

FORMATION

Gestion des données de la recherche

*

Contexte Science ouverte

26 octobre 2021, Dakar

Pascal Aventurier – IRD pascal.aventurier@ird.fr
Sophie Fortuno – CIRAD sophie.fortuno@cirad.fr



Ce(tte) œuvre est mise à disposition selon les termes de la [Licence Creative Commons Attribution - Partage dans les Mêmes Conditions 4.0 International](https://creativecommons.org/licenses/by-sa/4.0/).

Plan

Session de présentation

Introduction

1. Qu'est-ce que la Science ouverte ?
2. Introduction à l'ouverture des publications | open access
3. Introduction à l'ouverture des données de recherche | open research data
4. Introduction à l'ouverture du code source | open source

Cycle de vie de la donnée

1. Exposé
2. Atelier

Comment partager les données ?

1. Le Plan de Gestion de Données
2. Les principes FAIR
3. Les licences
4. Le Rôle des entrepôts (exposé plus animation)

Réutilisation des données

Aspects juridiques sur les données

Conclusion : Discussion autour des données en danger à OPSE





Introduction

*L'histoire et les grands concepts
de la science ouverte*



Qu'est ce que la science ouverte ?

DÉFINITION :

La science ouverte, c'est la diffusion sans entrave des publications et des données de la recherche.

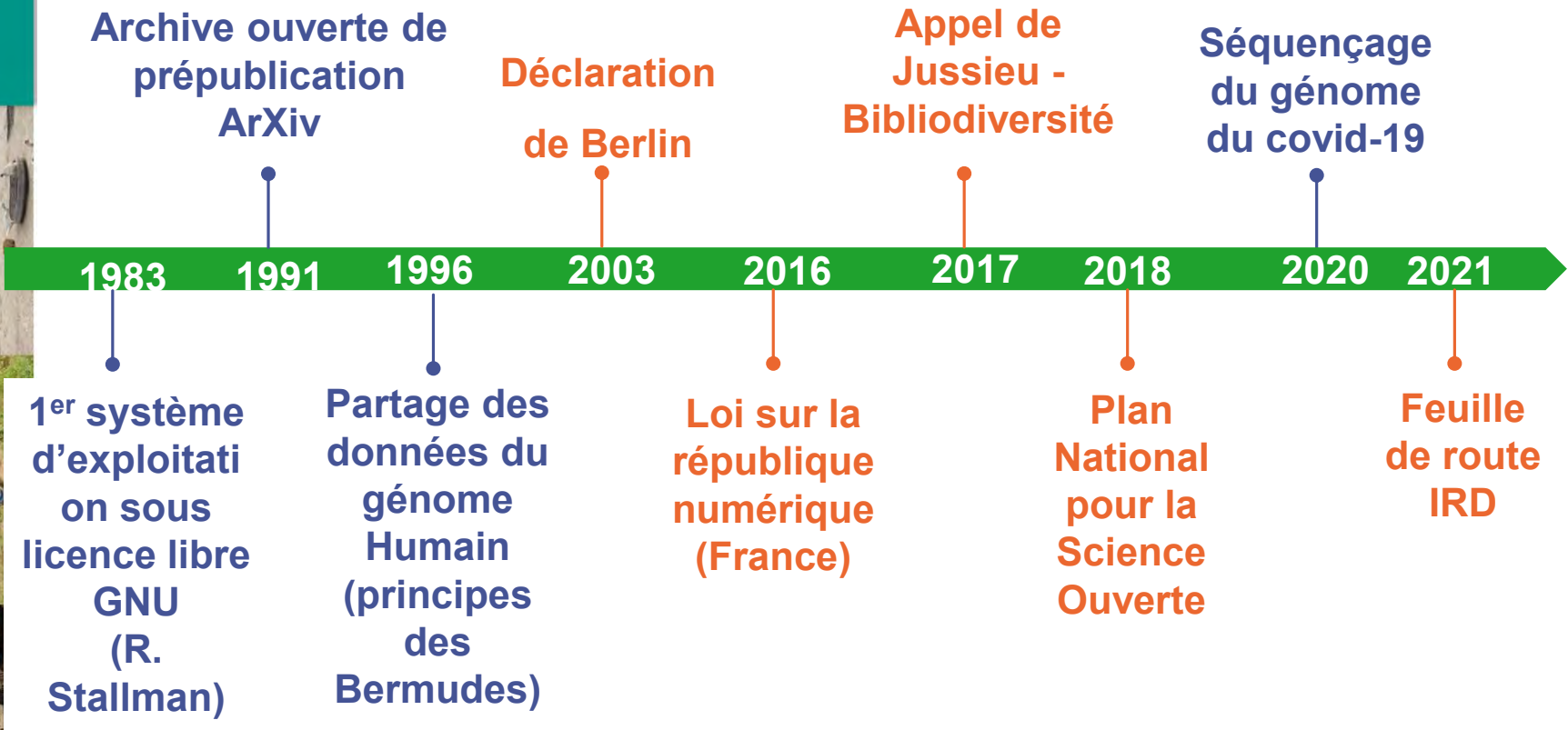
- Elle vise à construire un éco-système dans lequel la science est plus **cumulative**, plus **rapide** et d'accès plus **universel**.
- Elle favorise les **avancées scientifiques** et l'innovation.



Source : <https://www.ouvrirlascience.fr/plan-national-pour-la-science-ouverte/> juillet 2018

Quelques jalons historiques

Les déclencheurs scientifiques
Les pas politiques
Les faits les plus marquants



1

Pourquoi la science ouverte ?

les constats et motivations



Pourquoi la science ouverte ?

- **PRINCIPE**
Public money ? Public data !
- **SCIENTIFIQUE**
Une recherche de meilleure qualité
- **SOCIÉTÉ**
Une science plus ouverte et transparente,
plus accessible à la société



Crise de l'accès : une offre de plus en plus coûteuse

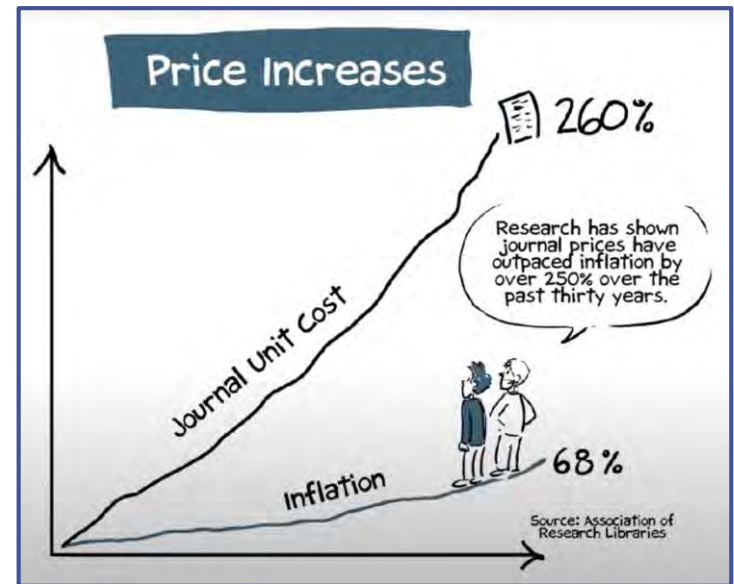


EN 2008

« la bibliothèque de Harvard était
abonnée à **98 900** périodiques

... contre **73 900** pour l'université Yale ...

La bibliothèque l'Indian Institute of
Science, était abonnée à **10 600**
périodiques. »

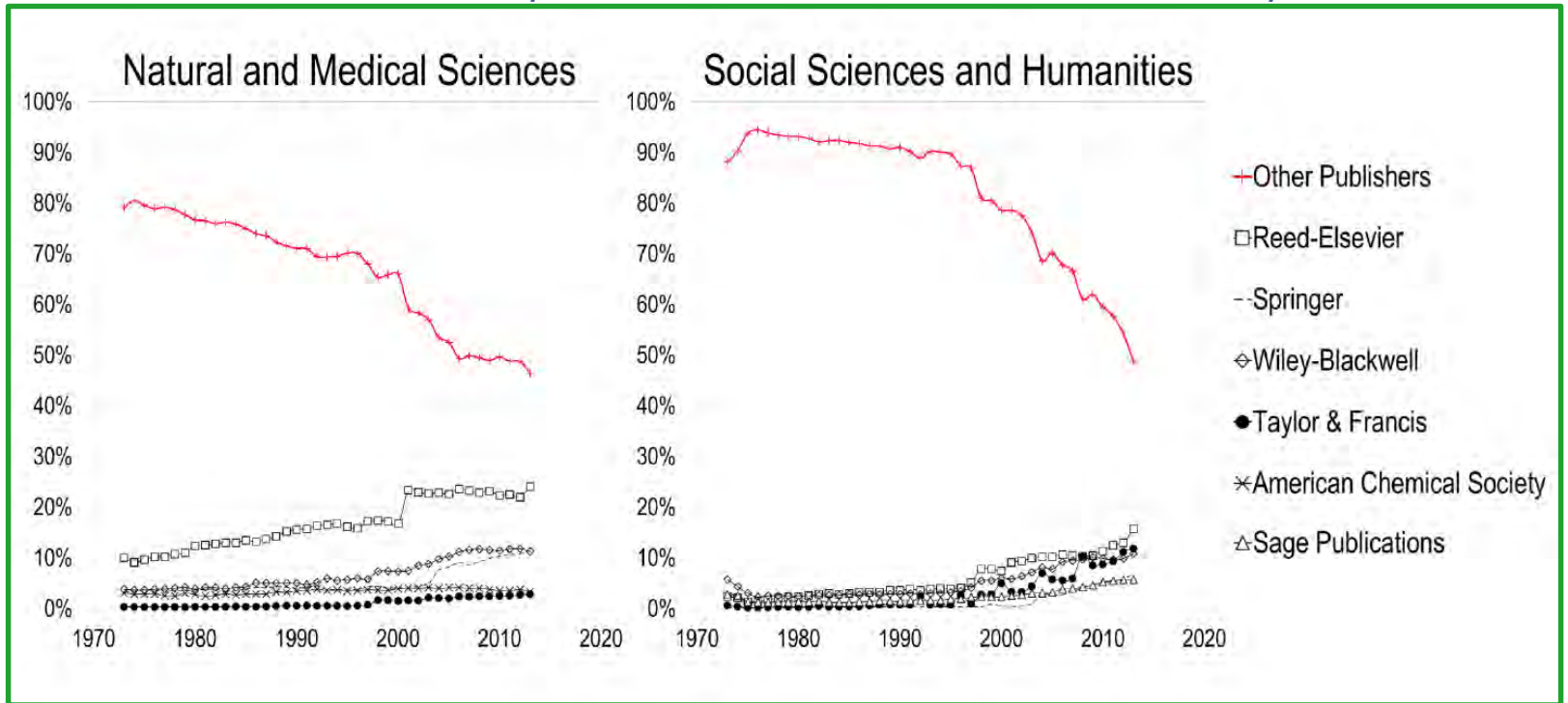


Source : <http://phdcomics.com/tv/?v=L5rVH1KGBCY>



Contrôle de l'édition savante par 5 grands éditeurs

Evolution de la part des éditeurs dans l'édition scientifique



Source : Vincent Larivière Journées Nationales de la Science Ouverte 2018

Les grandes puissances économiques utilisent aussi Sci-Hub !

