Semaine Data SHS

Traiter et analyser des données en sciences humaines et sociales

Du 9 au 13 décembre 2024

à la MSH Lyon St-Étienne, Lyon 7e / St-Étienne / distanciel

CONFÉRENCES ET ATELIERS MÉTHODOLOGIQUES

Contexte et objectifs

Dans le cadre de sa plateforme universitaire de données (PUD) PANELS, la MSH Lyon St-Etienne organise du 9 au 13 décembre 2024 sa semaine « Data SHS », qui a pour thématique générale « Traiter et analyser des données en sciences humaines et sociales ».

Cette semaine, inscrite dans le plan national de formations de PROGEDO, propose une série de présentations et d'ateliers pratiques autour des méthodes de traitement et d'analyse de données en sciences humaines et sociales.

Ces conférences et ateliers ont un double objectif : présenter en détails les fondements théoriques des méthodes d'analyse et de représentation de données SHS et expliciter les moyens de leur mise en œuvre pratique. Le but est de permettre aux participants d'adapter ces méthodes à leurs propres travaux, de façon la plus autonome possible. Ces présentations seront aussi l'occasion d'aborder les principes et outils utiles à la sciences ouverte et à la gestion de données « FAIR » en SHS.

Public : ce programme s'adresse aux chercheurs, enseignants-chercheurs, ingénieurs et techniciens, ainsi qu'aux doctorants, des laboratoires SHS du site Lyon St-Étienne.

Thématiques

5 grandes thématiques ont été retenues pour cette semaine de conférences et d'ateliers :

– **données de recherche** : actualités Progedo et Huma-Num, *Data papers* et valorisation (2 modules) ;
– introduction à la statistique : **analyse descriptive** (Excel), tests statistiques (R) (2 modules) ;
– **géomatique et cartographie**, humanités spatiales (2 modules) ;
– analyse textuelle outillée, **textométrie** : TXM et IRaMuTeQ (2 modules) ;
– **analyse qualitative** : méthodologies, outils et logiciels (1 module).

LUNDI 9 DÉCEMBRE 2024

Données de recherche : actualités Progedo et Huma-Num, *Data papers* et valorisation

Module 1 – Données de recherche : s'appuyer sur les infrastructures nationales

Intervenants : **Olivier Beaude**, directeur de l'IR* Huma-Num, **Nicolas Sauger**, directeur de l'IR* PROGEDO, **Alioscha Massein**, ingénieur d'études en statistiques (CNRS, MSH Lyon St-Étienne), **Julie Thomas**, maître de conférences en sociologie (Univ. St-Etienne, CMW - UMR 5283)

10h15-11h15 : **Actualités des infrastructures IR* Huma-Num et PROGEDO**

Deux infrastructures de recherche étoile (IR*) existent en sciences humaines et sociales : Huma-Num, acteur national dédié aux humanités numériques, et PROGEDO, qui favorise la culture des données quantitatives en SHS.

Huma-Num développe, avec les communautés et à partir d'un pilotage scientifique, une infrastructure numérique de niveau international (nœud français des ERIC DARIAH et CLARIN) pour les SHS. Elle structure, par l'intermédiaire de consortiums regroupant des acteurs de la recherche et d'un réseau de points de présence dans les maisons des sciences de l'Homme (MSH), l'accompagnement des chercheurs et laboratoires en matière d'infrastructure numérique pour les données de la recherche. Elle permet ainsi le développement, la réalisation et la préservation sur le long terme les programmes de recherche – leurs données et outils - dans un contexte de science ouverte et de partage des données.

La principale mission de PROGEDO est de développer la culture des données, impulser et structurer une politique des données d'enquêtes pour la recherche en sciences sociales. Son organisation est conçue pour les trois niveaux d'enjeux stratégiques où elle doit intervenir : européen, avec l'investissement dans les consortiums européens (ERIC) et le partage des grandes enquêtes ; national, avec les départements de la documentation et la diffusion des données par le département Quetelet-Progedo-Diffusion ; et régional, dans les MSH et les Universités avec les Plateformes Universitaires de Données (PUD).

Cette 1e partie sera l'occasion de vous présenter les deux infrastructures, leurs missions, les services qu'elles proposent à la communauté scientifique, ainsi que leurs actualités ou nouveautés en terme d'outils, services ou projets.

