

Bioinformatics in a tense geopolitical context

ANF Workflow et reproductibilité en Bioinformatique

Claire Toffano-Nioche, Thomas Denecker, Laurent Jourdren

Context



Le monde - 14 avril 2025

Trump made science news in 2025

- Massive budget cuts (total impact >2 billion dollars)
- Interruption of funding for targeted topics
- Censorship of new research projects
- Closure of certain data resources
- Loss of access to data and their operating platforms
- Cessation of data collection (climatology, clinical, gender, etc.)
- Suspension of research contracts on these topics
- Loss of scientific capacity (layoffs, non-renewal of contracts)
- Suspension of funding allocated to non-USA data resources and research projects
- Impact on international collaborations (particularly with Europe).
- ...

Scientific topics

- Biology and public health: microbiology, epidemiology, vaccination
- Climate change and the environment: climate data, weather, renewable energy
- Social sciences: gender studies, sexual orientation, diversity, inclusion

Main USA institutions targeted

- National Institutes of Health (NIH)
- Center for Disease Control and Prevention (CDC)
- National Science Foundation (NSF)
- United States Department of Agriculture (USDA)
- National Oceanic and Atmospheric Administration (NOAA)
- Environmental Protection Agency (EPA)
- Universities

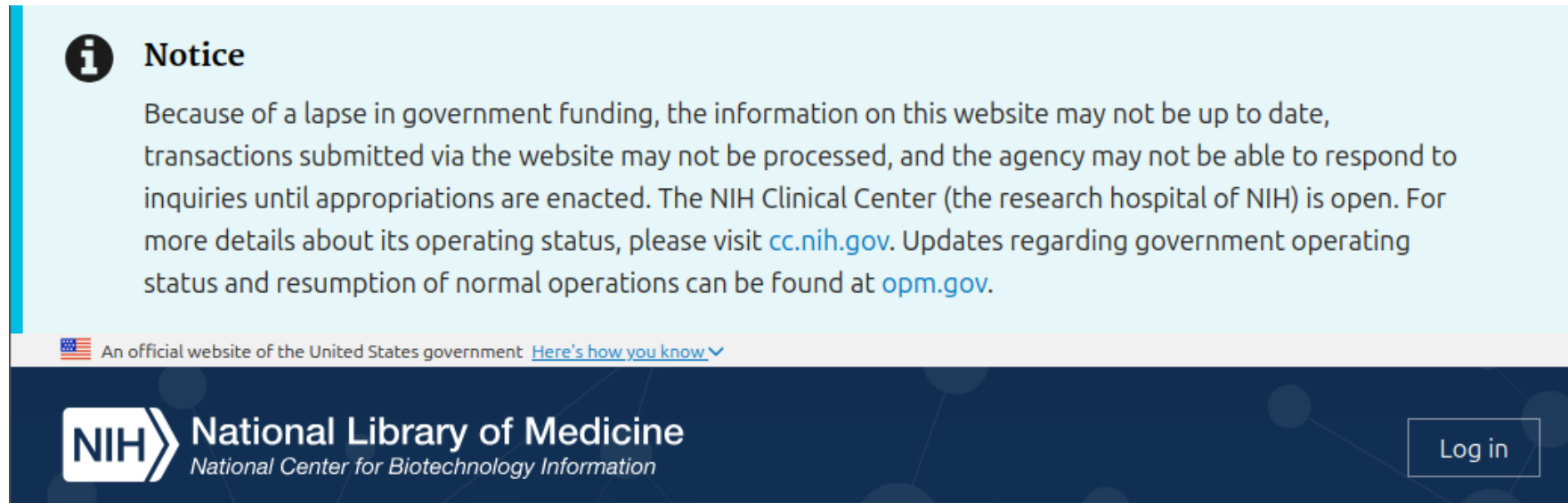
Withdrawal of major global organizations

- World Health Organization (WHO)
- Unesco, ...

Main biology and health data platform: **NCBI** (National Center for Biotechnology Information)

- 29 databases specializing in different areas of biology and health (ncbi.nlm.nih.gov).
- +14 tools for data analysis and visualization (BLAST, etc.).

