

CERES, rapport d'activités 2025

Félix Alié¹, Virginie Julliard^{1, 2}, Gaël Lejeune^{1, 3}



Résumé

Le CERES (Centre d'Expérimentation en méthodes numériques pour les Recherches en SHS) est une unité mutualisée de service rattachée à la Faculté des Lettres de Sorbonne Université. Il a pour vocation d'accompagner et de promouvoir l'usage des méthodes numériques dans les recherches en SHS à travers **deux objectifs principaux** :

1. Élaborer des solutions numériques visant à répondre aux besoins exprimés par les collègues pour la constitution, l'exploration, l'analyse et l'annotation de corpus, ainsi que pour la collecte, le traitement et la visualisation de données numériques ;
2. Formaliser et diffuser des méthodes, en encourageant une réflexion sur la place et les enjeux des outils informatiques dans les recherches en SHS.

Afin de remplir ces objectifs, le CERES met en œuvre **cinq types d'actions** :

1. Des ateliers de formation illustrant des usages variés d'outils numériques ;
2. Des journées dédiées à la collecte et à l'analyse de certains types de données (Hackathons) ;
3. Des accompagnements personnalisés pour le développement de méthodologies numériques et d'outils informatiques susceptibles de répondre à des problématiques transversales en SHS ;
4. Des journées d'étude consacrées à des thématiques transversales ;
5. Du conseil, voire des partenariats, pour l'intégration de volets numériques dans des réponses à des appels à projets (Émergence SU et Ville de Paris, ANR, ERC ...).

Pour assurer l'ensemble de ces missions, le CERES réunit une équipe interdisciplinaire composée d'enseignant·e·s, de chercheur·se·s et d'ingénieur·e·s aux compétences complémentaires (informatique, SHS, design).

¹ CERES

² GRIPIC, UR 1498, Sorbonne Université

³ STIH, EA 4509, Sorbonne Université

* **Site web** : <https://ceres-sorbonne.github.io/>

* **Mail** : lettres-ceres@sorbonne-universite.fr, virginie.julliard@sorbonne-universite.fr, gael.lejeune@sorbonne-universite.fr

Table des matières

1 CERES dans l'écosystème SU	1
2 Appui aux besoins numériques de la communauté scientifique de la Faculté des Lettres	2
2.1 Ateliers et Hacka/data-thons	2
2.2 Programme Doctoral	2
2.3 Suivi des doctorant·e·s	3
2.4 Accompagnements des chercheur·se·s	3
2.5 Conseils sur l'intégration d'un volet numérique dans les réponses aux appels à projets	3
2.6 Développements informatiques	4
2.7 Chantiers transversaux stratégiques	4
3 Le projet Semiosis	5

4 Analyse des pratiques numériques	6
4.1 Podcasts	7
4.2 Formations proposées aux personnels SU	7
4.3 Innovation méthodologique dans la FAIRisation des données de la recherche en SHS	7
5 Contrats, partenariats et dissémination	7
6 Diagnostic à 58 mois	8
7 Perspectives	10

1. CERES dans l'écosystème SU

L'unité CERES a été installée en janvier 2021 à la Faculté des Lettres par décision de la présidence de Sorbonne Université et après approbation du Décanat de la Faculté et du Conseil scientifique de l'université.

L'équipe opérationnelle CERES est constituée de : Virginie Julliard (Professeure en sciences de l'information et de la communication), directrice, Gaël Lejeune (Maître de Conférences en Informatique), directeur adjoint, Félix Alié (ingénieur en informatique) et Edouard Bouté (ingénieur en sciences humaines et sociales, jusqu'en septembre 2025). Cette équipe était complétée pour 2024 par des forces vives rémunérées sur fonds propres : un ingénieur infrastructure opérant sous forme de prestations. Le budget ne nous a pas permis de conserver notre ingénieur design à mi-temps et les deux assistants de recherche (étudiants) également à temps partiel. Ceci a fortement affecté l'activité de l'unité. Nous sommes donc passés de 3,25 équivalent temps plein à 1,8 en 2025.

Nous travaillons avec de nombreux chercheurs et chercheuses de SU, mais aussi avec différentes unités de l'alliance SU :

PLATEFORMES CERES compte désormais au nombre des plateformes de recherche de Sorbonne Université. Jusqu'alors, CERES était obligé de passer par le laboratoire STIH pour se faire payer les prestations réalisées (une en analyse de Corpus avec Sorbonne Nouvelle, une en structuration de jeux de données et accompagnement méthodologique avec l'UTC). Le passage au statut de plateforme nous permettra de mieux rationaliser et de mieux valoriser l'essaimage de nos propositions méthodologiques et services technologiques. Cette activité a permis d'abonder le budget CERES de 25.000€ par les prestations externes en 2025 et de 20.000€ par les prestations internes.

EHNE Nous avons développé un partenariat fructueux avec l'Encyclopédie d'Histoire Numérique de l'Europe, notamment via le projet Emergences Virapic sur le repérage de photographies historiques détournées¹.

SACADO Nous proposons une aide aux chercheurs qui veulent utiliser les ressources de calcul offertes par SACADO. Nous proposons, en accord avec SACADO, d'agir comme relais pour les collègues de la FL qui ont besoin d'accompagnement technique.

BSU Les missions de la BSU (numérisation des fonds, science ouverte, gestion des données) en font un partenaire privilégié de CERES. Depuis sa création, CERES a participé à des formations de la BSU autour des sciences ouvertes et des données.

OBTIC : Nous entretenons une relation étroite avec OBTIC de manière à pouvoir accompagner chacun les projets où nous avons l'expertise la plus spécifique. Notre complémentarité s'illustre également au travers de l'encadrement de la thèse de Caroline Parfait (bourse SCAI, soutenance en janvier 2025) et de la thèse de Jean-Baptiste Tanguy (financement région, soutenue en 2022). Le CERES accompagne aussi les thèses de l'OBTIC du programme doctoral Méthodes numériques pour

les SHS : Matteo Zibardi (directeur Andrea Del Lungo) et Marine Niger (directrice Florence Naugrette).

DRV Faculté de Lettres : La DRV nous accompagne sur le plan contractuel et juridique lors de la signature de partenariats et pour la facturation externe. Elle est un point de contact essentiel pour nous faire connaître des collègues de la Faculté des Lettres, nous mettant en lien avec des collègues que nous pourrions accompagner dans le cadre de dépôts de projet notamment. En outre, la DRV nous permet de bénéficier de quelques heures de travail d'un de ses gestionnaires pour permettre à CERES de passer des commandes.

SCAI : Les relations entre SCAI et CERES sont nombreuses et de différentes natures : l'UMS assure l'accompagnement méthodologique de deux boursier-e-s SCAI. CERES participe aux ateliers de lecture (*journal groupe* co-organisés par OBTIC et SCAI. Enfin, les membres de l'équipe CERES interviennent dans les formations SCAI au titre de leur expertise scientifique et méthodologique (SCAI Education ; écoles d'été).

Nous avons également été partie prenante de la proposition d'initiative Humanités Numériques à Sorbonne Université, initiative dont les activités ont démarré à la rentrée 2025.

2. Appui aux besoins numériques de la communauté scientifique de la Faculté des Lettres

2.1 Ateliers et Hacka/data-thons

Des ateliers mensuels sont organisés pour présenter aux collègues de la Faculté des Lettres des outils et méthodes existantes, susceptibles de les aider dans leurs recherches. Les suggestions des collègues concernant des outils qui les intéressent sont intégrées au calendrier des ateliers. Les ateliers réunissent dix à trente enseignant-e-s-chercheur-se-s et doctorant-e-s selon les séances. Ils constituent une occasion supplémentaire de faire se rencontrer des collègues aux compétences complémentaires et ainsi de favoriser le partage de savoir-faire à l'intérieur même de notre Faculté. Désormais, certain-e-s collègues se proposent spontanément pour prendre en charge des ateliers. Les ateliers organisés sont présentés dans le tableau 2 en page 15.

En plus de ces ateliers, CERES organise chaque année des journées "Datathon" durant lesquelles les collègues de la Faculté des Lettres sont invité-e-s à présenter les difficultés méthodologiques afférentes à leurs données de recherche. Des équipes sont constituées pour travailler en groupe sur des problématiques communes. Préparées en amont, ces journées sont l'occasion de lever rapidement des verrous méthodologiques ou techniques communs. Six datathons ont été organisés (un par semestre) depuis janvier 2022.

2.2 Programme Doctoral

Le programme doctoral Méthodes Numériques pour les thèses en LSHS, adossé à CERES, a démarré en 2021, conjoin-

1. Voir la vidéo de présentation sur le site de CERES

tement à l'UMS. Avec deux contrats doctoraux par an, il a permis de financer des projets innovants sur les méthodes numériques dans la Faculté des Lettres. Il est dirigé par un membre du comité de pilotage de CERES, le jury des bourses est composé de membres de différentes composantes de SU.

2.3 Suivi des doctorant-e-s

Le soutien méthodologique et technique de CERES s'adresse non seulement aux lauréat-e-s du programme doctoral susmentionné, mais également aux autres doctorant-e-s de la Faculté des Lettres. À titre d'exemple, nous avons accompagné une recherche doctorale sur le fonctionnement des algorithmes de modération de Twitter financée par SCAI. CERES a travaillé au développement d'un outil de collecte massive de tweets. Dans le cadre de cet accompagnement d'un doctorant SCAI, l'UMS a développé une fonctionnalité supplémentaire à cet outil, permettant de vérifier si les tweets collectés à une date précise sont toujours en ligne après différents laps de temps (1h, 12h, 24h, 2 jours, 1 semaine, etc.). Si les tweets sont supprimés, cette brique logicielle supplémentaire peut recueillir les informations relatives à la suppression. L'appareillage méthodologique développé dans ce cadre et les résultats auxquels il a donné lieu ont fait l'objet d'une publication dans une revue scientifique répertoriée². Nous avons ainsi accompagné une quinzaine de projets doctoraux de manière plus ou moins intensive selon les besoins ces quatre dernières années. En outre, une trentaine de doctorant-e-s issu-e-s de différentes disciplines ont assisté aux ateliers CERES, aux datathons ou encore au séminaire doctoral Méthodes Numériques pour les thèses en LSHS, animé par les doctorant-e-s CERES.

