



#dhnord2019

CORPUS ET ARCHIVES NUMERIQUES

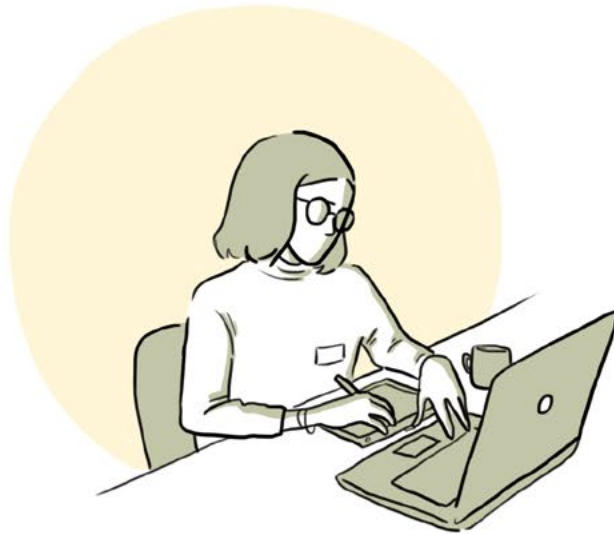
16-18 octobre
2019
Lille



LES DESSINS DU COLLOQUE

Keynote

Ces dessins ont été réalisés par
Adèle Huguet dans le cadre du colloque DHNord
organisé par la MESHS en octobre 2019.
Retrouvez les captations vidéo des conférences sur
<https://publi.meshs.fr/>
ou en cliquant sur les titres des illustrations.



«PAR MA FOI ! IL Y A PLUS
DE 15 ANS QUE JE FAIS [DU
NUMERIQUE] SANS QUE JE
N'EN SUSSE RIEN»

CORPUS, ARCHIVES
ET FORMATION DES
ETUDIANTS EN SHS

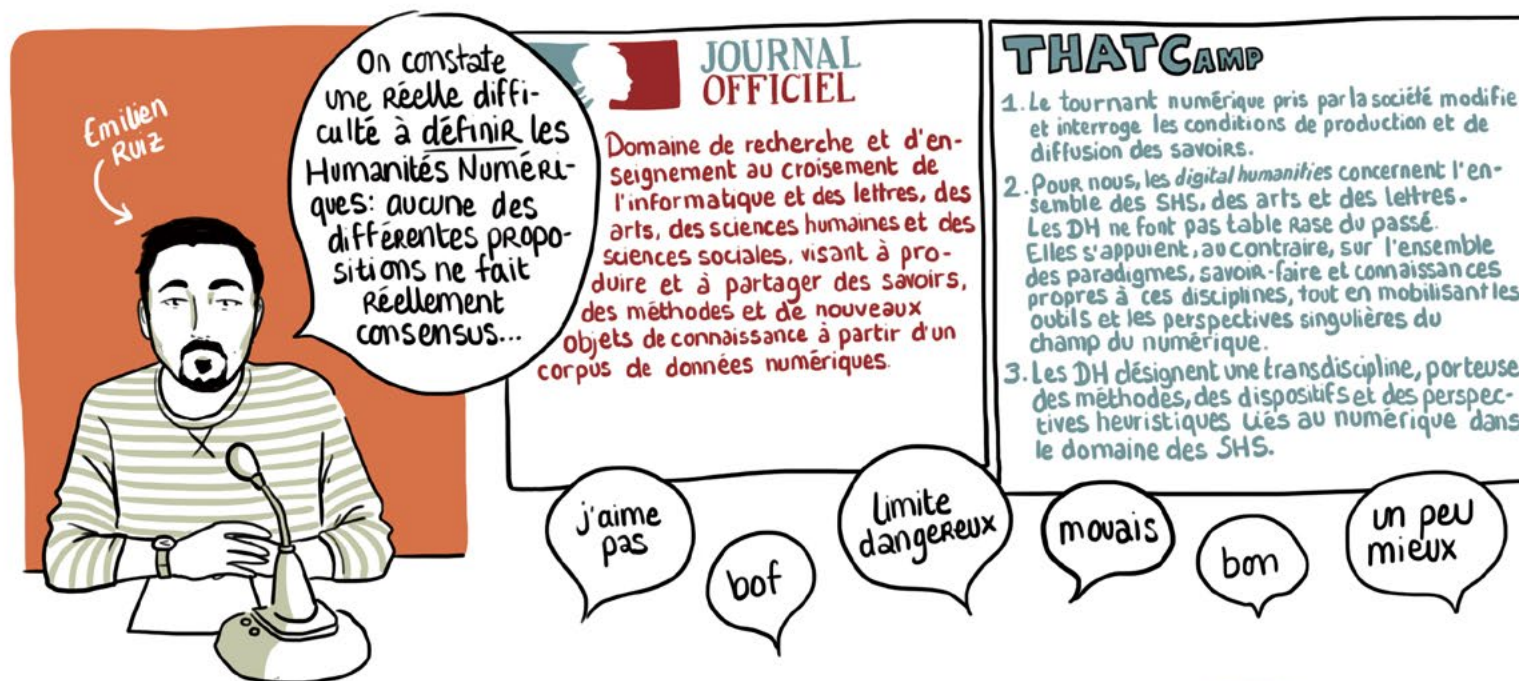
Par Emilien RUIZ,
historien,
assistant professor au
Centre d'histoire de
Sciences Po



