

Huma-Num IR*

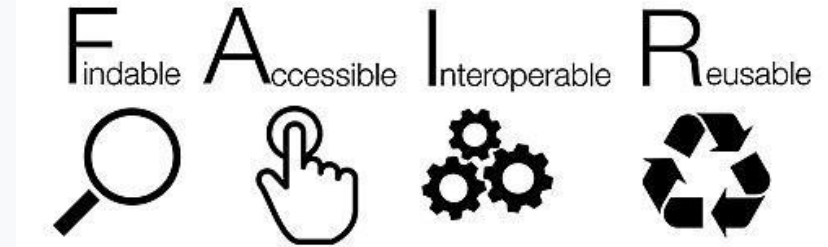
Choisir un entrepôt de confiance : les principes FAIR en pratique avec Nakala

Xiaou Wang - Référent Humanités Numériques

Qu'est-ce qu'un entrepôt de confiance

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
- Non-respect de la liberté académique



Références

Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Lamotte, F., et al. (2024). "Sélectionner un entrepôt thématique de confiance pour le dépôt de données". Report, Comité pour la Science Ouverte. <https://doi.org/10.52949/52>

Où trouver des entrepôts de confiance

Accès au portail Recherche Data Gouv

Sur cette page, Recherche Data Gouv se fait le relais des entrepôts thématiques de confiance proposés par le Collège des Données de la Recherche du Comité pour la science ouverte.

Le partage des données issues des activités de recherche implique de les mettre à disposition dans des entrepôts reconnus par les communautés disciplinaires, thématiques, institutionnels ou généralistes.

- Utilisez la **méthodologie d'identification** du Collège des Données.
- Consultez la **grille de critères d'analyse** retenus.
- Liste évolutive et non exhaustive.

Entrepôts de confiance

Sur cette page, Recherche Data Gouv se fait le relais des entrepôts thématiques de confiance proposés par le Collège des Données de la recherche du Comité pour la science ouverte.

Le partage des données issues des activités de recherche implique de les mettre à disposition dans des entrepôts reconnus par les communautés scientifiques, thématiques, institutionnels ou généralistes.

Pour faciliter le travail des équipes de recherche, le collège des Données de la recherche propose une méthode d'identification des entrepôts thématiques de confiance, ainsi qu'une première liste d'entrepôts découlant de la grille de critères d'analyse retenus à découvrir ci-dessous.

Cette liste est évolutive et a vocation à être progressivement complétée par le collège des Données de la Recherche. **Elle n'est par ailleurs pas exhaustive** : ainsi, ne pas figurer dans la liste n'implique pas que l'entrepôt concerné n'est pas de confiance.

⚠ N'oubliez pas, la dernière version à jour de la liste de référence est celle publiée sur le [site Ouvrir la Science](#).

■ Pour toute question relative aux entrepôts de confiance, contactez le [Collège des Données de la recherche du Comité pour la science ouverte](#).

(dernière mise à jour : avril 2025)

Rechercher sur la page Recherche

Disciplines

Recherche

Sciences humaines et sociales (11)

ArkeOpen Archéologie +

Array Express Biologie +

recherche.data.gouv.fr/fr/entrepots



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

Entrepôt national généraliste

Description détaillée



https://www.ouvrirlascience.fr/wp-content/uploads/2025/04/ListesdesEntrepotsdeConfiance_v2_2025-04-20.xlsx

Entrepôt de confiance et valorisation des données

Le 3 mars 2023, le Ministère de l'Enseignement supérieur et de la Recherche a confié au Collège des données de la recherche la double mission de :

Critères de confiance

Proposer une définition de critères de confiance permettant d'**apprécier la qualité d'un entrepôt** à l'aune des objectifs de « **FAIRisation** » des données.

Moissonnage & Visibilité

Œuvrer à l'identification d'entrepôts thématiques de confiance propres à **être moissonnés** par [Recherche Data Gouv](#).