2.4 Accompagnements des chercheur-se-s

Les accompagnements réalisés peuvent être plus ou moins longs. Cela va de l'appui ponctuel pour un-e chercheur-se qui a besoin d'une récolte de données ou de mieux comprendre le fonctionnement d'un outil numérique, à une collaboration poussée qui implique le développement d'outils dédiés, voire l'implication de CERES dans la participation au projet. Les personnes concernées par ces différents types d'accompagnements sont autant les chercheur-se-s de la Faculté des Lettres que les doctorant-e-s et les masterant-e-s (projets contractualisés, collectifs ou individuels). Nous avons, par exemple, développé un outil d'enrichissement linguistique de corpus, TXM4TXM, qui permet la lemmatisation et l'étiquetage morpho-syntaxique de corpus anciens à des fins de textométrie. Il est désormais utilisé par des chercheurs de la Faculté des Lettres et par des chercheurs extérieurs (Sorbonne Nouvelle notamment). L'ensemble des accompagnements réalisés sont présentés dans le tableau 6 en page 11.

2. Grison T., Julliard V., Alié F., Écrement V., « La modération abusive sur Twitter. Étude de cas sur l'invisibilisation des contenus LGBT et TDS en ligne », *Réseaux*, 2023/1 (N° 237), p. 119-149.

2.5 Conseils sur l'intégration d'un volet numérique dans les réponses aux appels à projets

Ces accompagnements ont parfois pris la forme d'une participation à l'élaboration de projets de recherche contractuels, qu'il s'agisse pour CERES de formuler des conseils sur l'intégration d'un volet numérique dans les réponses (identification de verrous méthodologiques et propositions pour les dépasser) durant la phase de rédaction, ou de s'engager comme partenaire pour développer des méthodes et des outils informatiques spécifiques si le projet était retenu. Entre 2022 et 2025, CERES avait été associé à un dépôt de projet CHIST-ERA, ainsi qu'à plusieurs dépôts de projet auprès de l'ANR. En 2022, un projet élaboré conjointement avec l'EHNE (Encyclopédie d'Histoire Numérique de l'Europe)³ a obtenu un financement de l'appel Émergence SU.

VIRAPIC, circulation des images historiques sur le Web, en collaboration avec l'EHNE Nous avons observé des convergences entre les travaux menés par l'EHNE et CERES autour des images et de leur circulation sur le Web. Le projet VIRAPIC (Histoire, photographies et viralités numériques) propose de restituer les différentes significations et usages de photographies historiques qui circulent dé/recontextualisées sur le Web, depuis leur contexte initial de production et de circulation jusqu'à leurs usages numériques contemporains. Il s'agit d'adjoindre un commentaire historique à ces photographies et d'étudier leur circulation numérique et leurs effets dans la construction d'une mémoire contemporaine des événements du passé. Ceci a permis d'éprouver les outils et méthodes d'annotation d'images mises en place par CERES, notamment dans PANOPTIC (cf section 2.7). Le projet a fait l'objet d'un dossier de 6 pages dans le magazine "La vie - Le Monde" au printemps de cette année⁴. Une vidéo a été réalisée pour détailler notre approche conjointe avec l'EHNE⁵. Enfin, une interview croisée EHNE/CERES est également disponible en ligne⁶.

En 2024, un projet élaboré conjointement avec Philomel⁷ a obtenu un financement de SOUND SU.

Transmettre et développer les compétences pédagogiques en études de genre dans l'enseignement supérieur avec Philomel Il s'agit pour CERES d'accompagner les réflexions sur la conception d'une plateforme de partage de ressource et de participer à la conception de cette plateforme. La plateforme sera officiellement lancée en janvier 2026.

CERES est par ailleurs associé au projet ANR JCJC DATARIGHTS qui a démarré en 2025⁸. Le projet inclut notamment un séminaire mensuel nommé GARNET⁹.

3. <https://ehne.fr/fr>

4. [urlhttps://boutique.lemonde.fr/bilan-atlas-et-geopolitique/3847-la-fabrique-de-l-opinion.html](https://boutique.lemonde.fr/bilan-atlas-et-geopolitique/3847-la-fabrique-de-l-opinion.html)

5. <https://www.youtube.com/watch?v=EwD7KS3tD1A>

6. <https://www.sorbonne-universite.fr/dossiers/intelligence-artificielle/virapic-quand-histoire-et-technologie-unissent-leur-force-cont>

7. <https://philomel.hypotheses.org/>

8. <https://datarights.huma-num.fr/>

9. <https://datarights.huma-num.fr/events/>

2.6 Développements informatiques

Le CERES travaille à la mise à jour régulière d'un catalogue des outils qu'il développe et plus largement des outils qui intéressent la communauté des chercheur-se-s de la Faculté des Lettres, et sur lesquels CERES se forge une expertise pour pouvoir accompagner les collègues (tableau 3 en page 16)

Lors de la première année d'existence de CERES, un premier outil important a été développé, RESTWEET, qui permet la collecte massive, et sur une longue durée, de données issues de la plateforme Twitter. L'intérêt principal de cet outil est d'intégrer une interface graphique permettant d'explorer, en temps réel, les données qui ont été collectées, à l'aide de graphiques, d'un grand nombre de filtres, et de la possibilité de réaliser des requêtes. Par ailleurs, cet outil permet de collecter, en plus du texte des tweets, les images et les vidéos publiées.

PANOPTIC est l'outil-phare de CERES. Nous avons beaucoup investi dans son développement parce que nous pensons qu'il nous permettra progressivement de gagner en autonomie. Il ouvre en effet des perspectives de collaborations particulièrement fructueuses parce qu'il permet de faire sauter des verrous méthodologiques auxquels sont confrontées de nombreuses recherches, quelle que soit leur perspective disciplinaire. Outil de visualisation et d'exploration de grands corpus d'images, il intègre des briques de *machine learning*, permettant la découverte d'images similaires, la création de groupes sémantiques, ou encore l'extraction automatique de texte. L'outil propose diverses options de filtrage, de recherche et de labellisation, qui permettent aux chercheur-se-s de constituer et d'exporter des sous-corpus (cf. Figure 1).

PELLIPOP est un outil qui permet de découper des vidéos en images fixes pour faciliter l'exploration et l'analyse de corpus de films et d'images en mouvement. Il ouvre également des perspectives de collaborations intéressantes en vertu des verrous méthodologiques qui pèsent sur l'analyse de gros corpus de vidéo. À cet égard, nous avons été associés au Consortium Huma-Num pour l'ANnotation, l'analyse Et l'archive de la Vidéo appliquées aux Activités Scientifiques (CANEVAS)¹⁰ par l'intermédiaire du projet ANR TRACTIVE auquel nous collaborons.

SCITOK est un outil pour faciliter l'acquisition, l'exploration et l'analyse des données TikTok développé par Yannick Zelle, dans le cadre de son stage au CERES et de son master à la Heinrich Heine University Düsseldorf. Là encore, cet outil ouvre des perspectives de collaborations fructueuses en ce qu'il reste précurseur pour l'étude scientifique de TikTok. L'outil a ainsi fait l'objet d'une présentation aux Troisièmes journées d'études de l'ARCOM¹¹, le 14 novembre 2024 et a rencontré l'intérêt du régulateur de la communication audiovisuelle et numérique dans le cadre de sa mission de régulation des plateformes numériques, des réseaux sociaux et des risques systémiques.

Deux autres outils, créés en amont au sein du laboratoire STIH, ont été perfectionnés par CERES. Il s'agit d'une part d'EUROPARSER qui permet de compiler et de formater des corpus issus de la base Europresse et, d'autre part, d'OCRES¹² qui permet la reconnaissance optique de caractères de multiples fichiers et leur exportation dans différents formats. Ce dernier outil a permis de mettre en valeur un prototype développé dans le cadre de la thèse de Jean-Baptiste Tanguy (co-encadrement : Glenn Roe, Motasem Alrahabi et Gaël Lejeune).

CERES a également participé, avec l'UMR IReMus, au développement d'une plateforme d'annotation sémantique collaborative en ligne reposant intégralement sur des données ouvertes en RDF/CIDOC-CRM (tant pour décrire la structure des sources annotées que les annotations). Ce service, qui repose sur une API authentifié par comptes ORCID¹³, a été éprouvé dans le cadre d'un outil d'annotation de partitions pour la musicologie. Un outil d'annotations de corpus iconographiques ouvert à toutes les disciplines est en cours de développement.

2.7 Chantiers transversaux stratégiques

En dehors des accompagnements sollicités par nos collègues, et des développements informatiques qui en résultent, CERES mène différents chantiers de fond pour développer des méthodologies innovantes visant à lever les verrous que rencontrent chercheur-se-s mobilisé-e-s dans l'UMS dans leurs recherches, sur leurs domaines d'expertise.

Modération L'accompagnement du projet doctoral relatif à la modération sur Twitter s'est étoffé au regard de l'intérêt de certain-e-s collègues de notre communauté, et a donné lieu à la création du chantier Modération. CERES explore actuellement les façons d'analyser les pratiques de modération dans d'autres réseaux sociaux numériques, à l'instar de FACEBOOK et de TIK TOK.