11h30-12h30 : **Projet de recherche « Danseà2 » : utiliser les données d'enquêtes nationales dans une recherche par méthodes mixtes**

La recherche Danseà2 : « Voulez-vous danser avec moi ? » de l'espace social des danses à deux à l'évolution des normes de genre dans la pratique de quatre danses en France a été financée par l'INJEP qui visait à soutenir des projets de recherche analysant les pratiques sportives à tous les âges, notamment en exploitant les données de l'enquête nationale de 2020 sur les pratiques sportives en France. La recherche se penchait sur des pratiques de loisir a priori répandues mais faiblement étudiées en sociologie du sport, les danses à deux. Ce sont pourtant des activités physiques mixtes du point de vue du genre et de l'âge, qui peuvent être pratiquées de manière régulière, autonome et/ou encadrée. Le projet se donnait pour but d'objectiver les types de publics des danses à deux comparativement aux autres danses, les conceptions des pratiques portées par les danseuses et danseurs au regard de leur trajectoire sociale et des modalités de leur pratique, ainsi qu'à identifier les leviers de transformation dont elles disposent concernant les évolutions actuelles relatives au genre. Pour cela, ce sont des méthodes mixtes qui ont choisies, croisant le traitement secondaire des données des enquêtes ENPPS 2020 et PC2018 : l'enquête pratiques culturelles des français·es de 2018, avec des entretiens biographiques et des observations directes et participantes effectuées au sein de quatre danses spécifiques (forró, lindy hop, salsa, tango).

En s'appuyant sur le traitement statistique de ces deux enquêtes nationales, nous avons pu constater des tendances que l'on retrouve et que l'on explicite avec le qualitatif. Les allers-retours entre les observations « statistiques » et celles sur le terrain ont permis à la fois de préciser les hypothèses (par ex. sur la question du couple), de valider nos observations, ainsi que d'ouvrir de nouvelles pistes d'enquête concernant des informations parcellaires ou trop généralistes dans les enquêtes nationales pour les besoins du projet de recherche.

Nous reviendrons dans un premier sur l'accès et l'utilisation des jeux des données que nous avons mobilisé dans le projet « Voulez-vous danser avec moi », avant de revenir sur la partie qualitative et l'imbrication des méthodes.

Horaires : 10h15-12h30 – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc) + distanciel [intervenants en visioconférences]

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/894276?lang=fr

Module 2 – *Data journals* et *data papers* en SHS : présentations, enjeux et retours d'expériences

Séance co-organisée par les Plateformes universitaires de données PANELS, PROGEDO-Loire et PUD-T.

Intervenants : **Victor Gay** (Univ. Toulouse, MSHS-T), **Julien Muller** (Univ. Lorraine, MSH Lorraine), **Sébastien Poublanc** (CNRS, FRAMESPA - UMR 5136), **Nicolas Sauger** (IR* PROGEDO, CSDP - Sciences Po), **Thibault Le Corre** (Univ. Montréal, Géographie-cités - UMR 8504), **Frédérique Mélanie-Becquet** (CNRS, LATTICE - UMR 8094), **Tessa Y. Vermeir** (Univ. Lyon 2, DDL - UMR 5596)

Le travail de production de données amène souvent de nombreux questionnements et choix dans la sélection de l'information, qui permettent de délimiter les perspectives et les limites d'un travail de recherche. Ce travail peut être valorisé à la fois par le dépôt des données et de sa documentation dans un entrepôt, mais également par la production d'un data paper afin de rendre visible et valorisable ce travail souvent ignoré du travail de recherche. Cette séance de la semaine Data-SHS se concentre donc sur la présentation et le questionnement de ces nouveaux supports que sont les data papers, et les data journals.

Programme :

14h-14h30 : Définitions et enjeux des data papers – Victor Gay (TSE)

14h30-15h30 : Revue des revues. Discussions autour d'une revue récente (*DEMC – Données, expériences, méthodes et codes*), d'une revue en projet (*La revue des données*) et de la revue *Humanités numériques* – Julien Muller, Stéphane Poublanc et Nicolas Sauger

15h45-17h : Retours d'expérience de chercheurs ayant publié des data papers – Thibault Le Corre, Frédérique Mélanie-Becquet, Tessa Y. Vermeir & Victor Gay.