NIH ex.:
BLAST,
PubMed,
Taxonomy, ...
(Oct. 2025)



The screenshot shows a notice on the NIH website. At the top, there is a light blue banner with a black circle containing a white 'i' icon, followed by the word 'Notice' in bold. The text below reads: 'Because of a lapse in government funding, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries until appropriations are enacted. The NIH Clinical Center (the research hospital of NIH) is open. For more details about its operating status, please visit cc.nih.gov. Updates regarding government operating status and resumption of normal operations can be found at opm.gov.' Below the notice is a grey bar with the text 'An official website of the United States government' and a link 'Here's how you know' with a dropdown arrow. At the bottom, there is a dark blue footer with the NIH logo (a white arrow pointing right with 'NIH' inside), the text 'National Library of Medicine' and 'National Center for Biotechnology Information'. On the right side of the footer is a 'Log in' button.

i Notice

Because of a lapse in government funding, the information on this website may not be up to date, transactions submitted via the website may not be processed, and the agency may not be able to respond to inquiries until appropriations are enacted. The NIH Clinical Center (the research hospital of NIH) is open. For more details about its operating status, please visit cc.nih.gov. Updates regarding government operating status and resumption of normal operations can be found at opm.gov.

 An official website of the United States government [Here's how you know](#) ▼

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

Talk objective

Private companies and/or the USA Patriot/Cloud Act (or other foreign countries)

Why ?

⇒ The applicable legislation depends on the location of the companies and/or the physical machines supporting them ...

USA law example

- **USA Patriot Act** (2001): anti-terrorism law, authorizes surveillance without a warrant
- **USA Cloud Act** (2018): allows USA authorities (FBI, NSA, ...) to access data stored by USA companies, even if that data is hosted outside the USA (cf. [CNLL](#))

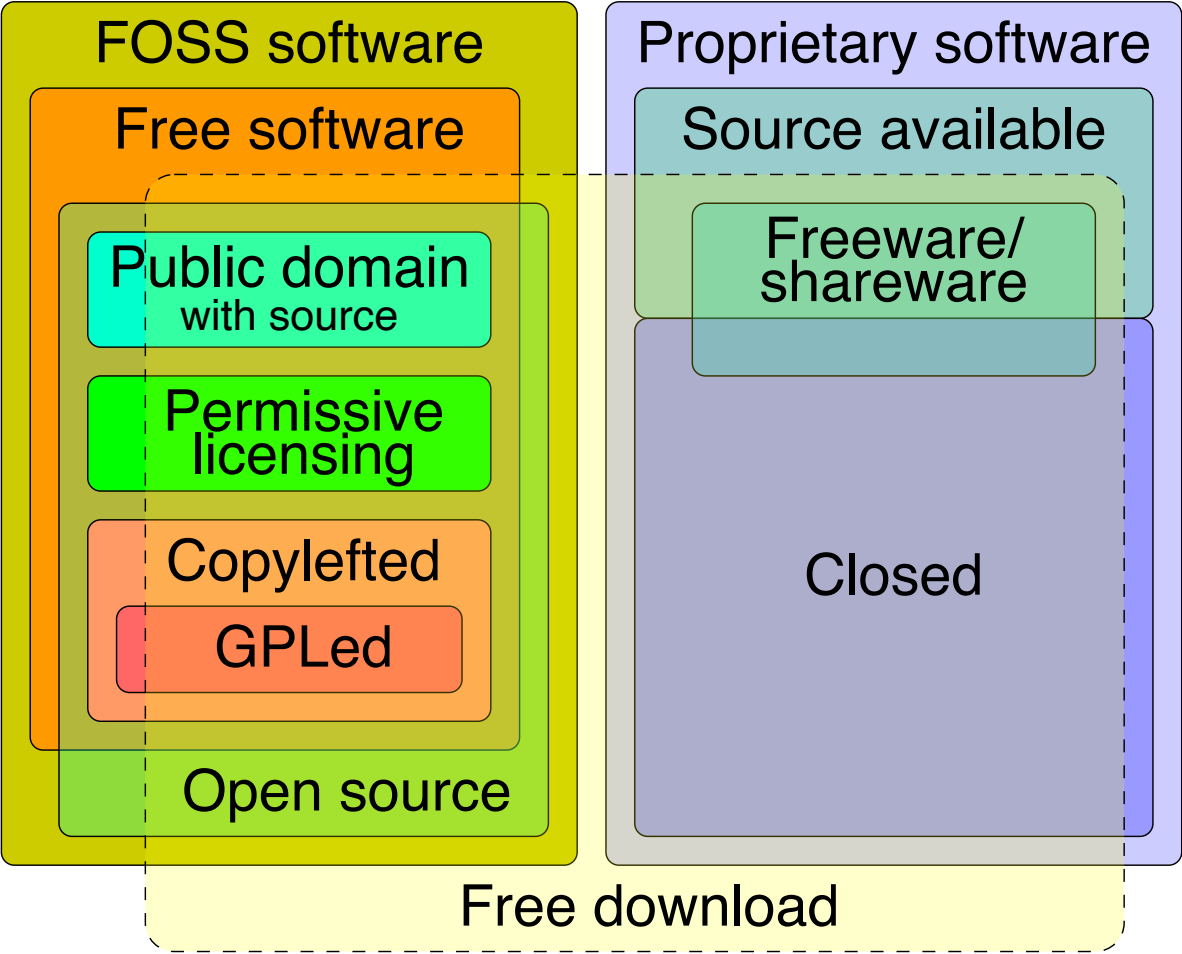
Software	Proprietary	Freeware	FLOSS
goal	make the user captive	free of charges, demo mode; if ok ⇒ proprietary	user sovereignty
return	profits for owners	more sophisticated versions sold (priv. comp., public serv.)	fair and equitable compensation for developers
development	salaried jobs	both	dynamic but limited community
example	Google, Windows (Microsoft), Discord, ...	antivirus, ...	internet (in its early days), Mozilla, Linux, Mattermost, ...

