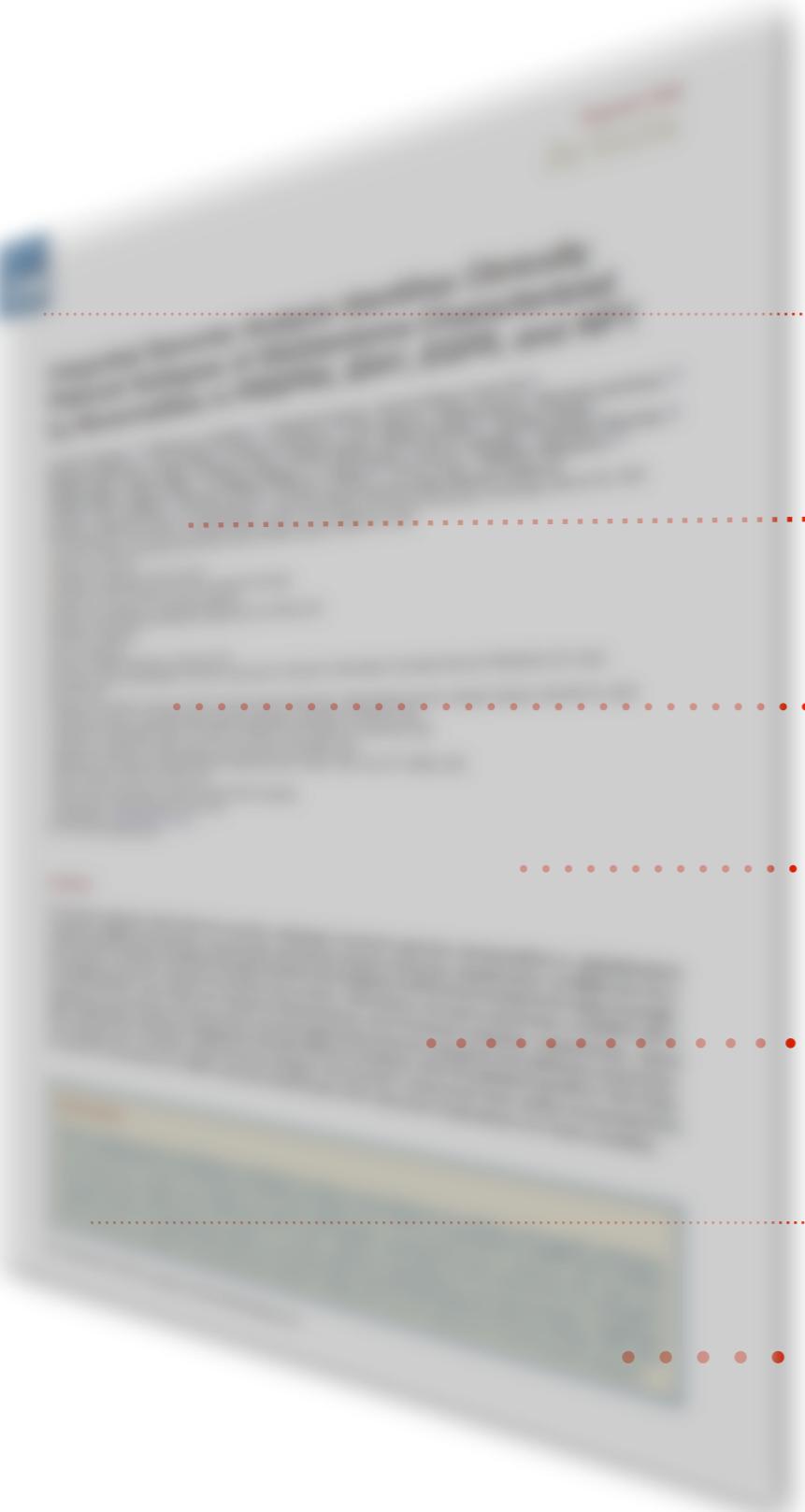


Source: <https://eagereyes.org/blog/2016/publicize-dont-just-publish>

the status quo tolerates poor communication of findings



often what is ***in principle***
reproducible, is not
practically reproducible



208,294,724
datapoints

124 pages
supplemental material

?? lines
unobtainable source code

?? version or architecture of
statistical analysis program (R)

enumerable R packages
and package dependencies

key R package “ClaNC”
no longer available

1231 citations

unidentified publication

- from journal with 5 year impact factor of 27
- article freely available for download
- data freely available for download

Source: <https://fr.slideshare.net/BrianMBot/tools-for-reproducible-research-in-an-increasingly-digital-world>

Reproducible scientific workflows

- Reproducibility implies repetition: requirements to **move back**
- **Retrace** steps, **question** or **change** assumptions and **move forward again.**

Source: <http://rd-alliance.org/>

What is reproducibility?

- **Computation reproducibility:**

- the analysis can be re-ran using your data and code/analysis scripts: we can reproduce your results

