

开放科学与科研数据管理初探

Introduction to Open Science & research data management

Zhengdong SHI
Research data & grant writing (SO, DMP) service
Directorate of Libraries, Information and Open
Science (DiBISO)
11/10/2022

提纲

- 什么是开放科学？(open access 开放获取)
- 科研数据的定义与分类
- 科研数据生命周期
- 科研数据共享FAIR原则 (FAIR principles)
- 科研数据管理计划 (data management plan)
- 科研数据知识库：科研数据的发表，查找与再利用

开放科学与科研数据管理初探： 一场促进科学发现的全球运动

什么是开放科学？

“开放科学指的是通过最新的数字进步技术来发展对出版物，还有尽可能地对数据、源代码和科研方法的开放访问，从而无阻碍地传播科研成果、方法和产品。它是公共资助机构管理研究项目成果一种方法与措施。”

法国开放科学计划

Second French Plan for Open Science (2021-2024)



Image : <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

开放科学的关键要素图（联合国教科文组织，2021年）

“COVID-19大流行病使人们关注开放科学的实践做法，如开放科学出版物、共享科研数据以及加强科研工作者与科学界以外人员的合作，提高研究效率并增强科研政策机构与社会之间的交流”。

Audrey Azoulay,
UNESCO Director-General

www.unesco.org/en/articles/unesco-sets-ambitious-international-standards-open-science

