

"Aussi ouvert que possible, aussi fermé que nécessaire"



Concept

Bonnes pratiques

CC BY-SA 4.0 -

Jörg Braukmann

Le principe "Aussi ouvert que possible, aussi fermé que nécessaire" met en avant l'importance de l'accès libre aux données de la recherche, tout en reconnaissant la nécessité de certaines restrictions, en particulier en conformité avec le RGPD (Règlement Général sur la Protection des Données) qui impose des règles strictes sur la protection de la vie privée. Ce principe vise à équilibrer ouverture et protection des données, en particulier sur les données dites sensibles et/ou personnelles.

Administrateurs des données, algorithmes et codes (ADAC)

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR
DE LA RECHERCHE
ET DE L'INNOVATION



Acteurs

Concepts

<http://www.numerique.gouv.fr>

La feuille de route 2021-2024 du MESRI sur la politique des données, des algorithmes et des codes sources prévoit la création d'un réseau d'administrateurs des données, des algorithmes et des codes sources de la recherche. Placés auprès de la gouvernance de leur établissement, ils sont des acteurs clé de la gouvernance des données de la recherche.



Lille
Open
Research
Data

DATAACTIVIST

Ateliers de la donnée & LORD

Lille
Open
Research
Data



Acteurs

Accompagnement

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

Au sein de l'écosystème Recherche Data Gouv, les ateliers de la donnée apportent, en proximité des chercheurs, un premier niveau d'expertise et de services à la préparation et la diffusion des données. Lille Open Research Data est l'atelier de la donnée lillois, labellisé depuis 2022. Il regroupe 12 établissements et organismes de recherche.



Lille
Open
Research
Data

DATAACTIVIST

Bonnes pratiques - codes et logiciels



Concept

Code et logiciels

Bonnes pratiques

CC BY-SA 2.0 -
gruntzooki

Les bonnes pratiques en matière de logiciel libre assurent la pérennité et la réutilisation des codes et logiciels. Parmi elles, on retrouve la publication du code source d'un logiciel sous une licence libre reconnue (GNU GPL, MIT...) permettant la modification et la redistribution, le maintien d'une documentation complète et à jour ou encore l'adoption d'un système de gestion de version (comme Git).

Captation de données par les éditeurs scientifiques



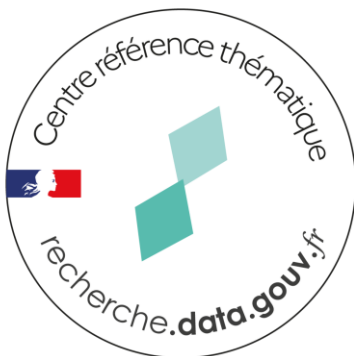
Enjeux

Acteurs

CC BY-SA 3.0 -
Vmenkov

Les éditeurs scientifiques associent de plus en plus les articles à un jeu de données, dans un souci d'intégrité scientifique, de reproductibilité et de transparence. Néanmoins, il arrive que les éditeurs demandent aux auteurs d'utiliser leur propre entrepôt commercial, ce qui pose de sérieux problèmes de souveraineté numérique, de respect du RGPD et de captation de données publiques par des acteurs privés.

Centres de Référence Thématique



Infrastructure

Acteurs

Accompagnement

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

Au sein de l'écosystème Recherche Data Gouv, les Centres de référence thématiques ont une portée nationale et disciplinaire. Ils accompagnent les équipes de recherche en matière de gestion et diffusion des données d'un domaine thématique scientifique. En novembre 2024, elles sont au nombre de 5 :

- Astronomie et astrophysique : Centre de Données astronomiques de Strasbourg
- Système Terre et Environnement : DATA TERRA
- Biologie-Santé : Institut français de bio-informatique
- Sciences Humaines et Sociales : Huma-Num et PROGEDO

Chaire Unesco pour la science ouverte



Temps forts

Institutions

unesco.org

La chaire UNESCO en science ouverte a été créée à la suite des recommandations de l'UNESCO publiées en 2021, et qui préconisaient un accès gratuit et inclusif aux publications scientifiques et données de la recherche. Dirigée par Vincent Larivière, professeur en bibliothéconomie et sciences de l'information à l'Université de Montréal, elle est soutenue dans son travail par des organisations savantes de dix pays. Elle a pour objectif la conduite de recherches sur les pratiques et les effets de la science ouverte sur la production scientifique.