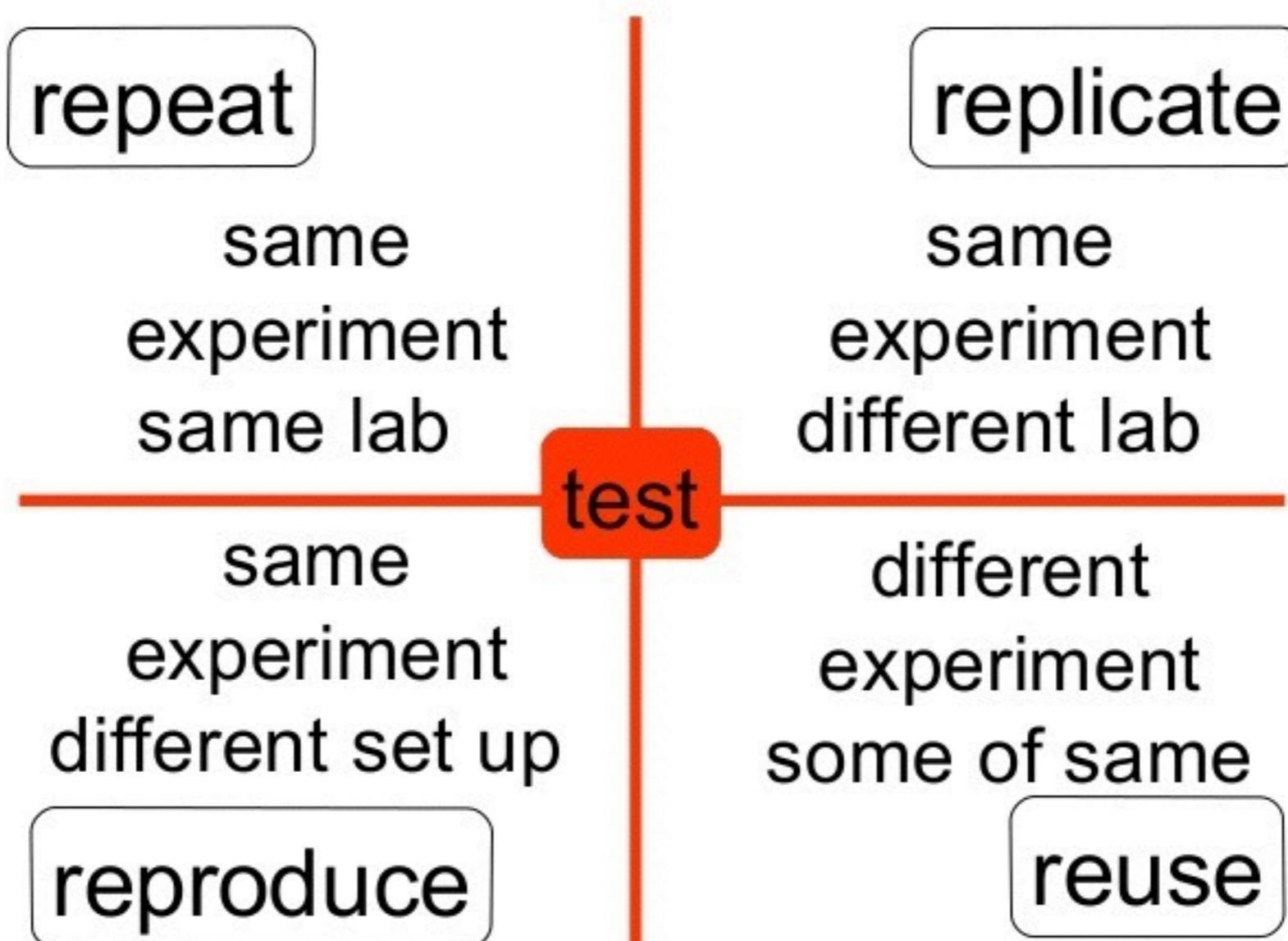
- **Empirical reproducibility:**

- Exact methods and analysis but collect new data and produce the same results
- **Results reproducibility** (replicability): you have enough information to re-run the analysis the way it was originally conducted

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5778115/>

<https://rd-alliance.org/>

Reproducibility, replicability, ...



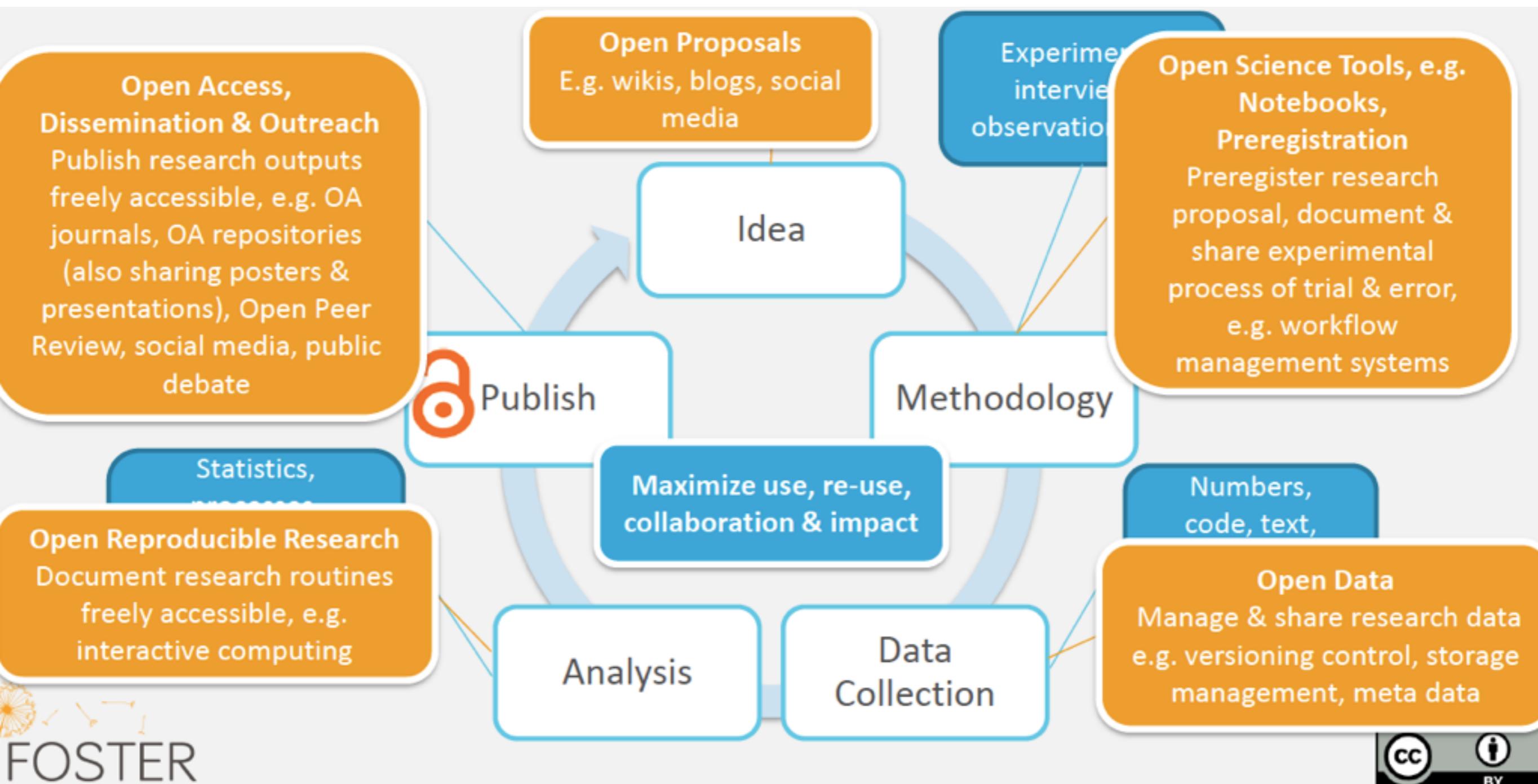
Drummond C Replicability is not Reproducibility: Nor is it Good Science, online
Peng RD, Reproducible Research in Computational Science *Science* 2 Dec 2011: 1226-1227.

Opening up the research life cycle

It just means that you carry out your research in a
more transparent and collaborative way