FOSS-FLOSS (Free/Libre/Open-Source Software) vs. Freeware

Free does not mean free of charge, it concerns only the user freedoms
think of free as in free speech, not as in free beer Richard Stallman
Freeware: can be copied without paying anyone, but comes without source and cannot be modified

3 categories = simplistic!
proprietary and freeware
software often rely on a mix of
proprietary and open source
code, and therefore a mix of
licenses
(e.g. Chrome vs. Chromium)

from wikipedia:



Free access to a resource can suddenly change its business model.

{features}

Write, collaborate, and create with Overleaf

Overleaf is the collaborative LaTeX editor that anyone can use.
Find out why we're loved by beginners and TeXperts alike.

Sign up for free



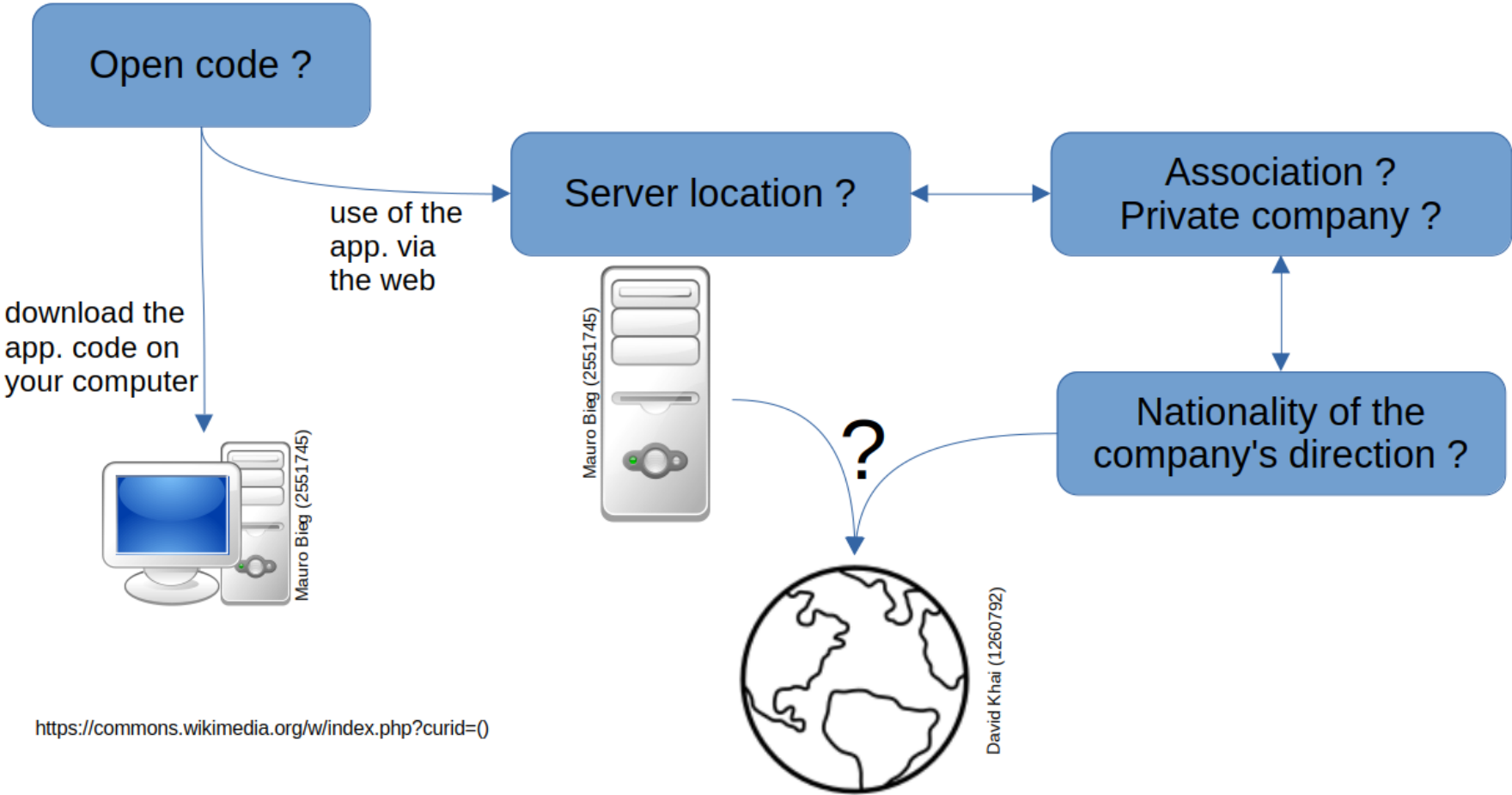
changed its policy in 2025 by limiting
collaborative writing of LaTeX
documents to a single collaborator in its
free version



