



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*



Normandie Université



Les entrepôts de données de recherche

17/03/2025



MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
ET DE LA RECHERCHE

*Liberté
Égalité
Fraternité*

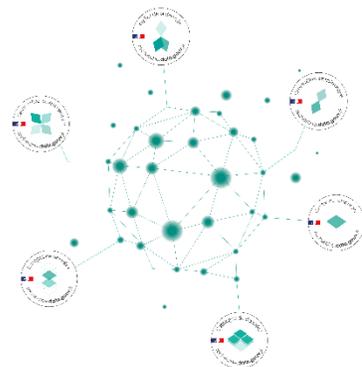
 **Ouvrir
la science !**



Recherche Data Gouv, du national au local

L'échelon national

L'écosystème Recherche Data Gouv



Accompagnement des équipes :

- Ateliers de la donnée
- Centres de référence thématiques
- Centres de ressources
- Centres de référence établissements

Dépôt, publication et signalement des données

- Un entrepôt
- Un catalogue

L'échelon régional

L'Atelier de la donnée en Normandie (ADN)*



Offre de formation normande sur les données de la recherche :

- Sensibilisation aux données de la recherche
- Diffuser ses données grâce aux *data papers*
- Les entrepôts de données de recherche

Organisation de journées d'études normandes sur les données de la recherche

L'échelon local

Les guichets des établissements



Accompagnement de proximité :

- Accompagnement dans la bonne gestion des données
- Relecture des plans de gestion de données
- Aide au dépôt et au signalement des données

Sommaire

1. Introduction
2. Qu'est-ce qu'un entrepôt ?
3. Comment choisir un entrepôt ?
4. Les questions à se poser avant de déposer des données dans un entrepôt
5. Exemples d'entrepôts
 - a) *Focus sur l'entrepôt Recherche Data Gouv*
6. Chercher un jeu de données dans sa discipline
 - a) *OpenAire*
 - b) *BASE*
 - c) *Dataset Search (Google)*
 - d) *Isidore*
7. Quelle réutilisation des données ?
8. Citer des jeux de données



Définitions des données de la recherche

- › Les données de la recherche sont des « **enregistrements factuels (chiffres, textes, images et sons)**, qui sont **utilisés comme sources principales pour la recherche scientifique** et sont généralement **reconnus par la communauté scientifique comme nécessaires pour valider des résultats de recherche** ».

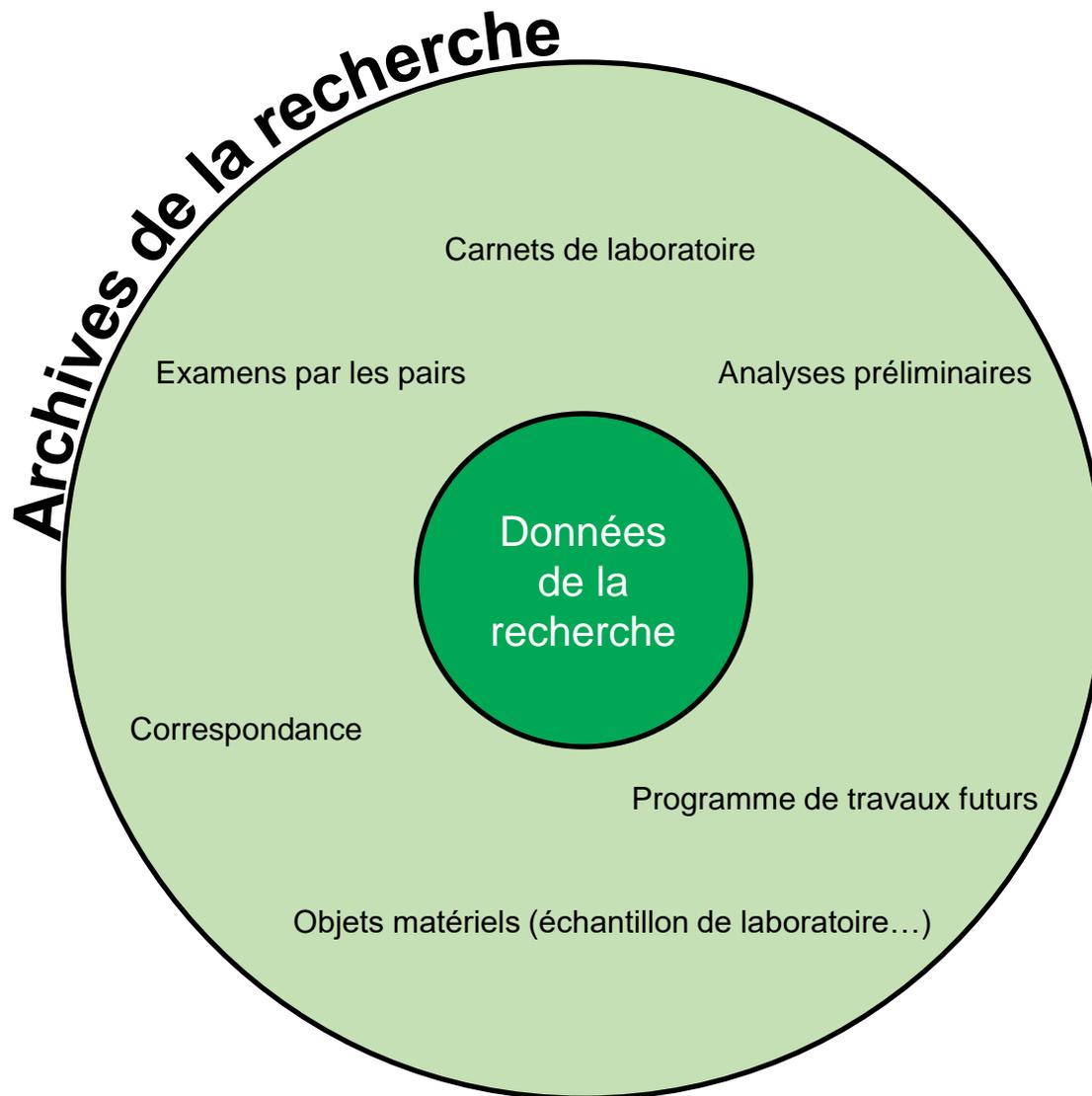
[OCDE \(2007\)](#)

