

Semaine Data SHS

Traiter et analyser des données en sciences humaines et sociales

Du 5 au 9 décembre 2022

à la MSH Lyon St-Étienne, Lyon 7^e et 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-Étienne organise du 5 au 9 décembre 2022 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 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.

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 :

- **données de recherche** : trouver, produire, mettre à disposition, en conformité (4 modules) ;
- **analyse qualitative outillée** : textométrie, CAQDAS (2 modules) ;
- plan de **gestion** de données, publication en **data papers** (3 modules) ;
- outils et méthodes de **visualisation** de données et réseaux (1 module) ;
- PROGEDO et ses services : **grandes enquêtes**, catalogues de données (2 modules).

■

LUNDI 5 DÉCEMBRE 2022

PROGEDO et ses services : grandes enquêtes, catalogues de données

Module 1 – Rechercher des données sur Quetelet-Progedo-Diffusion

Intervenants : Sébastien Oliveau, directeur de l'IR* PROGEDO, Lorraine Adam, ingénieure d'études (CNRS, PROGEDO – ADISP, Paris)

14h00-14h15 : Présentation générale de l'IR* Progedo

L'infrastructure de recherche étoile (IR*), PROGEDO a pour but de développer la culture des données, d'impulser et structurer une politique des données d'enquêtes pour la recherche en sciences sociales.

14h15-15h30 : Rechercher des données sur Quetelet-Progedo-Diffusion

Un nouveau catalogue réunit l'ensemble des références mises à disposition par les membres de Quetelet-Progedo Diffusion : L'Adisp et l'Ined. Plus de 1500 jeux de données sont actuellement accessibles via Quetelet-Progedo-Diffusion. Ils sont accessibles, gratuitement, aux chercheurs français et étrangers, doctorants, post-doctorants, et étudiants à des fins de recherche, de production scientifique et dans certains cas d'enseignement. Comment faire une recherche sur ce catalogue? Quels types de bases sont accessibles ? Sous quels formats ? Comment les commander ?

Horaires : 14h-15h30 – **Lieu :** Université St-Étienne, Campus Tréfilerie (salle du conseil)
+ retransmis en visioconférence à la MSH, Lyon 7^e (salle André Bollier)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/273393?lang=fr>

Module 2 – Le département « Données historiques, économiques et financières » de PROGEDO : recueillir et diffuser les bases de données historiques

Intervenante : Sandra Brée, chargée de recherche en histoire et démographie (CNRS, LARHRA) et responsable du département « Données historiques, économiques et financières » chez PROGEDO

Les bases de données historiques demandent des efforts et du temps pour être construites et constituent un patrimoine qu'il faut préserver et mettre en valeur. L'ambition de ce département est de faire vivre ces bases de données dont certaines sont inédites. Leur préservation et leur large diffusion sont rendues possibles grâce au soutien de l'équipe de l'Adisp-Progedo. Pour cela, les données doivent être documentées et Progedo peut aussi accompagner les chercheurs dans ces pratiques.

Le département a pour missions principales : permettre l'accès aux données historiques quantitatives ; soutenir les initiatives existantes ou en création ; recueillir les bases de données ; recenser les données existantes.

Horaires : 15h30-16h30 – **Lieu :** Université St-Étienne, Campus Tréfilerie (salle du conseil) ou MSH, Lyon 7^e (salle André Bollier) - intervenant en visioconférence

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/725361?lang=fr>

■

MARDI 6 DÉCEMBRE 2022

Plan de gestion de données, publication en *data papers*

Module 1 – Comment utiliser une boîte à outils avec le plan de gestion des données ?

Intervenant : Romain Boissat, ingénieur d'études en traitement de données (CNRS, Maison de l'Orient et de la Méditerranée - FR 3747)

Le plan de gestion des données (PGD ou DMP en anglais) est devenu un élément indispensable pour mener à bien ses recherches. Il est une obligation de la part des institutions (ANR, ERC, etc.) depuis 2019 pour tout projet financé. Au-delà de ces aspects réglementaires, il est un outil de suivi des décisions prises au sujet des données de la recherche d'un projet scientifique. Il est donc à la fois un document créé pour le futur, pour permettre de comprendre et réutiliser les données produites, mais également un outil de travail tout au long de la vie du projet. Il oblige à se

poser des questions méthodologiques, juridiques et techniques, à les déployer et à en garantir l'application jusqu'au terme du projet et même au delà. Le PGD est donc une sorte de guide d'utilisation des outils nécessaires à l'exploitation, la diffusion et la réutilisation des données de la recherche produites par la recherche.

