



Journées EVEille (Exploration et valorisation Electroniques des corpus en Sciences Humaines et Sociales)

Constitution des données de la recherche

Animées par Anne Réach-Ngô, Marine Parra et Régine Battiston

janvier-juin 2021

Organisé en cinq journées mensuelles de janvier à juin 2021, le colloque EVEille organisé par l'Institut de recherche en Langues et Littératures Européennes de l'Université de Haute-Alsace (UR 4363) 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. Le projet EVEille 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.

Comment certains projets de recherche ou de valorisation patrimoniale et culturelle peuvent-ils gagner à s'orienter en direction du numérique ? Dans quelle mesure les méthodologies employées s'ancrent-elles dans des pratiques plus anciennes qui ont fait leurs preuves ? Tous les projets se prêtent-ils à une telle approche ? Comment s'y prendre pour engager, poursuivre, relancer ou enrichir un projet de recherche à l'aide des humanités numériques ?

Contacts :

Projet-eveille@uha.fr

Inscription sur [Sciencesconf](https://www.sciencesconf.org/)

Carnet du projet [EVEille](https://www.eveille.uha.fr/)

Chaîne vidéo sur [Uha pod](https://www.uha.fr/)

 [@Eveille HN](https://www.facebook.com/EveilleHN)

 [@eveille_hn](https://www.instagram.com/eveille_hn)

 [@Eveille_HN](https://twitter.com/Eveille_HN)

9h RÉGINE BATTISTON, ANNE RÉACH-NGÔ et MARINE PARRA (Université de Haute-Alsace, ILLE), Ouverture du colloque

9h15-10h15 Conférence inaugurale

IAN JOHNSON (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, l'engagement discontinu des participants, les coûts de maintien, la compatibilité avec le travail des uns et 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 traiter ces difficultés pour arriver rapidement à un système efficace 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, nous trouverons des solutions plus immédiates, plus complètes, plus réactives et adaptables à l'évolution et aux besoins de nos projets, à moindre coût, à moindre risque, et plus stable à long terme que la réinvention de la roue et le sur-mesure. En outre, ces solutions devront fonctionner pour tout type de projets. Cette hypothèse sera examinée à la lumière d'exemples de divers projets élaborés sur une infrastructure commune : [Heurist](#).

10h30-11h45 Retours d'expérience animés par 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^e-XVII^e siècles »

Il s'agira ici d'offrir un 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^e et XVII^e siècles. Seront présentées 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 (IUT de Paris - Rives de Seine, CERILAC), « 10 000 poèmes du XVII^e siècle : naviguer dans un corpus massif à l'aide d'une base de données »

Au début du XX^e 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^e 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^e 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, 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 ([Les procès-verbaux du Bureau des longitudes](#)), 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 (Université de Genève, École normale supérieure de Paris), « 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. À partir de la présentation du projet [Artlas](#), cette intervention montrera quelques stratégies possibles pour les aborder en grand nombre, en circulation sur le long terme et à l'échelle mondiale, des approches statistiques et cartographiques à l'application plus récente d'algorithmes de vision artificielle, en passant par le web sémantique.

11h30 Discussions

11h45-12h30 Speed searching

Cette session vise à donner la parole à de jeunes chercheurs ou à des chercheurs confirmés qui débutent 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 numériques pour aborder le questionnement scientifique.

GERCÉLIA BATISTA DE OLIVEIRA MENDES (Université de Strasbourg, LiLPa), « Une base de données trilingue pour les termes juridiques et bioéthiques liés à la fin de vie »

SILYA BENAMMAR (Université de Haute-Alsace, ILLE), « Collecte de tweets en anglais afin d'y étudier l'utilisation et la compréhension de figures de style »

AUDREY DECK (Université de Strasbourg, LiLPa), « Collecte de tweets en allemand, alsacien et français pour la prédiction multilingue d'emojis »

Pause déjeuner

13h45-15h Atelier d'initiation aux outils numériques

Séance d'initiation à Heurist, animée par GUILLAUME PORTE (ingénieur d'études, Université de Strasbourg, ARCHE)

Heurist est un système intégralement en ligne et collaboratif permettant la création de bases de données sans connaissance particulière en informatique. Pensé par Ian Johnson, archéologue et spécialiste en humanités numériques, Heurist permet aux chercheur.e.s de se réapproprier leurs données en facilitant la structuration de celles-ci au moyen d'une interface de saisie préexistante et entièrement paramétrable. Générique sur le plan technique, Heurist est prévu pour une utilisation très spécifique en fonctions des enjeux d'une recherche et propose plusieurs manières d'interroger les données, depuis les fonctions de tri sur une typologie jusqu'à la projection sur carte, en passant par la visualisation de réseaux et de frises chronologiques. L'atelier permettra de découvrir l'interface et ses possibilités, depuis la création d'une base jusqu'à l'exploitation des données.

15h15-16h30 Projets collaboratifs en humanités numériques

15h15-15h30 ANNE RÉACH-NGÔ (Université de Haute-Alsace, ILLE), Présentation du projet Réson@nces Grand-Est

Le projet Réson@nces Grand-Est vise à réunir les initiatives des centres de recherches, SUAC, bibliothèques et centres culturels de la région Grand-Est pour envisager des partenariats en matière de valorisation des biens scientifiques et culturels qui relèvent des sciences humaines. Le cadre général de ce projet sera rapidement présenté et servira à nourrir les échanges, du point de vue de la gestion de projet avec les invités de la table ronde, qui présenteront les modalités pratiques de leurs collaborations au sein d'un projet spécifique, et ensuite lors des échanges avec les participants du colloque, qui sont invités à faire des propositions.

15h30-16h15 Table ronde « Collaborations entre ingénieurs et chercheurs, circulation des compétences et dynamique de projet », animée par ANNE RÉACH-NGÔ

La table ronde consacrée à l'épistémologie et la méthodologie de la recherche en environnement numérique partira de la présentation d'une expérience collaborative entre un chercheur en histoire médiévale, Thomas Brunner, et un ingénieur d'étude, Guillaume Porte, autour de la base Chirographum. Les échanges auront pour but de mettre en valeur les dynamiques intellectuelles, à la fois scientifiques et techniques, qui nourrissent l'élaboration d'un projet et conduisent le chercheur comme l'ingénieur à déplacer son point de vue pour mettre en œuvre une méthodologie adaptée à l'objet d'étude, aux techniques et outils à disposition.