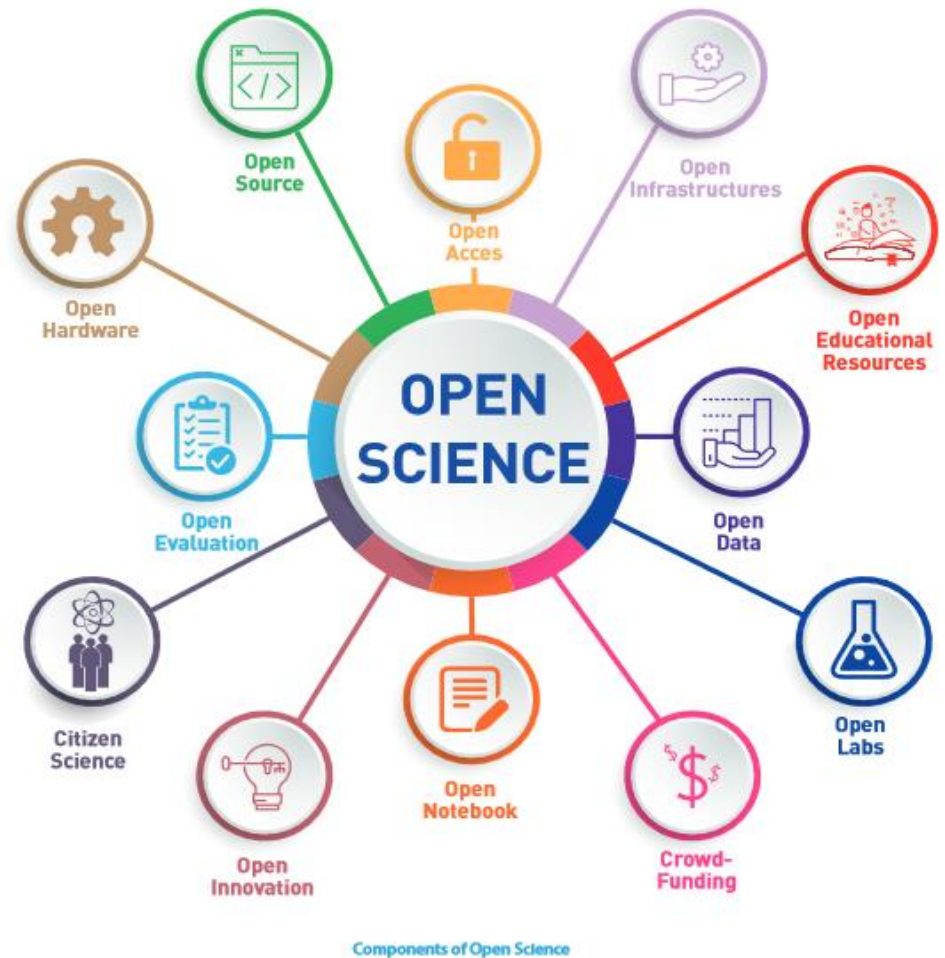


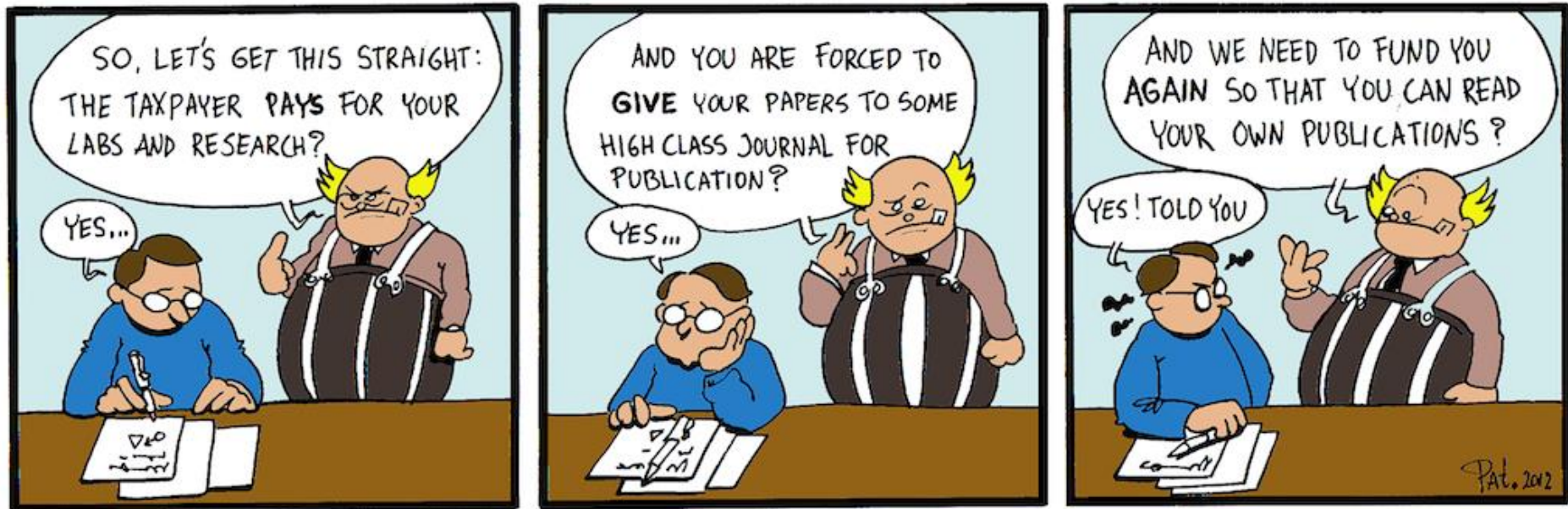
Image : Recommendation de l'UNESCO sur une science ouverte,
<https://fr.unesco.org/news/lunesco-etablit-normes-internationales-ambitieuses-science-ouverte>



Open Access 开放获取

为什么要开放获取？

Why we need open access?



