

Outils, infrastructures et traitement des données appliqués aux SHS

INSPÉ Lille HdF

Valentin de Craene¹

05/02/2025

1. Ingénieur d'études, chargé du traitement des données scientifiques, MESHS-CNRS (UAR 3185). @: valentin.de-craene@univ-lille.fr

Positionnement dans le champ de l'appui à la recherche régional de la MESHS (UAR 3185)

- **Unité d'Appui et de Recherche**, sous tutelle du CNRS et des universités des Hauts-de-France, depuis 2008.
- **Périmètre : 44 laboratoires en SHS affiliés** dans les Hauts-de-France (cf https://www.meshs.fr/page/laboratoires_affilies).
Spectre disciplinaire très large (objets et méthodes).
- **Actions :**
 - ➊ **Financement** de projets sur AAP (« émergents », « structurants », « inter-MSH », « soutien scientifique ponctuel », « *rewriting* » et « mobilités internationales »).
 - ➋ Soutien au **montage de projet** (pôle DEGESCI).
 - ➌ **Valorisation** et **communication** scientifique.
 - ➍ **Ingénierie de la donnée en SHS** et formation aux outils numériques (pôle HNO MAD dont service HN).

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

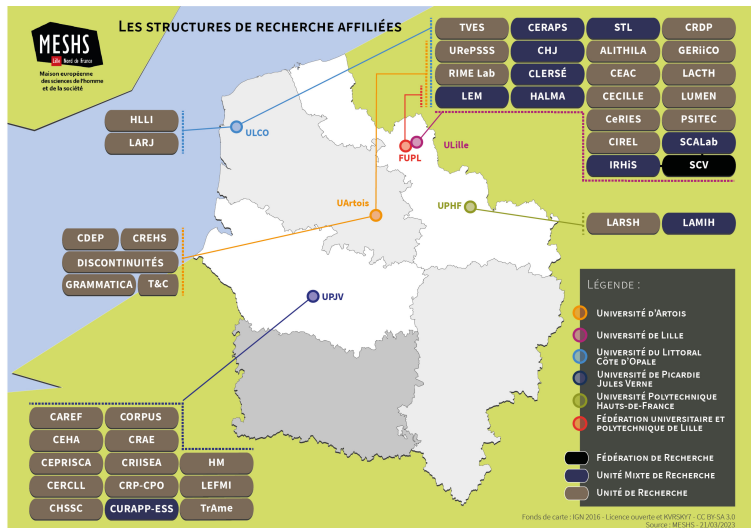


Figure – Cartographie des unités affiliées à la MESHS (2024)

Le service Humanités Numérique : expertises et structuration

- Action transversale Humanités Numériques au sein du **pôle Humanités Numériques Outils Méthodes et Analyse de données (HNOMAD)**.
- **Référent scientifique** : Philippe Useille (MCF en SIC au LARSH - UPHF).
- **Ingénierie** : Valentin de Cræne (IGE chargé du traitement des données scientifique - CNRS).



Figure – Logo du service « Humanités Numériques »

Positionnement dans le champ des « humanités numériques »

- 1 **Soutien ponctuel** : accompagnement personnalisé des chercheurs sur sollicitation.
- 2 **Relais régional de l'IR* Huma-Num** : formation aux outils et services numériques d'Huma-Num.
- 3 **Actions de formation en humanités numériques** : cycle de formation 2024 et 2025.
- 4 **Traitement des données au sein des projets** : formation et montée en compétences collective.
- 5 **Organisation du colloque DHNord** : formation par les pairs (liens avec pictorlA).