- › « Les données de la recherche sont des **informations, spécimens et matériaux produits, recueillis et documentés**. Elles sont **collectées ou exploitées à des fins de recherche et de preuves par les chercheurs et leurs équipes**. À ce titre, elles **constituent une partie des archives de la recherche**. ».

[Association des Archivistes Français \(2014\)](#)



Des données aux archives



Les entrepôts de données



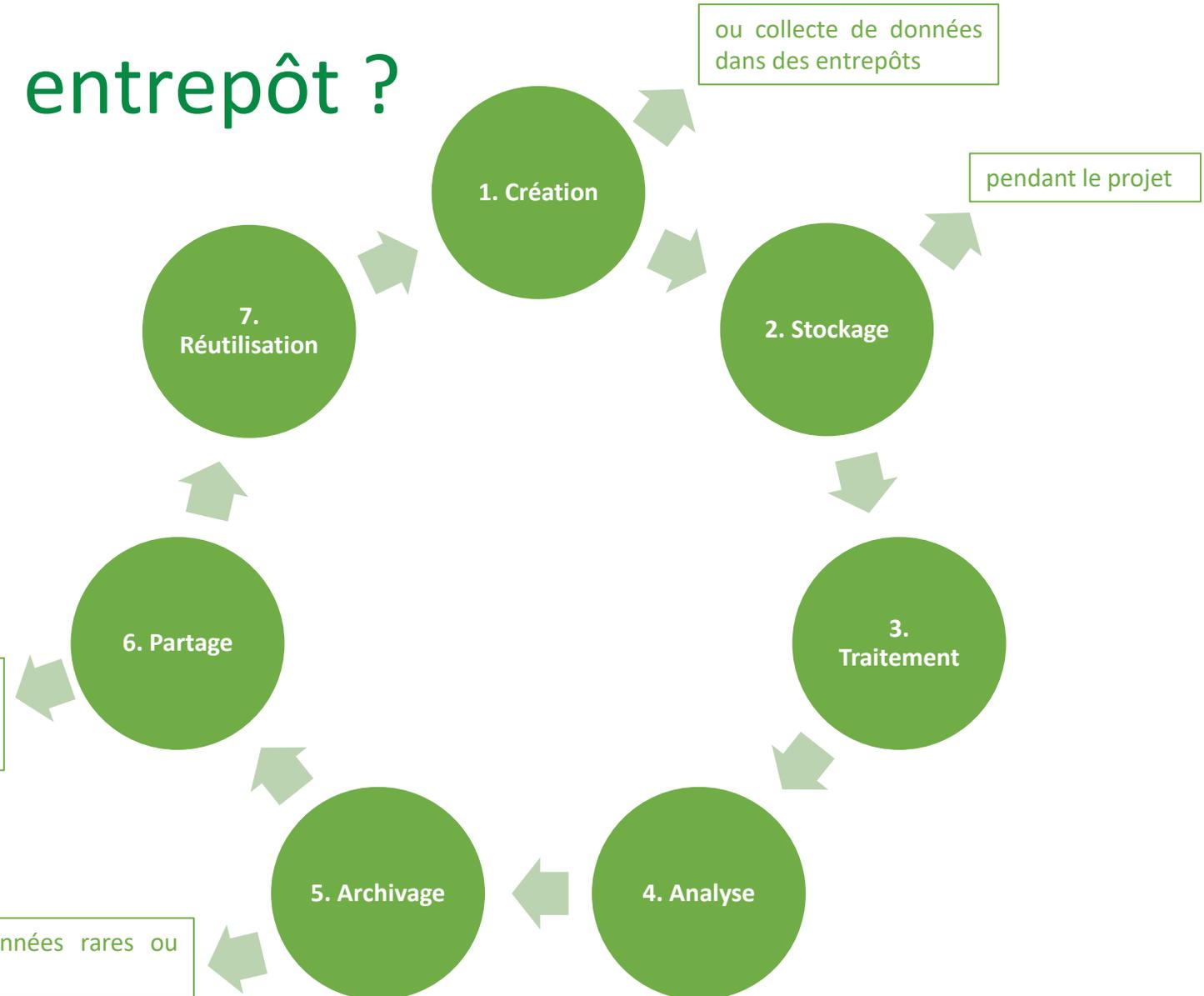
Qu'est-ce qu'un entrepôt ?

Le **Partage dans un entrepôt** est l'une des étapes du **cycle de vie des données** :

- À bien distinguer du **Stockage**
- Et de l'**Archivage**

à l'issue du projet, les données achevées doivent être immédiatement communicables (sauf exceptions légales) dans un **entrepôt**

conservation à long terme pour les données rares ou précieuses





Qu'est-ce qu'un entrepôt ?