GUILLAUME PORTE (ingénieur d'études, Université de Strasbourg, ARCHE) et THOMAS BRUNNER (Université de Strasbourg, ARCHE), Chirographum

Chirographum est une base de données en cours de développement mettant à disposition des chercheurs un type de documents médiévaux, les chirographes de juridiction du Nord de la France et de la Belgique actuels (actes juridiques [contrats divers, testaments] en plusieurs exemplaires validés par des juridictions laïques comme les municipalités ou parfois des cours seigneuriales). Contrairement aux bases de données textuelles concernant la période médiévale constituées depuis les années 1980, *Chirographum*, part non pas du texte mais du document pour répondre aux besoins de la recherche sur les pratiques de l'écrit médiéval.

16h15 Discussions et clôture de la journée



12 FÉVRIER 2021 JOURNÉE 2. BIBLIOTHÈQUES NUMÉRIQUES

9h RÉGINE BATTISTON, ANNE RÉACH-NGÔ et MARINE PARRA (Université de Haute-Alsace, ILLE), Ouverture de la journée

9h15-10h15 Conférence inaugurale

RICHARD WALTER (ingénieur de recherche, Sorbonne nouvelle, Thalim, CNRS/ENS), « Bibliothèque numérique et nouveaux usages »

La bibliothèque numérique s'est imposée comme un outil indispensable dans le monde académique et culturel. La bibliothèque a une définition communément admise (« une collection organisée de documents » *a minima*) mais en y adjoignant le terme numérique, cette définition est devenue très plurielle. Les missions dévolues aux bibliothèques « classiques » (constituer et valoriser une collection, conserver et communiquer un document) sont aussi prises en charge par les bibliothèques numériques mais dans des dimensions autres : les possibilités exponentielles et combinatoires de l'informatique font que les corpus deviennent de plus en plus étendus, avec des services nouveaux mais aussi des protocoles et des standards très hétérogènes. La notion même de « collection organisée » a été remise en question par les objectifs du web sémantique. Le domaine des bibliothèques numériques rassemble en somme une grande variété d'approches, de déclarations programmatiques et de résultats divers. Après un tour d'horizon historique avec de grands opérateurs comme Europeana ou Gallica, notre présentation examinera le succès du logiciel Omeka qui permet très « simplement » de mettre à disposition des corpus organisés avec des fonctionnalités multiples. L'apparition de nouvelles exigences *wysiwyg* ou *user friendly*, la généralisation des principes FAIR posent alors la question de la relation avec les usagers / utilisateurs qu'on pousse de plus en plus à être acteurs d'une bibliothèque numérique en les impliquant *via* du « collaboratif ». Enfin, cette multiplication de corpus et de bibliothèques numériques accessibles à tous impactera forcément les hiérarchies dans les différentes histoires culturelles.

10h30-11h45 Retours d'expérience animés par ANNE RÉACH-NGÔ

CHARLOTTE DESSAINT (Bibliothèque Ulm-Jourdan, École Normale Supérieure), « Notes de cours de l'ENS : l'édition d'un corpus de notes de cours sous Omeka ? »

Numérisées en 2015, les notes de cours de l'École normale rassemblent 177 manuscrits de notes préparatoires ou prises en séance à l'occasion des enseignements de grandes figures : Fustel de Coulanges, Georges Perrot, Jules Michelet, Paul Vidal de la Blache, etc. Ce corpus pluridisciplinaire éclaire d'une façon inédite une partie de l'histoire et des disciplines du XIX^e siècle. Depuis 2016, ces manuscrits font l'objet d'une transcription et d'une édition électronique, dans le cadre du projet [Notes de cours de l'ENS](#) hébergé sur la plate-forme EMAN. Conçue au sein du projet plus vaste de bibliothèque numérique scientifique des bibliothèques de l'ENS, l'édition de ce corpus au sein d'une bibliothèque numérique créée à partir du logiciel Omeka offre l'opportunité d'expérimenter la mise en ligne d'un corpus numérique adossé à une série d'outils d'exploitation et de valorisation scientifique dans le cadre d'une communauté d'usage et d'usagers.

ELISABETH GRESLOU (Université Grenoble-Alpes, Litt&Arts), « Du scientifique au technique : l'exemple de la correspondance de Proust »

Cette présentation portera sur la méthodologie et les moyens utilisés pour matérialiser un projet scientifique ambitieux. L'exemple de l'édition numérique de la correspondance de Proust, engagé dans le cadre du site [Corr-Proust](#), servira de support à la réflexion. Il permettra de montrer comment une collaboration fructueuse entre ingénieurs et chercheurs peut aboutir à une réalisation répondant au plus près aux objectifs scientifiques.

FLORIANE CHIFFOLEAU (ingénieur recherche/développement, INRIA), « Le projet DAHN, production d'une chaîne d'édition scientifique numérique pour un corpus d'égodocuments »

Le projet [DAHN](#) (Dispositif de Soutien à l'Archivistique et aux humanités numériques, 2020-2021) a pour objectif de mettre à disposition des chercheur.e.s et des institutions patrimoniales une chaîne d'édition scientifique qui permette de valoriser des fonds d'archives dans des formats numériques facilitant leur exploitation par la recherche. L'équipe Inria se concentre principalement, pour le développement de cette chaîne, sur des corpus d'égodocuments, de diverses époques et langues. Notre cas d'étude est un corpus épistolaire datant de la Première Guerre mondiale, à l'interface entre sphère privée et sphère publique. Actuellement, la chaîne d'édition se divise en plusieurs étapes et l'intervention aura pour but d'exposer ces étapes, de présenter les outils et méthodologies utilisés, de développer les résultats déjà obtenus lors de l'application de la chaîne et enfin, de soumettre de nouvelles perspectives d'utilisation.

ALIÉNOR SAMUEL-HERVÉ (chercheuse associée à la BnF, chargée de mission Geneanet), « Des outils de média(tisa)tion pour une histoire de l'horticulture : l'exemple du projet *Graines de culture* »

Pour donner à voir une recherche au long cours sur l'histoire de l'horticulture, j'ai décidé, dès la genèse du projet [Graines de culture](#), d'y inclure un volet « communication numérique », composé, entre autres, d'un carnet de recherche *Hypothèses* (sur lequel est publié au fur

et à mesure de sa constitution un Dictionnaire biographique des médaillés des expositions horticoles de Paris), et d'un compte Instagram (qui rassemble des planches botaniques extraites de Gallica). Cette intervention se propose d'explorer les perspectives et limites de ces outils de médiation et de médiatisation d'une recherche « en train de se faire », qui en constituent en même temps les produits.

