

Semaine Data SHS

Traiter et analyser des données en sciences humaines et sociales

Du 11 au 15 décembre 2023

à la MSH Lyon St-Étienne, Lyon 7^e et St-Étienne

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 11 au 15 décembre 2023 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.

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, exploiter (2 modules) ;
- L'**anonymisation** en question : méthodologie, protection des données (2 modules) ;
- Outils de **visualisation cartographique** (1 module) ;
- Créer et organiser des données : la **création de BDD** (1 module) ;
- **Données juridiques** : retour d'expériences (2 modules).



■

LUNDI 11 DÉCEMBRE 2023

Données de recherche : trouver, produire, exploiter

Module 1 – ELIPSS, un panel Internet pour les sciences sociales

Intervenants : **Mathieu Olivier**, chargé de production de données (Sciences Po Paris, CDSP - UAR828), **Alexia Ricard**, chargée de production de donnée (CNRS, CDSP - UAR828)

Le panel ELIPSS, conçu dès 2012 dans le cadre de l'EquipEx DIME-SHS, est un panel probabiliste qui repose actuellement sur un échantillon d'environ 2 600 personnes, invitées à répondre chaque mois par Internet à des questionnaires conçus par des équipes de recherche en sciences humaines et sociales. Cette présentation sera l'occasion de revenir sur l'origine du projet et d'exposer les caractéristiques actuelles du panel, mais aussi de mettre en avant son utilisation par la communauté scientifique. Ce sera l'occasion de rappeler les modalités de l'accès aux données, et de présenter des usages et travaux concrets réalisés à partir de ces données.

Horaires : 10h30-12h – **Lieu :** MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/345516?lang=fr>

Module 2 – Propriétés sociales et sens de la pratiques des pratiquant.e.s de yoga : enquête quantitative & qualitative

Intervenants : **Marie-Carmen Garcia**, professeur en sociologie (Univ. Lyon 1, L-VIS), **Brice Lefèvre**, maître de conférences (Univ. Lyon 1, L-VIS)

Il s'agira dans cette présentation d'aborder les données quantitatives concernant le sport et plus particulièrement les deux dernières enquêtes nationales de statistique publique ad hoc liées aux activités physiques et sportives (ENPPS). Les démarches globales de l'ENPPS 2010 et de l'ENPPS 2020 – disponibles sur Quetelet-Progedo-Diffusion – seront détaillées et permettront de discuter de la prise en main des données et de certains outils statistiques (analyse de tableaux croisés, régressions, typologies...).

Pour illustrer l'articulation quanti/quali, nous partirons à travers l'étude du yoga de l'approche quantitative permise par l'ENPPS 2020 (typologie des différents pratiquant.e.s) pour glisser vers une approche plus qualitative (observations participantes et entretiens semi-directifs).

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

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/244319?lang=fr>

■

MARDI 12 DÉCEMBRE 2023

L'anonymisation en question : méthodologie, protection des données

Module 1 – La protection des données personnelles dans les établissements d'enseignement supérieur et de recherche

Intervenant : **Nicolas Sanzano**, délégué à la protection des données (ENS de Lyon)

Le RGPD (Règlement général sur la protection des données) encadre les règles de protection des données à caractère personnel et fixe de nouveaux droits pour les personnes physiques, ainsi que de nouvelles obligations pour les responsables de traitement.

La présentation se déclinera en plusieurs parties : d'abord un retour sur les principes et notions qui entourent le RGPD, suivi des questions de mise en conformité par rapport à la réglementation. Nous verrons ensuite comment se répartit la responsabilité entre les acteurs, puis comment l'on peut transmettre des données à un tiers. Enfin, nous nous intéresserons aux droits individuels des personnes concernées par le RGPD, et nous conclurons sur les spécificités des données de la recherche publique

Objectifs :

- présenter le corpus juridique et les principes fondamentaux de la réglementation relative à la protection des données ;
- expliquer la mise en conformité des établissements et définir les notions clés
- comprendre les sanctions encourues ;
- apprendre à sécuriser les flux sensibiliser aux droits des personnes concernées.

Horaires : 10h-12h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

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

Module 2 – Pseudonymisation et anonymisation dans le cadre de traitement de données personnelles en SHS

Intervenant : **Thomas Soubiran**, ingénieur d'études en production et analyse de données (CNRS, CERAPS - UMR 8026)

Dans le cadre de traitement de données personnelles, le responsable du traitement (RdT) doit mettre en œuvre les mesures techniques et organisationnelles appropriées pour s'assurer et être en mesure de démontrer la conformité des traitements qu'il réalise. Le RdT doit ainsi, en autres obligations, garantir la confidentialité des données collectées. Pour ce faire, le RdT peut notamment avoir recours à la pseudonymisation des données. Dans d'autres cas, le RdT peut procéder à leur anonymisation qui n'a aucun caractère obligatoire et n'est qu'une possibilité parmi d'autres pour protéger les droits et libertés des personnes.