Visualisation de données CERES s'efforce, dans chacun de ses accompagnements, à penser avec les chercheur-se-s à la meilleure façon de servir leurs questions de recherche singulières. Une question transversale à de nombreuses recherches concerne l'exploration des données de la recherche et des grands corpus collectés, qui plus est lorsqu'ils intègrent différents registres sémiotiques (texte, image, etc.). Aussi, CERES a-t-il œuvré, d'une part, au développement d'une expertise dans la prise en main de différents outils de visualisation (NOTION, GEPHI, TABLEAU PUBLIC, VOYANT TOOLS) et, d'autre part, au développement d'outils spécifiques pour répondre aux besoins d'exploration auxquels les outils existants ne répondaient pas (RESTWEET intègre des outils d'exploration de corpus, PANOPTIC est dédié à l'exploration de corpus d'images, etc.).

Circulation de formes (textes/images) Ce chantier repose sur deux problématiques distinctes : le repérage de formes

catégorie/seminaire-garnet/

10. <https://canevas.hypotheses.org/>

11. <https://www.arcom.fr/actualites/troisieme-journee-detudes-de-larcom-jeudi-14-novembre-2024>

12. <https://ceres.huma-num.fr/ocr/>

13. <https://orcid.org/>

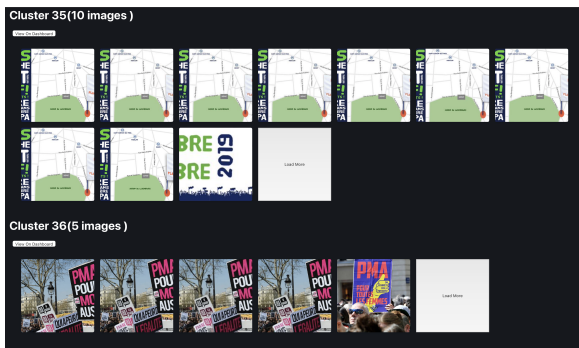


FIGURE 1. Prototype de PANOPTIC : vue des clusters

similaires (textes et images) dans les corpus et la restitution de phénomènes de circulation mis au jour par la recherche.

Pour aider les chercheur-se-s travaillant sur la circulation des images et des textes dans différents contextes (une thèse CERES sur le défigement linguistique et deux recherches sur les mobilisations dans les réseaux sociaux numériques contemporains, une recherche sur les dépêches d'agences de presse au tournant des XIXe et XXe siècle, une recherche sur les Mazarinades dans le contexte de la Fronde dans la France du XVIIe), CERES développe une expertise dans l'usage et l'articulation de différents logiciels de repérage de similitudes permettant de faciliter l'identification de phénomènes de reprises et de citations d'énoncés plurisémiotiques. Les résultats de ce travail enrichissent les fonctionnalités de PANOPTIC.

Un des enjeux du repérage de similitudes est d'identifier la circulation de formes plus ou moins défigées, d'énoncés ayant subi des variations à divers degrés. Pour permettre aux chercheur-se-s de rendre compte de phénomènes de circulation complexes (objet, acteur-riche, temps, espaces pluriels), CERES développe également une expertise dans l'usage et l'articulation d'outils de création de graphes divers. Ce travail pourrait s'étendre à d'autres recherches menées à la Faculté des Lettres (prosopographie, circulation des savoirs, etc.).

Enjeux juridiques et éthiques de l'exploitation des données de la recherche Dans le cadre d'une collaboration avec la BSU, CERES a produit, en 2022, un rapport sur le droit applicable aux données de la recherche, avec une focalisation sur les enjeux spécifiques aux travaux sur le Web (données personnelles, données sensibles, images et droit d'auteur, droit à l'oubli, etc.)¹⁴. Ce rapport parachève une enquête sur les questions animant les chercheur-se-s de la Faculté des Lettres que nous avons recueillies lors de l'atelier CERES du 2 décembre 2021 et s'efforce d'y répondre. Il a fait l'objet d'une restitution lors de l'atelier du 17 février 2022. Ces deux séances d'atelier CERES ont été conduites en partenariat avec la BSU.

Étude du traitement des données personnelles Dans le cadre de notre association avec l'ANR JCJC DATARIGHTS

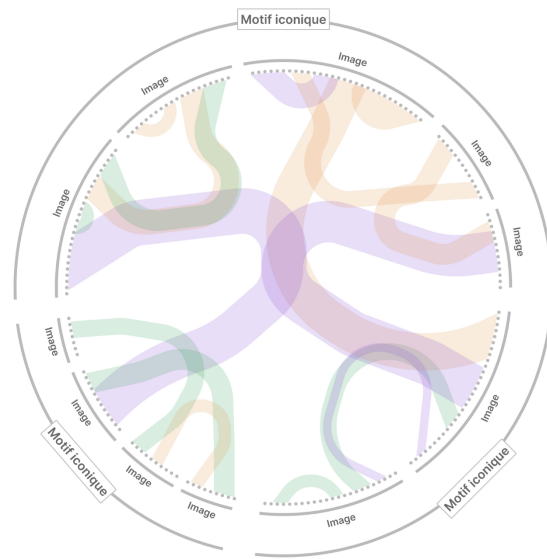


FIGURE 2. Graphe contraint : représentation des images co-publiées par un même compte Twitter et identification des images les plus circulantes au sein d'une controverse

(2025-2028), nous ouvrons un chantier pour réfléchir à l'appareillage méthodologique nécessaire à l'étude du traitement des données personnelles par les plateformes (Section 5).

3. Le projet Semiosis

Le projet SEMIOSIS (Sens, Expérimentation Méthodologique, Interprétation : Outils Scientifiques et Informatiques pour les SHS), soumis par Sorbonne Université pour l'UMS CERES, a fait l'objet d'un financement dans le cadre du Dialogue stratégique et de gestion 2020-2021. Coordination : Virginie Julliard Équipe : Félix Alié, Édouard Bouté, Thomas Bottini, V. Écrement, David Gödicke, Gaël Lejeune, Fred Pailler, Émile Rebours, Antoine Lebrun, Rebecca Bristow, Alexandre Bartz

Grâce à l'emploi de 7 ingénieur-es, pour des durées variables réparties sur 3 années, SEMIOSIS a permis de développer quatre axes principaux concernant les méthodes numériques dédiées à la recherche en humanités et en sciences sociales.

Identifier les gestes communs aux disciplines convoquant l'interprétation pour analyser des corpus (analyse musicale, sémiotique, linguistique, sociologie, info-com...)

Le premier objectif a consisté en une enquête auprès des chercheur-se-s mobilisant des méthodes numériques à la Faculté des Lettres. Après l'établissement d'une cartographie de ces recherches, une seconde phase de description des méthodes et des manières très concrètes de procéder a pu être engagée, principalement grâce à une série d'entretiens avec les chercheur-se-s et au suivi ethnographique dans la durée de certains des projets. Enfin, des ateliers de co-conception ont été organisés avec les chercheur-se-s de la Faculté des Lettres afin de

14. cf. <https://ceres.sorbonne-universite.fr/RapportJulienRossi/>

préfigurer et discuter le développement d'outils d'annotation et de visualisation adéquats.

Offrir aux SHS des outils d'annotation & de visualisation de corpus numériques multimédias Afin de développer plusieurs de ces outils, il s'est agi de nourrir le dialogue SHS/informatique (ML, TAL, IHM, Web sémantique), et produire un lexique et des objectifs communs à des disciplines, des approches du logiciel, et des pratiques de l'étude/enquête relativement distinctes. 3 outils principaux ont pu émerger ou évoluer grâce au financement de SEMIOSIS : 1. Le design d'une interface graphique pour l'application Restweet (développé sur le budget du CERES) : celle-ci permettait la collecte massive, et sur une longue durée, de données issues de la plateforme Twitter. L'interface graphique rend possible l'exploration des données qui ont été collectées, à l'aide de graphiques, d'un grand nombre de filtres, et de la possibilité de réaliser des requêtes supplémentaires. 2. Le prolongement du développement de Sherlock, une plateforme d'annotation sémantique collaborative en ligne reposant intégralement sur des données ouvertes en RDF/CIDOC-CRM. Ce service, qui repose sur une API authentifié par comptes ORCID, a été éprouvé dans le cadre d'un outil d'annotation de partitions pour la musicologie développé à l'IReMus. 3. le développement de Panoptic, un outil de visualisation, d'exploration et d'annotation de grands ensembles d'images (co-financé entre SEMIOSIS et des fonds propres du CERES) : Panoptic permet la navigation dans des corpus de dizaines/centaines de milliers d'images, en intégrant des briques de machine learning. Ces dernières permettant la découverte d'images similaires, la création de groupes sémantiques. L'outil propose diverses options de filtrage, de recherche et de labellisation, qui permettent aux chercheur-se-s de constituer et d'exporter des sous-corpus. Panoptic a fait l'objet de différents types de valorisation auprès de la communauté scientifique (2023-2024). Le prototype du logiciel est accessible sur Github. Il a permis l'accompagnement de plusieurs projets (Appel Emergence SU, ANR) et il a été présenté dans plusieurs articles, séminaires et conférences. 4. Par ailleurs, la dotation a permis de contribuer fortement à la mise en ligne de 4.000 documents numérisés de la Fronde au 17ème siècle (corpus dit des Mazarinades) avec un moteur de recherche et des fonctionnalités d'extraction associées, dans le cadre du projet Antonomaz. La solution technique finalement utilisée est TEI publisher2 qui permet d'exploiter des annotations riches en XML-TEI. Ceci correspond à une demande importante des collègues de la Faculté des Lettres. Nous accompagnons déjà deux projets, un en histoire et l'autre en langue française, dont les corpus seront diffusés via la même technologie.