- › « Un entrepôt de données (*Data repository, digital repository*) est un réservoir constitué majoritairement de données de recherche, brutes ou élaborées, qui sont décrites par des métadonnées de façon à pouvoir être retrouvées. » ([Pascal Aventurier, 2013](#))
- › Un entrepôt de données est « une infrastructure de stockage et de services facilitant le dépôt, la description, le partage en accès ouvert, la découverte et la réutilisation, par des humains ou des machines, de jeux de données [...]. Ces jeux de données sont associés à des métadonnées et sont conservés à moyen ou long terme. » ([Comité pour la science ouverte, 2024](#))



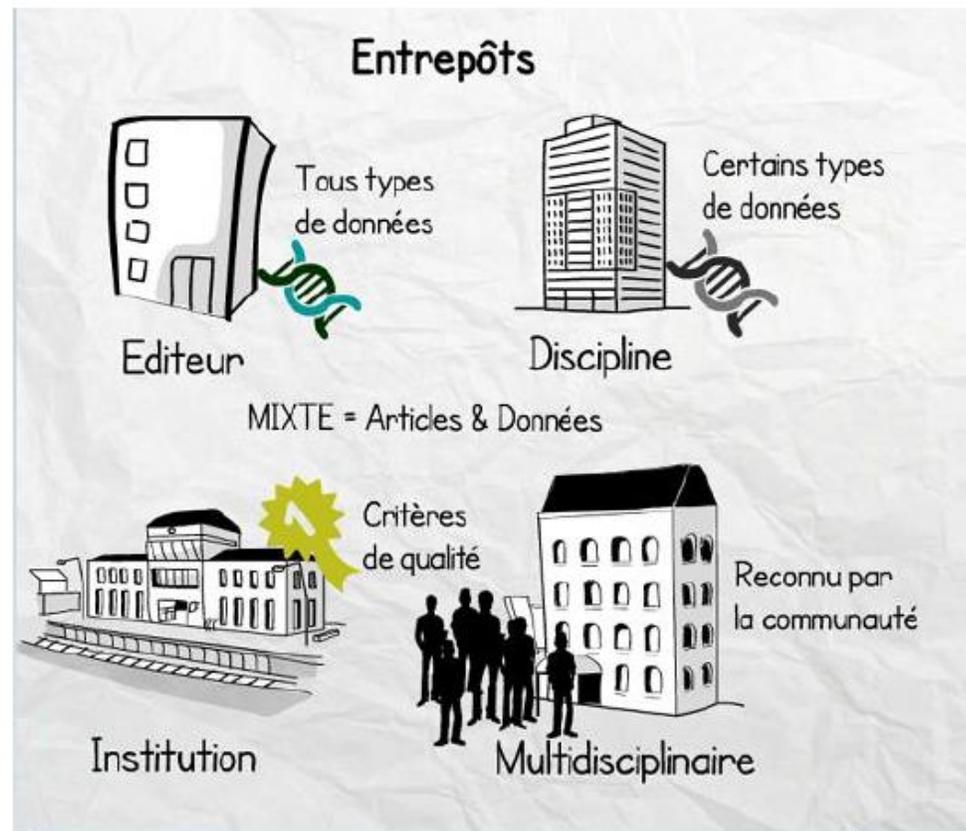
Qu'est-ce qu'un entrepôt ?

Pourquoi déposer dans un entrepôt ? Cela comporte de nombreux avantages.

- › Conserver ses données dans un environnement sécurisé
- › Donner de la visibilité aux données et faciliter l'accès pour les moteurs de recherche
- › Permettre l'interopérabilité des données grâce à l'utilisation de standards de métadonnées
- › Permettre la découverte, la réutilisation et la citation du jeu de données grâce à l'attribution d'un identifiant pérenne
- › Gérer les modalités de partage des données par l'attribution de licences de diffusion
- › Se conformer aux exigences des financeurs et institutions sur l'ouverture des données
- › Favoriser la reproductibilité de la recherche et son intégrité
- › Valoriser les données par leur réutilisation dans de nouvelles études et innovations

Qu'est-ce qu'un entrepôt ?

› Il existe plusieurs catégories d'entrepôts de données



Source : [La minute entrepôts](#) (Doranum)

➤ Comment choisir un entrepôt ?

Un contexte marqué par une offre en pleine croissance

Nombre d'entrepôts recensés en 2011 : **1895**

Nombre d'entrepôts recensés en 2023 : **6025**

(dont 163 en France)

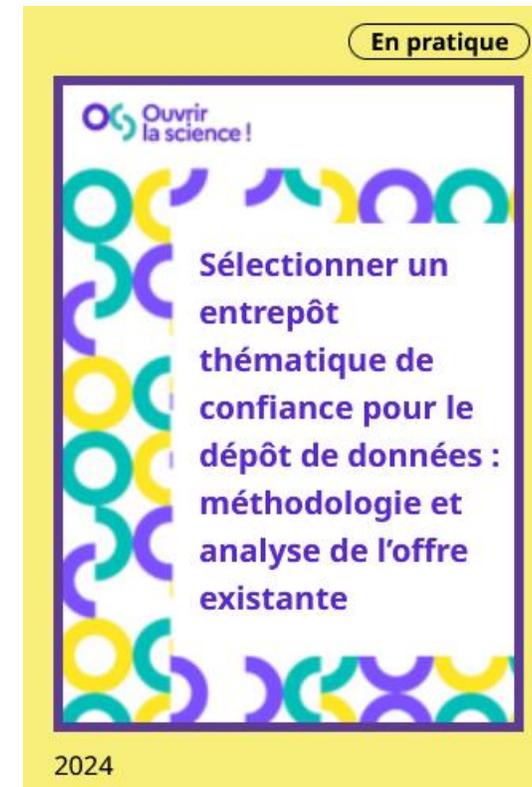
chiffres OpenDOAR

➤ Comment choisir un entrepôt ?