L'écosystème **Recherche Data Gouv** propose aux équipes de recherche une offre pluridisciplinaire de dépôt, publication et **signalement** de leurs données. Il est indispensable d'identifier les entrepôts thématiques et institutionnels en complémentarité avec ces services.

Principes FAIR en pratique avec Nakala - Entrepôt de confiance en SHS

Rechercher sur la page Recherche

Reinitialiser les filtres

Disciplines

Recherche

Sciences humaines et sociales (11)

Champs disciplinaires

Recherche

Anthropologie (1)

Archéologie (1)

Conditions de vie et société (1)

Institutions porteuses

11 résultats

Centre de Données Socio-Politiques (CDSP)	Sciences humaines et sociales +
COLlections de CORpus Oraux Numériques (CoCoON)	Sciences humaines et sociales +
Didómena	Sciences humaines et sociales +
Endangered LAnguages Repository (ELAR)	Sciences humaines et sociales +
Geovistory	Sciences humaines et sociales +
Long-term Access and Usage of Deeply Annotated Information (Laudatio)	Sciences humaines et sociales +
Nakala	Sciences humaines et sociales +

Principes FAIR en pratique avec Nakala - Accès

Authentification HumanID

<https://humanid.huma-num.fr/>



[Reset my password](#)

[Login with HumanID ?](#)

Your first visit on Huma-Num ?

[Create a Huma-Num account](#)

Log in with an external account ⓘ

[Log in with eduGAIN](#)

[Login with ORCID](#)

[Log in with HAL account](#)

[Log in with LinkedIn](#)

[Log in with Twitter](#)

[Log in with Google](#)



Gérer mes services Huma-Num

Manage my Huma-Num services



Search assistant in
Humanities and Social
Sciences

[Access](#)



Storage and file sharing
platform (Web and
WebDAV clients)

[Request access](#)



Share, publish and
enhance your scientific
data

[Request access](#)



Principes FAIR en pratique avec Nakala - Findable



Share, publish and enhance your scientific data

Upload my data in NAKALA

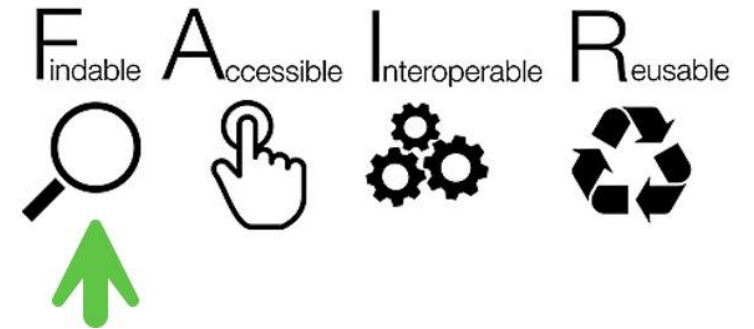
Search, cite and reuse scientific data

search in NAKALA...



Advanced search

<https://nakala.fr/>



F — Findable (Faciles à trouver)

Les données doivent pouvoir être retrouvées facilement :

- avec un titre clair,
- des métadonnées,
- un identifiant pérenne (DOI, URL stable...),
- et un référencement dans des catalogues ou entrepôts.

 **LIARArg: A dataset for fake news detection enhanced by argumentative features**

DOI : 10.34847/nkl.2cec5407

 Public



Principes FAIR en pratique avec Nakala - Findable



Mauvaise pratique : déposer un même jeu de données dans plusieurs entrepôts

LIARArg: A dataset for fake news detection enhanced by argumentative features

DOI : 10.34847/nkl.2cec5407 Public



LIARArg: A dataset for fake news detection enhanced by argumentative features

> <https://doi.org/10.6084/m9.figshare.32252700>



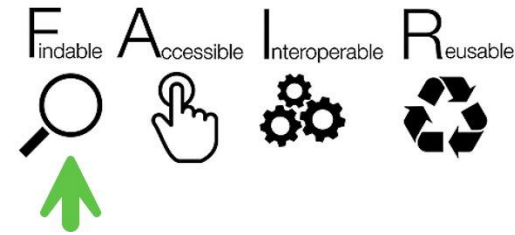
Cite

Download (46.76 MB)

Share

Embed

+ Collect



Le dépôt en double nuit à la découvrabilité et à la citation unique des données.

