

FAIRiser les données de CAHIER

Retour d'expérience, problématiques & enjeux



Fatiha Idmhand (Consortium CAHIER)



RENCONTRES
HUMA-NUM
2021

25-28 mai - 100% digitales



Mardi 25-05-2021

Le consortium CAHIER

(2011-2021)

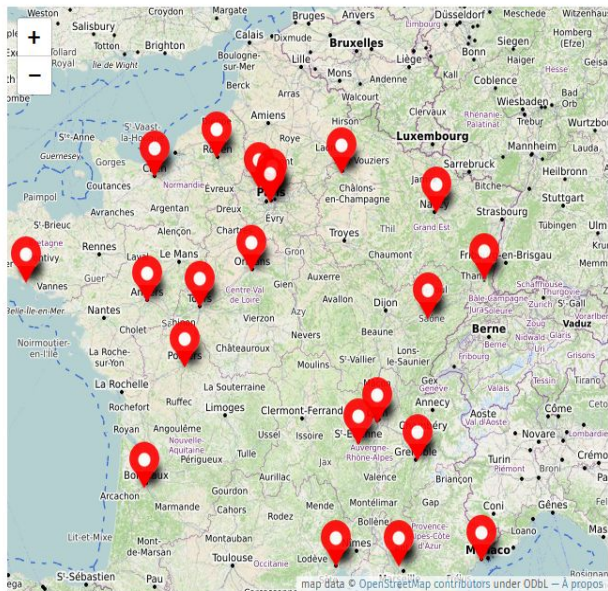


Un consortiumde “projets”



Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

Les projets membres du consortium



Carte des membres du consortium CAHIER par Laurence Rageot

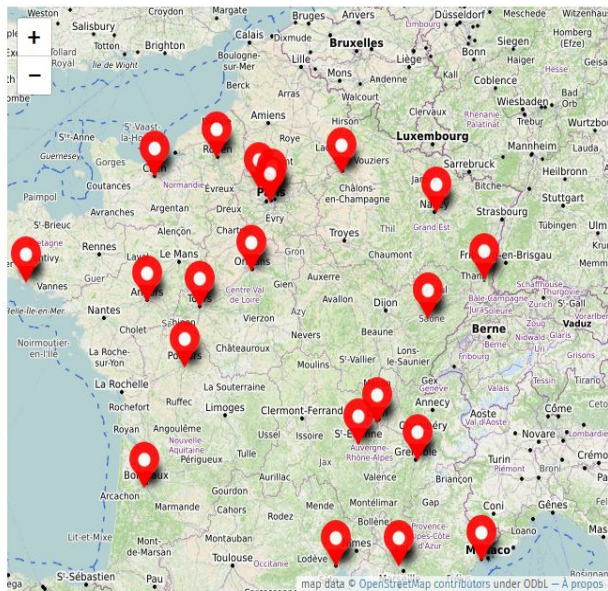
Consortiums de 65 projets de recherches sur des Corpus d'auteurs

1. **Domaines** : Textes et Langues
2. **Expertises scientifiques** : Philologie, Philologie numérique, Génétique des œuvres, Critique textuelle, Traitement automatique des textes, Édition “analogique” et numérique, Numérisation, Transcription, Encodage
3. **Expertises techniques** : Langages de balisage (XML et XML-TEI), Technologies et requêtes XML (XSLT, XQuery etc.), Langages de programmation (Python, Java, etc.)
4. **Outils et applications développés** : WebOai, TXM, mynkl...

Un consortiumde “projets”

Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

Les projets membres du consortium



Carte des membres du consortium CAHIER par Laurence Rageot

Le Consortium CAHIER est constitué de projets de recherche partenaires dont la vocation est de travailler conjointement à l'essor, à la valorisation et à la co-construction des nouvelles pratiques numériques dans le domaine des sciences humaines et sociales. Leurs objets d'études principaux sont les corpus d'auteurs et plus généralement les corpus textuels constitués en référence à l'oeuvre d'un auteur, d'une tradition éditoriale, d'une forme littéraire, d'un genre.

Afin de tenir compte de l'existant et de le faire évoluer tout en favorisant les projets nouveaux et les solutions originales, le consortium se propose de :

- Faire évoluer et consolider les formats de l'édition brute, critique, génétique, encyclopédique
- Favoriser les échanges de données et de méthodes
- Rétroconvertir ou faire migrer dans d'autres formats les corpus anciens, dispersés, déjà édités ou non ; de reprendre des programmes achevés afin de les intégrer dans des ensembles moissonnables et interrogeables
- Pérenniser les données par l'adaptation des formats aux exigences de l'archivage
- Favoriser le libre accès aux corpus
- Encourager la valorisation des productions (démonstrations, transferts à d'autres projets, tests comparatifs, communication des résultats, même provisoires)
- Former de nouvelles générations de chercheurs et d'ingénieurs à compétences mixtes (humanités et informatique) par des stages et des formations diplômantes
- Savoir adapter ou faire adapter les outils existants, provenant d'autres domaines (par exemple les SIG, les outils de visualisation, d'exploitation des réseaux sociaux appliqués aux documents relatifs aux humanités).

Adhérer

Après 10 ans de travail...

Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

10 ans de corpus d'auteurs

IUT, 1 Rue Jacques Ellul
33800 Bordeaux

Du 07 au 10 juin 2021

Programme: cahier10.sciencesconf.org
Contacts: cahier10@sciencesconf.org



Résultats

1. **Des corpus d'auteurs** sous forme
 - d'éditions critiques numériques de textes (Montaigne, Flaubert, Louis de Boissy, etc.)
 - de "bibliothèques virtuelles" numériques
2. **Des guides méthodologiques et des formations**
 - Guide juridique, Guide pour la FAIRisation, Guide pour l'édition numérique, etc.
3. **De la littérature grise** (articles scientifiques, conférences, etc.)
4. **Des outils** : TXM, WebOai, Thésaurus "Typologie textuelle", MynKl, et quelques plateformes éditoriales

10 ans de données... ... normalement “FAIR”



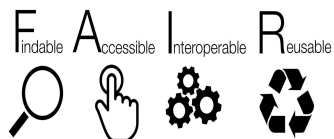
Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

10 ans
de corpus d'auteurs

Programme: cahier10@scienceconf.org
Contacts: cahier10@scienceconf.org

IUT, 1 Rue Jacques Ellul
33800 Bordeaux

Du 07 au 10 juin 2021

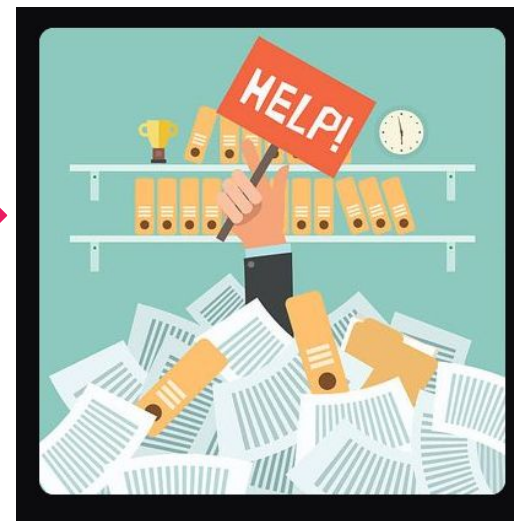


Données estimées

~500000 images

~327450 fichiers

~64 sites webs



Rendre les données de CAHIER “FAIR”

(2011-2021)



Quel outil de stockage?

Etat des lieux des outils et plateformes de dépôt de données

Déposer des données de recherche dans un entrepôt

- 1 - Qu'est-ce qu'un entrepôt de données de recherche
- 2 - Différentes catégories d'entrepôts de données
- 3 - Pourquoi déposer des données dans un entrepôt
- 4 - Les questions à se poser avant de déposer des données dans un entrepôt
- 5 - Comment choisir un entrepôt de données
- 6 - Déposer ses données en pratique : exemple du Dataverse

2 - Différentes catégories d'entrepôts de données

Il existe plusieurs catégories d'entrepôts de données : institutionnels, pluridisciplinaires, disciplinaires ou thématiques.

Un entrepôt de données peut, en effet, être mis en place et géré au niveau d'un établissement de recherche, on parle alors d'entrepôt institutionnel. Il peut également être créé et géré par une entité publique, associative, ou privée, cas d'entrepôt d'éditeurs destiné à accueillir les jeux de données sous-tendant les résultats présentés dans les articles scientifiques que l'éditeur publie.

Exemples d'entrepôts de données :

- Institutionnels (en France) : [Dataverse Cirad](#), [Datapartage](#) (INRAe), [DataSudis](#) (IRD)
- Pluridisciplinaires et internationaux : [Zenodo](#), [Dryad](#), [Figshare](#)
- Thématiques ou disciplinaires : [GenBank](#) (séquences génétiques), [TRY](#) (caractères botaniques), [GBIF](#) (biodiversité), [Pangaea](#) (sciences de la terre et de l'environnement), [WormBase](#) (nématologie), [Movebank](#) (mobilité animale), [West African Vegetation](#), [DataFirst](#) (enquêtes socio-économiques en Afrique), [Protocols.io](#) (protocoles), etc.
- Liés à un éditeur : [GigaDB](#) (Oxford Univ. Press), [Dataverse Ubiquity Press](#), [Dataverse Economics](#)

“Pour sélectionner votre entrepôt, vérifiez qu'il répond aux critères suivants :

- adapté au type de données que vous allez déposer
- répondant aux recommandations du bailleur de votre projet, de votre institution, ou de la revue dans laquelle vous publiez
- **reconnu dans votre discipline et par la communauté scientifique**
- **attribuant un identifiant numérique pérenne, univoque à chaque jeu de données**
- **assurant la conservation des données, c'est-à-dire leur pérennité**
- **gratuit** (la plupart des entrepôts) ou pratiquant des coûts de dépôt de données acceptables
- **proposant les modalités d'accès aux données**, adaptés à vos besoins : accès libre, après enregistrement, **accès restreint, sur demande, différé par un embargo**
- attribuant la licence de diffusion
- (...)”

Dedieu, L., Barale, M. 2020. **Déposer des données dans un entrepôt, en 6 points.** Montpellier (FRA) : CIRAD, 4.
<https://doi.org/10.18167/coopist/0070>

FAIRsharing.org
standards, databases, policies

The
Dataverse®
Project

Cat
OPIDoR

zenodo

 **nakala**

Etat des lieux des travaux de la Research Data Alliance



WG

23 Sep 2018

FAIR Data Maturity Model WG

Status: Recognised & Endorsed

Chair (s): Edit Herczog, Keith Russell, Shelley Stall

Secretariat Liaison: Stefanie Kethers

TAB Liaison: Karin Breitman

FAIR Data Maturity Model: core criteria to assess the implementation level of the FAIR data principles

CONTINUE READING

“Guidelines” publiées en avril 2020

Etat des lieux des projets de CAHIER



Guide pour la
FAIRisation de
vos données

trazado

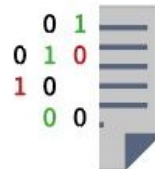
Le consortium CAHIER utilise Nakala pour stocker ses données afin qu'elles soient trouvables, accessibles, interopérables et réutilisables, cette procédure demande 4 étapes.