Public research is subsidized by governments, and its results are, in principle, made available to the public.

This is the objective of “open science” initiatives.

Questions about tools



<https://commons.wikimedia.org/w/index.php?curid=0>

The resources supporting the work of bioinformaticians can be divided into three parts:

- everyday tools: email, instant messaging, professional and social networks, videoconferencing, online marketplace...
- bioinformatics code development: OS, package manager, code versioning...
- database resource

Everyday tools

From all of us:

- «Bring Your Own Device Adoption» in the professional sphere of proprietary software that makes things easier (Google Docs, WhatsApp, Facebook, etc.)
- Addiction to convenience and difficulty with change

On behalf of our guardians:

- Contracts signed with Microsoft (Windows OS, Office suite, Outlook for emails) as part of the *marchés publics*
- Formation :
 - companies offer schools, colleges, high schools, and universities free versions of their software
 - employee training: the only training courses offered are for Microsoft tools (Excel/Word/PowerPoint)

Collaborative documents:



Common tools:



Reference catalog of recommended free software for the administration: [SILL](#), *Socle Interministériel de Logiciels Libres*, bottom-up enrichment by *tout agent du service publique*

Usage	FOSS Solution
Professional network	ORCID
Document sharing	Nextcloud , Resana
Source code sharing	GitLab
Word processing	OnlyOffice , LibreOffice , LaTeX , typst
Search for free software	SILL , Framalibre
Video conferencing	BigBlueButton , lasuitenumerique
Social network	Mastodon (e.g. sciences.re)
Instant messaging	Signal , Tchap
Email	Thunderbird
Web browsers	Firefox , Chromium
Search engines	DuckDuckGo , Qwant
Operating systems	Linux (home version Ubuntu)
Email and cloud service providers	CHATONS

[B.Albert](#), [C.Dillmann](#)
[et al. 2025](#)

Bioinformatics developement

Computer code constitutes data in itself (even if its volume is negligible compared to biological data)

Challenges for open science (free software, FAIR coding)

- Versioning, traceability, documentation
- Collaborative development environments
- Functional validity
- Cybersecurity

- **OS:** Windows < macOS $\Rightarrow$ Linux
- **Tool manager:** Anaconda () $\Rightarrow$ Conda (limited to conda-forge, bioconda), Pixi, Guix
- **Integrated Development Environment (IDE):** VSCode () $\Rightarrow$ VSCodium, Vim, JetBrains IDEs ()
- **Code versioning:** BitBucket < GitHub (Microsoft) < GitLab.com (inc.) $\Rightarrow$ sourcehut, Codeberg or your GitLab institution: CNRS: [src.koda](https://src.kodanetwork.org/), INRAe: [forge.inrae](https://forge.inrae.fr/), university: [sourcesup](https://sourcesup.univ-lille.fr/)
- **Workflow:** Galaxy, Nextflow () , Snakemake, ...
- **Matlab** () $\Rightarrow$ Scilab
- **Container:** Docker $\Rightarrow$ Podman, Singularity $\Rightarrow$ Apptainer

VSCode source code & MIT-licensed

Microsoft's Visual Studio Code **source code is open source** (MIT-licensed), but the **product** available for **download** (Visual Studio Code) is licensed under a **not-FLOSS license** and contains **telemetry/tracking**. Moreover, the integrated extension marketplace is not available in VSCodium (an alternative marketplace without all extensions must be used).

