

# FAIR Bioinfo 2022

## Best practice in your bioinformatic projects



P. Marin, M. Hiriart, P. Ruiz & N. Goué  
aubi@uca.fr

Université Clermont Auvergne, AuBi, Mésocentre

26 avril 2023



. This work is based on the IFB and I2BC formation offer

## Essay

# Why Most Published Research Findings Are False

John P. A. Ioannidis

## Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

factors that influence this problem and some corollaries thereof.

## Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

## It can be proven that most claimed research findings are false.

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. “Negative” research is also very useful.

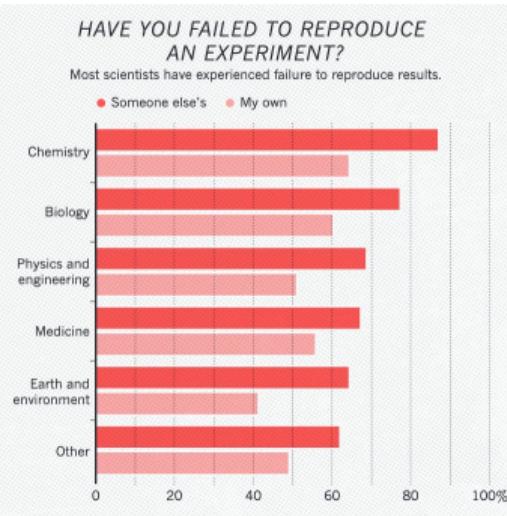
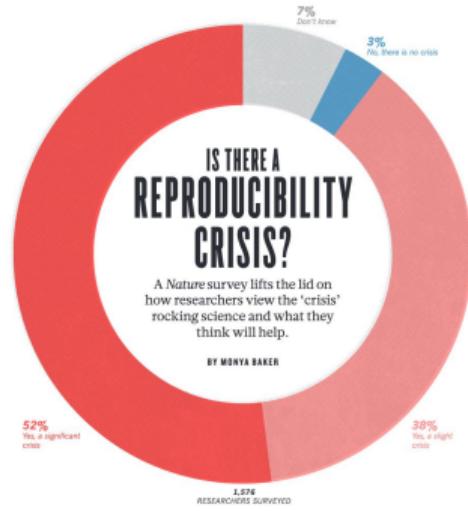
is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is  $R/(R+1)$ . The probability of a study finding a true relationship reflects the power  $1-\beta$  (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate,  $\alpha$ . Assuming that *c* relationships are being probed in the field, the expected values of the  $2 \times 2$  table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the  $2 \times 2$  table, one gets  $PPV = (1 - \beta)R/(1 - \beta)R + \alpha(1 - R)$ .

## Crisis elements

- Highlighted around 2005
- Since 2010 more articles related to the non reproducibility
- Medicine is one of the most impacted discipline

# Reproducibility crisis

2016

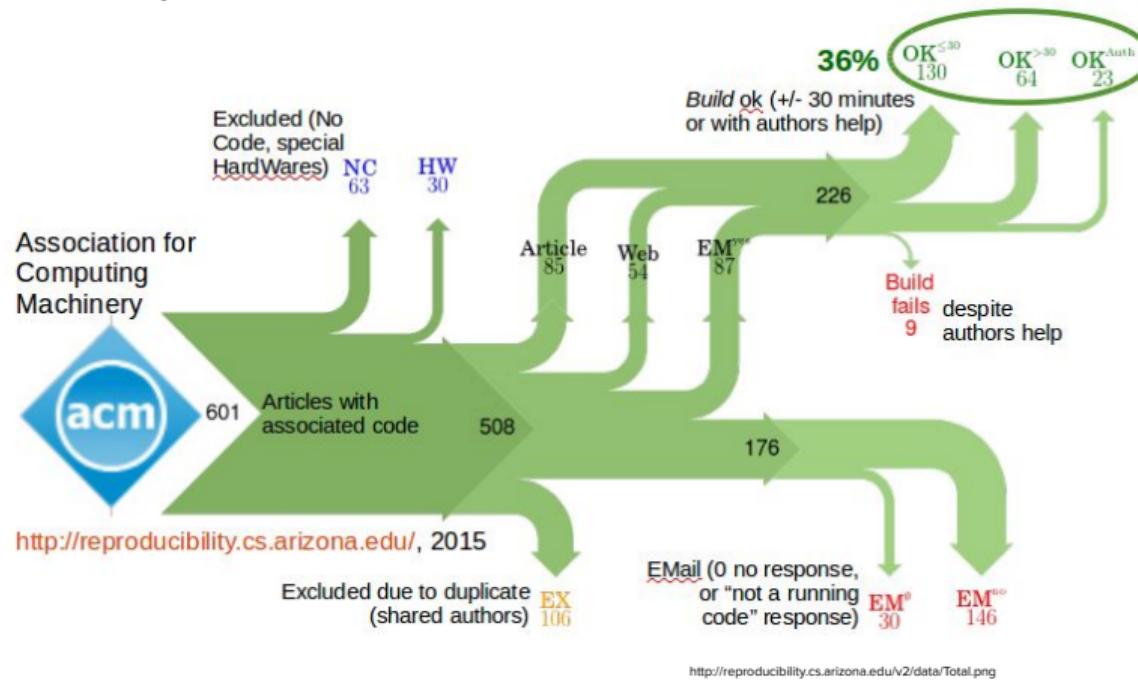


Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). <https://doi.org/10.1038/533452a>

4

<https://doi.org/10.1038/533452a>

## Also in computer sciences



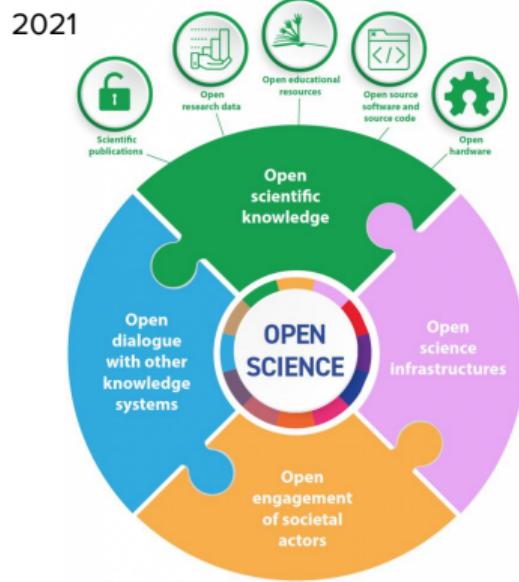
# Long term negative impact of retracted papers