Des conseils précis [sur le portail Science ouverte normand](#), reprenant les recommandations Recherche Data Gouv

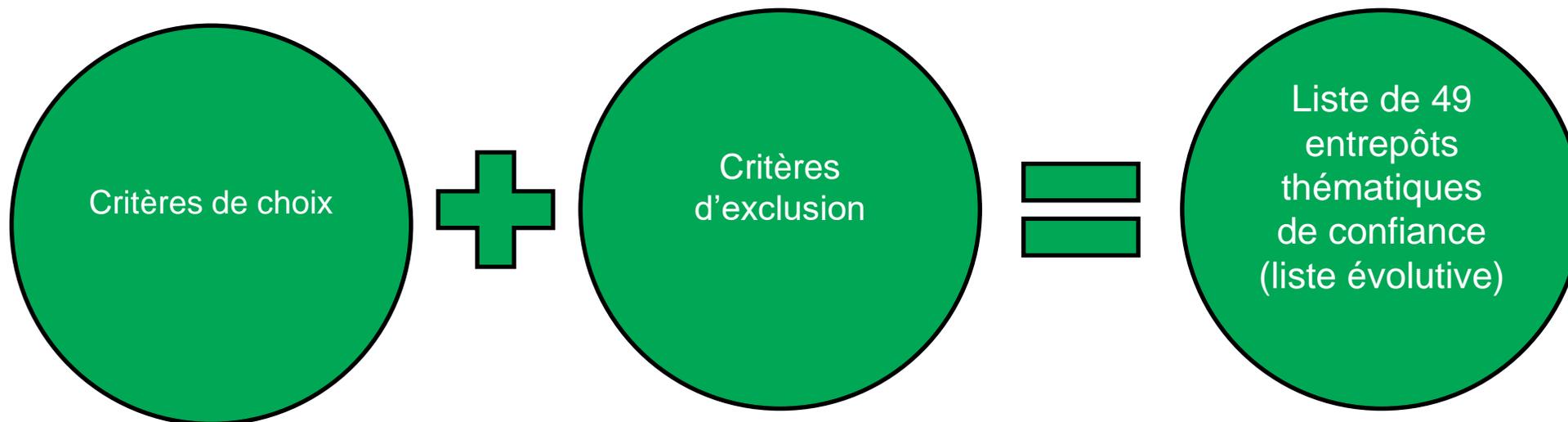
- Privilégier un **entrepôt thématique de confiance**
- Sans entrepôt thématique de confiance ou pour les projets interdisciplinaires, préférer **l'entrepôt Recherche Data Gouv**
- Eviter les entrepôts qui opèrent une **captation commerciale** des données de la recherche publiques

En cas de doute sur le choix à opérer, prenez contact avec nous.





Comment choisir son entrepôt ?



La liste est à retrouver sur le site du Comité pour la Science Ouverte (COSO) :
<https://www.ouvrirlascience.fr/selectionner-un-entrepot-thematique-de-confiance-pour-le-depot-de-donnees-methodologie-et-analyse-de-loffre-existante/>

Une présentation thématique est également disponible sur Recherche Data Gouv :
<https://recherche.data.gouv.fr/fr/entrepots>



Comment choisir un entrepôt non certifié : liste des critères de choix

Champ
disciplinaire

Données
acceptées

Identifiant
pérenne
fourni par
l'entrepôt

Pérennité des
données

Type de
modération

Possibilité
d'embargo

Limite de
volume



Comment choisir un entrepôt non certifié : liste des critères d'exclusion

Absence de
modération
des dépôts

Absence
d'identifiant
pérenne

Absence de
garantie sur la
pérennité de
l'infrastructure

Entrepôts
pratiquant la
cession de
droits

Politique
tarifaire
excessive

Localisation
des données
hors Union
européenne
pour certains
types de
données

Dépôt
restreint par
l'affiliation
institutionnelle

➤ Comment choisir un entrepôt ?



Un gage de confiance : la certification CoreTrustSeal



Les principes :

- Des critères généraux applicables à l'ensemble des entrepôts de données (généralistes ou disciplinaires) et définis par des experts
 - Une auto-évaluation
- Une évaluation externe par des experts indépendants

L'évaluation :

- 3 thèmes ; 16 critères
1. Infrastructure organisationnelle
 2. Gestion des objets numériques
 3. Technologie

Pour en savoir plus : <https://www.coretrustseal.org/> et <https://www.rd-alliance.org/groups/rda-france/forum/topic/coretrustseal-criteres-de-conformite/>

➤ Comment choisir un entrepôt ?

Des répertoires pour vous aider à trouver (et à choisir)





Comment choisir un entrepôt ? : Cat OPIDoR

Il s'agit d'un catalogue des services dédiés aux données de la recherche, il est collaboratif puisqu'il peut être actualisé par ses utilisateurs.



Il se limite aux services français.