Horaires : 14h-17h – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc) + distanciel

[retransmission en visioconférence commune entre la MSH Lyon St-Étienne, la MSH Loire et la MSHS Toulouse]

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/915441?lang=fr

MARDI 10 DÉCEMBRE 2024

Introduction à la statistique : analyse descriptive (Excel), tests statistiques (R)

Module 1 – Introduction aux statistiques descriptives avec Excel

Intervenant : **Alioscha Massein**, ingénieur d'études en statistiques (CNRS, MSH Lyon St-Étienne)

L'utilisation et la réutilisation des données quantitatives en SHS est un nécessaire dans une grande variété de projets de recherche. Qu'il s'agisse de produire des indicateurs, de comparer des effectifs, de produire des proportions, les SHS peuvent s'appuyer sur une analyse descriptive d'un terrain à partir de données produites dans le cadre du projet de recherche, ou de jeu de données mis à la disposition de la recherche et du grand public, comme par exemple dans l'entrepôt de données PROGEDO, qui donne un accès gratuit et simplifié aux grandes enquêtes nationales (INSEE, INED, etc.) et internationales.

Pour accompagner les personnes qui souhaitent se former à l'utilisation de ces jeux de données quantitatifs, la PUD PANELS propose un atelier de formation et d'introduction aux statistiques descriptives avec Excel. Nous y verrons :

1. L'entrepôt de données PROGEDO ;
2. Découverte du logiciel Excel ;
3. Une introduction à la statistique et aux principaux indicateurs ;
4. Les tris à plats et les tris croisés.

Cet atelier est ouvert à toute personne qui souhaite se former au traitement de données quantitatives, avec pour seule condition d'avoir Excel (ou équivalent libre) sur son ordinateur.

Horaires : 9h30-12h30 – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/834891?lang=fr

Module 2 – Réaliser des tests statistiques à l'aide de R et du package infer

Intervenant : **Lise Vaudor**, ingénieur de recherche en analyses statistiques (CNRS, EVS - UMR 5600, plateforme ISIG)

Cet atelier est dédié à la réalisation de tests statistiques avec le logiciel R, et s'adresse à toute personne souhaitant approfondir ses connaissances en statistique appliquée. La session débutera par une présentation théorique des tests d'hypothèse, en mettant l'accent sur le test de Student et le test du Chi-2, ainsi que sur leurs conditions d'application et interprétations. Ensuite, un temps sera consacré à la pratique, avec des exercices guidés que les participants pourront réaliser en ligne ou sur leur propre ordinateur, à condition d'avoir installé R et les packages nécessaires en amont. Ceux qui le souhaitent auront également la possibilité d'appliquer ces méthodes sur leurs propres jeux de données, facilitant ainsi une approche personnalisée des concepts abordés.

Horaires : 14h-16h – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/515671?lang=fr

MERCREDI 11 DÉCEMBRE 2024

Géomatique et cartographie, humanités spatiales

Module 1 – Introduction à la cartographie avec Carto

Intervenant : **Giovanni Pietro Vitali**, maître de conférences en humanités numériques (Univ. Paris-Saclay, DYPAC - UR 2449)

Cet atelier de cartographie numérique est conçu pour introduire les concepts fondamentaux de la cartographie numérique et l'utilisation d'applications spécifiques pour la création de cartes. Nous allons explorer les différents formats de données géographiques et les outils nécessaires pour gérer et analyser ces données efficacement. Nous utiliserons l'application carto.com pour la création des cartes. Carto est une plateforme de cartographie interactive qui permet aux utilisateurs de visualiser, analyser et partager des données géospatiales. Grâce à ses fonctionnalités intuitives et puissantes, Carto.com est idéal pour les étudiants et les enseignants qui souhaitent intégrer la cartographie dans leurs projets éducatifs et de recherche.

Pour profiter pleinement des fonctionnalités de Carto, nous vous encourageons à créer un profil gratuit. Carto.com offre des comptes gratuits spécifiquement pour les étudiants et les enseignants. Pour plus d'informations sur la façon de créer votre profil et les avantages offerts, veuillez consulter le lien suivant : https://docs.carto.com/faqs/carto-for-education. La procédure prend un peu temps, la commencer au moins deux semaines avant l'atelier.