La version archivée sur **Nakala** doit être la référence canonique.



Canonical archived version on Nakala:
<https://doi.org/10.34847/nkl.2CEC5407>

Principes FAIR en pratique avec Nakala - Accessible



Share, publish and enhance your scientific data

Upload my data in NAKALA

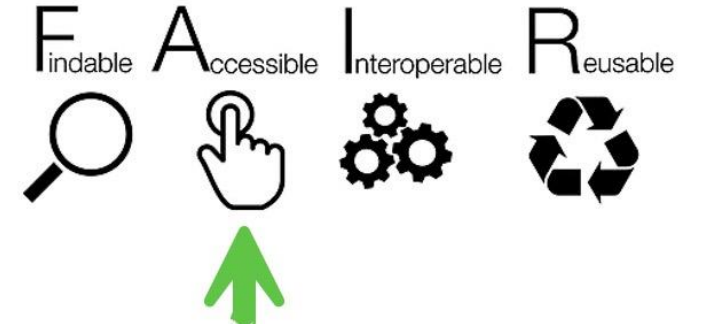
Search, cite and reuse scientific data

search in NAKALA...



Advanced search

<https://nakala.fr/>



A — Accessible (Accessibles)

Les données doivent être accessibles :

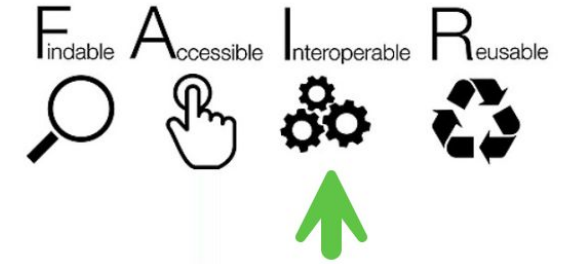
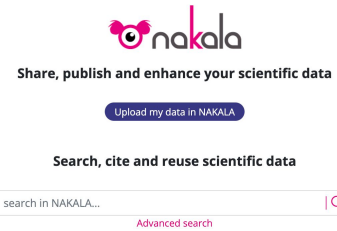
- via Internet,
- avec des protocoles standards,
- même si certaines restrictions existent.

Accessible ne veut pas forcément dire "ouvert à tous" :

- **certaines données peuvent être protégées,**
- mais les conditions d'accès doivent être claires.

<https://nakala.fr/10.34847/nkl.ead9ru8v>

Principes FAIR en pratique avec Nakala - Interopérable



Formats Standards

Utilisation de formats ouverts :

- CSV
- JSON
- XML...

Formats propriétaires ou peu interopérables :

- XLSX (Excel propriétaire, même si très répandu)
- DOCX
- PSD (Photoshop)

Vocabulaires Contrôlés

Utilisation de thésaurus et de terminologies partagées pour assurer une sémantique commune.

<https://fairsharing.org/search?q=thesaurus+economy>

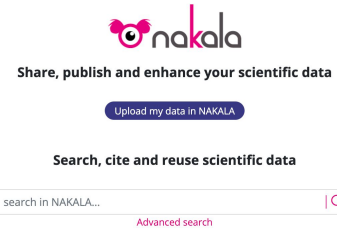
Métadonnées Normalisées

Schémas reconnus :

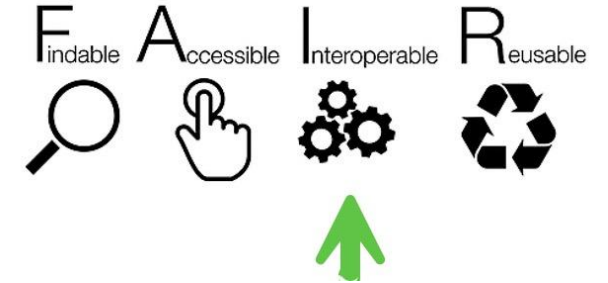
- Dublin Core
- Dublin Core Qualifié
- DataCite, etc.

https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf

Principes FAIR en pratique avec Nakala – Interopérabilité



<https://nakala.fr/>



Métadonnées Normalisées

Schémas reconnus permettant une interopérabilité sémantique :

- Dublin Core (DC)
- Dublin Core Qualifié (DCTERMS)
- DataCite
- Etc...