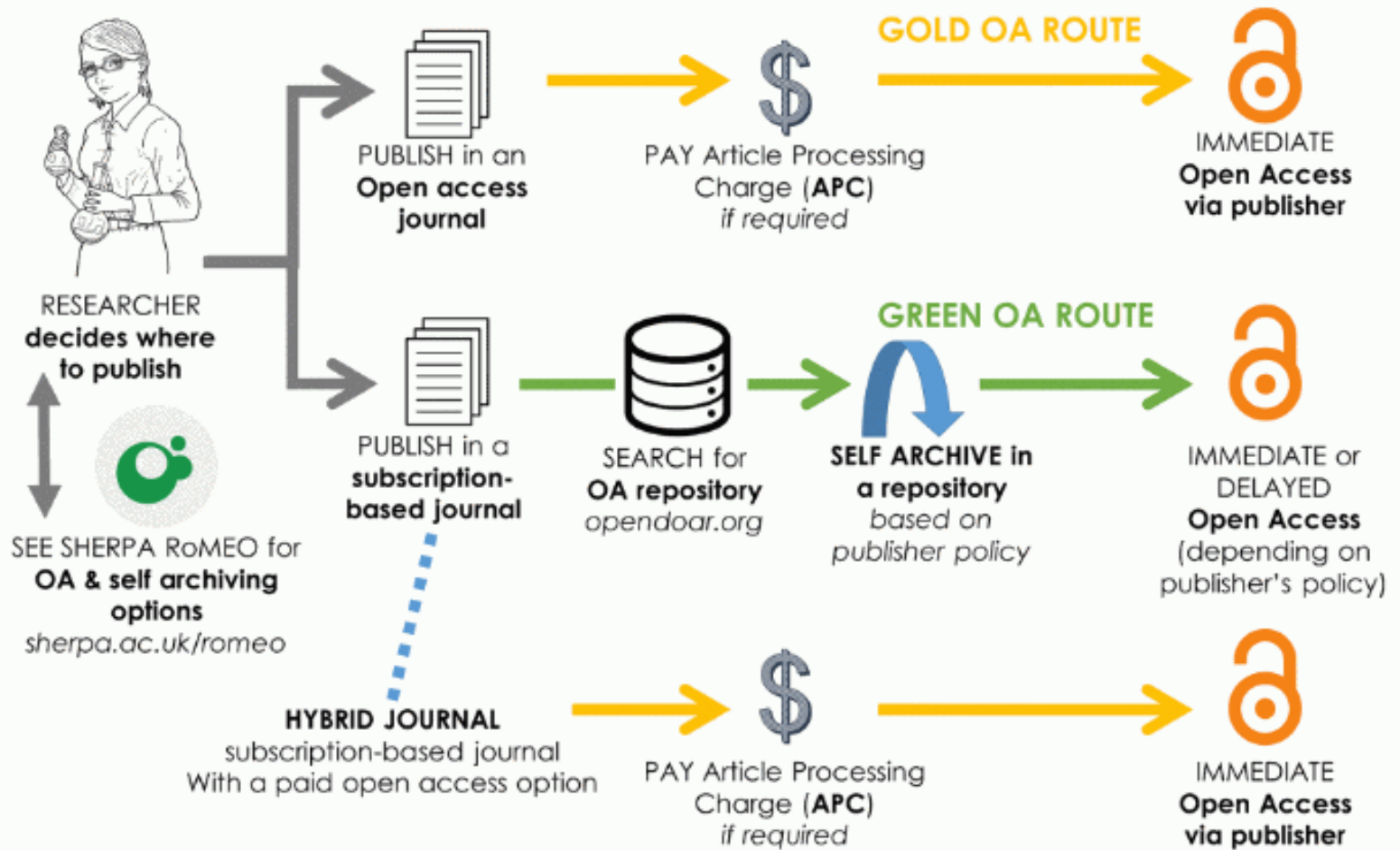
@by Patrick Hochstenbach
CC-BY

出版物开放获取 (Open access) 的益处



CC-BY Danny Kingsley & Sarah Brown

Open Access Publishing



出版物开放获取的分类

- 金色开放获取 (Gold Open Access 或者 voie dorée) : 开放期刊 (Open Access Journal)。比如 : Elsevier (爱思唯尔) 一些OA 期刊。
- 绿色开放获取 (Green Open Access 或者voie verte) : 自我存档 (self-archiving 或者auto-archivage)。 比如 : ArXiv, HAL档案库。
- 查找开放获取期刊, 就上DOAJ (Directory of Open Access Journals) : <https://doaj.org/>
- SHERPA/RoMEO (<https://v2.sherpa.ac.uk/romeo/>) : 一个在线资源平台, 汇总和分析世界各地的出版商开放存取政策。

arXiv.org

 DOAJ

Sherpa Romeo

université
PARIS-SACLAY

法国《数字共和国法》(2016)

Loi pour une République numérique (2016)

Selon l'article 30 :

vous avez le droit de mettre en ligne vos articles scientifiques, mais après un embargo (sciences dures 6 mois, SHS 12 mois)

根据法案第30条：

你有权将你的科学文章发表在网上，只要在**延迟期 (embargo)** 满后（自然科学6个月，社会科学12个月）。



Image : https://www.economie.gouv.fr/files/files/PDF/DP_LoiNumerique.pdf

注意

- 小心那些“掠夺性”出版商 (revues prédatrices, predatory journal)。

« 这类学术出版商是为了学术灌水或以此牟利而组织发表大量低品质的学术会议和期刊 »。

核实开放期刊是否为值得信赖期刊，谨防“掠夺性”期刊：

<https://thinkchecksubmit.org/>  **THINK**  **CHECK**  **SUBMIT**

资料来源： <https://www.wikiwand.com/zh-cn/%E6%8E%A0%E5%A5%AA%E6%80%A7%E9%96%8B%E6%94%BE%E7%8D%B2%E5%8F%96%E5%87%BA%E7%89%88>

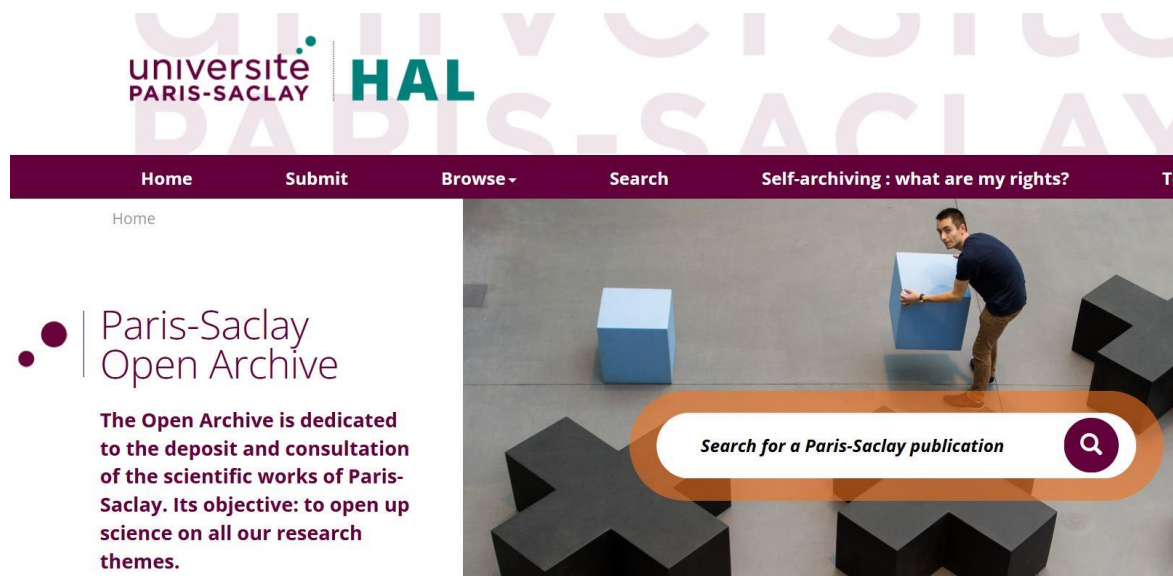
法国多学科开放档案库 HAL 数据平台

<https://hal.archives-ouvertes.fr/>



- Le portail HAL Paris-Saclay
巴黎-萨克雷大学开放档案库

<https://hal.universite-paris-saclay.fr/>



博士
论文
和科
研文
章可
以放
里面



开放研究数据： 一个开放科学的重要议题

Image: <https://fr.freepik.com/photos-vecteurs-libre/affaires>

为什么拥有一个良好的科学数据管理方式是如此重要？

- 提高研究效率，为你的研究工作节省时间
- 方便同事查找、理解你的数据，以及科研交流
- 提高科研透明度，以及验证和生产科研成果的能力。
- 实现了科研数据的再利用、开放与共享
- 进入科研机构 and 科研公司工作所需的必备知识和技能



Image from the Aalto University's webinar series on "Research data management".

科研数据的再利用只是冰山一角



Tip of the iceberg: Independent reviewers of clinical trial data have access to just a minuscule percentage of the actual information. *The Scientist*, 2012.