Explosion des téléchargements sur Sci-Hub à travers le monde

The image shows a YouTube video player interface. The video content includes a line graph titled "top countries last year" showing the number of downloads per month (M/TH) from 2011 to 2020. The y-axis ranges from 1 M/TH to 8 M/TH. The x-axis shows dates: 2011 Sep, 2013 May, 2015 Jan, 2016 Sep, 2018 May, and 2020 Jan. The graph shows a steady increase in downloads over time. A list of countries is shown on the left, with "United States" and "Germany" highlighted with orange boxes. Below the graph, a black box contains the text "that Sci-Hub makes sense only in some poor countries". The video player shows a progress bar at 15:19 / 52:19 and a video player interface with a woman speaking in the background.

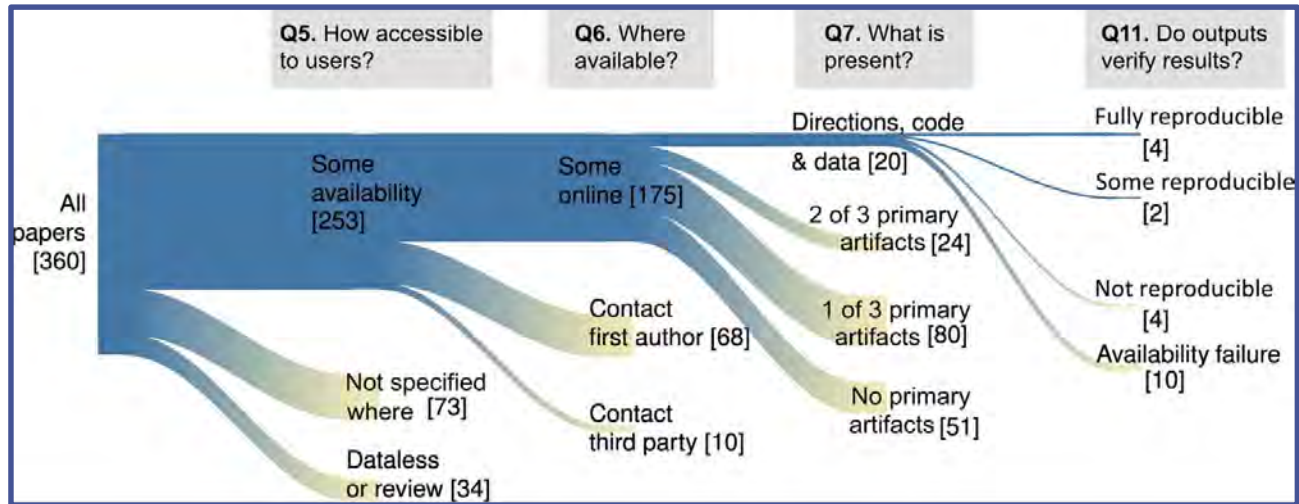
Country	Downloads per month (M/TH)
China	~7.5
India	~6.5
Brazil	~5.5
United States	~4.5
Indonesia	~3.5
Mexico	~2.5
Turkey	~2.0
Germany	~1.5
Iran	~1.0

Alexandra Elbakyan | Open knowledge: A silent revolution in science
311 visualizações • 1 de fev. de 2021

22 0 COMPARTILHAR SALVAR ...

Crise de la reproductibilité : un exemple en hydrologie

Nombre de papiers traités pour déterminer la disponibilité des données et le niveau de reproduction de l'expérience

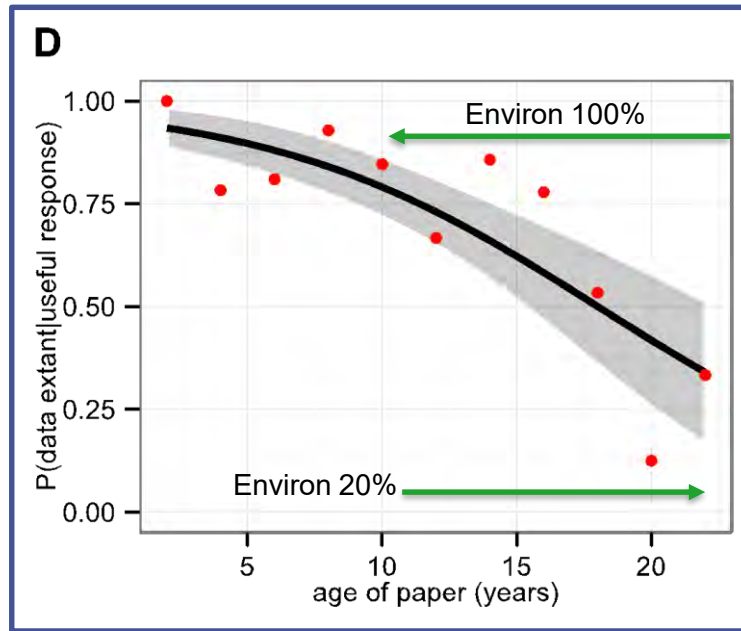


Source : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6390703/>

Moins de 2% des expériences sont reproductibles.

La bibliothèque des données brûle ...

*20 ans après leurs publications,
80% des données sous-jacentes
sont perdues*



La disponibilité des données diminue avec le temps.

« The availability of research data declines rapidly with articles age » (Nature, 2014)

Wikipédia s'appuie sur la connaissance et les savoirs scientifiques



WIKIPEDIA ANGLOPHONE CITE :

- 300 000 articles scientifiques
- 8 000 ouvrages de recherche



WIKIPEDIA
The Free Encyclopedia

Source : Etudes Wikipedia 2017



Sismologie : le partage de données est déjà la norme

En sismologie, la majeure partie des données sont ouvertes et permettent ainsi une **amélioration des connaissances** et de la mesure des risques sismiques à l'échelle mondiale.



IRIS
Incorporated Research Institutions for Seismology

Contact Staff Employment Sitemap Search

Data Management Center

- Seismograms
- Event Data
- Metadata
- Historical Data
- Products

IRIS Data Sources

Home » Data » Data Sources

Seismic Data Sources

Permanent Networks

Global Seismographic Network

The IRIS Global Seismographic Network (GSN) is one of the four major components of IRIS Consortium. The goal of the GSN is to deploy 128 permanent seismic recording stations uniformly over the earth's surface. [Learn more about the GSN](#)

FDSN

The [International Federation of Digital Seismograph Networks](#) (FDSN) is a global organization. Its membership is comprised of groups responsible for the installation and maintenance of broad-band seismographs either within their geographic borders or globally. [Learn more](#)

- List of [Permanent FDSN Network Codes](#)
- [Request a permanent FDSN Network Code](#)
- Data [shipment statistics](#) for FDSN and other data

Regional Networks

Several regional networks contribute data to the IRIS archive or have open access to their data sets. [List of regional networks](#)

Virtual Networks at the IRIS DMC

A virtual network is a group of stations and/or seismic networks. [List of virtual networks](#)

Une réutilisation citoyenne des données de la recherche



- **Initiative de consommateurs** collaborant avec une enquête scientifique
- **Contribution à la définition du nutriscore** utilisé dans les applications grand public



Exemple de réutilisation de données publiques ouvertes avec diffusion de code

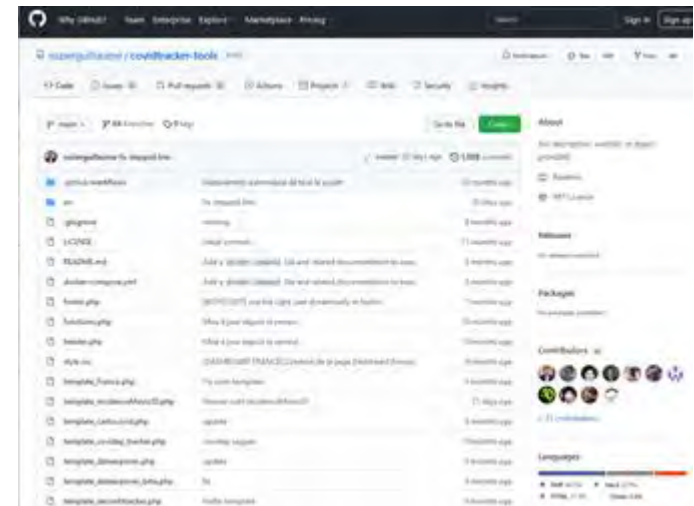
<https://www.covidtracker.fr>

CovidTracker.fr est un site web, d'agrégation de données publiques ouvertes de l'état français, sur le suivi de la pandémie de Covid-19 au niveau national. Site créé à l'initiative d'un citoyen.

Un ensemble d'outils ont été développés :

- prise de RDV de vaccination selon la disponibilité des stocks,
- rapports en ligne sur la pandémie du niveau national au niveau local, etc.

Le code est ouvert,
disponible sur une plateforme de partage
GitHub
Il peut être réutilisé ou enrichi.



Des incitations institutionnelles, nationales et internationales pour ouvrir les données et les publications

- **Recommandations de l'Unesco - nov 2021**
https://unesdoc.unesco.org/ark:/48223/pf0000378381_fre
- **Horizon Europe** https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/open-science_en
- **2e Plan national de la Science ouverte (France)**
<https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>
- Feuilles de route de l'IRD et CIRAD et autres...
 - IRD <https://www.ird.fr/une-feuille-de-route-pour-une-science-ouverte-et-partagee>
 - CIRAD <https://www.cirad.fr/collaborer-avec-nous/science-ouverte>



40° Conférence générale : les 193 États membres de l'UNESCO ont chargé l'Organisation d'élaborer un instrument normatif international sur la science ouverte, sous la forme d'une recommandation de l'UNESCO sur la science ouverte.

- Consultation des états membres (2020-2021)
- Adoption de la Recommandation (à venir) par les États membres en novembre 2021

[Open Science \(unesco.org\)](https://open.science.unesco.org/)



Composantes de la science ouverte

"Déclaration pour le partage et l'ouverture des données de la recherche pour le développement durable". 2019

Les participants du Colloque « Science Ouverte au Sud » se sont dotés des objectifs partagés suivants :

Déclaration pour le partage et l'ouverture des données de la recherche pour le développement durable, Dakar, 2019 <https://doi.org/10.5281/zenodo.3529678>

- a) un préalable : renforcer la connectivité et les services sur les données et les compétences dans la gestion, la valorisation et la réutilisation des données.
- b) des principes de bonne gestion des données,
- c) des principes de valorisation des données et des acteurs du partage
- d) des principes de gouvernance des données.

