

Semaine Data SHS

Traiter et analyser des données en sciences humaines et sociales

Du 7 au 11 décembre 2020

Événement en 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 7 au 11 décembre 2020 une semaine « Data SHS », qui aura pour thématique générale « Traiter et analyser des données en sciences humaines et sociales ». Cette semaine, inscrite dans le plan national de formations de la TGIR Progedo, propose une série de conférences et d'ateliers pratiques autour des méthodes de traitement et d'analyse de données en sciences humaines et sociales. Ces présentations seront aussi l'occasion d'aborder les principes et outils utiles à l'*Open science* et à la gestion de données « FAIR » en SHS.

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.

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

Thématiques

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

- **gestion de données** en SHS : retours d'expérience, débat et RGPD (3 modules) ;
- **textométrie** en SHS : outils, méthodes et applications (3 modules) ;
- panorama et écriture de l'objet « **Data papers** » (3 modules) ;
- grandes **enquêtes internationales** et bases de données quantitatives, TGIR Progedo (3 modules) ;
- **cartographie** et analyse spatiale appliquées aux SHS (1 module).

■

LUNDI 7 DÉCEMBRE 2020

Gestion de données en SHS : retours d'expérience, débat et RGPD

Module 1 – Les notions essentielles d'application du RGPD dans un cadre recherche

Intervenant : Céline Faure, statisticienne (CNRS, MSH Lyon St-Etienne - USR 2005)

Le monde de la recherche en sciences humaines et sociales implique souvent l'utilisation de données et notamment de données personnelles. Les faits de collecter, partager et protéger peuvent sembler contradictoires. Pourtant les acteurs de la recherche peuvent maintenir le niveau d'excellence de leurs recherches, tout en préservant les droits des personnes interrogées ou ayant fait l'objet d'une collecte. Pour cela, nous évoquerons les grands principes du règlement général de la protection des données applicable depuis le 25 mai 2018, et présenterons les modalités de mise en œuvre en s'appuyant sur les outils disponibles et les acteurs du site local (DPO).

Horaires : 11h-12h15 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/533153/lang-fr>

Module 2 – Culture numérique, culture des données : approches épistémologiques, politiques et concrètes

Intervenant : Eric Guichard, maître de conférences HDR en sciences de l'information et de la communication (ENSSIB, TRIANGLE - UMR 5206)

À partir d'exemples tirés de la Covid et de la téléphonie mobile, il sera montré en quoi la culture numérique relève de la culture de l'écrit contemporain, avec son lot de ségrégations et d'émancipations. Une précision sera également effectuée sur l'OPA intellectuelle des GAFAM et équivalents sur les illettrés du numérique que nous sommes (presque) tous. La présentation finira sur une proposition de moyens afin d'y remédier, de relier technologie de l'intellect et pensée critique. Si le temps le permet, un débat aura lieu avec les participants autour de l'évolution des épistémologies des sciences sociales, du fait du passage de l'écriture imprimée à l'écriture numérique. Les exemples renverront à du *small* comme à du *big data*, à partir de sources publiques. Les outils informatiques évoqués seront d'âges divers.

Horaires : 15h-17h – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/488288/lang-fr>

■

MARDI 8 DÉCEMBRE 2020

Textométrie en SHS : outils, méthodes et applications

Module 1 – La textométrie, une approche quantitative et qualitative des données textuelles

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

Cet exposé entend d'abord questionner les réalités concrètes des « données textuelles », et reconnaître la diversité des points de vue sur ces données, qui peuvent être très différents dans un contexte de production ou de recherche, et au sein même de la recherche entre sciences sociales et sciences humaines. À cette diversité de points de vue répond une diversité de méthodes et d'outils – CAQDAS, concordanciers, text mining –, au sein desquels situer la textométrie.

Le propos sera centré sur l'approche textométrique. Nous illustrerons cette approche en parcourant les principaux types de questions pour lesquels la textométrie a développé des outils : que peut-on faire avec un logiciel de textométrie ? Puis, prenant du recul, nous mettrons en évidence des principes sous-jacents et des éléments méthodologiques forts de l'approche textométrique, à discuter en regard des attentes de la recherche en SHS.

Le logiciel open-source TXM sera le principal support de l'exposé.