- 仅有10%的科研数据被公开
- 90%的科研数据储存在硬盘上，无法被再利用

法国开放科学计划

Second French Plan for Open Science (2021-2024)

目标：2021-2024年间，在法国普及开放科学

4 个主要内容：

- 推广出版物的开放获取
- 整理、共享和开放科研数据
- 开放和促进使用由科学研究产生的源代码
- 使开放科学实践成为默认准则



Image : <https://www.ouvri-lascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

法国和欧盟的研究资助机构，要求研究数据的开放使用

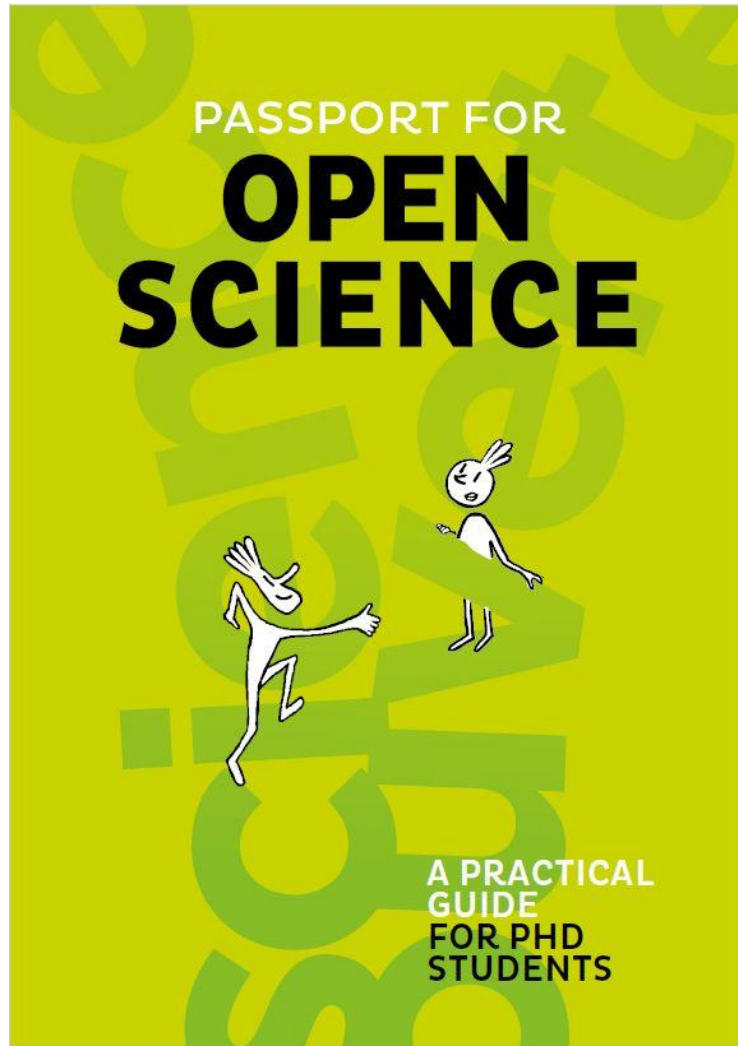


国家科学研究中心



- 将出版物存储在开放档案库的义务 (e.g. HAL 法国多学科开放档案库)
- 草拟数据管理计划的义务 (DMP)
- 鼓励开放研究数据
- “早期共享知识，数据，以及开放研究合作模式”，Horizon Europe (2021-2027)
- 数据“尽可能地开放，并在必要时封闭” (Data should be “as open as possible and as closed as necessary”)

博士教育课程中的开放科学实践指南



Example of the University of Lille, Ouvrir la science, MESRI
(Ministry of Higher Education, Research and Innovation)

Index

1. Planning an open approach to scientific work

- Using freely accessible resources p. 6
- Planning data management p. 8
- Working in a reproducible way:
For yourself, for others p. 11

2. Disseminating research

- Disseminating your publications in open access p. 16
- Making your thesis freely accessible p. 21
- Making research data open p. 25

3. Preparing for after your thesis, join the movement

- Deeply rooted public policies p. 30
- Evaluating research differently p. 32

Act now p. 34

Going further p. 35

Glossary p. 36

Sources p. 38

Legend

Underlined text is explained in the glossary.

This triangle ▼ refers to tools which are given as examples.