Cette présentation a pour but de présenter le PGD, son rôle et son fonctionnement à la lumière de plans existants sur des projets en cours d'ANR, ERC, etc. Il est bien entendu un outil qui liste les spécificités de chaque projet dont cette présentation souhaite rendre compte dans la mesure du possible. Elle guidera également les participants plus avancés vers les ressources et personnels disponibles sur le site de Lyon Saint-Etienne pour un accompagnement plus poussé à la rédaction de leur PGD.

Horaires : 10h30-11h30 – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/245585?lang=fr>

Module 2 – Les *data papers* en SHS : évolutions et perspectives

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

Cette séance propose une introduction aux *data papers* (articles de données) comme un nouveau type de publication scientifique. Quelle est leur finalité ? Quelle structure, quel contenu, et quelle différence avec un article « normal » ? Le module proposera une définition et analysera les *data papers* dans le contexte de la science ouverte, posera la question des liens avec les entrepôts de données et les plateformes de revues, et évoquera les enjeux des métadonnées et, d'une manière plus générale, des principes FAIR. Dans quel mesure les *data papers* contribuent-ils à la « FAIRisation » des données de recherche ? Quel est le développement des *data papers*, en particulier en SHS ? Quelles sont leurs perspectives, quelles sont les questions ouvertes ? Le module s'appuie sur une revue de littérature, sur des analyses scientométriques et conceptuelles, sur plusieurs événements dont la conférence DH Nord 2021 à Lille et sur une expérience pratique en tant qu'auteur, *peer reviewer* et rédacteur.

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

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/573353?lang=fr>

Module 3 – Rédiger et publier un *data paper* : deux retours d'expérience

Intervenants : **Louis Manière**, géomaticien (auto-entrepreneur), **Jean-Paul Rehr**, doctorant en histoire médiévale (Univ. Lyon 2, CIHAM) et ingénieur du ThEMA

Discutante : **Mariannig Le Béchec**, maître de conférences (Univ. Lyon 1, ELICO)

Les intervenants feront part de leurs expériences de publier un *data paper* (pourquoi ? comment ?) et les démarches et traitements réalisés sur leurs données pour cette publication.

16h00-16h30 : **Mettre en valeur ses données avec les *data papers* : retour d'expérience d'une publication en archéologie**

Ce *data paper* concerne des données géospatiales utilisées pour des travaux de recherche menés au sein des laboratoires Archéorient et HISOMA. Ces recherches sont destinées à élaborer un modèle de reconstruction des réseaux routiers hellénistiques et romains du désert oriental d'Egypte, avec une approche semi-empirique basée sur les itinéraires de voyageurs modernes. L'objectif de cet ensemble de données est d'être capable de reproduire ce réseau égyptien mais aussi de faciliter l'application de cette approche à d'autres domaines d'étude.

16h30-17h00 : **Présentation de 25 ans de données du projet ThEMA dans un *data paper***

Thesaurus Exemplorum Medii Aevi (ThEMA) est une base de données collaborative sur les exempla médiévaux. Les sermons médiévaux renferment des milliers d'*exempla*, ou histoires exemplaires, souvent utilisés pour transmettre doctrine et morale religieuses aux fidèles. Longtemps considérés comme une part négligeable des textes dogmatiques et moraux, ils sont désormais aussi approchés comme sources culturelles et historiques. Développée par l'historien Jacques Le Goff et ses héritiers, ThEMA comporte actuellement plus de 12 000 *exempla* indexés couvrant un long Moyen Âge s'étendant de l'Europe latine à Byzance et l'Asie, provenant de sources chrétiennes, juives, islamiques et bouddhistes. La conversion au format XML-TEI et l'engagement vers des « FAIR data » permettent un usage élargi des données.

Horaires : 16h-17h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/238668?lang=fr>

MERCREDI 7 DÉCEMBRE 2022

Outils et méthodes de visualisation de données et réseaux

Module 1 – Premiers pas en analyse des réseaux (Gephi)

Intervenant : **Giovanni Pietro Vitali**, maître de conférences en humanités numériques (Univ. Versailles St-Quentin-en-Yvelines, DYPAC - UR 2449)

L'objectif de cet atelier est de fournir aux participants les premières bases de l'analyse des réseaux par l'usage du logiciel libre Gephi. L'atelier sera divisé en trois parties :

1. Exemples d'analyse des réseaux (15 minutes) : cette première section illustrera quelques exemples typiques de visualisation de données appliquée à l'analyse des réseaux.
2. Préparez vos données (30 minutes) : une partie très importante du travail d'analyse des réseaux est la création d'un jeu de données. Dans cette partie de l'atelier, nous verrons comment formater les données pour composer une telle visualisation de données.
3. Faites votre réseau ! (60 minutes) : enfin, il est temps de créer votre propre réseau. Dans cette partie de l'atelier, il y aura un tutoriel sur l'éditeur graphique de réseaux Gephi. 15mn de questions pour terminer.