Crise de la reproductibilité



Concept

Enjeux

CC BY 2.0 - bfick

La crise de la reproductibilité désigne un problème croissant dans plusieurs disciplines scientifiques où les résultats de nombreuses études ne peuvent pas être répliqués. Elle est alimentée par divers facteurs (échantillons trop petits, biais de publication, méthodes statistiques inappropriées, manque de transparence dans la méthodologie, fraude...). Elle est souvent invoquée dans la promotion d'initiatives de science ouverte favorisant des pratiques de recherche plus robustes.

Cycle de vie de la donnée



Concept

CC BY-SA - Passeport pour la
Science Ouverte

Le cycle de vie de la donnée représente l'ensemble des étapes par lesquelles une donnée passe de sa création à sa suppression. Ces étapes comprennent en général : la collecte, le stockage, le traitement, l'utilisation, le partage, puis l'archivage ou la suppression définitive. Le cycle de vie assure que la donnée est gérée de manière organisée et sécurisée tout au long de son existence, mais aussi dans sa réutilisation.

Data Paper



Bonnes pratiques

CC BY -
priyanka

Un "data paper" est un type de publication scientifique qui décrit un ensemble de données ainsi que les méthodes de collecte, d'analyse et de traitement utilisées sur ce jeu de données. Contrairement aux articles traditionnels, qui se concentrent sur des résultats et leur interprétation, un "data paper" met l'accent sur la qualité, la provenance et l'accessibilité d'un jeu de données.

DMP OPIDoR



Outil

CC-BY-NC-ND - Inist-
CNRS

DMP OPIDoR est un outil d'aide à la rédaction en ligne de plans de gestion de données (PGD), développé par l'INIST (CNRS). Il donne accès à des modèles de PGD, des recommandations d'établissements, des PGD rendus publics et permet de rédiger directement en ligne son PGD.

Documentation des données



Concept

Bonnes pratiques

Archeologia.chodlik

L'insuffisance de la documentation est un défaut majeur des données ouvertes. Or, pour rendre ses données faciles à trouver, accessibles, interopérables et réutilisables, selon les principes FAIR, il est important de leur associer une bonne documentation (choix de métadonnées pertinentes, écriture d'un readme associé aux données, etc).

Données brutes



Concept

CC0 1.0

Il est essentiel d'ouvrir les données brutes, c'est-à-dire qui n'ont pas été modifiées, agrégées ou manipulées. Aux réutilisateurs après, d'effectuer les calculs et agrégations qu'ils souhaitent pour les rendre intelligibles par des humains.

Par exemple : des fichiers Excel avec des lignes fusionnées ou des totaux devront être transformés en fichiers .csv structurés, avec un enregistrement par ligne.

Données de la recherche



Concept

CC BY-SA - Passeport
pour la science ouverte

"Les données de la recherche sont définies comme 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, « Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics de 2007 »).

Données de la statistique publique



DATA
INFRASTRUCTURE

Acteurs

Exemples d'application

<http://www.progedo.fr>

Les données issues de la Statistique Publique sont constituées de grandes enquêtes et de bases de données issues de l'INSEE, de services statistiques ministériels et des institutions publiques. Elles intéressent particulièrement les sciences sociales

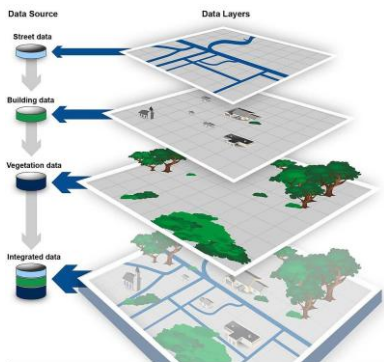
À Lille, la Plateforme Universitaire de Données de Lille (PUDL) accompagne les chercheurs pour l'accès à ces données sur Quetelet-Progedo-Diffusion et pour leur réutilisation.



Lille
Open
Research
Data

DATAACTIVIST

Données géographiques



Source: GAO. | GAO-15-193

Données

Exemples d'application

US Government
Accountability Office

La directive Inspire de 2007 impose aux autorités publiques de recenser les données géographiques, de créer des métadonnées normées et de suivre des standards européens. Le rapport Faure-Muntian en 2018 a rappelé l'importance de données géographiques de qualité pour que l'État exerce ses missions et sa souveraineté en matière de défense, de sécurité, de fiscalité, de protection de l'environnement ou de transport.