11h30 Discussions

11h45-12h45 *Speed searching*

Cette session vise à donner la parole à des jeunes chercheurs ou à des chercheurs confirmés qui débutent un projet et s'interrogent sur les choix à mener en matière d'outils et de méthodologie numériques. 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 numériques pour aborder le questionnement scientifique qui leur est propre.

ANNE RÉACH-NGÔ (Université de Haute-Alsace, ILLE) et RICHARD WALTER (ingénieur de recherche, Sorbonne nouvelle, Thalim, CNRS/ENS), Présentation du séminaire « Mon projet de recherche avec EMAN »

VÉRONIQUE LOCHERT (Université de Haute-Alsace, ILLE), « Spectatrix : un site pour comprendre la participation des spectatrices à la vie théâtrale du XVII^e siècle »

SILVIA BORASO (Université Ca'Foscari - Université Paris-Est), « Éditorialiser les citations d'un corpus de thèse sur EMAN, enjeux et possibilités »

ILARIA GIACOMETTI (Université Ca'Foscari - Sorbonne Université), « Sur les traces de Barbey d'Aureville : la réception italienne d'un écrivain hors normes »

RÉGINE BATTISTON et NIKOL DZIUB (Université de Haute-Alsace, ILLE), « Du colloque à la bibliothèque numérique : le projet *Amitiés/Inimitiés d'écrivains* »

Pause déjeuner

14h15-15h30 Atelier d'initiation aux outils numériques

Séance d'initiation à Omeka, animée par ANNE GARCIA-FERNANDEZ (ingénieure de recherche, Université Grenoble-Alpes, Litt&Arts, CNRS) et CAMILLE DESILES (ingénieure d'étude du projet COLLEX Fonte Gaia, Université Grenoble-Alpes, Litt&Arts)

Omeka est un logiciel permettant de créer et de gérer des bibliothèques numériques. Son objectif premier pourrait être défini comme la mise à disposition de fonds documentaires numérisés. On observe cependant une déformation de l'usage de l'outil, influencée par les pratiques et les besoins de recherche. Omeka est alors le support permettant seul ou en équipe d'organiser des données numériques, de les documenter, de les enrichir, de les partager avec l'équipe de recherche ou plus largement avec la communauté. « Mettre des données sur Omeka » constitue ainsi une phase de réflexion sur la donnée, sa structuration, les analyses qui en seront faites, mettant presque au second plan des aspects pourtant primordiaux de l'outil qui inclut intrinsèquement le respect des principes FAIR. Quant à l'aspect technique du travail, il est vite balayé par la facilité de prise en main de l'interface. L'atelier permettra de désacraliser l'outil en donnant des exemples d'usages et en proposant des démonstrations de certaines fonctionnalités choisies, des tout premiers pas aux opérations les plus complexes.

15h30-16h30 Projets collaboratifs en humanités numériques

Table ronde « Instances et acteurs des sciences humaines et sociales en environnement numérique », animée par MARINE PARRA

La table ronde visera à rassembler des responsables d'institutions et organisations qui interviennent à différentes échelles et suivant des modalités diverses pour accompagner la naissance et le développement de projets de recherche en environnement numérique. Dans quelle mesure un consortium, une plate-forme d'accompagnement à la publication, une Maison des Sciences de l'Homme ou un groupe d'ingénieurs dans un laboratoire peuvent-ils contribuer à nourrir la vie de la communauté scientifique et favoriser les échanges interprofessionnels au service des projets collaboratifs en humanités numériques ?

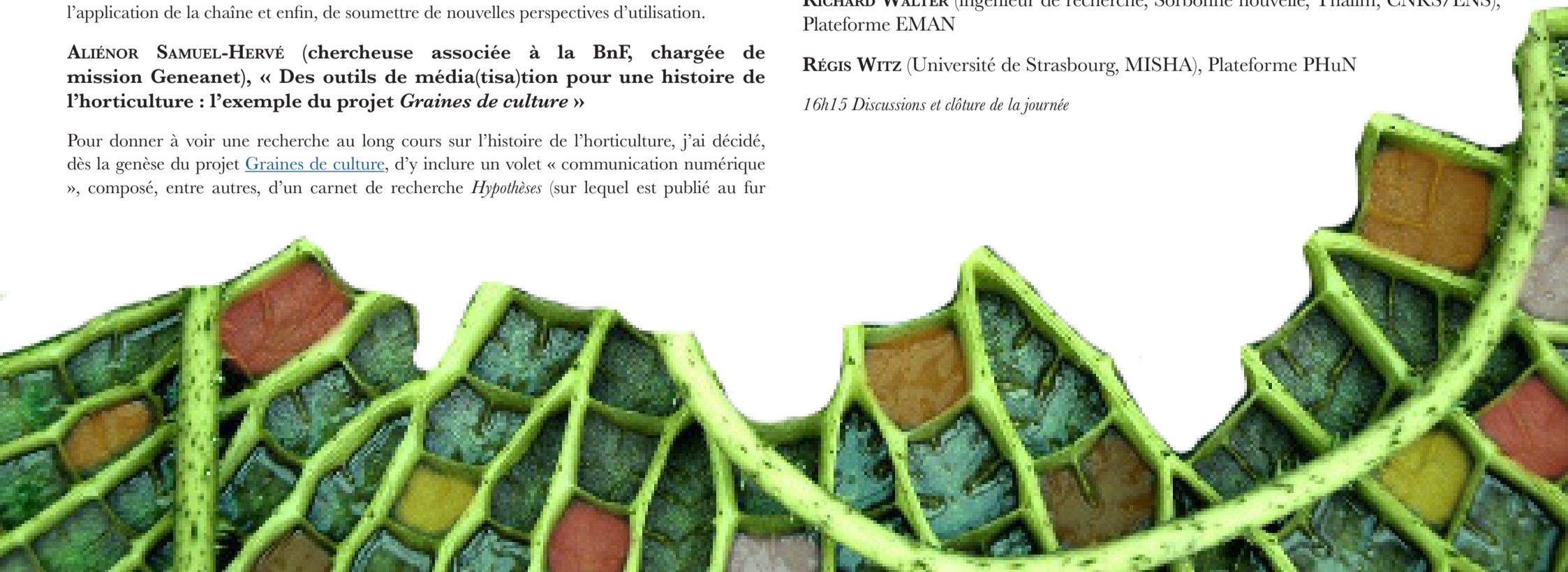
ANNE GARCIA-FERNANDEZ (ingénieure de recherche, Université Grenoble-Alpes, Litt&Arts, CNRS), Équipe ELAN