En particulier la rubrique consacrée aux entrepôts de données qui permet de naviguer :

- Par ordre alphabétique,
- Par domaine,
- Par thématique,
- Par localisation
- Par stade du cycle de vie.

Et trouver l'entrepôt qui vous convient.



Comment choisir un entrepôt ? : Re3Data

Re3data, pour Registry of Research Data Repositories, est un annuaire de dépôts de données à but non lucratif créé par plusieurs établissements universitaires et de recherche allemands.

Re3Data offre plusieurs possibilités de recherche pour trouver l'entrepôt le plus pertinent possible, entre autres :

- Par discipline
- Par identifiant pérenne
- Par types de données
- Par choix de la licence





Comment choisir un entrepôt ? : Re3Data

re3data.org

Search Browse ▾ Suggest Resources ▾ Contact

- Filter
- Subjects ▾
 - Content Types ▾
 - Countries ▾
 - AID systems ▾
 - API ▾
 - Certificates ▾
 - Data access ▾
 - Data access restrictions ▾
 - Database access ▾
 - Database access restrictions ▾
 - Database licenses ▾
 - Data licenses ▾
 - Data upload ▾
 - Data upload restrictions ▾
 - Enhanced publication ▾
 - Institution responsibility type ▾
 - Institution type ▾
 - Keywords ▾
 - Metadata standards ▾
 - PID systems ▾
 - Provider types ▾
 - Quality management ▾
 - Repository languages ▾
 - Software ▾
 - Syndications ▾
 - Repository types ▾
 - Versioning ▾

Search... Search [Toggle short help](#)

← Previous **1** 2 3 4 5 6 7 ... 126 Next →

Sort by ▾

Found 3131 result(s)

Unidata's RAMADDA
Unidata's Motherlode Repository

Subject(s) Geosciences (including Geography) Natural Sciences

Content type(s) Scientific and statistical data formats Source code Databases Images Structured graphics

Country United States

Our mission is to provide the data services, tools, and cyberinfrastructure leadership that advance earth-system science, enhance educational opportunities, and broaden participation. Unidata's main RAMADDA server contains access to a variety of datasets including the full IDD feed, Case Studies and other project data. RAMADDA is a content management system for Earth Science data. More information is available here: <https://ramadda.org/>

Visual Arts Data Service
VADS

Subject(s) Fine Arts, Music, Theatre and Media Studies Humanities Humanities and Social Sciences

Content type(s) Audiovisual data Images other

Country United Kingdom

VADS is the online resource for visual arts. It has provided services to the academic community for 12 years and has built up a considerable portfolio of visual art collections comprising over 100,000 images that are freely available and copyright cleared for use in learning, teaching and research in the UK. VADS provides: expert guidance and help for digital projects in art education; resource development and hosting for art education; project management and consultancy for art education; leadership in the innovative use of ICT in education through its research and development activities. VADS offers advice and guidance to the visual arts research, teaching and learning communities on all aspects of digital resource management from funding, through delivery and use, to preservation.

ZACAT
GESIS Online Study Catalogue

Subject(s) Social and Behavioural Sciences Humanities and Social Sciences



Les questions à se poser avant de déposer des données dans un entrepôt

- Quelle est la valeur scientifique des données et leur potentiel de réutilisation ?
- Avez-vous évalué le temps et l'effort nécessaires à la mise en forme des données et des métadonnées pour répondre aux exigences de l'entrepôt ?
- Quelles sont les obligations d'ouverture des données qui s'appliquent ?
- Avez-vous le droit de rendre publiques ces données ?
- Avez-vous obtenu l'accord de tous les contributeurs ?
- Avez-vous défini les conditions de réutilisation des données que vous avez produites ?

➤ Quelques exemples d'entrepôts

Quelques exemples d'entrepôts thématiques



Doranum – [exemples d'entrepôts de données en SHS](#)

Quelques autres exemples



Cliquez sur l'image ci-dessous afin de consulter les conseils de Couperin sur les entrepôts d'éditeurs :



Les entrepôts de données d'éditeurs commerciaux : quelle stratégie adopter ?



Focus sur l'entrepôt Recherche Data Gouv

Lors du lancement du deuxième plan national pour la science ouverte, il a été annoncé la mise en place d'un entrepôt de données au niveau national « [Recherche Data Gouv](#) ».

L'entrepôt a été inauguré le 8 juillet 2022.

« Recherche Data Gouv a été pensé pour soutenir les équipes de recherche dans leur travail de structuration des données pour les rendre Faciles à trouver, Accessibles, Interopérables, Réutilisables, c'est-à-dire conformes aux principes FAIR ».



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr

Devez-vous utiliser un entrepôt de données disciplinaire ou Recherche Data Gouv ?
Pour savoir où publier vos données, [cliquez ici](#)