Données personnelles et sensibles



Concept

Données

Cadre juridique

CC BY-SA 2.0 -
casey.marshall

Dans le RGPD, les données personnelles sont décrites comme des informations permettant d'identifier directement ou indirectement une personne, comme le nom, l'adresse, l'e-mail, ou les données biométriques. Les données dites sensibles incluent celles révélant l'origine ethnique, les opinions politiques, les croyances religieuses, la santé ou l'orientation sexuelle.

Le traitement de ces deux types de données nécessite des protections renforcées et le consentement explicite de la personne concernée.

DoRANum



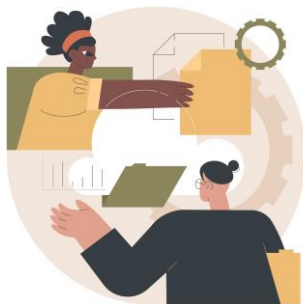
Outil

<https://doranum.fr/>

DoRANum est une plateforme de formation en ligne sur la gestion et le partage des données de la recherche, réalisée par l'Inist-CNRS et le GIS « Réseau Urfist » depuis 2015.

La plateforme propose des ressources généralistes, disciplinaires ou thématiques, retraçant l'ensemble des étapes du cycle de vie des données, ainsi que les enjeux juridiques ou éthiques associés aux données.

Droit d'accès aux documents administratifs

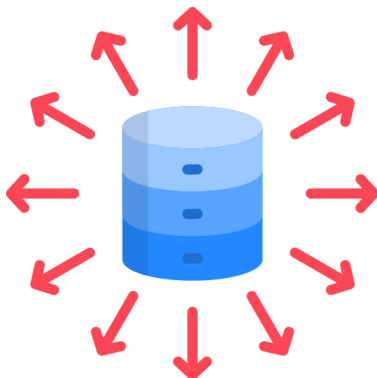


Cadre juridique

Vectorjuice sur Freepik

Prévu par loi CADA de 1978, aujourd'hui codifié dans le code des relations entre le public et l'administration, le droit d'accès concerne tous les documents sous toute forme produits ou reçus dans le cadre d'une mission de service public. Les personnes publiques ou privées sont tenues de les communiquer au public sauf si cela porte atteinte à un secret protégé. L'open data consiste à publier de manière volontaire les données sans attendre une requête, mais il est nécessaire de le demander explicitement aux communes de moins de 3500 habitants.

Entrepôts de données



Infrastructures

Bonnes pratiques

Awicon pour
Flaticon

Un entrepôt de données est une plateforme de stockage numérique qui permet de collecter, gérer et analyser des ensembles de données provenant de diverses sources. Ils peuvent être institutionnels, généralistes ou thématiques. En France, on dispose par exemple de Recherche Data Gouv (et sa déclinaison lilloise : Lillodata) ou encore Nakala pour les données en Sciences humaines et sociales.

De plus en plus de revues scientifiques exigent désormais le partage des données dans ces entrepôts comme condition pour la publication.

Agences de financement de la recherche



Enjeux

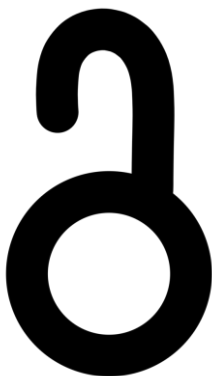
Acteurs

anr.fr

Les agences de financement de la recherche (Europe, ANR) demandent désormais aux porteurs de projets financés de respecter un certain nombre d'obligations liées à la Science ouverte et au data management.

Les données produites dans le cadre d'un projet financé doivent être gérées de façon FAIR, être aussi ouvertes que possible, et faire l'objet d'un Plan de gestion des données.

Format ouvert



Concept

Bonnes pratiques

CC0 1.0

La loi pour une République Numérique consacre l'utilisation de formats ouverts afin de faciliter l'interopérabilité et la réutilisation des données par les machines.

Aussi, pour rendre ses données FAIR, il faudra notamment privilégier des formats de fichiers ouverts ou largement utilisés, par exemple .csv à la place de .xlsx.