THOMAS LEBARBÉ (Université Grenoble-Alpes, Litt&Arts), Consortium CAHIER

RICHARD WALTER (ingénieur de recherche, Sorbonne nouvelle, Thalim, CNRS/ENS), Plateforme EMAN

RÉGIS WITZ (Université de Strasbourg, MISHA), Plateforme PHuN

16h15 Discussions et clôture de la journée



9h RÉGINE BATTISTON, ANNE RÉACH-NGÔ et MARINE PARRA (Université de Haute-Alsace, ILLE), Accueil des participants - SYLVIE RIVOT (vice-présidente Recherche de l'Université de Haute-Alsace), Ouverture de la journée

9h15-10h15 Conférence inaugurale

MARIE-LUCE DEMONET (Université de Tours, CESR), « Quand une édition numérique devient-elle “critique” ? »

Le site [Epistemon](#) a maintenant plus de vingt ans : quelles leçons tirer d'une expérience qui a commencé avec des transcriptions non-diplomatiques et se poursuit avec l'encodage fin des spécificités du texte, comme les différentes formes d'esperluette ? Jusqu'où aller dans la visualisation des variations (édition génétique) qui relèvent ou non de la critique textuelle ? Outre un état des lieux et une description des principes (évolutifs) qui ont guidé le développement de ce projet, l'idée est de faire le point sur quelques éditions critiques numériques actuellement disponibles, de commenter leurs choix, leur accessibilité, leur réutilisabilité par d'autres projets : bien souvent, plus l'édition est chargée d'information, moins elle est exploitable par les outils numériques et par autrui. Le renoncement à toute annotation étrangère à l'établissement du texte et à la cohérence linguistique peut décevoir, mais les formats mis à disposition, et en particulier le fichier xml/tei téléchargeable, laissent la possibilité de revenir en arrière, de simplifier autant que d'amplifier. Les questions de droits, de coût en ressources techniques et humaines — et de valorisation auprès d'un public non-universitaire — ayant un impact sur les réalisations, elles seront abordées à partir des éditions en cours de Rabelais (*Cinquième livre* manuscrit, vers 1560) et Montaigne (édition posthume des *Essais*, 1595) qui posent des problèmes d'authenticité, d'attribution et de chronologie : l'édition numérique peut tenter d'apporter des solutions.

10h30-11h30 Retours d'expérience animés par MARINE PARRA

SANDRA PROVINI (Université de Rouen-Normandie, CÉRÉDI), « L'édition critique numérique des Œuvres complètes de Michel d'Amboise : point d'étape »

Le site internet « Michel d'Amboise », volet numérique d'un projet d'édition critique des Œuvres complètes du poète à paraître aux éditions Honoré Champion, présente à ce jour quatre recueils encodés au format XML-TEI. Face aux vastes possibilités offertes par ce format, les éditeurs scientifiques ont fixé des limites — qui méritent d'être interrogées — à la finesse de l'encodage en fonction de plusieurs objectifs : affichage des textes dans une transcription diplomatique ainsi que dans une version corrigée et régularisée, mise en évidence du travail de réécriture et de recomposition effectué par l'auteur dans certains recueils, visualisation des phénomènes d'amplification dont Michel d'Amboise fait fréquemment usage dans ses traductions du latin, du néolatin et de l'italien.

PABLO RUIZ FABO (Université de Strasbourg, LiLPa), « DISCO et DISCOOver : Le Diachronic Spanish Sonnet Corpus et son interface de lecture circulaire »

Le corpus [DISCO](#) veut répondre au manque de ressources numériques pour les études quantitatives sur l'évolution du sonnet en espagnol. Il comporte 4 085 sonnets en TEI, par 1 204 auteurs d'Espagne et d'Amérique Latine, entre le XV^e et le XIX^e siècle. Les variables sociales des auteurs (dates, origine, sexe) sont encodées, ainsi qu'une annotation automatique des rimes, métrique et enjambement. Le corpus permet d'examiner la distribution de ces caractéristiques selon les périodes et régions. Son interface, [DISCOOver](#), permet de créer des sous-corpus, fournissant des données agrégées sur les rimes, les schémas métriques et rimiques et l'enjambement. À partir des résultats quantitatifs nous pouvons naviguer vers chaque poème et vice-versa, dans une lecture « circulaire ». Ceci aide à situer les caractéristiques d'un poème dans la production de son auteur ou de son époque et à évaluer l'originalité ou typicalité du texte.

EULALIE MONGET (Université de Strasbourg, LiLPa ; Junior Researcher au FIZ Karlsruhe — Leibniz Institute for Information Infrastructure), « L'apport du numérique aux études littéraires : le cas de la détection automatique de l'enjambement »

L'annotation est une tâche coûteuse et répétitive. Elle est pourtant nécessaire pour de nombreux travaux, allant de l'entraînement de systèmes à base d'intelligence artificielle à l'enrichissement automatique et semi-automatique de textes littéraires. Il s'agira ici de présenter un outil automatique d'annotation, qui repose sur des critères largement syntaxique, de l'enjambement dans la poésie anglaise. Cet outil a le potentiel d'alléger la charge de travail des annotateurs en réalisant, rapidement et à grande échelle, une première annotation des enjambements. Une étude de cas réalisée sur 632 poèmes des XVIII^e, XIX^e et XX^e siècles a montré que les résultats sont cohérents avec la littérature autour de l'enjambement, ouvrant la voie à son utilisation pour diverses tâches d'annotation. La typologie utilisée et ses limites seront également abordées.

11h15 Discussions

11h30-12h15 Speed searching

Cette session vise à donner la parole à de jeunes chercheurs ou à des chercheurs confirmés qui débutent un projet et s'interrogent sur les choix à mener en matière d'outils et de méthodologie numériques. 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 numériques pour aborder le questionnement scientifique qui leur est propre.