- **Complete environments for collaborative development:** version management, tasks, projects, bug reporting and tracking, continuous integration, code distribution, etc
- **American commercial solutions** (free or paid) **are largely predominant** for French academic repositories: github.com (~70%) and gitlab.com (~10%)
- **Institutional academic solutions** (CNRS, INRAE, Pasteur), but access is restricted to staff of the organization, and features are less extensive than commercial solutions
- **Forge platforms are not archives**, even though they are often used for this purpose

- **HAL** (French open archive): generalist archive, not specific to software, but widely used in France for code archiving
- **Software Heritage** ★: French initiative with international reach, trusted repository recommended by the Data College of the MESR Open Science Committee and Unesco (“archiving open software human heritage”)
- **RENATER SourceSup**, academic code repository
- **Zenodo**, repository for data and digital works (all types, without moderation or quality control)

Biological databases

- Genbank (nucleic acid sequences)
- SRA (Sequence Read Archive, raw sequencing data)
- Protein (protein sequences)
- Taxonomy (taxonomic classification of organisms)
- PubMed (bibliographic database of life sciences and biomedical articles)
- GEO (Gene Expression Omnibus, functional genomics data)
- ClinVar (genetic variants and their relationship to human health)
- ...
- many others (e.g. geolocated data) ...

Mirror

- [Genbank](#)@NCBI → [Ensembl](#)@EBI + [DDBJ](#) (JPN). International Nucleotide Sequence Database Collaboration ([INSDC](#)): synchronization between all of nucleic acid sequence data submitted to one of them.
- [SRA](#) (Sequence Read Archive) @NCBI → [ENA](#) (European Nucleotide Archive) @EBI
- [Protein](#)@NCBI → [Uniprot](#)@EBI+[SIB](#)

Partial mirror (with dependence on NCBI for the source)

- [Taxonomy](#): [ENA](#)@EBI mirror of the subset of taxa for which sequences are available
- [PubMed](#): 37 million articles, and [PMC](#) (full-text open articles). Partially covered by [Europe PMC](#), with differences.
- [GEO](#): [ArrayExpress](#) (EBI) until 2017

No equivalent outside the USA (human)

- [ClinVar](#) : genetic variants associated or not with diseases
- [OMIM](#): comprehensive collection of genes and phenotypes, updated daily
- [GnomAD](#): aggregation of exome and genome sequences from large-scale seq. projects.

Main bioinformatics actors in Europe

Outstation of the European Molecular Biology Laboratory (EMBL), founded in 1996.

Funded by member states (including ).

Wide coverage of biological data types: proteins, nucleic acids, biochemical reactions, metabolic pathways, etc.

Offer data exploitation by independent and different software platforms (genome browsers, analysis tools)

EMBL-EBI Home > Services > Data resources and tools

EMBL-EBI data resources and tools

EMBL's European Bioinformatics Institute maintains the world's most comprehensive range of freely available and up-to-date molecular data resources.

Search

Showing 1 - 15 out of 150 results

Sort by: Featured



[Europe PMC ORCID Tool](#)

A Europe PMC tool that allows article authors to search Europe PMC to build a bibliography and export it to their ORCID record.

TOOL | Training resources

Tags: Bibliography generation



[GWAS Catalog](#)

The NHGRI-EBI GWAS Catalog is a quality-controlled, manually curated, literature-derived collection of all published genome-wide association studies.

DATA RESOURCE | Web API | EMBL-EBI Terms of use | Training resources



[Identifiers.org](#)

Identifiers.org provides a manually maintained Registry of biomolecular data collections and a Resolver to map URIs to URLs.

DATA RESOURCE | Web API | CC-BY



[International Genome Sample Resource](#)

A deep catalog of shared human genetic variation in population groups worldwide that follows from the 1000 Genomes Project.

DATA RESOURCE | EMBL-EBI Terms of use



intergovernmental EU organisation that brings together life science resources (DB, software tools, training materials, cloud storage, HPC), based on national nodes (IFB for France)

→ see Bonus: ELIXIR Position Paper: The importance of open data infrastructures for European competitiveness in life sciences

Service Delivery Plan

Bioinformatics services with international scope, supported and certified by national nodes + by ELIXIR. Shared certification criteria: impact, best practices, open access, sustainability, etc.

- 615 services, including 38 FRA
- 187 data resources, including 18 FRA

Core Data Resources

Data resources considered essential to biology. Extremely demanding certification process.

- 31 core data resources, including 1 FRA (Orphadata + 2 certifications in progress, 1 other in the pipeline)

Conclusion

Data/code ownership

The data stored on US platforms has been produced and deposited by an international research community of data producers. In principle, the **data belongs to the producers** and not to the data platform.