Souvent, les SHS se sentent étrangères au numérique.
En témoignent certaines réactions au sein des réseaux de chercheurs...



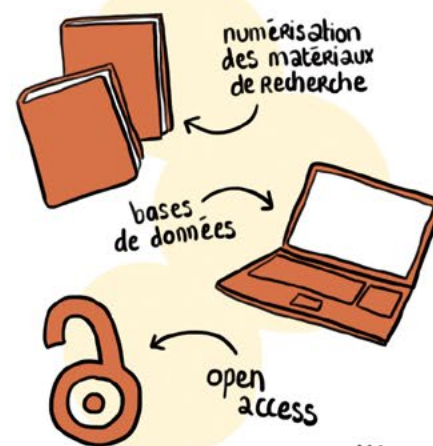
Pourtant, les pratiques numériques offrent de belles opportunités pour l'enseignement, à condition que les chercheurs s'en emparent.



Finalement, en un sens, "les Humanités Numériques n'existent pas"...

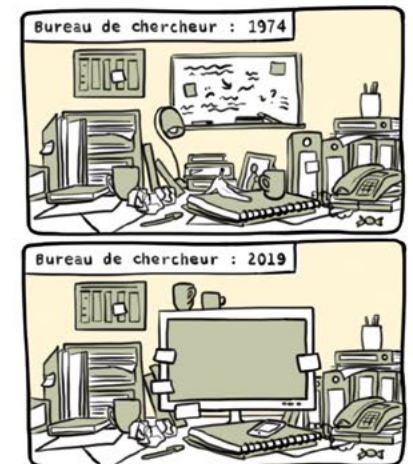


... mais "elles sont ce que nous en faisons quand on se réunit pour discuter"



C'est un éventail très large, ce qui ne facilite pas leur définition, mais offre une grande force opérationnelle et ouvre les possibles en sciences humaines.

Les masters se revendiquant des Humanités Numériques sont plus souvent issus des sciences de l'information et de la communication que des lettres ou de l'histoire, et encore moins de la sociologie, par exemple.



Les DH ne doivent plus être une spécialité. D'ailleurs, dans les faits, tous les chercheurs travaillent déjà avec un ordinateur!

Adapter les méthodes pédagogiques aux nouvelles pratiques, se tenir au courant des derniers outils, continuer à se former... Finalement faire une vraie place aux DH dans l'enseignement des SHS n'est possible que si les équipes enseignantes s'impliquent et ménagent des temps et des espaces consacrés à échanger sérieusement sur ces sujets.

