

# Projet EVEille

(Exploration et Valorisation Electronique des corpus de l'ILLE)

Animé par Anne RÉACH-NGÔ, Marine PARRA  
et Régine BATTISTON

(janvier-mai 2021)

Organisé en cinq journées mensuelles de janvier à mai 2021, le colloque EVEille organisé par l'Institut de recherche en Langues et Littératures Européennes de l'Université de Haute-Alsace (EA 43 63) vise à lancer, au sein de l'ILLE et en collaboration avec d'autres acteurs du monde des SHS et des Humanités numériques, une réflexion collective sur les usages du numérique, les atouts et innovations mais aussi les obstacles à sa mise en œuvre, ainsi que les évolutions à venir des pratiques de recherche. Il entend prendre ses distances avec le discours du « tout-numérique » et revenir à une réflexion sur un humanisme numérique fondé sur la mise en commun des expérimentations intellectuelles, des explorations informatiques et leur appropriation par les usagers, qu'ils soient chercheurs, bibliothécaires et documentalistes ou encore responsables de centres culturels.

L'objectif de cet espace de réflexion n'est pas de défendre un nécessaire virage des SHS en direction des humanités numériques, malgré les fortes incitations des organismes de financement et d'évaluation de la recherche qui par la pression imposée en dénaturent les méthodes et les questionnements. Il s'agit en revanche de donner aux participants qui s'interrogent sur l'intérêt des humanités numériques les moyens d'examiner si une telle orientation peut – ou non – être pertinente dans le cadre de leurs recherches ou dans le traitement des biens et objets patrimoniaux et culturels dont ils ont la charge. Il s'agit également de permettre aux porteurs de projets en humanités numériques de présenter leurs travaux et d'échanger sur la mise en réseau de ces différents projets, qui restent encore parfois mal intégrés aux dynamiques d'équipes ou à l'environnement culturel local.

## Conférence inaugurale 9 h 15 - 10 h 15

**Ian JOHNSON** (Honorary Associate, Faculty of Arts and Social Sciences, University of Sydney)

*Concevoir et collaborer sur des bases de données sans réinventer la roue*

La construction d'une base de données pour la recherche nécessite une approche holistique et une réflexion importante qui dépasse le cadre strict de l'outil. Il faut prévoir les ressources techniques nécessaires à l'élaboration du système, les délais de mise en place et de mise à jour, les dépendances vis-à-vis d'individus dont l'engagement n'est pas garanti, les coûts de maintien, la compatibilité avec le travail des autres et la pérennité du support (infrastructure, logiciel, sites web), le tout dans un environnement technique toujours en flux.

Il s'agira ici d'indiquer comment éviter ces pièges pour arriver rapidement à des systèmes efficaces en bâtissant sur des fondations sûres : l'expérience d'autres chercheurs et l'élaboration mutualisée de modèles et de systèmes d'information adaptés aux besoins (divers) des chercheurs en sciences humaines. En suivant ce chemin, des solutions plus immédiates, plus complètes, plus réactives et adaptables à l'évolution et besoins de nos projets, à moindre coût, à moindre risque, et plus stable à long terme que la réinvention de la roue et du sur mesure sont trouvés, et fonctionnent pour tout type de projets. Cette hypothèse sera examinée à la lumière d'exemples de divers projets élaborés sur une infrastructure commune : Heurist.

## Retours d'expérience 10 h 30 - 11 h 45

Table ronde animée par Anne RÉACH-NGÔ & Marine PARRA

**Romain WEBER**, chercheur indépendant

*Enjeux et difficultés d'une base de données bien construite pour l'étude d'un corpus : l'exemple de la réalisation d'une base bibliographique de canards des XVI<sup>e</sup>-XVII<sup>e</sup> siècles*

Retour d'expérience sur la réalisation d'une base de données bibliographique visant à cerner la pertinence de la notion de « canard » (brochures de faits divers) aux XVI<sup>e</sup> et XVII<sup>e</sup> siècles. Les différentes étapes de construction de la base de données depuis la rétroconversion des bibliographies papier et les problématiques qui ont surgi : quelles données récolter, quel type de transcription réaliser, comment structurer l'information et quels résultats obtenus.

