Open Science, « DATA&CODE, how to be FAIR ? »

Guichet de la donnée UP : A support service to improve the management of scientific data from our laboratories' research projects

https://scienceouverte.univ-poitiers.fr/

Research data : precious assets

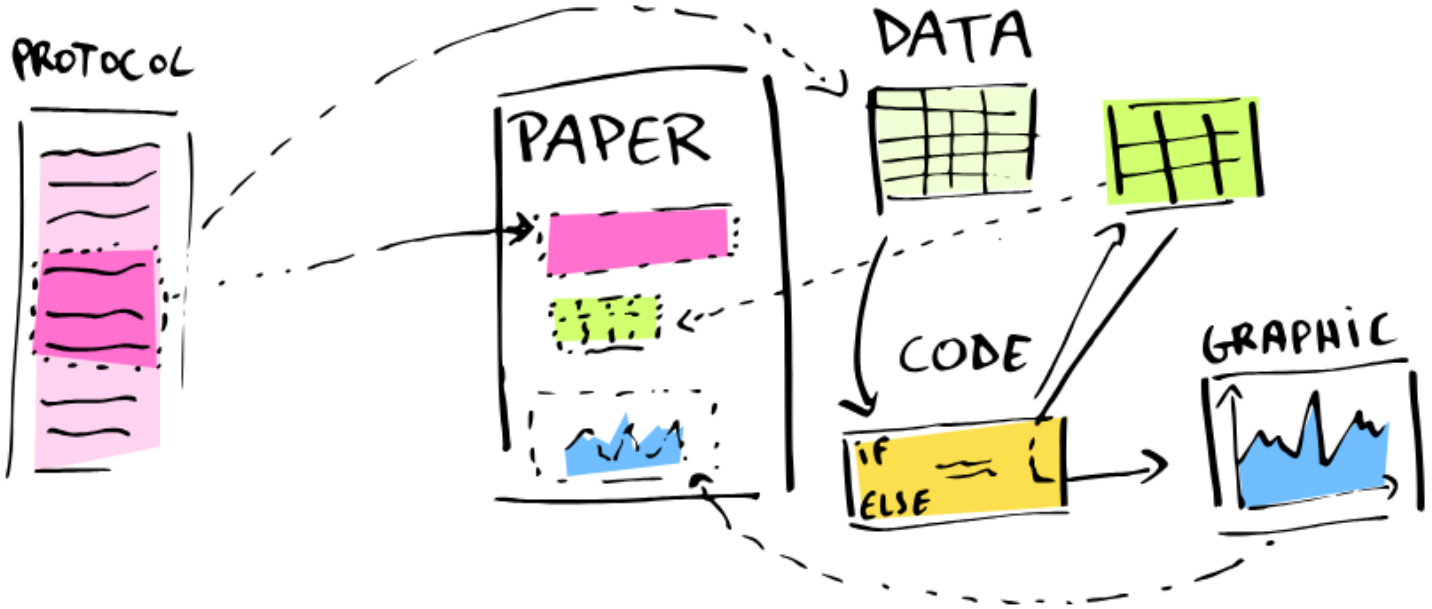
- Imperative for research
  - Testing hypotheses
  - Continuing previous work
- Production and exploitation
  - Costly : financially, in time
  - Complex : technical, organisational

- Accessibility of publication data

« publications provide access to around 10% of data, with the rest remaining available but unused on computer hard drives »

https://doranum.fr/enjeux-benefices/

Research data&code : precious assets



« The European Commission report ***Cost of not having FAIR research data*** *published in 2019 estimates that the cost of poor management of research data amounts to 3 billion euros for France, due to wasted time, failure to optimization of storage costs, licensing fees, research duplication issues, lack of cross-fertilization. »*