CAROLINE DORLÉANS (Université de Tours), « Élaborer sa culture numérique en Licence de Lettres en vue d'un master, quel impact sur l'élaboration du projet ? »

MICHELA LAGNENA (Université Ca'Foscari - Sorbonne-Nouvelle), « Du livre à l'écran : initiation à l'édition critique numérique des *Lettres amoureuses* d'Étienne Pasquier »

PHUONG QUYEN TRAN THI (Sorbonne-Nouvelle), « Fouille d'opinion et recherche de l'intelligence émotionnelle : étude d'un corpus linguistique du Cahier de doléances à l'ère numérique, à travers la textométrie »

MERYAM TALHAOUI (Sorbonne-Nouvelle), « Modélisation de la politique éditoriale des planches graphiques du projet d'Édition Numérique CRitique et Collaborative de l'Encyclopédie de Diderot et d'Alembert (ENCCRE) »

Pause déjeuner

13h45-15h Atelier d'initiation aux outils numériques

Séance d'initiation à l'édition critique en environnement numérique animée par PASCAL JOUBAUD (Ingénieur, projet CorNum)

À partir de l'exemple d'un projet en cours, celui de l'édition des Œuvres complètes d'un poète de la première moitié du XVI^e siècle, Michel d'Amboise (dir. S. Provini), il s'agira de présenter la démarche scientifique et les outils adoptés pour construire cette édition numérique XML-TEI. En partant des problèmes spécifiques posés par l'œuvre : multiplicité des formes poétiques, réécriture en profondeur de certains textes, traductions de textes latins et italiens —, des choix effectués par les éditeurs critiques et des fonctionnalités souhaitées pour l'édition en ligne, nous verrons quelles modalités d'encodage ont été choisies et quels outils ont été développés pour le site internet.

15h15-16h30 Projets collaboratifs en humanités numériques

Table ronde « Éditer et enseigner », animée par RÉGINE BATTISTON

L'édition critique numérique constitue une pratique propice à l'enseignement de la littérature, de même qu'elle offre un cadre favorable à l'expérimentation problématisée du traitement des données textuelles et iconographiques. La table-ronde sera consacrée à des projets engagés cette année par des enseignants-chercheurs et ingénieurs qui désirent initier leurs étudiants à l'acte d'éditorialisation en contexte numérique.

CLAIRE SICARD (Université de Tours, CESR), *Fleur des Tanneurs*

Fleur des Tanneurs est un carnet *Hypothèses*, support du travail collectif d'étudiants de L3 de l'Université de Tours. Il prend la forme d'une anthologie d'épigrammes du XVI^e siècle, pensée pour favoriser l'exploitation pédagogique des textes recueillis, en particulier dans la classe de français de 2^{de}, où la poésie du Moyen Âge et de la Renaissance est désormais au programme.

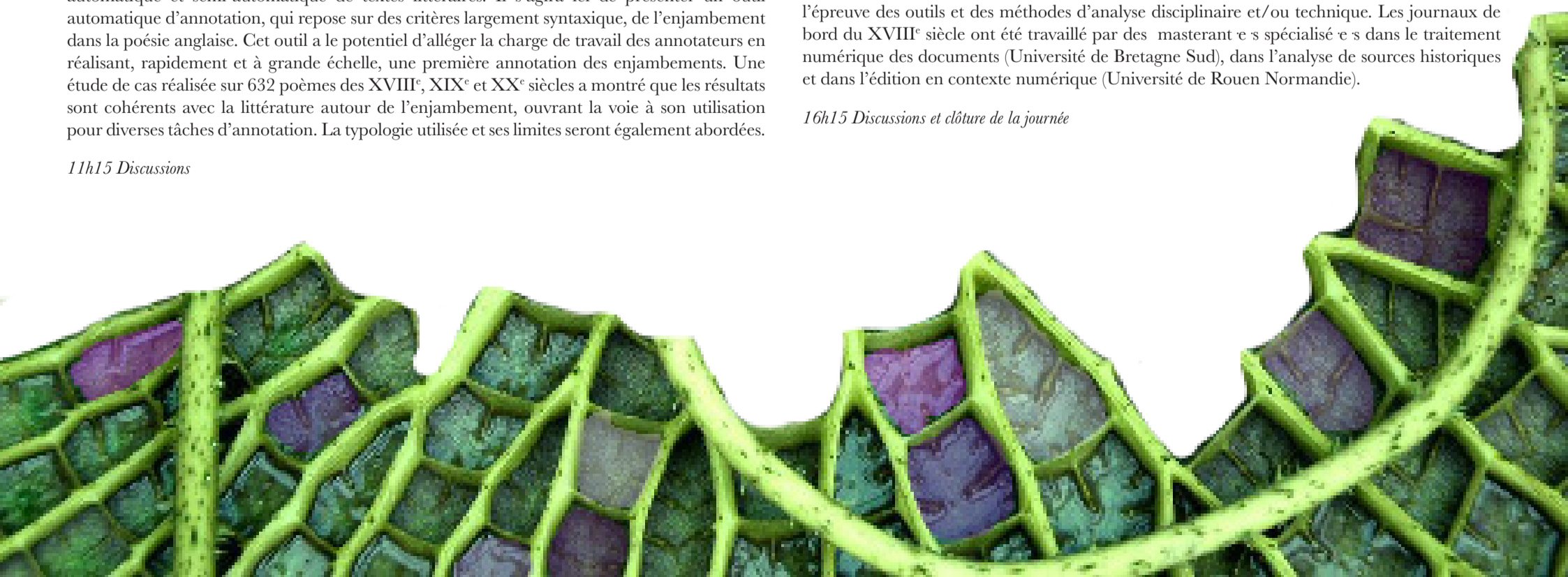
MAGDA CAMPANINI (Université Ca' Foscari) et ANNE RÉACH-NGÔ (Université de Haute-Alsace, ILLE), *Tragiques Inventions*

Tragiques Inventions est une bibliothèque numérique EMAN conçue et réalisée par des masterants de Lettres et d'Ingénierie documentaire. Elle vise à interroger la médiation numérique des récits tragiques de la Première Modernité. Née d'une collaboration avec l'université Ca'Foscari, elle est désormais ouverte à d'autres enseignants et étudiants français qui s'intéressent au geste d'éditorialisation en environnement numérique.

BENOÎT ROUX (Université de Rouen Normandie, ERIAC) ET MARINE PARRA (Université Bretagne Sud, TEMOS), *ÆTeiCab*

ÆTeiCab (Archives éditorialisées en TEI de CArnets de Bord) est une expérience d'éditorialisation collaborative et interdisciplinaire de pièces d'archives courtes. Ce projet permet aux participants de se confronter à des sources souvent méconnues et de mettre à l'épreuve des outils et des méthodes d'analyse disciplinaire et/ou technique. Les journaux de bord du XVIII^e siècle ont été travaillés par des masterants spécialisés dans le traitement numérique des documents (Université de Bretagne Sud), dans l'analyse de sources historiques et dans l'édition en contexte numérique (Université de Rouen Normandie).