**Miriam SPEYER**, ATER à l'IUT de Paris - Rives de Seine (CERILAC EA 4410, Université de Paris)

*10000 poèmes du XVII<sup>e</sup> siècle : naviguer dans un corpus massif à l'aide d'une base de données*

Au début du XX<sup>e</sup> siècle, Frédéric Lachèvre fait, dans sa Bibliographie des recueils collectifs de poésies publiés entre 1597 et 1700, l'inventaire des compilations poétiques parues tout au long du XVII<sup>e</sup> siècle. Or, la bibliographie matérielle, quoique très précieuse, limite fortement les angles d'approche, tandis que la base de données permet, elle, de tenir compte de toute la plasticité de ce corpus hétérogène. À l'exemple des recueils collectifs de poésies du XVII<sup>e</sup> siècle, l'intervention cherchera à présenter la genèse et les apports heuristiques de ce type d'outil numérique, tout particulièrement dans un travail de thèse, et d'en esquisser les prolongements possibles.

**Julien MÜLLER**, ingénieur d'études du projet ANR BDL (Bureau des Longitudes) /1795-1932 (Observatoire de Paris, Bureau des Longitudes, MSH Lorraine, Archives Henri-Poincaré)

*Les procès-verbaux du Bureau des longitudes (1795-1932). De la valorisation d'un corpus numérisé à la construction d'outils d'exploration des données pour l'histoire des sciences*

Initié en 2016, le projet ANR « Le Bureau des longitudes (1795-1932) – De la Révolution française à la Troisième République » arrivera prochainement à son terme. Conçu initialement dans une optique de valorisation d'archives numérisées entièrement retranscrites et mises en ligne sur un site web dédié (bdl.ahp-numerique.fr), le projet s'est étoffé progressivement afin de fournir aux historiens des sciences aussi bien qu'à un public plus large d'amateurs d'histoire, des outils facilitant l'exploration de ce vaste corpus de plus de 2,5 millions de mots (bases de données, index, focus, dossiers thématiques, etc.). À travers une courte présentation du site web et des bases de données accessibles en ligne, nous évoquerons l'intérêt qu'a pu présenter l'usage d'outils numériques dans le cadre de ce projet, tout en portant un regard critique sur les solutions mises en place.

**Béatrice JOYEUX-PRUNEL**, professeure à l'université de Genève (chaire des Humanités numériques, directrice du Centre d'excellence Jean Monnet IMAGO à l'École normale supérieure de Paris et du projet Artlas (<https://artlas.huma-num.fr>))

*Étudier la circulation des images avec le numérique*

L'outil numérique est le plus souvent utilisé pour éditer et étudier des textes (on pense au fameux « distant reading » mis en valeur par Franco MORETTI). Or les images méritent elles aussi de ne pas être étudiées que par des moyens monographiques. Cette intervention montrera quelques stratégies possibles pour les aborder en grand nombre, en circulation sur le long terme et à l'échelle mondiale, les approches statistiques et cartographiques à l'application plus récente d'algorithmes de vision artificielle, en passant par le web sémantique.

11 h 30 : discussions

## Speed searching 11 h 45 - 12 h 30

Cette session vise à donner la parole à de jeunes chercheurs ou à des chercheurs confirmés qui débent un projet et s'interrogent sur les choix à mener en matière d'outils et de méthodologie. Le défi d'une présentation de 5 minutes "montre en main" les incitera à caractériser la singularité de leur projet de recherche et à formuler leurs attentes en matière de base de données pour aborder le questionnement scientifique.

**Gercélia BATISTA DE OLIVEIRA MENDES**, doctorante à l'Unistra (LilPa)

*Une base de données trilingue pour les termes juridiques et bioéthiques liés à la fin de vie*

**Silya BENAMAR**, doctorante à l'UHA (ILLE)

*Collecte de tweets en anglais afin d'y étudier l'utilisation et la compréhension de figures de style*

**Audrey DECK**, masterante à l'Unistra (LilPa)

*Collecte de tweets en allemand, alsacien et français pour la prédiction multilingue d'émojis*

**Tinghan CHEN**, masterant à l'Unistra (LABO)

*L'analyse stylistique et thématique d'un corpus de presse satirique en ligne et développement d'un classifieur*

## Atelier 13 h 45 - 15 h

Séance d'initiation au site collaboratif de constitution de base de données Heurist, animé par Guillaume PORTE (Unistra).

## Projets collaboratifs 15 h 15 - 16 h 30

15 h 15 - 15 h 30

Introduction par Anne RÉACH-NGO

15 h 30 - 16 h

Table ronde animée par Guillaume PORTE

Collaborations entre ingénieur et chercheur, circulation des compétences et dynamique de projet

16 h - 16 h 30

Discussion

Échanges entre ingénieurs et chercheurs (titre provisoire)