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

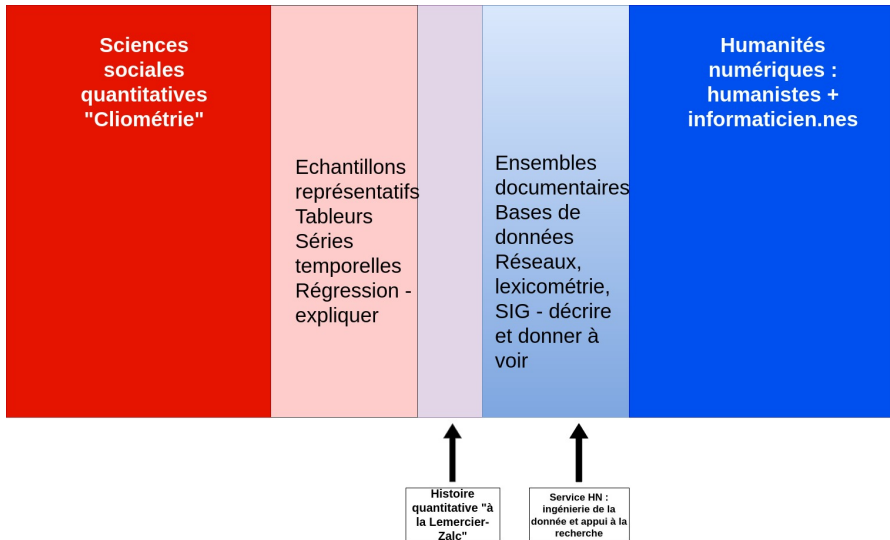


Figure – « Le spectre des humanités numériques » (dans les sciences historiques), d'après C. Lemerrier et Zalc

Cycle de vie des données et outils : une approche par la pratique

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

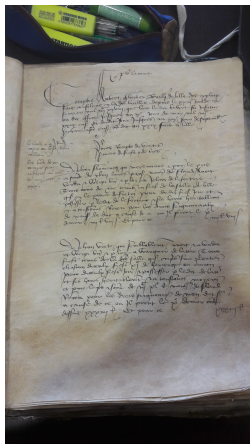


Figure – Registre du bailliage de Lille (1429-1441, ADN, B 6234)

Collecte et gestion des données de recherche

- Numérisation et **transcription** des sources : avec e-Scriptorium (<https://msia.escriptorium.fr/>)



Figure – Logo d'e-Scriptorium

- Nettoyage et **normalisation** des données : langage de programmation Python - *package Pandas*



Figure – Logo de Python

Structuration, traitement et analyse des données de recherche

- **Base de données relationnelle** : import sous Heurist des fichiers CSV (*Comma Separated Values = tableur*)



Figure – Logo d'Heurist

- **Visualisation des relations entre justiciables** : Gephi et GephiLite



Figure – Logo de Gephi

Gestion bibliographique et veille scientifique

- **Gestion des données bibliographiques** : Zotero et connecteurs web



Figure – Logo de Zotero

- **Prise de note sémantique** : écriture en Markdown et éditeur (ex : Obsidian)



Ouverture, pérennisation et réutilisation des données

- Dépôt des publications : HAL



Figure – Logo de HAL

- Dépôt et mise à disposition des données : entrepôt de données Nakala



Figure – Logo du Nakala

Quelques exemples de projets accompagnés par le service HN



Accueil Découvrir le projet Recherche avancée Index Le dictionnaire en graphe (à venir) Recherche rapide Rechercher

FERME Gé

Accueil Découvrir le projet Recherche avancée Index Le dictionnaire en graphe (à venir) Recherche rapide Rechercher

Dictionnaire Numérique de la Ferme générale

← Liste des notices publiées

Pourquoi *réduire* la Ferme générale en un dictionnaire ? Pour rassembler ce qui est éparé, comprendre une organisation complexe dont le champ d'action est à la fois fiscal, financier, économique, commercial, politique, juridique, administratif, colonial et international, pour saisir les effets d'une telle structure sur un territoire lui-même bigarré par des privilèges de toute nature. L'approche lexicale répond au défi auquel les historiens se trouvaient confrontés : saisir cet objet d'histoire dans sa globalité. Ne serait-ce que la fonction première de cette organisation, la fiscalisation des denrées et des marchandises, nécessitait une mise en répertoire. Tout au plus avait-on exploré les mécanismes des gabelles, du monopole de la vente du tabac, de quelques douanes, mais la fiscalité du vin, des eaux-de-vie, de la viande, des poissons, des cartes à jouer, des cuirs, des huiles, des métaux, du papier, ou encore des produits coloniaux comme le café ou le sucre, demeurait presque entièrement méconnue, tout comme son évolution.