Confronter annotations machiniques & humaines pour mieux cerner ce qui est irréductible à l'interprétation humaine Un point d'innovation dans les SHS consiste à saisir non pas comment des outils numériques viendraient remplacer des techniques "manuelles", mais plutôt à saisir comment ces outils déplacent les modalités et les limites de l'interprétation

humaine face à un corpus donné. Ce type de déplacements se produit à de multiples niveaux, par exemple en offrant à l'interprétation un corpus exhaustif là où elle aurait du se contenter de portions minuscules de textes ou d'images quelques années plus tôt (à ce propos, le chercheur É. Ollion explique très bien l'importance de cette transformation au micro de X. De La Porte écouter le podcast). Un autre de ces déplacements, particulièrement travaillé dans le cadre de Semiosis, consiste à donner une plus grande mobilité au moment de l'interprétation, voire à démultiplier ce moment. Panoptic offre par exemple les moyens d'assembler des groupes images issues d'un corpus de différentes manières : soit directement à partir de ce que l'on peut en interpréter grâce à la connaissance de leur contexte de production, soit à partir de propriétés strictement formelles des images. La méthode engagée dans le second cas repousse le travail d'interprétation après les opérations de tri à proprement parler, ou dans d'autres cas permet de les relancer avec de nouvelles questions. La nature médiatique des corpus numériques/numérisés, leurs transformations, leurs remix, leur circulation, prend place dès lors au cœur de l'interprétation.

Rendre compte de faits sociaux, culturels et politiques contemporains Les recherches que nous avons menées grâce à ces outils ont permis d'atteindre le 4e objectif. Les outils développés grâce à la dotation SCSP ont permis à CERES d'accompagner des recherches sur la « haine en ligne », telle que la recherche doctorale sur le fonctionnement des algorithmes de modération de Twitter de T. Grison (soutenance prévue en décembre 2024). L'étude portant sur la montée de la droite réactionnaire de V. Julliard & f. Pailler (cf. "The Womb, the Banknote and the Trolley. Elements of French Anti-Gender Visual Culture"), ainsi que celle portant sur les violences policières de É. Bouté (cf. "Militantisme transplateforme : la répression policière des Gilets jaunes entre Facebook et Twitter") ont aussi fait l'objet de publications reposant en partie sur les méthodes et les outils développés dans le cadre de Semiosis.

4. Analyse des pratiques numériques

De manière à identifier les pratiques numériques de notre communauté, et les besoins en termes d'appui méthodologique, nous avons :

1. réalisé une cartographie des travaux de recherche mobilisant des méthodes numériques à la Faculté des Lettres ;
2. conduit une enquête auprès des collègues de la Faculté des Lettres pour saisir leurs activités de recherche spécifiques à l'exploitation des méthodes numériques ;
3. ouvert des permanences mensuelles à la Maison de la Recherche, parfois avec des collègues de la BSU, pour accueillir les collègues de la Faculté des Lettres souhaitant être accompagné-e-s dans l'exploitation ou le développement de méthodes numériques dans leurs recherches.

Pour réaliser la cartographie des travaux de recherche mobilisant les méthodes numériques à la Faculté des Lettres, nous avons commencé par identifier les laboratoires de la Faculté, puis nous avons recherché les publications reliées à ces laboratoires sur HAL, CAIRN, SCHOLAR et IDREF. La cartographie nous permet d'avoir une vision globale de l'usage des méthodes numériques par la communauté scientifique de la Faculté des Lettres, d'identifier les collègues mobilisant de telles méthodes, ainsi que des personnes-ressources pour approfondir l'enquête de terrain. Cette cartographie est disponible en ligne¹⁵. Une version miniature est disponible en annexe de ce document (Figures 3 et 4, pages 18 et 19).

4.1 Podcasts

Afin de faire connaître les différentes actions sur les méthodes numériques menées à la Faculté des Lettres, nous avons réalisé des interviews de chercheurs et chercheuses spécialistes d'outils numériques utiles aux SHS. Les podcasts réalisés dans ce cadre sont listés dans le tableau 5 page 17.

4.2 Formations proposées aux personnels SU

Nous avons animé un module pédagogique dans le cadre de la formation portée par la BSU dédiée aux sciences ouvertes. Ce module, intitulé « Modéliser les connaissances en SHS : pourquoi ? comment ? jusqu'où ?avec le CIDOC-CRM », articule une partie méthodologique (le rôle de l'ingénieur-e dans les phases amont des projets de recherche en SHS produisant des données) et une partie plus technique (la prise en main de l'ontologie sémantique CIDOC-CRM¹⁶). Sa fréquence est pour l'instant annuelle.

4.3 Innovation méthodologique dans la FAIRisation des données de la recherche en SHS

Afin de répondre aux enjeux des sciences ouvertes et de la FAIRisation des données de la recherche tout en visant une innovation méthodologique dans le périmètre des humanités numériques appliquées à l'analyse musicale, nous avons entrepris des actions à la fois techniques et méthodologiques. Nous avons développé une compétence pointue dans l'application et l'extension de l'ontologie CIDOC-CRM, qui est un standard international pour la description des objets du patrimoine immatériel et matériel et la pérennisation des connaissances produites à leur propos. L'innovation portée repose sur le fait de convoquer cette ontologie dès les premières phases des projets analytiques, ce qui confère un caractère « nativement FAIR » aux données de la recherche générées par les chercheur-se-s. Différentes actions ont été menées, dans le cadre de la conception, du développement, de la maintenance et de la valorisation de l'infrastructure Web SHERLOCK (Émergence SU 2019-2021), dédiée à la production, à la gestion et à la diffusion de données scientifiques en SHS :

- Conversion de données scientifiques « historiques » existantes à l'Institut de Recherche en Musicologie en données sémantiques RDF ;

- Construction d'un modèle ontologique générique pour l'annotation de contenus multimédias en SHS ;
- mise en œuvre d'interfaces de saisie ergonomiques pour les chercheurs et chercheuses ;
- Déploiement d'un SPARQL endpoint sur un serveur du TGIR Huma-Num du CNRS
- Mise en œuvre de procédés techniques transférables pour la publication de données ouvertes. Toutes ces actions s'inscrivent dans le paradigme des sciences ouvertes.

5. Contrats, partenariats et dissémination

Il nous apparaît que CERES facilite l'identification de contacts de collègues qui déploient des méthodes numériques dans leur recherche, en interne à SU comme en externe. Nous donnons ici quelques exemples des partenariats de CERES les plus emblématiques noués à ce jour.

Réseaux Pictoria et Canevas CERES fait partie de deux consortiums de recherche nationaux très actifs dans le domaine des Humanités Numériques. D'une part, PICTORIA¹⁷ dédié à l'analyse de corpus d'images via des outils d'intelligence artificielle. D'autre part, CANVAS¹⁸ dédié lui à l'instrumentation des recherches en SHS sur les corpus vidéo. Dans ce cadre, nous participons activement aux journées d'études et autres événements pilotés par ces réseaux. Ceci nous a permis de faire connaître nos développements logiciels dans ces deux domaines : PANOPTIC¹⁹ pour les images, SCITOK²⁰ et PELLIPOP²¹ pour les vidéos.

ANR DATARIGHTS Le projet DATA RIGHTS, piloté par Julien Rossi (CEMTI, Paris 8) s'intéresse à l'impact des algorithmes utilisés par des plateformes numériques (comme Facebook, TikTok) et d'autres organisations sur la société et la démocratie. Ces algorithmes, souvent opaques, compliquent la recherche critique et le contrôle démocratique. Le projet présenté propose une méthode innovante pour les sciences sociales, utilisant les droits d'accès des individus à leurs données personnelles (prévus par le RGPD). L'objectif est de permettre à la recherche et à la société civile de mieux comprendre et exploiter les données obtenues, par exemple pour analyser les biais algorithmiques ou la circulation de l'information sur les réseaux sociaux. CERES sera responsable d'une partie du volet formation d'une part et du chantier analyse des données d'autre part.

ANR TRACTIVE Le projet ANR TRACTIVE²², qui a démarré en mars 2022, vise à caractériser et quantifier la repré-

17. <https://pictoria.hypotheses.org/>

18. <https://canevas.hypotheses.org/>

19. <https://cat.opidor.fr/index.php/Panoptic>

20. <https://cat.opidor.fr/index.php/SciTok>

21. <https://cat.opidor.fr/index.php/Pellipop>

22. Towards a computational multimodal analysis of film discursive aesthetics (responsable scientifique pour Sorbonne Université). Coordinatrice : Lucile Sassatelli, Université Côte d'Azur. ANR, Appel à projets générique, Axe : Révolution numérique : rapports au savoir et à la

15. <https://ceres.sorbonne-universite.fr/>

16. <https://www.cidoc-crm.org/>

sensation du genre et l'objectivation des femmes dans les films et les médias visuels en concevant une analyse du discours multimodal basée sur l'intelligence artificielle. La conception de modèle d'apprentissage profond incluant l'expertise des SHS sur des corpus de films, de séries télévisées et de vidéos collectées sur les réseaux sociaux numériques devra permettre notamment d'identifier des modèles discursifs visuels complexes tels que les régimes scopiques genrés (male et female gaze), qui véhiculent des représentations biaisées. Les chercheur-se-s en SHS rattaché-e-s au laboratoire GRIPIC (partenaire SU du projet) apportent leur expertise en matière d'analyse de la production du genre dans les fictions audiovisuelles. Le CERES apporte son expertise pour faciliter la prise en main, par la doctorante de l'équipe SHS, de l'outil d'annotation de vidéo développé par l'INSERM dans le cadre de ce projet, pour initier l'équipe SHS au fonctionnement du machine learning dans la reconnaissance de formes visuelles, et pour l'accompagner dans ses réflexions méthodologiques et épistémologiques sur la question de savoir quels apports des SHS résistent à l'incorporation dans l'entraînement des modèles d'apprentissage profond. Ce projet fournit par ailleurs un cadre stimulant pour se forger une expertise méthodologique dans le traitement de grands corpus de vidéo.