Formations



Accompagnement

<https://lord.univ-lille.fr>

Lille Open Research Data propose plusieurs types de formations et ateliers à destination des chercheurs et personnels d'appui à la recherche autour des données, codes et logiciels de la recherche.

Webinaires science ouverte, Formation continue des personnels, Formations doctorales dans le cadre du collège doctoral, formations sur mesure dans les laboratoires, etc.

Identification des plantes



Exemples d'application

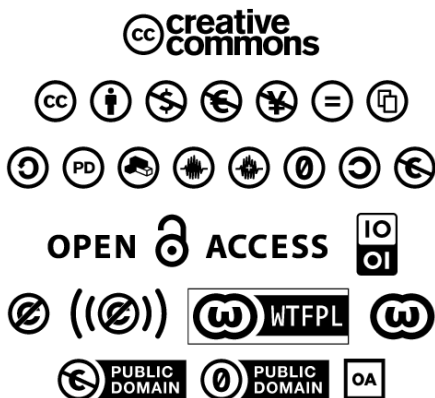
MF

L'open data peut aussi servir dans vos randonnées ou balades en forêt.

Par ex., Pl@ntNet est une application mobile de sciences participatives permettant d'identifier des plantes à partir d'une photo.

Son service est fondé sur plusieurs bases de données ouvertes et les photos prises par les utilisateurs sont reversées en open data pour faciliter l'identification automatisée d'espèces végétales.

Licences ouvertes



Infrastructure

Bonnes pratiques

CC BY - www.alain.les-hurtig.org

Les licences ouvertes sont des outils juridiques permettant aux auteurs de partager leurs œuvres (au sens légal du terme) tout en définissant les conditions de leur utilisation, facilitant ainsi la diffusion des créations tout en respectant le droit d'auteur.

Parmi elles, on retrouve les licences Creative Commons, la Licence Ouverte 2.0 d'Etatlab ou encore la licence ODbL (Open Database Licence). Ces licences permettent notamment le partage des données de la recherche.



Lille
Open
Research
Data

DATAACTIVIST

Lillodata



LILLODATA



Dépôt des données de recherche
produites et collectées par la
communauté scientifique de l'Université
de Lille.

[voir la fiche](#)
.....

Outil

Bonnes pratiques

<https://ent.univ-lille.fr>

Lillodata est l'espace institutionnel de l'Université de Lille sur l'entrepôt Recherche Data Gouv. Lillodata est destiné aux jeux de données produits et collectés par la communauté scientifique de l'Université de Lille.



Lille
Open
Research
Data

DATAACTIVIST

Loi pour une République Numérique



Cadre juridique

Temps forts

Service d'information du
Gouvernement (SIG)

La Loi pour une République Numérique (2016) vise à moderniser la législation française face aux évolutions numériques. Elle impose notamment un principe d'ouverture des données par défaut à toutes les administrations.

Lorsqu'elles correspondent à des données publiques, les données de recherche sont donc soumises à un principe d'ouverture par défaut.

Métadonnées



Concept

Données

Bonnes pratiques

CC BY-SA 2.0 -
quinn.anya

Une métadonnée est une donnée décrivant une autre donnée, facilitant son identification, sa gestion et sa réutilisation. Elle comprend des informations telles que le titre, l'auteur, la date de création, et le sujet d'un document.

Un exemple de schéma de métadonnées (norme de description) est le Dublin Core, qui propose un ensemble standard de 15 éléments permettant de décrire divers types de ressources numériques (documents textuels, images, vidéos, jeux de données...).

L'open data en France



Concept

Domaine public

En France, l'Open Data est une démarche consistant à rendre accessibles et réutilisables les données publiques de manière libre et gratuite. Depuis la loi pour une République numérique de 2016, les administrations ont pour obligation de publier leurs données non personnelles. L'entrepôt gouvernemental data.gouv.fr centralise notamment ces données pour stimuler leur exploitation dans divers domaines et les rendre accessibles aux citoyens et aux entreprises pour leur réutilisation.

Pandémie de Covid-19



Exemple
d'application

Temps forts

Mika Baumeister
sur Unsplash

Début 2020, la nécessité d'une diffusion rapide des connaissances produites pour lutter contre la pandémie a conduit à une adoption massive des données et des preprints, pour permettre aux chercheurs de partager rapidement leurs résultats.