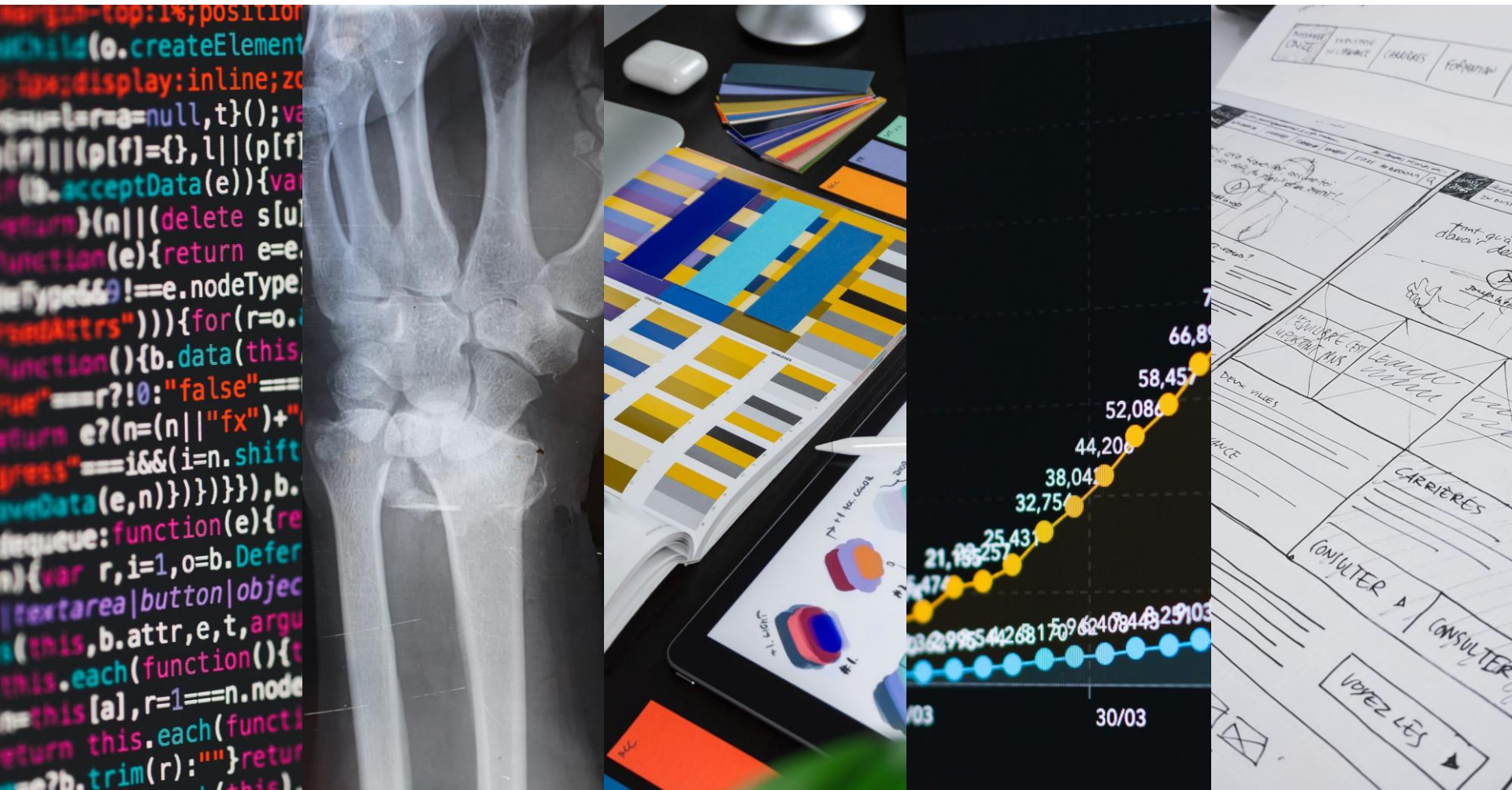
Image sous licence CC BY SA : https://www.ouvrirlascience.fr/wp-content/uploads/2020/11/Passport-for-Open-Science-A-Practical-Guide-For-PhD-Students_30-10-2020_WEB.pdf



科研数据的定义

<https://fr.freepik.com/photos-vecteurs-libre/donnees>

科研数据的定义与分类



Code, medical data, drawings, statistical models, production processes...

代码、医疗数据、图形、统计模型、生产流程...

Image from the Aalto University's webinar series on "Research data management".

科研数据

是为验证研究结果过程中收集、观察、生成或创建的任何信息。

Research data

is any information that has been collected, observed, generated or created to validate research findings.

科研数据的分类

性质或形式 Nature or form

- 文本 (问卷、采访记录、程序、协议、记录本, 等等。) questionnaires, interview transcripts, procedures, protocols, logbooks)
- 电子记录 (测试或调查的答案、医疗数据、测量和记录。 answers to tests or surveys, medical data, measurements and records)
- 多媒体 (声音、视频、图像。 sounds, videos, images), 开发代码 development codes (源代码, 算法, 脚本。 source codes, algorithms, scripts)

管理程度 Level of elaboration

- 原始数据 (收集的数据, 但没有结构化和格式化) 。 **Primary** data (data collected, but not organised and formatted)
- 经过处理的数据 (它已经被程序化, 结构化, 准备好进行分析) 。 Processed data (it has been organised, structured to be ready for analysis) ;
- 经分析的数据 (如出版物中出现的) 。 **Analysed** data (as they appear in publications).

数据 “ 尽可能地开放， 并在必要时封闭 ”

(‘As open as possible, as closed as necessary’)

-  • 非匿名的个人数据
-  • 不属于自己的数据
-  • 正在申请专利的数据
-  • 传播会造成安全问题的数据



研究数据的生命周期

Data life cycle model (UK Data Archive)