Figure – Projet ANR FermeGé : page d'accueil du dictionnaire numérique (<https://fermege.meshs.fr/>)

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

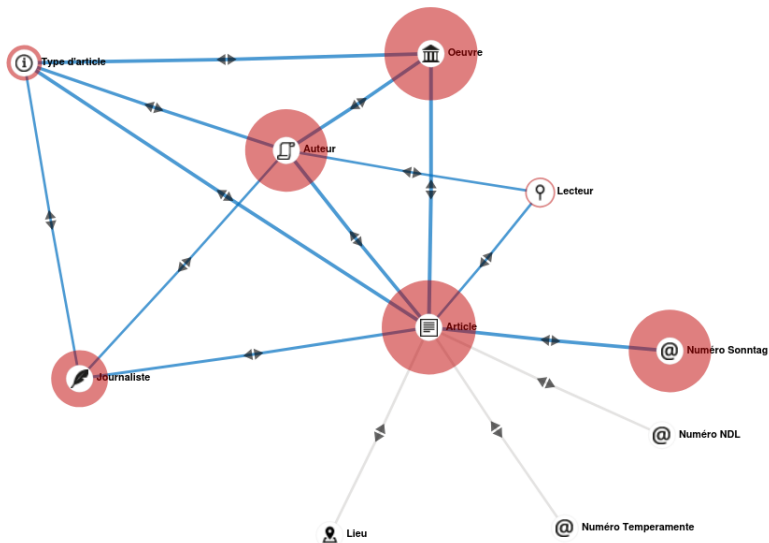


Figure – Projet émergent PERARTEM : une base de données Heurist pour la revue littéraire Sonntag (année 1977) - modèle relationnel

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

Items

In tem sit. ImaginIRL

Title ▾ Ascending ▾ SORT

showing results



[Conférence Sinn Féiner, Mr Michael Collins, secrétaire de la Conférence et porte-parole de Mr Farnon de Valera]. [photographie de presse] / [Agence Mourate]
Image de Michael Collins à la conférence de Londres



[De p. à d.] Cathal O'Shannon, W. J. (Le Thomas) Johnson parlementaires irlandais le 12 décembre 1922. à Dublin]. [photographie de presse] / [Agence Bol]
Image des parlementaires irlandais Cathal O'Shannon et Thomas Johnson. Troisième homme non identifié.



[De p. à d.] John Bagwell et Thomas Grattan [Le Grattan] Emond parlementaires irlandais à Dublin le 12 décembre 1922]. [photographie de presse] / [Agence Bol]
Image des parlementaires irlandais à Dublin. De gauche à droite, John Bagwell et Thomas Grattan



[De gauche à droite] M. (Robert) Barton, M. de Valera, le comte Plunkett, M. Arthur Griffith, M. Austin Stack. [2 juillet 1921, Holyhead, Parc de Galles]. [photographie de presse] / [Agence Bol]
De gauche à droite, M. Robert Barton, M. de Valera, le comte Plunkett, M. Arthur Griffith, M. Austin Stack à Holyhead



[Dublin ? soldats britanniques]. [photographie de presse] / [Agence Bol]
Négatif d'une photographie parue dans "La Dépêche (Toulouse, 1870-1944)" des soldats britanniques à Dublin.



[Portrait de M. Arthur Griffith.]



[La notice britannique à la recherche de M. (Farnon) De Valera à bord du Pionier à Dublin]. [photographie de presse] / [Agence Bol]



[Dublin, devant la] fusillade devant la maison des douanes (25 mai 1921). [photographie de presse] / [Agence Bol]
Image de la fusillade devant la maison des douanes

Figure – Projet émergent ImaginIRL : explorer la réception de l'indépendance irlandaise (1920) dans la presse francophone - galerie Omeka S