Des préoccupations ont cependant émergé concernant la qualité et la fiabilité des résultats publiés. Ce contexte a accentué les débats sur l'importance de l'accès libre, ainsi que sur la nécessité de garantir des normes élevées dans la recherche.



Lille
Open
Research
Data

DATAACTIVIST

Plan de gestion de données



Concept

Données

Designed by Freepik -
storyset

Le plan de gestion des données (PGD ou "Data Management Plan" - DMP - en anglais) est un document décrivant la manière dont les données de recherche seront collectées, organisées, stockées, partagées et archivées. Il vise à assurer leur qualité, leur pérennité et leur réutilisation. Le PGD détaille aussi les aspects juridiques, éthiques et techniques, garantissant une gestion responsable et transparente des données. Ce document constitue un livrable pour les porteurs de projets financés par l'ANR ou la Commission européenne.

Plan national pour la science ouverte



Temps forts

Institutions

enseignementsup-
recherche.gouv.fr

Le Plan National pour la Science Ouverte (PNSO) est un document officiel du Ministère français de l'Enseignement Supérieur et de la Recherche, lancé en 2018 et renforcé en 2021. Il définit des orientations stratégiques pour promouvoir la science ouverte en France. Il repose sur trois axes : la généralisation de l'accès ouvert aux publications, la structuration et l'ouverture des données de recherche, ainsi que la promotion de la science ouverte dans les pratiques et l'évaluation des chercheurs.



Lille
Open
Research
Data

DATAACTIVIST

Politique publique de la donnée



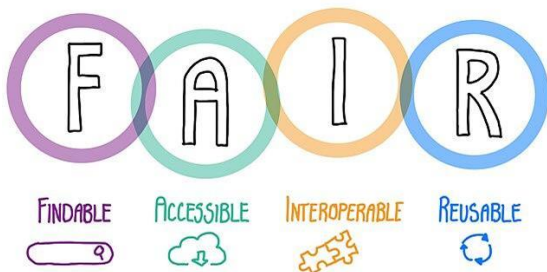
Grands principes

pojok d by Frepik

Le rapport au Premier ministre du député Bothorel sur la politique de la donnée (2020) a montré que les politiques liées aux données dépassent les seuls enjeux techniques :

« la donnée et les codes source ne sont pas juste un enjeu « tech », mais d'abord et surtout une question politique, démocratique, scientifique et économique » .

Principes FAIR



Concept

Données

Bonnes pratiques

CC BY 4.0 -
Heidi Seibold

Les principes FAIR (anglais : Findable, Accessible, Interoperable, Reusable) sont un ensemble de bonnes pratiques qui visent à rendre les données de la recherche ouvertes **F**aciles à trouver, **A**ccessibles via des protocoles ouverts, **I**nteropérables (compatibles avec d'autres systèmes) et **R**éutilisables. Ils impliquent l'utilisation de métadonnées standardisées, de formats ouverts et de licences claires pour favoriser une gestion, un partage et une diffusion efficaces des données de la recherche.

Recherche data gouv



RÉPUBLIQUE
FRANÇAISE

*Liberté
Égalité
Fraternité*

recherche.data.gouv.fr



**Un écosystème au service du partage
et de l'ouverture des données de recherche**

Infrastructures

Accompagnement

Outil

Ouvrir.lascience.fr

Recherche Data Gouv est la plateforme nationale fédérée dédiée au partage et à la valorisation des données de la recherche scientifique. Sa création a été annoncée dans le second Plan National pour la Science Ouverte (2021) et elle fut officiellement lancée en juillet 2022.

Recherche Data Gouv propose des outils pour le dépôt des données (entrepôt, catalogue) et un réseau complet d'accompagnement, en local (ateliers de la donnée) et au niveau national (notamment : centres de référence thématiques, centres de ressources).



Lille
Open
Research
Data

DATAACTIVIST

Règlement Général sur la Protection des Données (RGPD)



Cadre juridique

Temps forts

Dooffy sur Pixabay

Le RGPD est une législation de l'UE entrée en vigueur en mai 2018. Ses principales dispositions incluent le consentement explicite des individus pour le traitement de leurs données, le droit d'accès, de rectification, de suppression, de notification en cas de violation, et la désignation dans chaque entreprise d'un délégué à la protection des données (DPO). Il régit le traitement des données personnelles, qu'il faut anonymiser en cas d'ouverture en s'assurant que les individus ne sont pas réidentifiables par croisement de données.