Instability of data resources

The current crisis has **raised** awareness of **the dependence** on the vagaries of a single country's policy and the need to secure and sustain these data resources.

Access and redistribution

Data resources are based on **different** types of **licenses** which authorize or prohibit downloading, keeping a local copy, academic or commercial reuse, complete or partial redistribution, etc.

Data sensitivity

- Some databases contain **personal data** (metadata, health data, genomic or transcriptomic sequences)
- Consent forms specify a destination and conditions of availability

- Some resources are based on USA legislation and **cannot be deployed on European servers due to the GDPR** (more restrictive ; e.g. population genomics)

And their associated tools ?

🔧 The futility of data without tools

The value of data lies in the ability to exploit it, which relies on software platforms

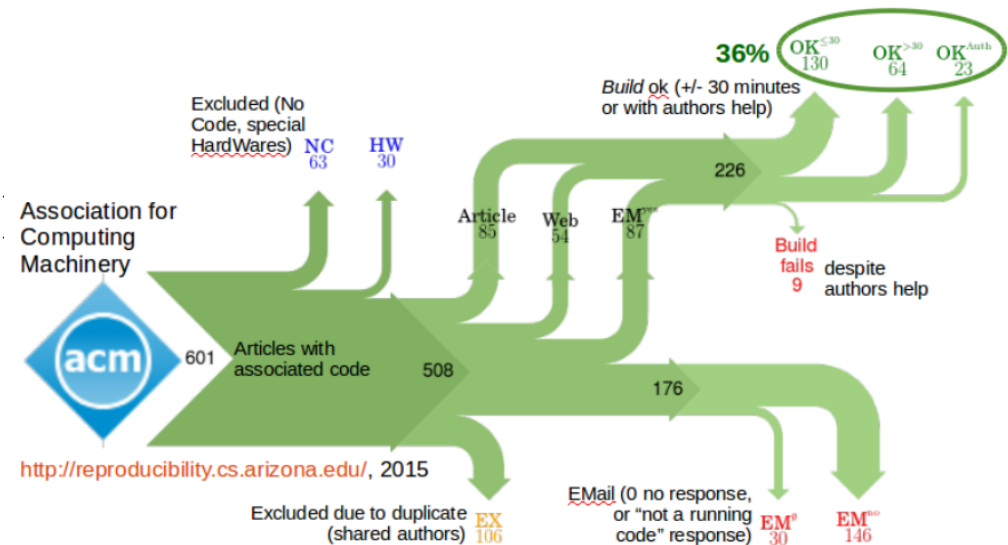
→ It is **not enough to download a copy** to be able to reuse it

These platforms are complex and are **based on decades of development** by teams of bioinformaticians

→ re-implementation is not feasible within a reasonable timeframe and with reasonable resources

👻 Enough to give you nightmares

- OVH data center fire (2021)
- Loss of access to GitHub repositories (2023)
- Tools installed on Amazon web services
- and not to mention the installation
- ...



Example of National initiatives



<https://lasuite.numerique.gouv.fr/>



<https://www.opendesk.eu>

Example of European initiatives



European Digital Infrastructure Consortium - EDIC <https://digital-strategy.ec.europa.eu/en/policies/edic>



Forum of organizations that fund biological data resources (globalbiodata.org)

The GBC envisions a world in which biodata, essential to science, are freely and openly available to all, with long-term access assured through the tools and services offered by global biodata resources.

Strategy 2023-2026

- Expand and strengthen the commitments of funders at the global level
- Strengthen dialogue among funders to ensure the sustainability of data resources
- Demonstrate and defend for the necessity, value, and impact of biological data resources
- Develop new models for collaborative cross-funding
- **Identify and maintain a list of Global Core BioData Resources** (GCBRs)
- Characterize the diversity of biological data resources to strengthen their funding and sustainability

GBC Members

- National Institutes of Health ()
- National Science Foundation ()
- Wellcome ()
- State Secretariat for Education Research & Innovation SERI ()
- United Kingdom Research & Innovation ()
- Research Council of Norway ()
- Chan Zuckerberg Initiative ()
- Genome Canada ()
- European Molecular Biology Laboratory ()
- African Academy of Sciences

GBC Observers

- Commonwealth Scientific and Industrial Research Organisation
- European Commission ()
- European Research Council ()
- Institut National de la Santé Et de la Recherche Médicale (Inserm) ()
- South African Medical Research Council ()

Example of Mondial initiatives

The Data Rescue Project started in Feb. 2025 as a coordinated effort of three data organizations, including members of [IASSIST](#), [RDAP](#), and the [Data Curation Network](#)



Data Rescue Project



International Association for Social Science
Information Service & Technology



RDAP
ASSOCIATION

**DATA
CURATION
NETWORK**

Our goal is to serve as a clearinghouse for data rescue-related efforts and data access points for public USA governmental data that are currently at risk. We want to know what is happening in the community so that we can coordinate focus. Efforts include: data gathering, data curation and cleaning, data cataloging, and providing sustained access and distribution of data assets.