https://www.ouvrirlascience.fr/les-donnees-de-la-recherche-un-sujet-strategique/

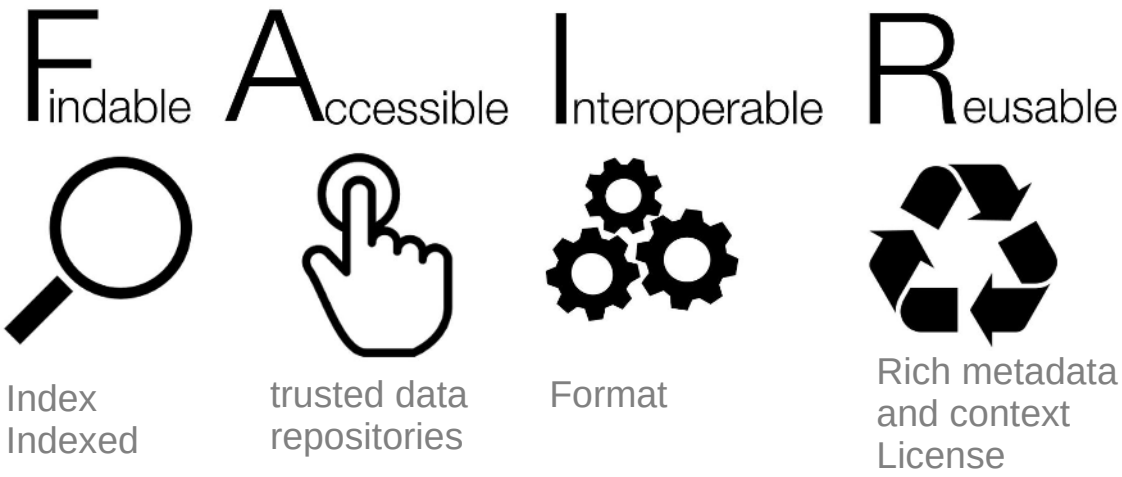
Which DATA ?

- Scientific data from research projects

  « factual records (numbers, spreadsheets, texts, images and sounds), used as primary sources for scientific research and are generally recognised by the scientific community as necessary to validate research results. »

  https://www.oecd.org/fr/science/inno/38500823.pdf

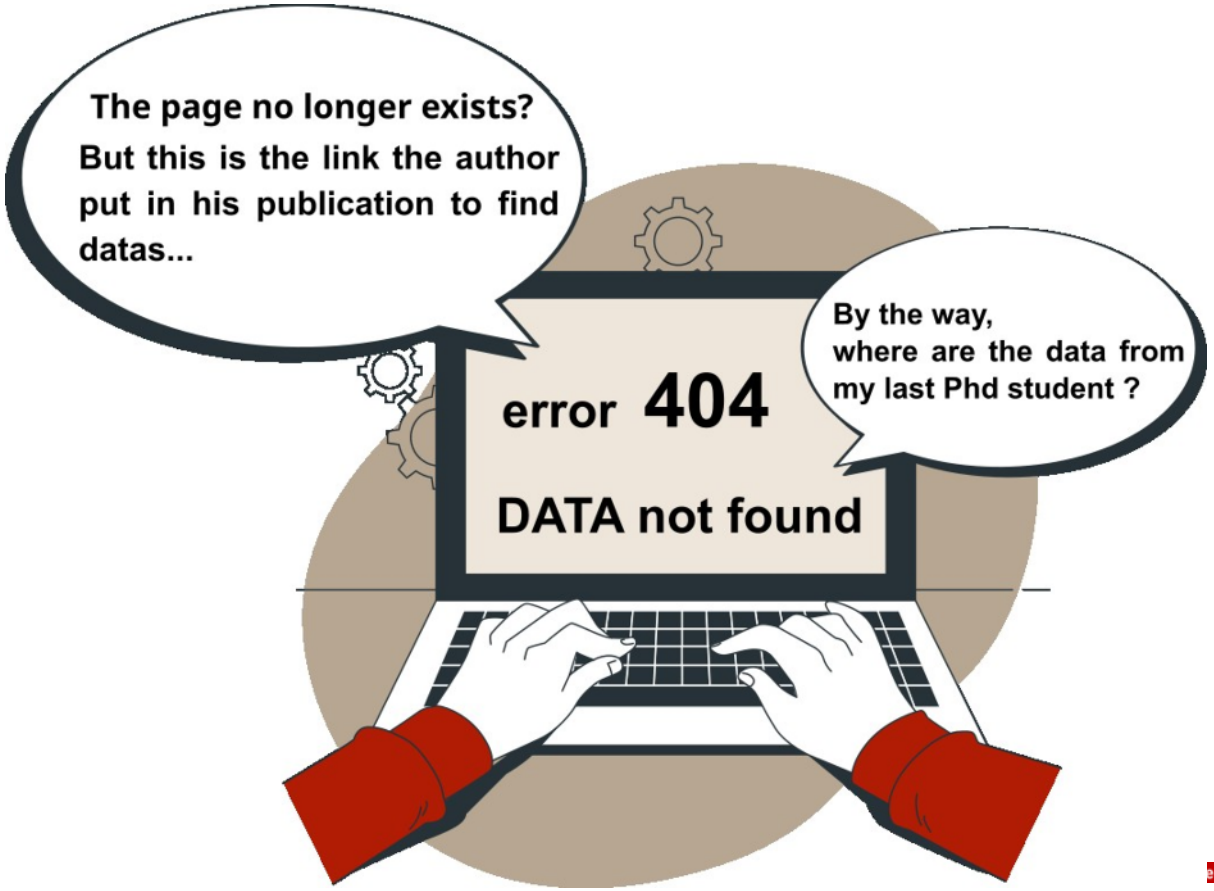
FAIR, what does it mean ?