ANR Ecole Ce projet ANR ²³, déposé au printemps 2022 et qui démarre officiellement en avril 2023, s'intéresse à l'analyse des pratiques de lecture entre le Moyen-Âge central (13^{ème} siècle) et la fin de l'Ancien régime (18^{ème} siècle). CERES apporte ici ses compétences en termes de reconnaissance de caractères et d'analyse de structure de documents. L'appui de CERES permettra notamment d'aider à l'encadrement et à l'émulation au sein de l'équipe, d'informaticiens recrutés, ingénieurs et post-doctorants, et idéalement d'assurer une maintenance au-delà de l'échéance du projet.

Projet ALLIANCE avec l'Université de La Rochelle L'université de la Rochelle (ULR), notamment par le travail de Cécile Chatraine-Braillon et d'Antoine Doucet, cherche à développer le suivi des recherches en humanités numériques et à consolider des projets en HN. Souhaitant s'inspirer de CERES, ils nous ont conviés en mars 2022 à une journée d'études commune à La Rochelle dédiée aux Unités de Service en HN ²⁴. Une deuxième journée aura lieu en juin 2023, cette fois à Paris. Notre partenariat a également permis à un ingénieur de recherche affilié au CERES d'être recruté en Post-Doctorat à l'ULR. Une convention d'hébergement à CERES permet à une ingénieure ULR de continuer de travailler avec l'UMS.

Conservation des compétences développées à la Faculté des Lettres Notre pratique, qui consiste à permettre aux ingénieur-e-s CERES de déployer leurs compétences auprès

de partenaires de l'UMS et à constituer des équipes modulaires impliquant des laboratoires dans SU et au-delà, présente plusieurs avantages. D'une part, cela permet de construire l'expertise en humanités et méthodes numériques de nos jeunes collègues, d'autre part, cela permet la conservation des compétences développées à la Faculté des Lettres. Dans le cadre du projet ANTONOMAZ, par exemple, une chaîne de traitement pour faire de la reconnaissance de caractères (OCR) dans des imprimés du 17^{ème} siècle a été développée. L'implication de CERES a permis que cette chaîne de traitement, initialement utilisable en ligne de commande, bénéficie désormais d'une interface web facilitant son utilisation par les collègues de la FL ²⁵.

À la suite du projet ANTONOMAZ, nous contribuons en partenariat avec l'équipe d'ECRISOI ²⁶ (projet impliquant le CELFF) à une réflexion, pilotée par Karine Abiven (Université de Roue, Ex SU) pour CERES sur la mise en valeur de taxonomies multilingues sur le thème de l'autobiographie.

6. Diagnostic à 58 mois

Depuis quatre ans et demi d'existence, le CERES a acquis une expertise sur 33 outils, développé 8 outils en propre (recensés sur le catalogue OPIDOR du CNRS ²⁷), accompagné plus de 30 projets de recherche. L'UMS est associée à 7 projets financés. Plus d'une centaine de nos collègues de la Faculté des Lettres ont bénéficié de l'appui du CERES ou participé à nos activités.

Indéniablement, les **forces** du CERES tiennent à son positionnement, son expertise et son fonctionnement. L'équipe CERES associe des disciplines, des compétences et des statuts variés, des chercheur-se-s de la Faculté des Lettres acculturé-e-s aux méthodes numériques encadrant des ingénieur-e-s sensibilisé-e-s aux problématiques scientifiques des SHS au profit de la communauté. Ce fonctionnement permet de conduire un travail réellement interdisciplinaire et d'œuvrer au développement de l'« innovation méthodologique » au sein de la Faculté des Lettres. CERES articule ainsi intimement soutien technique, d'une part, et accompagnement méthodologique et scientifique, d'autre part. Le positionnement de CERES parle à nos collègues de la Faculté des Lettres. Si l'UMS encourage le déploiement des méthodes numériques, c'est de manière raisonnée. Ainsi, l'équipe CERES questionne toujours la pertinence de l'usage de méthodes numériques pour des questions de recherche particulières. Il n'est en effet pas évident que les méthodes numériques offrent une plus-value dans tous les cas. Il n'est pas non plus évident que l'outil dont les chercheur-se-s ont entendu parler soit le plus adapté à la recherche menée. L'équipe CERES se fait ensuite facilitatrice du partage de connaissances, demandant à des chercheur-se-s identifié-e-s de faire bénéficier la communauté de leurs expertises (dans les ateliers et les podcasts, notamment). En

culture, 2022-2026. Budget SU : 153 552€. Cf. <https://anr.fr/Project-ANR-21-CE38-0012>

²³. <http://www.univ-paris3.fr/anr-ecole-747549.kjsp>

²⁴. <https://www.univ-larochelle.fr/actualites/journees-detudes-autour-des-humanites-numeriques/>

²⁵. <https://ceres.huma-num.fr/ocr/>

²⁶. <https://ecrisoi.univ-rouen.fr/accueil>

²⁷. <https://cat.opidor.fr/index.php/CERES>

outre, l'équipe CERES arbitre entre formation à des outils existants et développements informatiques rapides et peu coûteux ou demandant plus d'investissement en temps, selon le bénéfice que peut en tirer la communauté. Enfin, l'équipe CERES conduit une réflexion épistémologique sur le statut de l'outil informatique dans les recherches en SHS qui permet d'entretenir les échanges scientifiques avec nos collègues de la Faculté des Lettres.

Par ailleurs, le fonctionnement très souple de l'UMS, son ancrage à la Maison de la Recherche et la possibilité de rencontrer l'équipe en « poussant la porte » du bureau S007 permet de répondre très vite aux demandes. Le CERES a également trouvé un modèle efficace et pertinent : offrant son appui aux collègues de la Faculté des Lettres, notamment aux jeunes doctorant-e-s, encourageant les collègues aux besoins les plus importants à déposer des projets pour bénéficier d'une implication en temps plus conséquente de la part de l'UMS, et facturant ses services aux entreprises.

Les **faiblesses** sont évidemment corrélées et sont de deux types. En premier lieu, le danger d'un étouffement dans une situation où nos capacités sont déjà largement employées, or la disponibilité de l'équipe support reste un élément central de la qualité et de la disponibilité des « prestations » de CERES. L'octroi d'un budget supplémentaire dans le cadre du dialogue stratégique de gestion, en 2021 et 2022, a permis de recruter un ingénieur supplémentaire sur 24 mois (ingénieur infrastructure), un designer à 20% ainsi qu'un mi-temps étudiant et ainsi renforce nos capacités de réponses aux besoins immédiats des collègues sans obérer le travail de fond réalisé par les deux ingénieurs rémunérés sur le budget principal de CERES. Le retour, en 2023 et 2024, au budget initial de l'unité mutualisée de service, constitue une difficulté puisque dans le même temps l'activité de CERES s'est notablement développée. La recherche de fonds propres pour faire ces recrutements nécessaires a des limites. D'abord, certains accompagnements tout à fait nécessaires au rayonnement scientifique de notre université, ceux qui concernent les doctorant-e-s notamment, peuvent plus difficilement donner lieu à un « retour sur investissement » pour l'unité mutualisée de services, dans la mesure où ils ne peuvent pas déboucher sur des dépôts de projets contractualisés à Sorbonne Université, comme c'est le cas pour les titulaires. Par ailleurs, bien que l'UMS commence à obtenir des budgets par ses propres moyens, ceux-là ne couvrent pas encore les besoins, et nous nous retrouvons en manque de ressources humaines. Ceci peut conduire CERES à sélectionner entre les projets à accompagner, au risque de créer de la frustration dans notre communauté. De plus, sachant que les services proposés, en particulier les accompagnements scientifico-techniques, sont une spécificité de CERES au sein de l'écosystème Sorbonne Université, au point d'inspirer certains instituts, les opportunités de faire progresser les Humanités Numériques à la Faculté des Lettres pourraient en pâtir. Il devient nécessaire, pour CERES, de recruter des ingénieur-e-s supplémentaires pour répondre aux demandes, maintenir les outils que le CERES a déjà développés et conti-

nuer de développer des outils utiles aux recherches de la Faculté des Lettres. Le CERES est confronté à une deuxième difficulté. Les conditions de recrutement de l'université (salaire et durée des contrats notamment), ne facilitent pas l'attrait de CERES auprès des profils informatiques susceptibles d'intéresser l'UMS. Les réseaux des membres de l'équipe ont permis de trouver d'excellents candidat-e-s, qui acceptent ces conditions dégradées, dans la mesure où le projet du CERES les passionnent. Toutefois, ajouter à des salaires inférieurs au secteur privé des contrats relativement courts fragilise la structure (plusieurs profils seniors, très intéressants, ont ainsi fini par renoncer à un recrutement en cours). Notre budget actuel couvre un poste d'ingénieur et si nous avons considérablement augmenté nos ressources issues des prestations (45.000€ en 2025 moins les 20% allant en fonctionnement), c'est encore trop juste pour couvrir véritablement un second poste d'ingénieur, une fois déduits les frais de fonctionnement et de matériel.

L'équipe CERES développe une expertise méthodologique dans plusieurs domaines qui offrent de nombreuses **opportunités** d'accompagnement de recherches pouvant bénéficier d'un financement :

- la reconnaissance de similitudes d'images et le regroupement (*clustering*) de ces images piloté en fonction de la question de recherche ;
- le multilinguisme via le traitement de corpus textuels dans différentes formes du Français (anciennes ou modernes), de langues peu dotées ou encore de corpus hétérogènes ;
- Les données bruitées, notamment pour des données textuelles issues de reconnaissance optique de caractères (OCR) ou de reconnaissance d'écriture manuscrite (HTR).

Il nous semble qu'en termes d'opportunités, le soutien aux collègues nouvellement recruté-e-s à la Faculté des Lettres serait une piste à creuser. Ceci afin de faire du soutien aux projets numériques un axe fort pour attirer des profils humanités numériques dans toutes les disciplines de la Faculté des Lettres. Le fait que nous ayons réalisé à ce jour un soutien technique à une dizaine de laboratoires va dans ce sens.