Objectifs :

1. Collecte et analyse des documents : comprendre l'importance de la collecte et de l'analyse préliminaire des documents géographiques ;
2. Collecte des données : Apprendre à collecter et à organiser des données géospatiales pour la cartographie ;
3. Création de cartes : utiliser Carto.com pour créer des cartes interactives et visuellement attrayantes ;
4. Concepts clés des SIG : maîtriser des concepts essentiels tels que les couches (layers), les points, les lignes, les polygones, et les différents formats de fichiers géospatiaux (GeoJSON, KML, Shapefile, etc.).

Horaires : 9h30-12h30 – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)

Inscription (avant le 27 novembre) > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/878345?lang=fr

Module 2 – Eléments d'analyse spatiale – Venez avec vos données !

Intervenant : **Hélène Mathian**, ingénieur de recherche (CNRS, EVS - UMR 5600)

L'atelier se propose d'introduire quelques éléments qui structurent les démarches d'analyse des données géographiques. Le champ de l'analyse spatiale cherche à caractériser les organisations spatiales et s'appuie sur la construction d'indicateurs ou de statistiques adaptées. Nous aborderons quelques unes de ces méthodes, comme la caractérisation des semis de points, les calculs d'autocorrélation spatiales, la caractérisation d'empreintes, etc.

La séance se décomposera en 2 parties : la 1e partie sera consacrée à la mise en œuvre de quelques exemples d'analyse d'organisations spatiales ; la 2e partie sera consacrée à la formalisation de déroulements méthodologiques à partir des données et des questions des participant.es.

Pré-requis : l'environnement de travail de l'atelier sera le logiciel QGIS. Les participants devront l'avoir installé. Ils/elles devront avoir les bases pour charger des données géographiques et être familiarisés avec les formats de données géographiques (shp ou gpkg). Il sera apprécié que les participants arrivent avec leur propres données et leurs questions pour interagir dans la 2e partie. Un espace d'interaction sera proposé aux inscrits avant l'atelier.

Horaires : 14h-17h – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/619485?lang=fr

JEUDI 12 DÉCEMBRE 2024

Analyse textuelle outillée, textométrie : TXM et IRaMuTeQ

Module 1 – TXM, un logiciel de textométrie open-source pour l'analyse de données textuelles en sciences humaines et sociales

Intervenant : **Bénédicte Pincemin**, chargée de recherche en linguistique (CNRS, IHRIM - UMR 5317)

Cet exposé présente le logiciel TXM du point de vue de ses possibilités d'usage en recherche. Une démonstration donne un aperçu de la palette des principales fonctionnalités disponibles. Une fois cette vue d'ensemble concrète posée, on peut d'une part situer la textométrie par rapport à d'autres approches d'analyse de données textuelles (text mining, CAQDAS,...), et d'autre part dégager les caractéristiques de TXM au sein des logiciels de textométrie, complémentaires les uns des autres. Dans le domaine très vaste des humanités numériques, on entend ainsi aider à comprendre à quels types de besoins et d'attentes la méthode textométrique en général, et le logiciel TXM en particulier, peuvent correspondre.

Nous vous invitons à télécharger et installer le logiciel avant la séance à cette adresse : https://txm.gitpages.huma-num.fr/textometrie/files/software/TXM/0.8.3/ - Si vous êtes confronté à un problème d'installation, vous pouvez contacter l'équipe TXM à cette adresse : textometrie (c) groupes.renater.fr

Horaires : 9h30-12h30 – **Lieu** : Université St-Étienne, Campus Tréfilerie (salle K201, bâtiment K, 2e étage)

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/867759?lang=fr

Module 2 – Utiliser TXM : fonctionnalités d'analyse

Intervenant : **Bénédicte Pincemin**, chargée de recherche en linguistique (CNRS, IHRIM - UMR 5317)

L'objectif de cet atelier pratique est de prendre en main le logiciel de textométrie open-source TXM et d'apprendre à (mieux) l'utiliser, en fonction des attentes et besoins communiqués par les participants. Il pourra poser quelques bases pour les débutants et faire connaître des fonctionnalités ou astuces d'analyse basiques ou avancées répondant à des besoins concrets exprimés (la durée de la séance ne permet pas de faire le tour du logiciel).