It applies to all research disciplines

Opening up the research life cycle



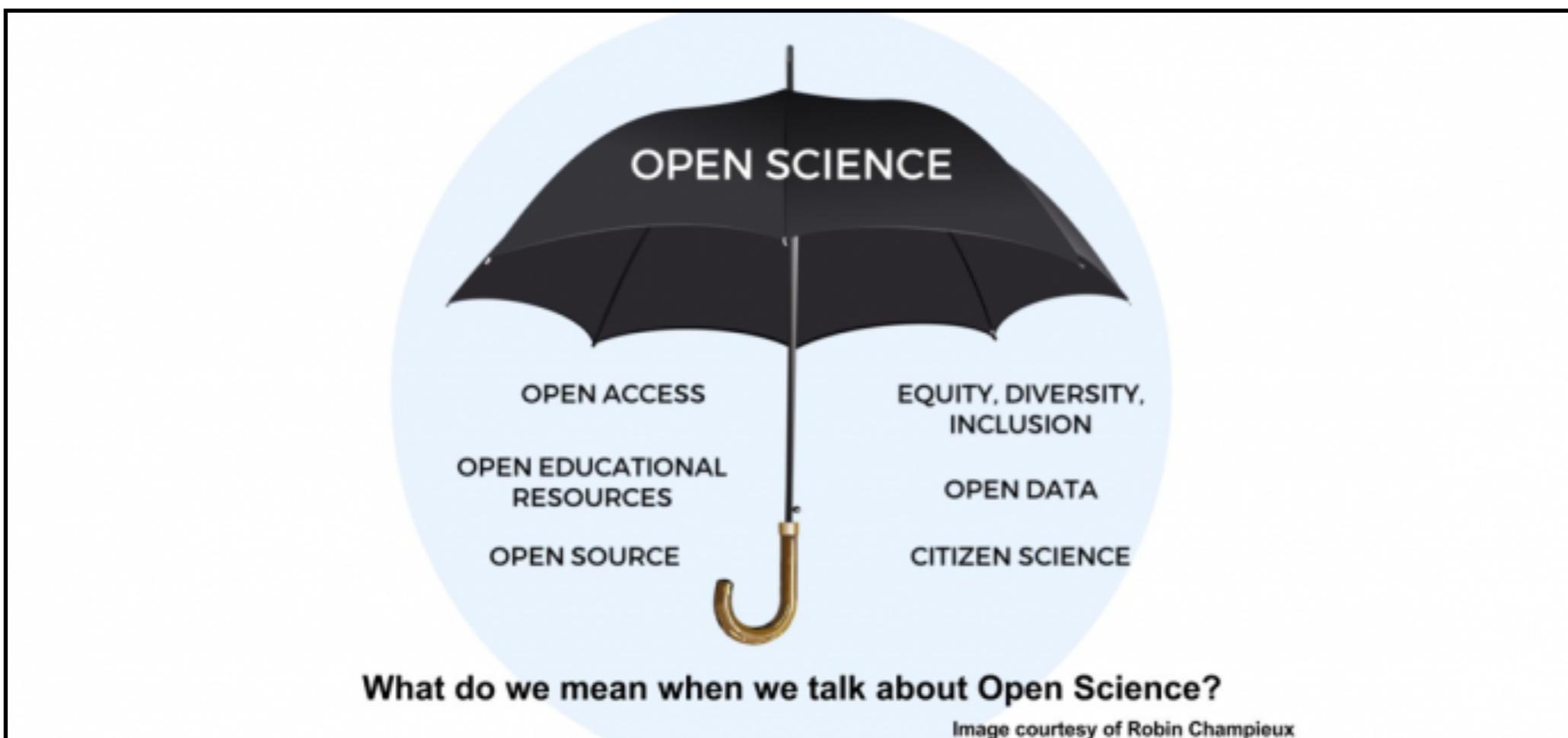
What is Open Science?

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.

Open Science is **transparent** and **accessible** knowledge that is shared and developed through **collaborative** networks (Vicente-Sáez & Martínez-Fuentes 2018).

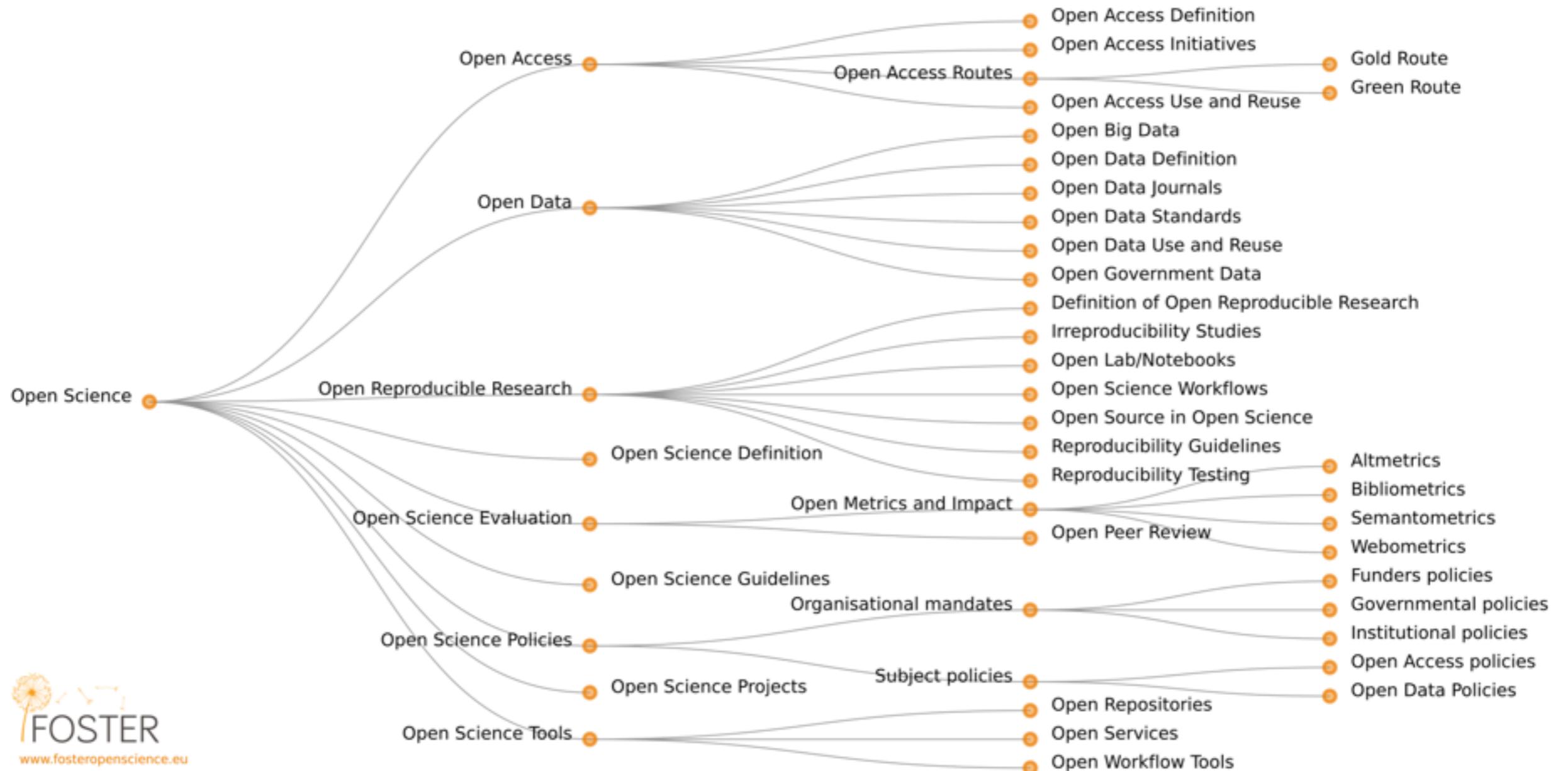
<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>