Au niveau national : l'IR* Huma-Num

Comité de pilotage

Conseil Scientifique

Équipe de direction



Olivier Baude
DR Huma-Num IR*



Ariane Allet
IE - Secrétaire générale
Responsable du pôle PAC



Nicolas Larrousse
IR - Directeur adjoint
Responsable du pôle CONIC



Stéphane Pouyllau
IR - Directeur adjoint
Responsable du HN Lab



Hélène Jouguet
IE - Responsable du pôle
DONNEES

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

PAC

Pôle Pilotage, administration, communication

Responsable :
Ariane Allet



Alain Belvindrah
AI - Gestionnaire financier



Pauline Busonera
IE - CDD
Chargée de communication

ACCES

Pôle Architecture, Conception, Connaissances, Enrichissement, Service



Adrien Desseigne
IE - Concepteur et développeur d'applications web



Bruno Vanderaert
IE
Ingénierie logicielle



Corentin Fragnaud
IE CDD
Ingénierie logicielle

CONIC

Pôle Coordination Nationale et Internationale des Communautés



Edward Gray
CR CDD - Ingénieur coordination Europe ERIC DARIAH / Huma-Num



Adeline Joffres
IR - Responsable de la coopération internationale

HSTC

Pôle Hébergement, Stockage, Traitement, Calcul



Gérald Foliot
IR - Expert en infrastructures



Thomas Audo
IE - Administrateur de l'infrastructure numérique



Julie Erismann
IE - Administratrice systèmes et réseaux



Elina Gusing
IE CDD
Administratrice de bases de données et d'application

DONNEES

Pôle DONNEES et coordination de l'accompagnement utilisateur



Michel Jacobson
IR - Expert en archivage numérique
Co-responsable du pôle



Claire Carpentier
AI - Chargée du traitement des données scientifiques



Sara Tandar
IE - Chargée du traitement des données scientifiques



Huma-Num Lab

Responsable :
Stéphane Pouyllau



Léa Maronet
Doctorante



Adam Faci
Post-doctorant



Antoine Silvestre De Sacy
IE CDD Chargé du traitement des données scientifiques

Moyens informatiques

Gérald Foliot
IR - Expert en infrastructures

Thomas Audo
IE - Administrateur de l'infrastructure numérique

Novembre 2023



- **Huma-Num** : fusion du Très Grand Equipement Adonis et de l'Infrastructure de Recherche Corpus en 2013 -> ex-TGIR, labellisée IR* (Infrastructure de Recherche Etoile) en 2022
- **Missions** :
 - 1 Mise en oeuvre d'une **infrastructure numérique**
 - 2 développer une **culture de la donnée** en SHS
 - 3 proposer une **offre de services** pour les projets de recherche par le biais des **relais Huma-Num** (dans les MSH)

Accès centralisé aux services d'Huma-Num via HumanID

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

The screenshot shows the HumanID portal interface. At the top, there is a navigation bar with the HumanID logo and four links: "Mes services", "Mes identifiants", "Mon mot de passe", and "Mes informations". Below this, the main heading is "Paramètres du compte personnel" with a sub-link "Gérer mes services Huma-Num". The main content area displays a grid of service cards, each with a logo, a brief description, and an "accéder" button. The services shown are:

- isidore**: Votre assistant de recherche en Sciences Humaines et Sociales. Button: accéder.
- ShareDocs**: Plateforme de stockage et de partage de fichiers (Web et clients WebDAV). Button: accéder.
- nakala**: Partager, publier et valoriser vos données scientifiques. Button: accéder.
- __Stylo__**: Un éditeur de texte simplifiant la rédaction et l'édition d'articles scientifiques en SHS. Button: accéder.
- GitLab**: Plateforme de forge basé sur git. Button: Demander l'accès.
- Mattermost**: Service de discussion d'équipes. Button: Demander l'accès.

Figure – Portail d'accès et de demande d'accès aux services d'Huma-Num

Focus sur ShareDocs

- Outil de stockage collaboratif pour données hétérogènes d'un projet
- *Versionning* des fichiers
- Gestion fine des droits d'accès (partage par utilisateurs, groupes, liens web,...) et compte projet
- Outils intégrés d'OCR, *SpeechToText* et transcodage video

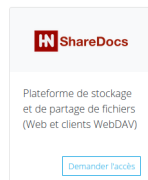


Figure – Onglet d'accès à ShareDocs

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

The screenshot shows the ShareDocs interface. At the top, there is a '+ Nouveau' button and a search icon. Below this is a navigation bar with 'Mes fichiers' and several icons. The main area is divided into a left sidebar and a main table.