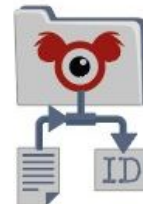
1° Évaluer le degré d'ouverture de ses projets



2° Confronter ses métadonnées aux attentes du consortium



3° Compléter et corriger les métadonnées si nécessaire



4° Déposer sur Nakala, obtenir un identifiant pérenne et l'associer aux documents publiés sur son propre site web (ou sur un site web institutionnel)

Lancement des "FAIR lines" en mars 2020

Nakala

Nakala



Partager, publier et valoriser vos données scientifiques

Je dépose mes données dans nakala

Rechercher, citer et réutiliser des données scientifiques

Rechercher dans NAKALA...



1. Toutes les données peuvent être déposées : adapté à l'hétérogénéité des données de CAHIER: textes (transcrits ou océrisés), vidéos, audios, images (photos et textes numérisés), sources nativement numériques
2. Les données reçoivent un identifiant pérenne : citabilité
3. Les données sont stockées dans un endroit sûr et fiable : l'infrastructure Huma-Num (In2p3)
4. Elles sont formatées en RDF ce qui est conforme aux standards du web sémantique
5. Nakala propose un point d'accès sparql facilite l'accès, les requêtes interactives et l'interopérabilité
6. Nakala propose un point d'accès via API qui facilite(ra) la construction de services ou d'outils

Nakala



Partager, publier et valoriser vos données scientifiques

Je dépose mes données dans nakala

Rechercher, citer et réutiliser des données scientifiques

Rechercher dans NAKALA...



FAIR Data Maturity Model: specification and guidelines.

RDFAIR Data Maturity Model Working Group (2020). Research Data Alliance. DOI:

<https://doi.org/10.15497/rda00041>

FAIR data maturity model In FAIR data maturity model Indicators

| 2020-04-08/version 0.05 | | 2020-04-08/version 0.05 | | |
|-------------------------|--------------|-------------------------|---|-------------------|
| Principe | Sub-principe | ID | Traduction | Degré de priorité |
| Findable | F1 | RDA-F1-01M | Les métadonnées sont identifiées par un identifiant permanent | Essentiel |
| Findable | F1 | RDA-F1-01D | Les données sont identifiées par un identifiant permanent | Essentiel |
| Findable | F1 | RDA-F1-02M | Les métadonnées sont identifiées par un identifiant unique au niveau mondial | Essentiel |
| Findable | F1 | RDA-F1-02D | Les données sont identifiées par un identifiant unique au niveau mondial | Essentiel |
| Findable | F2 | RDA-F2-01M | Des métadonnées riches sont fournies pour permettre la découverte | Essentiel |
| Findable | F3 | RDA-F3-01M | Les métadonnées comprennent l'identifiant des données | Essentiel |
| Findable | F4 | RDA-F4-01M | Les métadonnées sont proposées de telle sorte qu'elles peuvent être diffusées, récoltées et indexées | Essentiel |
| Accessible | A1 | RDA-A1-01M | Les métadonnées contiennent des informations permettant à l'utilisateur d'accéder aux données | Important |
| Accessible | A1 | RDA-A1-02M | Les métadonnées peuvent être consultées manuellement (c'est-à-dire avec une intervention humaine) | Essentiel |
| Accessible | A1 | RDA-A1-02D | Les données peuvent être consultées manuellement (c'est-à-dire avec une intervention humaine) | Essentiel |
| Accessible | A1 | RDA-A1-03M | L'identificateur de métadonnées se transforme en un enregistrement de métadonnées | Essentiel |
| Accessible | A1 | RDA-A1-03D | L'identificateur de données est un objet numérique | Essentiel |
| Accessible | A1 | RDA-A1-04M | Les métadonnées sont accessibles par le biais d'un protocole standardisé | Essentiel |
| Accessible | A1 | RDA-A1-04D | Les données sont accessibles via un protocole standardisé | Essentiel |
| Accessible | A1 | RDA-A1-05D | Les données peuvent être consultées automatiquement (c'est-à-dire par un programme informatique) | Important |
| Accessible | A1.1 | RDA-A1.1-01M | Les métadonnées sont accessibles via un protocole d'accès libre | Essentiel |
| Accessible | A1.1 | RDA-A1.1-01D | Les données sont accessibles via un protocole d'accès libre | Important |
| Accessible | A1.2 | RDA-A1.2-02D | Les données sont accessibles via un protocole d'accès qui prend en charge l'authentification et l'autorisation | Utile |
| Accessible | A2 | RDA-A2-01M | Les métadonnées resteront disponibles lorsque les données ne sont plus disponibles | Essentiel |
| Interoperable | I1 | RDA-I1-01M | Les métadonnées utilisent une représentation des connaissances exprimée dans un format standardisé | Important |
| Interoperable | I1 | RDA-I1-01D | Les données utilisent une représentation des connaissances exprimée dans un format standardisé | Important |
| Interoperable | I1 | RDA-I1-02M | Les métadonnées utilisent une représentation des connaissances compréhensible par la machine | Important |
| Interoperable | I1 | RDA-I1-02D | Les données utilisent une représentation des connaissances compréhensible par la machine | Important |
| Interoperable | I2 | RDA-I2-01M | Les métadonnées utilisent des vocabulaires conformes aux principes FAIR | Important |
| Interoperable | I2 | RDA-I2-01D | Les données utilisent des vocabulaires conformes aux principes FAIR | Utile |
| Interoperable | I3 | RDA-I3-01M | Les métadonnées comprennent des références à d'autres métadonnées | Important |
| Interoperable | I3 | RDA-I3-01D | Les données comprennent des références à d'autres données | Utile |
| Interoperable | I3 | RDA-I3-02M | Les métadonnées comprennent des références à d'autres données | Utile |
| Interoperable | I3 | RDA-I3-02D | Les données comprennent des références qualifiées à d'autres données | Utile |
| Interoperable | I3 | RDA-I3-03M | Les métadonnées comprennent des références qualifiées à d'autres métadonnées | Important |
| Interoperable | I3 | RDA-I3-04M | Les métadonnées comprennent des références qualifiées à d'autres données | Utile |
| Reusable | R1 | RDA-R1-01M | Une pluralité d'attributs précis et pertinents sont fournis pour permettre la réutilisation | Essentiel |
| Reusable | R1.1 | RDA-R1.1-01M | Les métadonnées comprennent des informations sur la licence en vertu de laquelle les données peuvent être réutilisées | Essentiel |
| Reusable | R1.1 | RDA-R1.1-02M | Les métadonnées font référence à une licence de réutilisation standard | Important |
| Reusable | R1.1 | RDA-R1.1-03M | Les métadonnées font référence à une licence de réutilisation compréhensible par la machine | Important |
| Reusable | R1.2 | RDA-R1.2-01M | Les métadonnées comprennent des informations sur la provenance selon des normes communautaires spécifiques | Important |
| Reusable | R1.2 | RDA-R1.2-02M | Les métadonnées comprennent des informations sur la provenance selon une langue intercommunautaire | Utile |
| Reusable | R1.3 | RDA-R1.3-01M | Les métadonnées sont conformes à la norme d'une communauté | Essentiel |
| Reusable | R1.3 | RDA-R1.3-01D | Les données sont conformes à la norme d'une communauté | Essentiel |
| Reusable | R1.3 | RDA-R1.3-02M | Les métadonnées sont exprimées conformément à la norme d'une communauté compréhensible par la machine | Essentiel |
| Reusable | R1.3 | RDA-R1.3-02D | Les données sont exprimées conformément à la norme d'une communauté compréhensible par les machines | Important |