| Article  | Year of retraction | Citing Articles before retraction | Citing Articles after retraction | Total cites (journals indexed by Web of Science) |
|--|--------------------|-----------------------------------|----------------------------------|--|
| 1. Primary Prevention of Cardiovascular Disease with a Mediterranean Diet. N ENGL J MED; APR <b>2013</b> . Estruch R, et al.   | 2018               | 1919                              | 816                              | 2735   |
| 2. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. LANCET; FEB 28 <b>1998</b> . Wakefield AJ, et al.                         | 2010               | 642                               | 867                              | 1509   |
| 3. Visfatin: A protein secreted by visceral fat that mimics the effects of insulin. SCIENCE; JAN <b>2005</b> . Fukuhara A, et al.  | 2007               | 232                               | 1192                             | 1424   |
| 4. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. PLANT J; MAR <b>2003</b> . Voinnet O, et al. | 2015               | 896                               | 375                              | 1271   |
| 5. Lysyl oxidase is essential for hypoxia-induced metastasis. NATURE; APR <b>2006</b> . Erler JT, et al.   | 2020               | 977                               | 81                               | 1058   |

Retraction Watch : Top 10 most highly cited retracted papers  
<https://retractionwatch.com/the-retraction-watch-leaderboard/top-10-most-highly-cited-retracted-papers/>

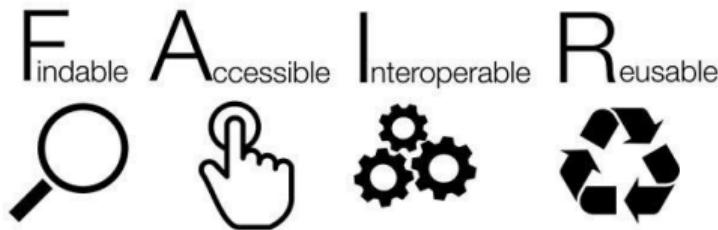
6

# A way out: Open science and FAIR principles



Graphic on page 11. [UNESCO Recommendation on Open Science](#). CC BY IGO 3.0 C. Green

2016



Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).  
<https://doi.org/10.1038/sdata.2016.18>

7

## FAIR history

- Born in 2016 with *The FAIR Guiding Principles for scientific data management and stewardship*
- How to build, stock, share, use and publish data
- Make criteria to better use our data

