**RENCONTRES DE L’ÉCOLE DOCTORALE D’HISTOIRE**

### DE L’UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE

### Doctorantes et doctorants face à leurs données

### Collecte, traitement, valorisation

**Le samedi 27 avril 2024, de 14h à 17h**

**Salle Marc Bloch (17 rue de la Sorbonne, 75005 Paris)**

### Coordinatrice : Romane Carballo

### Grégory Spadacini (ANHIMA), *Tropy, une super database pour l’historien ?*

Tropy est une base de données gratuite développée en 2016 par et pour les historiens du *Center for History and New Media* de l’Université George Mason. Conçu pour aider les scientifiques dans la gestion des fonds d’archives, le logiciel permet de catégoriser des données brutes selon des critères spécifiques, rendant l’accessibilité et l’exploitation plus efficientes. L’objectif est d’alléger l’identification des interactions, telles que les liens de causalité, les proportions et les tendances chronologiques. Cette nouvelle catégorie de logiciels vise à simplifier l’organisation et la conception des connaissances, au-delà du seul stockage. C’est le travail de l’historien qui s’en trouve amélioré et étendu. La facilité d’usage doit conduire à une réflexion sur le redimensionnement des thèmes et des questions traités par l’historien. Et conjointement, à une réflexion sur le rapport aux sources. Si le logiciel a été pensé principalement à destination d’historiens modernistes et contemporanéistes, voire d’archéologues et de numismates, son utilisation semble pouvoir s’élargir au-delà de ces champs et servir à gérer des données littéraires. Moyennant quelques contournements, l’algorithme de Tropy supporte très bien ce type d’information et peut être employé plus largement encore. C’est ce qui encourage à le définir comme une *super-database.*

### Romane Carballo (CHS), *La carte et le désespoir : manipuler des données spatiales avec ou sans QGIS*

Cette communication repose sur une expérience personnelle de confrontation avec QGIS, un logiciel libre de manipulation de données spatiales et de construction des cartes et de Systèmes d’Informations Géographiques (SIG) très utilisé en sciences sociales. D’une grande utilité, le logiciel peut s’avérer difficile à prendre en main et surtout trop complexe pour certains usages. Il s’agira donc de présenter les principaux apports de QGIS mais aussi des alternatives moins lourdes comme Khartis et le Géoportail. On cherchera également à démontrer l’intérêt du recours à la dimension spatiale dans notre rapport aux données. Ces logiciels sont ainsi un moyen de visualiser la répartition spatiale de nos données par le biais de cartes assez simples à produire, pourvu que les données soient formatées correctement. Ils permettent également de géolocaliser certains objets, en se basant sur des cartes existantes ou sur des données contenues dans nos archives, et même de construire un Système d’Information Géographique, base de données regroupant de nombreuses couches d’informations spatiales. Une fois le SIG construit et/ou ses données géolocalisées, c’est enfin un moyen de mesurer des distances entre des objets spatiaux, de mesurer des surfaces…

### Jean-Paul Gagey (SIRICE), *Visualiser les migrations vers l’Extrême-Orient russe (1897-1926)*

Dans le cadre d’un travail de thèse consacré à une histoire spatiale de la gestion des migrations internes vers l’Extrême-Orient russe entre xixe et xxe siècles, les publications des résultats des premiers recensements conduits dans l’empire tsariste (1897) puis en URSS (1926) sont des sources primordiales. Ces fastidieuses tables listant consciencieusement jusqu’au plus petit foyer de peuplement sont toutefois difficilement exploitables, sinon même lisibles, en l’état. Toutefois, grâce au logiciel *Abbyy FineReader*, leur transformation dans un format propre à leur exploitation numérisée est rendue plus aisée, ouvrant la voie à des traitements quantitatifs. Il devient alors envisageable de produire des visualisations statistiques (diagrammes de Sankey) ou cartographiques (avec le logiciel QGIS) dont l’utilité dépend du type d’étude migratoire retenu, entre modèle convergent ou divergent. Loin de se suffire à elles-mêmes, ces représentations conduisent tant à une meilleure appropriation de l’espace étudié qu’à susciter de nouvelles questions en permettant de localiser les lieux où se nouent les interactions entre espace, populations, pouvoirs et technologies. La combinaison des logiciels et des méthodologies nécessaires aux traitements informatisés relève ainsi des bricolages et braconnages qui font l’atelier de l’historien et dont la présentation constituera le premier apport de cette communication au séminaire de l’école doctorale. Par contraste avec l’exploitation numérique qui s’offre désormais à lui, ce même historien en vient en retour à se demander quelle utilisation pouvait effectivement être tirée par les contemporains de ces volumineuses publications démographiques, amenant par conséquent à une interrogation sur les processus cognitifs et perceptifs à l’œuvre dans la production de l’espace colonial. À ce titre, le second apport de cette communication sera de proposer, outre l’aspect purement technique du traitement des données, une réflexion sur l’heuristique de la carte, laquelle apparaît avant tout comme un outil suscitant des questions plutôt que d’être toujours une façon d’y répondre.