https://www.go-fair.org/fair-principles/

Benefits of good data management

- Avoid losing them
- Comparing results
- Meta-analysis
- ...



Findable - Accessible

Solution

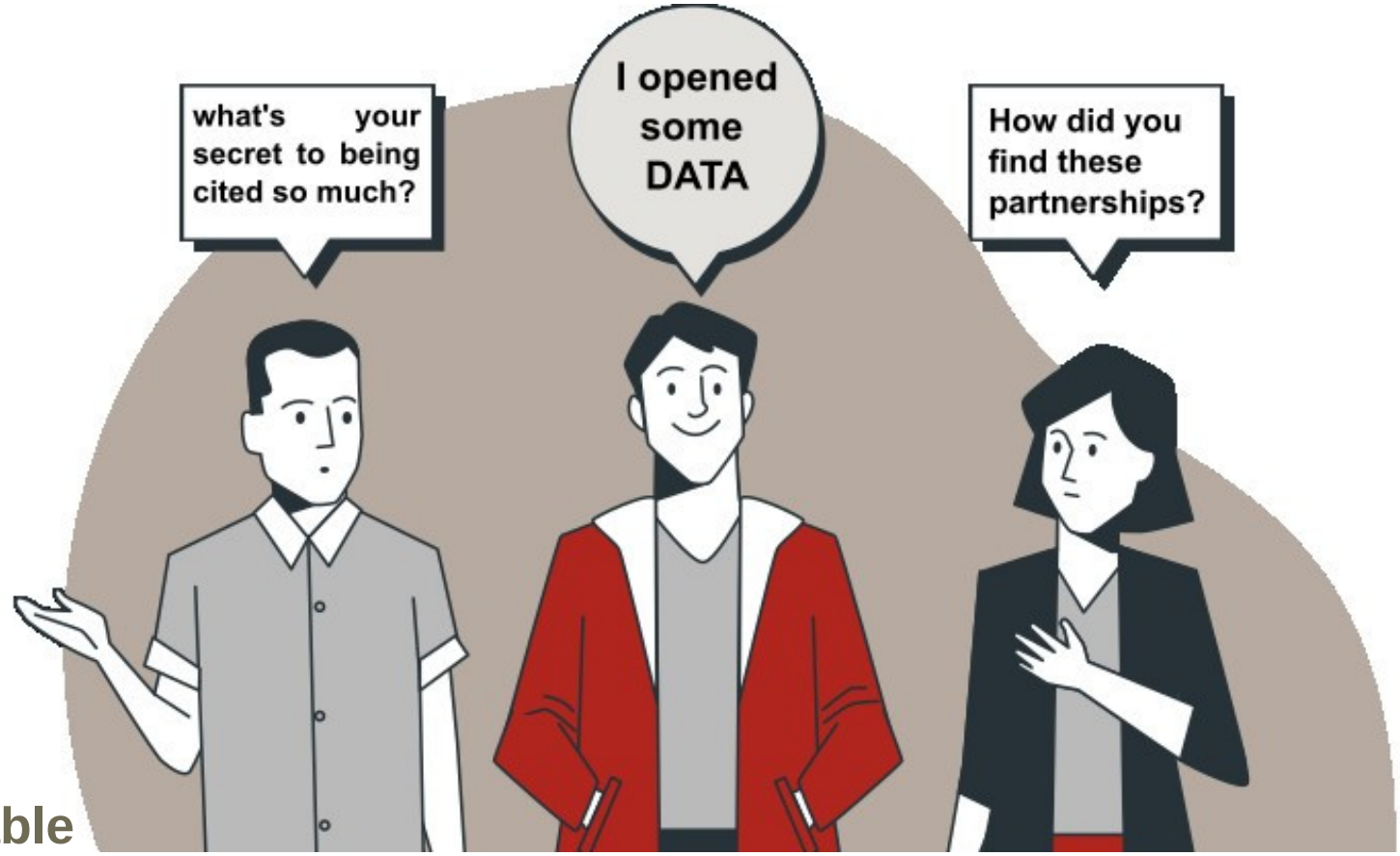
- Using trusted data repositories
  - France : Recherche Data Gouv
  - France (example SHS) : Nakala / TGIR Huma-Num
    - SHS-Linguistique : Ortolang et Cocoon
  - European : Zenodo
  - List of thematic repositories

