

Science ouverte et humanités numériques

Enjeux, stratégies et perspectives pour la valorisation du patrimoine

Nicolas Fressengeas

23 octobre 2025

Université de Lorraine



Science ouverte et patrimoine numérique

Patrimoine numérique : ouvrir, c'est préserver

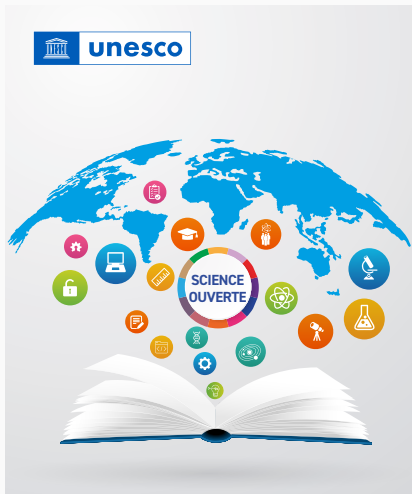
Un patrimoine visible, interopérable & réutilisable

Science ouverte et patrimoine

Science ouverte et patrimoine

Science ouverte

Ouvrir la science : plusieurs visions



Recommandation de l'UNESCO sur une science ouverte

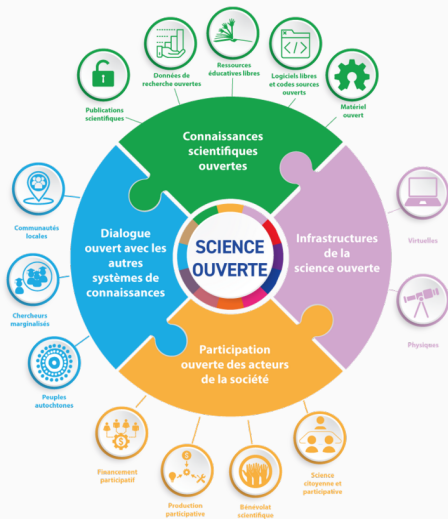


Deuxième Plan national pour la science ouverte



**GÉNÉRALISER
LA SCIENCE OUVERTE
EN FRANCE 2021-2024**

Ouvrir la science. Toute la science?



Reprenons en main la diffusion de nos écrits

- Une question d'infrastructure
- Une question de modèle économique
- Une question de temps de travail
- Une question de changement culturel



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire
- Transformation collective — globale — en cours

Reprenons en main la diffusion de nos écrits

- Une question d'infrastructure
- Une question de modèle économique
- Une question de temps de travail
- Une question de changement culturel



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire
- Transformation collective — globale — en cours

Les données sont un trésor souvent inexploité

- Une question d'intégrité
- Une question d'efficacité
- Une question de responsabilité
- Une question de souveraineté



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre partage, ouverture et protection du patrimoine scientifique et technique

Les données sont un trésor souvent inexploité

- Une question d'intégrité
- Une question d'efficacité
- Une question de responsabilité
- Une question de souveraineté



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre partage, ouverture et protection du patrimoine scientifique et technique

Les données sont un trésor souvent inexploité

- Une question d'intégrité
- Une question d'efficacité
- Une question de responsabilité
- Une question de souveraineté



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre partage, ouverture et protection du patrimoine scientifique et technique

Les logiciels sont dans toutes les disciplines

- Une question de reproductibilité
- Une question de rayonnement
- le troisième côté du triangle de la confiance



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre licences libres et valorisation économique

Les logiciels sont dans toutes les disciplines

- Une question de reproductibilité
- Une question de rayonnement
- le troisième côté du triangle de la confiance



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre licences libres et valorisation économique

Les logiciels sont dans toutes les disciplines

- Une question de reproductibilité
- Une question de rayonnement
- le troisième côté du triangle de la confiance



Une question d'évaluation des activités de recherche

- Valorisation du travail nécessaire

Une question d'injonctions qui semblent contradictoires

- Entre licences libres et valorisation économique

Science ouverte et patrimoine

Patrimoine et numérique

Un patrimoine qui peut être numérisé

- 3D — Architecture
- 2D — Peinture
- 1D — Son
- 4D — Vidéos
- ...



Quid du patrimoine immatériel ?

Selon le Ministère de la Culture

Le *patrimoine culturel immatériel* recouvre les **expressions** et traditions orales, les **pratiques sociales**, les **rituels**, les événements festifs, les savoirs et pratiques relevant des **arts du spectacle**, les **savoir-faire** artisanaux ou encore les connaissances en lien avec la nature et l'univers.