Open Science pillars



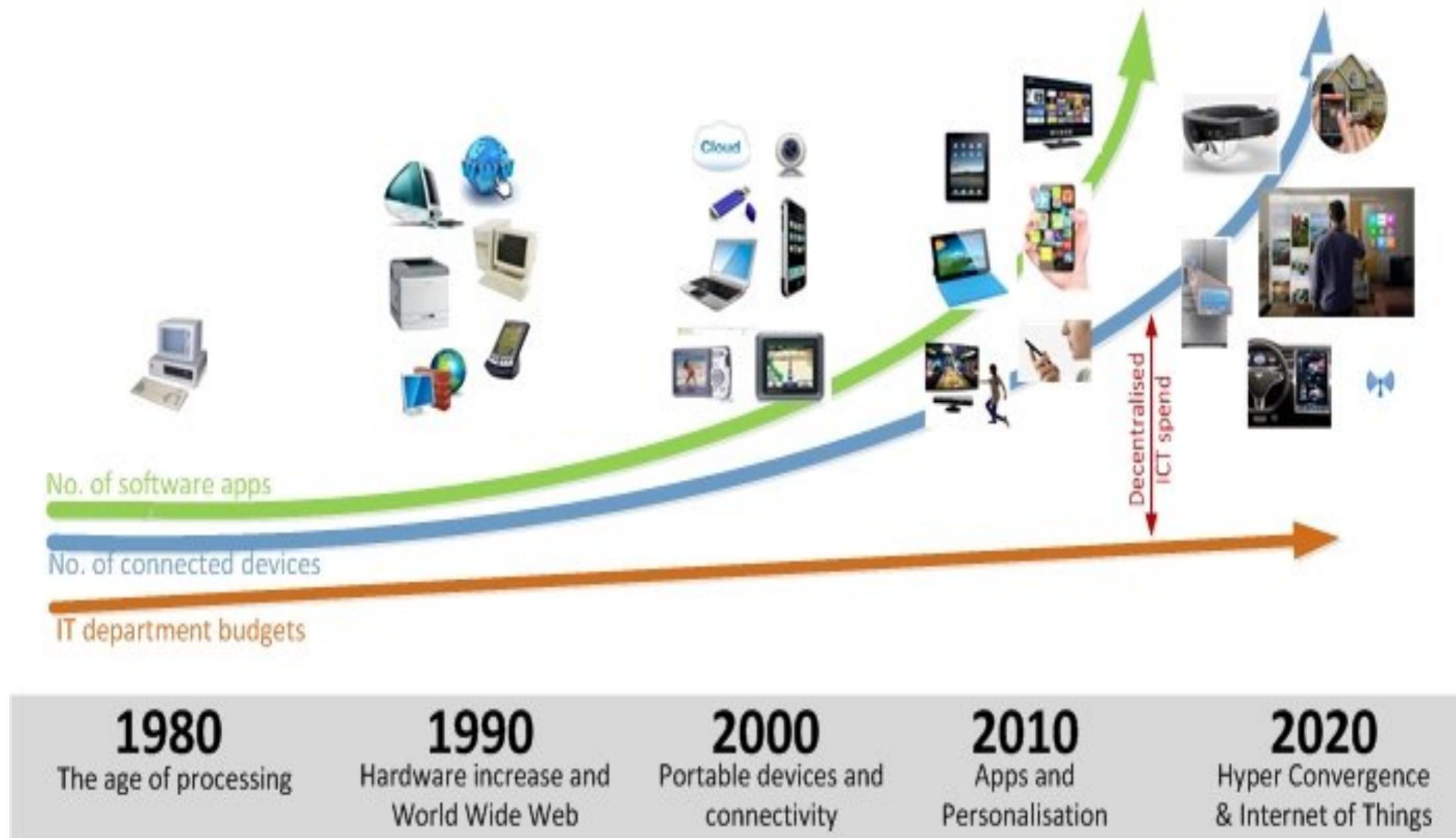
Source: <http://blogs.plos.org/neuro/2018/01/31/open-science-sharing-is-caring-but-is-privacy-theft-by-david-mehler-and-kevin-weiner/>

Open Science Taxonomy



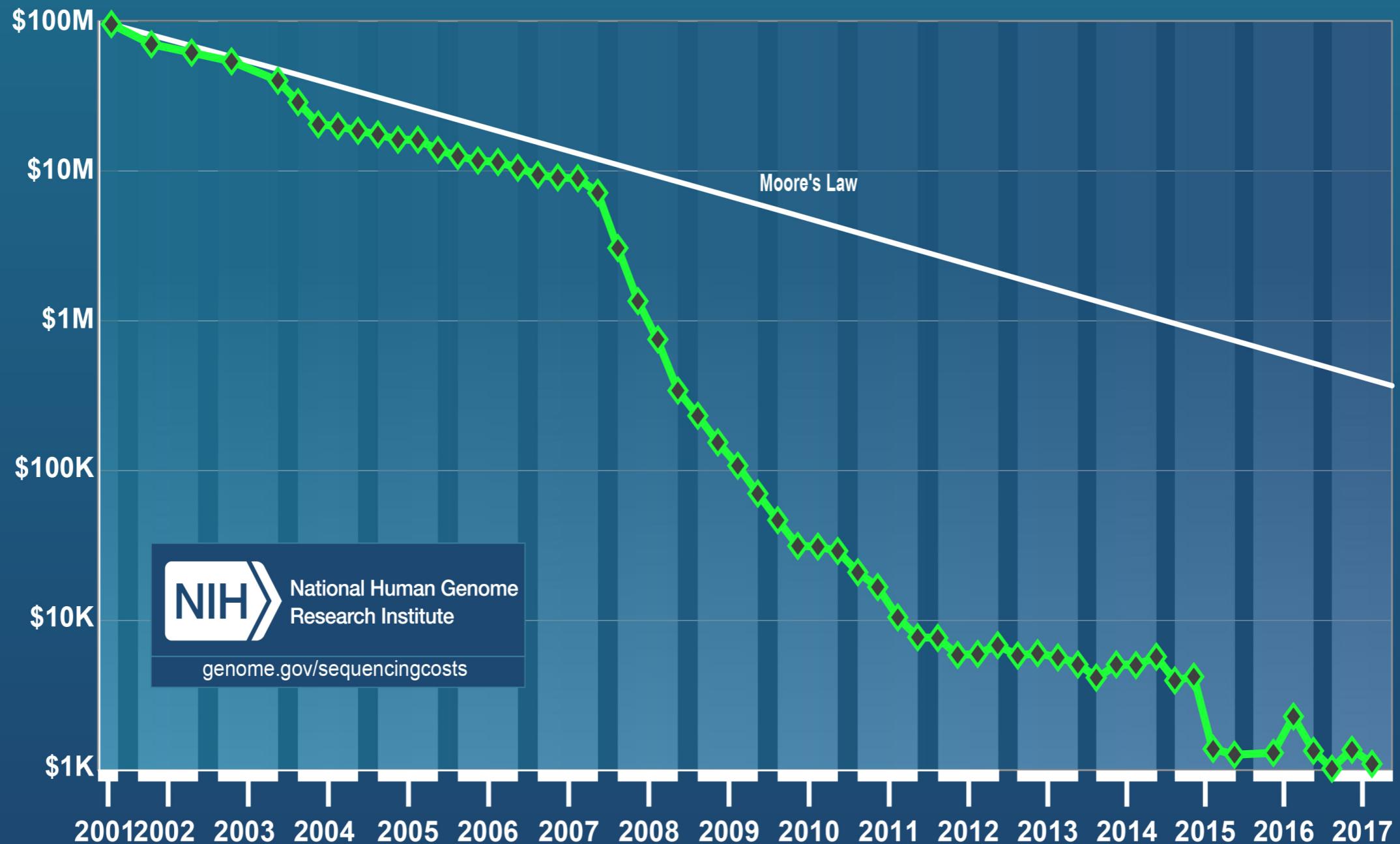
Digital interconnected world

Technology Timeline



<https://www.linkedin.com/pulse/technology-increase-vs-department-budgets-sam-errington/>

Cost per Genome



Open collaborations

- Different sets of skills: diversified ideas, different experiments, etc
- Participation of different entities (research groups, researchers, etc.)
- Fostering inclusivity, participation and social impact

"While one's subjective experience is true, it may not be the totality of the truth"

Brian Bot

Many available tools to facilitate scientific collaborations

Synapse

Synapse

Register Sign in Support



Organize your digital research assets

Create a free Synapse Project to store your research data, code, and results.