. <https://doi.org/10.1038/sdata.2016.18>

# SCIENTIFIC DATA

Amended: Addendum

OPEN

## SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

## Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.<sup>#</sup>

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

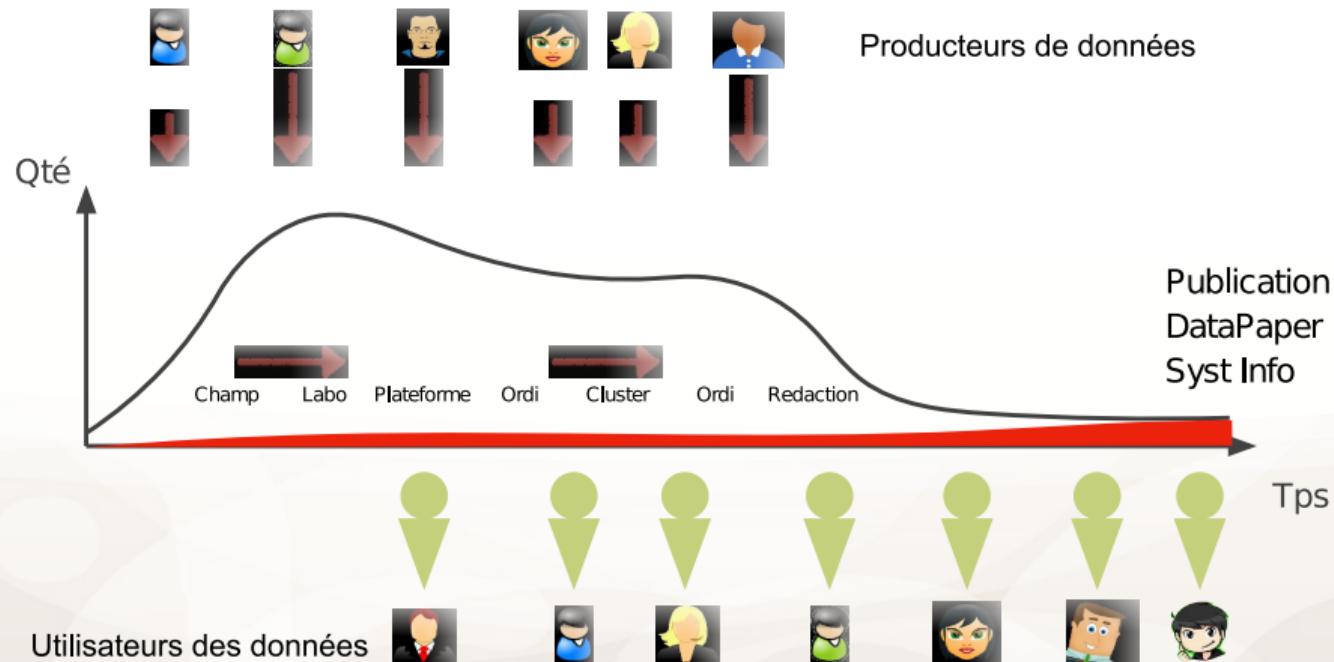
<https://doi.org/10.1038/sdata.2016.18>

## Apply FAIR TO

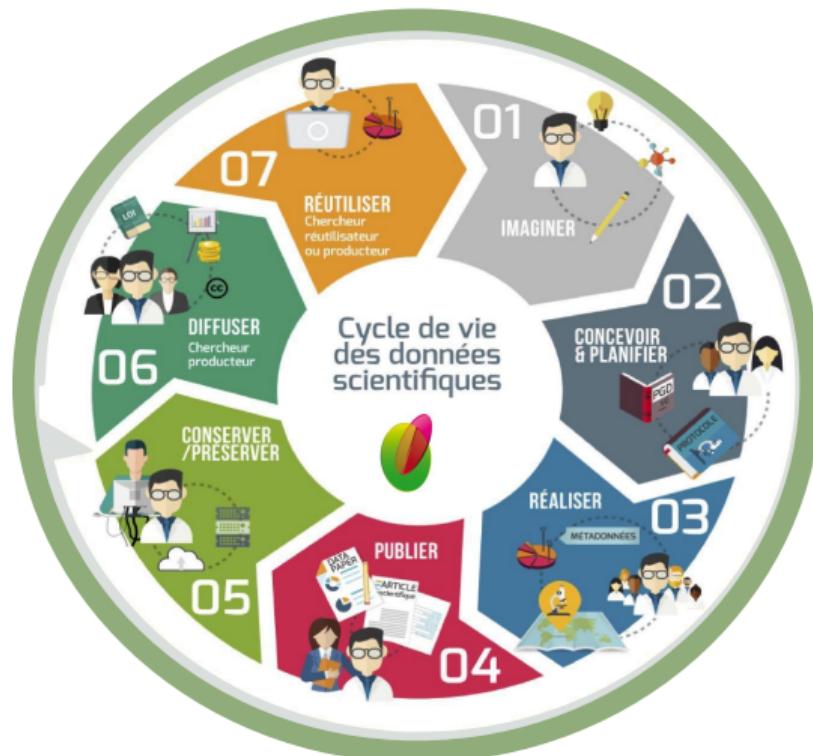
### ■ Your DATA

- Data lifecycle
- Data Management Plan (DMP)
- Metadata
- Data storage

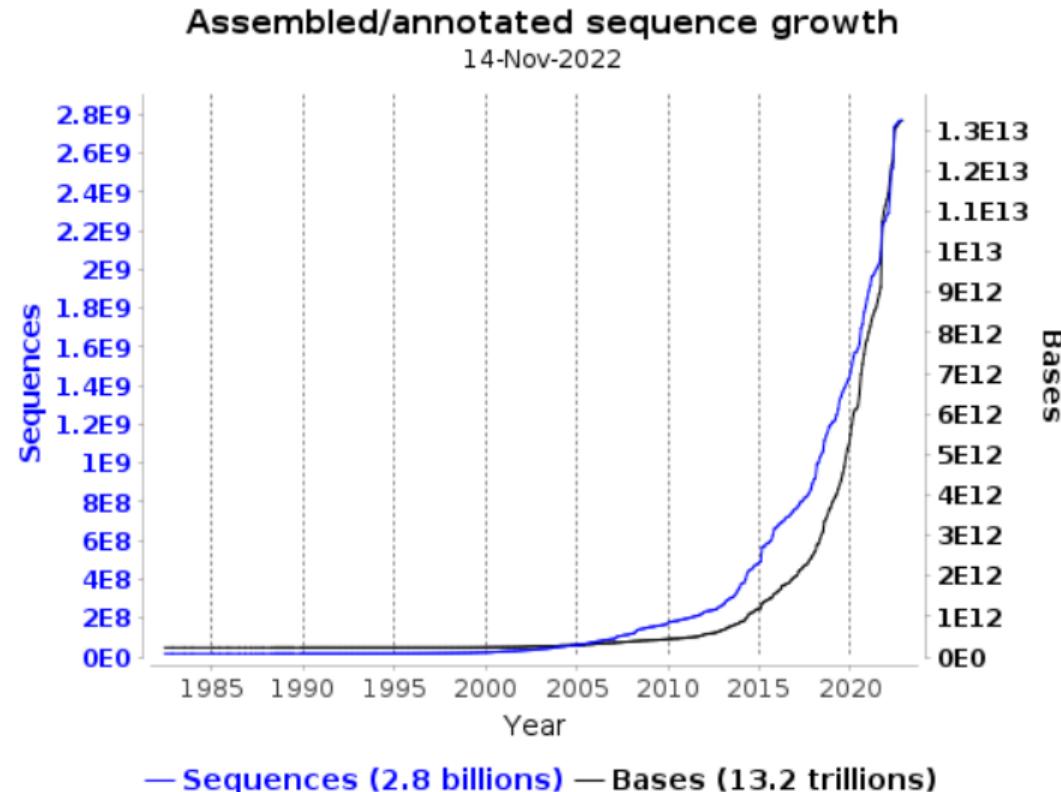
# Focus on the data lifecycle



# Focus on the data lifecycle



# Focus on the data lifecycle



<https://www.ebi.ac.uk/ena/browser/about/statistics>

# Focus on the data exchange

## Transfert de vos données de recherche



### Comment transmettre vos données ?





## Cost of not having FAIR research data

Cost-Benefit analysis for FAIR research data

## Apply FAIR TO

- Your DATA

- Data lifecycle
- Data Management Plan (DMP)
- Metadata
- Data storage

- Your scripts, environment...

- Objective of this training

# FAIR principles

F  
indable



By 維基小霸王 - Own work, CC BY-SA 4.0.  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

PID  
Repository

8

<https://doi.org/10.1038/sdata.2016.18>

# 404

This is not the  
web page you  
are looking for.



## To be Findable

- (meta)data are assigned a globally unique and persistent identifier
- data are described with rich metadata
- metadata clearly and explicitly include the identifier of the data it describes
- (meta)data are registered or indexed in a searchable resource

# FAIR principles

F  
indable



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

A  
ccessible



<https://nitsfirstworldproblems.tumblr.com/post/147555650875/i-can-t-reach-the-top-shelves-of-the-kitchen>

PID  
Repository

Protocols  
(free, open, auth.)

9  
<https://doi.org/10.1038/sdata.2016.18>

## To be Accessible

- (meta)data are retrievable by their identifier using a standardized communication protocol
- the protocol is open, free, and universally implementable
- the protocol allows for an authentication and authorization procedure, where necessary
- metadata are accessible, even when the data are no longer available

# FAIR principles

**F**indable



By 維基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

**A**ccessible



<https://nitsfirstworldproblems.tumblr.com/post/147555650875/i-can-t-reach-the-top-shelves-of-the-kitchen>

**I**nteroperable



By Unknown author - Popular Science Monthly Volume 88, Public Domain  
<https://commons.wikimedia.org/w/index.php?curid=22614407>

PID  
Repository

Protocols  
(free, open, auth.)

Standards  
(format, vocabulary)

10  
<https://doi.org/10.1038/sdata.2016.18>

## To be Interoperable

- (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- (meta)data use vocabularies that follow FAIR principles
- (meta)data include qualified references to other (meta)data

# FAIR principles

**F**indable



By 雜基小霸王 - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=88894774>

**A**ccessible



<https://nillsfirstworldproblems.tumblr.com/post/147555650875/i-can-t-reach-the-top-shelves-of-the-kitchen>

**I**nteroperable



By Unknown author - Popular Science Monthly Volume 88, Public Domain  
<https://commons.wikimedia.org/w/index.php?curid=22614407>

**R**eusable



By Sun Ladder - Own work, CC BY-SA 3.0,  
<https://commons.wikimedia.org/w/index.php?curid=5746428>

PID  
Repository

Protocols  
(free, open, auth.)