Personal laptop and usb key are not enough

Findable - Accessible

Benefits of good data management

- Increase your chances of being cited (+25%)
- Encouraging collaboration



Reusable

Colavizza G & al. (2020) The citation advantage of linking publications to research data. PLOS ONE (https://doi.org/10.1371/journal.pone.0230416)

Benefits of good data management

- Easily meet the expectations of colleagues and publishers
- Avoid appropriation by certain publishers

Solution

- Data are released with a clear and accessible usage licence (etalab 2.0 / CC-BY 4.0)
- Data have persistent identifier (DOI)
- (Meta)data meet domain-relevant community standards
- Data are described with rich metadata and context

Do you have :
- metadata standards ?
- file-naming convention ?
- folder structures ?
- version control
- standardized units, values ?
- readme file ?

Filenames example

Case 1

Ma thèse v1.doc
Ma thèse v1_relu PR.doc
Ma thèse v2.doc
Thèse 02/01/2024 v2.doc
Thèse 03-01-2024 v3.doc
Thesis_final.docx
Thesis_final_corrrction-2.docx
Thesis_final_last_ok.docx
Thesis_final_last_ok_v7.odt
Thesis_correction_après_soutenance.odt
Thesis_finale-final-last_v42.pdf
Thesis_finale-final-last_v33_PDF-A.pdf

Case 2

MaThese_backup-01.tex
MaThese_backup-02.tex
MaThese_backup-03.tex
MaThese_backup-04.tex
MaThese_backup-05.tex
MaThese_backup-06.tex
MaThese_backup-07.tex
MaThese_current.tex
MaThese_current.pdf

Do you know GIT ?

Scientific data management: national strategy

French Plan for Open Science

" open data to encourage its re-use by researchers, teachers, citizens and businesses, so that it can create scientific and economic values "

La feuille de route 2021-2024 du MESRI relative à la politique des données, des algorithmes et des codes sources

Open Science : national strategy

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE

Ouvrir la science !