Cette présentation reviendra donc sur la mise en œuvre de la pseudonymisation et de l'anonymisation dans des traitement de DCP en SHS ainsi que sur les difficultés propres à la seconde. La suppression de données directement identifiantes n'est, dans les faits, généralement pas suffisante pour garantir la confidentialité des données, les autres renseignements collectés recelant encore suffisamment d'information pour singulariser les personnes ou apparier les fichiers. Des altérations plus ou moins substantives des données peuvent alors se révéler nécessaires. Mais cela parfois au détriment de l'utilité même des données et sans pour autant toujours garantir un niveau de protection suffisant. À ce titre, on verra que d'autres mesures sont envisageables, notamment dans le cadre de la mise à disposition de données.

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

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

MERCREDI 13 DÉCEMBRE 2023

Outils de visualisation cartographique

Module 1 – Premiers pas dans la cartographie numérique

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

Ces dernières années, la recherche en sciences humaines a vu une utilisation de plus en plus importante de la cartographie numérique comme méthode d'investigation transversale dans les recherches en histoire, littérature, linguistique, études culturelles, etc. Les Humanités Spatiales nous ont permis d'interpréter, ou de réinterpréter, les phénomènes culturels à travers de des nouveaux supports d'analyse. L'objectif de cette intervention est de présenter comment il devient possible, par le biais d'une visualisation de données géoréférencées, de proposer une nouvelle façon de lire et d'analyser les questions disciplinaires à travers une approche interdisciplinaire. Toutefois, une partie importante de cet exposé sera consacrée à une réflexion autour des outils nécessaires à la création numérique des modèles cartographiques. L'objectif consiste à présenter ce que peuvent être les différences entre un modèle de mapping réalisée avec des applications ou du code et les intégrations possibles entre les cartes numériques et d'autres visualisations de données comme, par exemple, l'analyse de réseau.

Chaque participant.e devra ouvrir un compte CARTO au moins deux semaines avant l'atelier.

Compte totalement gratuit pour les enseignants, chercheurs et étudiants.

Instructions : <https://docs.carto.com/faqs/carto-for-education>

Horaires : 9h-13h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/351727?lang=fr>

■

JEUDI 14 DÉCEMBRE 2023

Créer et organiser des données : la création de BdD

Module 1 – Construire et explorer : comment Heurist simplifie la création de bases de données pour la recherche

Intervenant : Vincent Ventresque, ingénieur d'études (Univ. St-Étienne, TRIANGLE - UMR 5206 et IHRIM - UMR 5317), Fabienne Vial-Bonacci, ingénieure d'études (CNRS, IHRIM - UMR 5317)

L'accélération du nombre de projets nécessitant la collecte et l'organisation de données de format hétérogène a fait émerger un besoin croissant en matière de mise en place de bases de données. Ce besoin s'est manifesté par la nécessité d'un outil à la prise en main facile, permettant notamment aux chercheurs de travailler sans être restreints par un éventuel contexte de ressources d'ingénierie limitées.

Heurist, système de gestion de base de données créé en 2005 par le professeur Ian Johnson, initialement pour la discipline archéologique, s'est avéré constituer l'une de ces réponses aux besoins ainsi exprimés par l'ensemble des disciplines en sciences humaines et sociales. Intégrant des fonctionnalités de travail collaboratif, de visualisation de données textuelles, géographiques, ou encore temporelles, mais aussi de mise en ligne simple mais effective des données, Heurist est avant tout un outil offrant une flexibilité suffisante pour assurer une forme d'autonomie dans le travail des chercheurs.

Cette présentation étayera ainsi ses qualités en matière de flexibilité et de simplicité d'usage, tout en portant un regard critique alimenté par les retours d'expérience de projets de recherche. La communication s'appuiera notamment sur le projet ANR Rey's Ornament Image investigation (ROIi) porté par les laboratoires IHRIM et Hubert Curien, et ayant utilisé Heurist pour la conception d'une large base de données d'ornements dans le cadre d'une analyse des pratiques éditoriales de Marc Michel Rey (1720-1780).