Standards  
(format, vocabulary)

Metadata  
License  
Origin

11  
<https://doi.org/10.1038/sdata.2016.18>

## To be Reusable

- meta(data) are richly described with a plurality of accurate and relevant attributes
- (meta)data are released with a clear and accessible data usage license
- (meta)data are associated with detailed provenance
- (meta)data meet domain-relevant community standard

# To be Reusable

## A point on the licences

- number different types of licences from proprietary to totally free AND open
- more than 50 different kinds
- No licence = authors' rights
- France = support for free and open data

# To be Reusable

## A point on the licences

- number different types of licences from proprietary to totally free AND open
- more than 50 different kinds
- No licence = authors' rights
- France = support for free and open data

**Recherche dans GitHub par type de licence**

Vous pouvez filtrer les référentiels en fonction de leur licence ou famille de licences à l'aide du qualificateur `license` et du mot clé de licence exact :

| Licence   | Mot clé de licence |
|---|--------------------|
| Academic Free License v3.0                          | afl-3.0            |
| Licence Apache 2.0                                  | apache-2.0         |
| Licence Artistic 2.0                                | artistic-2.0       |
| Licence logicielle Boost 1.0                        | bsl-1.0            |
| Licence BSD « simplifiée » à 2 clauses              | bsd-2-clause       |
| Licence BSD « nouvelle » ou « révisée » à 3 clauses | bsd-3-clause       |

# To be Reusable

## A point on the licences

- number different types of licences from proprietary to totally free AND open
- more than 50 different kinds
- No licence = authors' rights
- France = support for free and open data

The screenshot shows the homepage of the Joinup Licensing Assistant. At the top left is the European Commission logo. The top navigation bar includes links for "Interoperable Europe", "Interoperability Solutions", "Sign in", and "Get started". A search icon is also present. The main header features the text "Joinup Licensing Assistant" over a background of binary code. Below the header, there's a diagram illustrating the licensing process: a circular icon with a person head, a circular icon with a document and a star, a circular icon with a globe, and a circular icon with a folder. A red checkmark is placed over the folder icon. To the right of the diagram is a button labeled "SUBSCRIBE TO THIS SOLUTION". A small box in the bottom left corner displays the EUPL logo and the text "Topic: Legal".

# To be Reusable

## A point on the licences

- number different types of licences from proprietary to totally free AND open
- more than 50 different kinds
- No licence = authors' rights
- France = support for free and open data

[Overview](#)

[Members](#)

[About](#)

[JLA - Compatibility Checker](#)

[JLA - Find and compare software licenses](#)

[REPORT ABUSIVE CONTENT](#)

A unique tool allowing everyone to compare and select open licences based on their content.

[Select licence terms below](#)

|               | Can             | Must          | Cannot         | Compatible     | Law              | Support |
|---------------|-----------------|---------------|----------------|----------------|------------------|---------|
| Use/reproduce | Incl. Copyright | Hold liable   | None N/A       | EU/MS law      | Strong Community |         |
| Distribute    | Royalty free    | Use trademark | Permissive     | US law         | Governments/EU   |         |
| Modify/merge  | State changes   | Commerce      | GPL            | Licensor's law | OSI approved     |         |
| Sublicense    | Disclose source | Modify        | Other copyleft | Other law      | FSF Free/Libre   |         |

Pierre MARIN (Université Clermont Auvergne, AuBi, Mésocentre)

FAIR Bioinfo 2022

26 avril 2023

25 / 30

# To be Reusable

## A point on software archive

- Git-like (github, gitlab) are web services not archive
- Software are fragile

# To be Reusable

## A point on software archive

- Git-like (github, gitlab) are web services not archive
- Software are fragile

The screenshot shows the Software Heritage website. At the top, there is a navigation bar with links for Mission, Archive, Communauté, Grants, Soutien, A propos, and a search icon. The main header features the Software Heritage logo (a stylized orange and yellow starburst) and the text "Software Heritage". Below this, a sub-header reads "préserve le code source des logiciels, pour les générations actuelles et futures". To the right, there is a photograph of a modern library or archive interior with white bookshelves and a staircase. The central part of the page has a large red banner with the text "Nous construisons l'archive universelle des logiciels". At the bottom left is the "Collect Preserve Share" logo, which consists of a stylized orange and yellow starburst with arrows pointing outwards. To the right of the logo, there is explanatory text about collecting and preserving software source code, and another section about conserving and making it accessible.

Software Heritage

préserve le code source des logiciels, pour les générations actuelles et futures

Nous construisons l'archive universelle des logiciels

Collect  
Preserve  
Share

Nous collectons et préservons les logiciels sous forme de code source parce qu'ils sont le support indissociable des connaissances techniques et scientifiques de l'humanité tout entière et que nous ne pouvons pas prendre le risque de les perdre.

Nous conservons et rendons accessible tous les logiciels que nous collectons car c'est uniquement en les partageant que nous pouvons

INSTITUT  
Clermont Auvergne

# FAIR tools

**F**indable



**A**ccessible



**I**nteroperable



**R**eusable



Data



Software  
and  
analyses



[FAIRsharing.org](https://fairsharing.org)

standards, databases, policies



CeCILL

12

ie

# A complete integrated FAIR environment

The Galaxy project



# A complete integrated FAIR environment

The Galaxy project

## Galaxy

Galaxy is an open-source platform for FAIR data analysis that enables users to :

- Use tools from various domains (that can be plugged into workflows) through its graphical web interface.
- Run code in interactive environments (RStudio, Jupyter...) along with other tools or workflows.
- Manage data by sharing and publishing results, workflows, and visualizations.
- Ensure reproducibility by capturing the necessary information to repeat and understand data analyses.

# A complete integrated FAIR environment

## The Galaxy project

The screenshot shows the Galaxy web interface running locally at `localhost:8080/datasets/edit`. The interface includes a navigation bar with links like Connexion - CAS, Workflow, Visualize, Données partagées, Admin, Aide, Utilisateur, and a search bar. The main area features a circular genome annotation plot with concentric rings of colored data (red, orange, blue, green) against a grey background. The plot is labeled with genomic intervals from 0 bp to 1200 bp. To the left is a sidebar with a 'Tools' section containing a search bar and a 'Upload Data' button, followed by a list of tool categories: Get Data, Send Data, Collection Operations, Lift-OVER, Text Manipulation, Convert Formats, Filter and Sort, Join, Subtract and Group, Fetch Alignments/Sequences, Operate on Genomic Intervals, Statistics, Graph/Display Data, Phenotype Association, amr, taxonomy, and trimming. On the right is a 'History' panel titled 'amr\_test' showing two datasets: '58 : genome\_annotation\_plot.pn' and '25 : Enterococcus\_faecalis.fa'. The total disk usage is listed as 284.5 GB.