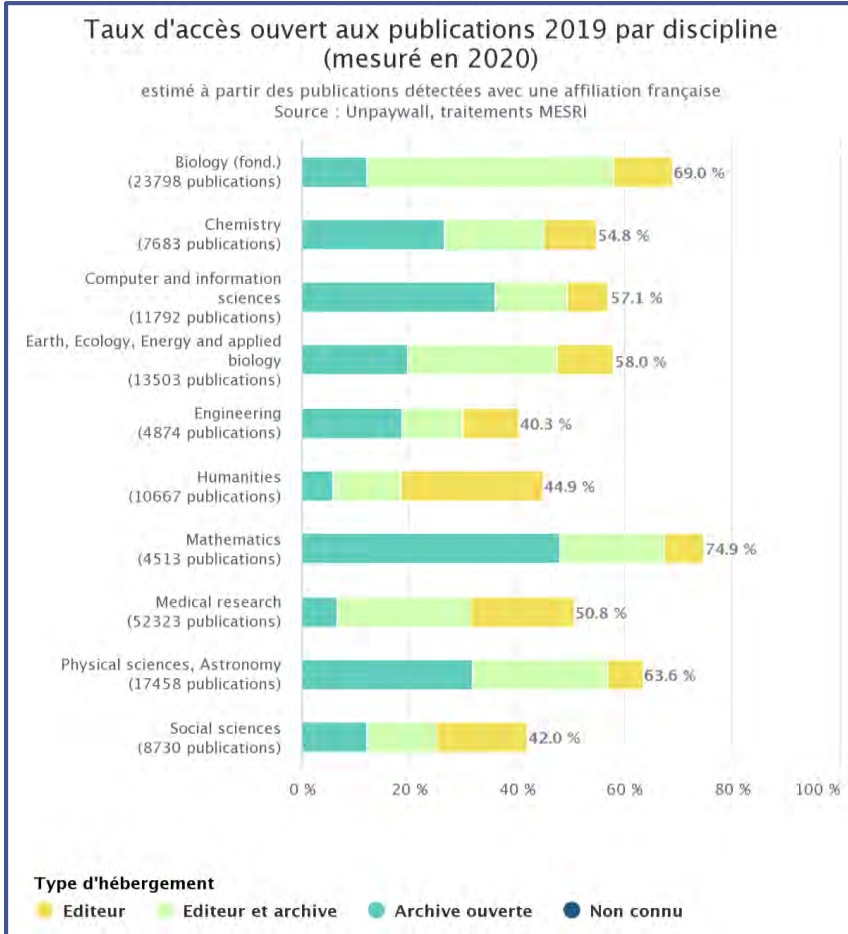


Introduction au libre accès

Les publications scientifiques en accès libres



Baromètre de la science ouverte en France



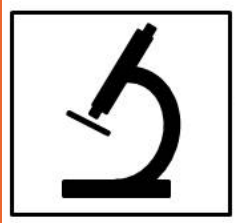
• **56 % des publications en accès libre**

(+ 7% en 1 an)

• **86.816 / 1.700.227 publications "France"**

Comment publier en libre accès ?

Votre document



Voie verte

* Site de Preprints

Version non validée de l'article

* Archive ouverte

- Dépôt autre type de document
- Dépôt article

Version acceptée auteur, embargo de 6-12 mois (loi française)

Journal - Accès sur abonnement

Voie dorée

€ Payé par les auteurs

Voie diamant

€ Payé par les institutions

Centre de Ressources virtuelles des Rivières du Sud, Sénégal



Centre de ressources virtuel des Rivières du Sud

Accueil

Collections ▾

Dépôts

En savoir plus ▾

Ouvrir une session

🏠 Accueil du CRV / Parcourir par titre

Recherche



Parcourir par titre

PARCOURIR

Tout le CRV

Communautés et Collections

Titres

Auteurs

Sujets

Date de publication

Territoires

MON COMPTE

Ouvrir une session

S'inscrire

0-9 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Ou entrer les premières lettres :

Aller

Voici les éléments 1-20 de 750



15 villes signent la déclaration de l'ONUSIDA pour mettre fin à l'épidémie du sida.

CNLS (Conseil National de Lutte contre le Sida du Sénégal). Magazine d'information. Dossier CTA. + Vie du Programme. (2016)

Une cérémonie de signature d'une déclaration d'engagement entre le secteur privé et le Conseil National de Lutte contre le Sida pour mettre fin à l'épidémie du sida au Sénégal.

2,6-Quinolinylnyl dérivatives as potent VLA-4 antagonistes

Lassoie, Marie-Agnès; Gassama, Abdoulaye; Broeders, Fabienne; Collart, Philippe (Elsevier, 2007)

Une nouvelle série de dérivés 2,6-quinolinylnyl a été préparée conduisant à de puissants antagonistes de VLA-4 / VCAM-1 nanomolaires puissants.

4. Agronomie et pédologie du bassin de l'anambé.

Sénégal. Ministère du développement rural et de l'hydraulique. Société de Développement Agricole et Industriel SODAGRI (1977)

L'aménagement du bassin de l'anambé a pour objectif la mise en valeur des terres consacrées à la riziculture. Pour atteindre cet objectif des études agronomiques et pédologiques ont été effectuées.

40ème anniversaire 1963 - 2003 de l'institut de technologie alimentaire. "Assurer un service de qualité pour promouvoir l'émergence de produits locaux compétitifs".

Sénégal. Ministère de l'Agriculture et de l'Élevage. ITA Institut de Technologie Alimentaire. (2003)

<https://rivieresdusud.uasz.sn/>

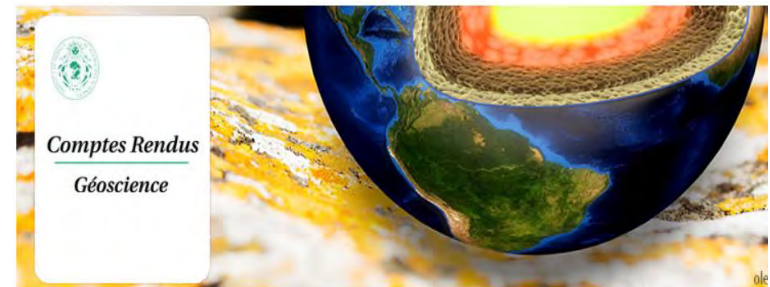
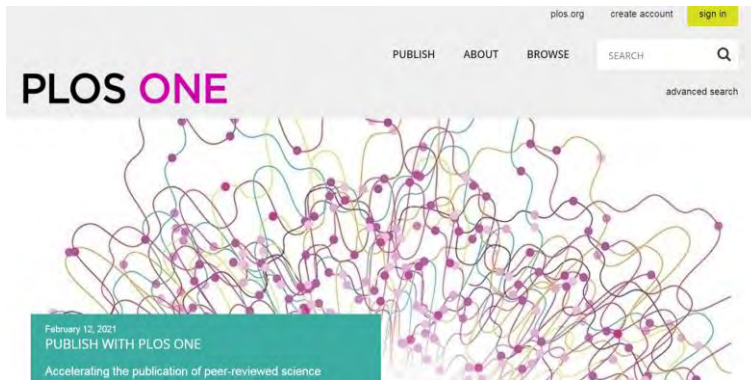
Publier dans une revue en libre accès

Voie dorée

<ul style="list-style-type: none"> • Les auteurs conservent leurs droits • Fonctionnement d'une revue classique 	<ul style="list-style-type: none"> • Coût des frais de publications payés par l'auteur (moyenne 2000 \$) • Présence d'éditeurs prédateurs
---	---

Voie diamant

<ul style="list-style-type: none"> • Les auteurs conservent leurs droits • Fonctionnement d'une revue classique • Gratuite pour les auteurs 	<ul style="list-style-type: none"> • Offre réduite
---	---



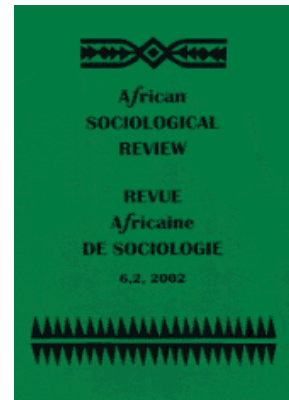
DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

17000 Revues en Libre Accès
<https://doaj.org>



538 revues en Afrique dont 278 en libre accès
<https://www.ajol.info/index.php/ajol>

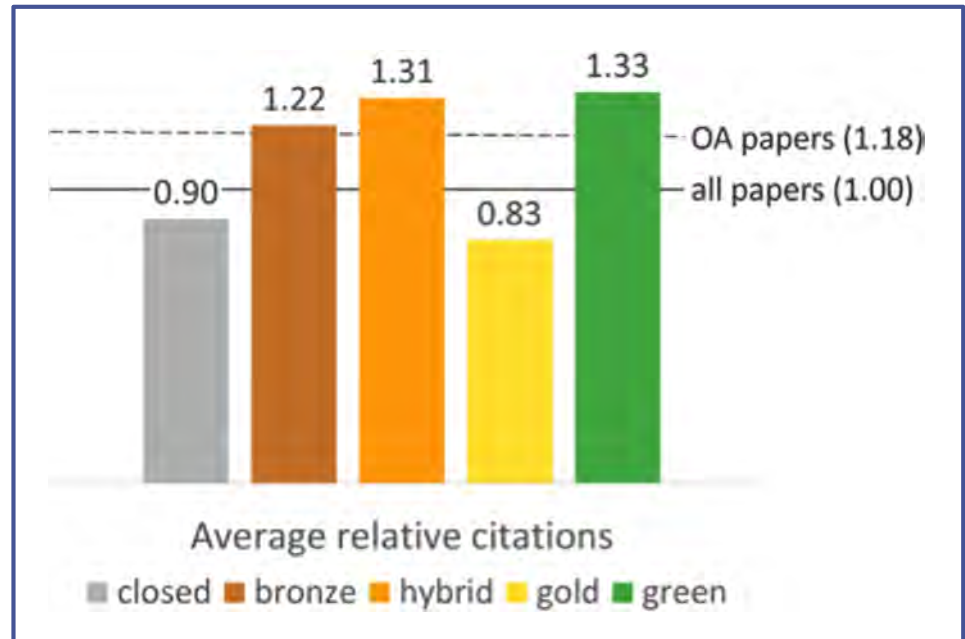
6 revues éditées au Sénégal, mais pas en libre accès



Les articles en libre accès sont plus cités



- **18% de citation en plus**
- **Sur 70 études jusqu'en 2015 :**
 - 46 montre un avantage de citation
 - 17 ne montrent aucun avantage
 - 7 ne peuvent pas conclure



Source : <https://doi.org/10.7717/peerj.4375>



Publier / faire évaluer un Preprint



- * Rapidité
- * Possibilité de publier aussi l'article dans un journal



- * Pas de validation qualité ?
- * Prise en compte dans les systèmes d'évaluation ?



bioRxiv
THE PREPRINT SERVER FOR BIOLOGY



* PCI = Recommandation de Preprints

Introduction au données de la recherche

Open research data



Données de la recherche

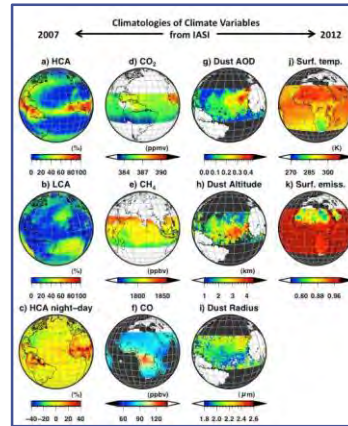
Une expression qui renvoie à de multiples objets... et pas de consensus mais...

