

« Dix ans avec CAHIER »...

Assemblée générale 2021



consortium
cahier



Huma-Num

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



HN
Huma-Num

 consortium
cahier

 e-ana

 Université
de Nouvelle-
Aquitaine

 Université
de Bordeaux

 Université
d'Auvergne

 Université
de MONTAIGNE

Dix ans ...

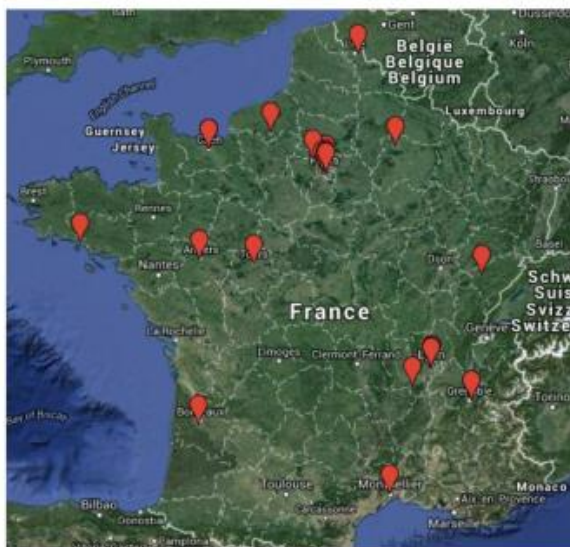
- « Il vaut mieux dix ans glaner. Qu'une seule année moissonner. »
(Proverbe Français)



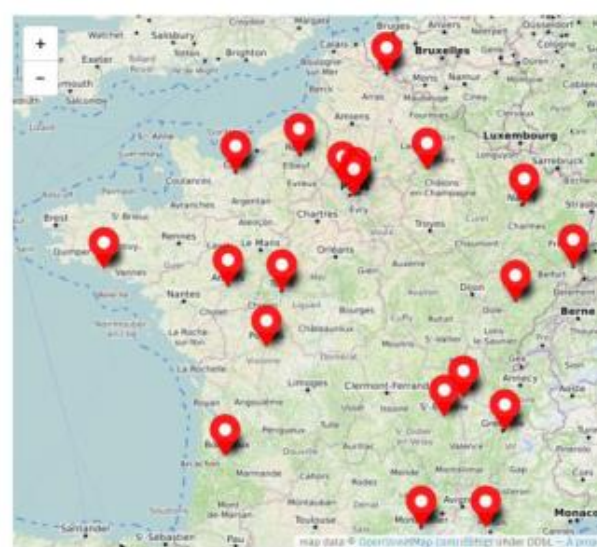
Dix ans pour tisser un réseau national



12... 65... 58...



Carte du réseau en 2014



Carte du réseau en 2017



Carte du réseau en 2021



Un réseau de projets...

Cf. Guide « *Les publications numériques de corpus d'auteurs - Guide de travail, grille d'analyse et recommandations* ».

❖ **32 projets d'archives éditorialisées** : projets offrant une scénarisation minimale des ressources en lien avec la question de recherche. C'est le cheminement scientifique qui organise le « vrac » des documents.

La plupart des projets qui ont opté pour ce type de publications s'inscrivaient dans les domaines de la génétique des œuvres et de la critique textuelle.

❖ **26 projets d'éditions enrichies** : publications qui proposent un texte profondément enrichi d'informations documentaires et contextuelles. Type d'édition suit nécessairement des pratiques harmonisées et se réfère à des standards comme ceux du consortium TEI.

Les projets ayant opté pour ce type d'éditions s'inscrivaient dans le domaine de la philologie numérique.

❖ **Aucune édition de lecture** (« reading edition ») n'a été proposée.