# A complete integrated FAIR environment

## The Galaxy project

The screenshot shows the Galaxy web interface running on localhost:8080. The main panel displays a workflow titled "staramr" which scans genome assemblies against three databases: ResFinder, PlasmidFinder, and PointFinder. The workflow has a single input file, "25: Enterococcus\_faecalis.fa", listed under the "genomes" section. Below the workflow, there are several configuration sliders for BLAST parameters:

- Percent identity threshold for BLAST: 98.0
- Percent length overlap of BLAST hit for ResFinder database: 60.0
- Percent length overlap of BLAST hit for PointFinder database: 95.0
- Percent length overlap of BLAST hit for PlasmidFinder database: 60.0

On the right side of the interface, there is a "History" panel showing the results of the workflow. It contains two entries: "amr\_test" and "58 : genome\_annotation\_plot.png". The "amr\_test" entry shows a file size of 94.8 GB and contains a file named "25 : Enterococcus\_faecalis.fa". The "58 : genome\_annotation\_plot.png" entry shows a file size of 0.0 GB.

# A complete integrated FAIR environment

## The Galaxy project

The screenshot shows the Galaxy web interface running a workflow titled "abromics\_SR\_PE\_workflow".

**Workflow Options:**

- Search tools: search tools
- Upload Data: button
- Get Data
- Send Data
- Collection Operations
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- amr
- taxonomy
- trimming

**Workflow Details:**

Workflow: abromics\_SR\_PE\_workflow

History Options: Send results to a new history (No selected)

Workflow steps (outputs):

- 1: R1\_fastq
- 2: R2\_fastq
- 3: abromics\_SR\_PE\_trimming
- 4: abromics\_SR\_PE\_taxonomy
- 5: abromics\_SR\_PE\_assembly
- 6: abromics\_assembly\_antimicrobial\_detection
- 7: abromics\_assembly\_annotation
- 8: abromics\_SR\_PE\_QC\_metrics
- 9: abromics\_depth\_amr\_gene\_workflow

**History:**

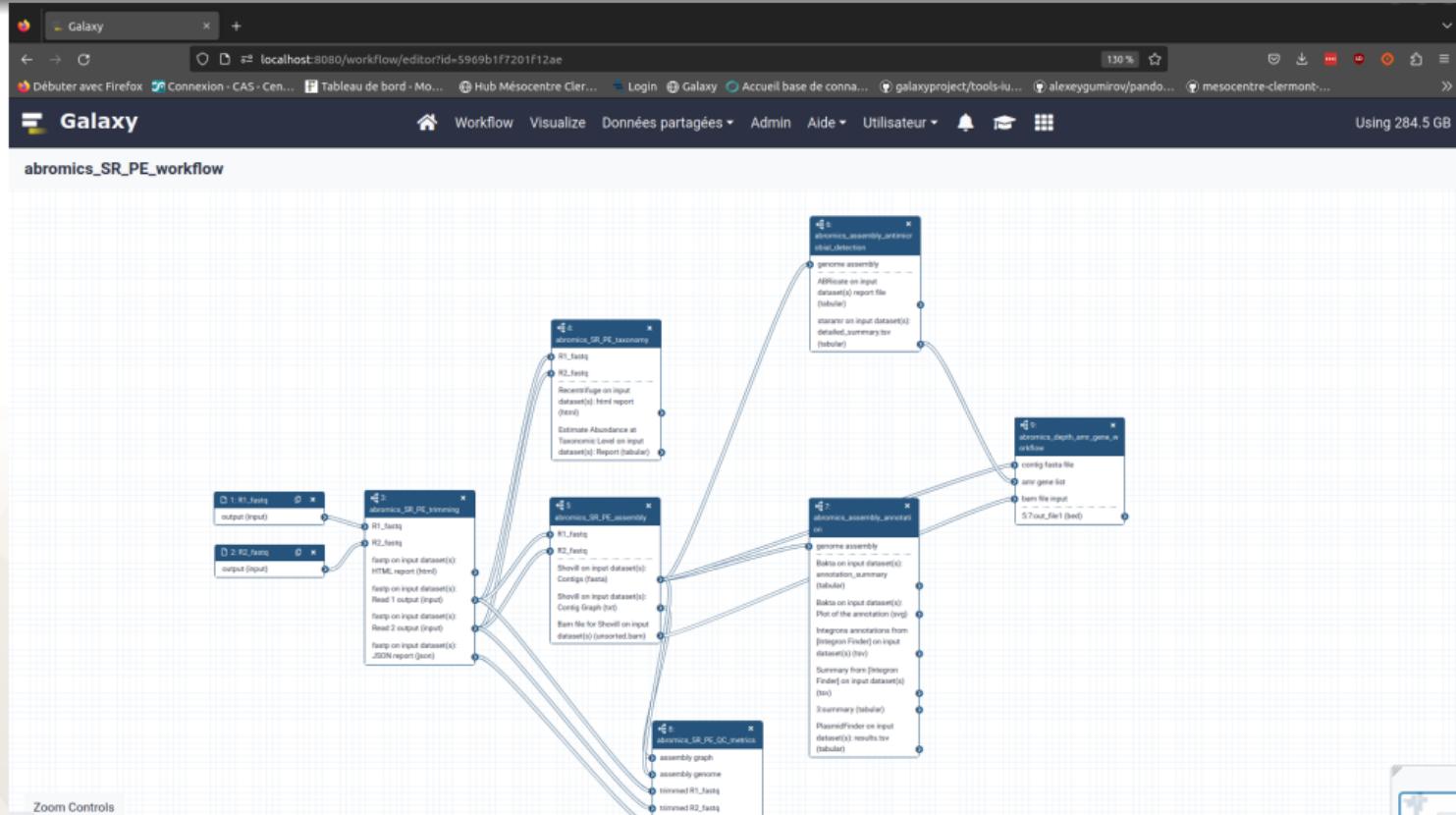
Using 284.5 GB

History: amr\_test

- 94.8 GB (9 files, 53 datasets, 3 annotations)
- 58 : genome\_annotation\_plot.png
- 25 : Enterococcus\_faecalis.fasta

# A complete integrated FAIR environment

## The Galaxy project



# A FAIR approach at any level



*"One practice to rule them all, One practice to find them, One practice to bring them all and in the FAIR bind them."*

# A FAIR approach at any level



*"One practice to rule them all, One practice to find them, One practice to bring them all and in the FAIR bind them."*

## Improvement for who ?

At each level a good practice help you

- Long term efficiency to a bioinformatician
- stop wasting time for a beginner

# How to integrate FAIR concepts in my work ?

Some tools for reproducible research



- FAIR apply on the data
- FAIR apply to the code

# FAIR session with AuBi

## Objectives

- Discover FAIR practices
- Discover tools for best practices
- Learn tools and best practices