Indépendamment de la future création d'une formation en Humanités Numériques, nous avons entamé un recrutement prometteur de mi-temps étudiants en puisant dans les formations de master de SU. Ceci présente de multiples avantages : (I) pour l'étudiant qui travaille sur son lieu d'enseignement et sur un sujet connexe à sa formation plutôt que dans un emploi moins qualifié, (II) pour la formation où il est intégré puisqu'il peut ainsi être moteur pour ses camarades et (III) pour l'institution qui dispose ainsi de profils qualifiés, déjà formés qui sont plus à même d'entrer dans une carrière d'ingénieur Humanités Numériques.

Pour rappel, **les menaces** que nous avons identifiées durant ces dernières années concernent l'impossibilité, parfois, de récolter certains financements (ex. facturation de service

avec un partenaire slovène qui a échoué en 2023) ou des opportunités manquées du fait de difficultés liées à la gestion administrative des projets. Nous pouvons citer en exemple une proposition ferme de co-tutelle de thèse, financée sur 4 ans, avec l'ADAPT CENTER (Irlande) qui a elle aussi échoué faute de réactivité du côté de notre université (2022 puis 2023). Il y a aussi une certaine opacité sur les modalités d'utilisation et de report d'une année sur l'autre des différentes prestations.

7. Perspectives

Au vu du développement de l'activité de CERES et des résultats que nous avons obtenus en termes d'acquisition d'expertise, de développement d'outils et de visibilité au sein de notre institution et au-delà, nous identifions quatre perspectives pour CERES, au service de la communauté des chercheur-se-s de la Faculté des Lettres et du rayonnement de Sorbonne Université.

- Étant donné les demandes croissantes observées ces deux dernières années : développer notre expertise méthodologique autour de la visualisation de données complexes et de l'analyse de nouveaux corpus (documents sonores et vidéos circulant sur le Web).
- Consolider cinq champs d'expertise qui singularisent CERES au sein de l'écosystème de Sorbonne Université et pour lesquels l'UMS est sollicitée par différents services de Sorbonne Université, les chercheur-se-s de la Faculté des Lettres et au-delà, autant que par le secteur privé : (I) la collecte de données Web, (II) l'appui technique à la FAIRisation des données de la recherche en LSHS, l'appui technique à l'analyse (III) de corpus manuscrits, (IV) de corpus multilingues et (V) de corpus multimodaux (textes, graphiques, images).
- Nous avons amélioré notre visibilité externe, notamment par la participation active à différents réseaux et l'accueil de personnes hors SU dans nos ateliers de formation. Nous souhaitons aller plus loin sur différents plans : accompagnement des doctorant-e-s et des collègues pour des communications, des publications et des dépôts de projets ; accueil de stagiaires étranger-e-s susceptibles d'être intéressé-e-s par les cursus de Sorbonne Université ; participation à des projets nationaux et internationaux ; participation à des initiatives de mise en réseau de structures proches de CERES (UMS, labs, etc.). Ceci pour faire connaître les outils CERES et les recherches auxquelles l'UMS participe.
- En lien avec le VP International, un programme d'échange international pourrait être créé dans lequel CERES accueillerait des doctorant-e-s étrangers (nous avons en particulier développé des liens avec la Slovénie et le Luxembourg) et leur offrirait l'accompagnement proposé aux doctorant-e-s de la Faculté des Lettres. Ceci afin de constituer un pôle d'attraction pour des chercheur-se-s déployant des méthodes numériques, en France et à l'étranger.
- Nous comptons participer activement au déploiement des futures formations en méthodes numériques, en particulier dans la licence et le master Humanités Numériques. Le module du master Recherche et Développement du CELSA que nous avons mis en place pourrait contribuer à nourrir les réflexions pour une offre de formation transversale autour des méthodes numériques.

TABLE 6. Tableau des Accompagnements

Chercheur.euses	Nom/type projet	Laboratoire	Discipline	Date de démarrage
GENRE EN COURS	Plateforme Web	Sound/Philomel	Sociologie, études de genre	19 août 2025
Marine Tiger	Reconnaissance d'écriture manuscrite (HTR)	CELFF	Humanités Numériques	12 mai 2025
Anatole Grimaldi	Étude de données	CREM (extérieur)	Economie	12 mai 2025
Thibault Clément	Étude de données	HDEA	Etudes Américaines	12 mai 2025
Alice Bravard	Étude de données	CRHXIX	Histoire	12 mai 2025
BNF/Pictoria	Développement spécifique Panoptic	BNF/Pictoria (extérieur)	Divers	12 mai 2025
Diane Grenat	Étude de données	GRIPIC	SIC	12 mai 2025
Rimane Karam	Stylométrie	Orient et Méditerranée	Langue Arabe	12 mai 2025
Lorelei Riahi-Castillo	Étude de données	ECPAD/Ministère des armées (extérieurs)	Divers	12 mai 2025
HDEA	Site web de la bo	HDEA	Etudes Anglophones	12 mai 2025
Richard Walter EMAN	Dév Web	THALIM	Humanités Nuémriques	12 mai 2025
Priscille-Laëta ATTELEYN	Ponctuel	GRIPIC	SIC	20 mars 2025
Marie Guermeur	Thèse Marie G.	GRIPIC	SIC	11 décembre 2024
Thomas DeBretagne	Ponctuel : collecte d'utilisateurs twitter		Littérature	18 novembre 2024
Philomel (Hécate/Irene)	Philomel : plateforme	Sound/Philomel	Sociologie	22 octobre 2024
Philomel (Hécate/Irene)	Ponctuel / Plateforme Philomel	Sound/Philomel	Informatique, pédagogique	26 septembre 2024
Sound/Philomel	Benchmark Philomel Plateformes ressources sur le genre	Sound/Philomel	Ergonomie du Web	13 septembre 2024
Priscille-Laëta ATTELEYN	OCR images	GRIPIC	SIC	2 juillet 2024
Hécate Vergopoulos	Presse HDR	GRIPIC	SIC	19 juin 2024
Clément Mabi, Anne Bellon	Hackathon YouScribe	UTC/Costech (Alliance Sorbonne - ext)	SIC	12 juin 2024
Emilie Trébuchet	aide ponctuelle	Inrap / UMR 7324 Citères-LAT (ext)	Archéologie	12 juin 2024
Stavros Lazaris	Accompagnement ponctuel	Orient et Méditerranée	Sirènes	5 juin 2024
Lucille Lamache		GRIPIC	SIC	5 juin 2024
Ambre Abid-Dalençon		GRIPIC	SIC	5 juin 2024
Sébastien Appiotti (et al.)	Europresser l'Eurovision	GRIPIC	SIC	5 juin 2024
Louis Le Corno	Doctorant/Ponctuel	Carism (extérieur)	SIC	5 juin 2024
Alessandra Balotti	OCR Langue Nordiques	HDEA	Langues Nordiques	4 juin 2024

Chercheur.euses	Nom/type projet	Laboratoire	Discipline	Date
Emilie Seris	OCR + annotation	Rome et ses renaissances	Latin	3 juin 2024
Caroline Parfait	Epimethee	OBTIC	Informatique, Linguistique, Littérature	15 mai 2024
Caroline Parfait	OCR pour thèse de Caroline	OBTIC	Informatique, Linguistique, Littérature	4 mai 2024
Alessandra Ballotti		REIGENN	Etudes Nordiques	3 mai 2024
Julien Chamboredon	Aide ponctuelle / organisation documentaire des données de recherche	GRIPIC	SIC	10 avril 2024
Maya Mazzacane	Politique des drogues	GRIPIC	SIC	10 avril 2024
Priscille Atteley	Thèse seconde main	GRIPIC	SIC	10 avril 2024
Amélie Peresson	Retranscription automatique entretien	GRIPIC	SIC	9 avril 2024
Martin Cluseret (masterant)	Retranscription automatique entretien	GRIPIC	SIC	9 avril 2024
Sébastien Appiotti	Visionné l'Eurovision	GRIPIC	SIC	9 avril 2024
Sandrine Lage	Aide analyse thèse	GRIPIC	SIC	8 février 2024
Richard Walter; Elsa Courant	EBDF	CELLF	Littérature	18 janvier 2024
Sylvie Jouet	Zotero (accompagnement ponctuel)	STIH	Linguistique	18 janvier 2024
Anna Arzoumanov	Liberté d'expression (Libertééééééé !)	STIH	Linguistique	18 janvier 2024
Valérie Toureille (Cergy Université)	Projet Européen Femmes dans l'histoire & images	Extérieur (CYU)	Histoire	18 janvier 2024
José Da Costa (Summit)	SUMMIT-KITTANCE	SUMMIT	Entreprise	18 janvier 2024
Pauline Haas (P13)	La littérature de Sony Labou Tansi	Extérieur (Lat-tice)	Linguistique	18 janvier 2024
Labos Clément	Labos Clément	/	Entreprise	18 janvier 2024
Lucas Jamin	Master Occitan	STIH	Linguistique	18 janvier 2024
Alain Serafin	TEISHS	SUMMIT	/	18 janvier 2024
Sylvie Jouet	Verbe Prendre – accompagnement régulier	STIH	Linguistique	18 janvier 2024
Jeong-won Cho	Accompagnement ponctuelle	GRIPIC	SIC	7 décembre 2023
Karine Abiven, Laelia Veron	Transfuge de classe	STIH	Linguistique	7 décembre 2023
Julien Chamboredon	Accompagnement ponctuelle	GRIPIC	SIC	7 décembre 2023
Nelly Quemener, Virginie Julliard	Projet	GRIPIC	SIC	7 décembre 2023
Mathieu Marly	Projet ANR VIRAPIX	EHNE	Histoire	7 décembre 2023
Victor Ecrement	Analyse d'articles scientifiques	IEA Nantes	NLP sur données scientifiques	7 décembre 2023