Focus sur l'entrepôt Recherche Data Gouv

The screenshot shows the homepage of the Recherche Data Gouv repository. At the top, there is a navigation bar with the French Republic logo and the text 'recherche.data.gouv.fr'. The navigation menu includes 'Recherche', 'À propos', 'Guide d'utilisation', 'Support', 'Français', 'S'inscrire', and 'Se connecter'. A yellow maintenance banner states: 'Maintenance – L'entrepôt Recherche Data Gouv sera indisponible le 05/07/2023 entre 12h00 et 14h00 en raison d'une opération technique.' Below this is a teal banner with the text 'Un écosystème au service du partage et de l'ouverture des données de la recherche' and the subtext 'FÉDÉRER, ACCOMPAGNER, PARTAGER, OUVRIR, RÉUTILISER'. The main content area features the title 'Recherche Data Gouv' and 'Génération datapaper', followed by '(Recherche Data gouv)'. A statistics bar shows '610 821 téléchargements' and includes 'Statistiques', 'Contact', and 'Partager' links. The main text describes the repository as a sovereign solution for sharing and opening up data produced by communities that do not have a recognized disciplinary repository. It is based on the Dataverse software. Data should be deposited in the space assigned to an institution which one of the contributor belongs to. A generic space is available for data produced by institutions which do not have such a space yet themselves. Tests should be performed in the sandbox. The latest news and events on the Recherche Data Gouv platform are mentioned. At the bottom, there is a carousel of logos for partner institutions: 'Espace Générique', 'Arts et Métiers Institute of Technology', 'cirad Centre de Coopération Internationale en Recherche Agronomique pour le Développement', and 'INRAE Data INRAE'. A search bar at the bottom left contains the text 'Chercher dans cette collection...' and a magnifying glass icon, followed by a 'Recherche avancée' link.



Chercher un jeu de données dans sa discipline





Chercher un jeu de données dans sa discipline : Open AIRE

[OpenAIRE explore](#) est une plateforme européenne qui indexe les références des publications et des données de recherche d'entrepôts ou de sites web de fournisseurs d'information.

[OpenAIRE explore](#) contient des données de recherche issues de plus d'une centaine de fournisseurs :

- Agence Datacite d'enregistrement de DOI de jeux de données déposés dans des entrepôts de toutes sortes)
- Entrepôts :
 - pluridisciplinaires (Figshare...)
 - thématiques (Omics Discovery Index, Pangaea...)
 - publics internationaux (Zenodo...)
 - institutionnels (Dataverse du Cirad...)
 - d'éditeur commercial (Mendeley Data d'Elsevier)





Chercher un jeu de données dans sa discipline : Open AIRE

OpenAIRE | EXPLORE

Search Deposit Link Data sources

Sign in

Advanced search

Search by title, author, abstract, DOI, orcid...

RESEARCH PRODUCTS (230,000,395) PROJECTS (3,298,154) DATA SOURCES (125,987) ORGANIZATIONS (198,272)

Filters

Access (100)

- Open Access (77,135,915)
- Closed Access (38,174,695)
- Restricted (2,631,782)
- UNKNOWN (94,109)
- Open Source (84,227)
- Embargo (67,645)

[View all >](#)

Type (4)

- Publications
- Research data
- Research software
- Other research products

Year range

1800 - 2033 >

[This year](#) | [Last 5 years](#) | [Last 10 years](#)

230,000,395 Research Products [DOWNLOAD RESULTS](#)

Results per page: 10 | Sort by: Relevance | 1 2 3 4 5 >

A Reassessment of the Genus Strombidinopsis (Ciliophora, Choreotrichida) with Descriptions of Four New Planktonic Species and Remarks in its Taxonomy and Phylogeny

[Publication](#) » Article • 1991 • Cambridge University Press (CUP)

Denis H. Lynn; David J. S. Montagnes; T. Dale; G. L. Gilron; +1 Authors

DOI: 10.1017/s0025315400053170

Strombidinopsis species were collected from the Gulf of Alaska, Gulf of Mexico, Caribbean Sea, Gulf of Maine, and Barents Sea. Six strains were conservatively placed in six species: Strombidinopsis chiorhax sp. nov.; Strombidinopsis batos sp. nov.; Strombidinopsis sphaira sp. nov.; Strombidinopsis cercionis sp. nov. ...

[Journal of the Marine...](#) | [Link to](#) | [Share](#) | [Cite](#) | [Claim](#) | 44

Аксиологические основания применения искусственного интеллекта

[Publication](#) » Article • 2020 • Zenodo

DOI: 10.5281/zenodo.3830980

В статье рассмотрен философский аспект развития государственных программ цифровизации разных стран. Дана характеристика категории «искусственный интеллект». Показаны особенности внедрения слабого искусственного интеллекта в социальные системы, в частности, в России, Китае и странах Запада. На примере...

[ZENODO](#) | [Link to](#) | [Share](#) | [Cite](#) | [Claim](#) | 0

[HELP](#)

<https://explore.openaire.eu/search/find/datasets>





Chercher un jeu de données dans sa discipline : BASE

BASE (Bielefeld Academic Search Engine) est l'un des moteurs de recherche les plus puissants au monde.