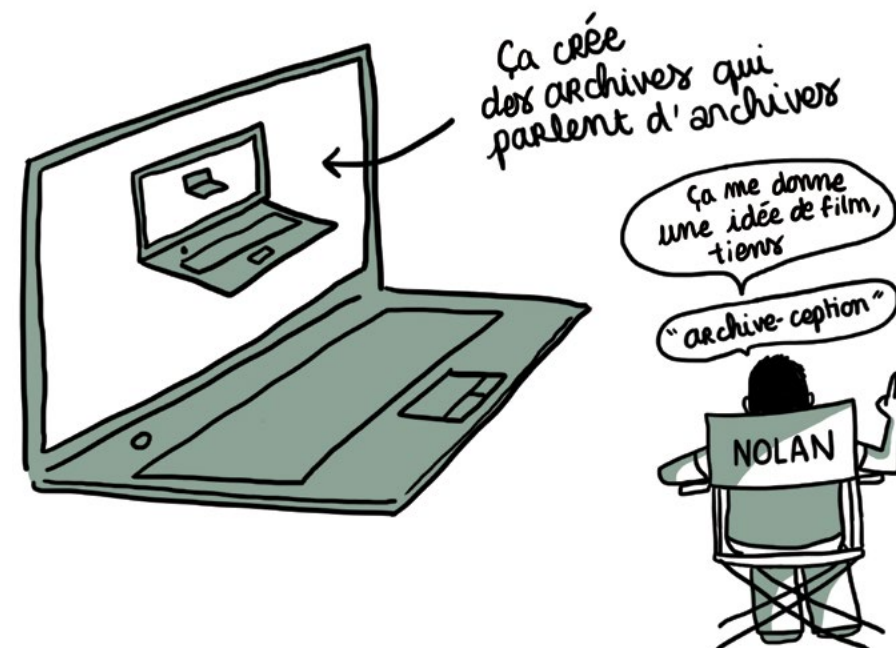
Jour 1

Jeudi 17 Octobre



LE GOÛT DU CORPUS
À L'ÈRE NUMÉRIQUE

Réécrire
Le goût des archives
de façon participative



LE GOÛT DU CORPUS
À L'ÈRE NUMÉRIQUE

Des réticences face
à l'usage du
numérique...



...et pas toujours
par ceux qu'on pense

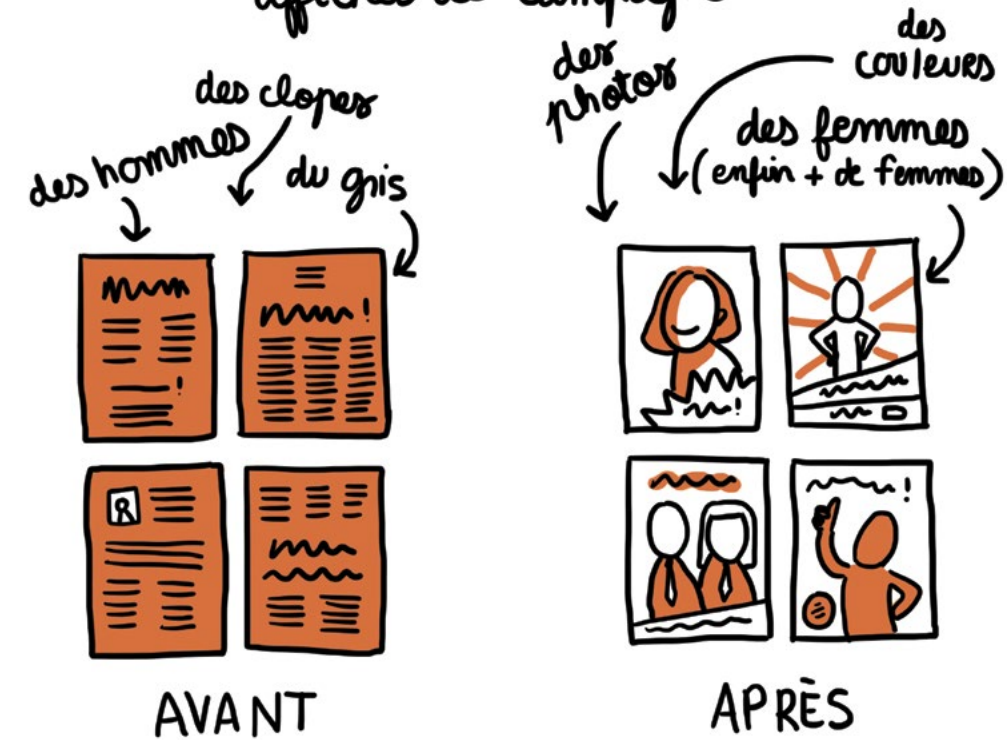
ARCHELEC, LES ARCHIVES
ÉLECTORALES FRANÇAISES
DE LA V^ÈME RÉPUBLIQUE

Du papier au numérique :
reflet fidèle ou distorsion ?



ARCHELEC, LES ARCHIVES
ÉLECTORALES FRANÇAISES
DE LA V^ÈME RÉPUBLIQUE

Le relooking des
affiches de campagne :