Publications, relues, corrigées et nettoyées des erreurs humaines (ou de celles qui sont issues de la reconnaissance automatique de caractères)



Défi de « CAHIER après CAHIER »

Internationaliser le réseau

Déployer les trois types de publications qui correspondent à différents « états du projet scientifique »



Dix ans, trois axes, des livrables



Axe 1

La construction de données « interoperables », FAIR

L'objectif de CAHIER était d'acculturer la communauté scientifique à l'utilisation de standards et de normes ouvertes facilitant l'interopérabilité et donc, la réutilisation des données

Pour atteindre cet objectif, CAHIER a encouragé

- l'acquisition de données de qualité (image et texte)
- le partage de normes de transcription
- le testé de dispositifs d'affichage « du mode texte et du mode image » de façon à opérer des choix pertinents en fonction des publics mais également de la qualité des données produites.



Bilan de l'Axe 1

- ❖ 6 publications de type guides méthodologiques
- ❖ 61.107 ressources ouvertes dont 57.073 images et 4.034 fichiers XML-TEI
 - o 1604 images
 - o **1793 fichiers XML-TEI disponibles sur Ortolang** (1 projet a utilisé Ortolang)
 - o 167 images et 167 fichiers XML-TEI disponibles sur Zenodo (1 projet a utilisé Zenodo)
 - o **~55.302 images et ~16.030 fichiers de données à ce jour dont 2077 fichiers XML TEI** (6 projets ont utilisé Nakala)
 - o 439 fichiers XML-TEI sont, par ailleurs, disponibles sur github (2 projets ont utilisé github)
- ❖ 3 083 résultats (au 26/10/21) signalées dans Isidore
- ❖ 3 outils de communication à destination de la communauté (1 blog, 2 comptes sur les réseaux sociaux, 2 listes de diffusion mail)



Défi de « CAHIER après CAHIER »

Accroître le nombre
de fichiers totalement « FAIR »

Poursuivre l'accompagnement



Axe 2

L'approche critique des pratiques de publication numérique de corpus d'auteurs

Le consortium s'était également donné comme objectif de **travailler sur l'évaluation des pratiques d'édition numérique** en fonction des projets.:

- ❖ de discuter les critères (modes d'accès, volumétrie, outils utilisés, adéquation aux objectifs du consortium) et choix en vue d'harmoniser des pratiques.
- ❖ d'interroger les conditions de réalisation des éditions numériques de façon à travailler sur les éventuels verrous et tenter de les lever.



Bilan de l'Axe 2

Un pas vers l'évaluation...

- ❖ 1 guide « Les publications numériques de corpus d'auteurs - Guide de travail, grille d'analyse et recommandations ».

Plusieurs actions pour faire le bilan et échanger sur les pratiques

- ❖ 1 colloque scientifique

et

- ❖ 23 participations à des colloques scientifiques

- ❖ 1 ouvrage collectif (sous presse)

- ❖ **39** publications recensées sur HAL pour promouvoir ces travaux



Défi de « CAHIER après CAHIER »

Promouvoir encore ces grilles et travaux (traduction, présentation publique, etc)

Développer un service?



AXE 3

Les outils numériques des sciences des textes

Le consortium s'était donné pour but de tester les prototypes qui « circulent dans le domaine des humanités numériques » afin:

- ❖ d'en mesurer l'efficacité pour la construction et/ou l'exploitation de corpus numériques
- ❖ de former aux outils repérés ou utilisés par la communauté scientifique afin d'étendre son panel méthodologique
- ❖ soutenir, le cas échéant, des développements informatiques.

Bilan de l'Axe 3

- ❖ **14** formations ou écoles thématiques
- ❖ 3 outils numériques développés pour la communauté scientifique de CAHIER et au-delà
- ❖ soutien au développement d'au moins 3 outils développés par d'autres communautés scientifiques

Défi de « CAHIER après CAHIER »

Poursuivre les formations et travailler sur le public cible?

Développer d'autres outils pour rendre d'autres services?