### Antonin Gay-Dupuy (CHS), *Les recensements en bidonville : organiser et traiter une source hétérogène (1950-1996)*

La résorption des bidonvilles en France métropolitaine engage un nombre important d’administrations à de multiples échelles. Celles-ci s’engagent alors dans des recensements nombreux des populations résidentes, tant pour évaluer la population à reloger que pour mieux l’appréhender. En tout, j’ai pu relever 174 recensements nominatifs de plus d’une centaine de bidonvilles pour près de 35 000 individus adultes et enfants, répartis sur l’ensemble de la période et les quatre territoires étudiés. Les pratiques sont très hétérogènes et si certaines ne sont jamais recensées, d’autres le sont plusieurs fois. La constitution du corpus, riche mais très hétérogène, m’a mené à plusieurs questionnements. En premier lieu, la nature des sources explique le refus de l’échantillonnage d’un côté et l’impossibilité de centraliser l’ensemble de la donnée en une seule base. Cette double contrainte limite les possibilités d’une histoire quantitative des populations résidentes et nécessite l’emploi d’autres données qualitatives, relativement rares comme des photographies, des rapports ou encore des témoignages.

### Rémy Péru-Dumesnil (IHMC), *Identifier les invisibles en Guyane : homogénéiser une population par la généalogie et le recoupement des sources*

Mes travaux sur l’accueil des sinistrés martiniquais arrivés en Guyane consécutivement à l’éruption de la montagne Pelée du 8 mai 1902 m’ont amené à constituer une base de données de 233 individus à partir d’archives administratives. Cependant, force a été de constater que leur recensement en a invisibilisé un grand nombre : noms et dates de naissance absents compromettaient l’homogénéité de la population. Pour pallier ce manque d’unité, la généalogie fournit des méthodes de recoupement des sources et de déduction des données en croisant l’état civil, les registres matricules militaires ou encore les actes notariaux. Toutefois, notons que les particularités du contexte colonial antillo-guyanais limitent la reproductibilité de la démarche. La problématique sera, justement, de déterminer son champ d’efficacité afin de la conceptualiser plus finement. Pour explorer cette question, un autre cas sera abordé : la redécouverte de Lionel Sugat (1855-1943), compositeur guyanais, dont les 250 œuvres pour piano furent retrouvées aux Archives Territoriales de la Guyane en 2020. Les méthodes de généalogies appliquées au contenu des partitions du compositeur mirent à jour ses cercles familiaux, sociaux et professionnels, révélant en partie les grandes lignes de son parcours. Inévitablement, l’exploitation de ces données dans le cadre de commémorations publiques ou pour l’érection de monuments interroge l’interaction entre les sphères universitaire et de la société civile. Dans quel contexte, à quel dessin et par quelles institutions peuvent-elles être exploitées ? Quels sont les lieux d’accès les plus désignés pour les restituer aux lecteurs ? Le rôle des archives et de l’université seront les points de départ de cette réflexion, qui fera office d’ouverture à la demi-journée, sur l’usage et l’accessibilité des données au-delà de la sphère universitaire.

### Table ronde conclusive autour de l’usage des données entre collecte, traitement et valorisation