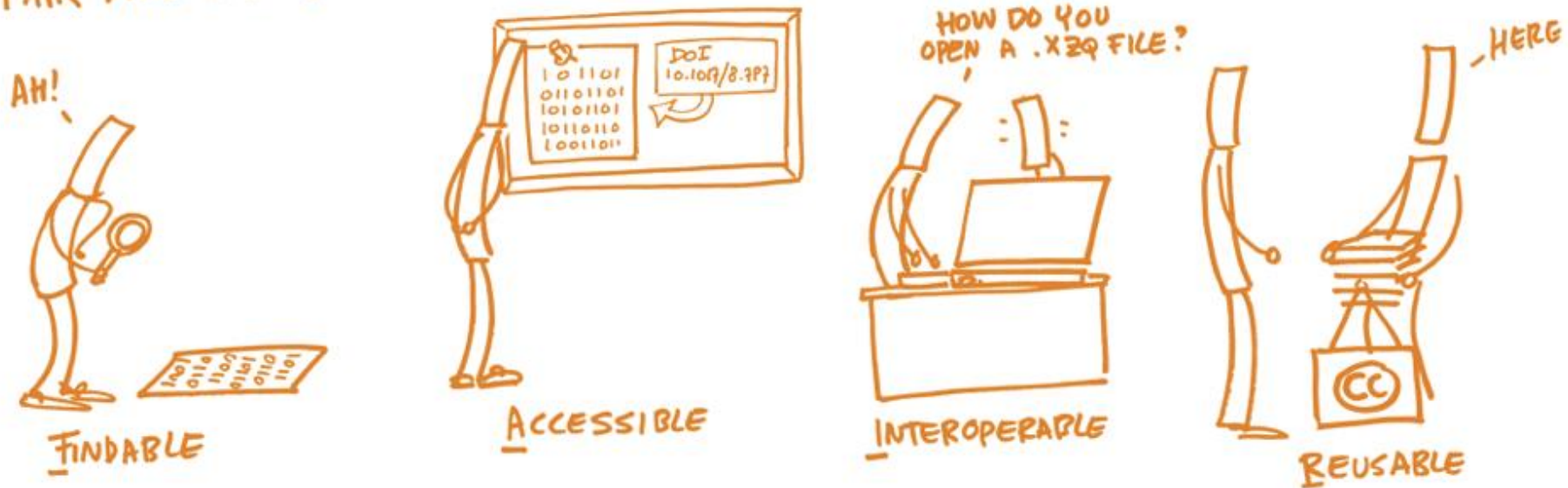
Source: https://www.researchgate.net/figure/Data-life-cycle-model-UK-Data-Archive_fig1_333466327

什么是科研数据共享FAIR原则？What is FAIR data?

RESEARCH DATA – OPEN BY DEFAULT



FAIR DATA PRINCIPLES



- 可发现 Findable = 其他人可以找到你的科研数据
- 可访问 Accessible = 你的科研数据可以提供给他人使用
- 可互操作 Interoperable = 你的数据可以与其他数据整合，或者可以被人和机器轻松使用和读取
- 可再利用 Reusable = 你的数据可用于新的研究

FAIR 原则



Source: Australian National Data Service (ANDS), https://www.ands.org.au/working-with-data/fairdata/training?utm_source=home-page-carousel&utm_medium=Web-banner&utm_campaign=fair-data-training

小指南：

How FAIR are your data?

怎样让你的数据变成FAIR数据？

https://www.cessda.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf

How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable

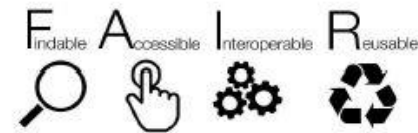
Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

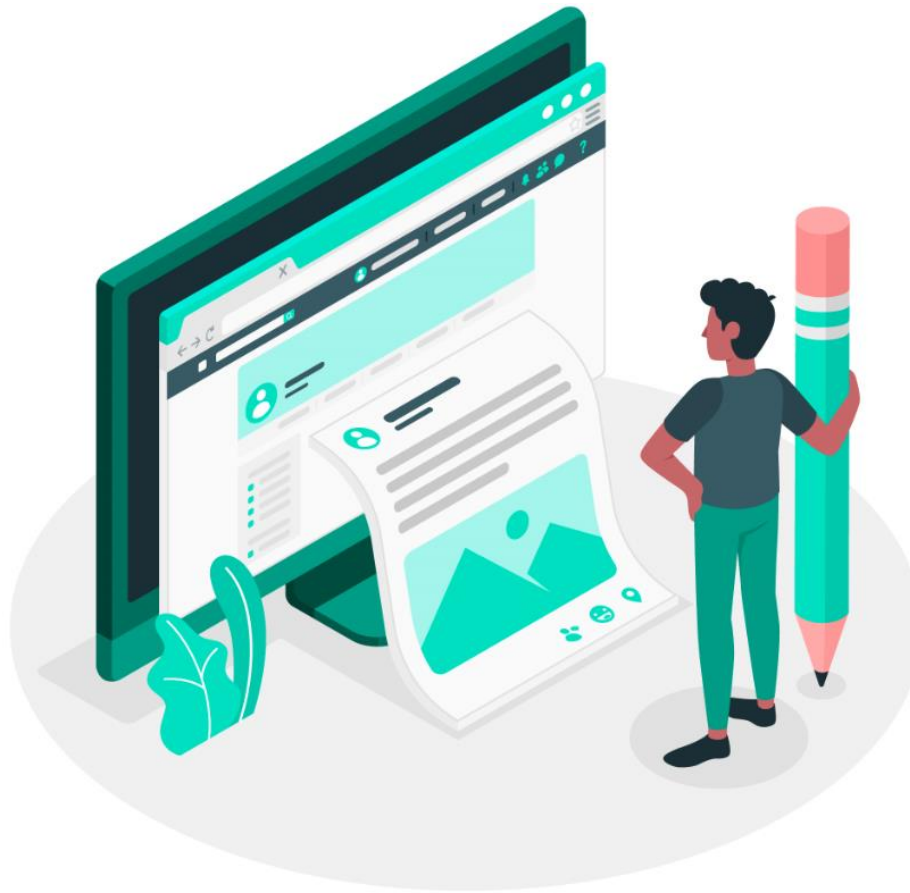
Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards



'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, [EUDAT](#). Image CC-BY-SA by [SangeyaPundir](#)

科研数据管理计划 Data Management Plan (DMP)



科研数据管理计划

Data Management Plans (DMPs)

科研数据管理计划（DMPs）是一份书面文件，它描述了您在研究项目过程中预期获得或产生的数据，您将如何管理、描述、分析和存储这些数据，以及您在项目结束时将使用何种机制来分享和保存您的数据。

A data management plans (DMPs) is a written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.