OPEN SCIENCE | OUR ACTIONS | RESOURCES | NEWS | WHO ARE WE ? | FR - EN

First Open Science Awards Ceremony for PhD Thesis

NEWS FROM THE COMMITTEE 07/11/2024

For the first time, the French Ministry of Higher Education and Research (MESR) has awarded the Open Science PhD thesis prizes. This prize complements the open science prizes already in place for research data and open access research software. As part of the second French plan for open science, these prizes aim to reward open science practices in research

The open science PhD thesis prize, in particular, aims to acknowledge open science practices among young researchers. It aims to encourage doctoral students to train in open science and to apply its principles in their research work. The prize rewards PhD theses that have already been defended and for which the practice of open science has contributed to the quality of scientific work.

The prizes were awarded during the French Doctoral Day, held in Lille on November 6, 2024, by Sylvie Pommier, president of the jury, professor at ENS Paris-Saclay, and by Marin Dacos, national coordinator of open science at the Ministry of Higher Education and Research.

https://www.ouvrirlascience.fr/first-open-science-awards-ceremony-for-phd-thesis/

Scientific data management : Poitiers University strategy

Roadmap for open science UP (2022)

- *1: Support open access publishing and distribution*
- *2: Contribute to the development of business models for scientific publishing*
- *3: Promote the efficient management and openness of research data*
- *4: Open up and promote the source codes produced by research*
- *5: Open science and evaluation*
- *6: Support and training in open science*

Data management : advice from funders

Example ANR

"The Agency draws the attention of coordinators to the importance of considering the issue of data management and sharing from the outset of the project."

"It calls for the development of a data management plan [...] a document that summarises the description and evolution of data sets, and prepares for the sharing, re-use and long-term preservation of data."

Data Management Plan (DMP)

An invitation to master the entire data life cycle of your research project

Questions :

- *Data production (plateforme, instrument, simulation, protocol)*
- *Data processing and analysis (tool, software, code)*
- *Storage and backup (data protection, sharing)*
- *Documentation and data quality (contextualization, fiabilization)*
- *Legal and ethical requirements (Ethics Committee, RGPD, PPST)*
- *Data sharing and long-term preservation (repository)*

Data Management Plan (DMP)

A tool for project management

Examples (ANR template) :

What data (for example the kind, formats, and volumes), will be collected or produced?
Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.

How will new data be collected or produced ?
Explain which methodologies or software will be used

What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?
Indicate which metadata will be provided to help others identify and discover the data.
Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
Indicate how the data will be organised during the project

Data Management Plan (DMP)

A tool for project management

Examples (ANR template) :

What data quality control measures will be used?

Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.

How will data and metadata be stored and backed up during the research?

Describe where the data will be stored and backed up during research activities and how often the backup will be performed.

How will data security and protection of sensitive data be taken care during the research ?

Explain how the data will be recovered in the event of an incident.
Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.
Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed

« Google drive » is not recommended

Main target

- **Researchers** : Need to efficiently write the Data Management Plan (DMP) for their funded projects (ANR, Europe, AAP UP)
  - PGD as a project management tool
    - Preventing and anticipating blocking problems
      - Storage space, sharing with project members, security, RGPD, time spent consolidating and distributing data...
    - Facilitating exchanges and collaboration
      - Minimize friction between all project members

Guichet de la donnée UP

A single gateway to a range of support services for efficient management of scientific data

- Coordinated experts network
  - DPO, integrity referent, SCD, SPVR, DSI, MSHS, PUD, DRInnov, labo. members
- From advisory services to concrete assistance
- Courses and events
- tools

Guichet de la donnée UP

Goals

Encourage and Facilitate the writing of Data Management Plans (DMP)

- Individualized advice and proofreading
- Simplify answers to all DMP questions by making recommendations from the establishment available on the Opidor DMP platform