16h15 Discussions et clôture de la journée



25 JUN 2021 JOURNÉE 4. ARCHIVES
ET FONDS PATRIMONIAUX NUMÉRISÉS

9h RÉGINE BATTISTON, ANNE RÉACH-NGÔ et MARINE PARRA (Université de Haute-Alsace, ILLE), Ouverture de la journée

9h15 Réflexions inaugurales

FRÉDÉRIC CLAVERT (Université du Luxembourg, Centre for Contemporary and Digital History) et CAROLINE MULLER (Université Rennes 2, Tempora), « Une nouvelle relation aux archives ? Quelques réflexions autour du *Goût de l'archive à l'ère numérique* »

En 1989, Arlette Farge publiait *Le goût de l'archive*. Traduit en plusieurs langues, ce « petit livre », selon ses propres dires, a eu un succès constant depuis. Investigant la relation matérielle, intime de l'historien.ne à l'archive, Arlette Farge posait, finalement, la question des conditions de l'interprétation des archives. Le livre en ligne *Le goût de l'archive à l'ère numérique* se propose d'investiguer à nouveau cette question des conditions matérielles de l'interprétation des archives dans le contexte nouveau d'un monde de données. Il documente la relation de l'historien.ne à ses sources primaires telle qu'elle s'est transformée depuis l'émergence d'internet, des politiques massives de numérisation et de l'apparition de données nativement numériques. En résulte une série de réflexions sur les pratiques numériques historiennes que cette présentation exposera.

10h00-11h30 Retours d'expérience animés par MARINE PARRA

LAURENCE MACÉ (Université de Rouen, CÉRÉDI/IRIHS), « Les archives de la censure à travers le prisme du numérique dans le projet ECuMe »

Le projet *ECuMe* (Edition Censure Manuscrit) vise à constituer un prototype de bibliothèque numérique d'archives de la censure pour l'édition génétique des textes d'Ancien Régime. À partir d'un corpus littéraire et patrimonial (manuscrits soumis à la censure ; éditions imprimées autorisées et clandestines ; archives de la censure et de la police ; témoignages issus de correspondances et sources périodiques etc.), il s'agit de reconstituer le processus de censure et de mesurer son incidence dans la genèse des textes d'Ancien Régime, à travers le manuscrit conçu comme lieu d'une transaction entre auteur et autres instances : imprimeur, censeur(s), voire (dans le cas particulier des manuscrits de la Comédie Française) comédiens-acteurs. *ECuMe* est un projet prototypique développé sur la plateforme EMAN (CMS Omeka) autour d'un objet patrimonial (ou de son fantôme s'il est perdu), le manuscrit soumis à la censure, transcrit comme les éditions et tous les documents qui, avec elles constituent une série.

ARIANNA SFORZINI (ingénieure de recherche du Projet ANR FFL, CNRS/ENS Lyon, Triangle), « Le projet Foucault *Fiches de lecture*. Repenser les archives d'un philosophe à l'âge du numérique »

La description du fonds Michel Foucault conservé à la BnF et des modalités de sa constitution met en évidence la spécificité du corpus des fiches de lecture foucauldienne et les défis non seulement techniques mais aussi théoriques que l'exploitation numérique de ces fiches pose aux chercheurs. Elles constituent en effet un exemple sans doute inédit d'archives de travail d'un philosophe, qui redoublent et mobilisent la réflexion que Foucault a lui-même menée sur les archives de la pensée. Le projet *Foucault Fiches de Lecture / Foucault's Reading Notes* invite à se demander quels outils et quelles fonctions numériques pourraient être les plus utiles aux chercheurs à l'avenir ?

ISEUT DE KERNIER et PIERRE EMMANUEL GUILLERAY (Conservateurs, bibliothèques de Metz et de Nancy), « *Limédia*, une bibliothèque numérique développée en réseau »

Limédia est la bibliothèque numérique du Sillon Lorrain. Elle rassemble actuellement quatre villes (Thionville, Metz, Nancy et Épinal) et comprend trois sites distincts (Galeries, Kiosque et Mosaïque). *Limédia Galeries* accueille les collections patrimoniales numérisées, avec des expositions virtuelles. *Limédia Kiosque* abrite la presse régionale ancienne, abordée sous l'angle de l'histoire de la presse. *Limédia Mosaïque* offre des ressources numériques payantes aux abonnés des médiathèques partenaires, avec plusieurs sélections thématiques chaque semaine. Le point commun de ces trois sites est l'important travail de médiation réalisé sur chacun des sites et sur les réseaux sociaux, mais aussi le développement du logiciel qui est piloté en réseau par les bibliothèques elles-mêmes.

11h15 Discussions

11h45-12h15 *Speed searching*

Cette session vise à donner la parole à de jeunes chercheurs ou à des chercheurs confirmés qui débute un projet et s'interrogent sur les choix à mener en matière d'outils et de méthodologie numériques. 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 numériques pour aborder le questionnement scientifique qui leur est propre.

NINA PÖSCH (Université de Haute-Alsace, CRESAT), « Constitution de sources sur les archives révolutionnaires avec Citavi, une étape préparatoire à la rédaction de la thèse »

CLÉMENT FONTANNAZ (Université de Strasbourg - Université de Lausanne), « Étudier la revue *Le Phare* numérisée sur Gallica. Perspective pour une étude biographique de Jules Humbert-Droz »

Pause déjeuner

14h-15h15 Atelier d'initiation aux outils numériques

Séance d'initiation à Tropy, animée par BENOÎT ROUX (Université de Rouen Normandie, ÉRIAC)

Jusqu'à récemment, il n'existait pas de logiciel *ad hoc* permettant aux chercheurs en sciences humaines de gérer leurs photographies d'archives. Le développement de Tropy par le *Center for History and New Media* (George Mason University) est venu combler ce vide en 2017, et ce, tout en réaffirmant la nécessité pour la communauté scientifique d'élaborer ses propres outils. Cette solution open source et multiplateforme simplifie le classement des clichés, leur association à des métadonnées, leur transcription et leur annotation. L'atelier proposera une découverte, à partir d'exemples concrets, de l'interface et de ses fonctionnalités, des premiers pas jusqu'à l'exploitation des données.

15h15-16h30 Projets collaboratifs en humanités numériques

Table ronde « Corpus d'auteur et exploitation des archives numérisées », animée par RÉGINE BATTISTON

La numérisation des archives et des collections patrimoniales, engagée maintenant depuis plusieurs décennies, a transformé les pratiques de conservation, de communication et de valorisation des fonds. Si l'enjeu de l'accessibilité est bien au cœur des politiques de numérisation, il convient de mesurer également l'impact de ces pratiques dans la redéfinition même des fonds, dans leur identité comme dans la construction des savoirs et représentations auxquels ils participent. La table ronde invitera quelques représentants de fonds d'archives à examiner ce lien entre archives et figures d'auteur à l'aune d'un tel phénomène.