Get credit for your research

Mint a DOI for your work - and describe exactly what you did using Synapse provenance.



Collaborate

Share your Project with your collaborators, or make it Public!

 Sign in with Google

Sign in with a Synapse account

Username or Email Address

 **Synapse**
@SageSynapse

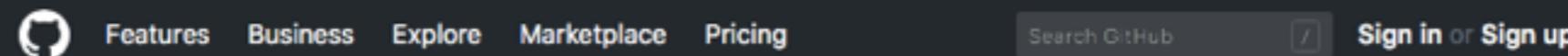
The winners of the @DR_E_A_M Parkinson's Disease Digital Biomarkers Challenge have been announced: businesswire.com/news/home/2018... Congratulations to all the participants! Learn more about the challenge here: synapse.org/DigitalBiomark...

Jan 17, 2018

<https://www.synapse.org>

GitHub



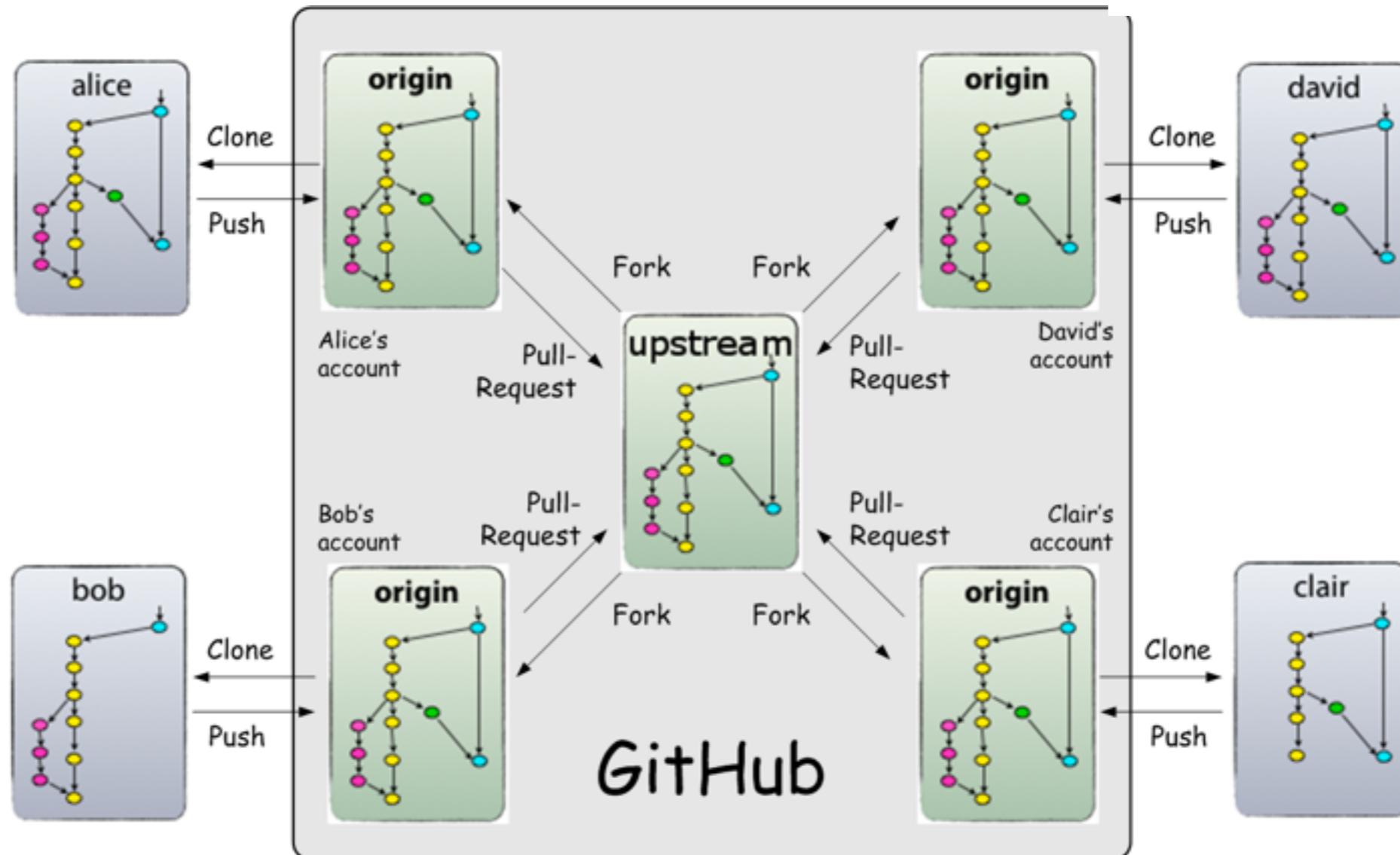
Built for developers

GitHub is a development platform inspired by the way you work. From **open source** to **business**, you can host and review code, manage projects, and build software alongside millions of other developers.

<https://github.com>

open code

distributed



Project example

broadinstitute / gatk

Watch 132 Star 486 Fork 195

Code Issues 916 Pull requests 75 Projects 3 Wiki Insights

Official code repository for GATK versions 4 and up <https://software.broadinstitute.org/gatk>

genomics spark science dna ngs sequencing genome bioinformatics gatk

3,612 commits 447 branches 33 releases 76 contributors BSD-3-Clause

Branch: master New pull request Create new file Upload files Find file Clone or download

Commit	Message	Date
 tomwhite	Make HaplotypeCallerSpark extend AssemblyRegionWalkerSpark (#5386) ...	Latest commit 626c887 4 days ago
 .github	Add slightly modified version of GATK3 github issue template (#4796)	5 months ago
 docs	cited CGA in Mutect docs (#5228)	2 months ago
 gradle(wrapper)	prevent SeekableByteChannelPrefetcher double-wrapping and thread leak(#...)	2 years ago
 hooks	Fixed the git version for the output jar on docker automatic builds (#...)	a year ago
 resources_for_CI	adding a new servicekey.json for travis (#5308)	26 days ago
 scripts	ScatterIntervals produces interval_list instead of intervals (#5392)	5 days ago

