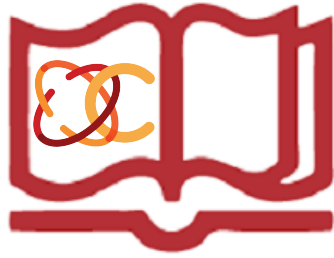


Traiter et diffuser les données des sciences des textes

La perspective du consortium CAHIER

F. Idmhand - T. Lebarbé

<http://cahier.hypotheses.org/>

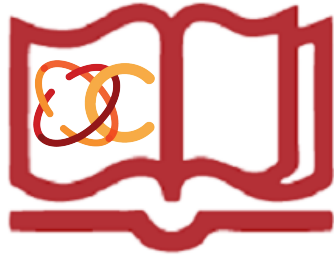


Corpus d' Auteurs pour les Humanités :

Informatisation, Edition, Recherche

- **Corpus**
=> données modélisées et structurées
- **Auteurs**
=> diversité des auteurs et des périodes
- **Informatisation**
=> processus de prod. et de conversion numérique
- **Edition**
=> matérialisation
- **Recherche**
=> manipulation / traitement

« Il faut pré-réfléchir les données »



Les « esprits » de CAH...

Informatisation

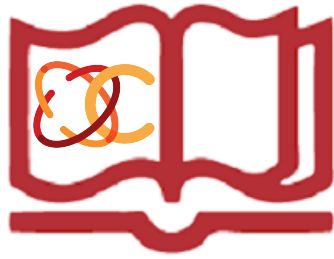
Esprit « Turing » : systématiser avec intelligence pour automatiser ... (au moins partiellement)

Edition

Esprit « érudit »: l'édition numérique pour « Mettre au jour, produire » (lat. *edere*)

Recherche

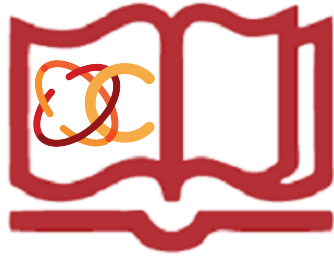
Esprit « Montaigne »: la recherche par la « pensée » pour le renouvellement



Construire le réseau scientifique

- Par les formations
 - Pour les membres
 - Ouvertes à tous
- Par les rencontres
 - = vivier de chercheur
 - = une représentation locale
 - = un mapping des MSH
(presque)

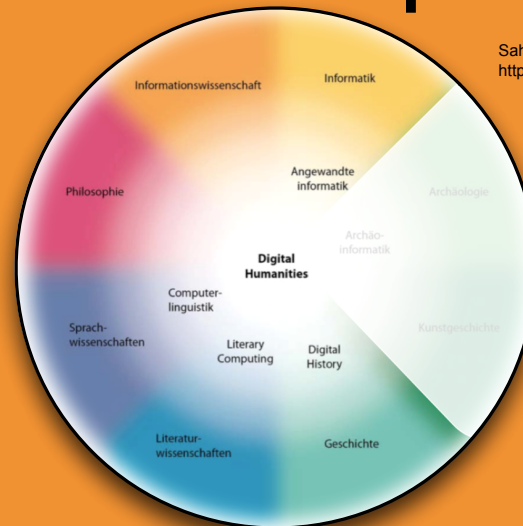




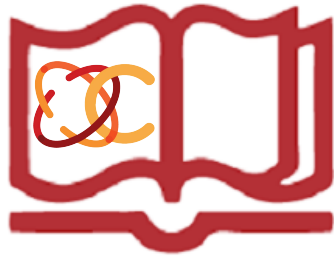
Contribuer aux bonnes pratiques

Une couverture disciplinaire large

- Les formations
 - par CAHIER
 - soutenues
 - inscriptions
- L'atelier annuel
- Lien entre les « communs » et la discipline → convergence autour du matériau

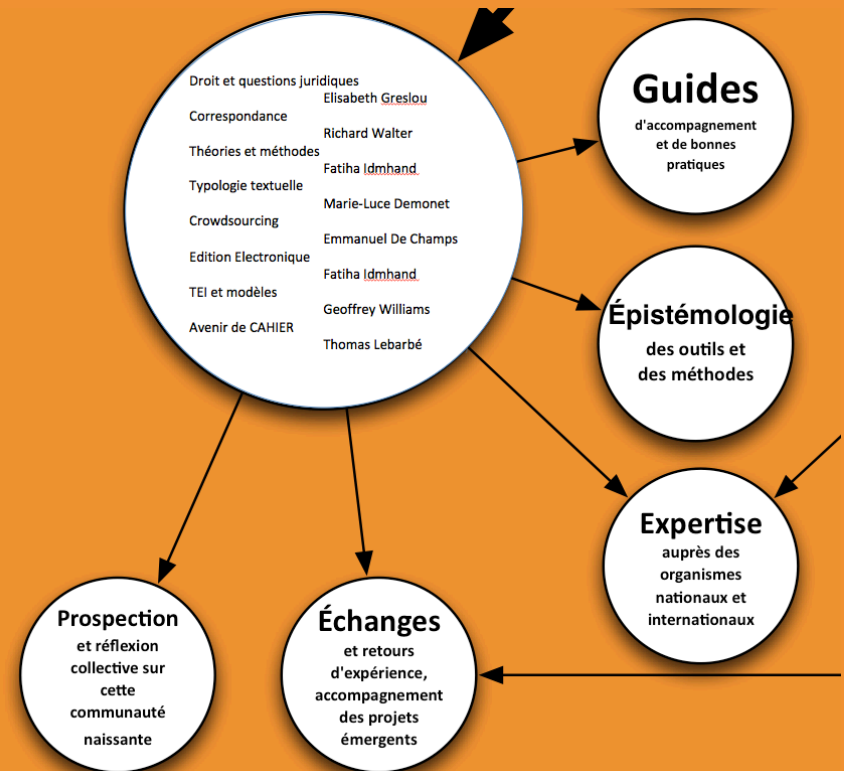


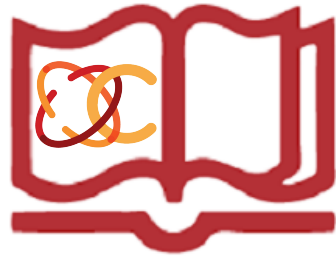
Sahle P., *DH Studieren !*, DARIAH-DE Working Papers, 2013,
<http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2013-1.pdf>



Construire des connaissances

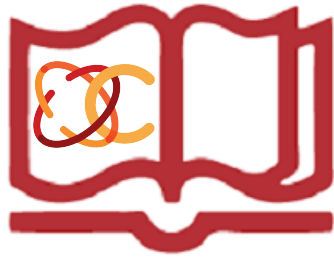
- Les GdT
- Les projets des chercheurs et leurs questions de recherche
- Les partenariats avec les associations (TEI, Humanistica, ADHO)





CAHIER un consortium singulier

- Deux principes : la demande d'adhésion et un dépôt de projet
- Une singularité : le corpus et non la période.
- Des corpus d'auteurs et la convergence des approches philologiques, littéraires et génétiques des textes
- Engager les sciences du textes à relever le défi de l'informatisation des textes



CAHIER

Les « 2 âges »

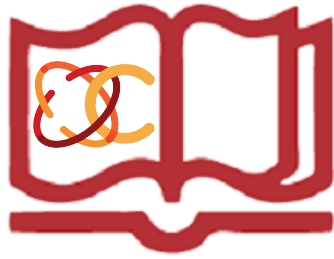
Temps 1 (1ère labellisation 2011-15) : CRÉER LES DONNÉES

- Accompagner la conversion numérique des sources, compréhension, mettre en place des méthodes...
- Organiser la « montée en compétences » par la formation, la dissémination de "bonnes pratiques" et la création de la communauté
- Résultats : 60+ projets et 300+ personnes en France

Temps 2 (2de labellisation 2015-19) : TRAVAILLER SUR LES DONNÉES

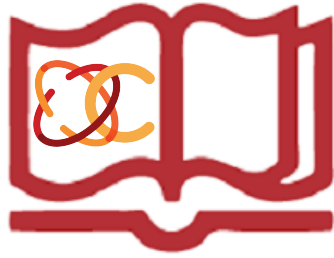
- Consolider la communauté créée et générer de nouvelles connaissances à partir des données
- Poursuivre la montée en « compétence » & asseoir les expertises de cette communauté scientifique experte du domaine des sciences des textes assistées par l'informatique > diffusions de guides)

Bientôt le 3^{ème} âge ?



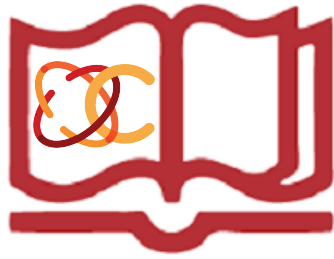
Traiter : R2CAHIER

- **Construire les nouvelles épistémologies des sciences des textes**
 - créer la « DATA SCIENCE des textes littéraires » et une sciences des métadonnées
 - repenser et formaliser des méthodes, approches théoriques et pratiques (exp: le « personnage » en questions)
 - reconsidérer les heuristiques usuelles pour aller vers les nouvelles formes/méthode de l'herméneutique
- **Premiers résultats du GT :**
 - un nouveau concept (le crowdreading comme nouvelle approche)
 - Le développement d'une d'interface pour assister l'analyse des textes et des personnages (avec Caen)
 - NB : lien avec un autre *GT Typologie des genres*



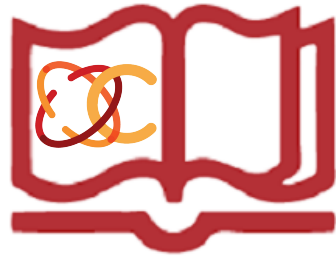
Diffuser : EVENT

- suite à DH Cracovie (2016) :
 - évaluer l'édition numérique
 - comment, avec quelles méthodes?
 - quel guide pour CAHIER (par rapport à d'autres communautés: MLA, TEI)?
- **Premiers résultats du GT :**
 - GRILLE D'ANALYSE DES EDITIONS NUMERIQUES : outil d'évaluation et d'auto "évaluation" des éditions numériques
 - NB : lien avec un autre GT qui démarre à l'automne : GT-TEI



Traiter et diffuser DATA_CAHIER

- Nouveau GT qui accompagne le stockage et la mise à disposition des données de CAHIER. Le but est le stockage de toutes les données de CAHIER sur Nakala et leur exposition sur Isidore
 - **Deux priorités :**
 - qualité des données
 - disponibilité des données (accessibles, consultables et téléchargeables et s'il s'agit d'un logiciel dit "open source", que le code soit accessible)
 - réutilisabilité
 - **Outils prioritaires: Isidore, WebOai et Nakala**
 - accompagnement personnalisé des projets pour préparer leur TEIHeaders en vue de générer un entrepôt oai et d'exposer les données sur <http://weboai.cahier.huma-num.fr/> et Isidore
 - accompagnement personnalisé des projets pour le stockage de leurs données sur Nakala
- **Premiers résultats du GT :** <http://weboai.cahier.huma-num.fr/>



MERCI !

Contact /

Coordination :

Thomas Lebarbé

thomaslebarbe@me.com

Fatiha Idmhand

fatiha.idmhand@yahoo.es

Laurence Rageot

laurence.rageot@univ-tours.fr