Nakala



Partager, publier et valoriser vos données scientifiques



Je dépose mes données dans nakala

Rechercher, citer et réutiliser des données scientifiques

Rechercher dans NAKALA...



Nakala est...

- mieux que  car il fournit des identifiants pérennes aux données, le point d'accès sparql et le formatage des données en rdf ;
- aussi bien et fiable que  (CERN) mais avec le point d'accès sparql, le formatage des données en rdf et l'API;
- **un service public.**

Dans l'absolu, utiliser Nakala n'empêche pas le recours à d'autres solutions

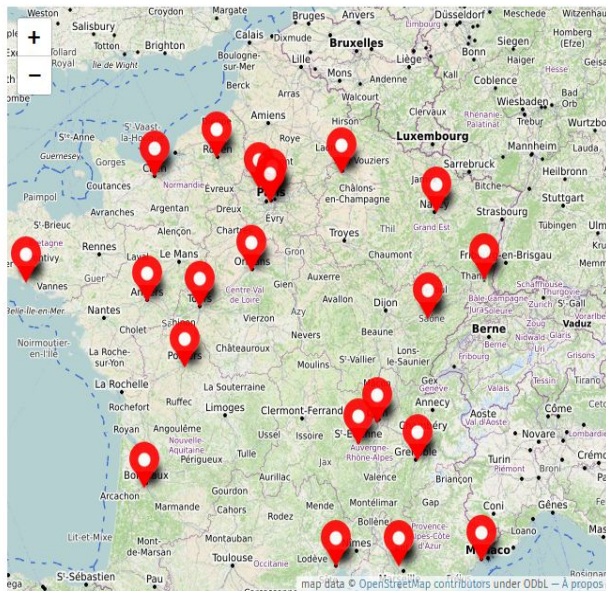


Méthodes et solutions apportées aux problèmes rencontrés

CAHIER est un consortium qui a réuni des projets pendant 10 ans



Les projets membres du consortium



Carte des membres du consortium CAHIER par LaurenceRageot

Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation. Édition. Recherche

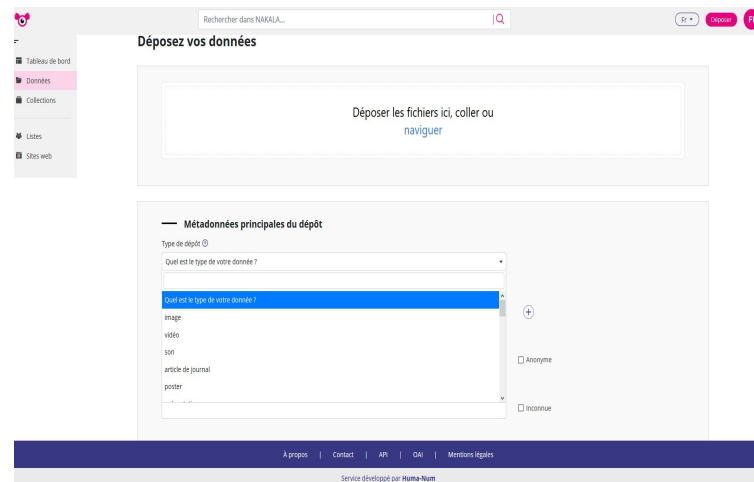
CORPUS

- Archives e-LV
- Base de français médiéval
- BIBLINDEX
- Bibliothèques Virtuelles Humanistes
- Boissy
- Bouvard et Pécuchet
- Charles Fontaine
- Corpus 14
- Corpus 14 – Ortolang
- Correspondance Bayle
- Correspondance de Flaubert
- Création de corpus de livrets d'opéra sous l'ancien régime, l'Académie royale de musique de Paris
- De piscibus (Ichtya)
- E-stampages
- E2E
- E2E
- Édition électronique de la correspondance de Flaubert
- ENCCRE
- Etudes sur la Renaissance d'Horace
- Jean-François Séguier
- Journaux d'Alexandre Dumas
- LAFABREV
- Les Écrits de l'abbé de Saint-Pierre
- Malaterra
- Manuscrits de Stendhal
- Marc-Michel Rey
- Mauriac en ligne
- Mercure galant
- Montédite
- Montesquieu. Bibliothèque & éditions
- NENUFAR
- POR FAVOR
- Privilèges de librairie
- Projet Réticence
- Projets numériques du Groupe D'Alembert
- Renaissances d'Horace
- Schola Rhetorica
- Tacitus On Line
- Testaments de Poilus