DÉFINITION :

“Le terme de données de la recherche désigne les **données** sous forme de **faits**, **d’observations**, **d’images**, de **résultats de programmes informatiques**, **d’enregistrements**, de **mesures** ou **d’expériences** sur lesquels un argument, une théorie, un test ou une hypothèse, ou un autre produit de la recherche est basé. »

“Elles peuvent être **brutes**, **nettoyées** ou **traitées**, ...”

d’après ANDS, « What Is Research Data », 2017.
https://www.ands.org.au/data/assets/pdf_file/0006/731823/Whatis-research-data.pdf



Source : <https://www.futura-sciences.com/planete/actualites/meteorologie-secrets-iasi-instrument-meteorologie-41551/>

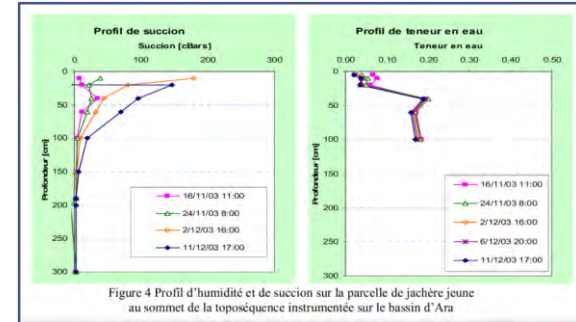
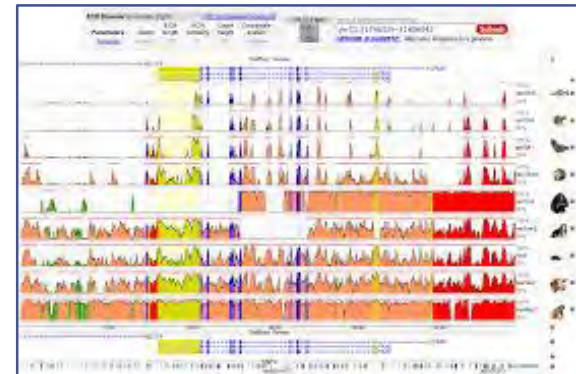


Figure 4 Profil d’humidité et de succion sur la parcelle de jachère jeune au sommet de la toposequence instrumentée sur le bassin d’Ara

Source : <https://www.semanticscholar.org/paper/Caract%C3%A9risation-des-processus-hydrologiques-sur-le-S%C3%A9guis-Galle/64388d1c3162ca8567ebf07aa23ae8221c2eae4/figure/3>



Source : https://www.google.com/url?sa=i&url=http%3A%2F%2Fpedagogix-tagc.univ-mrs.fr%2Fcourses%2FBIUS_bio%2Fevolution%2Fpdf_files%2F2.4.genomique_dias.pdf&psl=AOVaw5JfPYSH4-GB4dm0_Xc98&ust=1613834897115000&source=images&cd=vfe&ved=0CAMQIB1qFwoTKDxoNq8u4CFOAAAAAAdAAAAAABAQ

Jeu de données hydrométriques déposé dans l'entrepôt Datasuds

HSM UMR HydroSciences Montpellier (CNRS, Université de Montpellier, IRD) → site web

DataSuds > UMR HydroSciences Montpellier > ADHI: African Database of Hydrometric Indices

Contact Partager

ADHI: African Database of Hydrometric Indices
Version 4.0

Tramblay, Yves; Rouché, Nathalie, 2020, "ADHI: African Database of Hydrometric Indices" <https://doi.org/10.23708/LXGXQ9>, DataSuds, V4, UNF:6:+hcxglYr5VFf5f5sklSsw== [fileUNF]

Citer le jeu de données

Statistiques du jeu de données
454 Téléchargements

Pour en savoir plus, consultez le document Data Citation Standards.

Description This database contains different hydrometric indices, describing runoff characteristics, seasonality, floods and low flows, for about 1500 rivers covering all regions of Africa between 1950 and 2018. (2020-08-10)

Sujet Computer and Information Science; Earth and Environmental Sciences

Mot-clé Africa, River, Discharge, Runoff, Flows, Afrique, Rivière, Débit, Ecoulement

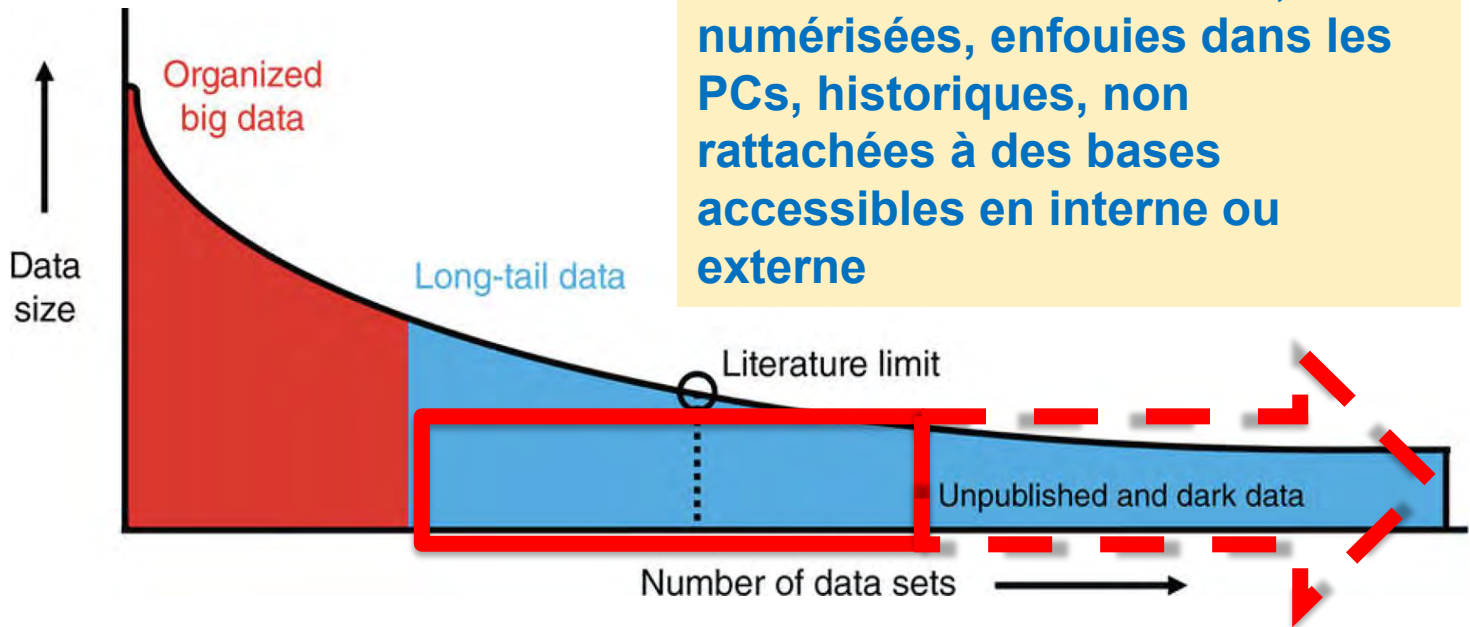
Related Publication Tramblay Y., Rouché N., Paturol J.-E., Mahé G., Boyer J.-F., Amoussou E., Bodian A., Dacosta H., Dakhlaoui H., Dezetter A., Hughes D., Hanich L., Peugeot C., Tshimanga R., Lachassagne P.: ADHI: The African Database of Hydrometric Indices (1950-2018). Earth System Science Data. doi: 10.5194/essd-2020-281

<https://doi.org/10.23708/LXGXQ9>

La longue traîne des données

Observatoires, pôles de données, bases de données accessibles sur le web

Données non structurées, non numérisées, enfouies dans les PCs, historiques, non rattachées à des bases accessibles en interne ou externe



Source : distribution des données de la recherche (Ferguson et al., 2014)



Préparer au plus tôt les données pour leur diffusion et leur réutilisation

- Pratiques de gestion des données spécifiques aux disciplines
- Certaines communautés sont organisées autour de dispositifs de gestion et de diffusion déjà existants
- **Priorité sur les données** qui ne font pas l'objet d'une **préservation**, d'une diffusion par l'intermédiaire de services en ligne dédiés, un risque de pertes



Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Mass.: The MIT Press, 2015, pp.9-10.

Un principe-clé

“ouvert autant que possible, fermé autant que nécessaire”



OUVERT AUTANT QUE POSSIBLE



FERMÉ AUTANT QUE NÉCESSAIRE POUR :

- Données **issues des activités de recherche en santé** impliquant des personnes humaines (Loi Jardé)
- Données **sensibles** (protection industriel, secret défense, mettant en danger écosystème)
- Données **personnelles** non anonymisées ou pseudo-anonymisées



Les dispositifs existants pour ouvrir ses données

En France et ailleurs...



- Entrepôts **généralistes** : Zenodo (EU), Figshare, Dryad
- Entrepôts **disciplinaires** : Nakala (SHS, FR)
- Entrepôts **institutionnels (FR)** : DataSuds, Cirad Dataverse, Data INRAE, Portail des données marines



DataSuds IRD

- Portails et bases de données des **dispositifs d'observation et d'expérimentation (FR)**: Pôles de données des sciences de la Terre, de la biodiversité
- Entrepôts des **éditeurs** : Elsevier, Mendeley data, etc.