Trois formats de moissonnage sur Nakala

- Trois formats de métadonnées sont moissonnables :
 - En **Dublin Core** : oai_dc,
 - Exemple d'une collection exposée en Dublin Core : https://api.nakala.fr/oai2?verb=ListRecords&metadataPrefix=oai_dc&set=10.34847/nkl.d972955s
 - En **Dublin Core qualifié** (DCTerms) : oai_qdc,
 - Exemple d'une collection exposée en DCTerms : https://api.nakala.fr/oai2?verb=ListRecords&metadataPrefix=oai_qdc&set=10.34847/nkl.d972955s
 - Au format de **DataCite** : oai_datacite,
 - Exemple d'une collection exposée en DataCite : https://api.nakala.fr/oai2?verb=ListRecords&metadataPrefix=oai_datacite&set=10.34847/nkl.d972955s

Principes FAIR en pratique avec Nakala – Interopérabilité



Share, publish and enhance your scientific data

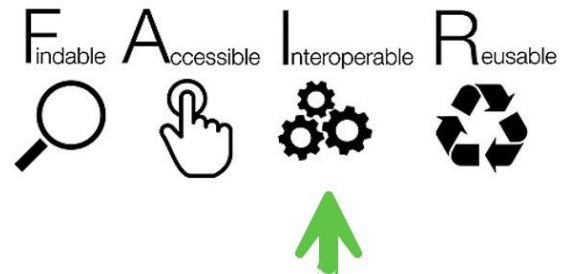
Upload my data in NAKALA

Search, cite and reuse scientific data

search in NAKALA... | Q

Advanced search

<https://nakala.fr/>



Métadonnées Normalisées

Schémas reconnus permettant une interopérabilité sémantique :

- Dublin Core (DC)
- Dublin Core Qualifié (DCTERMS)
- DataCite
- Etc...

Mappage de métadonnées dans Nakala

propriété NAKALA	signification	oai_dc	qdc	datacite
nkl:title	Titre	dc:title	dcterms:title	title
nkl:creator	Auteur	dc:creator "surname, givenname [ORCID]" ou "[s.n.]" si vide ou encore ∅ si vide et qu'il existe un dcterms:creator	dcterms:creator "surname, givenname [ORCID]" ou "[s.n.]" si vide ou encore ∅ si vide et qu'il existe un dcterms:creator	creator

Principes FAIR en pratique avec Nakala – Interopérabilité et valorisation



Share, publish and enhance your scientific data

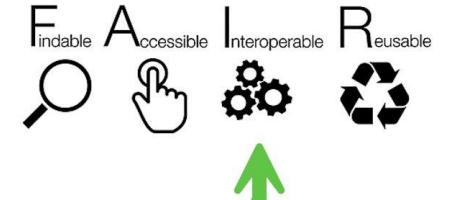
Upload my data in NAKALA

Search, cite and reuse scientific data

search in NAKALA... | Q

Advanced search

<https://nakala.fr/>



Signalable dans (importance des métadonnées riches)

Gallica

Les collections de NAKALA peuvent être "moissonnées" par [GALICA](#).

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la [demande](#).

Les collections de NAKALA peuvent être "moissonnées" par **Gallica**.

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la **demande**.

OpenAire

Les collections de NAKALA peuvent être "moissonnées" par [OpenAire](#).

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la demande.

Les collections de NAKALA peuvent être "moissonnées" par **OpenAire**.

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la **demande**.

Google Data Search


Les collections de NAKALA peuvent être "moissonnées" par [Google Data Search](#).

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la demande.