Défi 1

Maîtriser le cycle de vie des données

Nakala 2: une interface dorénavant “user friendly”



- ❖ Fonctionnalités simples et claires: Nakala peut être pris en main rapidement par les projets “débutants”
 - ❖ Idéal pour les collègues qui travaillent dans des laboratoires disposant de peu de moyens ou sans personnel d’accompagnement
- **Solution apportée par CAHIER** : “Guide pour la FAIRisation des données des Corpus d’auteurs” (<https://halshs.archives-ouvertes.fr/halshs-03037748>) dont le but est d’accompagner et de formuler des recommandations

Défi 2

Traiter la quantité de données

Interface / API

Une API “machine friendly”



Nakala - API 1.0.0

[Base URL: api.nakala.fr]

Cette API vous permet de créer, gérer et administrer vos dépôts sur la plateforme NAKALA.

Déposez vos données, exposez vos données, valorisez vos données, etc.

Authentifiez-vous

search Recherche données et collections de Nakala

GET /methods/search Récupération des adresses associées aux données de Nakala

GET /search Recherche des données Nakala

datas Gestion des données

GET /datas/{identifiant} Récupération des informations sur une donnée

PUT /datas/{identifiant} Modification des informations d'une donnée

DELETE /datas/{identifiant} Suppression d'une donnée

POST /datas Dépot d'une nouvelle donnée

GET /datas/{identifiant}/files Récupération des métadonnées des fichiers associés à une donnée

POST /datas/{identifiant}/files Ajout d'un fichier à une donnée

DELETE /datas/{identifiant}/files/{fileIdentifiant} Suppression de fichier à une donnée

- ❖ Idéal pour les dépôts de grandes quantités de données (après 10 ans, CAHIER a accumulé les données!)
 - ❖ MAIS : fonctionnalités qui ne sont pas à la portée de tous, compétences en programmation nécessaires
 - ❖ API = frein pour les collègues qui travaillent dans des laboratoires disposant de peu de moyens ou sans personnel d'accompagnement
- **Solution apportée par CAHIER** : sur la base du “Guide pour la FAIRisation des données des Corpus d’auteurs” (<https://halshs.archives-ouvertes.fr/halshs-03037748>), **une application web** destinée à faciliter les dépôts en masse

Défi 3

Traiter la quantité et préserver la qualité des données

L'API de Nakala comme service

Dialoguer avec l'API de Nakala: l'application "myinkl"



- **Répondre au problème ponctuel** : le dépôt en masse ne doit pas être un frein à la FAIRisation des données y compris en l'absence de personnel dédié dans les institutions
- **Préserver le travail du chercheur** et notamment la qualité des métadonnées préparées (rien n'est supprimé grâce aux *mappings* et à l'utilisation "extensive" des dcterms : CAHIER fournit cette analyse de l'outil, cette étude a été facilitée par **la collaboration avec l'équipe d'Huma-Num**)
- **Préserver la richesse de l'encodage XML-TEI** (les métadonnées attendues par Nakala sont ajoutées mais rien n'est supprimé)
- **Préserver toutes les données et les processus mis en place par les projets** : tous les "avatars" sont déposés (jpeg, pdf, transcriptions, etc.)
- **Faciliter la "citabilité"** : les identifiants pérennes que fournit Nakala sont récupérés automatiquement

Pas de stockage d'information, on se connecte à l'application avec sa clé API (compte nakala)

Résultats

Dépôt d'images

Le projet “Bouvard et Pécuchet”

| Fichiers | Visualisation |
|---|---|
| <ul style="list-style-type: none">g226_3_f_033_r____.jpegg226_3_f_033_r____-ver3.xmlg226_3_f_034_r____.jpegg226_3_f_034_r____-ver3.xml |  |

— — —

Dépôt de fichiers XML-TEI

Le projet “Basnage”


* V2_DU_LettreK

ID : 11280/c90ded25

Publiée

Auteur : Henri Basnage de Beauval

Ceci est la V2 du Dictionnaire universel, revu et corrigé en 1701 par Basnage de Beauval. Cette version a été obtenue avec GROBID dictionaries, après adaptation du système et reentraînement en juillet 2020.

| Fichier | Visualisation |
|---|--|
|  | <pre data-bbox="1236 333 1835 737"><?xml version="1.0" encoding="UTF-8"?> <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="applic <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="applic schematypens="http://purl.oclc.org/dsdl/schematron"> <TEI xmlns="http://www.tei-c.org/ns/1.0"> <teiHeader> <fileDesc> <titleStmt> <title type="main" xml:lang="fre">Dictionnaire universel contenant généralement tous les mots français tant vieux que modernes et les termes des sciences et des arts,... 1701. - Une édition numérique</title> <title type="sub">Lettre </title> <author key="Basnage, Henri (1657-1710)" corresp="http://www.isni.org/isni/000000008073257"> <forename>Henri</forename> <surname>Basnage</surname> </author> <author key="Furetière, Antoine (1619-1688)" corresp="http://www.isni.org/isni/0000000108580490"></pre> |
| ID : 11280/c90ded25/873c9a1a31cbdff3e523ae9085159a17bf2ce7fd | |

Dépôts de fichier XML-TEI et d'images

Le projet “Bouvard et Pécuchet”

Rechercher dans NAKALA...