Chercheur.euses	Nom/type projet	Laboratoire	Discipline	Date
Gabriela Parussa	TXM4TXM ; CoDiF	Sorbonne Nouvelle CLESTHIA	Linguistique	7 décembre 2023
Thomas Sähn	Projet BD (SYSPHER)	CELISO	Linguistique	7 décembre 2023
Léa Andolfi	Thèse/Projet tractive	GRIPIC	SIC	7 décembre 2023
Eva Lacroix	Projet MAD	STIH	Linguistique	7 décembre 2023
Anouch Bourmaysan	UNDERGREW	STIH	Linguistique	7 décembre 2023
Vincent Moriniaux	Mcdo	Médiation	Géographie culinaire	7 décembre 2023
Rimane Karam	Accompagnement thèse	Orient et Méditerranée	Linguistique	7 décembre 2023
Fred Pailler	Exploration de plusieurs corpus de memes	C2DH (Univ du Luxembourg)	SIC	7 décembre 2023
Adélie Laruncet	Accompagnement thèse	GRIPIC	SIC	7 décembre 2023
Barbara Van Doosselaere, Margot Mahoudeau	Carto DRV (cartouche)	DRV	/	7 décembre 2023
Gilles Couffignal	Vers Métrique	STIH	Linguistique	7 décembre 2023
Sébastien Appiotti, Lisa Boltz	AutoMinet	GRIPIC	SIC	7 décembre 2023
Julien Bezançon	Accompagnement ponctuel	STIH	Linguistique	7 décembre 2023
Gilles Couffignal	accompagnement ponctuel	STIH	Linguistique	15 mars 2023
Adelaide MARINE-GOUGEON	accompagnement ponctuel	CRHXIX	Histoire	15 mars 2023
Killian Moreau				15 mars 2023
Aurore Aymerie	accompagnement ponctuel	Muséum	Biologie	13 mars 2023
Philippe Monneret ; Xin Li	Projet GG	STIH	Linguistique	13 mars 2023
M. Chen	Thèse aide ponctuelle			5 février 2023
Laurie-Anne Laget	Projet - Vie Littéraire Torre / Accompagnement	CRIMIC	Littérature	30 janvier 2023
Kong xiangfeng	Thèse - Pouvoir / Ponctuel	STIH	Linguistique	26 janvier 2023
Caroline Richard	SCHOLIB	Rome et ses renaissances	Linguistique	26 janvier 2023
Marie-Lise Buisson	Thèse - yuka / Ponctuel	GRIPIC	SIC	26 janvier 2023
Renaud Debailly	Droit à l'oubli	GEMASS	Sociologie	26 janvier 2023
Gilles Couffignal	Langue Occitane	STIH	Linguistique	26 janvier 2023

Chercheur.euses	Nom/type projet	Laboratoire	Discipline	Date
Jean-Louis Jean-nelle, Françoise SImonet-Tenant, Jean-Christophe Igalens	Ecrit soi	CELLF	Littérature	1 juin 2022
Pierre-Marie Chauvin	Sackler	GEMASS	Sociologie	7 avril 2022
Claire Crignon	Collecte de données		Histoire de la médecine, Philosophie de la médecine	7 avril 2022
Vincent Moriniaux	aide ponctuelle	Médiation	Géographie	7 avril 2022
Marie-Cécile Souleyreau	Restitution de données	Centre Roland Mousnier	Histoire	7 avril 2022
F. Viale	Analyse NLP	STIH	Linguistique italienne	7 avril 2022
François-Joseph Ruggiu, Cyril Grange, Michael Gasperoni	OCR	Centre Roland Mousnier		7 avril 2022
Virginie Julliard, Fabien Guilloux, Anne Piéjus, Sébastien Appiotti	SHERLOCK	GRIPIC, IReMus	SIC	7 avril 2022
Renaud Debailly, Hugo Jeaningros	Droit à l'oubli pour les malades du cancer	GEMASS	Sociologie	7 avril 2022
Karine Abiven, Gaël Lejeune, Jean-Baptiste Tanguy	Site Antonomaz	CELLF, OB-TIC, STIH	Informatique, Linguistique, Littérature	7 avril 2022
Virginie Julliard, Thibault Grison, Kenza Benabdelouhab	Modération en ligne	GRIPIC	SIC	7 avril 2022
Virginie Julliard, Caroline Marti, Nelly Quemener, Sébastien Appiotti, Thibault Grison, Kenza Benabdelouhab	Analyse de Données	GRIPIC	SIC	7 avril 2022
Guillaume Métayer	Analyse de données	CELLF	Littérature	7 avril 2022

NOM	PRENOM	QUOTITE	FONCTION	Contrat	Coût	Provenance crédits
ALIE	FELIX	100%	Directeur Technique	12 mois	50 100	CERES
BOUTE	Edouard	100%	Ingénieur SHS	8 mois (fin : 31/08/25)	34 000	CERES
GODICKE	David	100%	Ingénieur en Informa- tique	1 mois (à partir 01/12/25)	3 500	Reliquats CERES

TABLE 1. Effectif de l'unité CERES pour 2025 (en gras les deux ingénieurs rémunérés sur fonds propres)

Intitulé	Thèmes	Date
Europresse	Presse ; 4 europresse ; txm ; corpus	2024-12-05
TXM (Textométrie)	textométrie ; TXM	2024-11-14
Excel pour les SHS	excel ; données ; tableur	2024-10-17
Web Scraping	scraping	2024-05-30
Musiconis	iconographie ; corpus	2024-04-25
Panoptic	images	2024-04-11
Gephi	gephi ; graphes de réseaux ; visualisation	2024-03-28
Tableau Public	visualisation ; tableau public	2024-02-08
Anonymisation des données	réseaux sociaux ; rgpd ; données personnelles	2023-06-22
Gephi	gephi ; graphes de réseaux ; visualisation	2023-03-02
Tableau Public	visualisation ; tableau public	2023-03-02
Voyant Tools	Encodage de Corpus ; TAL	2023-02-09
Excel pour les SHS	excel ; données ; tableur	2022-12-08
Antconc & TXM	Textométrie ; Corpus	2022-11-10
Notion	notion ; annotation	2022-10-06
ARCGIS : cartes interactives	arcgis ; sig ; cartographie	2022-03-17
Scraping	Scraping ; données ; Python	2022-01-22
PGD, RGPD et archivages des données personnelles dans les recherches en SHS	données personnelles ; rgpd	2021-11-19
CrowdTangle	crowdtangle ; réseaux sociaux	2021-11-18
Iramuteq	iramuteq ; textométrie	2021-09-09
TXM (Textométrie)	textométrie ; TXM	2021-10-21

TABLE 2. Ateliers CERES

	Description de l'outil, de la ressource
AntConc	Boîte à outils d'analyse de corpus pour l'analyse de texte
Audacity	Logiciel d'édition de fichiers audio permettant de réaliser des montages
Biopython	Librairie Python permettant entre autres de comparer des chaînes de caractères entre différents textes pour trouver des similarités (logiciel d'alignement permettant la reconnaissance de chaînes identiques)
CorText Manager	Application web d'extraction d'informations et analyse socio-sémantique pour les sciences humaines et sociales
Daniel	Daniel est un outil de classification d'articles de presse fondé sur une recherche des mots et groupes de mots les plus saillants d'un document. Combiné à un lexique spécialisé, il peut réaliser une extraction d'information, cette option a notamment été utilisée pour la veille épidémiologique multilingue
Europresse	Site web permettant d'accéder et de faire des requêtes auprès d'une vaste base de données d'articles de presse et de contenus web
Excel	Logiciel tableur particulièrement propice à la réalisation de calculs et de graphiques à partir de données numériques, mais aussi à partir d'objets textuels. Permet de traiter des fichiers .csv. L'usage des tableaux croisés dynamiques et des graphiques dynamiques est particulièrement utile pour la recherche en SHS
Framapad	Traitement de texte collaboratif
Gephi	Logiciel permettant de réaliser des cartographies de réseaux (graphes) et d'accompagner à leur analyse
Graphcommons	Outil de création et d'analyse de graphes à la main ou à partir de bases de données. Inclut des fonctions "narratives" pour faire des présentations interactives.
Hedgedoc	Logiciel d'édition collaborative de documents markdown en libre accès à héberger sur son propre serveur. Il n'a pas d'interface de gestion de fichiers arborescente, ce qui incite à créer des réseaux de documents grâce à des liens hypertexte
Iramuteq	Analyse statistique automatique et classification de corpus de textes. Permet de réaliser des analyses multidimensionnelles de textes et de questionnaires
Notion	Outil de traitement de texte et de stockage de données sous formes de tableaux dynamiques. Un mélange entre un word, un excel et un wiki, tout en étant très simple
Obsidian	Obsidian est un logiciel de prise de notes qui fonctionne avec des fichiers Markdown. Son intérêt principal consiste en ce qu'il permet aux utilisateurs de créer des liens internes entre les différents fichiers de notes, puis de visualiser les connexions sous forme de graphes
Octoparse	Outil qui permet de scraper et de crawler avec une interface graphique
Tableau Public	Tableau public est un logiciel en ligne gratuit qui permet de créer des visualisations interactives, hautement personnalisables et accessibles en ligne. Sa particularité est que les visualisations ne sont pas construites selon des diagrammes standards, mais de manière modulaire, permettant de tester différentes manières de représenter et spatialiser les métadonnées
TEI Publisher	Outil de publication de contenu sous format XML, intéressant pour interroger les métadonnées à l'aide de filtres réalisés sur les documents publiés
Trafilatura	Trafilatura est un package Python et un outil en ligne de commande conçu pour rassembler du texte sur le Web. Il comprend des composants de découverte, d'extraction et de traitement de texte. Ses principales applications sont le crawling, le téléchargement, le scraping et l'extraction de textes, de métadonnées et de commentaires
Treeverse	Outil de lecture des tweets et de ses commentaires sous forme d'arborescence. L'outil affiche également les temps de réponses entre chaque commentaire
Twint	Outil de collecte des données publiées sur Twitter, sans passer par l'API, en fonction de certains mots-clés et différentes caractéristiques des tweets
TXM	Logiciel d'analyse textuelle de corpus qui est plus basé sur les caractéristiques lexicométriques que le logiciel Iramuteq
Voyant Tools	Plein de petites visualisation et analyses à partir d'un corpus de texte. Permet d'avoir une bonne vue d'ensemble et différentes façons d'aborder le corpus