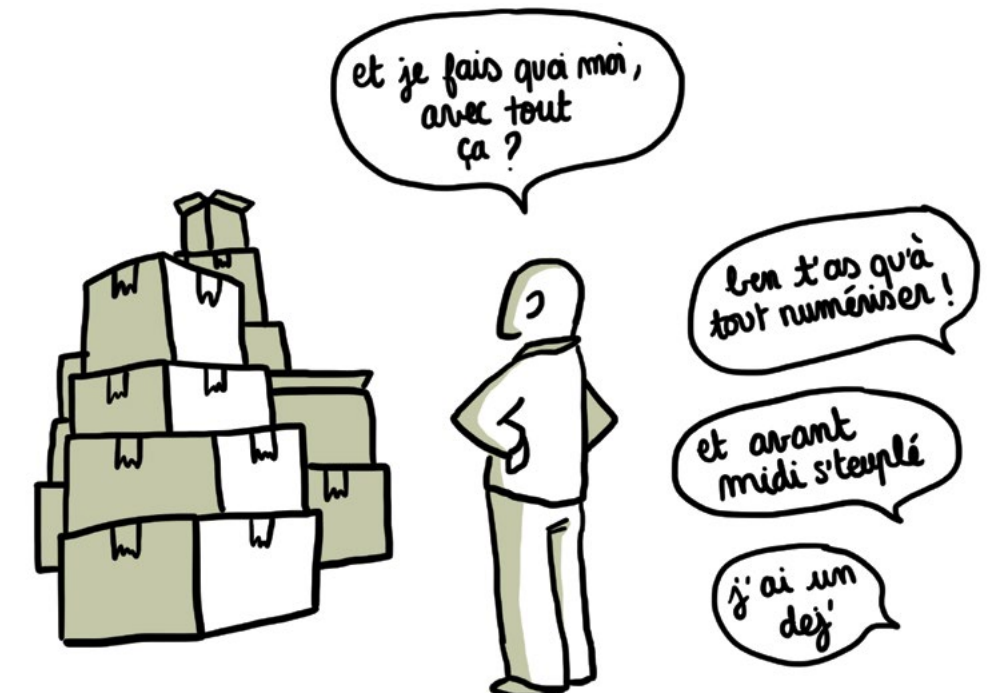
ARCHELEC, LES ARCHIVES
ÉLECTORALES FRANÇAISES
DE LA V^ÈME RÉPUBLIQUE

collection d'historien :
les pin's politiques



DU CORPUS ARCHIVISTIQUE AU CORPUS
NUMÉRIQUE : LES SOUBASSEMENTS
DU WEB SÉMANTIQUE

Seul face aux archives :



DU CORPUS ARCHIVISTIQUE AU CORPUS
NUMERIQUE: LES SOUBASSEMENTS
DU WEB SÉMANTIQUE

le secret:
découper les sources



ARCHIVAGE DU WEB,
UN ENJEU DE GOUVERNANCE
(D'INTERNET)

Un problème:
la nature éphémère
des données



ARCHIVAGE DU WEB,
UN ENJEU DE GOUVERNANCE
(D'INTERNET)

Tensions géopolitiques:
pendant ce temps,
chez Internet
Archive...



L'APPORT DE LA DIPLOMATIQUE
À LA QUALIFICATION
DES TRACES NUMÉRIQUES

Mais où est la seconde
moitié des données du
Grand débat national ?



LES TECHNOLOGIES DU WEB
POUR LA VALORISATION D'UN
PATRIMOINE INDUSTRIEL TEXTILE
EN MOUVEMENT DANS LES
HAUTS-DE-FRANCE

Grand prix de l'acronyme
le plus tiré par les cheveux :



LE CONSORTIUM «CAHIER» :
CORPUS D'AUTEURS POUR LES
HUMANITÉS, INFORMATISATION,
ÉDITION, RECHERCHE

ça c'est un
bel
acronyme

oui bon
ça va,
hein
DENIM

Réunion de travail
au consortium :



LES TECHNOLOGIES DU WEB
POUR LA VALORISATION D'UN
PATRIMOINE INDUSTRIEL TEXTILE
EN MOUVEMENT DANS LES
HAUTS-DE-FRANCE

Entretiens semi-directifs
avec les acteurs du textile



LE CONSORTIUM «CAHIER» :
CORPUS D'AUTEURS POUR LES
HUMANITÉS, INFORMATISATION,
ÉDITION, RECHERCHE

archivistes = amarchistes ?



DES SOURCES NUMÉRIQUES
À LA NUMÉRISATION DES
PRATIQUES HISTORIENNES

Aller aux archives
aujourd'hui :
le safari-photo



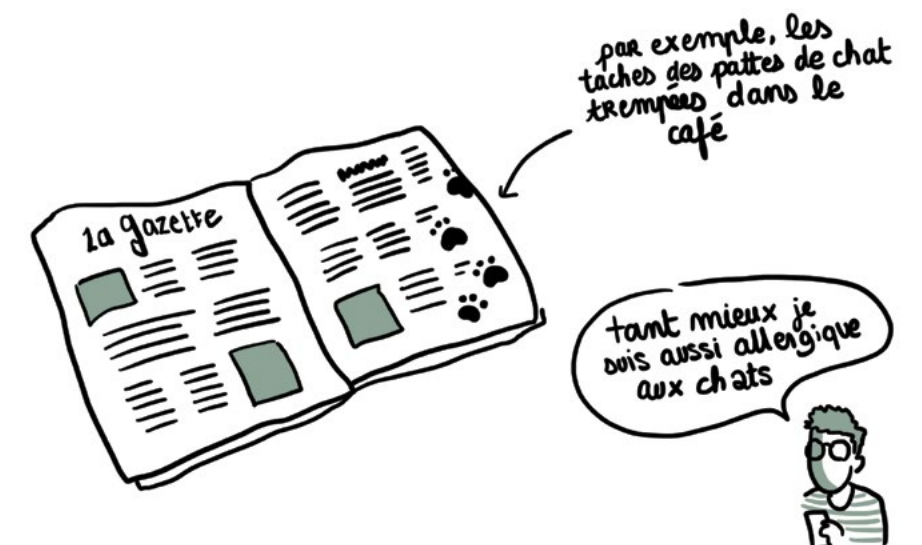
DES SOURCES NUMÉRIQUES
À LA NUMÉRISATION DES
PRATIQUES HISTORIENNES

Qui sont
les jeunes chercheurs ?