Risques



Enjeux

Bonnes pratiques

CC BY-SA -
Passport pour la
Science Ouverte

La mauvaise gestion et l'absence de partage des données de la recherche ont des conséquences personnelles et collectives fortes : perte de données, fuite de données, impact sur la qualité de la recherche produite, etc. L'impact financier est également important : une étude de 2018 a estimé que l'impact de la mauvaise gestion des données se chiffrait à 10 milliards € pour l'Europe.

(European Commission, Cost of not having FAIR research data, 2018,
<https://data.europa.eu/doi/10.2777/02999>)

Secret statistique



Concept

Cadre juridique

Designed by Freepik -
storyset

Le secret statistique interdit la communication de données issues de traitements à finalités statistiques dès lors qu'elles sont individuelles ou susceptibles d'identifier des personnes.

Le secret statistique est opposable à toute réquisition judiciaire et toute utilisation des données à des fins de décision à l'égard des personnes concernées, comme le contrôle fiscal.

Secrets protégés



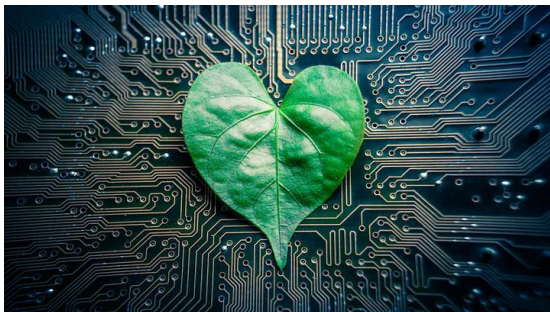
Concept

Cadre juridique

Vectorjuice sur
Freepik

Certains documents administratifs (et donc des données) ne peuvent être communiqués lorsque leur contenu porte atteinte aux secrets suivants définis par la loi : défense nationale, conduite de la politique extérieure, monnaie et le crédit public, secret médical, vie privée, secrets industriels et commerciaux, secret des affaires...

Sobriété numérique



Grands principes

Bonnes pratiques

Adobe Stock –
peach_adobe

Les volumes de données produites par la recherche double tous les 3 ans. Pour limiter l'impact environnemental de la recherche publique, il est nécessaire de mettre en place de bonnes pratiques de gestion des données (réutilisation de données, limitation des versions stockées, low code, etc.)

Software Heritage



Software Heritage

Infrastructures

Bonnes pratiques

Code et logiciels

Domaine public

Software Heritage est une archive ouverte visant à collecter, préserver et partager le code source des logiciels. Fondée en 2016 par Inria et soutenue par l'UNESCO, elle vise à archiver et à préserver le patrimoine des logiciels développés, en les rendant accessibles pour la recherche et la réutilisation. Le projet repose sur un dépôt durable, garantissant que le code source, les historiques de version et la documentation soient sauvegardés. Il est connecté à l'archive ouverte HAL depuis 2018.



Lille
Open
Research
Data

DATAACTIVIST

Transparence et intégrité scientifique



Grands principes

Enjeux

Uniconlabs pour
Flaticon

La démarche de Science ouverte s'inscrit dans un mouvement allant vers une plus grande transparence du processus scientifique.

Dans ce contexte, la diffusion de données FAIR bien documentées, mais également des méthodes, protocoles, codes sources, est une des garantie pour une recherche plus intègre.

Visibilité du chercheur



Enjeux

Bonnes pratiques

Busracavus

Le partage des données – comme toute pratique de Science ouverte – permet d'augmenter la visibilité des travaux scientifiques.

Ainsi, une étude de 2025 a montré qu'un article partageant des données est associé à **14,3%** de citations supplémentaires

(Giovanni Colavizza, Lauren Cadwallader, Iain Hrynaskiewicz, « An analysis of the effects of open science indicators on citations in the French Open Science Monitor », 2025 -

<https://arxiv.org/abs/2508.20747>)

Titre

Mot clé 1

Mot clé 2

Mot clé 3

Source et crédit

Texte