La séance sera principalement consacrée à l'ergonomie générale du logiciel TXM et à ses possibilités d'analyse une fois l'import du corpus réalisé (telles que listes de mots ou d'expressions, concordances, mots associés, répartition des mots, cartographie du corpus...). Il est possible d'apporter ses données pour avoir des orientations sur la façon de les entrer dans TXM.

Les participants sont invités à venir avec un ordinateur sur lequel TXM est déjà installé (cf site textometrie.org). Ce n'est pas obligatoire mais vivement conseillé pour tirer le meilleur parti de l'atelier. Il ne sera pas possible de s'occuper d'installation en séance. Les attentes et questions pour préparer l'atelier sont à communiquer via l'adresse de contact de l'équipe TXM : textometrie (c) groupes.renater.fr.

Horaires : 14h-17h – **Lieu** : Université St-Étienne, Campus Tréfilerie (salle K201, bâtiment K, 2e étage)

Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/322525?lang=fr

VENDREDI 13 DÉCEMBRE 2024

Analyse qualitative : méthodologies, outils et logiciels

Module 1 – Travailler sans CAQDAS : est-ce (encore) possible ?

Intervenants : **François Féliu**, ingénieur d'études (CNRS, CMW - UMR 5283)

Bien souvent l'expression « ouvrir la boite noire » est utilisée pour discuter d'un logiciel, d'un outil technologique afin d'expliciter son fonctionnement et son utilité pour la recherche. Dans cet atelier, à l'inverse, il s'agira d'ouvrir « ma boite noire » de l'analyse de données qualitatives en sociologie. En prenant appui sur plusieurs enquêtes collectives, je reviendrai sur mes différentes manières de travailler, sans CAQDAS. J'ouvrirai « ma boite à outils » afin de revenir, avec vous sur les différents bricolages effectués durant une enquête : Comment passer d'un entretien à une donnée exploitable ? Comment en rendre compte à ses collègues ? Je vous décrirai, dans le contexte d'enquête collective, mes manières de produire, traiter et analyser des données qualitatives : utilisation de logiciels de traitement de textes (Word, OpenOffice), de messageries instantanées, d'un drive universitaire, d'Excel.

Horaires : 10h-12h30 – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)
Inscription > formulaire en ligne : https://enquetes.msh-lse.fr/index.php/381112?lang=fr

Comité d'organisation

Alioscha Massein (CNRS, MSH Lyon St-Etienne)
Céline Faure (CNRS, MSH Lyon St-Etienne)
Loïc Bonneval (Univ. Lyon 2, CMW)
Orline Poulat (CNRS/Univ. St-Etienne, MSH Lyon St-Etienne)
Alexandra Dugué (Univ. Lyon 2, MSH Lyon St-Etienne)
Amélie Hugot (CNRS, MSH Lyon St-Etienne)

Informations pratiques

Les personnes intéressées pourront s'inscrire à un ou plusieurs modules (durée 1h à 3h). La participation est gratuite, mais sur inscription obligatoire (nombre de places limité).

Localisation
Les séances se tiendront pour la plupart en présentiel, avec distanciel possible pour certaines.
Lundi : hybride – MSH Lyon St-Etienne, 14 av. Berthelot, Lyon 7e (salle André Bollier) + distanciel
Mardi, Mercredi, Vendredi : MSH Lyon St-Etienne, 14 av. Berthelot, Lyon 7e (salle André Bollier)
Jeudi : Univ. Jean Monnet, Campus Tréfilerie, rue du 11 novembre, St-Étienne (salle KI201, bâtiment K, 2e étage)

Inscriptions
Merci de vous inscrire, avant le 3 décembre 2024, à l'aide des formulaires en ligne spécifiés pour chacun des modules (cf programme détaillé ci-dessus).

Contacts
Contenu des ateliers, organisation : Alioscha Massein (MSH Lyon St-Etienne) - alioscha.massein@msh-lse.fr

Site web : www.msh-lse.fr/semaine-data-shs-2024
Twitter : @mshlyon #SemaineDataSHS