L'équitation de tradition française est une pratique fondée sur l'harmonie des relations homme-cheval, la légèreté et l'absence de contraintes. Elle est le seul PCI concernant l'ensemble de la France, avec le Repas gastronomique des Français.

Le patrimoine nativement numérique

L'Internet **libre** est (*une partie de*) notre patrimoine de données numériques

- Enjeu de la préservation
- Enjeu de la diversité des données
- Enjeu de la fermeture de certains jardins numériques



Le patrimoine logiciel s'organise

- Préservation
- Référencement
- Organisation et architecture



Science ouverte et patrimoine

Le patrimoine scientifique

Le patrimoine scientifique *invisible* de nos établissements

Écrits scientifiques : publication et ouvrages

- Un patrimoine **accessible** ?
 - Une accessibilité limitée aux deux tiers en France
- Un patrimoine **découvrable** ?
 - L'enjeu de l'éparpillement et du référencement

Données produites par la recherche

- Des données FAIR ?
- Des données ouvertes ?
- Un patrimoine découvrable ?

Logiciels et codes sources

- Concerne toutes les disciplines
- Moins d'un article sur cinq fait référence aux logiciels utilisés



Le patrimoine scientifique *invisible* de nos établissements

Écrits scientifiques : publication et ouvrages

- Un patrimoine accessible ?
- Un patrimoine découvrable ?

Données produites par la recherche

- Des données FAIR ?
 - Accessibles, documentées, intre-opérables, réutilisables... ?
 - Un article sur cinq fait référence aux données qu'il mentionne
- Des données ouvertes ?
- Un patrimoine découvrable ?

Logiciels et codes sources

- Concerne toutes les disciplines
- Moins d'un article sur cinq fait référence aux logiciels



Écrits scientifiques : publication et ouvrages

- Un patrimoine accessible ?
- Un patrimoine découvrable ?

Données produites par la recherche

- Des données FAIR ?
- Des données ouvertes ?
- Un patrimoine découvrable ?

Logiciels et codes sources

- Concerne toutes les disciplines
- Moins d'un article sur cinq fait référence aux logiciels utilisés
- Des logiciels ouverts ?

Écrits scientifiques : publication et ouvrages

Données produites par la recherche

Logiciels et codes sources

Mettons notre patrimoine scientifique en valeur

- Ouverture
- Accessibilité
- Interopérabilité
- Réutilisabilité
- Référencement

L'enjeu patrimonial national

Le patrimoine des publications françaises

- Un enjeu qui n'est plus à démontrer
- Un patrimoine dont la visibilité n'est pas si évidente

Les données de la recherche française

- Le catalogue de Recherche Data Gouv
Ouverture prochaine

Les logiciels de la recherche française

- Le catalogue des logiciels libres de la recherche

<https://logiciels.catalogue-esr.fr/>



VEILLE n° 16/09/2025

Pour la première fois à l'échelle d'un pays entier, la France, une étude montre que les pratiques de science ouverte sont associées à une augmentation des citations

Les auteurs de l'étude ont choisi d'étudier la France, un des pays leader dans le suivi des indicateurs de la science ouverte, et d'utiliser les données du Baromètre français de la science ouverte qui couvre le territoire national depuis 2018...



recherche.data.gouv.fr



SCIENCE OUVERTE 12/26
à l'Université de Lorraine

La bibliothèque d'Alexandrie

<https://openalex.org>



Les données de la recherche mondiale?

OpenAlex aborde le sujet

Les logiciels de la recherche?

Inclus dans Software Heritage



Ouvrir, c'est préserver

Ouvrir, c'est préserver

Préserver, c'est choisir

Un patrimoine écrit en danger

Order #813

Manuscript Publication time: **February 2026**

Proposal available until **02-06-2024**

There are 4 coauthors per this manuscript.

Price from 500 USD.

♥ to Favorites

Close this page

Topic of the article:

Drug addiction in children

To buy 1 place 760 USD. **free**

Cntr.

✓ Add to cart

To buy 2 place 670 USD. **free**

Cntr.

✓ Add to cart

To buy 3 place 590 USD. **free**

Cntr.

✓ Add to cart

To buy 4 place 500 USD. **free**

Cntr.

✓ Add to cart

Key words:

Scopus

Q4

Percentile 1..20

Medicine

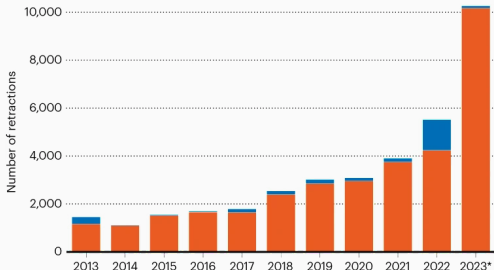
Pharmaceuticals

Drug Discovery

A BUMPER YEAR FOR RETRACTIONS

Retraction notices in 2023 have passed 10,000, largely because of more than 8,000 retractions by Hindawi.