Après « CAHIER »



La suite du consortium

- ❖ 4 + 4 + 2 (2019-2021) : **crise sanitaire et confinements**
- ❖ Etude des 3 scénarios recommandés pour l'« après CAHIER »
- ✓ Réseau au sein du RnMsh
- ✓ Réseau « Méthodes et pratiques »
- ✓ Nouveau (X) projet (S) de consortium

- *Discussions difficiles en raison de la distance et du manque de préparation, par CAHIER, de cette concertation*
- *Accélération entre juin et septembre*

« CAHIER après CAHIER »

- ❖ Relever les défis identifiés
- ❖ Elargir la communauté dans le domaine des sciences humaines, et à l'échelle internationale
- ❖ Préciser les questions du projet scientifique

Défi de « CAHIER après CAHIER »

Rédiger le nouveau projet avant juin
2022



« CAHIER » en 2021



3 groupes de travail, 3 défis

- ❖ Décrire les genres des textes
- ❖ Penser la (Ré)utilisabilité
- ❖ FAIRiser les données : stockage des données et Plan de gestion des données

3 groupes de travail, 3 défis

- ❖ Décrire les genres des textes
- ❖ Penser la (Ré)utilisabilité
- ❖ FAIRiser les données : stockage des données et Plan de gestion des données

Groupe de travail « Typologie Textuelle »

(novembre 2020-novembre 2021)

(<https://openthese.humanum.fr/openthese/api/theso/Typologie>)

Typologie textuelle

- Réunions (en visioconférence et en présentiel lors de l'AG et du colloque de Bordeaux) :
- 27 novembre 2020
- 17 décembre 2020
- 27 janvier 2021
- 4 mars
- 26 avril
- 5 mai (crash-test)
- 10 juin (Bordeaux)
- 24 novembre 2021, Paris, en mode hybride

Travaux

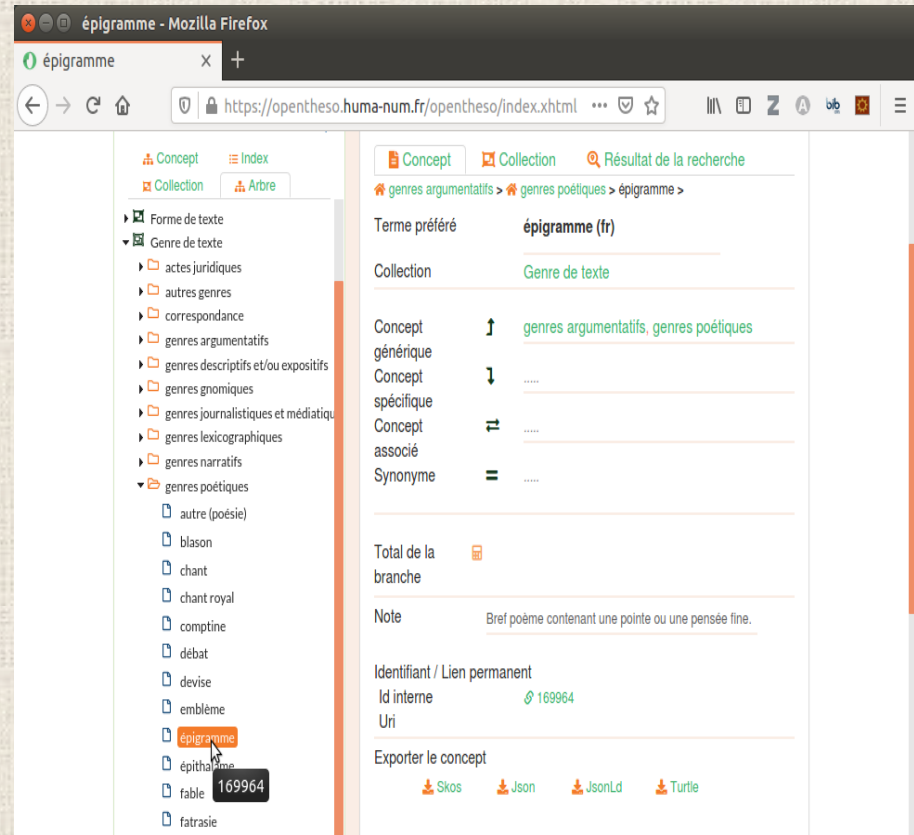
Rappel de la création et objectifs:

Groupe de travail créé en 2015 [et non 2019 comme dans le rapport général de Cahier], il s'est donné pour objectif de développer une ontologie des typologies de textes à l'aide de l'outil OpenTheso (développé par Miled Rousset, <https://github.com/miledrousset/opentheso>).