Left Sidebar (Navigation):

- Mes fichiers (selected)
- ANR_FermeGé
- Documents
- Images_MESHS
- Imagin_IRL_éme...
- MemoireM2TNAH
- MESHS
- ressources_form...
- Stage MESHS D...
- hnTools_watchFol...
- Marie-Sophie Winter
- Bibliothèques
- Favoris
- Mes documents parta...
- Liens partagés
- Fichiers supprimés

Main Table (File List):

<input checked="" type="checkbox"/>	Nom	Type	Icons	Date de ...	Image prop...
<input type="checkbox"/>	ANR_FermeGé			Déc. 6, 2022	
<input type="checkbox"/>	Documents			Nov. 11	
<input type="checkbox"/>	Images_MESHS			Avril 12	
<input checked="" type="checkbox"/>	Imagin_IRL_émergent_MESHS_2024_2025			Nov. 26	
<input type="checkbox"/>	MemoireM2TNAH			Déc. 6, 2022	
<input type="checkbox"/>	MESHS			Avril 15	
<input checked="" type="checkbox"/>	ressources_formations_donnees			Avril 29	
<input type="checkbox"/>	Stage MESHS DE CRAENE Valentin			Juill. 27, 20...	
<input checked="" type="checkbox"/>	grediPlus_recommandations_images	md		Avril 12	

Focus sur Nakala

- Entrepôt sécurisé de données numériques labellisé par le ministère destiné à accueillir les données de la recherche en SHS
- Accessibilité directe aux données et métadonnées
- Citabilité garantie avec génération des DOI



Figure – Logo de l'entrepôt de données Nakala

Outils, infrastructures et traitement des données appliqués aux SHS

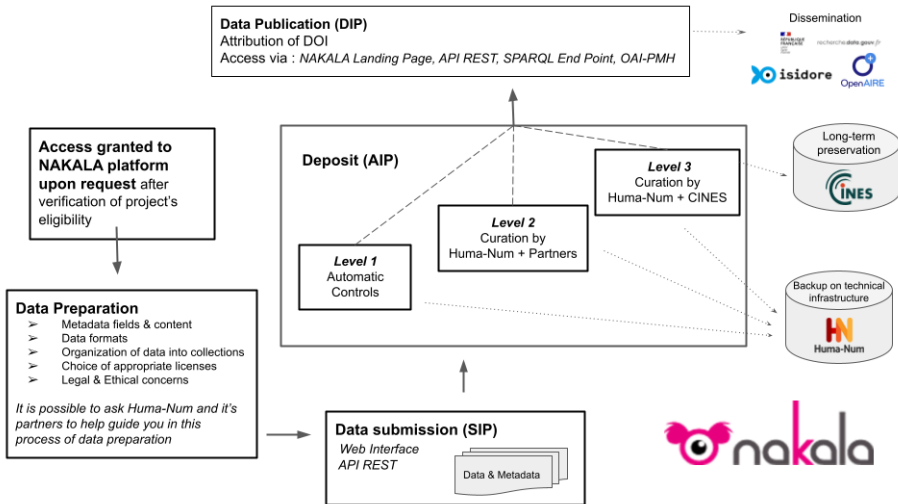
Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?



Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

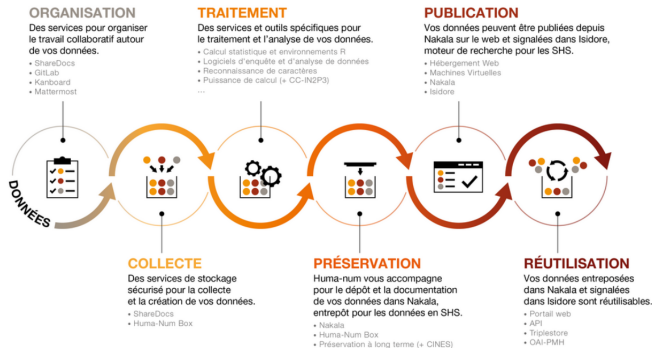


Figure – Répartition des services d'Huma-Num suivant le cycle de vie de la donnée