TABLE 3. Compétences outillées du CERES

Outil	Usage
Europarser	Outil permettant de transformer des exports HTML du site Europress, vers des formats directement exploitables dans des logiciels de traitement textuel comme Iramuteq ou TXM
OCRes	Outil permettant de transformer des fichiers PDF en fichiers textes exploitables
Panoptic (en développement)	Outil de visualisation et d'exploration de grands corpus d'images. Il permet de naviguer dans un corpus à l'aide d'outils de machine learning tels que la recherche d'image similaires, la création de groupes sémiotiques, ou encore l'extraction automatique de texte. L'outil propose diverses options de filtrage, de recherche et de labellisation, permettant la création et l'exportation de sous-corpus
Restweet	Outil de collecte massive sur le long terme de données Twitter. L'outil intègre une interface graphique permettant d'explorer les données à l'aide de graphiques, d'un grand nombre de filtres et d'options de recherche. En plus du texte, Restweet permet de réaliser de la collecte d'images et de vidéos

TABLE 4. Outils développés par le CERES

Intitulé	Interviewé-e
Europresse (1)	T. Grison (SCAI/GRIPIC)
Europresse (2)	G. Lejeune (STIH/CERES)
Clustering	S. Poder et C. González Gallardo
Extraction de données sur Instagram	M. Buffard (GRIPIC/SCAI)
Visualisation de sources multiples	A. Bartz (Projet Antonomaz)
RGPD et sciences sociales	J. Rossi (CEMTI/Paris 8)
Iramuteq (1)	A. Wrona (GRIPIC)
Iramuteq (2)	L. Bolz (GRIPIC)
Iramuteq (3)	R. Debailly (GEMASS)
TXM	G. Lejeune (STIH/CERES)

TABLE 5. Podcasts CERES

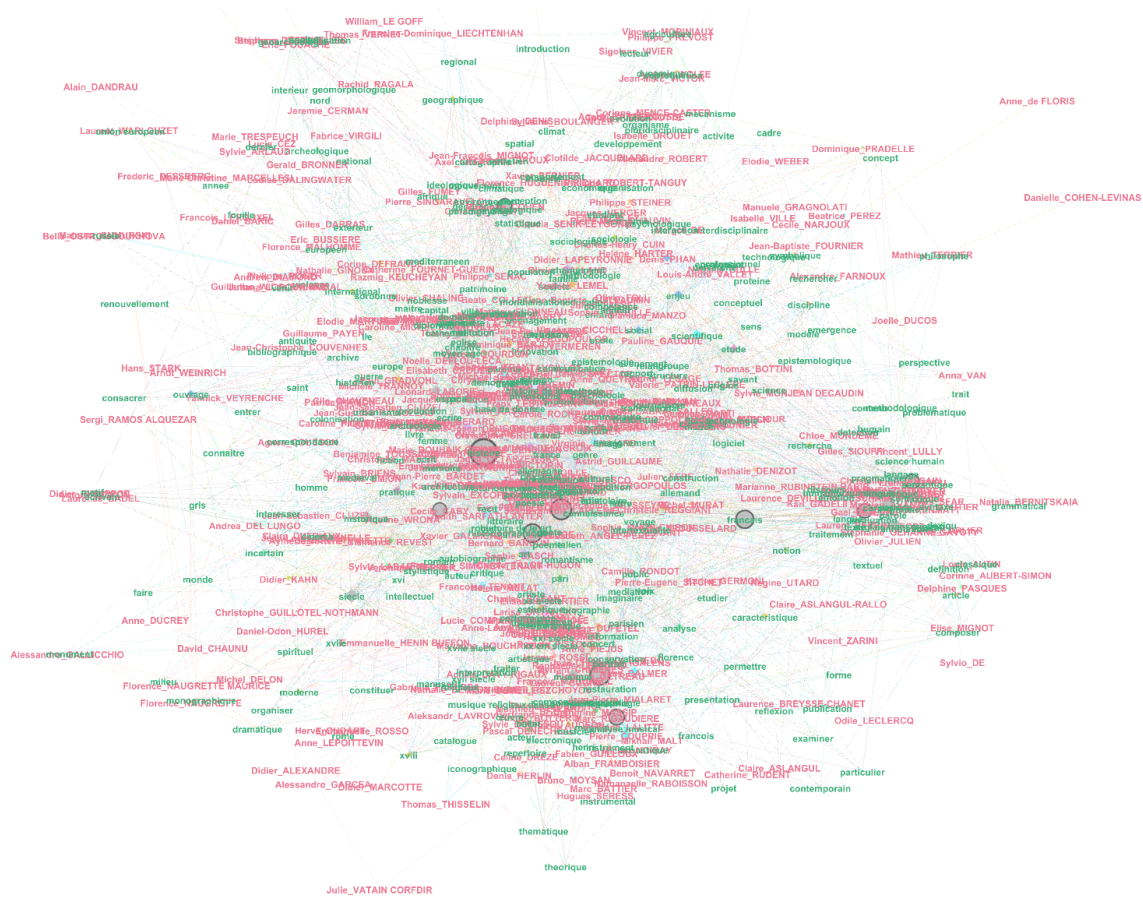


FIGURE 3. Cartographie des usages des méthodes numériques à la Faculté des Lettres

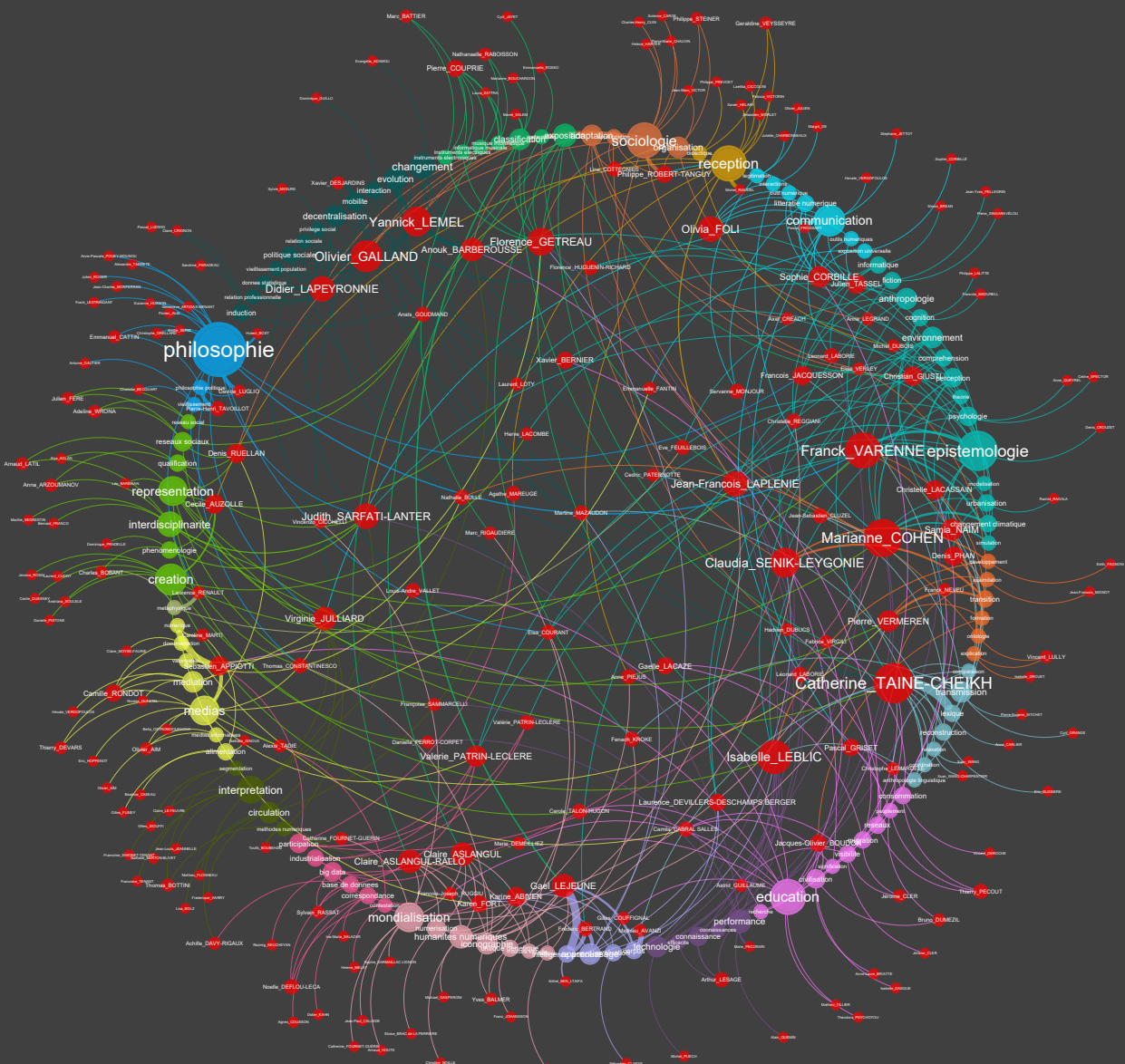


FIGURE 4. Usages des méthodes numériques à la Faculté des Lettres (Cartographie circulaire)