Quelques entrepôts de Données au Sénégal

À compléter par la salle

Sénégalais

- Geo Senegal Infrastructure nationale de gestion des données géographiques: <https://geosenegal.sec.gouv.sn/>
- ANADS : <http://anads.ansd.sn/index.php/catalog/229>).
- Geoportail service climatique Sénégal : <https://retd1.teledetection.fr/climap/>
- Sénégal le portail de données <https://senegal.opendataforafrica.org/>

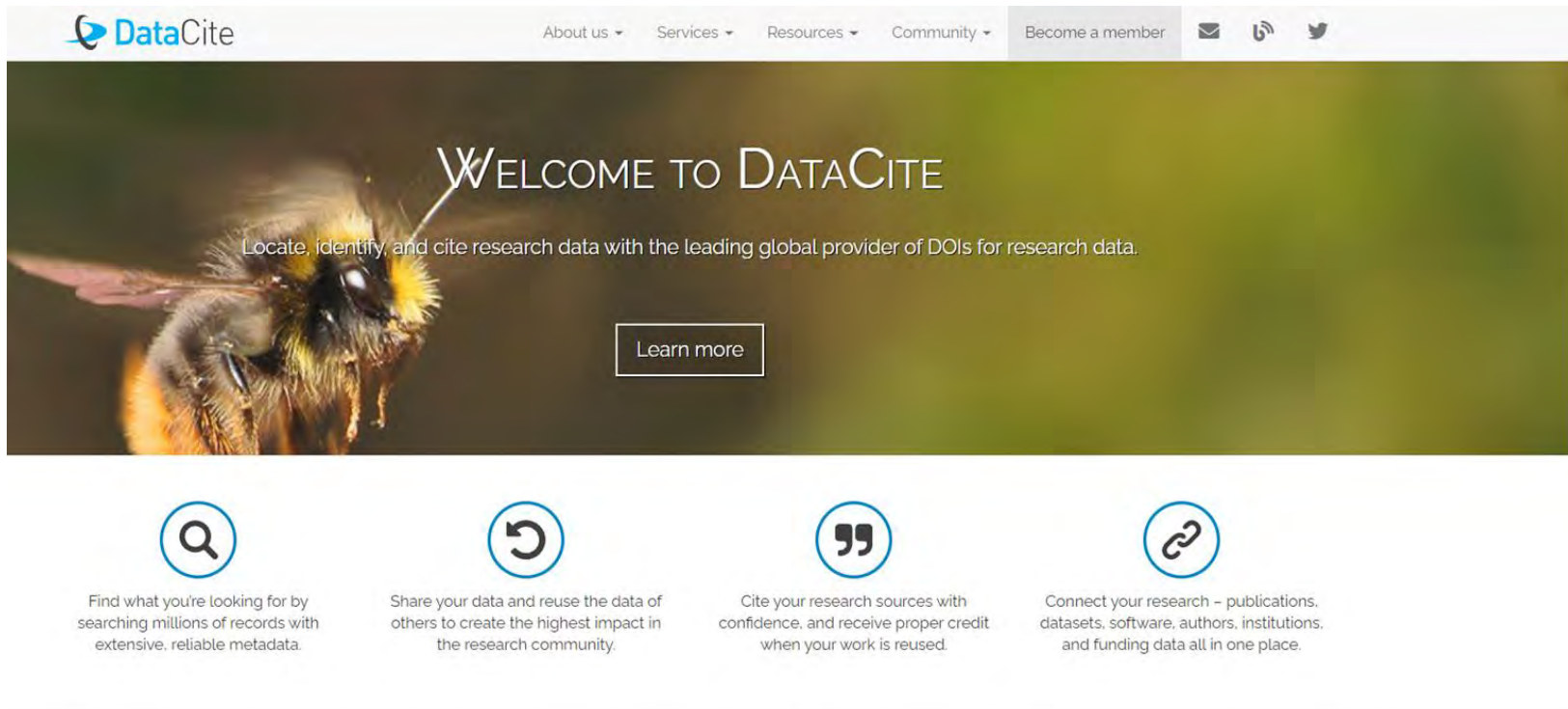
Internationaux

- West African Science Service Centre on climate change and adapted land use Wascal Data Repository and portals (<https://wascal.org/wascal-data-repository-and-portals/>)
- - [Human Heredity and Health in Africa](https://h3africa.org/) (H3Africa) : (<https://h3africa.org/>)
- - INDEPTH Data repository (<https://www.indepth-ishare.org/index.php/home>)
- WorldPop Open Population Repository (<https://wopr.worldpop.org/>)
- BAOBAB (<https://baobab.sedoo.fr/>)
- Global Biodiversity Information Facility <https://www.gbif.org/country/SN/summary>

Datacite

Consortium international qui attribue le DOI et encadre l'identification pérenne d'objets numériques

<https://datacite.org>



DOI Digital Object Identifier



Identifier les données de manière pérenne

- Méthode standardisée pour l'identification permanente d'un objet numérique diffusé sur internet
- L'attribution et la gestion des DOI est sous la responsabilité d'agences externes internationales

<https://doi.org/10.5555/YRFU1371>

DOI resolver prefix suffix

<https://www.crossref.org/documentation/member-setup/constructing-your-dois/>

DataCite : recherche « Senegal » et 2021

<https://datacite.org>

DataCite Commons

× Q

Works People Organizations

Pages Support Sign In

1,113 Works

Publication Year

 2021

Work Type

 Dataset

License

 CC-BY-NC-4.0
 CC-BY-4.0
 CC0-1.0

Language

 English
 Wolof
 French

20m West Africa's Mangrove Forest Extent Data Products for 2017 Created from Sentinel-2 and Sentinel-1 Satellite Data

Xue Liu, Temilola E. Fatoyinbo, Nathan M. Thomas, Wendy Guan, Yanni Zhan, Pinki Mondal, David Lagomasino, Marc Simard & Carl C. Trettin

Content published 2021 in [Harvard Dataverse](#)

2017 mangrove forest extent mapping data products at 20m resolution for 11 countries in West Africa created using a combination of Sentinel-2 and Sentinel-1 satellite imagery (Senegal, The Gambia, Guinea Bissau, Guinea, Sierra Leone, Liberia, Cote D'Ivoire, Ghana, Togo, Benin, and Nigeria). This dataset includes 9 GeoTIFF files (Senegal and The Gambia combined, Togo and Benin combined).

DOI registered January 5, 2021 via DataCite.

[Dataset](#)

<https://doi.org/10.7910/dvn/9evoeo>

24 2 3.5K Citations

Introduction à l'ouverture du code source

Logiciels de la recherche

Code source ouvert | Open source

Les logiciels de recherche sont des logiciels scientifiques originaux développés à des fins d'analyse, de simulation, de visualisation, etc dont le code source est disponible.

Selon la définition de l'[Open Source Initiative](https://opensource.org/), les logiciels open source doivent :

- être distribués sous forme de codes sources
- être partagés sous une licence permettant la modification, la dérivation et la redistribution



<https://opensource.org/osd>

Introduction à l'ouverture du code source

Logiciels de la recherche

Enjeux

- La recherche s'appuie sur des logiciels. Pour exploiter ou reproduire cette recherche, il faut avoir accès au code source complet
 - L'ensemble qui a produit le résultat de la recherche est constitué du code, des données [et de l'environnement technique].

Le libre accès au code source des logiciels, contribue à la transparence, et à un meilleur impact, de la recherche

- Le partage des logiciels utilisés pour la recherche est une condition nécessaire, mais non suffisante de la reproductibilité
 - Les logiciels sont perçus comme des « boîtes noires »
 - Une description écrite rigoureuse d'un logiciel ne permet pas de rendre compte de tous les résultats y compris l'erreur.

Introduction à l'ouverture du code source

Logiciels de la recherche

Enjeux

- **Rôle d'une forge logicielle**

Une forge est une plateforme qui permet la gestion, le contrôle de version et le suivi global des modifications d'un logiciel :

- services de stockage à distance
- maintenance, partage et collaboration sur des logiciels de recherche

Exemples de forges : GitHub (Microsoft), GitLab, Bitbucket,...

Une forge s'avère centrale pour établir un flux de travail de recherche ouvert et reproductible



Introduction à l'ouverture du code source

Logiciels de la recherche

Enjeux

- Importance d'associer un identifiant pérenne (DOI) au logiciel produit
 - Certains entrepôts s'intègrent à GitHub pour archiver le logiciel et fournir un DOI lorsque les développeurs effectuent une publication officielle sur GitHub.

Save and reference research software

Home / Save and reference research software

A pillar of Open Science

"Software should be considered a legitimate and citable product of research." — Software citation principles

"Non-reproducible single occurrences are of no significance to science"
— Karl Popper, The Logic of Scientific Discovery, 1934

"Sometimes, when you do not have the software, you do not have the data."
— Christine Borgman, JNSO 2018

The diagram shows three cylindrical icons representing different types of repositories. At the top is 'Open Source Repositories' with a code symbol (</>). At the bottom left is 'Open Access Repositories' with a document icon. At the bottom right is 'Open Data Sets Repositories' with a bar chart icon. These three icons are connected by lines to a central globe icon, forming a triangular structure.

Roberto Di Cosmo, *How to use Software Heritage for archiving and referencing your source code: guidelines and walkthrough*, <https://www.softwareheritage.org/save-and-reference-research-software/>

Introduction à l'ouverture du code source

Logiciels de la recherche

Enjeux

- Le partage ouvert des logiciels de la recherche permet aux développeurs de recevoir un crédit pour leurs efforts :
 - soit par une citation directe (via DOI)
 - soit par des articles sur les logiciels publiés

Exemples de revues spécialisées

- Journal of Open Research Software
- Journal of Open Source Software
- Journal of Statistical Software ...

Optimal Design Generation and Power Evaluation in R: The skpr Package

Tyler Morgan-Wall, George Khoury

Vol. 99, Issue 1

 Paper

 R package (skpr)

 R replication code



Journal of Statistical Software

[Information on Mission](#) [Information for Authors](#) [Style Guide](#) [Volumes](#) [About](#)

Introduction à l'ouverture du code source

Logiciels de la recherche

En France > une ambition inédite

Ouvrir et promouvoir les codes sources produits par la recherche

« Le deuxième Plan national pour la science ouverte s'enrichit d'un nouvel axe dédié aux **codes sources produits par la recherche**, qui représentent **un enjeu majeur pour la reproductibilité** des résultats scientifiques. Ils devront être développés et maintenus de façon durable, valorisés en tant que productions scientifiques et **diffusés de manière privilégiée sous licence libre.** » Juillet 2021

Les données et les logiciels de la recherche dans les politiques nationales



Valoriser et soutenir la diffusion sous licence libre des codes sources

- Soutenir Software Heritage et proposer la standardification du Software Heritage Identifier (SWHID) qui complètera les DOI pour les logiciels
- Créer un prix science ouverte du logiciel libre pour la recherche



<https://www.softwareheritage.org/?lang=fr>

Cycle de vie des données de recherche

Définition

Le cycle de vie des données de recherche (Research data lifecycle) décrit le processus d'utilisation des données de leur création à la publication et à leur réutilisation ultérieure ⁽¹⁾

- Le cycle de vie des données permet de savoir à quelle étape on se situe
- Il est important de savoir comment chaque étape a été traitée dans un jeu de données
- Les différents étapes sont reprises dans la rédaction d'un plan de gestion de donnée
- Structure une gouvernance des données

Introduire les étapes => puis atelier d'expression sur la base du cdvie expliqué
6 à 7 groupe + 1 restitution. Travailler par post it sur les étapes.