Horaires : 10h-11h30 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/844118/lang-fr>

Module 2 – Lexicométrie et textométrie appliquées à l'analyse de corpus en SHS : retour d'expérience sur l'analyse des données textuelles appliquée à la géographie

Intervenant : Silvia Flaminio, chargée de recherche (Univ. Lausanne, Institut de Géographie et de durabilité)

Après une brève introduction sur les enjeux de la lexicométrie et de la textométrie, la présentation proposera un retour d'expérience sur l'analyse de plusieurs corpus textuels (issus d'entretiens, de presse, de publications scientifiques). Les fonctions de deux logiciels libres d'analyse des données textuelles (Iramuteq et TXM) seront évoquées ainsi que leurs potentialités pour des études en sciences sociales, notamment en géographie.

Horaires : 14h-15h30 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/914341/lang-fr>

Module 3 – Analyse de graphes avec attributs textuels pour l'étude des réseaux sociaux

Intervenant : Antoine Gourru, doctorant en informatique et mathématiques (Univ. Lyon 2, ERIC - EA 3083)

Un graphe est un objet mathématique défini par un ensemble d'entités, appelées nœuds, connectées par des liens. Par exemple, les réseaux sociaux forment des graphes : un ensemble d'utilisateurs se « suivent », se « connectent ». L'étude des graphes et de leurs propriétés est un domaine à part entière depuis le XVIII^e siècle. En plus des liens, les nœuds possèdent la plupart du temps un ensemble d'attributs (âge d'un utilisateur, son sexe). Un attribut particulièrement intéressant à prendre en compte dans la modélisation est le texte (en général les publications), central dans l'analyse des réseaux de type Twitter, ou Facebook. De nombreuses méthodes d'analyse de données et d'apprentissage automatique permettent d'analyser les graphes avec attributs textuels, qui sont des objets a priori complexes.

Le module sera découpé en trois parties : 1) une introduction théorique sur les graphes ; 2) une session pratique sur leur manipulation/visualisation grâce au logiciel Gephi ; 3) une présentation des travaux récents conduits au laboratoire ERIC sur les graphes avec attributs textuels.

Prérequis : il n'est pas nécessaire de connaître un langage de programmation.

Outils/logiciels à télécharger : La partie pratique nécessite l'installation préalable des outils suivants : Java Runtime Environment (JRE) Java 7 ou 8 : <https://www.java.com/fr/download/>
Logiciel Gephi : <https://gephi.org/users/install/>

Horaires : 15h45-17h45 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/659976/lang-fr>

MERCREDI 9 DÉCEMBRE 2020

Panorama et écriture de l'objet « Data papers »

Module 1 – Les *data papers* dans les stratégies de publication des chercheurs

Intervenant : Cherifa Boukacem-Zeghmouri, professeur en sciences de l'information et de la communication (Univ. Lyon 1, ELICO - EA 4147)

Le propos de cette intervention est de situer la place des *data papers* dans les stratégies de publication des chercheurs, selon les disciplines. En se fondant sur les résultats d'une recherche observant les mutations de la communication scientifique, il s'agira de montrer comment le *data paper* a acquis de la visibilité et de la légitimité dans certaines communautés, et contribue aujourd'hui à diversifier les vecteurs de la communication scientifique.

Horaires : 10h-12h – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/979324/lang-fr>

Module 2 – *Data papers*, une introduction

Intervenant : **Joachim Schöpfel**, maître de conférences en sciences de l'information et de la documentation (Univ. Lille, GERIICO - UR 4073)

L'intervention introduira le nouveau concept des articles de données (*data papers*) : Quelles sont leurs fonctions ? Quelle structure, quel contenu, quelle taille ? Quelles consignes aux auteurs ? Elle donnera aussi un aperçu du paysage actuel : Combien d'articles de données, dans quelles revues et dans quels domaines ? Chez quels éditeurs ? En libre accès ou pas ? Un aspect intéressant des *data papers* est leur lien avec les entrepôts de données qui développent des fonctionnalités pour une génération automatique d'un *data paper*, y compris l'envoi à une plateforme de revue.

Horaires : 14h30-16h30 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/426869/lang-fr>

Module 3 – Rédiger des *Data papers* à l'aide de Jupyter Notebook

Intervenant : **Romain Mularczyk**, ingénieur d'études en Gestion de données (Univ. Lyon 2, MSH Lyon St-Etienne, USR 2005)

Dans ce mini-atelier/démo, il vous sera présenté les fonctionnalités de Jupyter Notebook, une application open-source permettant de créer et partager des documents contenant du code informatique pouvant être exécuté. Cela sera également l'occasion d'une démonstration de ses capacités en vue de la rédaction d'un *Data paper*.