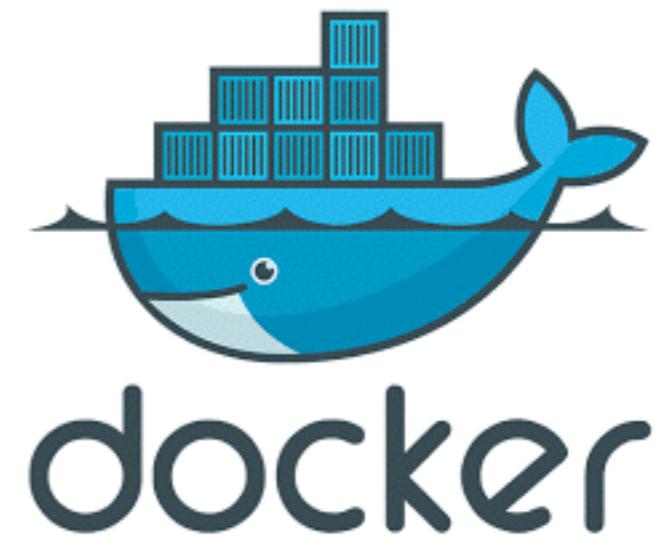
Open code and open source

- Sharing code for transparency
- Sharing under a defined licence allows the community to repeat, replicate, reuse and reproduce analysis
- Code can be used as a starting point for another analysis
- Needs detailed documentation (readme)

Open Source projects



Other platforms



AFRICAN OPEN SCIENCE PLATFORM

HOME ABOUT ▾ RESEARCH, SCIENCE (INCL. DATA) EVENTS ACKNOWLEDGEMENTS CONTACT

African Open Science Platform: a vision of Agenda 2063

Source: October 2018 – AlphaGalileo is the world's independent business to business service of breaking research news for the media

Agenda 2063 envisions an African continent that is peaceful, prosperous and integrated. The emerging view is that the continent would have developed elaborate intra- and inter-information systems and processes to exploit the digital revolution for the attainment of



Search ...

BigBlueButton

Meetings

Open Data

Data refers to genomes, numbers, but also geospatial coordinates, text, images, multimedia items, and other types of information that can be used to answer questions or solve problems.

Different types of data are being collected by researchers and scientists

All of this data is potentially useful and powerful

Open data

- Open data is data that is made freely and easily available to anyone to use, reuse and distribute.

“Opening” data means maximizing that potential

Open data repositories for Biomed (EGA, GEO, RSA, etc.)