(1) Source : https://www.inist.fr/wp-content/uploads/donnees/co/module_Donnees_recherche_7.html
Projet Open OPSE | Formation à la gestion des données de la recherche, 26 octobre 2021

TP Le cycle des données et vous.



<https://data.ird.fr/utiliser-des-donnees/>



Pourquoi partager les données ?

Contextes et enjeux

1. Politiques nationales / internationales
2. Politiques des financeurs, exigences des éditeurs
3. Réutilisation des données des autres
4. Diffusion et partage des connaissances



Contexte français

France : 2eme Plan national sur la science ouverte, 2021

Un cadre pour les institutions de recherche, les universités....

Les données et les logiciels de la recherche dans les politiques national



IRD

rendre nos productions faciles à trouver, accessibles, interopérables et réutilisables, qui est une condition sine qua non à l'avancée des connaissances scientifiques et à leur réutilisation par les communautés de recherche et les acteurs socio-économiques. C'est un facteur d'innovation au Sud.

Cirad

L'ambition du Cirad est de gérer et partager des données et des logiciels de recherche [selon les principes FAIR](#) (Facile à trouver, Accessible, Interopérable, Réutilisable). Dans le cadre du mouvement de l'open data et de l'open source, ces pratiques sont devenues des éléments clés d'une politique de science ouverte visant une recherche transparente, intègre, responsable et répondant aux exigences des bailleurs de la recherche.

Les financeurs

Politique des bailleurs de fonds

Ligne de conduite adoptée par des organismes qui financent la communauté de recherche et ses projets de recherche.
(Source : Foster).



VÉRIFIER LES RECOMMANDATIONS
Votre financeur, éditeur, institution, vos partenaires ou encore votre discipline peut vous guider vers un entrepôt.

Les financeurs

Politique des bailleurs de fonds

Horizon Europe, programme-cadre de l'Union européenne pour la recherche et l'innovation (2021-2027), renforce la politique initiée par Horizon 2020 sur l'ouverture de la recherche et **exige un accès ouvert immédiat** :

- **aux publications**, licence ouverte CC-BY
- **aux données FAIR** mises à disposition sur la base du principe « aussi ouvert que possible et aussi fermé que nécessaire ».
- **Plan de gestion des données**, obligation contractuelle.

Le principe de réciprocité en matière de science ouverte est également encouragé dans tous les accords d'association et de coopération avec des pays tiers.



Les éditeurs adoptent des politiques sur les données

4 types de Data policy . Springer Nature

Policy Types



Process



Nature

Humanities & Social Sciences Communications


View all journals Search Q Login

Explore content About the journal Publish with us Sign up for alerts RSS feed

nature > humanities and social sciences communications

Would you like to join our Editorial Board?

If you would like to apply to join our Editorial Board, please send your CV/resume and details of your publication history to: hsscomms@springernature.com. **Only those successful will be contacted.**



Announcements <https://www.nature.com/palcomms/>

Availability of materials and data

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims.

Therefore, a condition of publication is that authors are required to make materials, data and associated protocols **promptly available to readers without undue qualifications.**

...

Supporting data must be made available to the **journal and peer-reviewers** at the time of submission for the purposes of evaluating the manuscript.

the journal reserves the right to refuse publication in cases where authors do not provide adequate assurances that they can comply with the publication's requirements for sharing materials.

.....

All original articles should include a [data availability statement](#).

Une recherche plus cumulative

Partage du séquençage génome humain



- **Création de valeur** économique et de santé publique
- Coût de **3,6Md\$** pour le décodage
- Impact économique de plus de **700Md\$**

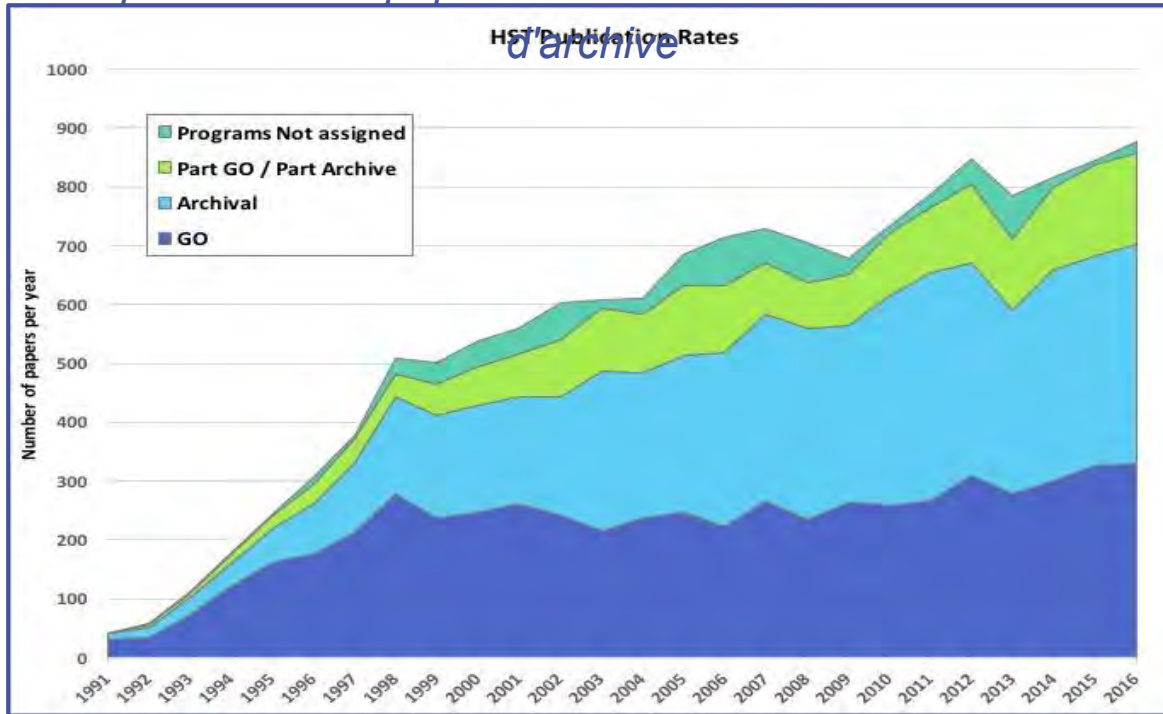


The Human Genome Project - <https://www.genome.gov/human-genome-project>

Une recherche réutilisable

Astronomie : les données « archives » produisent plus de publications que la collecte initiale

Importance des papiers écrits sur la base de données



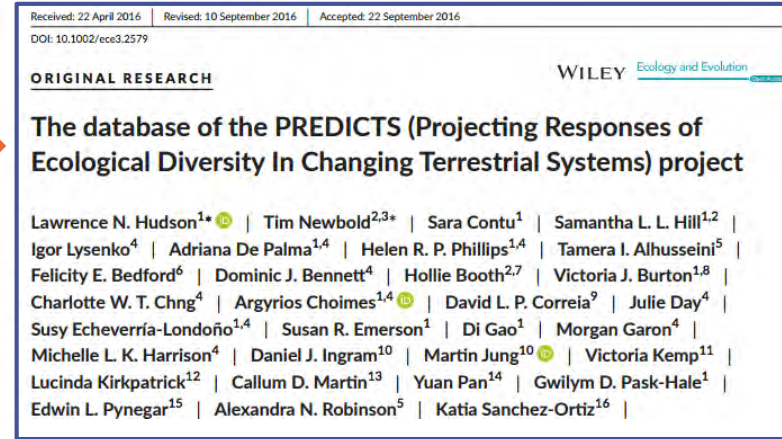
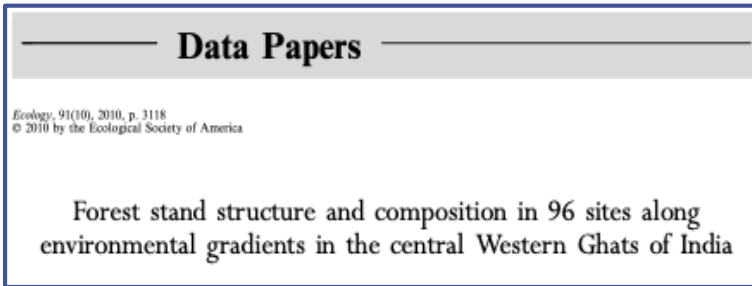
Source : <https://zenodo.org/record/1035837>



Une recherche plus visible

L'effet boule de neige des data papers

L'exemple d'un de nos collègues dans le domaine de l'écologie



- * **24 citations Google Scholar depuis 2010**
- * **21 dans des articles indépendants sans co-signature**

... 500 co-auteurs, 56 citations

Source : <https://www.futura-sciences.com/planete/actualites/meteorologie-secrets-iasi-instrument-meteo-climatologie-41551/>

Quand les données ne sont pas bien partagées...



<https://www.youtube.com/watch?v=N2zK3sAtr-4>

Mise en place de bonnes pratiques de gestion des données

PLAN DE GESTION DES DONNEES

- Au **début** d'un projet de recherche et l'appuie **pendant et post-projet**
- Mise en place de **bonnes pratiques** de gestion
- A toutes les étapes du **cycle de vie** des données

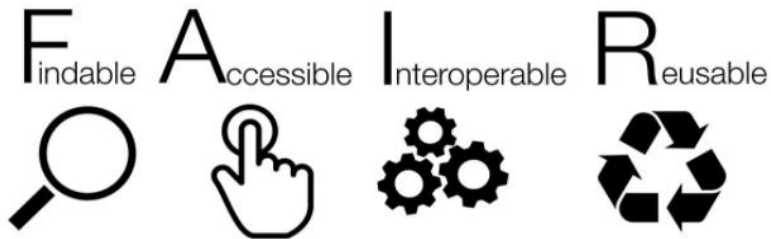


EN SAVOIR PLUS SUR :

<https://data.ird.fr/plan-de-gestion-de-donnees/>

<https://coop-ist.cirad.fr/gerer-des-donnees/>

... pour des données faciles à trouver et à réutiliser



- **Garantir** une recherche et un téléchargement des données
- **Faciliter** la découverte et l'évaluation de l'adéquation d'un jeu de données aux besoins du chercheur
- **Augmenter** la capacité de réutilisation des données



Présentation d'un jeu de données dans l'entrepôt DataSuds

The screenshot shows the DataSuds interface for the 'Base de données insulaires mondiale (BIM)' dataset. Annotations point to various features:

- Identifiant:** Points to the DOI logo and the dataset title 'Base de données insulaires mondiale (BIM)'.
- Métadonnées:** Points to the 'Description', 'Auteur', 'Mots-clés', 'Relevé Publications', and 'Remarques' sections.
- Licence:** Points to the Creative Commons Attribution (CC BY) license logo.
- Export citation:** Points to the 'Export citation' button in the top right corner.
- Accès fichiers:** Points to the 'Fichiers' tab and the list of files below.