DES SOURCES NUMÉRIQUES
À LA NUMÉRISATION DES
PRATIQUES HISTORIENNES

Avec la numérisation, on
perd certains aspects des
archives



L'IMPACT DES PRATIQUES NUMÉRIQUES
SUR LES SOURCES : D'UNE DÉMARCHE IMPLICITE
À DES PROJETS EXPLICITES

Numériser les
archives de l'Internationale
communiste :

Ça, d'accord.

mais ces braves
chercheurs sont-ils
au moins propriétaires
de leurs moyens de
production
informatiques ?



L'IMPACT DES PRATIQUES NUMÉRIQUES
SUR LES SOURCES : D'UNE DÉMARCHE IMPLICITE
À DES PROJETS EXPLICITES

Une plateforme
archivistique, c'est comme
une bouteille de vin : ça
se bonifie avec le temps



surtout avec
les archives de la
MSH de Dijon !

hahaha

la bonne
blague

(astuce : le vin est aussi très bienvenu
pour négocier avec les archivistes russes)

ADJEHI

Keynote

TRAVERSER LES CORPUS DE PRESSE NUMERIQUES

UN TRAVAIL D'ARTISAN ?

Par Juliette DE MAEYER,
professeure agrégée en
communication à l'Université
de Montréal



La numérisation des corpus de presse ouvre de nouveaux horizons pour les études des médias.



Depuis, les archives ont été numérisées et on dénombre en 2010 plus de 500 références!

En 1998, on comptait 58 références au quotidien canadien Globe and Mail dans les thèses canadiennes



Les algorithmes d'OCR permettent de convertir l'image d'une page de journal en une suite de caractères modulables et sélectionnables.

* Reconnaissance Optique des Caractères

Cela permet par exemple d'étudier la viralité dans la presse: comment les textes circulent-ils entre les différents journaux?



Ce type d'étude aurait été extrêmement laborieux, voire impossible, sans la lecture algorithmique.

Toutefois, ces logiciels font parfois (souvent ?) des erreurs, et il arrive que les chercheurs aient de mauvaises surprises...



... Un an plus tard, j'obtenais enfin les PDFs des pages de journaux qu'ils contenaient, stockés dans des DVDs...



Malheureusement, à cause des bavures d'encre et de la mise en page en colonnes, impossible de convertir par OCR... J'y avais mis tout mon argent, et le corpus était inutilisable



Le travail de l'historien se rapproche alors de celui du philologue: comment obtenir un texte à partir d'un objet ?

Franklin Ford était un journaliste américain qui, au XIXe s., a théorisé sur l'avenir des médias, et en particulier sur le rôle des nouvelles technologies.



En étudiant ses lettres et articles, Juliette de Maeyer et ses collègues y ont décelé un potentiel original...

Grâce à la numérisation de ces documents, ils ont donc créé un bot: un compte Twitter au nom de Ford, qui publie des extraits de texte de manière aléatoire.

cela permet d'attirer l'attention sur des choses que l'on n'avait pas vues au départ



Ce bot a la particularité de conserver les erreurs d'OCR, et produit donc parfois des phrases à la syntaxe incorrecte.

Bon sang!
Serais-tu un robot lui aussi?!

Ces erreurs ne sont pas dénuées de sens: elles sont les traces du processus de remédiation, de l'incorporation et de la modification d'un média par un autre.*

REMÉDIATION

*Bolter & Grusin, *Remédiation*, MIT press, 1999

Elles sont des indices historiographiques qui pourront servir aux historiens de demain à étudier les outils d'aujourd'hui et leurs limites.

On peut les comparer avec le "ETA OIN S HROU" qu'ajoutaient les linotypistes pour signaler qu'une ligne comprenait une erreur, et devenus un outil des historiens des médias

C'est pourquoi les chercheurs du projet Franklinfordbot ont choisi de conserver ces erreurs dans leur nouvel algorithme.

il s'agit de ne pas masquer, et au contraire d'accepter et de mettre en avant ces imperfections, un peu comme le fait l'art japonais du kintsugi, qui consiste à recoller les céramiques brisées avec de la colle dorée

pas comme ça

plutôt comme ça

La numérisation s'apparente à un travail d'artisan, marqué par les défauts et les irrégularités du matériau brut dont il est issu.