- Faire avancer les principes FAIR de CAHIER en offrant une ressource accessible pour l'indexation des données textuelles, l'ambition étant d'associer à chaque concept un identifiant stable.
- 5 actions cette année:
 - Redistribution de la hiérarchie
 - Identifiants handle
 - Un brouillon d'article (première rédaction par Ioana Galleron)
 - Un guide de bonnes pratiques (préparé par Alexei Lavrentiev)
 - Des tests

1. Reprise de la hiérarchie des genres

- La hiérarchie des hypergenres ou micro-thésaurus a été entièrement redistribuée en neuf groupes, chacun sur trois niveaux au maximum, afin de bien identifier les nœuds, les classes et les propriétés



The screenshot shows a web browser window titled "épigramme - Mozilla Firefox" with the URL "https://opentheso.huma-num.fr/opentheso/Index.xhtml". The page displays a hierarchical tree of genres on the left and a detailed view of the concept "épigramme" on the right.

Left Panel (Hierarchy):

- Concept
- Index
- Collection
- Arbre
- Forme de texte
- Genre de texte
 - actes juridiques
 - autres genres
 - correspondance
 - genres argumentatifs
 - genres descriptifs et/ou expositifs
 - genres gnomiques
 - genres journalistiques et médiatiques
 - genres lexicographiques
 - genres narratifs
 - genres poétiques
 - autre (poésie)
 - blason
 - chant
 - chant royal
 - comptine
 - débat
 - dévisé
 - emblème
 - épigramme** (highlighted)
 - épithalame
 - fable
 - fatrasie

Right Panel (Concept Details):

- Concept
- Collection
- Résultat de la recherche
- genres argumentatifs > genres poétiques > épigramme >
- Terme préféré: **épigramme (fr)**
- Collection: **Genre de texte**
- Concept générique: genres argumentatifs, genres poétiques
- Concept spécifique:
- Concept associé:
- Synonyme: =
- Total de la branche: 1
- Note: Bref poème contenant une pointe ou une pensée fine.
- Identifiant / Lien permanent
 - Id interne: 169964
 - Uri:
- Exporter le concept
 - Skos
 - Json
 - JsonLd
 - Turtle

2. Attribution d'identifiants handle aux concepts

- de nouveaux concepts ont été ajoutés avec leurs définitions (365 à ce jour)
- les identifiants handle fournis par Huma-Num ont été intégrés au thésaurus (mai 2021) et sont accessibles sur le serveur (<https://opentheso.huma-num.fr/opentheso/api/theso/Typologie>)
 - Exemple pour le concept « épigramme »

Mozilla Firefox

opentheso.huma-num.fr

https://opentheso.huma-num.fr/opentheso/

Concept Index
Collection Arbre

discours
dissertation
épigramme
libelle
oraison funèbre
pamphlet
sermon
genres descriptifs et/ou expositifs
genres gnomiques
genres journalistiques et médiatiques
genres lexicographiques
genres narratifs
genres poétiques
autre (poésie)
blason
chant
chant royal
comptine
débat
devise
emblème
épigramme
épithalame
fable

Concept Collection Résultat de la recherche

genres argumentatifs > genres poétiques > épigramme >

Terme préféré **épigramme (fr)**

Collection **Genre de texte**

Concept générique ↑ **genres argumentatifs, genres poétiques**

Concept spécifique ↓

Concept associé ↔

Synonyme =

Total de la branche

Note **Bref poème contenant une pointe ou une pensée fine.**

Identifiant / Lien permanent

Id interne 169964

Uri <https://hdl.handle.net/20.500.11942/crto5sfm5na2>

IdHandle <https://hdl.handle.net/20.500.11942/crto5sfm5na2>

Exporter le concept

Skos Json JsonLd Turtle

https://hdl.handle.net/20.500.11942/crto5sfm5na2

Créé : 2021-03-06 Dernière modification 2021-03-06

3. Rédaction d'un article collaboratif

- Publication en revue avec la traduction en anglais
- « Décrire un corpus d'auteurs : un thésaurus pour les types de textes »
 - Les étapes de réalisation
 - Les principes théoriques
- Plan:
 - État des lieux
 - L'organisation intellectuelle du thésaurus
 - Exemples