Exigences techniques : ordinateur (pas de tablette) avec au moins 15 Go d'espace libre sur le disque dur et au moins 8 Go de RAM (16 Go recommandés). Vous devez installer un gestionnaire de tableaux tel qu'OpenOffice et au moins deux navigateurs web (recommandé : Chrome et Firefox). Un compte Google est recommandé (google earth et google My maps). Installez Gephi au préalable : <https://gephi.org/users/download/>

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

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/681695?lang=fr>

Données de recherche : trouver, produire, mettre à disposition, en conformité

Module 1 – Méthodologies et outils géomatiques mises en oeuvre dans le projet Vélotactique (urbanisme cycliste et crise sanitaire)

Intervenant : **Nicolas Ovtracht**, ingénieur (CNRS, LAET)

les aspects relatifs aux enquêtes et aux bases de données spatiales (SIG) mises en oeuvre dans le cadre du projet ANR « VELOTACTIQUE : De l'urbanisme tactique cyclable au changement durable : une comparaison internationale des politiques publiques et des pratiques en contexte de pandémie ».

L'objectif du projet VELOTACTIQUE (<https://cyclops.hypotheses.org/velotactique>), coordonné par le laboratoire LAET est de collecter des données qui apporteront aux décideurs comme aux chercheurs des premiers éléments de connaissance sur la mise en oeuvre et la réception de l'urbanisme tactique cycliste dans le cadre de la crise sanitaire, ainsi que sa contribution à la transition énergétique. Le projet s'appuie sur une approche exploratoire et comparative, destinée à collecter sur 12 mois des données sur les politiques, leurs évolutions, leur réception auprès des usagers et les changements observés dans les pratiques. Il prévoit une analyse des politiques publiques au sein des métropoles étudiées, une cartographie des pistes cyclables, et l'observation de la réception des aménagements et mesures auprès de cyclistes avec un focus sur les nouveaux utilisateurs du vélo.

Le projet vise à analyser les impacts de la crise sanitaire sur les aménagements cyclables dans plusieurs métropoles en France, en Suisse, au Québec et en Colombie.

Horaires : 10h-11h – **Lieu :** MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/546915?lang=fr>

Module 2 – Accéder à des données confidentielles : le Centre d'accès sécurisé aux données (CASD)

Intervenante : **Florence Goffette-Nagot**, directrice de recherche (CNRS, GATE - UMR 5824)

Le CASD est une « bulle sécurisée », c'est-à-dire une infrastructure de stockage et de traitement des données où des données confidentielles sont sanctuarisées. 477 sources de données y sont disponibles, avec entre autres des données de l'INSEE, des ministères de la Justice, de l'Éducation nationale, de l'Agriculture et de l'alimentation. Une équipe dédiée gère l'accès des utilisateurs.

Cette intervention aura pour but de présenter les modalités d'utilisation du CASD, depuis la demande d'accès jusqu'à l'exportation des résultats d'analyse, sur la base d'une expérience d'utilisation pour des recherches en économie.

Horaires : 14h-15h – **Lieu** : MSH, Lyon 7e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/986763?lang=fr>

Module 3 – Le consortium ICPSR pour la recherche politique et sociale

Intervenants : **Amélie Vairalles**, chargée de communication, et **Nicolas Sauger**, professeur associé et directeur du CDSP (Sciences Po Paris)

Le Consortium interuniversitaire pour la recherche politique et sociale (ICPSR) gère un référentiel de données multidisciplinaire à grande échelle en sciences sociales, hébergé à l'Université du Michigan depuis 1962. Son catalogue de données contient plus de 16.000 études associées à environ 85.000 ensembles de données.

Cette séance présentera 3 aspects de ce Consortium : panorama des sources, process pour récupérer les bases de données, exemple de cas d'utilisation.

Horaires : 15h-16h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/465332?lang=fr>

JEUDI 8 DÉCEMBRE 2022

Données de recherche : trouver, produire, mettre à disposition, en conformité

Module 4 – Les notions essentielles pour se conformer au RGPD

Intervenante : **Céline Faure**, ingénieur d'études en statistiques (CNRS, MSH Lyon St-Étienne)

La recherche en sciences humaines et sociales implique souvent l'utilisation de données et notamment de données personnelles. Avec la mise en application du règlement général de la protection des données applicable (RGPD), le 25 mai 2018, les acteurs de la recherche doivent veiller à préserver les droits des personnes ayant fait l'objet d'une collecte.