Les dossiers de Bouvard et Pécuchet
ID : 10.34847/nk.61631c5 [Publiée](#)

Auteurs : Stéphanie Dord-Crouseil et Gustave Flaubert

Les dossiers de Bouvard et Pécuchet, le dernier roman - posthume et achevé - de Gustave Flaubert (1821-1880), constituent un ensemble patrimonial imposant (2 400 feuillets conservés à la Bibliothèque municipale de Rouen), cohérent, d'importance scientifique et culturelle reconnue. Ce chantier documentaire a servi à rédiger le premier volume de l'œuvre et aurait dû être réutilisé pour la composition...

[Voir plus](#) ▾

Fichiers

- g226_3_f_033_r_...jpeg
- g226_3_f_033_r_...ver3.xml**
- g226_3_f_034_r_...jpeg
- g226_3_f_034_r_...ver3.xml

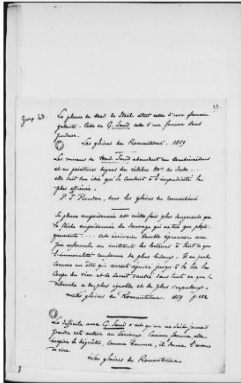
Visualisation

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:tei="http://www.w3.org/2001/TEI-Include" xmlns:dpp="http://www.dossiers-flaubert.fr/ns/1.0" xmlns:svg="http://www.w3.org/2000/svg" xmlns:math="http://www.w3.org/1998/Math/MathML" xml:lang="fr" g226_3_f_033_r_...>
<filedesc>
<titlestmt>
<title>Cote : g226_3_f_033_r_... | ID_folio : 1997 | ID_Transcription : 2846 | ID_ina
</title>
<urlbase>Stéphanie Dord-Crouseil</urlbase>
<respinit uri="mailto:sd">
<resp>transcription du manuscrit</resp>
<name>Julie Robert</name>
</respinit>
<respinit uri="mailto:sd">
<resp>Validation de la transcription</resp>
<name>Stéphanie Dord-Crouseil</name>
</respinit>
<respinit uri="mailto:sd">
<resp>Script de conversion des transcriptions saisies dans la base de données</resp>
<name>Stephane Tournayre</name>
</respinit>
</filedesc>
```

Fichiers

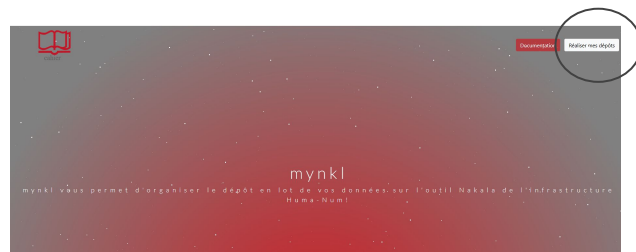
- g226_3_f_033_r_...jpeg
- g226_3_f_033_r_...ver3.xml
- g226_3_f_034_r_...jpeg
- g226_3_f_034_r_...ver3.xml

Visualisation



Dépôt de données en masse

A l'aide de l'application
“myinkl”



Exemple : déposer en une fois **2403 fichiers** décrivant, à l'aide d'une **trentaine de métadonnées** les **18736 fichiers images** du fonds Julio Córdazar

Mon profil

| | |
|--------------|----------------------|
| Identifiant | consortiumcahier |
| Prénom | Consortium |
| Nom | Cahier |
| Adresse mail | <input type="text"/> |

CC

[✎ Éditer mon profil](#)

Ma clé d'api

Clé d'api

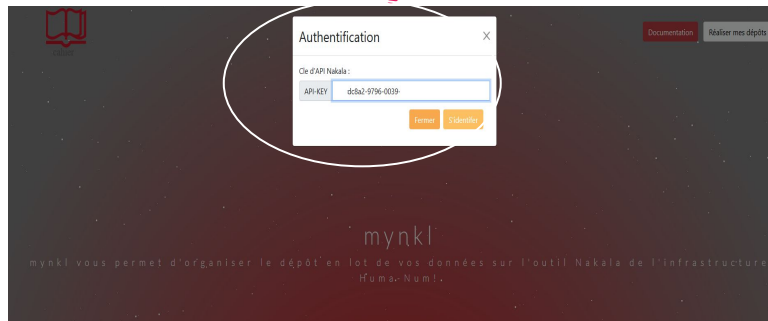
dc8a2-9796

[🔄 Générer une nouvelle clé d'api](#)

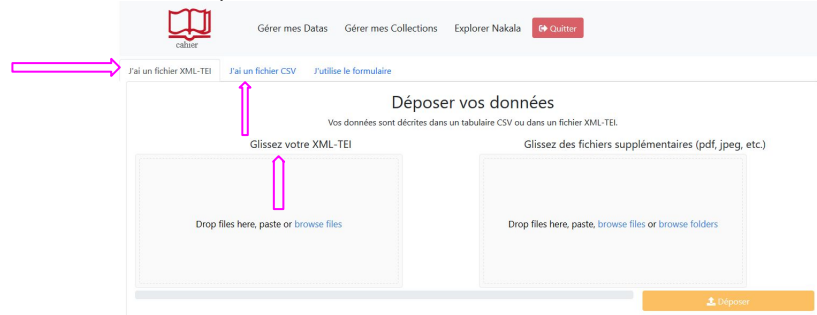
→ Développement de l'application par Ala Eddine Laouir (M2 Informatique, UHA)

Dépôt de données en masse... à l'aide de *myinkl*

A l'aide de l'application
“myinkl”



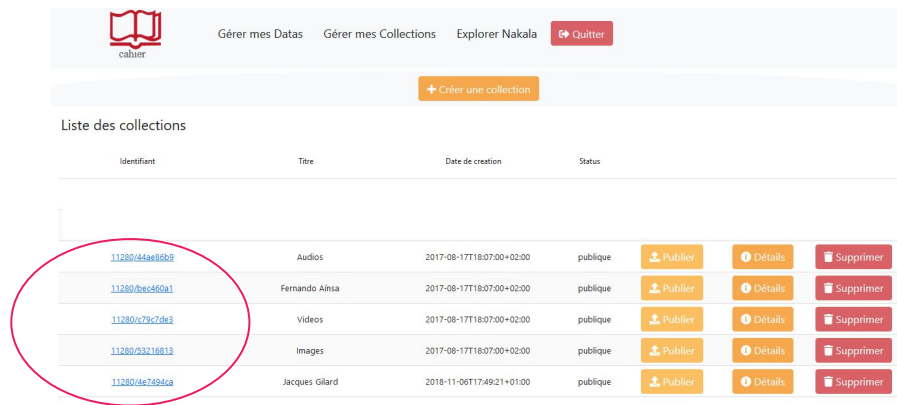
La connexion à Nakala passe par le compte Nakala de l'utilisateur: pas de stockage de données, pas de collecte d'informations privées



Si les données ont été préparées selon les recommandations de CAHIER, alors il n'y a pas de problèmes: l'application assure le mapping et dépose l'ensemble des données sur Nakala via l'API. Le dépôt est possible selon différentes collections

Récupérer les identifiants et citer mes données... à l'aide de *myinkl*

A l'aide de l'application
“myinkl”



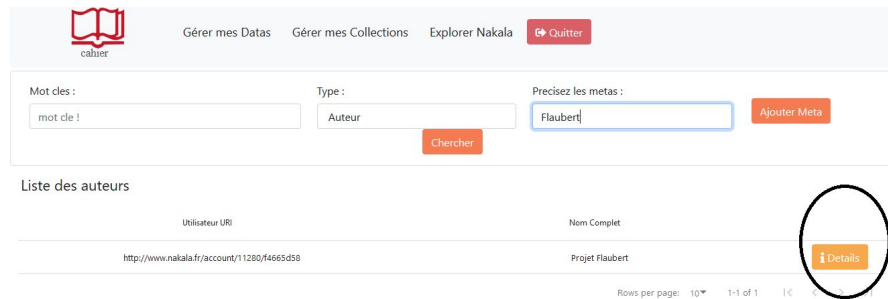