KATRIN BEDENIG (directrice des archives Thomas Mann à la bibliothèque de l'École Polytechnique de Zurich), « On the digitization of the holdings of the Thomas Mann Archive at ETH Zurich »

The entire manuscript holdings, press holdings, photo holdings and the author library have been recorded and digitized in 3 projects and are available for metadata research via online databases. Where copyright permits, the digital copies are also available online. In the case of current copyrights, consultation is possible in the reading room.

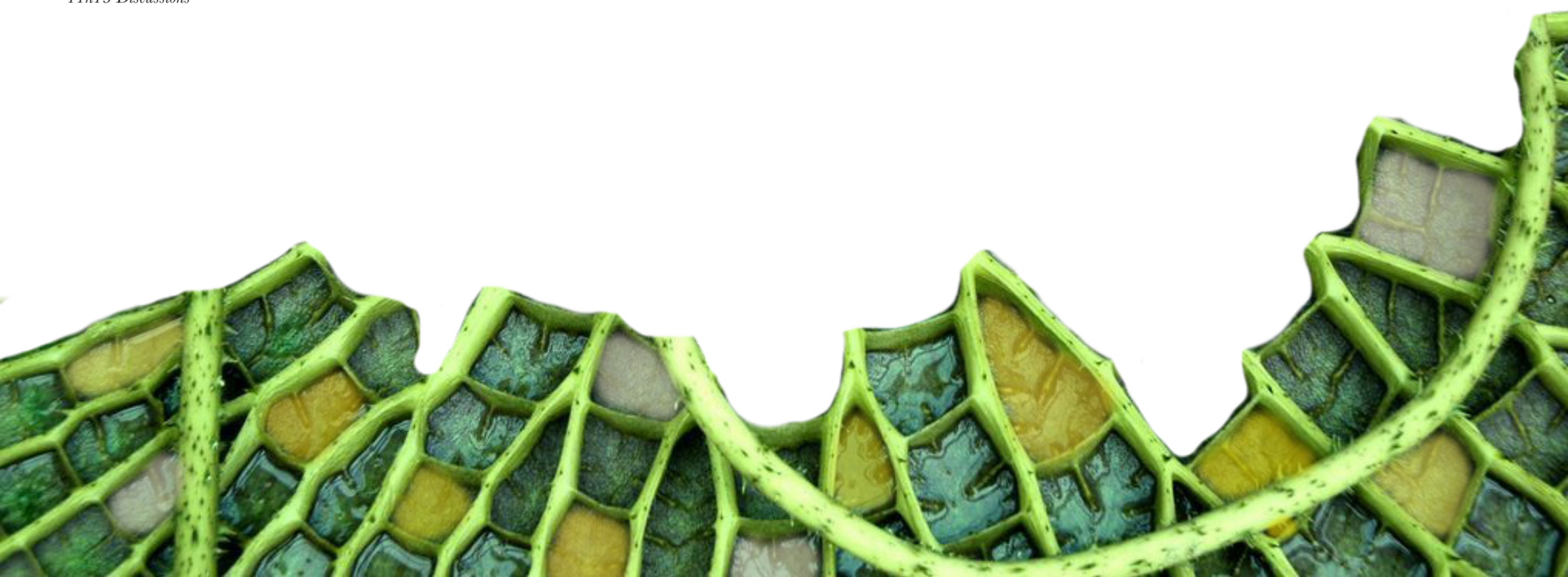
TOBIAS AMSLINGER (directeur des archives Max Frisch à la bibliothèque de l'École Polytechnique de Zurich), « Digitisation in the Max Frisch Archive »

Since 2014, the holdings of the Max Frisch Archive have been digitised. For legal reasons, the digital copies can only be viewed in the reading room. In future, selected documents will be published as digital editions.

AMBRE PHILIPPE (directrice de la fondation Catherine Gide), « Les archives d'André Gide »

La fondation Catherine Gide travaille à la sauvegarde et la mise en valeur du patrimoine d'André Gide, elle soutient la recherche en littérature et les projets interdisciplinaires autour de l'écrivain et son univers. Elle a été fondée en 2007 par la fille du Prix Nobel. Depuis 2016, elle met progressivement en ligne et à disposition du public ses archives.

16h15 Discussions et clôture de la journée



28 MAI 2021 JOURNÉE 5. VISUALISATION DES RÉSULTATS DE LA
RECHERCHE, GRAPHS ET SCHÉMAS DE DONNÉES

9h RÉGINE BATTISTON, ANNE RÉACH-NGÔ et MARINE PARRA (Université de Haute-Alsace, ILLE), Ouverture de la journée

9h15-10h15 Conférence inaugurale

PIERRE CUBAUD (CNAM, CEDRIC), « Visualisation pour les bibliothèques numérisées »

Portée par les progrès de l'informatique graphique, la visualisation est devenue un outil incontournable pour l'analyse et le suivi de grands volumes de données. L'application de ces procédés aux grands corpus documentaires est ancienne mais son usage systématique est assez récent, et surtout restreint aux outils d'analyse de contenu des collections. Nous pensons — avec d'autres — que la visualisation peut avoir un rôle plus important encore dans les interfaces, pour un intermédiaire encore à inventer entre les outils de recherche et les outils de lecture. La visualisation peut ainsi apporter une réponse aux besoins de déambulation et d'appropriation rapide des collections numérisées. En appui de notre thèse, nous présenterons quelques travaux réalisés au CEDRIC depuis une trentaine d'année pour divers projets collaboratifs issus des sites ABU et Cnum : des interfaces 3D de déambulation, un mur de pages pour le contrôle qualité dans une usine de numérisation et un espace de conception pour les corpus de correspondances.

10h30-11h45 Retours d'expérience animés par ANNE RÉACH-NGÔ

LAURENT CAPRON (Centre Jean Pépin, CNRS), « Cartographie des philosophes antiques : voir moins, c'est voir mieux ! »

L'objectif premier du projet « [Cartographie des philosophes antiques](#) » était d'offrir une visualisation spatiale des philosophes de l'Antiquité qui montrerait la vacuité de la représentation inconsciente et commune des philosophes concentrées dans les villes d'Athènes puis de Rome, et mettrait en valeur la riche diversité des provenances et des centres culturels antiques. Mais il a rapidement fallu faire face à la maigreur des données disponibles et fiables en comparaison du nombre de noms de philosophes connus. La structuration des données selon des critères scientifiques bien définis, intégrant la part d'incertitude quant aux informations disponibles, nous a obligés à choisir certains modes de visualisation plutôt que d'autres. Plutôt que de pouvoir tout visualiser, nous avons préféré mieux montrer la réalité de nos connaissances.

THOMAS MOHNIKE (Université de Strasbourg, Mondes germaniques et nord-européens) et LUDOVIC STRAPPAZON (Université de Strasbourg, DNU), « Cartographier le savoir narratif – approches numériques aux mythes du Nord en circulation culturelle »

Le projet [Mythes du Nord](#) cherche à identifier les unités du savoir narratifs (mythes) dans des textes narratifs traitant ou venant du Nord de l'Europe et au-delà avec des approches computationnelles et à cartographier leur grammaire discursive. Nous utilisons des logiciels d'analyse développés ou déployés dans notre « laboratoire des mythes » et des logiciels de visualisation comme Gephi pour créer des cartes des univers narratifs, les horizons d'attente (Jauss) et les changements de la grammaire du savoir social dans le temps.

ALEXANDRE ZEITLER (Université de Strasbourg, Mondes germaniques et nord-européens), « Cartographier les réseaux du savoir dans le Finnmark norvégien : méthode(s) et enjeux »

À partir des savoirs écologiques locaux recueillis par les autorités norvégiennes pour déterminer les droits de propriété et d'usage sur les îles de Stjernøya et de Seiland en Norvège, j'esquisserai dans un premier temps un cadre méthodologique pour cartographier les réseaux de production du savoir. Pour ce faire, mon étude prendra appui sur les outils des humanités numériques : de la cartographie GIS à la visualisation Gephi en passant par les outils de *text mining*. Les premiers résultats obtenus avec cette méthode d'analyse pour le corpus primaire laissent entrevoir trois clusters de savoirs (le renne, l'activité minière et la pêche) produits par les communautés des deux îles pour construire leurs identités. Afin de donner une vue d'ensemble du potentiel de cette méthode, j'exposerai dans un second temps les résultats obtenus à partir d'un corpus secondaire : les déclarations du Conseil sámi.

LAURENT MILLET-LACOMBE (MetaindeX), « MetaindeX : statistiques d'un corpus avec Kibana »

L'outil open-source [MetaindeX](#) propose un environnement de travail en ligne avec les services de base que pourraient attendre une institution ou un chercheur pour le catalogage, la consultation et l'étude de ses données. Il se base sur ElasticSearch pour le stockage et l'interrogation des données, et intègre l'outil Kibana pour la création de graphiques statistiques complexes que l'utilisateur configure lui-même. La présentation portera sur la manière d'utiliser Kibana au travers de quelques exemples simples et didactiques, qui s'appuient sur un corpus de démonstration importé via MetaindeX.

11h30 Discussions

11h45-12h30 *Speed searching*

Cette session vise à donner la parole à de jeunes chercheurs ou à des chercheurs confirmés qui débutent un projet et s'interrogent sur les choix à mener en matière d'outils et de méthodologie numériques. 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 numériques pour aborder le questionnement scientifique qui leur est propre.

MARIE-LOU SOLBACH (Université de Haute-Alsace, ILLE), « De la visualisation à la rédaction de thèse : cartographie d'enquête(s) au sein d'un corpus de polars polaires »

CLAIRE DAVRIL (Université de Strasbourg, ACCRA), « Une réflexion par l'image. Propos parallèle sans nombre de signes ? »

GEOFFREY KOENIG (Université de Strasbourg, Service historique de La Défense), « L'idéologie nazie et sa langue. La lexicométrie comme outil d'exploration de la *Lingua Tertii Imperii* »

Pause déjeuner

13h45-15h Atelier d'initiation aux outils numériques

Séance d'initiation à l'analyse de réseaux, animée par PABLO RUIZ FABO (Université de Strasbourg, LiLPa)

Les réseaux constituent une structure pratique pour représenter des données relationnelles de façon à donner un aperçu de groupes et interactions. Dans l'atelier nous regarderons quelques caractéristiques des réseaux ainsi que des outils qui permettent de les tracer et analyser, comme Cytoscape ou Gephi. Une application des réseaux à l'analyse textuelle avec l'outil CorText Manager sera également montrée.

15h15-16h30 Projets collaboratifs en humanités numériques

Table ronde « Genèse et visualisation », animée par ANNE RÉACH-NGÔ

En mettant au cœur de l'étude le processus plus que le produit fini et le geste créateur plus encore que l'effet sur le lecteur, l'analyse des phénomènes de genèse littéraire pose pleinement la question des modes de restitution des résultats de la recherche. Les méthodes et outils qui relèvent de la visualisation constituent à cet égard une entrée éclairante pour interroger la publication de ces enquêtes et la manière dont elles viennent renouveler le regard sur une œuvre, un corpus, un auteur, un fait artistique.

MATHILDE LABBÉ (Universités de Nantes, L'AMo), *Éditer L'Étoile du Sud*

Le projet « [Éditer L'Étoile du Sud](#) » se concentre sur un objet dont la genèse comporte des zones d'ombre : si l'on dispose de l'avant-texte rédigé par Paschal Grousset pour ce roman édité sous le nom de Jules Verne, le travail de ce dernier sur le texte du premier n'est accessible par aucun document connu. Il s'agit ici d'étudier un cas de pluriauctorialité contrainte et, par la suite, de rendre visible la complexité de ce palimpseste.

FRANZ JOHANSSON (Sorbonne Université, ITEM), *Robinson de Paul Valéry : édition génétique*

[Robinson de Paul Valéry : édition génétique](#) se propose de reconstituer le dossier de genèse d'un écrit demeuré, comme bien d'autres projets chez son auteur, inachevé : le conte Robinson. Elle aspire également à déterminer la cohérence des différents éléments et le jeu complexe de leurs interactions.

HAKIM USOOF (University of Peradeniya, Sri Lanka, Dept. of Statistics & Computer Science), CHRISTOPHE LEBLAY (Université de Turku, Finlande, Département des langues et Traduction) ET MINJING ZHONG (Université de Xiangtan, Chine, Département de français), *GenoGraphiX*

Le programme [GenoGraphiX](#) a été développé à partir de la génétique textuelle, de la théorie mathématique des graphes et des sciences de l'informatique. Ce programme met l'accent sur l'utilisation de la visualisation des données scripturales afin de rendre compréhensible l'abondance et la complexité de ces données pour l'apprentissage et l'enseignement de l'écriture ainsi que pour la recherche.

16h15 Discussions et clôture de la journée