Les collections de NAKALA peuvent être "moissonnées" par **Google Data Search**.

Le moissonnage n'est pas automatique, il est nécessaire d'en faire la **demande**.

Principes FAIR en pratique avec Nakala – Interopérabilité et valorisation



Share, publish and enhance your scientific data

Upload my data in NAKALA

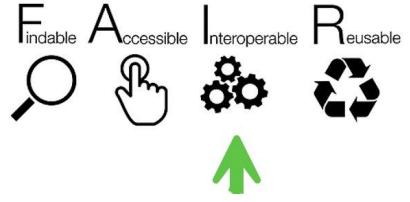
Search, cite and reuse scientific data

search in NAKALA... | Q

Advanced search

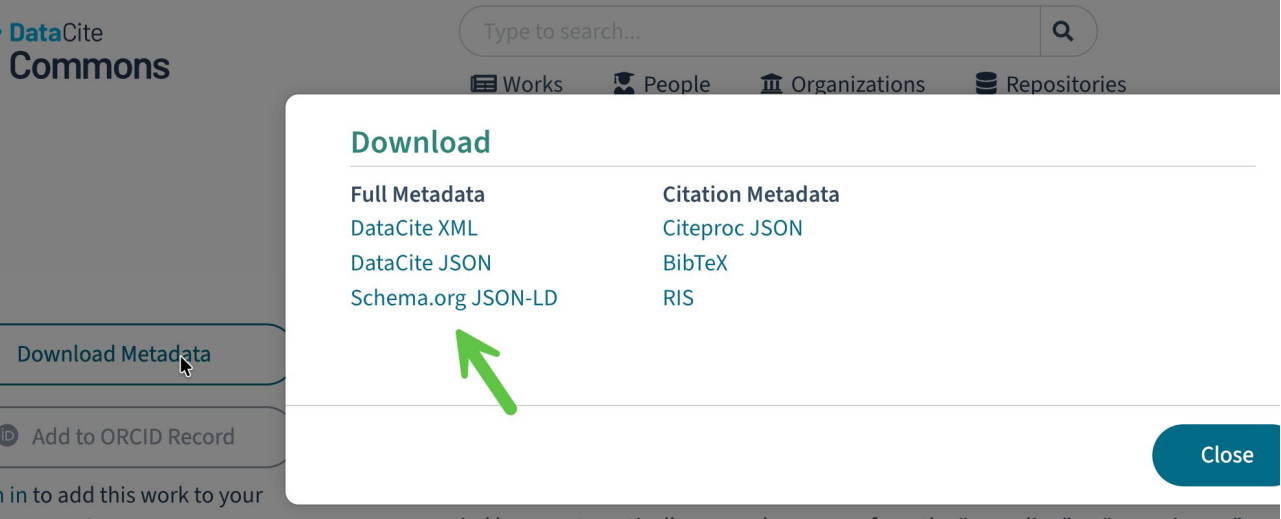
<https://nakala.fr/>

F_{indable} A_{ccessible} I_{nteroperable} R_{eusable}



↑

Exposition automatique des métadonnées dans d'autres formats par des catalogues



The screenshot shows the DataCite Commons interface. A search bar at the top contains the text "Type to search...". Below the search bar are navigation tabs for "Works", "People", "Organizations", and "Repositories". A dropdown menu is open, titled "Download", and lists the following options:

Full Metadata	Citation Metadata
DataCite XML	Citeproc JSON
DataCite JSON	BibTeX
Schema.org JSON-LD	RIS

A green arrow points to the "Schema.org JSON-LD" option. Below the menu, there are buttons for "Download Metadata", "Add to ORCID Record", and "Close".



Schema.org => schéma préféré de Google

D'où le moissonnage automatique par **Datasearch Google**.

Principes FAIR en pratique avec Nakala – Valorisation par site statique

Nakala : Dépôt de données

<https://nakala.fr/>



Share, publish and enhance your scientific data

Upload my data in NAKALA

Search, cite and reuse scientific data

search in NAKALA...