Nous évoquerons les grands principes du RGPD et sa mise en œuvre : Quels changements le RGPD implique t-il ? Quelles sont les données concernées (données personnelles) ? Quel traitement est concerné ? Que dois-je faire pour préparer cette mise en conformité ? Comment identifier mon DPO ? Où puis-je m'informer ? Un quizz interactif vous permettra de tester les connaissances acquises lors de la séance. Nombre de places limité à 15 personnes.

Horaires : 11h-12h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/727256?lang=fr>

VENDREDI 9 DÉCEMBRE 2022

Analyse qualitative outillée : textométrie, CAQDAS

Module 1 – Le Vrai débat : comment traiter un corpus de 7 millions de mots ?

Intervenants : **François Robert** (TRIANGLE - UMR 5206, membre associé), **Vincent Ventresque**, ingénieur d'études (Univ. St-Etienne, TRIANGLE - UMR 5206)

Au sein du mouvement des Gilets jaunes, un collectif a ouvert une plateforme de démocratie participative, *Le Vrai débat* (30 janvier - 3 mars 2019). Cette initiative constitue une expérience inédite et un témoignage de première importance pour comprendre le mouvement des Gilets jaunes, témoignage qui a pris la forme d'un texte collectif de plus de 7 millions de mots. Face à la difficulté de traiter un tel corpus, les animateurs du site *Le Vrai débat* ont demandé l'aide de chercheurs et de laboratoires de recherche pour analyser leurs données. Le groupe de travail

Gilets jaunes du laboratoire TRIANGLE qui s'était constitué dès le début du mouvement, a fourni, en exclusivité, aux animateurs du Vrai débat, une première lecture de ces données. C'est la méthodologie de traitement de données qui sera ici présentée. Elle combine l'utilisation de plusieurs logiciels notamment, TXM, IRAMUTEQ.

Horaires : 14h-15h – **Lieu** : MSH, Lyon 7e (salle Elise Rivet, 4^e étage)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/788184?lang=fr>

Module 2 – Analyse quantitative et qualitative de corpus textuels avec TXM

Intervenant : Alexey Lavrentev, ingénieur de recherche (ENS de Lyon, IRHIM - UMR 5317)

La plateforme TXM (<http://textometrie.org>) fournit un ensemble d'outils d'analyse de données textuelles développés dans le cadre de la textométrie. Cette méthode se caractérise par un équilibre entre les approches quantitative (listes de fréquence, analyse de correspondances, classification, spécificités, progression, cooccurrences) et qualitative (concordances, retour au texte, plans textuels, annotation). L'exemple d'un corpus de la littérature russe du début du XX^e siècle sera utilisé pour illustrer le passage des analyses statistiques à leur interprétation et vérification qualitative.

Matériel facultatif : un ordinateur avec TXM installé. Téléchargement et documentation sur <https://txm.git-pages.huma-num.fr/textometrie/files/software/TXM/0.8.2>

Horaires : 14h-15h30 – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/337167?lang=fr>

Comité d'organisation

Céline Faure (CNRS, MSH Lyon St-Étienne)
Vannina Goossens (CNRS, MSH Lyon St-Étienne)
Loïc Bonneval (Univ. Lyon 2, CMW)
Alexandra Dugué (Univ. Lyon 2, MSH Lyon St-Étienne)

Orline Poulat (CNRS, MSH Lyon St-Étienne)
Alioscha Massein (CNRS, MSH Lyon St-Étienne)
Amélie Hugot (CNRS, MSH Lyon St-Étienne)

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

Les séances se tiendront pour la plupart en présentiel, avec distanciel possible (visioconférence) :
Lundi : Université Jean Monnet, Campus Tréfilerie, Maison de l'université, 10 rue Tréfilerie, St-Étienne (salle du conseil)
Mardi à Vendredi : MSH Lyon St-Étienne, 14 av. Berthelot, Lyon 7^e (salle André Bollier ou Elise Rivet)

Inscriptions

Merci de vous inscrire, avant le 28 novembre 2022, à l'aide des formulaires en ligne spécifiés pour chacune des conférences ou ateliers (cf programme détaillé ci-dessus).
Pour les sessions en distanciel, un lien de connexion à la visioconférence vous sera communiqué ultérieurement.

Contacts

Contenu des ateliers, inscriptions, organisation : Céline Faure (MSH Lyon St-Etienne) - celine.faure@msh-lse.fr

Site web : www.msh-lse.fr/semaine-data-shs-2022

Twitter : [@mshlyon](https://twitter.com/mshlyon) #SemaineDataSHS2022