4. Un Guide de bonnes pratiques

- Historique du thésaurus
- Prise en main d'openTheso
- Classement et hiérarchie des concepts
- Intégration des identifiants handle dans:
 - XML/TEI (TEI header)
 - Omeka (DC)

Publication imminente sur HAL:

https://docs.google.com/document/d/10Ay3pfTAhf0KSC_LKjIBed7yNYnl6TZFjMUUEFI

5. Des tests sur corpus

Exemple de header TEI pour un roman

- `</fileDesc>`
- `<profileDesc>`
- `<textClass>`
- `<keywords scheme="https://opentheso.humanum.fr/opentheso/api/theso/Typologie">`
- `<term type="domaine"`
`ref="https://hdl.handle.net/20.500.11942/crteb7oodgosm"`
`>littérature</term>`
- `<term type="factualite"`
`ref="https://hdl.handle.net/20.500.11942/crt22aacdhvi7"`
`>fiction</term>`
- `<term type="forme"`
`ref="https://hdl.handle.net/20.500.11942/crtnufdew11fb">prose</term>`
- `<term type="genre"`
`ref="https://hdl.handle.net/20.500.11942/crt97p3ya2hot">roman de`
`formation</term>`
- `[...]`
- `<term type="inscription"`
`ref="https://hdl.handle.net/20.500.11942/crthy6j6jf9nq">imprimé</term>`
- `</keywords>`
- `</textClass>`
- `</profileDesc>`
- `</teiHeader>`

Exemple pour Dublin Core:

The screenshot shows the 'Projeto Guarnido' Dublin Core metadata editor. The interface includes a top navigation bar with 'Projet Guarnido', 'Extensions', 'Apparence', 'Utilisateurs', 'Paramètres', and 'Se déconnecter'. A left sidebar contains a menu with items like 'Tableau de bord', 'Contenus', 'Collections', 'Types de contenu', 'Mots-clés', 'NAKALA Import', 'Simple Pages', 'CSV Import', 'NAKALA Export', 'Neatline', 'Expositions', and 'OAI-PMH Harvester'. The main content area is titled 'Description de rdfr:description' and contains a text input field with the value 'Genre: Roman'. Below this is another text input field with the URL 'https://hdl.handle.net/20.500.11942/crthuyig/pwt'. The right sidebar contains several buttons: 'Sauvegarder les changements', 'Voir la page publique', and 'Supprimer'. There are also checkboxes for 'Public' and 'Mis en avant', and a 'Collection' dropdown menu with 'Récit' selected.

2022

- Finaliser l'article
- Mettre à jour le guide
- Améliorer l'intégration des concepts dans les métadonnées
- Tester dans d'autres corpus
- Traductions (concepts et définitions): en anglais en priorité
- Intégrer de nouveaux concepts
 - Exemples:
 - Polémique, traité,
- Alignement avec les autres référentiels (Rameau...)
- Collaborations ELTEC? COSME?

Groupe de travail « Reuse »



consortium
cahier

Groupe de travail (ré)utilisabilité

créé en décembre 2020 et animé par Anne Garcia-Fernandez, Elisabeth Greslou et Richard Walter



Objectifs du groupe

- Mener une réflexion sur les actions et les outils permettant que les données, au-delà de leur seule exposition, soient concrètement (ré)utilisables et (ré)utilisées.
- Proposer des recommandations concernant le R de FAIR dans le cadre des projets de recherche

Ont participé au groupe de travail

Julie Aucagne, Céline Bohnert, Marguerite Bordry, Camille Desiles, Marie Dupond, Francine Filoche, Ioanna Galeron, Camille Koskas, Constance Krebs, Mathilde Labbé, Sylvie Le Clech, Françoise Leriche, Stéphane Loret, Jean-Sébastien Macke, Gwenaëlle Patat, Danielle Pouban, Pierre Willaime.