The screenshot shows the Myinkl application interface. At the top, there is a navigation bar with a logo and buttons for "Gérer mes Datas", "Gérer mes Collections", "Explorer Nakala", and "Quitter". Below this is a button to "Créer une collection". The main section is titled "Liste des collections" and contains a table with the following columns: "Identifiant", "Titre", "Date de creation", "Status", and three action buttons: "Publier", "Détails", and "Supprimer". The first row of the table is circled in red.

| Identifiant | Titre | Date de creation | Status | Publier | Détails | Supprimer |
|--------------------------------|----------------|---------------------------|----------|---------|---------|-----------|
| 11280/44e888b2 | Audios | 2017-08-17T18:07:00+02:00 | publique | Publier | Détails | Supprimer |
| 11280/0ec460a1 | Fernando Ainsa | 2017-08-17T18:07:00+02:00 | publique | Publier | Détails | Supprimer |
| 11280/c79c74e2 | Videos | 2017-08-17T18:07:00+02:00 | publique | Publier | Détails | Supprimer |
| 11280/33216812 | Images | 2017-08-17T18:07:00+02:00 | publique | Publier | Détails | Supprimer |
| 11280/4e7494ca | Jacques Gilard | 2018-11-06T17:49:21+01:00 | publique | Publier | Détails | Supprimer |

L'identifiant attribué par Nakala apparaît dans l'interface. Il est possible de mettre à jour les données en y intégrant le DOI.

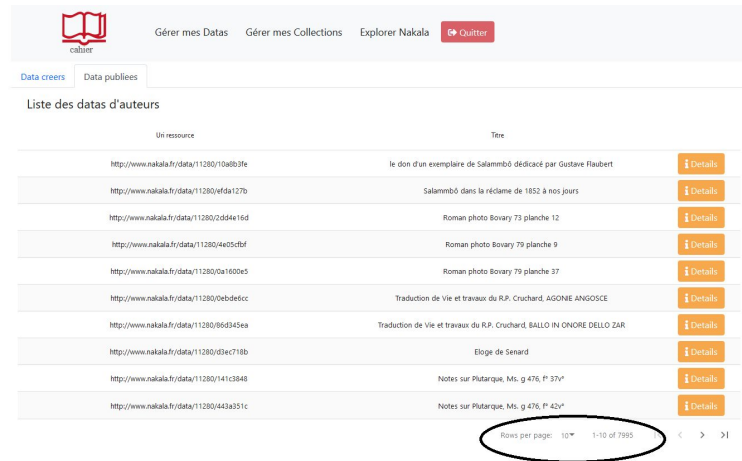
Explorer Nakala... à l'aide de *myinkl*

A l'aide de l'application
“myinkl”



The screenshot shows the top navigation bar with the 'cultural' logo and links for 'Gérer mes Datas', 'Gérer mes Collections', 'Explorer Nakala', and a 'Quitter' button. Below the navigation is a search section with three input fields: 'Mot cle :' containing 'mot cle !', 'Type :' containing 'Auteur', and 'Précisez les metas :' containing 'Flaubert'. There are 'Ajouter Meta' and 'Chercher' buttons. Below the search section is a table titled 'Liste des auteurs' with columns 'Utilisateur URI' and 'Nom Complet'. The first row shows 'http://www.nakala.fr/account/11280/46665d58' and 'Projet Flaubert'. A 'Details' button is circled in red on the right side of the table. At the bottom right, there is a pagination control showing 'Rows per page: 10', '1-1 of 1', and navigation arrows.

Fouiller les données de Nakala tout en tenant compte des dépôts habituels : l'application retourne des résultats classés en fonction des “habitudes de dépôt”.



The screenshot shows the top navigation bar with the 'cultural' logo and links for 'Gérer mes Datas', 'Gérer mes Collections', 'Explorer Nakala', and a 'Quitter' button. Below the navigation is a section with 'Data creés' and 'Data publiées' tabs. Below this is a table titled 'Liste des datas d'auteurs' with columns 'Un ressource' and 'Titre'. The table lists several data items with their URIs and titles, each with a 'Details' button. The first row shows 'http://www.nakala.fr/data/11280/10a03fe' and 'le don d'un exemplaire de Salammbô dédié par Gustave Flaubert'. The last row shows 'http://www.nakala.fr/data/11280/443a351c' and 'Notes sur Plutarque. Ms. g 476, P 42v'. A 'Details' button is circled in red on the right side of the table. At the bottom right, there is a pagination control showing 'Rows per page: 10', '1-10 of 7995', and navigation arrows.