Horaires : 16h45-17h30 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/515198/lang-fr>

■
JEUDI 10 DÉCEMBRE 2020

Grandes enquêtes internationales et bases de données quantitatives, TGIR Progedo

Module 1 – La TGIR Progedo et son département « Données historiques, économiques et financières »

Intervenant : **Sandra Brée**, chargée de recherche en histoire et démographie (Univ. Lyon 2, LARHRA - UMR 5190) et responsable du Département « Données historiques, économiques et financières » de la TGIR Progedo

Cette session vous présentera PROGEDO, la Très Grande Infrastructure de Recherche chargée d'impulser et structurer une politique publique des données pour la recherche en sciences sociales. Pour développer la culture de données dans les communautés de recherche SHS, PROGEDO travaille à faire connaître le vaste ensemble de statistiques et d'enquêtes publiques produites en France et œuvre à en faciliter et organiser l'accès.

L'accent sera ensuite mis sur le département « Données historiques, économiques et financières » de PROGEDO, qui a pour ambition de recenser les bases de données quantitatives existantes dans le domaine, de les valoriser et d'inciter à partager ses propres bases de données quantitatives afin qu'elles puissent être utilisées par le plus grand nombre. Seront présentés les missions et les objectifs de ce nouveau département créé en 2020.

Horaires : 9h30-10h30 – **Lieu :** visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/767825/lang-fr>

Module 2 – La plateforme de données des acteurs du territoire de la Métropole de Lyon : data.grandlyon.com

Intervenant : **Emmanuel Gastaud**, chef de produit plateforme data (Métropole de Lyon, Service Données Métropolitaines)

La Métropole de Lyon considère la donnée comme un actif à haute valeur ajoutée : c'est une matière première indispensable pour une métropole innovante. Sa stratégie de la donnée répond à plusieurs enjeux qui seront détaillés et s'appuie sur un écosystème riche : des partenaires territoriaux du secteur privé, du secteur académique, du secteur public, des laboratoires d'innovation... qui peuvent être producteurs ou ré-utilisateurs des données.

La plateforme data.grandlyon.com est tout à la fois un outil de collecte, de normalisation et de publication des données du territoire. Elle permet, par un système de licences de réutilisation, de répondre aux exigences de l'*open data*, à la mise à disposition de données d'intérêt général, et à la réutilisation de données en mode privé au sein de projets partenariaux. Différents exemples de réutilisation des données seront abordés.

Les rubriques et fonctionnalités du portail data.grandlyon.com seront présentées, en particulier les fonctionnalités de recherche et de navigation dans les quelques 950 jeux de données disponibles. Seront également détaillées les différentes possibilités d'accès à la donnée, téléchargement ou utilisation de web services notamment (API). En fonction des attentes des participants, un focus pourra également être fait sur l'architecture technique de la plateforme.

Prérequis : intérêt pour la consultation et l'utilisation de jeux de données du territoire de la Métropole de Lyon, et pour les questions d'ouverture et de partage des données.

Horaires : 11h-12h – **Lieu** : visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/475584/lang-fr>

Module 3 – Mises en données de la ville : quelle structuration ? quels dispositifs ? quelles médiations ?

Intervenants : **Valentyna Dymytrova**, maître de conférences en sciences de l'information et de la communication (Univ. Lyon 3, ELICO - EA 4147) et **Françoise Paquienséguy**, professeur en sciences de l'information et de la communication (Sciences Po Lyon, ELICO - EA 4147)

Les deux chercheuses présenteront quelques éléments d'un projet de recherche ANR et des réflexions qui l'ont suivi dans le cadre de la rédaction de l'ouvrage « Données urbaines et smart cities » paru récemment aux Éditions des Archives contemporaines, disponible en *open access* (www.archivescontemporaines.com/books/9782813003263).

En se fondant sur une analyse de huit portails, Françoise Paquienséguy montrera que les portails *open data* métropolitains sont la marque : 1) d'un processus de facilitation de la réutilisation des données ; 2) des stratégies de développement et de promotion du territoire ; et 3) de la transition numérique des villes par et avec les données vers une ville intelligente, créative, et/ou durable.

À son tour, à partir d'une enquête de terrain menée auprès des professionnels des *data*, Valentyna Dymytrova restituera les chaînes de transformation des données qui accompagnent la création de nouveaux services informationnels et montrera comment les producteurs de ces services deviennent de nouveaux médiateurs urbains. En ce sens, la division du travail propre à la fabrique de la *Smart city* basée sur l'*open data* soulève des enjeux fondamentalement politiques.