Advanced search

NakalaPress : Valorisation par site statique

<https://liararg.nakala.fr/>



Dataset • Fact-checking • Argumentation

LIARArg

An argumentative dataset for automated fact-checking

The need for automated fact-checking has become urgent with the rise of misleading content on social media. Recently, Fake News Classification (FNC) has evolved to incorporate the justifications provided by fact-checkers in order to explain their decisions.

In this work, we argue that an argumentative representation of fact-checkers' justifications can improve the precision and explainability of fake news classification systems.

To address this challenging task, we present LIARArg, a novel linguistic resource composed of **2,832 news claims** and their corresponding justifications.

Average number of claims and premises across the labels.

Label	Items	Claims	Premises
True	443	1.2	2.3
Mostly-True	529	1.4	2.5
Half-True	522	1.5	2.6
Barely-True	476	1.5	2.7
False	471	1.4	2.5
Pants-on-fire	391	1.4	2.4
Total	2,832	1.4	2.5

LIARArg extends the LIAR-PLUS dataset with six veracity labels by adding argumentative structures, resulting in the first fake news classification dataset annotated with argumentative components (*claim* and *premise*) as well as fine-grained relations (*attack*, *support*, *partial support*, and *partial attack*).

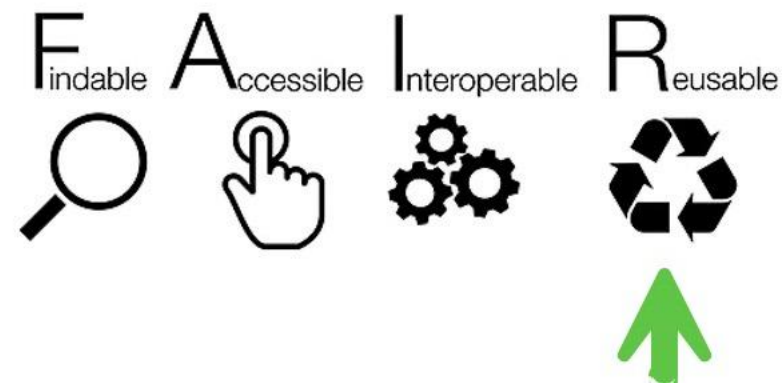
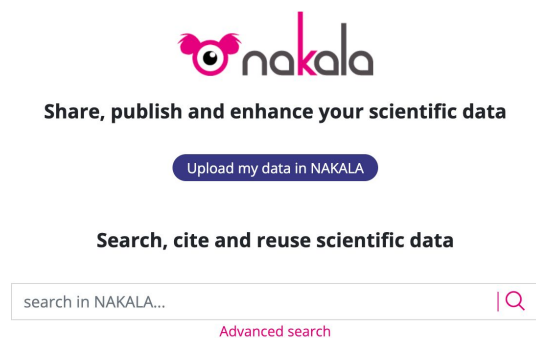
Associated Publication

Wang, X., Cabrio, E., & Villata, S. (2025). *When automated fact-checking meets argumentation: Unveiling fake news through argumentative evidence*. *Argument & Computation*, 16(3), 405-424.

Dataset Citation

Principes FAIR en pratique avec Nakala – Réutilisable

<https://nakala.fr/>



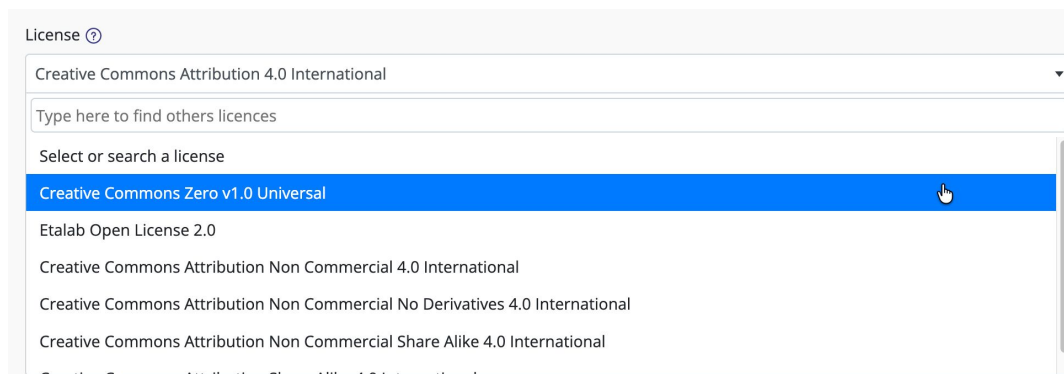
R — Reusable (Réutilisables)