Journal articles Conference papers



onature

*As of 8 December 2023

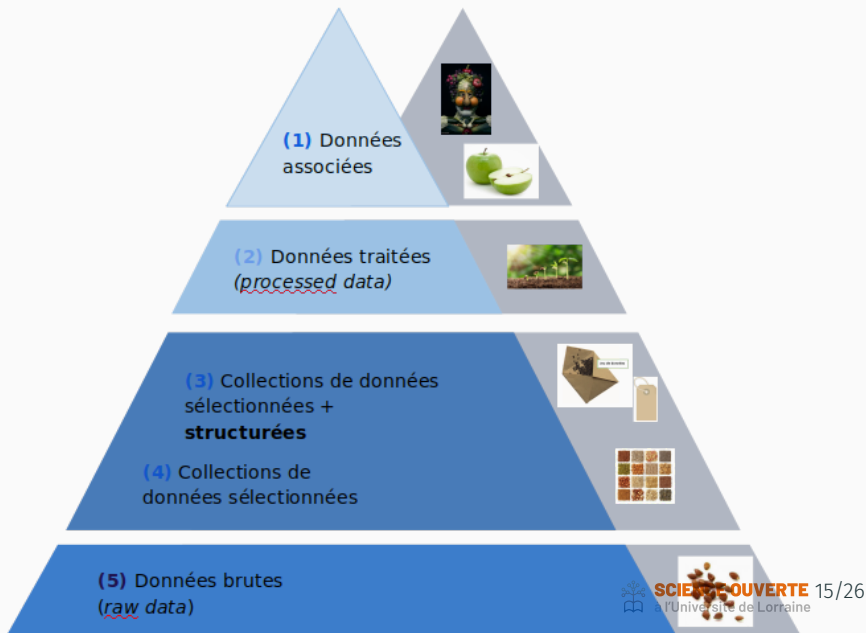
Nature, 12 décembre 2023

The Conversation 2024, Rizqy Amelia Zein,
doi : 10.64628/AAN.5jhd6cscg



SCIENCE OUVERTE 14/26
à l'Université de Lorraine

La pyramide de la donnée





Une archive mondiale et de tous les temps

- Construire la bête : le graphe cryptographique unifié
- Au-delà des bases de données traditionnelles : la puissance des graphes
- Applications et perspectives issues des graphes
- Travaux futurs avec l'IA : instaurer la confiance grâce à la transparence
- Préserver nos logiciels communs : une responsabilité partagée

Ouvrir, c'est préserver

Préserver, c'est documenter

Pas de patrimoine sans métadonnées

Les données sur les données...

- Pour comprendre
- Pour réutiliser
- Pour classer
- Pour archiver
- Pour retrouver
- Pour indexer



Ouvrir, c'est préserver

Préserver, c'est copier

Corée du Sud, 26 septembre 2025, presque 1 Po perdu



THE SOUTH KOREAN DATA CENTER FIRE

how a single fire crashed a nation's digital services

EAG

<https://eticaag.com/south-korean-data-center-fire-crashed-digital-services/>

Des risques systémiques aux USA

- **Fermeture** de l'accès aux ressources
- **Arrêt** ou suspension de la collecte
- **Dégradation** des infrastructures et perte de compétences
- **Privatisation** et tarification des services

Une dépendance scientifique et technique

- 66% des logiciels issus de la recherche française sont sur GitHub ou GitLab
- 70 bases de données suspendues ou en situation critique parmi les 450 bases US identifiées comme indispensables aux recherches françaises

Des risques systémiques aux USA

- Fermeture de l'accès aux ressources
- Arrêt ou suspension de la collecte
- Détérioration des infrastructures et perte de compétences
- Privatisation et tarification des services

Une **dépendance** scientifique et technique

- **66%** des logiciels issus de la recherche française sont sur GitHub ou GitLab
- **70** bases de données suspendues ou en situation critique
parmi les 450 bases US identifiées comme indispensables aux recherches françaises

La **redondance**, pilier de la sécurité des données numériques

- En local, l'architecture RAID
- Dans les Data Centre labellisés de l'ESR : sur plusieurs sites

La redondance, c'est politique

- À patrimoine international, redondance internationale
- Software Heritage développe ses miroirs

La **préservation**, c'est l'ouverture

- Le patrimoine numérique doit pouvoir être copié
- L'ouverture est la clef de la légalité
 - Licences Creative Commons
 - Licences logicielles libres