Citer un jeu de données

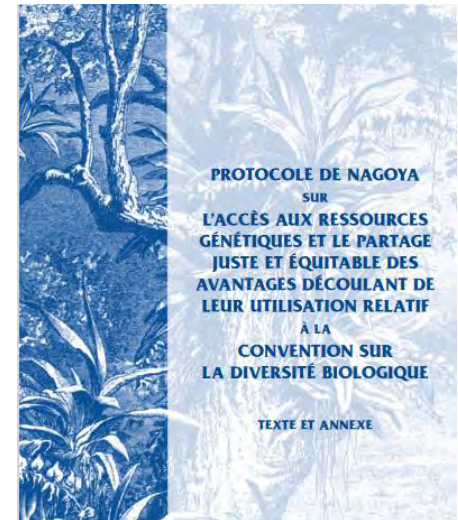
- Articles
- Rapports de projet
- Rapports d'activité**
- Dossiers de candidature

Contextes éthique, juridique

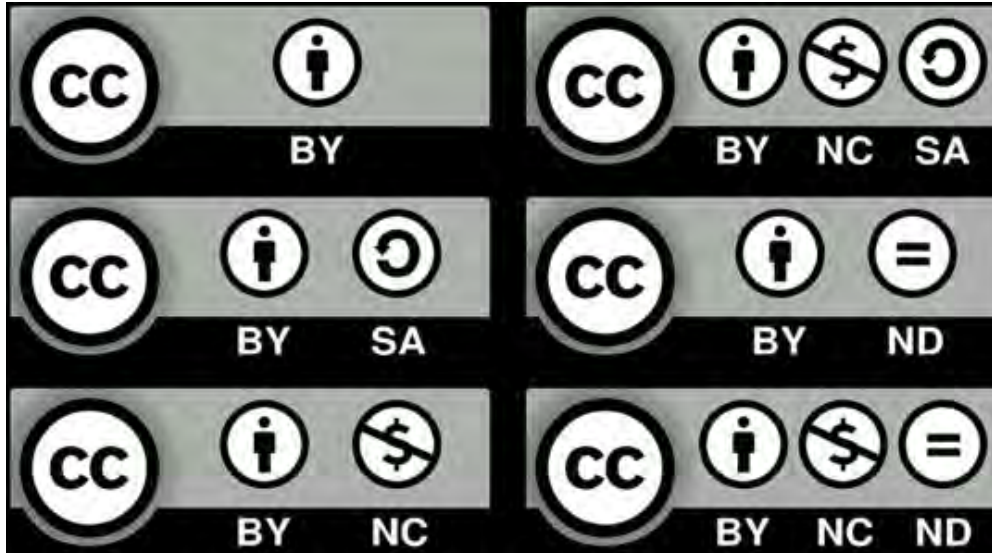
- ❑ CNERS, Comité National d'Ethique pour la Recherche en Santé
<https://cners.sn>
- ❑ CDP - Commission de Protection des Données Personnelles du Sénégal, <https://www.cdp.sn>

- ❑ Protocole de Nagoya

- ❑



Exemple licences internationales sur les données



BY- attribution : Toutes les licences Creative Commons obligent ceux qui utilisent vos oeuvres à vous créditer (citation)

NC – Pas d'utilisation commerciale : Vous autorisez les autres à reproduire, à diffuser et à modifier votre œuvre, pour toute utilisation autre que commerciale,

SA – Partage dans les mêmes conditions : Vous autorisez les autres à reproduire, diffuser et modifier votre œuvre, à condition qu'ils publient toute adaptation de votre œuvre sous les mêmes conditions que votre oeuvre.

ND- Pas de modification : Vous autorisez la reproduction et la diffusion uniquement de l'original de votre oeuvre

Entrepôts de données de la recherche

Service en ligne (outil) : dépôt, description, conservation, recherche et diffusion des jeux de données.



Types d'entrepôts : disciplinaires / institutionnels / généralistes / infrastructures nationales, internationales / commerciaux / créés par des éditeurs scientifiques /

A quoi sert un entrepôt de données ?

Visibilité, partage et accès aux données des Unités et projets de recherche

Maîtrise de la diffusion des données : licences et niveaux d'accès

Valorisation : susciter des collaborations entre recherche publique et secteur privé

Juridique et éthique : vérification des accords et suivi des cadres éthiques

Choix d'un entrepôt de données

La thématique Certains entrepôts sont incontournables dans leurs disciplines : renseignez-vous auprès de vos **collègues** !

La recommandation

Le **financeur** de la recherche ou le **journal** qui publie l'article associé fournit parfois des listes d'entrepôts recommandés.

Le **plan de gestion de données du projet** peut donner des instructions

La visibilité

L'entrepôt doit permettre de décrire les données selon les [principes FAIR](#)

La durabilité

Un entrepôt sera d'autant plus durable qu'il est soutenu par une ou plusieurs institutions publiques.

Les entrepôts certifiés par le [Core Trust Seal](#) sont à privilégier.

Autre critères : les services proposés , l'usage et les fonctionnalités

ne cédez jamais tous vos droits ! Vérifiez les conditions du dépôt dans l'entrepôt et les licences de réutilisation

Quelques entrepôts de données au Sénégal

Liste à compléter en séance

Sénégalais

- Geo Senegal Infrastructure nationale de gestion des données géographiques: <https://geosenegal.sec.gouv.sn/>
- ANADS : <http://anads.ansd.sn/index.php/catalog/229>).
- Geoportail service climatique Sénégal : <https://retd1.teledetection.fr/climap/>
- Sénégal le portail de données <https://senegal.opendataforafrica.org/>

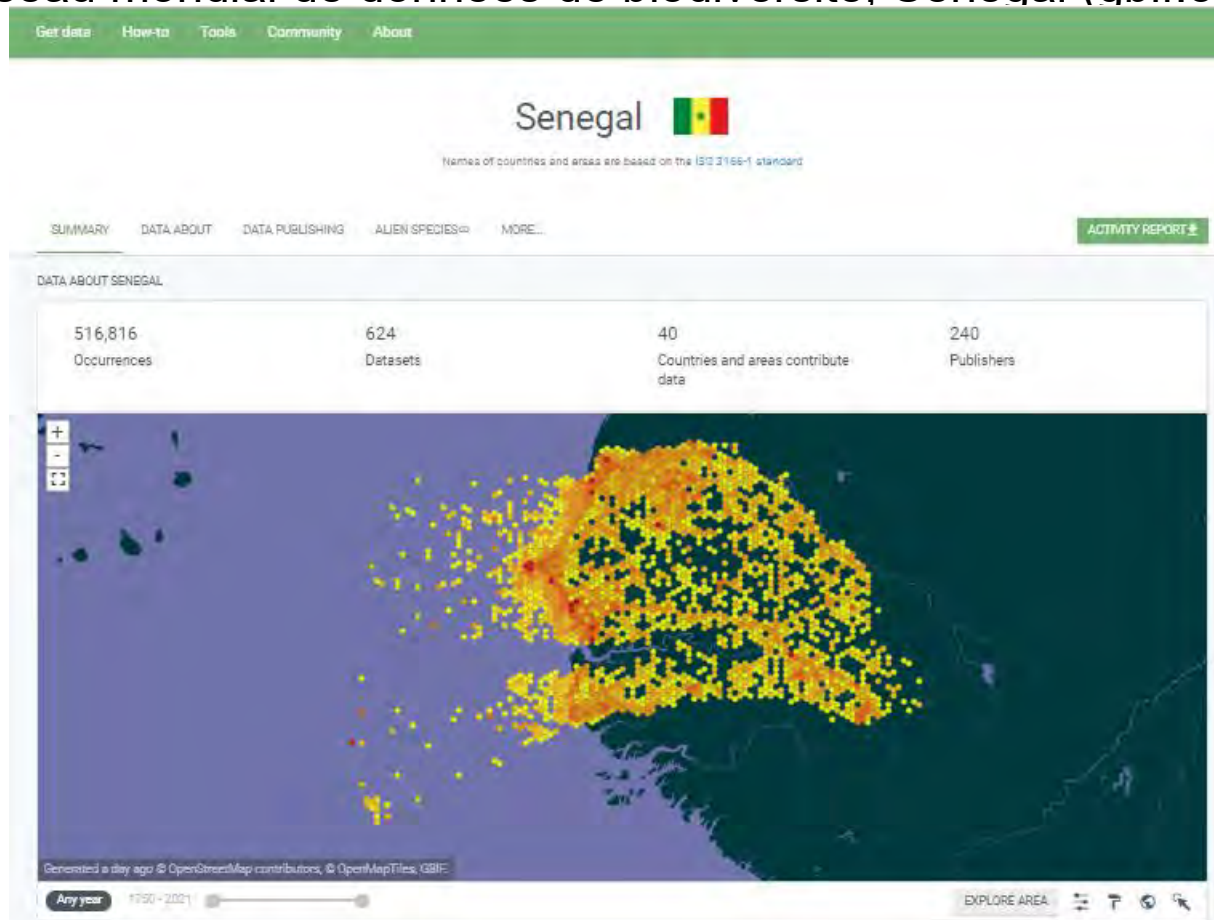
Internationaux

- West African Science Service Centre on climate change and adapted land use Wascal Data Repository and portals (<https://wascal.org/wascal-data-repository-and-portals/>)
- - [Human Heredity and Health in Africa](https://h3africa.org/) (H3Africa) : (<https://h3africa.org/>)
- - INDEPTH Data repository (<https://www.indepth-ishare.org/index.php/home>)
- WorldPop Open Population Repository (<https://woopr.worldpop.org/>)
- BAOBAB (<https://baobab.sedoo.fr/>)

Entrepôt GBIF focus Sénégal

GBIF | Global Biodiversity Information Facility

- Réseau mondial de données de biodiversité, Senegal (gbif.org)





Registre d'entrepôts de données
 Recense 2749 entrepôts de données
 dans le monde
 Recherche par filtre

Filter

[Reset all](#)

- Subjects** ⊞
- Content Types** ⊞
- Countries** ⊞
 - Benin (1)
 - Burkina Faso (1)
 - Denmark (1)
 - European Union (1)
 - Germany (1)
 - Senegal (1)**
 - United Kingdom (1)
- Data access** ⊞
- Data access restrictions** ⊞
- Database access** ⊞
- Data licenses** ⊞
- Data upload** ⊞
- Data upload restrictions** ⊞
- Enhanced publication** ⊞
- Institution responsibility type** ⊞
- Institution type** ⊞

Search...