In Oct. 2025, the Data Rescue Project **Portal** displays **1139 datasets** across **86 governments** compiled through the efforts of over 250 volunteers.

This crisis has opened our eyes to existing or emerging alternative solutions.

Exemple: the International Court of Justice will replace MS Office by OpenDesk

However, these alternatives often come with their own set of challenges and limitations:

- ▶ Multiple solutions for the same need
- ▶ Lack of maturity or features compared to established proprietary solutions
- ▶ Resistance to change from users accustomed to proprietary tools



Therefore, it is crucial to find a balance between achieving digital sovereignty and ensuring usability, efficiency, and user satisfaction

To organize this training, we used:

- Framaform (instead of Google forms) 😊
- Grist (instead of Google Sheets) 😐
- Visio (instead of Zoom) 😍
- Etherpad 😊 and next Docs (instead of Google Docs) 😄
- Resana and sDrive (instead of Google Drive) 😞
- Tchap (instead of slack) 😄
- Typst (local) & Libre office (instead of Google slides) 😊 / 😞
- GitLab (instead of GitHub) 😐 / 😄

Almost final approval for our other projects.

Special acknowledgments

- Jacques van Helden for slides about Trump impacts and biological databases counterparts.
- Christine Dillmann for slides about FOSS solutions

Bibliography

- [B.Albert, C.Dillmann, ..., C.Noûs](#) 2025 Résister au mouvement anti-science américain : libérons-nous des GAFAM et sécurisons nos données scientifiques
- [M.Watin-Augouard](#) 2025 Qui exerce l'autorité dans l'espace numérique ?
- [N.Smyrnaio](#) 2023 Les GAFAM, entre emprise structurelle et crise d'hégémonie
- [Annales des Mines Enjeux numériques N°31](#) 2025 Données géolocalisées, souveraineté et indépendance stratégique
- [Web magazine Web ENSAE-Alumni](#) 2021 “GAFA, Reprenons le pouvoir !” - Entretien avec J.Toledano
- [ELIXIR Position Paper](#) 2025 The importance of open data infrastructures for European competitiveness in life sciences

Bonus

- Everyday tools & {fr .eu .un} legislation
- Digital Europe Programs
- ELIXIR Position Paper: Key recommendations
- Bioinformatics resources: who, where, & how

- **Fr**: law for a Digital Republic (Oct. 7, 2016, Art. 16): includes the **default online publication of public data**, recommends the use of formats that promote interoperability, and mandates the **use of free software and open formats** in public administration information systems in order to preserve control, sustainability, and independence
- **Fr**: Prime Minister's data circular (Apr. 27, 2021) to support the digital transformation of the State: better understand and **use free software** and digital commons in the administration, develop and support the opening up of public sector source codes, and rely on **free and open source software** to enhance the attractiveness of the State as an employer for digital talent
- **Eu**: General Data Protection Regulation (**GDPR**, EU 2016/679): concerns the protection of individuals with regard to the processing of personal data and the movement of such data
- **UN** (United Nations): Global Digital Compact adopted at the Future Summit (New York, 2024): ambitious global digital pact defining principles, objectives, and actions aimed at supporting an **open, free, and secure digital future** for all, which respects human rights, bridges digital divides, and reaffirms multi-stakeholder governance of the internet

Objectives: high-performance computing, artificial intelligence, cybersecurity and trust, advanced digital skills, deployment and better use of digital capabilities

2 regulations (EU) of the European Parliament in 2022 (application in 2024):

▸ **DMA** “Digital Markets Act”: to regulate the economic activity of large digital platforms within the European Union.

In 2023, 6 companies were designated as “gatekeepers” by the European Commission:

- Alphabet (Google, Chrome, Android, Youtube)
- Amazon
- Apple (iOS, Safari, App Store)
- ByteDance (TikTok)
- Meta (Facebook, Instagram, Whatsapp, Messenger)
- Microsoft (Windows, LinkedIn)

▸ **DSA** “Digital Services Act”: to regulate digital services

Based on [ELIXIR Hub](#) Published 23 Jun 2025

The importance of open data infrastructures for European competitiveness in life sciences

- Deliberately constructive rather than critical wording.
- Importance of open data resources: research, innovation, competitiveness.
- Underestimation of their value and underfunding.
- Strategic importance of data, and proposed actions.
- Need to ensure the sustainability of data resources.
- Importance of data “FAIRness” for its use by AI.
- Strategy to position Europe as a world leader in open science.