Les consortiums Huma-Num

- Regroupement et mise en relation de différents unités et équipes de recherche autour de thématiques labellisées ou d'objet commun d'études
- Budget mis à disposition des consortium pour production d'études (guide de bonnes pratiques, guides juridiques,...) ou de services

Les consortiums-HN labellisés



Pictoria



AR:IA:NE



SOL



Time Machine
PROJECTS



MASA^{plus}



distam



canevas



musica²



CONSORTIUM 3DHIN



CORLI

Conclusion : l'IA
et la recherche,
vers un nouveau
paradigme ?

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

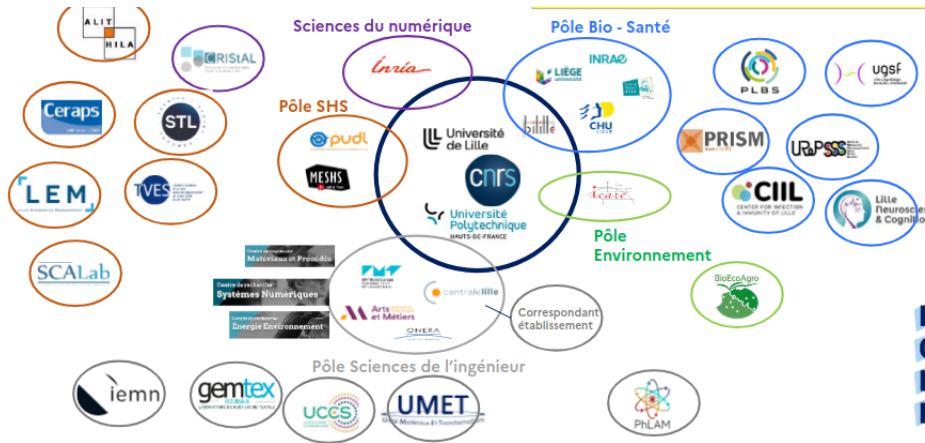
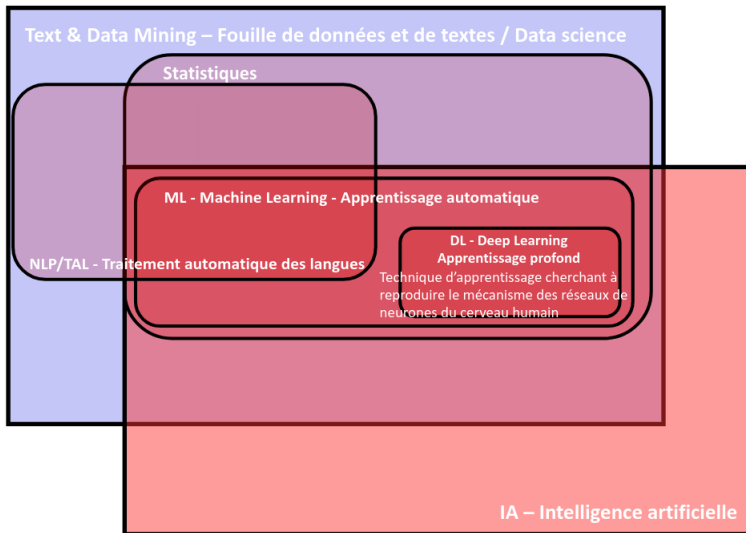


Figure – Cartographie des acteurs de l'atelier de la donnée LORD (Lille Open Research Data)

Historicité de l'intelligence artificielle

- **1950-60** : contexte de la **cybernétique** (sciences cognitives mais difficultés de passer à l'échelle comme en traduction)
- **1970-80** : « hiver de l'IA » car manque de financements
- **1980** : renouveau avec engouement autour des **systèmes experts** adossés à des bases de données
- **1980-2010** : 2e « hiver de l'IA »
- **2010** : révolution de l'**apprentissage artificiel profond**, apport des GAFAM (corpus de données type BERT)
 - 2018 : prix Turing remis à Yann LeCun pour ses travaux sur le *deep-learning*
- **2022-...** : émergence de l'intelligence artificielle générative

Comprendre l'IA et ses usages



Outils,
infrastructures et
traitement des
données
appliqués aux
SHS

Valentin de
Cræne

La MESHS et
son service HN :
brève
introduction

Cycle de vie des
données et
outils : une
approche par la
pratique

Infrastructures
de recherche et
d'appui en SHS

Conclusion : l'IA
et la recherche,
vers un nouveau
paradigme ?