En somme, la dématérialisation des médias nous oblige à les considérer dans toute leur matérialité.

Mouais...

Alors votre métaphore je veux bien...

N'empêche que la matérialité de ma table basse reste un peu plus palpable que celle du Télé-7 jeux

Jour 2

Vendredi 18 octobre



QUAND L'HERMÉNEUTIQUE SE FAIT
HERMÉTIQUE: LA PHILOGIE
SANSKRITE FACE AUX CORPS
ÉLECTRONIQUES

Les outils informatiques de retranscription
sont mal adaptés au travail
des philologues

Non mais Thierry
comment tu l'as
conçu ton logiciel ?

Pas moyen de
browser en TEI

Et le tag XML ?

156 Mo pour
un texte ?!

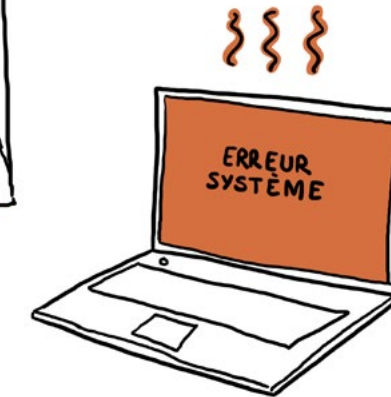
हिंदाहं वलपि
यन्यने वापिक ?!



CONSTITUTION D'UN CORPUS
POUR L'ARCHIVAGE DE LA
MÉMOIRE MINÈRE

Transcription automatique
d'entretiens avec d'anciens
mineurs :

Et là ché louque un carbon qui l'au
d'puis queq's années qu'on était emmolo
par les annellonis par les coursours
ch'est in me rulle pièche ch'ti
pis j'y dis qu'ch't de l' pièche
min l'autre y a mingé de
x'vaque inlogée, j'y dis
quo qu'te lolo ch'to ?! T'in
compris ti ? j'sais faneé dis



DEUX SIÈCLES DE SOURCES DISPARATES
SUR L'INDUSTRIE TEXTILE EN FRANCE :
COMMENT AUTOMATISER LES TRAITEMENTS
D'UN CORPUS NON-UNIFORME ?

Numérisation:
quel logiciel pour des
écrits manuscrits ?