Sommaire du vademecum



1. Introduction
2. Pratiques et verrous de la réutilisation des données
 - 2.1 La visibilité des données
 - 2.2 Les obstacles juridiques
 - 2.3 Les facteurs humains
 - 2.4 Les verrous techniques
 - 2.5 Bilan
3. Penser la réutilisation au début du projet
 - 3.1 Préparer le terrain
 - 3.2 Méthodologie et organisation
 - 3.3 L'organisation interne
 - 3.4 Volet formation
 - 3.5 Faire un état des lieux juridique
 - 3.6 Plan de gestion de données et réutilisabilité
 - 3.7 Deux états des données
 - 3.8 Choisir ses outils
 - 3.9 Stockage
 - 3.10 Choix de communication externe

4. Penser la réutilisabilité tout au long du projet

- 4.1 Gestion de projet : grands principes
- 4.2 En pratique
- 4.3 Méthodologie
- 4.4 Trucs & astuces

5. Bilan

Annexes

- A Check-List réutilisabilité
- B Bibliographie
- C Glossaire des notions et des outils

Pour plus d'infos, le wiki du groupe de travail :

<https://groupes.renater.fr/wiki/gt-re-utilisabilite-cahier/index>

Et la page sur le site de CAHIER :

<https://cahier.hypotheses.org/activites/groupe-reutilisabilite>

Groupe de travail
[Data_CAHIER]

FAIRiser

FAIRiser vos données : mémorandum

umcahierconsortium
consortiumcahiercons

Plan de Gestion de Données du Consortium CAHIER

PGD V1: 30/10/2021, PGD Consortium CAHIER de Huma-Num

Ceci est le Plan de Gestion des Données du consortium CAHIER de l'Infrastructure Huma-Num. Il s'agit d'un plan de gestion des données de structure. De 2011 à 2021, CAHIER a travaillé à la numérisation et à la construction de données sur des corpus d'auteurs. Ce plan de gestion des données offre une vue générale des actions menées par le consortium sur les données au bout de dix ans.

CAHIER a donc rédigé deux types de plans de gestion des données : un Plan dit « de structure » et un plan par projet volontaire.

Les plans proposés par CAHIER s'inspirent des recommandations de l'Agence Nationale de la Recherche et du guide publié en 2016 par le programme Horizon Europe [1]. Les modèles mis en ligne par la plateforme DMP OPIDoR (<https://opidor.fr/>), outils d'aide à la création en ligne de plans de gestion de données (Data Management Plan ou DMP), ont été étudiés et adaptés aux besoins de la communauté scientifique du Consortium CAHIER. Le modèle de PGD proposé par CAHIER sera mis en ligne sur Opidor.

Auteurs du plan de gestion des données :

IDMHAND, Fatima, IdHAL : fatima-idmhand ; ORCID : [0000-0001-7135-8152](https://orcid.org/0000-0001-7135-8152), Université de Poitiers, Institut des textes et Manuscrits Modernes (ITEM, UMR5132), Paris, France
Rôle dans le consortium CAHIER : Coordinatrice adjointe

L'HERMITE, Laurène, Université de La Rochelle, Centre de recherches en histoire internationale et atlantique (EA1103), La Rochelle, France
Rôle dans le consortium CAHIER : Chargée de PGD

Version du plan de gestion des données :

PGD V1: 30/10/2021, PGD Consortium CAHIER de Huma-Num
Une seule version de ce PGD est actuellement prévue

Après « Dix ans avec CAHIER »...
il est temps d'imaginer « CAHIER après CAHIER »

Rendez-vous en 2022....

To be continued



consortium
cahier

The logo features the letters 'H' and 'N' in a bold, sans-serif font. The 'H' is dark red and the 'N' is yellow, both set against a white background within a red-bordered box.

Huma-Num