# FAIR session with AuBi

## Objectives

- Discover FAIR practices
- Discover tools for best practices
- Learn tools and best practices
- 5 sessions for courses and practices
  - Day 1 : Introduction to FAIR and building training environment
  - Day 2 : Code versioning with Git
  - Day 3 : Environment managment using conda and docker/singularity
  - Day 4 : Workflow managment using snakemake
  - Day 5 : Documentation using Rmarkdown or Jupyter

## Contents

- Introduction to FAIR practices

## Contents

- Introduction to FAIR practices
- Code control using Git 

  - Git environment
  - Gitlab and Github  

## Contents

- Introduction to FAIR practices
- Code control using Git 

  - Git environment
  - Gitlab and Github 

- Encapsulation process
  - Conda environment and packages use 
  - Containers as docker & singularity 
  - Reproducible workflow using snakemake 

## Contents

- Introduction to FAIR practices
- Code control using Git 

  - Git environment
  - Gitlab and Github 

- Encapsulation process
  - Conda environment and packages use 
  - Containers as docker & singularity 
  - Reproducible workflow using snakemake 
- Literate programming and documentation
  - Markdown syntax 
  - Rmarkdown for R 
  - Jupyterlab for Python 

# Biosphere a virtual environment for training

IFB Biosphère RAINBio myVM DATA Support [fr] Se connecter

SE CONNECTER



# Biosphere a virtual environment for training

## BIOSPHERE FEATURES

With IFB-Biosphere, you get:

- A unified user portal ([Biosphere portal](#)) to deploy all bioinformatics environments on all clouds
  - Single sign-on, with your academic credentials ([Sign in](#)).
- Pre-defined bioinformatics environments, available with an one-click deployment from the [RAINBio catalogue](#)
- Infrastructure as Code: most configurations rely on public git repositories available in [Biosphere Commons](#).
- 8 cloud sites with more than 10,000 vCPU and 40 TB RAM ([System status](#))
  - High-availability thanks to the different sites usable equally.
- Modular cloud environments:
  - Single virtual machine (VM) to bunch of VMs.
  - Usual VM: up to 64 vCPUs-250 GB RAM.
  - BigMemory VM: up to 3 TB RAM.
  - HighFrequency VM : up to 3.8GHz vCPU.
  - ManyCores VM: up to 255 vCPUs in a VM
- Bring-Your-Own-Tools
  - Admin rights in the VM (all apps).
  - Deploy with your own container image (some apps).
  - Configure some apps with your own git repository (Infrastructure as Code).
- Useful public biological reference databases
- On-demand resources (CPU, RAM, storage, IFB experts) to support:
  - Training events, university courses, scientific schools, workshops, hackathons.
  - Scientific projects.

## BIOSPHERE PORTAL

The Biosphere portal provides high-level cloud interfaces:

# Biosphere a virtual environment for training

**IFB Biosphère RAINBio myVM DATA Support pierre.marin@uca.fr (edu) GAIN**

## RAINBIO - APPLIANCES BIOINFORMATIQUES DANS LE CLOUD

Catalogue des appliances bioinformatiques dans le cloud, filtrez-les en utilisant les termes présents dans l'ontologie EDAM, ou en langage naturel.

**App Store (59)** Appliances Outils Topics

|  |  |  |  |   |   |
|--|--|--|--|---|---|
| <b>Galaxy</b><br>bioconda, Docker, Galaxy portal<br>Informatics, Bioinformatics, Comparative genomics, Functional                | <b>AnalysesSV</b><br>bcftools, BEDTools, BWA, Jupyter, Matplotlib, pandas, SAMtools<br>DNA polymorphism, Genetic variation, Genotyping experiments, GWAS | <b>ANF MetaBioDiv</b><br>Bioconductor, DESeq2, devtools (R), ggplot2, gridExtra, phyloseq, r<br>Bioinformatics, Computational biology, Data management, Taxonomy | <b>Askomics</b><br>AskOmics<br>Data integration and warehousing, Data visualisation  | <b>BactComparativeGenomics</b><br>ImageJ2, Jupyter, MACS2, Matlab, Nextflow, pandas, Prokka<br>Imaging, Mathematics, Statistics and probability, CHP-seq, Workflows, Data archiving | <b>Bacterial Genomics</b><br>HMMER, Insygh, SGE - GridEngine, Ubuntu, Web interface<br>Protein folds and structural domains, Sequence comparison, Sequence composition, omics |
| <b>Bioimage</b><br>Bureau virtuel, Icy, ImageJ-Fiji, X2Go, Xfce<br>Informatics, Data visualisation, OF, Imaging                  | <b>BioPipes</b><br>bioconda, cwltool, Docker, Nextflow, Snakemake<br>Informatics, Bioinformatics, Workflows  | <b>bistor</b><br>bioconda, Bowtie2, FastQC, Snakemake<br>Bioinformatics, Genomics, Mapping, Sequence alignment, Data quality management                          | <b>CentOS 7</b><br>Ansible, bioconda, Docker<br>Informatics, Bioinformatics  | <b>CentOS 7 Desktop</b><br>Ansible, bioconda, Bureau virtuel, Docker<br>Informatics, Bioinformatics   | <b>CoursAnalysesNanoporeSG</b><br>bandage, Jupyter<br>Data architecture, analysis and design, Mathematics, Statistics and probability   |
| <b>Cytoscape</b><br>Bureau virtuel, Cytoscape, X2Go, Xfce<br>Bioinformatics, Data visualisation, Molecular interactions, Pathway | <b>Debian 10</b><br>Ansible, bioconda, Docker<br>Bioinformatics, Informatics   | <b>Debian 11</b><br>Ansible, bioconda, Docker<br>Bioinformatics, Informatics   | <b>DRomics</b><br>Bioconductor, DESeq2, DRomics, R - base, RStudio, Shiny<br>Bioinformatics, Computational biology, Data management, Taxonomy  | <b>DRomicsInterpreter</b><br>Bioconductor, DESeq2, DRomics, R - base, RStudio, Shiny<br>Bioinformatics, Computational biology, Data management, Taxonomy                            | <b>EBAME-2022 oDNA</b><br>BAMtools, BEDTools, bioconda, Bowtie2, metaDMG-cpp, python3, R<br>Data management, Bioinformatics, Sequence alignment, Computer science, Sequencing |
| <b>EBAME-2022 MetaTOR</b><br>Bowtie2, CheckM, hictst, Pairwise, SAMtools<br>Sequence alignment, Phylogenetics, Population        | <b>EBAME-Anvio</b><br>Anvio  | <b>EBAME-Quince</b><br>bam-readcount, BEDTools, BWA, CONCOCT, DESeq2, Diamond, FastQC<br>Transcriptomics, Informatics  | <b>ETBII AnalyseMultivariée</b><br>BioCStyle, Butcher, clusterProfiler, ComplexHeatmap, DESeq2, DT<br>Literature and language, Transcriptomics | <b>ETBII Réseaux</b><br>BINGO (Cytoscape), Bureau virtuel, compositions (R), Cytoscape, Networkx<br>Transcriptomics, Sequencing   | <b>formation_CIRI</b><br>DESeq2, FastQC, HISAT2, RStudio, Trimomatic<br>Transcriptomics, Whole genome sequencing  |