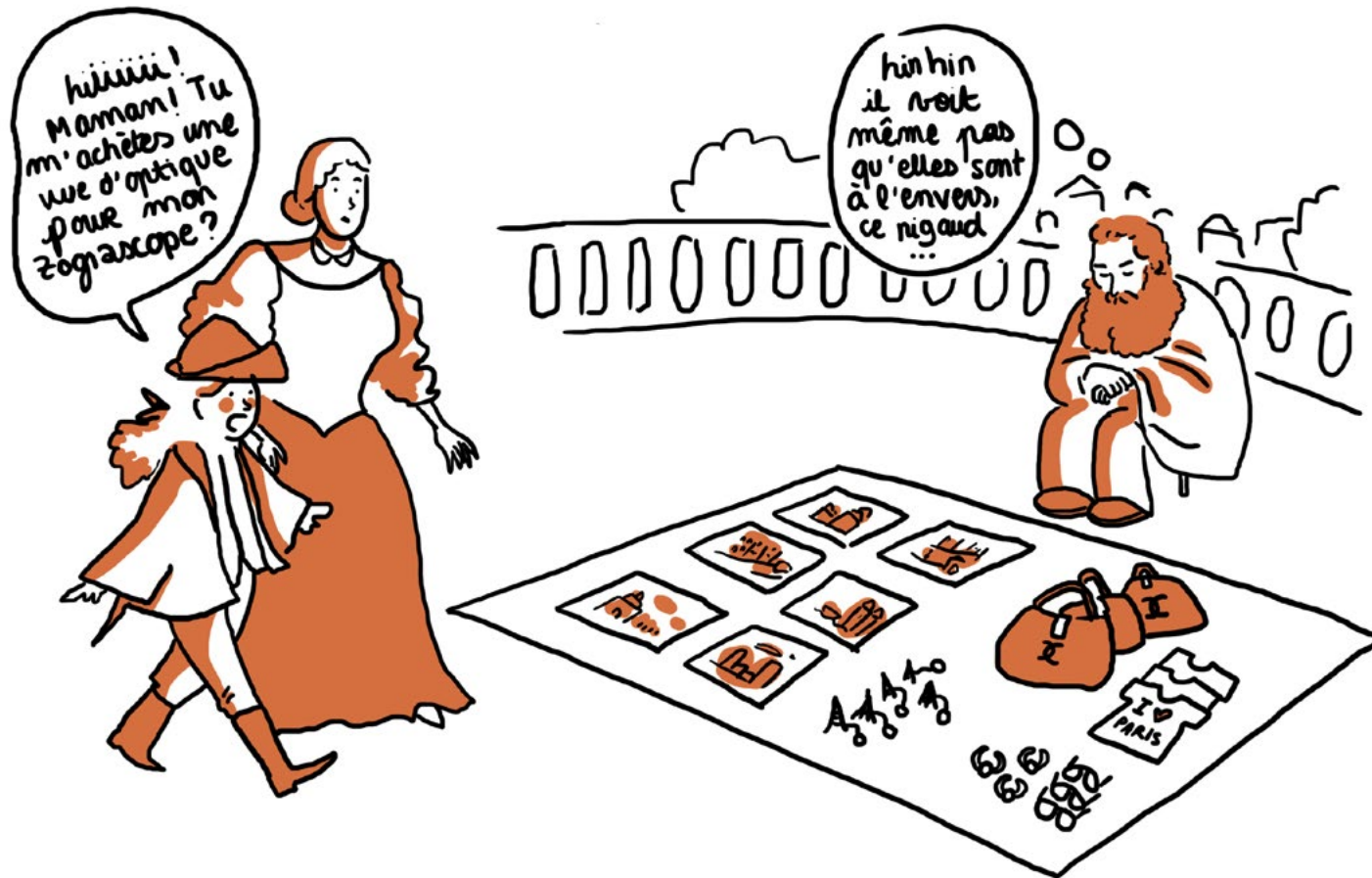


J'ai développé un
logiciel capable de
reconnaître et retranscrire
les cartes postales
écrites par des médecins

qui dit mieux ?

UN OCÉAN D'IMAGES:
ETABLIR UN CATALOGUE
RAISONNÉ D'ESTAMPES
À L'ÈRE DU NUMÉRIQUE

La contrefaçon d'estampes
au XVIII^e siècle :



UN OCÉAN D'IMAGES:
ETABLIR UN CATALOGUE
RAISONNÉ D'ESTAMPES
À L'ÈRE DU NUMÉRIQUE

Routine bien-être des
chercheurs en humanités
numériques :



LE MUSÉE COMME SERVICE
D'INFORMATION. POUR UNE
POLITIQUE DES INTERFACES
MUSÉALES

Tentative d'harmonisation
culturelle européenne :



LE MUSÉE COMME SERVICE
D'INFORMATION. POUR UNE
POLITIQUE DES INTERFACES
MUSÉALES

Difficile appropriation
des interfaces muséales par
le public :

ben...

il est pété...

je swipe et ça
change pas...



POUR UN REGARD À 360°
DES CORPUS VISUELS:
PRATIQUES DE MISE À
DISPOSITION ET DE
RÉUTILISATION

Comment s'amuser
pendant un colloque
universitaire :

je vous propose
d'inventer
des acronymes
tous ensemble !



WOUHOOOU !

YOUPI !

TROP
COOL !

POUR UN REGARD À 360°
DES CORPUS VISUELS:
PRATIQUES DE MISE À
DISPOSITION ET DE
RÉUTILISATION

Avantage du numérique:
le zoom très puissant qu'on
peut faire sur les œuvres



LE SAVIEZ-VOUS ?

EN ZOOMANT VRAIMENT
BEAUCOUP SUR LA JOCONDE,
ON PEUT APERCEVOIR
UNE CHOUETTE HULOTTE
ATTAQUÉE PAR DES
ÉTOURNEAUX

ben sang, voilà
qui remet en
cause toute
ma recherche...



L'ÉDITION NUMÉRIQUE
COLLABORATIVE ET CRITIQUE
DE L'ENCYCLOPÉDIE DE
DIDEROT ET D'ALEMBERT
(ENCRE)

17 volumes numérisés
intégralement

y a pas photo,
c'est clairement
plus pratique



L'ÉDITION NUMÉRIQUE
COLLABORATIVE ET CRITIQUE
DE L'ENCYCLOPÉDIE DE
DIDEROT ET D'ALEMBERT
(ENCRE)

Un outil de recherche
par domaine :



LE PROGRAMME DES REGISTRES
DE LA COMÉDIE FRANÇAISE :
UN CORPUS NUMÉRIQUE EN EXTENSION

15h10:
Quand l'ordinateur plante
au milieu d'une conférence

Ben alors?

ça fait pourtant
partie du jeu,
au théâtre!

L'IMPROVISATION!



PROJET eBALZAC : CONSTRUIRE
UNE BIBLIOTHÈQUE HYPERTEXTUELLE
DESSOURCES INTELLECTUELLES

Quand Balzac découvre
Twitter :

140 caractères ?!