Definition from Stanford Libraries: <https://library.stanford.edu/research/data-management-services/data-management-plans>

科研数据管理计划，到底是什么？

The DMP: what is it?

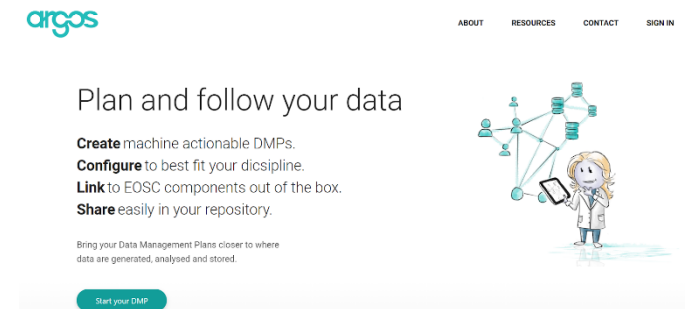
- 描述你将如何收集、组织、分析、保存和分享数据。
- 一份规划性文件，记录数据的产生，以及你将如何处理和管理这些数据。
- 创建DMP是任何使用或产生数据的研究项目人员的良好做法。是实现正确研究数据管理的第一步。
- 该计划帮助你系统地管理你的数据，它将节省你的时间。
- 一份“活的”文件，你可能需要随着你的研究过程的进展而不断地记录科研数据。
- 资助者需要计划文件。请提前查看他们的资助指导方针。

它通常涉及以下主题：

1. 哪些数据将被收集，以及如何被收集。
2. 数据将如何被记录。
3. 如何处理任何伦理和法律问题。
4. 在研究过程中，数据将如何被储存和备份。
5. 任何在项目结束后保存（部分）数据的计划。
6. 任何分享或提供访问（部分）数据的计划。
7. 谁将负责数据管理以及可能需要的相关额外资源。

DMP 撰写工具 - Writing tools

- Viewing DMPs :
 - [DMP Templates \(dmptool.org\)](https://dmptool.org)
 - https://dmponline.dcc.ac.uk/public_plans
 - Argos - OpenAIRE's "machine-actionable" tool for creating Data Management Plans (<https://argos.openaire.eu/home>)
 - [DMP OPIDoR \(opidor.fr\)](https://opidor.fr)



OPIDoR DMP: 主页 the homepage



[Public DMPs](#) [DMP Templates](#) [Help](#) [More ▾](#)

[English\(UK\)](#) ▾

✖ Suite à un changement de configuration de l'Accès institutionnel, nous avons été contraint, pour l'ensemble des comptes DMPOPIDoR, de **supprimer le lien avec les comptes institutionnels**.

Suite à cette opération, vous pourrez relier à nouveau votre compte institutionnel à votre compte DMP OPIDoR en vous connectant avec votre mot de passe, puis en cliquant sur "Lier vos identifiants institutionnels" depuis votre **Profil**.

Nous nous excusons pour la gêne occasionnée.

Welcome!

DMP OPIDoR guides you through the drafting and implementation into practice of data or software management plans.



Accessible by any member of the High Education and Research community as well as its French or foreign partners



Customizable by any research organization willing to implement a data policy



Enriched by examples and recommendations relevant to research environments



Collaborative: it facilitates exchanges between partners of a same project and support staff

Thanks to your feedback DMP OPIDoR will continue to evolve. The development work is a combined international effort around DMP Roadmap open source software.

[Join the DMP OPIDoR user's community](#)

Create an account, sign in and let you guide!

[Discover DMP OPIDoR](#)

[Sign in](#)

[Create account](#)

Institutional Access

[Sign in](#)

Individual Access

* Email

* Password

[Forgot password?](#)

☐ Remember me

[Sign in](#)

OPIDoR DMP: 建立一个DMP

[My Dashboard](#)[Create plans](#)[Public DMPs](#)[DMP Templates](#)[Help](#)[More ▾](#)[English\(UK\) ▾](#)[Stela Suhan ▾](#)

Université Paris-Saclay

✖ Suite à un changement de configuration de l'Accès institutionnel, nous avons été contraint, pour l'ensemble des comptes DMP OPIDoR, de **supprimer le lien avec les comptes institutionnels**.

Suite à cette opération, vous pourrez relier à nouveau votre compte institutionnel à votre compte DMP OPIDoR en vous connectant avec votre mot de passe, puis en cliquant sur "Lier vos identifiants institutionnels" depuis votre **Profil**.

Nous nous excusons pour la gêne occasionnée.

Project title

Choose a template

You can either choose a template provided by your organisation, another organisation or a funder template. The default template is **Science Europe : modèle structuré**.

[Find the list of the available templates](#)

Université Paris-Saclay (Your organisation)

Other organisation

Funder

No template available.

[Create a plan](#)[Next](#)[Use Default Template](#)

DMP OPIDoR : 科研项目信息

Projet Test

Renseignements sur le projet

Produits de recherche

Modèle choisi

Rédiger

Partager

Télécharger

* Titre du projet

Projet Test

☐ projet de test, d'entraînement ou à des fins de formation

Financier

Agence nationale de la recherche (ANR)

Numéro de subvention

Résumé du projet

B *I*    

Identifiant

Chercheur Principal

Nom

Mireille Brenel

Identifiant ORCID

Courriel

mireille.brenel@universite-paris-saclay.fr

Téléphone

Contact Pour Les Données

☒ Identique au Chercheur Principal

Sélection des recommandations du plan

Pour vous aider à rédiger votre plan, DMP OPIDoR peut vous proposer des recommandations provenant de différents organismes.