Outils, infrastructures et traitement des données appliqués aux SHS

Valentin de Craene

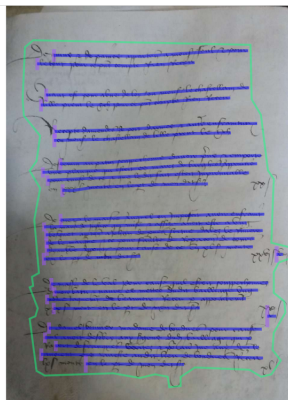
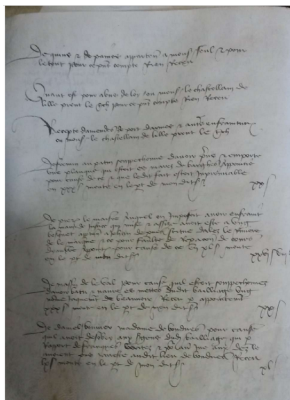
La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?

Intégration du deep-learning dans les projets de recherche : l'exemple de l'Handwritten Text Recognition



- 1 De quins et de panies appartenant a monseigneur seul et pour
- 2 le tout pour ce present compte rien receu
- 3 Quant est pour abus de loy ou monseigneur le chastelain de
- 4 Lille prend le tierche pour ce present compte rien receu
- 5 Recepte d'amendes de port d'armes et autres enfreintures
- 6 ou monseigneur le chastelain de Lille prend le tierche
- 7 De Fremin au Patit souppechonné d'avoir pris et emporté
- 8 une planque qui estoit es mettes de Barghes appointé
- 9 pour cause de ce et que le dit fait estoit improuvable
- 10 en XXX sous monté en part de mon seigneur XX sous
- 11 De Pierre le Maistre auquel on imposito avoir enfrainct
- 12 la main de justice qui mise et assize avoit estéa ung
- 13 bosquet appartenant a Jehan de Pernes situé l'alez le riviere
- 14 de la marque et ce pour faulte de reparation de cours
- 15 Appointé pour cause de ce en XL sous monté
- 16 en le part de mon dit seigneur
- 17 De Massin de la Val pour cause qu'il estoit souppechonné
- 18 d'avoir batu et marié es mettes dudit bailliage ung
- 19 nommé Jacquemet de Beauovne receu par appointment
- 20 XXX sous monté en le part de mon dit seigneur
- 21
- 22
- 23 De Daniel Bennier madame de Bondues pour cause
- 24 qu'il avoit desobey aus sergens dudit bailliage qui par
- 25 rapport de franques veritez et pour l'an jue aux dez le
- 26 avois pris et arresté au dit lieu de Bondues receu
- 27 LX sous monté en le part de mon dit seigneur