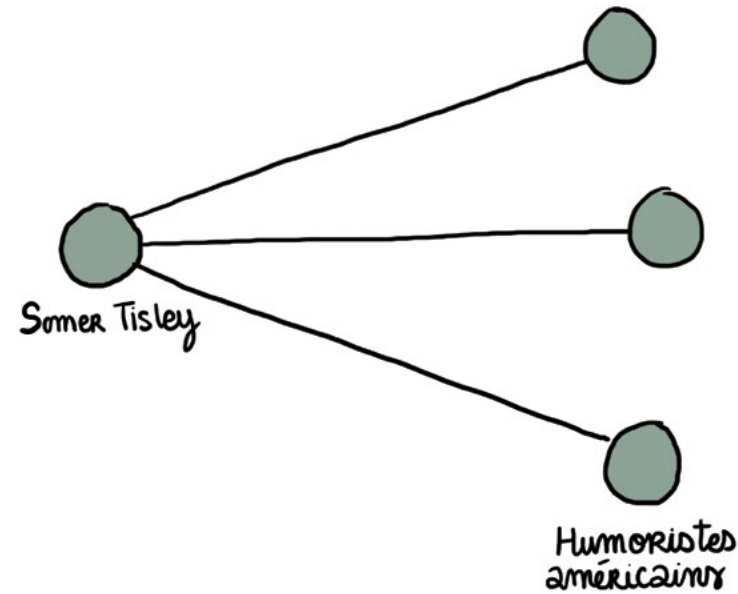
vous vous fichez
de moi ?

vous savez combien
elle en fait la
Comédie Humaine ?



PROJET EBAZAC : CONSTRUIRE
UNE BIBLIOTHÈQUE HYPERTEXTUELLE
DES SOURCES INTELLECTUELLES

Outil de détection des
co-occurrences de mots entre
les textes :



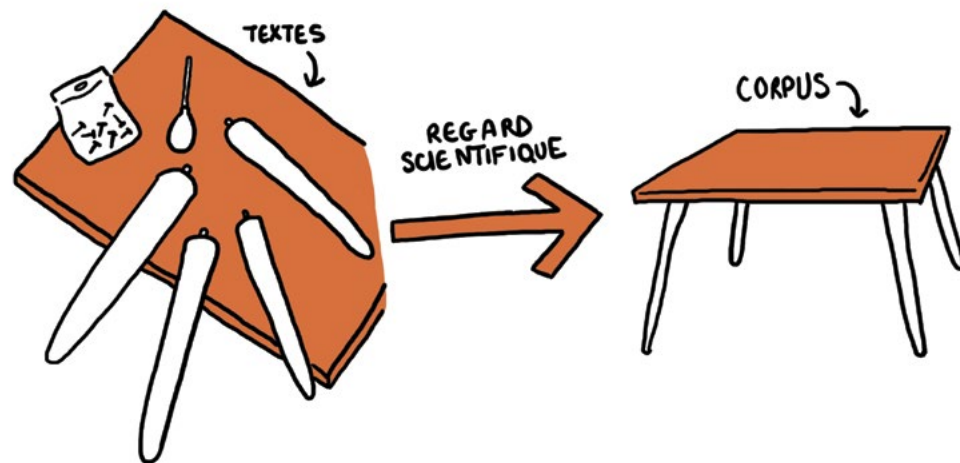
ÉLABORER DES CORPUS NUMÉRIQUES:
LES COLLABORATIONS ENTRE
LABORATOIRES, BIBLIOTHÈQUES
ET CENTRES D'ARCHIVES

Adapter les outils numériques
aux compétences des chercheurs :



ÉLABORER DES CORPUS NUMÉRIQUES:
LES COLLABORATIONS ENTRE
LABORATOIRES, BIBLIOTHÈQUES
ET CENTRES D'ARCHIVES

Comment fabrique-t-on
un corpus numérique ?
Le retour de la métaphore
de la table Ikea



ÉLABORER DES CORPUS NUMÉRIQUES:
LES COLLABORATIONS ENTRE
LABORATOIRES, BIBLIOTHÈQUES
ET CENTRES D'ARCHIVES

La communication entre
bibliothécaires et chercheurs :



CRÉDITS

Illustrations :

Adèle Huguet

<https://adeledessinecom.wordpress.com/adele.huguet@outlook.fr>

Mise en page :

Services HNOMAD et communication MESHS

Inspiration :

Colloque DHNord 2019, organisé par la MESHS
<https://www.meshs.fr/page/dhnord2019>

Le Colloque DHNord 2019 bénéficie du soutien du Conseil Régional des Hauts-de-France et du Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation

