Illustration d'après *Loop*, d'Aldo Frullini.

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
Carnet du projet EVEille
Chaîne vidéo sur Uha pod

@Eveille HN
@eveille_hn
@Eveille_HN

ILLE Institut de recherche en Langues et Littératures Européennes UNIVERSITÉ HAUTE-ALSACE
MULHOUSE ALSACE AGGLOMÉRATION
UNIVERSITÉ HAUTE-ALSACE
EVEille Exploration & Valorisation Electroniques de corpus en SHS
NOVATRIS Centre de compétences transfrontalières UNIVERSITÉ HAUTE-ALSACE

22 janvier 2021 Journée 1. Base de données

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 XVIe-XVIIe 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 XVIe et XVIIe 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 XVIIe siècle : naviguer dans un corpus massif à l'aide d'une base de données »

Au début du XXe 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 XVIIe 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 XVIIe 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'émojis »

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

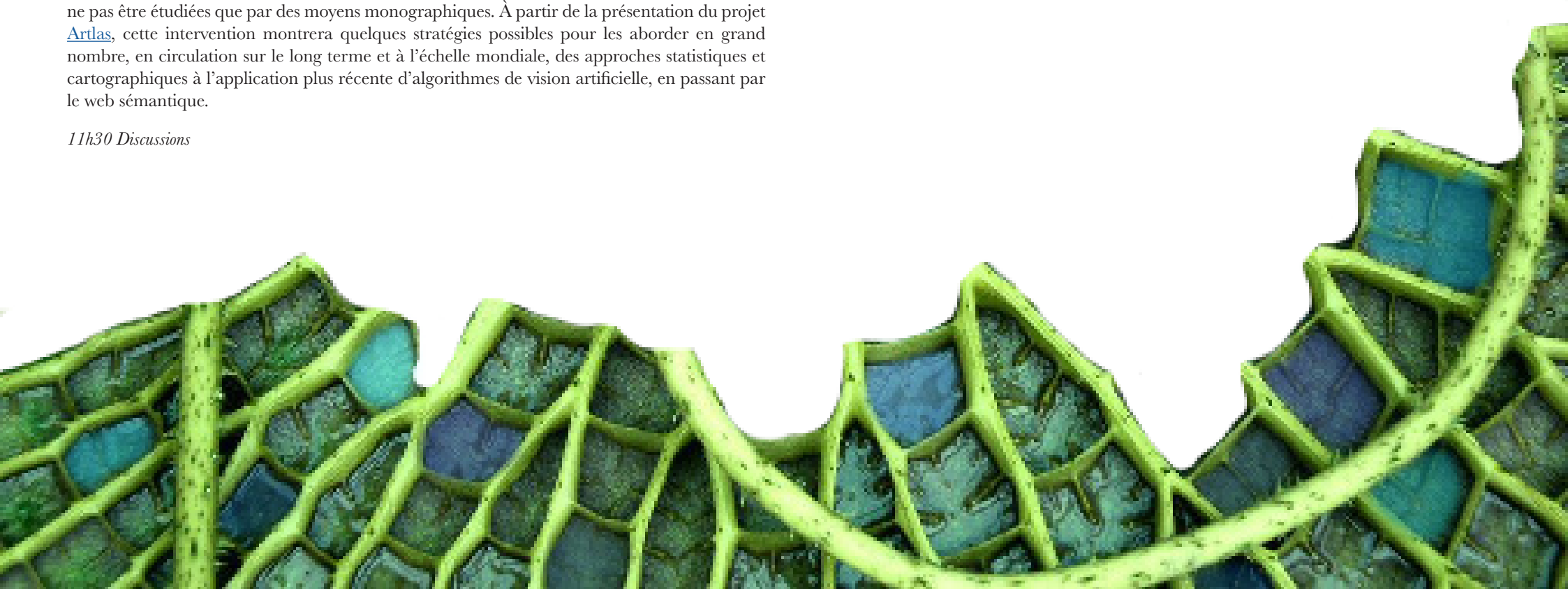
15h15-15h30 Anne Réach-Ngô (Université de Haute-Alsace, ILLE), Présentation du projet Réson@nces Grand-Est

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 siences 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 médiévales.

16h15 Discussions et clôture de la journée