I. Open data resources are the foundation for life science innovation: Europe should invest in open data as a critical enabler of innovation

- Explore **transitional funding mechanisms** to support critical European resources following reduction in USA support.
- **Recognise data resources as infrastructure, not research projects**, with suitable funding instruments and lifecycle management.

II. To reach their full potential, AI investments must be anchored in sustainable data infrastructure

- Anchor investments in AI and data spaces in **sustainable, FAIR-compliant and openly accessible datasets**— particularly in life sciences, where data complexity is highest and potential impact greatest.

III. Secure Europe's data sovereignty through investing in world-leading data resources, a skilled workforce and national capacity aligned with European infrastructure

- Reinforce Europe's sovereignty by **strengthening Europe's role as a net data contributor**, ensuring others continue to align with European infrastructures and open data values.
- **Invest in human capital** by sustaining a digital workforce with expertise in managing, curating and exploiting complex datasets.
- Amplify national efforts by **incentivising alignment of national capacities with EU-level strategies** (e.g. the EU Life Science Strategy, the ESFRI and the Common European Data Spaces including EOSC)

IV. Europe should play the role of global leader in open science, supporting multilateral collaboration and championing the benefits of open data

- Support multilateral science by providing stable, **long-term funding** for international open data infrastructures with strong European participation.
- **Leverage diplomacy through openness** by positioning Europe as a principled and dependable steward of global scientific collaboration.
- Lead through action by **making open science and shared data infrastructures central, well-funded pillars of the next Framework Programme**—signalling to the world that Europe values open data and global collaboration.

containerization

Software	Cie/author	country	license - code
Docker	Docker Inc.	USA	proprietary software
Apptainer	Gregory Kurtzer	USA	BSD-3

email

Software	Cie/author	country	license - code
Thunderbird	Mozilla Foundation	USA	MPL-2.0 (Mozilla Public-License)

e-store

Software	Cie/author	country	license - code
App Store	Apple	USA	proprietary software
Amazon	Amazon	USA	proprietary software

IDE

Software	Cie/author	country	license - code
VSCode	Microsoft	USA	freeware
VSCodium (open source part of VSCode)	Microsoft	USA	MIT
vim	Bram Moolenaar	NLD	VIM(GPLv2)
JupyterLab	Fernando Pérez	COL	BSD-3
RStudio	Posit Software	USA	AGPLv3

online chat

Software	Cie/author	country	license - code
Matrix	Matrix.org Foundation C.I.C.	GBR	Apache License 2.0
Discord	Discord Inc.	USA	proprietary software
Mattermost	Mattermost Inc.	USA	server MIT, other AGPLv3
Slack	Salesforce	USA	proprietary software

OS

Software	Cie/author	country	license - code
Android	Alphabet Inc.	USA	Apache License 2.0
iOS	Apple	USA	proprietary software
Windows	Microsoft	USA	proprietary software
Linux	GNU	USA	GPLv2

package manager

Software	Cie/author	country	license - code
Conda	Anaconda Inc.	USA	freeware (conda BSD)
Pixi	prefix.dev	DEU	BSD-3
Guix	GNU	FRA	GPLv3+

search engines

Software	Cie/author	country	license - code
Google	Alphabet Inc.	USA	proprietary software
Chrome	Alphabet Inc.	USA	freeware
Chromium	Alphabet Inc.	USA	BSD-3
Safari	Apple	USA	freeware
Firefox	Mozilla Foundation	USA	MPL-2.0 (Mozilla Public-License)
DuckDuckGo	DuckDuckGo, Inc.	USA	Apache License 2.0
Qwant	QWANT	FRA	GPLv3

share content

Software	Cie/author	country	license - code
YouTube	Alphabet Inc.	USA	proprietary software

software forge

Software	Cie/author	country	license - code
GitHub	Microsoft	USA	proprietary software
GitLab CE	GitLab Inc.	USA	MIT
GitLab EE	GitLab Inc.	USA	proprietary software
Git	Linus Torvalds	FIN	GPLv2

translation

Software	Cie/author	country	license - code
DeepL	DeepL SE	DEU	MIT

workflow system

Software	Cie/author	country	license - code
Nextflow	Seqera Labs	ESP	Apache License 2.0
Snakemake	Johannes Köster	DEU	MIT
Galaxy	Galaxy community	Galaxy community	MIT (from 2021-04-07; Academic Free License v3 before)