# Biosphere a virtual environment for training



**Biosphere**

RAINBio myVM DATA

[Support](#)

## VOS PARAMÈTRES

**Informations personnelles**

|  |   |
|--|---|
| Adresse électronique                             | pierre.marin@uca.fr   |
| Affiliation fournie par la fédération d'identité |   |
| Prénom   | Pierre  |
| Nom  | MARIN   |
| Ville et code postal                             | 63170   |
| Formation initiale (optionnel)                   | <ul style="list-style-type: none"> <li>• Biology</li> <li>• Computer Science</li> </ul>   |
| Pubkey   | <ul style="list-style-type: none"> <li>• ssh-rsa<br/>AAAAB3NzaC1yc2EAAAQABAAQDXgGQYRHqk4klU6XoeNqYVFqL14F5WkO3U2xSCDDQW8MpVQTCRqbzTq02GpIMF6014Q0gnC5-W-bTTJkZgnsigSziViug31MOdviZH8oWKXW6c0TZ1InMr0WjbgUBdB7RApTmHL4Laihvz/Ed/IASOTbKAQMLBtm4k3uL7l2RvFMdZcYVTKSbvcilEGShw8OHOHRicNEW3giiCIIUFSTZ5KEZ99/4KgarJRNcIXybzza+mtYGETfIQhJkyRD+R6DNr/hU8aJasgGpPdPwT/9ARVs9/avoF7e0JMUuDxWT9D139EPe44ony/s3k93HK4mpqGV piemari@hpclingo1</li> <li>• ssh-rsa<br/>AAAAB3NzaC1yc2EAAAQABAAQDXgGQCFgYeYnhKbf+mcckwcqzJVBGvRTgZwfhGWeWwfLyJsgDM9fWV+4ZucxtXLrxja69d9r2309E+IdUDoFYRvBtLQhX9X472DqbZ2duuDHvPKtgYaAtK4rjhjhih32VuxSq7HeffrV5csTTbPJhunCTE6MIGLNVGH5lJxCwI-WtDvvaHYCdse2:YRAJeFsqAHtubRQucE2EPKpqRixzS5cg8q2T3kdiYjaquovCvQVQBlzswfkNyvQdg38LMf4JDqOW26ls7WEtavMD1X3EKUEmxm+TldNoc1hnNjgg8XEN/h</li> </ul> |

Connecté.e en tant que pierre.marin@uca.fr

Langues

[en] English

[fr] Français

**Paramètres**

Groups

Quota

[Se déconnecter](#)

<https://biosphere-france.bioinformatione.fr/cloudweb/account/settings/>

# SSH key use

## The principle

- Secure Shell (SSH)
- Securized protocol
- Encryption of informations between users
- Authentification without password
- Assymetric system based on keys

## The keys

- Paired key based
- A public key for everyone to share
- A private key for myself (never share)

# Installing a bash terminal

Git bash for non unix user



--distributed-even-if-your-workflow-isnt



Search entire site...

[About](#)

[Documentation](#)

[Downloads](#)

GUI Clients

Logos

[Community](#)

The entire [Pro Git book](#) written by Scott Chacon and Ben Straub is available to [read online for free](#). Dead tree versions are available on [Amazon.com](#).

## Downloads



macOS



Windows



Linux/Unix

Older releases are available and the [Git source repository](#) is on GitHub.

### GUI Clients

Git comes with built-in GUI tools ([git-gui](#), [gitk](#)), but there are several third-party tools for users looking for a platform-specific experience.



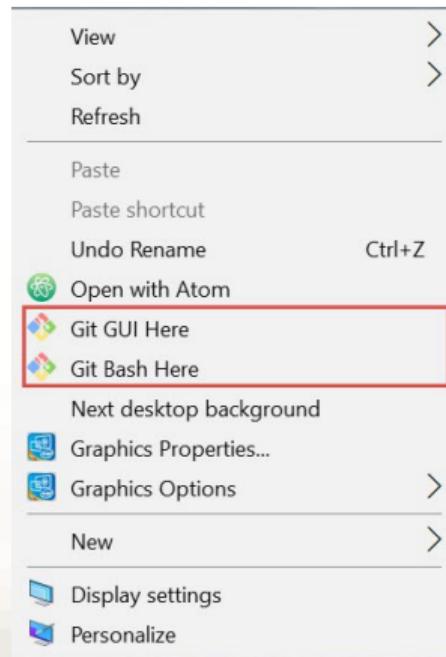
### Logos

Various Git logos in PNG (bitmap) and EPS (vector) formats are available for use in online and print projects.

[View Logos →](#)