Les données doivent pouvoir être réutilisées dans le futur par d'autres chercheurs.

Cela nécessite :

- une documentation claire,
- une licence explicite,
- une description de la méthodologie,
- des informations sur la qualité et le contexte de production.

Plus de 400 licences disponibles dans
Nakala



Conclusions

Un entrepôt de confiance respecte les principes FAIR et permet de :



Trouvables

rendre les données
identifiables et donc
facilement trouvables



Accessibles

garantir l'accessibilité des
données, sans pour
autant les rendre
nécessairement ouvertes
à tous



Interopérables

assurer l'interopérabilité
des données, essentielle à
leur valorisation et à leur
réutilisation



Réutilisables

préciser clairement les
conditions de réutilisation
des données

Contacts



Guichet unique

Pour toutes les demandes liées au cycle de vie des données (en me mettant en copie)

✉ recherche.datazur@univ-cotedazur.fr



Questions spécifiques aux SHS et à Huma-Num

Xiaoou Wang

☎ 04 89 15 20 39

✉ Xiaoou.WANG@univ-cotedazur.fr

📍 MSHS Sud-Est, bureau 125 (1er étage)

Prochain Webinaire



ATELIER ANONYMISATION DES DONNÉES

Apprenez à protéger les données sensibles tout en préservant leur valeur pour la recherche.



SENSIBILISER

Comprendre les enjeux éthiques, juridiques et techniques liés à la protection des données.



DÉCOUVRIR

Explorer des méthodes et outils d'anonymisation adaptés aux données de recherche.



PRATIQUER

Manipuler un outil libre, gratuit et 100% hors ligne lors d'une démonstration concrète.



DATE
24 juin 2026
de 14 h à 15 h



PUBLIC
Chercheurs, doctorants, ingénieurs, bibliothécaires, archivistes, étudiants... et toute personne manipulant des données sensibles dans un cadre académique.

PROTÉGER, PARTAGER, VALORISER
Des données anonymisées, une recherche responsable



which entity categories should be replaced.

Run spaCy NER Copy anonymized text Show audit report Clear Detected 9 unique entities using fr_core_news_sm.

1. Paste Text

263 characters

Extrait d'entretien :

La docteure Marie Dupont a rencontré Jean Martin à Paris le 12 mars 2022. Marie travaille avec l'Université de Lyon. Son email est marie.dupont@example.com. La participante a aussi mentionné +33 6 12 34 56 78 et https://example.org/project.

2. Replace Categories?

Model: fr_core_news_sm

12-mars-2022

Replace Emails
1 unique value(s)
marie.dupont@example.com

Replace Phone numbers
1 unique value(s)
6-12-34-56-78

Replace URLs
1 unique value(s)
https://example.org/project

3. Highlighted Entities

9 detected, selected categories stay highlighted

Extrait d'entretien :

La docteure Marie Dupont a rencontré Jean Martin à Paris le 12 mars 2022. Marie travaille avec l'Université de Lyon. Son email est marie.dupont@example.com. La participante a aussi mentionné +33 6 12 34 56 78 et https://example.org/project.

4. Anonymized Preview

Selected categories only

Extrait d'entretien :

La docteure [PERSON_1] a rencontré [PERSON_2] à Paris le 12 mars 2022. [PERSON_3] travaille avec l'Université de Lyon. Son email est marie.dupont@example.com. La participante a aussi mentionné +33 6 12 34 56 78 et [URL_1].