La redondance, pilier de la sécurité des données numériques

- En local, l'architecture RAID
- Dans les Data Centre labellisés de l'ESR : sur plusieurs sites

La redondance, c'est politique

- À patrimoine international, redondance internationale
- Software Heritage développe ses miroirs

La préservation, c'est l'ouverture

- Le patrimoine numérique doit pouvoir être copié
- L'ouverture est la clef de la légalité
 - Licences Creative Commons
 - Licences logicielles libres

La redondance, pilier de la sécurité des données numériques

- En local, l'architecture RAID
- Dans les Data Centre labellisés de l'ESR : sur plusieurs sites

La redondance, c'est politique

- À patrimoine international, redondance internationale
- Software Heritage développe ses miroirs

La préservation, c'est l'ouverture

- Le patrimoine numérique doit pouvoir être copié
- L'ouverture est la clef de la légalité
 - Licences Creative Commons
 - Licences logicielles libres

Ouvrir, c'est préserver

Préserver, c'est interpréter

Une donnée numérique n'est rien sans son format

Seule une donnée numérique dans un format ouvert est un patrimoine

- Il ne sert à rien de conserver une donnée sans son format
- Conserver des données dans un format propriétaire est illusoire

C'est conserver un patrimoine dans une cave dont on ne possède pas la clef

Patrimoine numérique : l'enjeu de l'ouverture des formats

La patrimoine numérique inclut les formats ouverts



Une donnée numérique n'est rien sans son format

Seule une donnée numérique dans un format ouvert est un patrimoine

- Il ne sert à rien de conserver une donnée sans son format
- Conserver des données dans un format propriétaire est illusoire

C'est conserver un patrimoine dans une cave dont on ne possède pas la clef

Patrimoine numérique : l'enjeu de l'ouverture des formats

La patrimoine numérique inclut les formats ouverts



Un patrimoine visible & utilisable

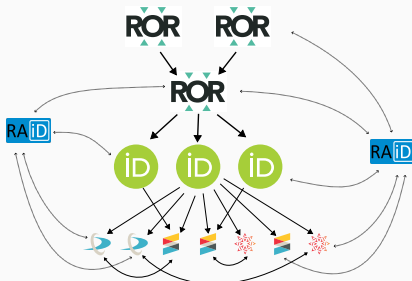
Un patrimoine visible & utilisable

Identifier

L'enjeu des identifiants persistants internationaux

Le graphe du patrimoine scientifique...

- Identification **p**érenne
- Identification **u**nivoque
- **L**iens entre producteurs et productions
- **F**ormats ouverts et normalisés
- Circulation des données



Sans oublier la patrimoine culturel

ISBN,ISSN, ARK...

L'enjeu des identifiants persistants internationaux

Le graphe du patrimoine scientifique...

- Identification pérenne
- Identification univoque
- Liens entre producteurs et productions
- Formats ouverts et normalisés
- Circulation des données



Sans oublier la patrimoine culturel

ISBN, ISSN, ARK...



L'enjeu des identifiants persistants internationaux

Le graphe du patrimoine scientifique...

- Identification pérenne
- Identification univoque
- Liens entre producteurs et productions
- Formats ouverts et normalisés
- Circulation des données



Sans oublier la patrimoine culturel

ISBN, ISSN, ARK...

Un patrimoine visible & utilisable

Référencer

Pour assurer une visibilité internationale

- avec les **identifiants appropriés**
- avec les **protocoles appropriés**
- *sans payer Google*

Pour assurer une découvrabilité optimale

Avec les **métadonnées appropriées**

Les clefs de l'écoystème numérique de la science ouverte

Pour les écrits

- Leur associer un DOI (de préférence de Crossref)
- Les ouvrir, en archive ouverte et/ou chez l'éditeur

Pour les données

- Utiliser un entrepôt de confiance de Recherche Data Gouv

Pour les logiciels et codes sources

- Utiliser une forge pour le développement
- S'assurer qu'elle soit moissonnée par Software Heritage
- Déposer sur HAL les versions abouties
- Les déclarer dans le catalogue national des logiciels de la recherche académique



Pour les écrits

- Leur associer un DOI (de préférence de Crossref)
- Les ouvrir, en archive ouverte et/ou chez l'éditeur

Pour les données

- Utiliser un entrepôt de confiance de Recherche Data Gouv

Pour les logiciels et codes sources

- Utiliser une forge pour le développement
- S'assurer qu'elle soit moissonnée par Software Heritage
- Déposer sur HAL les versions abouties
- Les déclarer dans le catalogue national des logiciels de la recherche académique