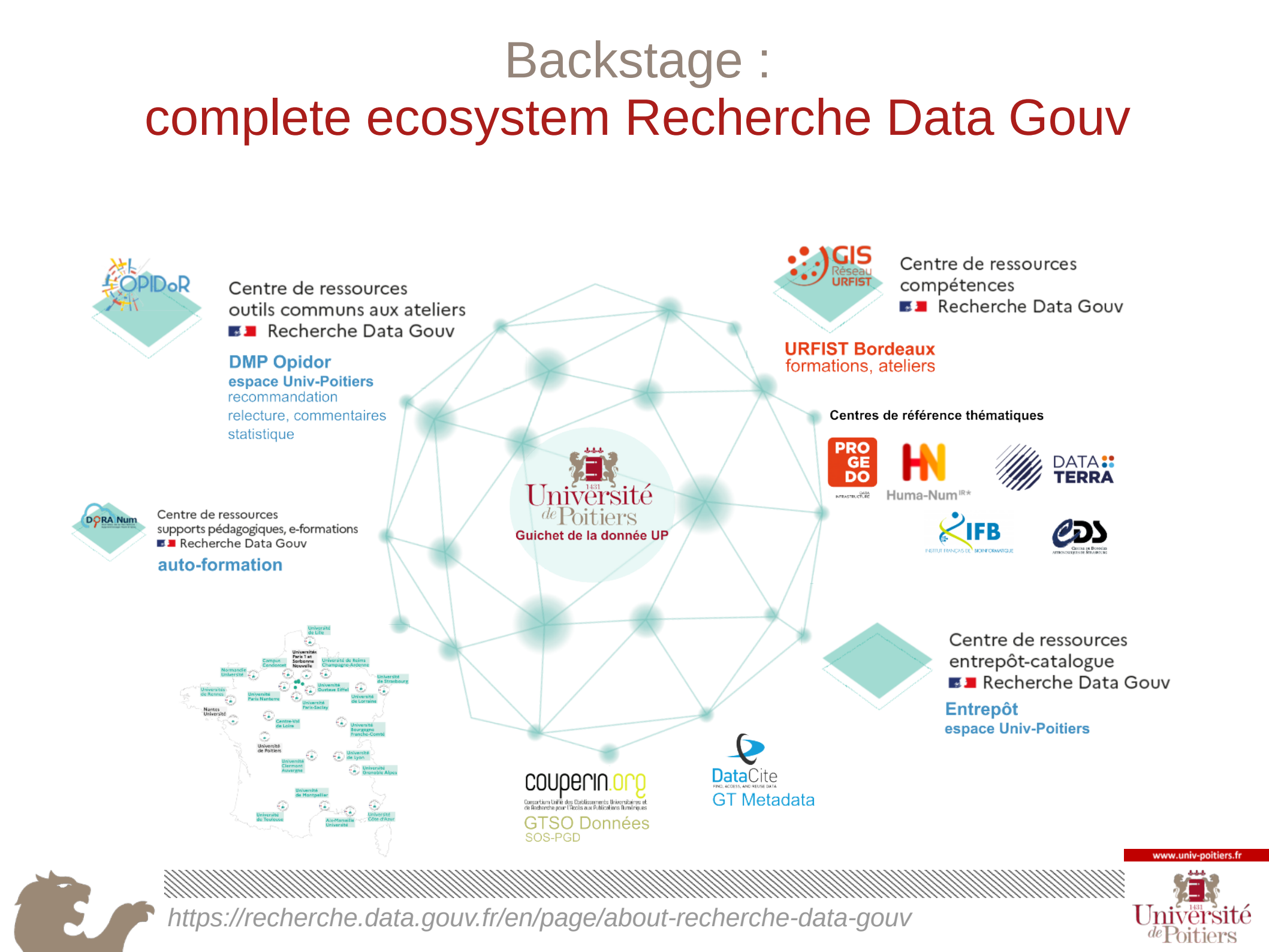
Supporting data FAIRization

- Help with selecting and depositing data in the appropriate data repository
- Help in selecting a data sharing and distribution license
- Advice on metadata, file formats & naming, data processing chains

Raising awareness and training in Open Science

- Courses and events

Backstage :
complete ecosystem Recherche Data Gouv



https://recherche.data.gouv.fr/en/page/about-recherche-data-gouv

What about research software / script ?
same ambition, other requirement

Dedicated page : https://scienceouverte.univ-poitiers.fr/piliercodeslogiciel/

Pilier Codes sources et logiciel

- Objectifs
- Le saviez-vous ?
- À lire sans modération
- Formation
- Plan de gestion de logiciels

Objectifs

L'objectif est de garantir que les **codes sources et logiciels produits par la recherche publique française soient développés, maintenus, préservés et valorisés de manière durable**.