Search

[Toggle short help](#)

← Previous **1** Next →

Sort by ▾

Found 1 result(s)

West African Vegetation

a vegetation data network for West Africa



Subject(s)

- Biology
- Plant Systematics and Evolution
- Plant Ecology and Ecosystem Analysis
- Ecology of Agricultural Landscapes
- Soil Sciences
- Life Sciences
- Plant Sciences
- Agriculture, Forestry, Horticulture and Veterinary Medicine
- Agriculture, Forestry, Horticulture and Veterinary Medicine

Content type(s)

- Raw data
- Scientific and statistical data formats
- Structured graphics

Country

- Germany
- European Union
- Denmark
- United Kingdom
- Burkina Faso
- Senegal
- Benin

This is a database for vegetation data from West Africa, i.e. phytosociological and dendrometric relevés as well as floristic inventories. The West African Vegetation Database has been developed in the framework of the projects "Sustainable Use of Natural Vegetation in West Africa" (SUN, <http://www.sunproject.dk/>) and "Biodiversity Transect Analysis in Africa" (BIOTA, <http://www.biota-africa.org/>).

ResearchGate



Data

Woody and Herbaceous Vegetation Change across the Savannas of West Africa, 1982-2013

May 2020

DOI: [10.3334/ORNLDAAC/1738](https://doi.org/10.3334/ORNLDAAC/1738)

Project: [Africa Carbon Project](#)

Julius Anchang · Lara Prihodko · Armel Kaptue · [Show all 10 authors](#) · Niall P Hanan

Research Interest ⓘ 0.8

Citations 0

Recommendations 1

Reads ⓘ 21

[See details](#)

Overview

Stats

Comments

Citations

References



Request file

Share ▾

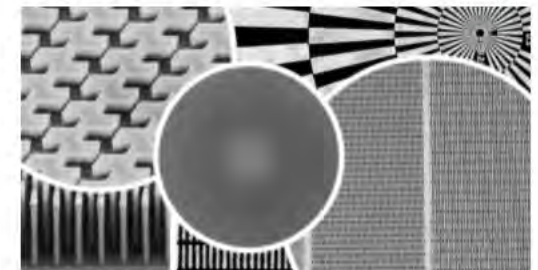
More ▾

Description

The WAVEtrends dataset is a 0.05 degree (5.55 km) vegetation change product, spanning the West African Sudano-Sahel region. It provides pixel-wise information on concurrent woody and herbaceous vegetation trends over a 32-year period (1982-2013). Change in woody vegetation was derived using long-term rain use efficiency (RUE) sensitivity, i.e., the per-pixel comparison of the difference of mean RUE between the first and last decades of the 32-year time series. Herbaceous vegetation change was defined by short-term RUE sensitivity, i.e., comparing the slope of the RUE relationship (productivity vs. precipitation) between both decades using per-pixel Analysis of Covariance (ANCOVA). Categorical vegetation change was then determined for each pixel using the direction of the change and a significance level of $p < 0.05$. The use of RUE (the amount of biomass produced per unit of precipitation) for vegetation trend analysis in savanna regions relies on the assumption that rainfall is a significant

XRnanotech GmbH

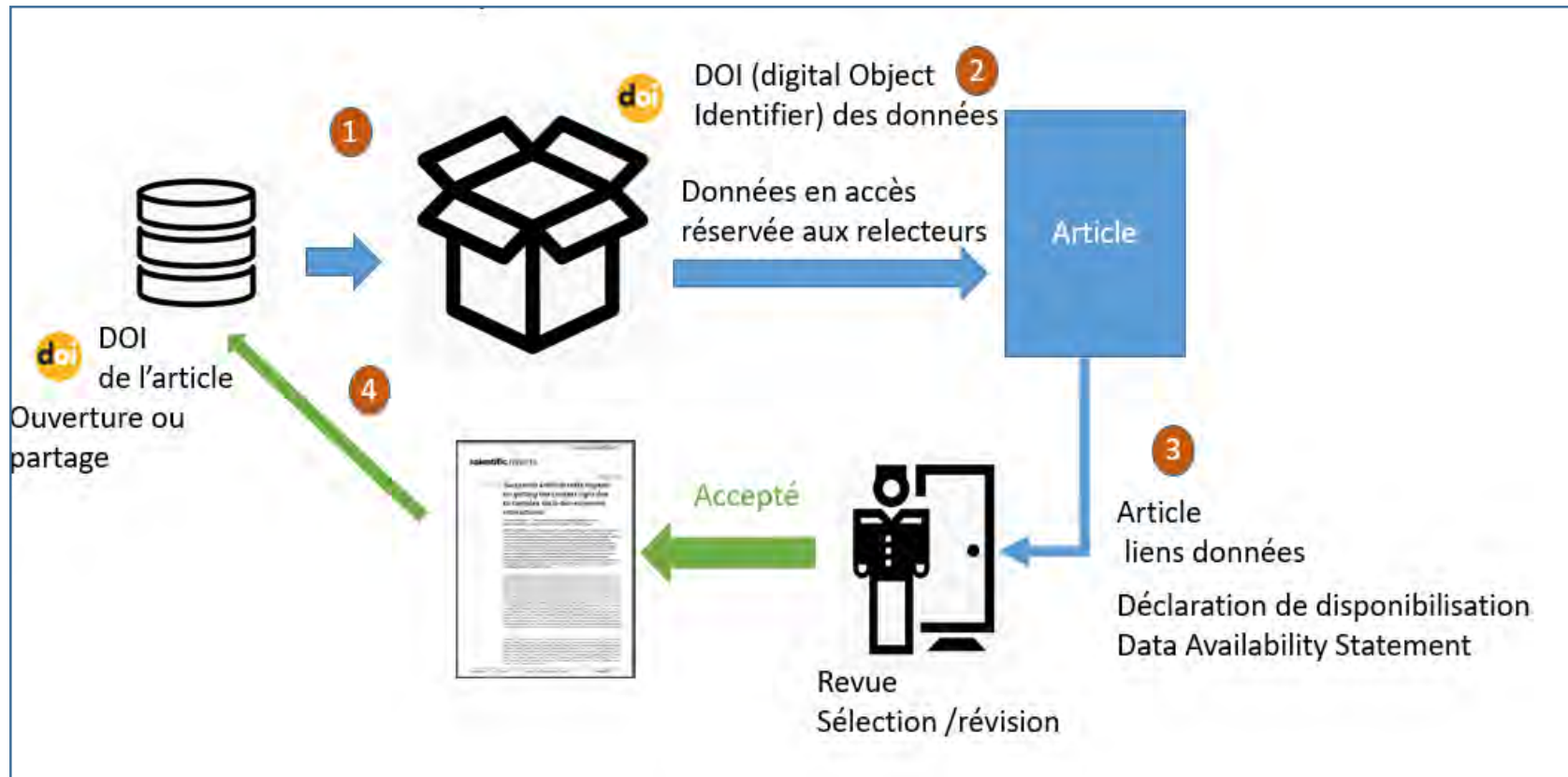
Ad ▾



Nanostructured optics with Swiss precision

XRnanotech offers innovative solutions in nanofabrication e.g. gratings, zone plates, test targets, fab services & more.

Circuit de dépôt d'un article et des données



l'icône : The noun project <https://thenounproject.com/>

Article – données exemple de PNAS

● LETTER

Replications provide mixed evidence that inequality moderates the association between income and generosity

Stéphane Côté^{a,1} and Robb Willer^{b,c,d}

^aRotman School of Management, University of Toronto, Toronto, ON M5S 3E6, Canada; ^bDepartment of Sociology, Stanford University, Stanford, CA 94305; ^cDepartment of Psychology, Stanford University, Stanford, CA 94305; and ^dGraduate School of Business, Stanford University, Stanford, CA 94305

Author contributions: S.C. and R.W. designed research; S.C. performed research; S.C. analyzed data; and S.C. and R.W. wrote the paper.

The authors declare no competing interest.

Published under the [PNAS license](#).

Data deposition: Supplementary information, including references and data, has been deposited at Harvard Dataverse (<https://dataverse.harvard.edu/loginpage.xhtml?redirectPage=%2Fdataset.xhtml%3FpersistentId%3Ddoi%3A10.7910%2FDVN%2FIDJAWT>).

¹To whom correspondence may be addressed. Email: Scote@rotman.utoronto.ca.

www.pnas.org/cgi/doi/10.1073/pnas.1918979117

PNAS

<https://www.pnas.org/content/pnas/early/2020/03/30/1918979117.full.pdf>

Données - article exemple de PNAS

CIRAD Dataverse > UMR TETIS >

Sénégal – Données entomologiques du projet VMERGE 2014-2016

Version 1.1



Fall, Assane Gueye; Biteye, Biram; Ciss, Mamadou; Seck, Momar Talla; Apolloni, Andrea; Fall, Moussa; Tran, Annelise; Gimonneau, Geoffrey, 2019, "Sénégal – Données entomologiques du projet VMERGE 2014-2016", <https://doi.org/10.18167/DVN1/LINKQD>, CIRAD Dataverse, V1, UNF:6:E4qG20NvrCTE/jwalyQhTg== [fileUNF]

Cite Dataset +

Learn about Data Citation Standards.

Access Dataset ▾

Contact Owner

Share

Dataset Metrics ⓘ

28 Downloads ⓘ

Description ⓘ

Collections entomologiques de moustiques réalisées dans le cadre du projet VMERGE, Sénégal, 2014 - 2016
Mosquito collections conducted in the framework of VMERGE project, Senegal, 2014 - 2016 (2019-04-12)

Subject ⓘ

Agricultural Sciences

Keyword ⓘ

culicidae, Rift Valley fever virus, Sénégal, entomologie, écologie

Related Publication ⓘ

Biteye B., Fall A.G., Ciss M., Seck M.T., Apolloni A., Fall M., Tran A., Gimonneau G. 2018. Ecological distribution and population dynamics of Rift Valley fever virus mosquito vectors (Diptera, Culicidae) in Senegal. *Parasites and Vectors*, 11:10 p. doi: 10.1186/s13071-017-2591-9

Files

Metadata

Terms

Versions

<https://doi.org/10.18167/DVN1/LINKQD>

Merci pour votre attention

Plus d'informations sur le site d'OPSE

<https://lped.info/wikiObsSN/?OpSe>

Pascal Aventurier – IRD pascal.aventurier@ird.fr

Sophie Fortuno – CIRAD sophie.fortuno@cirad.fr