Les clefs de l'écoystème numérique de la science ouverte

Pour les écrits

- Leur associer un DOI (de préférence de Crossref)
- Les ouvrir, en archive ouverte et/ou chez l'éditeur

Pour les données

- Utiliser un [entrepôt de confiance](#) de Recherche Data Gouv

Pour les **logiciels** et codes sources

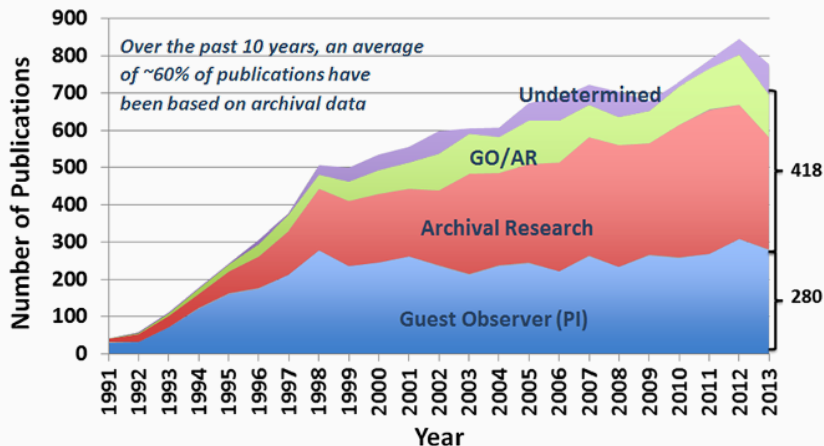
- Utiliser une **forge** pour le développement
- S'assurer qu'elle soit moissonnée par **Software Heritage**
- Déposer sur **HAL** les versions abouties
- Les déclarer dans le [catalogue national des logiciels de la recherche académique](#)



Un patrimoine visible & utilisable

Réutiliser

Quelques exemples de réutilisation




Quelques exemples de réutilisation



HTR-United est un catalogue qui répertorie des jeux de données d'entraînement hautement documentés utilisés pour les modèles de transcription ou de segmentation automatiques.

HTR-United normalise les descriptions des jeux de données à l'aide d'un schéma, propose des recommandations pour l'organisation des dépôts de données et fournit des outils pour le contrôle de la qualité et la documentation continue. C'est un écosystème ouvert et transparent hébergé sur GitHub, conçu pour faciliter sa maintenance. HTR-United a été créé pour aider les projets à accéder rapidement à diverses données de vérité terrain pour entraîner des modèles sur des petits corpus.

 Explorer le catalogue

 Enregistrer un dataset

Quelques exemples de réutilisation

[Données](#)[API](#)[Réutilisations](#)[Organisations](#)[Démarrer sur data.gouv.fr](#) ▾[Nouveautés](#)[Nous écrire](#)

[Accueil](#) > [Jeux de données](#) > [Système d'Information Géographique de la France de la Troisième République \(1870–1940\)](#)

Système d'Information Géographique de la France de la Troisième République (1870–1940) TRF-GIS

La base de données TRF-GIS correspond à un système d'information géographique de la France de la Troisième République. Elle met à disposition nomenclatures et shapefiles annuels correspondant aux cir...

Description

La base de données TRF-GIS correspond à un système d'information géographique de la France de la Troisième République. Elle met à disposition nomenclatures et shapefiles annuels correspondant aux circonscriptions administratives de France métropolitaine de 1870 à 1940. Elle décrit les circonscriptions administratives générales (départements, arrondissements, cantons) ainsi que les circonscriptions militaires, judiciaires, pénitentiaires, électorales, académiques, ecclésiastiques, et les inspections du travail. Elle met aussi à disposition des nomenclatures annuelles établissant une correspondance entre chaque commune contemporaine et les circonscriptions auxquelles elle appartenait.

Pour plus de détails sur le contenu et la construction de la base de données TRF-GIS, se référer au data paper suivant : [Gay, Victor. 2021. "Mapping the Third Republic. A Geographic Information System of France \(1870–1940\)." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 54\(4\), 189–207 !\[\]\(de5c2ae460dd39f4872d99dd907de767_img.jpg\)](#). Celui-ci est aussi disponible en Open Access sur HAL à l'adresse suivante : <https://hal.archives-ouvertes.fr/hal-02951461>.

Quartier Richelieu

Plongez dans l'histoire du Paris du XIX^e siècle



4217 documents

Issus de 26 institutions

Représentant 1118 points d'intérêt

Dans 331 lieux

[Entrer dans le site](#)

Diaporama en accès ouvert



<https://doi.org/10.5281/zenodo.17390489>

Gardons le contact



<https://fressengeas.net>