- Process of sharing biomedical data not comprehensively documented

Data Management life cycle

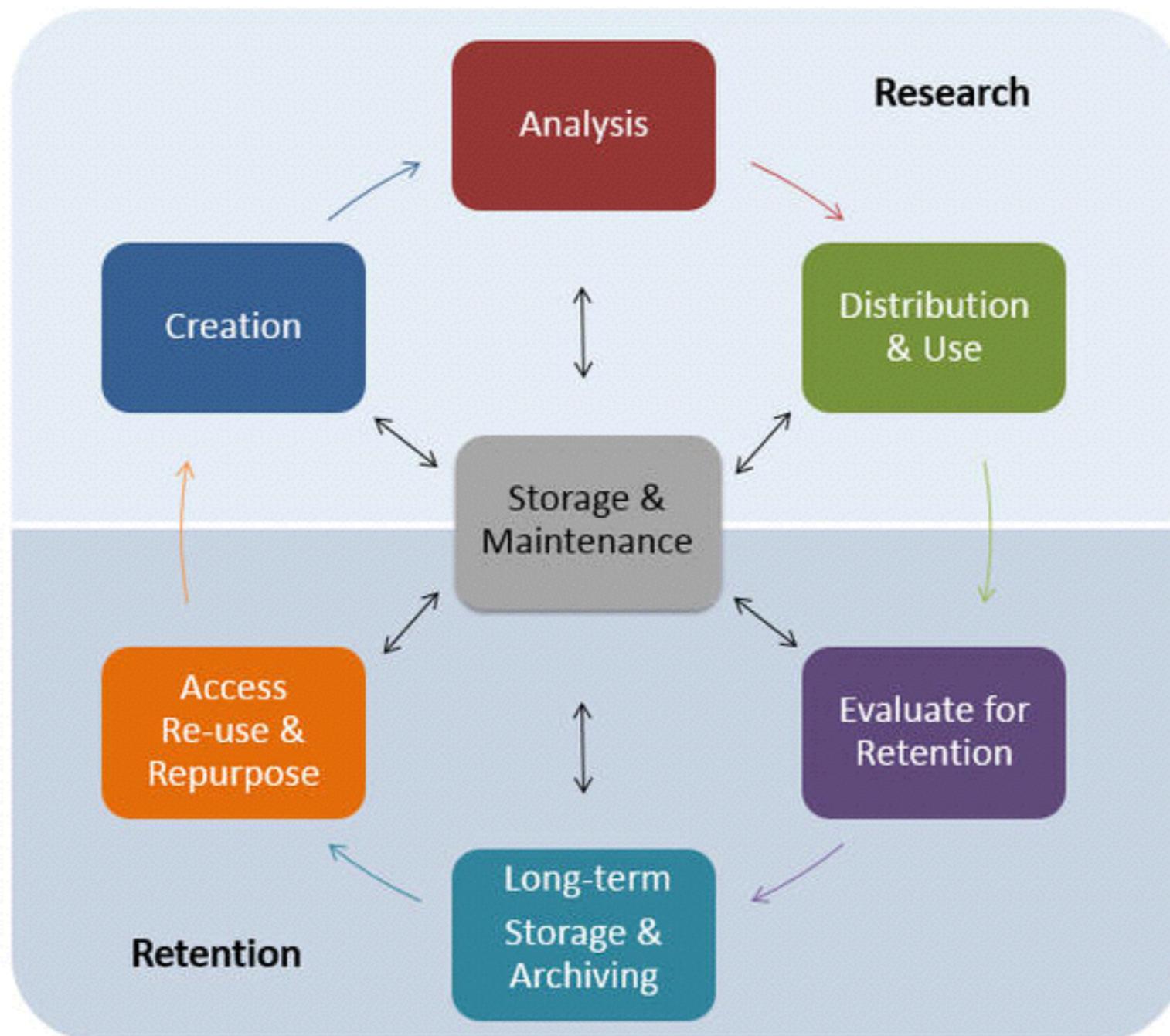


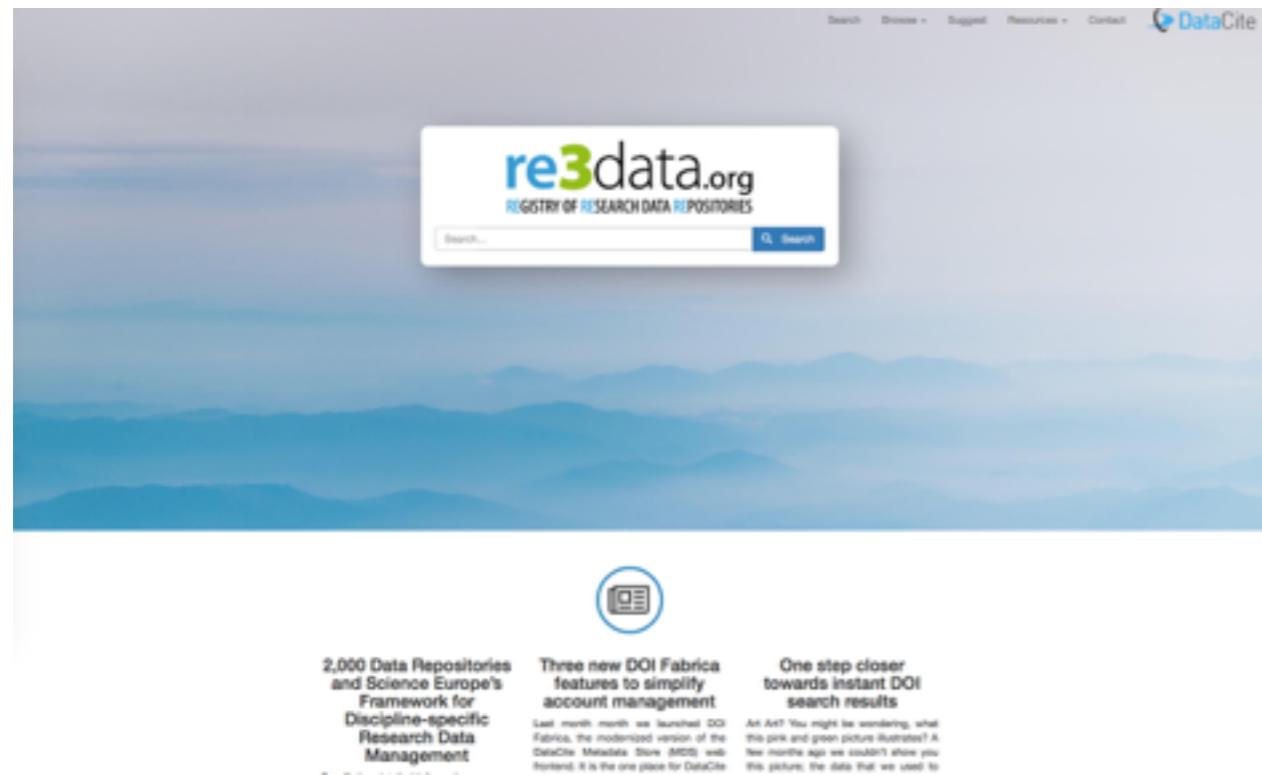
image acquired from the [Harvard Biomedical Data Management Website](#)

Open Data

The Open Knowledge Foundation defines the following key factors that make data “open”:

- **Access & availability** - data is available to all in a convenient and modifiable form
- **Re-use & redistribution** - terms of use allow for reusing, remixing and redistributing the data
- **Universal participation** - there are no restrictions on who may do any of the above with the data

Re3Data



Re3Data is a global catalog of online data repositories. You can browse by subject area, country or the type of content (e.g. images, code, audiovisual, etc).

Has information about access restrictions, permanent identifiers as well as links to policies for those repositories

<https://www.re3data.org>

FAIR Data

FAIR research principles advocate for **Findable, Accessible, Interoperable, and Reusable** data.

- data should be curated with indexed metadata which are searchable
- metadata should be retrievable using the stable identifiers
- should be described with standard vocabularies as well as using stable, recognised identifiers
- metadata must adhere to community standards, and include provenance and data usage information

FAIR data doesn't necessarily mean Open

Open access



Source: <http://openaccess.cz/en/>

Format: Abstract ▾

Genome Res. 2016 Feb;26(2):271-7. doi: 10.1101/gr.196295.115. Epub 2015 Dec 1.

H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa.

Mulder NJ¹, Adebiyi E², Alami R³, Benkahla A⁴, Brandful J⁵, Doumbia S⁶, Everett D⁷, Fadlelmola FM⁸, Gaboun F⁹, Gaseitsiwe S¹⁰, Ghazal H¹¹, Hazelhurst S¹², Hide W¹³, Ibrahimi A¹⁴, Jaufeerally Fakim Y¹⁵, Jongeneel CV¹⁶, Joubert F¹⁷, Kassim S¹⁸, Kayondo J¹⁹, Kumuthini J²⁰, Lyantagaye S²¹, Makani J²², Mansour Alzohairy A²³, Masiga D²⁴, Moussa A²⁵, Nash O²⁶, Ouwe Missi Oukem-Boyer O²⁷, Owusu-Dabo E²⁸, Panji S¹, Patterson H²⁹, Radouani F³⁰, Sadki K³¹, Seghrouchni F³², Tastan Bishop Ö³³, Tiffin N³⁴, Ulenaga N³⁵; H3ABioNet Consortium.

Collaborators (53)

Author information

Abstract

The application of genomics technologies to medicine and biomedical research is increasing in popularity, made possible by new high-throughput genotyping and sequencing technologies and improved data analysis capabilities. Some of the greatest genetic diversity among humans, animals, plants, and microbiota occurs in Africa, yet genomic research outputs from the continent are limited. The Human Heredity and Health in Africa (H3Africa) initiative was established to drive the development of genomic research for human health in Africa, and through recognition of the critical role of bioinformatics in this process, spurred the establishment of H3ABioNet, a pan-African bioinformatics network for H3Africa. The limitations in bioinformatics capacity on the continent have been a major contributory factor to the lack of notable outputs in high-throughput biology research. Although pockets of high-quality bioinformatics teams have existed previously, the majority of research institutions lack experienced faculty who can train and supervise bioinformatics students. H3ABioNet aims to address this dire need, specifically in the area of human genetics and genomics, but knock-on effects are ensuring this extends to other areas of bioinformatics. Here, we describe the emergence of genomics research and the development of bioinformatics in Africa through H3ABioNet.

PMID: 26627985 PMCID: PMC4728379 DOI: 10.1101/gr.196295.115

[Indexed for MEDLINE]

Free PMC Article

Send to ▾

Full text links



Save Items

Add to Favorites ▾

Similar articles

Review Development of Bioinformatics Infrastructure for Genomics Research [Glob Heart. 2017]

Assessing computational genomics skills: Our experience in the H3AF [PLoS Comput Biol. 2017]

Developing expertise in bioinformatics for biomedical research in Africa [Appl Transl Genom. 2015]

Designing a course model for distance-based online bioinformatics training [PLoS Comput Biol. 2017]

Review H3Africa and the African life sciences ecosystem: building sustainable infrastructure [OMICS. 2014]

See reviews...

See all...

Cited by 20 PubMed Central articles ▾

Benefits of Open Access for you and society



Source: <https://blogs.bournemouth.ac.uk/research/researcher-toolbox/research-outputs/bu-open-access-mini-guide/>

Types of Open Access

	Green	Gold
Timing 	Often deferred open access, in accordance with publisher embargo periods (normally 12 to 24 months)	Immediate open access with articles made freely available at the point of publication
Funder/HEFCE requirements	From 1 April 2016, Green OA will be required to ensure articles are returnable in a future REF	May be required by funders, especially where publisher does not allow self-archiving
Licensing 	Offers more flexibility in license options, may allow author to limit re-use of work, particularly for commercial purposes	Often published under a license which minimises barriers to re-use and dissemination, provided author's work is credited
Cost 	No additional cost to author	Usually requires payment of article processing charge (APC).