Choisir au maximum 6 organismes dont vous souhaitez afficher les recommandations.

Trouver les recommandations d'autres organismes ci-dessous

[Voir la liste complète](#)

Enregistrer



科研数据知识库 Data repositories

Image Source: https://fr.freepik.com/vecteurs-libre/gros-centre-donnees-rack-salle-serveurs-processus-ingenierie-travail-equipe-technologie-informatique-stockage-nuage_4103163.htm

如何查找数据库，将你的科研数据公布，存放在数据识库中

Deposit your data in a repository

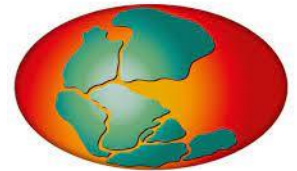
Generalist/Disciplinary?



DRYAD



- Find a repository: [re3data](#), [FAIRsharing](#)



PANGAEA.

- Licenses attribution
 - 知识共享许可协议Creative Commons (CC-BY), (<https://creativecommons.org/licenses/?lang=zh>)
 - Choose a licence
- Assign a PID = Persistent identifiers (永久标识符)
- DOI数字对象（唯一）标识符 for data and dataset/ORCID 开放的研究员和贡献者ID for author



ORCID

Connecting Research
and Researchers

université
PARIS-SACLAY

Recherche.data.gouv.fr : 法国国家科研数据库



recherche.data.gouv.fr

[Submit / Publish a dataset](#)

[About](#) [The Ecosystem members](#) [Data](#) [Online help](#) [News](#)

An ecosystem for sharing and opening research data

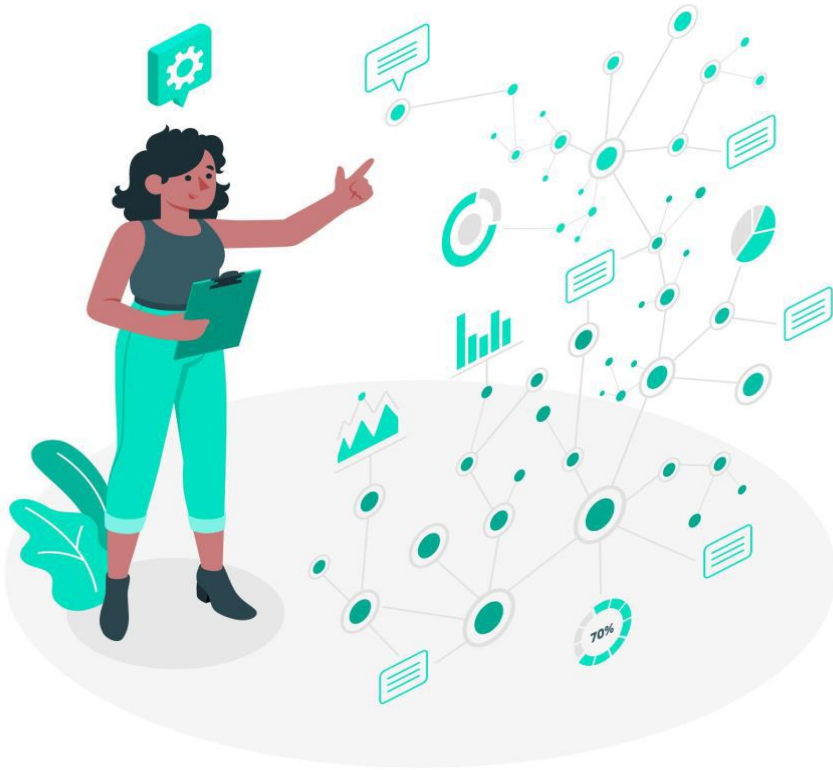
FEDERATE, SUPPORT, SHARE, OPEN, REUSE

[Find a dataset](#)

[Find data](#)

[Submit / Publish a dataset](#)

<https://recherche.data.gouv.fr/fr>



更多信息...

<https://fr.freepik.com/photos-vecteurs-libre/graphique>

- 图书馆网页 Université Paris Saclay
[Research data | Université Paris-Saclay \(universite-paris-saclay.fr\)](https://universite-paris-saclay.fr/research-data)
- 培训课程 Parcours interactif de DoRANum
<https://doranum.fr/enjeux-benefices/parcours-interactif-sur-la-gestion-des-donnees-de-la-recherche/>
- 科研数据生命周期 Cycle de vie des données
<https://doranum.fr/enjeux-benefices/le-cycle-de-vie-des-donnees-de-recherche/>
- 发表科研数据 Diffuser ses données
<https://www.universite-paris-saclay.fr/recherche/science-ouverte/les-donnees-de-la-recherche/ouvrir-ses-donnees-comment-ca-marche>
- 科研数据知识库 Entrepôts
<https://doranum.fr/depot-entrepots/choix-entrepot-depot-donnees/>
- 元数据 Métadonnées
<https://doranum.fr/metadonnees-standards-formats/>
- 科研数据的储存与存档 Stocker/archiver ses données
<https://doranum.fr/stockage-archivage/>

联系我们 Contact us!

图书馆提供关于科研数据的服务：

帮助科研数据的管理与传播

- Support for data management and data dissemination

支持FAIR 数据指导原则

- Support for FAIR Data Guiding Principles

数据管理计划审读

- Data management plan relecture

donnees-recherche@universite-paris-saclay.fr