Horaires : 14h-15h30 – **Lieu** : visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/369453/lang-fr>

VENDREDI 11 DÉCEMBRE 2020

Cartographie et analyse spatiale appliquées aux SHS

Module 1 – Magrit, exemples pratiques à la réalisation de cartes thématiques géostatistiques : mini-atelier de démonstrations partagées

Intervenants : **Hervé Parmentier**, assistant ingénieur cartographe/géomaticien (ENS de Lyon, EVS - UMR 5600)

Magrit est une application en ligne de cartographie thématique statistique développée par Matthieu Viry, Timothée Giraud et Nicolas Lambert au sein du pôle géomatique de l'UMS RIATE. Les concepteurs sont ingénieurs en traitement, analyse et représentation de données.

L'objectif de ce mini-atelier démo est de montrer à partir d'exemples existants comment réaliser des cartographies unies et bivariées partant de variables statistiques quantitatives, des fonds de carte, aux données attributaires, en passant par les représentations spatiales, puis mises en cartes et en pages conformes aux règles de la cartographie, jusqu'enfin leurs exports au format PDF ou SVG. Nous essayerons ainsi par cette approche d'analyser les potentialités de cet outil cartographique simplifiant à la fois cet art, ses techniques et cette science.

Prérequis : connaissances/pratiques en système d'information géographique et cartographie.

Horaires : 10h-10h45 – **Lieu** : visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/533189/lang-fr>

Gestion de données en SHS : retours d'expérience, débat et RGPD

Module 3 – Table ronde : retours d'expérience autour du plan de gestion de données en sciences humaines et sociales

Intervenants :

Delphine Cavallo, responsable diffusion et traitement numérique de la recherche (Aix-Marseille Univ., TELEMME)

Lauriane Locatelli, docteure en lettres et littératures anciennes, post-doctorante (Projet Hairesis, programme IMPULSION IDEX Lyon, HISOMA - UMR 5189)

Cynthia Pedroja, chargée du traitement de données scientifiques (CNRS, Campus Condorcet, CESR - UMR 7323)

Adeline Levivier, ingénieure d'études (École Française d'Athènes)

Axelle Janiak, ingénieure d'études (CNRS, CEIAS - UMR 8564)

Cette table ronde sera l'occasion d'inviter des intervenantes venant de différents horizons et ayant contribué à la réalisation d'un plan de gestion de données (PGD) pour nous raconter leurs expériences. Après une présentation des contextes de création de chacun des PGD, différents thèmes et questions seront abordés. Allant du plus concret avec les étapes de mises en place, les acteurs et appuis, le budget, ainsi que les outils utilisés pour l'archivage et la diffusion des données, jusqu'au ressenti plus personnel : quelles ont été les difficultés rencontrées ? Les avantages procurés ? L'utilité et l'adaptation aux SHS ? Quels conseils donnés aux futurs contributeurs de PGD ? La parole sera également donnée aux participants pour poser leurs questions et étoffer le débat.

Horaires : 14h-16h – **Lieu** : visioconférence

Inscription > formulaire en ligne : <http://enquetes.msh-lse.fr/index.php/337516/lang-fr>

Comité d'organisation

Agathe Déan (CNRS, MSH Lyon St-Etienne)

Romain Mularczyk (Univ. Lyon 2, MSH Lyon St-Etienne)

Loïc Bonneval (Univ. Lyon 2, CMW)

Céline Faure (CNRS, 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 2h). La participation est gratuite, mais sur inscription obligatoire (nombre de places limité).

Localisation

En raison de la situation sanitaire, l'intégralité des séances se tiendra en visioconférence.

Inscriptions

Merci de vous inscrire, avant le 30 novembre 2020, à l'aide des formulaires en ligne spécifiés pour chacune des conférences ou ateliers (cf programme détaillé ci-dessus).

Un lien de connexion à la visioconférence vous sera communiqué ultérieurement.

Contacts

Contenu des ateliers, inscriptions, organisation : Agathe Déan (MSH Lyon St-Etienne) - agathe.dean@msh-lse.fr

Site web : www.msh-lse.fr/semaine-data-shs-2020

Twitter : [@mshlyon](https://twitter.com/mshlyon) #SemaineDataSHS2020