Conclusion

Rendre les données de CAHIER “FAIR” est possible....

mais il reste des problèmes qui demeurent actuellement sans solutions...

(2011-2021)



Une histoire de choix

Les choix techniques d'hier ont des conséquences aujourd'hui

Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

2005-2011

Choisir une
solution
technique

2011-2016

Revenir en
arrière ou
persister

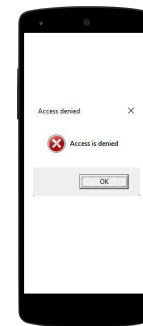
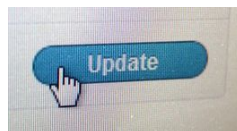
2016-2021

Attendre

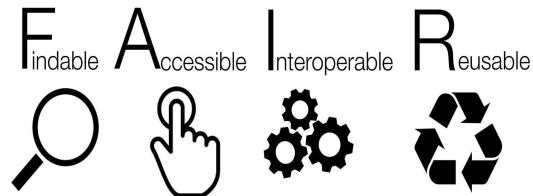
Comment
FAIR?



Mon nouveau site web!



Le site web : the “FAIR” illusion



“J’expose donc je suis”



Des réponses (presque vraies) à la question:

F : “oui mes données sont trouvables, elles sont en ligne sur le site [“www.monprojet.monlabo.fr”](http://www.monprojet.monlabo.fr)”

A : “mais enfin c’est accessible et pérenne, mon site est chez Huma-Num : www.monprojet.huma-num.fr”

I : “t’inquiète, derrière c’est du XML / c’est du Dublin Core, donc c’est interoperable!”

R : “tu peux tout télécharger depuis mon site www.monprojet.monlabo.huma-num.fr si tu veux! Sinon envoie un mail à soljeanpat@monlabo.univ.fr pour demander l’autorisation”

FAIR data ou Terra utopia?

“Data are **capta**” (Johanna Drucker “Humanities Approaches to Graphical Display”, 2011)... “so they are **MY** data! Et après tout ce travail, pourquoi je partagerais ?”

OUVRIR

LA SCIENCE

SCIENCE OUVERTE COMITÉ GROUPES BLOG AGENDA RESSOURCES LETTRES

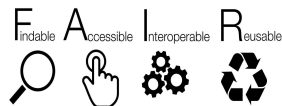
FR - EN

Accueil > Ressources > FAIR Principles



TEXTES FONDATEURS

Les principes FAIR sont un ensemble de principes directeurs pour gérer les données de la recherche visant à les rendre faciles à trouver, accessibles, interopérables et réutilisables par l'homme et la machine.



https://vidensportal.deic.dk/sites/default/files/uploads/A%20oFAIRy%20otale%20obook%20digital_ny.pdf

Merci pour votre attention

— — —

En savoir plus sur Nakala :

<https://documentation.huma-num.fr/nakala/>

En savoir plus sur CAHIER :

<https://cahier.hypotheses.org/le-consortium>

Contact:

laurence.rageot@univ-tours.fr

Suivre le colloque des 10ans

<https://cahier10.sciencesconf.org/>

Fatiha Idmhand

fatihaidmhand[at]yahoo.es

Le consortium CAHIER
Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche



10 ans
de corpus d'auteurs

IUT, 1 Rue Jacques Ellul
33800 Bordeaux

Du 07 au 10 juin 2021

Programme: cahier10.sciencesconf.org
Contacts: cahier10@sciencesconf.org



Merci à

Le consortium CAHIER

Corpus d'Auteurs pour les Humanités
Informatisation, Édition, Recherche

Ala Eddine Laouir
Andrés Felipe Echavarría Pelaez
Julie Laurent
Laurent Passion

Ioana Galleron

et bien sûr

Laurent Capelli
Adrien Desseigne



10 ans de corpus d'auteurs

IUT, 1 Rue Jacques Ellul
33800 Bordeaux

Du 07 au 10 juin 2021

Programme: cahier10.sciencesconf.org
Contacts: cahier10@sciencesconf.org