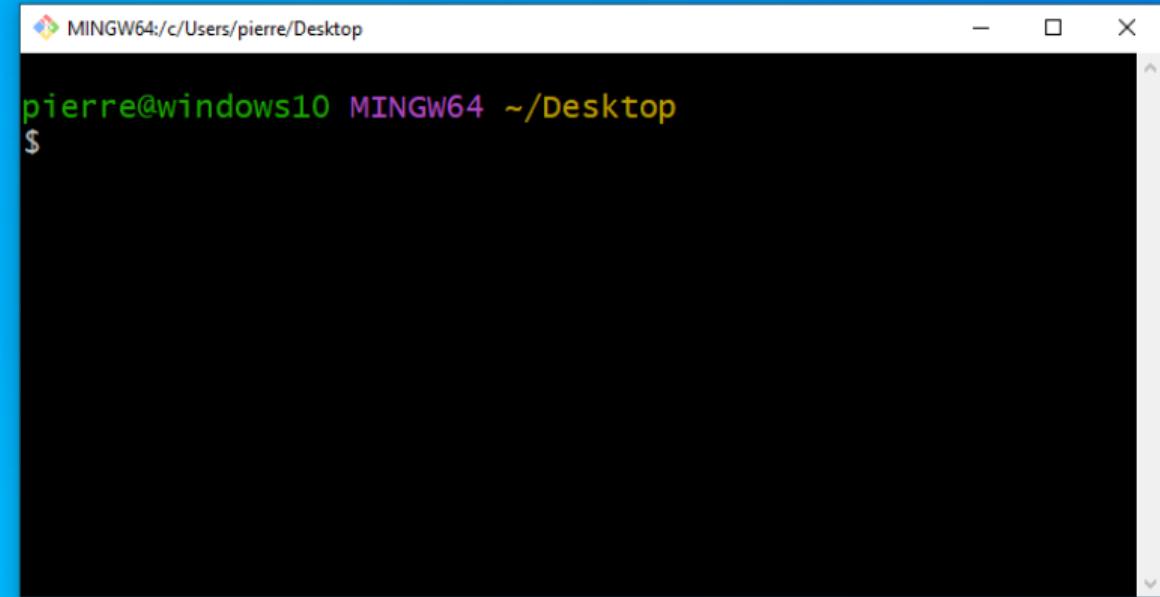
# Installing a bash terminal

Git bash for non unix user



# Installing a bash terminal

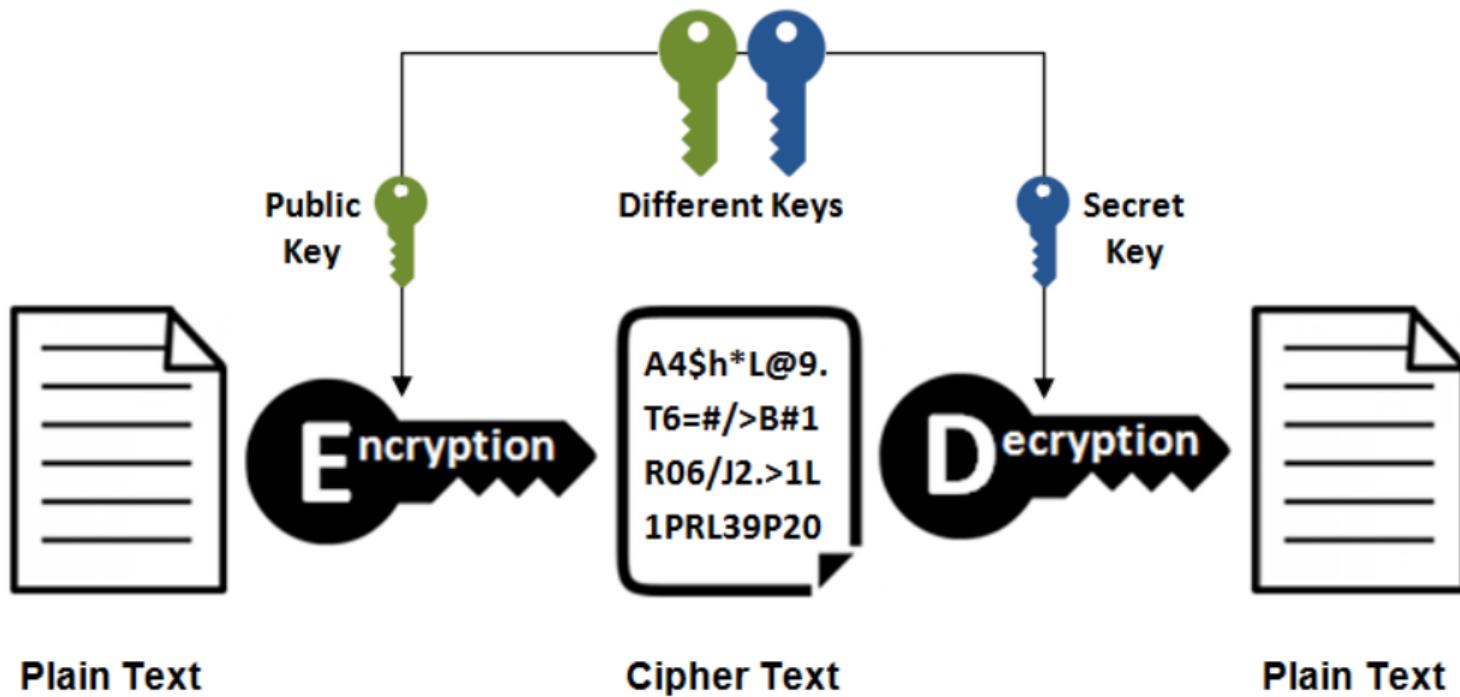
Git bash for non unix user



```
MINGW64:c/Users/pierre/Desktop
pierre@windows10 MINGW64 ~/Desktop
$
```

## SSH key use

## Asymmetric Encryption



# SSH key use

```
ssh-keygen -t dsa
```



# SSH key use

```
ssh-keygen -t dsa
```

```
$ ssh-keygen -t dsa
```

Generating public/private dsa key pair.

Enter file in which to save the key (/Users/tdd/.ssh/id\_dsa): /Users/tdd/.ssh/id\_dsa\_

Enter passphrase (empty for no passphrase):

Enter same passphrase again:

Your identification has been saved in /Users/tdd/.ssh/id\_dsa\_ga.

Your public key has been saved in /Users/tdd/.ssh/id\_dsa\_ga.pub.

The key fingerprint is:

65:31:7e:ee:49:3a:66:cd:92:7b:02:2b:bf:b3:1a:79 tdd@CodeMagic.local

The key's randomart image is:

+--[ DSA 1024]----+

```
 |          |
 |          o   |
 |         . o  |
 |         + .  |
 |         o o  |
 |           S  o |
 |             o  |
```

# SSH key use

```
cat /home/pierre/.ssh/id_dsa_ga
```

# SSH key use

```
cat /home/pierre/.ssh/id_dsa_ga
```

```
$ cat ~/.ssh/id_dsa_ga
-----BEGIN DSA PRIVATE KEY-----
Proc-Type: 4,ENCRYPTED
DEK-Info: DES-EDE3-CBC,6ED59B013D8A361F
pB5eHHpvXxoz6i1jFz1KANv9W6SeHw664PV/1A90acR/Mw/ERQvTQKo3TaLaFhkb
NwhQFyxZZty2hn6xrv5UIAGTpjk+P2+waRmSno1Vg1x0epCp45kvFRv9AiXsOpt4
RgzPC5+a6kjPf8EtyozGoQ==
-----END DSA PRIVATE KEY-----
```

```
cat /home/pierre/.ssh/id_dsa_ga.pub
```

```
$ cat ~/.ssh/id_dsa_ga.pub
ssh-dss AAAAB3NzaC1kc3MAAACBAJhbQcZK81FMvpw7trbFj51Sqjd9nKBu2xkw/kvUAQ1PQPaIRL0iq92fx
```