Source: <https://blogs.bournemouth.ac.uk/research/researcher-toolbox/research-outputs/bu-open-access-mini-guide/>

Preprints

Peer-reviewed publications take weeks, months and even years from the initial submission to be published

A preprint is a draft that has not yet been peer reviewed for "formal publication"

Make your research findings publicly available immediately and get feedback

Preprints

why preprint publications?:

- It's free and fast - published within 24 hours.
- Your preprint receives a permanent, archival-standard link (DOI).
- Establish precedent for your research.
- Publish before you submit for peer-review, or at the same time.
- Readers can leave comments and ask questions.

Source: <https://peerj.com/about/preprints/what-is-a-preprint/>

Preprint servers examples:

- <https://www.preprints.org>
- <https://peerj.com/about/preprints/what-is-a-preprint/>
- <https://www.biorxiv.org>



A free preprint service for African scientists

[Powered by OSF Preprints](#)

or

[See an example](#)

Browse by subjects

[Architecture](#)[Arts and Humanities](#)[Business](#)[Education](#)[Engineering](#)[Law](#)[Life Sciences](#)[Medicine and Health Sciences](#)[Physical Sciences and Mathematics](#)[Social and Behavioral Sciences](#)

When and How to Comply

1 Preparing a manuscript

Address copyright

[show me](#)**2** Accepted for publication

Post it to PubMed Central and track it in My NCBI

[show me](#)**3** Reporting to NIH

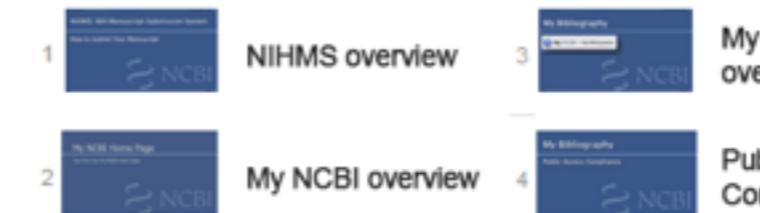
Include PMCID in citations

[show me](#)

Overview:

To advance science and improve human health, NIH makes the peer-reviewed articles it funds publicly available on [PubMed Central](#). The NIH public access policy requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to PubMed Central immediately upon acceptance for publication. [\[more\]](#)

[Show me specific instructions for my publication](#)

Public Access Policy
Video Training

When and How to Comply

- 1** Preparing a manuscript Address copyright
[show me](#)
- 2** Accepted for publication Post it to PubMed Central and track it in My NCBI
[show me](#)
- 3** Reporting to NIH Include PMCID in citations
[show me](#)

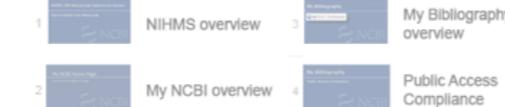
Overview:

To advance science and improve human health, NIH makes the peer-reviewed articles it funds publicly available on [PubMed Central](#). The NIH public access policy requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to PubMed Central immediately upon acceptance for publication. [\[more\]](#)

[Show me specific instructions for my publication](#)



Public Access Policy Video Training



nature

International weekly journal of science

[Home](#) | [News & Comment](#) | [Research](#) | [Careers & Jobs](#) | [Current Issue](#) | [Archive](#) | [Audio & Video](#) |

[Archive](#) > [Volume 514](#) > [Issue 7524](#) > [Editorial](#) > [Article](#)

NATURE | EDITORIAL

عربي



Code share

Papers in Nature journals should make computer code accessible where possible.

29 October 2014



NIH GDS Policies

NIH Genomic Data Sharing Policy

Genomic research advances our understanding of factors that influence health and disease, and sharing genomic data provides opportunities to accelerate that research through the power of combining large and information-rich datasets. To promote robust sharing of human and non-human data from a wide range of genomic research and to provide appropriate protections for research involving human data, the National Institutes of Health (NIH) issued the *NIH Genomic Data Sharing Policy* (GDS Policy) on August 27, 2014 in the *NIH Guide Grants and Contracts* (available at <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html>), and in the *Federal Register* (available at <https://federalregister.gov/a/2014-20385>) on August 28, 2014. The GDS Policy and related documents are available at:

- [Preamble to the NIH GDS Policy](#)
- [NIH GDS Policy PDF](#)
- [Supplemental Information to the NIH GDS Policy](#)
- [NIH Intramural Investigator Responsibilities under the NIH GDS Policy](#)
- [NIH Guide Notice on Implementation of the GDS Policy for NIH Grant Applications and Awards](#)
- [NIH Guide Notice on Development of Data Sharing Policy for Sequence and Related Genomic Data](#)
- [GDS Policy Overview](#)

Scientific Data Sharing

- > [Genomics and Health](#)
- > [Scientific Data Management](#)

[Scheme finder](#)[Funding guidance](#)[Develop your research career](#)

Policy on data, software and materials management and sharing

As a charity, Wellcome works to ensure that the results of the research we fund are applied for the public good. This includes creating an environment that enables and incentivises researchers to maximise the value of their research outputs, including data, software and materials.

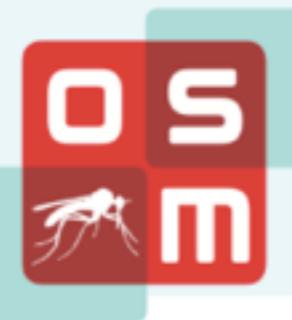
We expect our researchers to manage research outputs in a way that will achieve the greatest health benefit. This may involve making outputs widely available or using intellectual property (IP) as a tool to help protect and commercialise an original idea, product or technology.

There is international consensus on the need to share and preserve research datasets in a way that maximises their long-term value. Key documents such as the [UK concordat on open research data \(2016\)](#) articulate this.

Making data available in a timely and responsible way ensures other research can verify it, build on it and use it to advance knowledge and make health improvements. Similarly, making software or materials – such as antibodies or cell lines – available to the research community supports reproducibility and can underpin further research.

Funding guidance

[Applying for funding](#) >[Starting a grant](#) >[Grants finance and administration](#) >[Research practice and policies](#) >[Ending a grant](#) >[Research outputs](#) >[Funding guidance overview](#) >



OPEN SOURCE MALARIA

Looking for New Medicines

Also Check Out



[Lab Notebook](#)



[Project Wiki](#)



[Molecule Database](#)
Chrome Only



Catch us on the Daily Show?

If you're here to see more about the **Breaking Good** project, as featured on the Daily Show

[Click here!](#)

The Open Source Malaria project is trying a different approach to curing malaria. Guided by open source principles, everything is open and anyone can contribute.

[Read More](#)

Recent Contributors

MFernflower

32

mattodd

26



Activity



Join the Team



Meet Us

[View All](#)

[Collaboration Request](#)

[Series 3](#)

[Series 4](#)

[PR](#)

[Writing](#)

[question](#)

[Non-Science](#)

[Being Synthesised Now](#)

[Synthetic Chemistry Needed](#)

<http://opensourcемalaria.org>