Figure – Interface d'E-Scriptorium, un outil open source d'HTR

Outils, infrastructures et traitement des données appliqués aux SHS

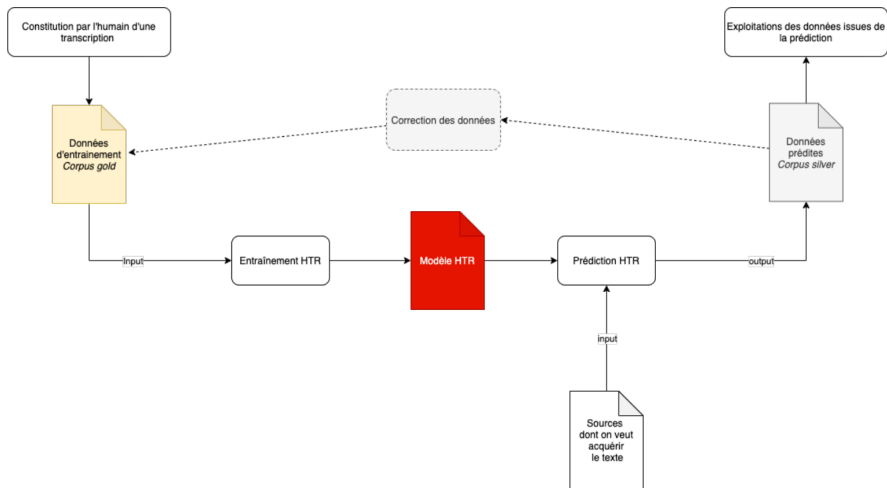
Valentin de Craene

La MESHS et son service HN : brève introduction

Cycle de vie des données et outils : une approche par la pratique

Infrastructures de recherche et d'appui en SHS

Conclusion : l'IA et la recherche, vers un nouveau paradigme ?



Intégration de l'IA dans les infrastructures et consortiums de recherche



Figure – Logo du consortium Huma-Num Pictoria



Figure – Logo du consortium Huma-Num ArlAne

Comprendre et comparer l'IA générative : le projet ComparIA

RÉPUBLIQUE FRANÇAISE
L'avenir s'écrit ensemble

compar:IA BETA
Le comparateur d'IA conversationnelles

✎ Nous aider à améliorer compar:IA

Accueil Liste des modèles d'IA À propos Partenaires FAQ Événement 7 février

Inscrivez-vous à la journée compar:IA qui aura lieu à la BNF le 7 février 2025 — en savoir plus

Ne vous fiez pas aux réponses d'une seule IA

Discutez avec deux IA à l'aveugle pour croiser leurs réponses

J'accepte les modalités d'utilisation [↗](#)
Les données sont partagées à des fins de recherche

[Commencer à discuter](#)

- 1 Je discute avec deux IA anonymes**
Échangez aussi longtemps que vous le souhaitez
- 2 Je donne mon avis**
Vous contribuez ainsi à l'amélioration des modèles d'IA
- 3 Les modèles sont démasqués !**
Apprenez en plus sur les modèles d'IA et leurs caractéristiques

Figure – Page d'accueil du site web : <https://www.comparia.beta.gouv.fr/>

Contacts

E-mail : valentin.de-craene@univ-lille.fr

Pôle HNOMAD (humanités numériques, PUDL, indicateurs scientifiques) :
hnomad@meshs.fr

Des questions ou besoin d'orientation ?

Suivez librement nos **formations en Humanités Numériques** et nos
permanences HN chaque mois

Outils,
infrastructures et
traitement des
données
appliqués aux
SHS

Valentin de
Cræne

La MESHS et
son service HN :
brève
introduction

Cycle de vie des
données et
outils : une
approche par la
pratique

Infrastructures
de recherche et
d'appui en SHS

Conclusion : l'IA
et la recherche,
vers un nouveau
paradigme ?

CYCLE DE FORMATION HUMANITÉS NUMÉRIQUES 2025



PROGRAMME

- #1 17/01 : Introduction aux humanités numériques : outils, méthodes et infrastructures de recherche en SHS
- #2 21/02 : Markdown et Obsidian : introduction au "personal knowledge management" en SHS
- #3 21/03 : Zotero : gestion des données bibliographiques pour la recherche en SHS
- #4 18/04 : Open Refine: traitement et normalisation des données sérielles
- #5 16/05 : Introduction aux bases de données relationnelles avec Heurist
- #6 20/06 : La Science Ouverte par la pratique : HAL et les entrepôts de données

Formations mensuelles le vendredi de 10h à 12h, en format hybride, organisées et dispensées le service Humanités numériques (MESHS-CNRS)

Contact : valentin.de-craene@univ-lille.fr

Publics concernés
Etudiants en Master 2, doctorants,
personnels ITA, chercheurs,
enseignants-chercheurs
des laboratoires affiliés



Sur inscription
Merci de vous rendre sur la
page de chaque formation
pour trouver le lien
d'inscription