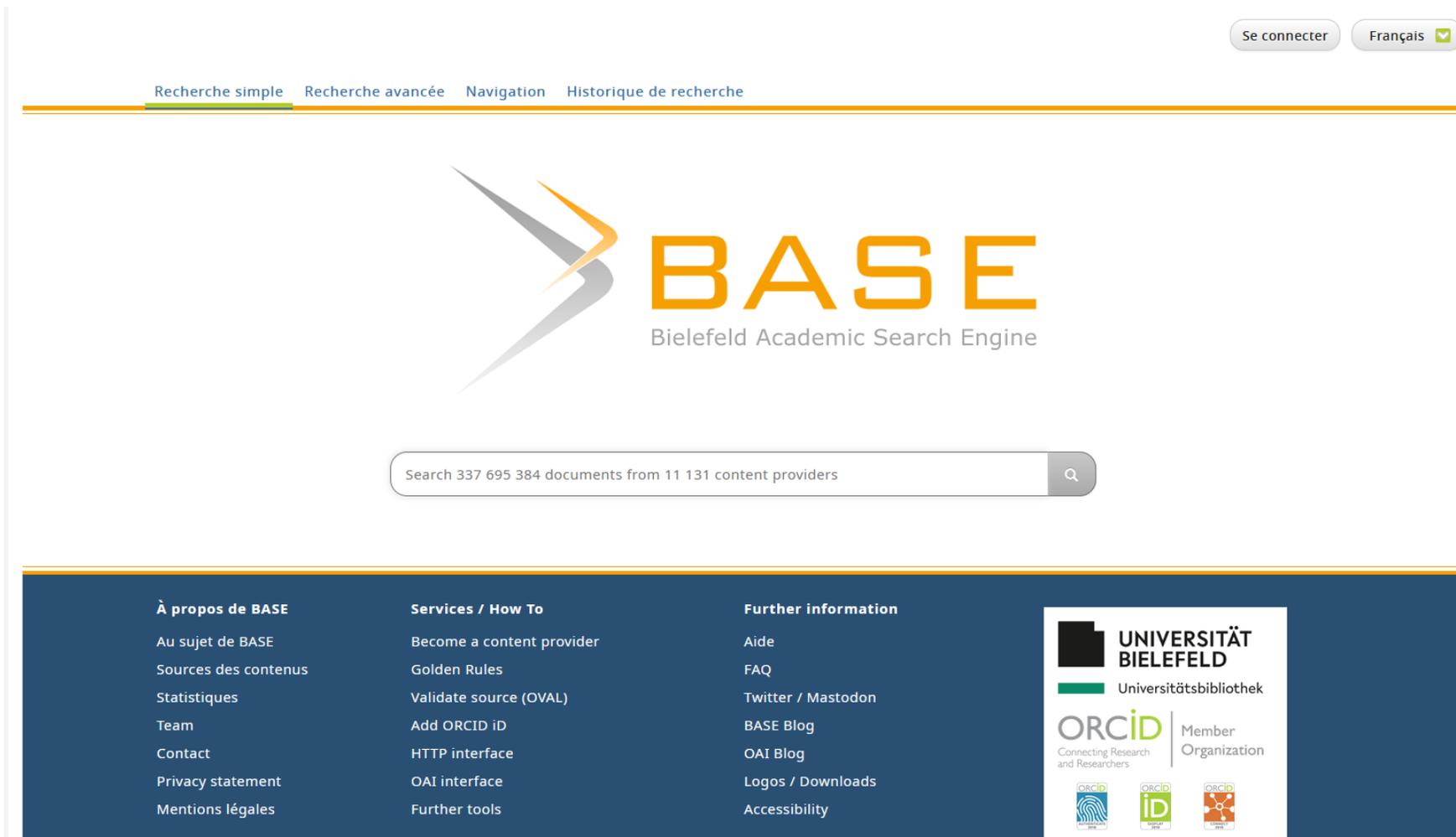
Parmi les sources d'information : près de 200 entrepôts de données internationaux, nationaux, institutionnels, publics ou privés, parmi lesquels DataCite Metadata Store, Zenodo, Pangaea, Figshare...

L'index est [continuellement enrichi](#) par l'intégration de nouvelles sources OAI ainsi que par des sources locales.





Chercher un jeu de données dans sa discipline : BASE



The screenshot shows the homepage of the BASE (Bielefeld Academic Search Engine) website. At the top right, there are buttons for "Se connecter" and "Français" with a dropdown arrow. Below this is a navigation menu with links for "Recherche simple", "Recherche avancée", "Navigation", and "Historique de recherche". The main content area features the BASE logo, which consists of a stylized orange and grey arrow pointing right, followed by the text "BASE" in large orange letters and "Bielefeld Academic Search Engine" in smaller grey text below it. A search bar is positioned below the logo, containing the text "Search 337 695 384 documents from 11 131 content providers" and a search icon. The footer is a dark blue section with three columns of links. The first column is titled "À propos de BASE" and includes links for "Au sujet de BASE", "Sources des contenus", "Statistiques", "Team", "Contact", "Privacy statement", and "Mentions légales". The second column is titled "Services / How To" and includes links for "Become a content provider", "Golden Rules", "Validate source (OVAL)", "Add ORCID iD", "HTTP interface", "OAI interface", and "Further tools". The third column is titled "Further information" and includes links for "Aide", "FAQ", "Twitter / Mastodon", "BASE Blog", "OAI Blog", "Logos / Downloads", and "Accessibility". On the right side of the footer, there are logos for "UNIVERSITÄT BIELEFELD" (Universitätsbibliothek) and "ORCID" (Member Organization), along with smaller ORCID logos at the bottom.



Chercher un jeu de données dans sa discipline : Google Dataset Search

Google Dataset Search est une version du moteur de recherche de Google qui peut être spécifiquement utilisée pour rechercher des ensembles de données dans des domaines scientifiques, sociaux, médicaux etc.

Parmi les sites et entrepôts de données indexés :

- Entrepôts institutionnels (Dataverse du Cirad)
- Entrepôts thématiques (GBIF sur la biodiversité)
- Sites de données administratives (data.gouv.fr)
- Bases de données publiques internationales (OpenAIRE explore)
- Réseaux sociaux (ResearchGate)





Chercher un jeu de données dans sa discipline : Google Dataset Search

Google



Dataset Search

Rechercher des ensembles de données



Essayez [coronavirus covid-19](#) ou [water quality site:canada.ca](#).