L'université de Poitiers encourage sa communauté scientifique à participer à l'écosystème national qui relie les codes, les données et les publications, en suivant la trajectoire engagée par les recommandations du deuxième Plan National pour la Science Ouverte (2021) et celles de la Feuille de route du MESRI (2021-2024).

Conformément à sa propre feuille de route, l'université s'engage à apporter son soutien en vue de définir une politique d'ouverture des logiciels libres qu'elle produit, ainsi qu'à identifier la forge sur laquelle publier ses codes.

Le saviez-vous ?

Nous avons rédigé de la **documentation** sur :

- Je veux poser une licence sur mon code : comment faire ?
- L'aspect juridique des licences logicielles
- Les grandes catégories et compatibilités des licences entre-elles
- Je veux utiliser un moyen de calcul ou de stockage mutualisé : comment faire ?

Une **question sans réponse** dans cette documentation ?

- Contactez le service « **guichet de la donnée** »

Il est possible de **déposer le code source** de son logiciel dans **HAL & Software Heritage**.

- Comment déposer le code source d'un logiciel dans HAL https://doc.archives-ouvertes.fr/deposer/deposer-le-code-source/
- Un tutoriel en français et en anglais : https://hal.science/hal-01872189

« Open Source == more transparent and reproducible research + creating a wide community »
« If code has no license, it should only be read, not executed..." » B.Mercier (XLIM lab.)

PhD programs

Code

A3.4 - Licences logicielles - il faut une licence sur mon code ! [In-person training] (2026-02-05)

Research Data

A3 - Maîtriser la gestion des données de recherche [In-person training] (2026-01-08)
A3 - Maîtriser la gestion des données de recherche [Remote learning] (2026-04-28)

A3.1 - Introduction au RGPD [Remote learningl] (2026-01-12)

A3.2 - Introduction to Research Data and Data Management Plan [Remote learning] (25-03-11)

Interfaces

website **scienceouverte.univ-poitiers.fr**

En Bref

La Science ouverte (SO) est la diffusion sans entrave des résultats, des méthodes et des « produits » de la recherche scientifique. Elle vise donc à **développer l'accès** ouvert, sans barrière technique ou juridique, des **publications scientifiques** et, autant que possible, aux **données** et aux **codes sources**.

L'université de Poitiers est activement engagée dans ce mouvement international de démocratisation des accès aux savoirs et du partage.

Elle accompagne l'ensemble de sa communauté scientifique dans cette démarche et développe une politique ambitieuse et équilibrée de science ouverte dans le respect des spécificités disciplinaires et des usages des communautés.

A travers son Guichet de la Donnée UP, l'université propose un service d'accompagnement facilitant et améliorant la gestion des données scientifiques des projets de recherches de nos laboratoires.

DÉCOUVRIR LE GUICHET DE LA DONNÉE UP

Offre d'Accompagnement

Un réseau d'interlocuteurs de proximité et coordonnés au profit de la réussite de votre projet de recherche.

Liste de raccourcis vers les services les plus utilisés

RÉDIGER EFFICACEMENT SON PGD — ACCÉDER AU SERVICE

DÉPOSER FACILEMENT SOUS HAL — EN SAVOIR PLUS

DEMANDER UN CONSEIL — CONTACTER LE GUICHET

PILIER PUBLICATION — DÉCOUVRIR

PILIER DONNÉES — DÉCOUVRIR

PILIER CODE SOURCE & LOGICIEL — DÉCOUVRIR

PILIER EVALUATION — DÉCOUVRIR

PILIER FORMATION — DÉCOUVRIR

mail
donnees-recherche@univ-poitiers.fr

TEAM / Network
35+ members UP

Thanks

Behind the Guichet Unique door, a forest of expertise