Découvrez le projet Recherche d'ensembles de données.

<https://datasetsearch.research.google.com/>



Chercher un jeu de données dans sa discipline : ISIDORE

ISIDORE est un moteur de recherche permettant l'accès aux données numériques des sciences humaines et sociales (SHS).

Il permet de rechercher dans le texte intégral de plusieurs millions de documents (articles, thèses et mémoires, rapports, **jeux de données**, pages Web, notices de bases de données, description de fonds d'archives, etc.), des signalements événements (séminaires, colloques, etc.).

ISIDORE offre également une boîte à outils (ISIDORE à la demande) qui permet — sur demande, d'enrichir, de catégoriser et d'annoter à l'aide de référentiels scientifiques des données de la recherche.





Chercher un jeu de données dans sa discipline : ISIDORE



Votre assistant de recherche
en Sciences Humaines et Sociales

Documents ▾ Rechercher dans les 10 398 919 documents de ISIDORE...

[Recherche avancée](#)

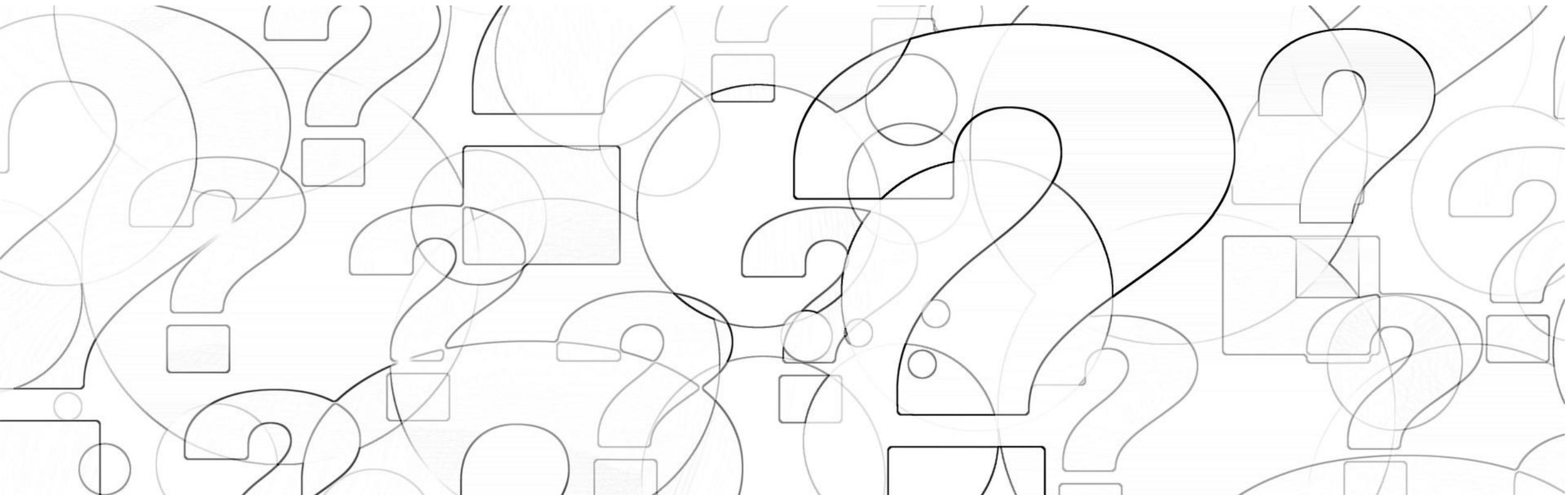
[i démonstration](#)

<https://isidore.science/>





Temps d'échange : des questions ? des suggestions ?



Quelle réutilisation des données ?

Avant de réutiliser les données que vous avez trouvées, vérifiez toujours la licence appliquée à ces données.

Vous pouvez réutiliser librement les données sous licence **Etalab** ou sous licence **CC-BY 4.0** : elles sont dites permissives car la **seule contrainte est de mentionner la paternité et la date de la dernière mise à jour**.

Pour le reste, il est **possible de reproduire, redistribuer, modifier, exploiter à titre commercial**.

Attention donc aux conditions fixées par les licences CC-BY-SA, CC-BY-NC, CC-BY-NC-SA, CC-BY-ND et CC-BY-NC-ND !



Citer des jeux de données

- **Zenodo** : fonction « cite as »
- **Zotero** : le type de document « dataset » est à présent accepté
- Avec un simple DOI, l'outil magique : **DOI Citation Formatter**
<https://citation.crosscite.org/>

Cite as

Banda, Juan M., Tekumalla, Ramya, Wang, Guanyu, Yu, Jingyuan, Liu, Tuo, Ding, Yuning, Artemova, Katya, Tutubalina, Elena, & Chowell, Gerardo. (2023). A large-scale COVID-19 Twitter chatter dataset for open scientific research - an international collaboration [Data set]. In Epidemiologia (Version 162, Vol. 2, Number 3, pp. 315–324). Zenodo. <https://doi.org/10.5281/zenodo.7834392>

Start typing a citation style...





Nous contacter



Normandie Université

<https://science-ouverte.normandie-univ.fr/>



UNIVERSITÉ
CAEN
NORMANDIE

donnees-recherche@unicaen.fr



learningcenter@em-normandie.fr



donnees-recherche@univ-lehavre.fr



scd-chercheurs@univ-rouen.fr



eugenie.gourichon@insa-rouen.fr
christelle.vandepitte@insa-rouen.fr





Merci de votre attention !