Horaires : 13h30-15h30 – **Lieu :** Université St-Étienne, Campus Tréfilerie (salle du conseil)

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

■

VENDREDI 15 DÉCEMBRE 2023

Données juridiques : retour d'expériences

Module 1 – Données juridiques : jurimétrie, méthodes empiriques, saisie dynamique

Intervenants : Sarah Schmalian, doctorante (Univ. Grenoble-Alpes, CRJ), Vincent Rivollier, maître de conférences (Univ. Savoie Mont-Blanc, CMW - UMR 5283), Delphine Neven, doctorante (Univ. Grenoble-Alpes, ODENORE)

Retours d'expériences sur des problématiques spécifiques relatives aux données juridiques.

9h15-10h : **Le recours à la jurimétrie dans l'analyse du contentieux administratif**

La présente contribution vise à présenter l'intérêt, la mise en œuvre ainsi que les questionnements – pratiques et théoriques – du recours à la jurimétrie pour l'analyse du contentieux administratif. Plus spécifiquement, nos travaux portent sur l'évolution des devoirs et des pouvoirs assignés au juge compétent dans le cadre d'une voie de droit ayant vocation à permettre, dans l'urgence, la protection des libertés fondamentales.

La contribution suivra les étapes suivantes :

- 1/ Présentation de l'objet de recherche : le référé-liberté
- 2/ L'objectif du recours à la jurimétrie : une posture de rupture épistémologique
- 3/ La détermination des variables : un processus inductif et déductif
- 4/ Quid de la fiabilité des résultats.

10h-11h : **Les données et les méthodes empiriques utilisées dans le cadre de la recherche en matière de réparation du dommage corporel**

La présente proposition vise à dresser la synthèse des données et des méthodes empiriques mises en œuvre dans un champ juridique particulier : le droit de la réparation du dommage corporel, c'est-à-dire la réparation des préjudices subis par les victimes d'accidents de la circulation, d'accidents médicaux, d'accidents du travail ou de maladies professionnelles, et de tout type d'accident.

11h20-12h : **Saisie dynamique de données juridiques : la réalisation d'une enquête ethnographique dans le cadre d'un travail de thèse en droit**

La recherche vise à questionner la dimension juridique des facteurs de non-recours au droit à vivre dans un logement social adaptés identifiés à partir de l'expérience de personnes pour qui le fait de bénéficier d'un tel logement est conçu, dans le droit positif, comme un droit subjectif et opposable (bénéficiaires potentiels du dispositif juridique « DALO »).

Cet objet est appréhendé en deux temps :

1/ identifier, à partir de données collectées dans le cadre d'une enquête ethnographique, des « facteurs de non-recours » au logement social. Plus exactement, il s'agit d'identifier ce qui constitue des obstacles au bénéfice effectif d'un logement social adapté pour les personnes pour qui le fait de bénéficier d'un tel logement est conçu, dans le droit positif, comme un droit subjectif et opposable.

2/ questionner la dimension juridique des facteurs de non-recours préalablement identifiés. Concrètement, pour chaque facteur, il s'agit de chercher à opérer des corrélations entre ce qui, dans les faits, apparaît comme un obstacle au bénéfice effectif de la prestation et le contenu du « droit ». Plus précisément, il s'agit de se demander dans quelle mesure le « droit » engendre ou participe à engendrer l'obstacle identifié. Une fois la causalité établie – si elle est pertinente –, je suis alors en mesure d'identifier quelle(s) est/sont la/les strate(s) du droit à l'origine de l'obstacle et de réfléchir aux voies, juridiques ou non, susceptibles de conduire au dépassement de l'obstacle.

Horaires : 9h15-12h – **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)
[retransmission en duplex entre la MSH Lyon St-Étienne et la MSH Alpes]

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

Module 2 – Données juridiques : analyse automatisée du langage, justice algorithmique

Intervenants : **Isabelle Sayn**, directrice de recherche en droit (CNRS, CMW - UMR 5283) ; **Romain Rambaud**, **Alya Hafsaoui**, **Caroline Bligny**, **Frédérique Letué**, **Marie-José Martinez**, **Gilles Bisson**, **Annette Casagrande** et **Thomas Lebarbé** (Univ. Grenoble-Alpes, projet JADE)

Retours d'expériences sur des problématiques spécifiques relatives aux données juridiques.

14h-15h : **Quels apports de l'analyse automatisée du langage naturel pour la connaissance de la production des juridictions**

Les juridictions françaises produisent plus de 3 millions de décisions par an, pour l'essentiel des décisions de première instance et d'appel. Si les décisions des juridictions supérieures sont connues en raison de leur participation à la construction de la règle de droit, les décisions des juridictions dites inférieures sont longtemps restées non explorées, sous réserve de travaux de recherche spécifiques, le coût que représente de telles explorations étant très lourd. Dans ces conditions, que peut apporter l'analyse automatisée du langage naturel pour la connaissance de ces décisions ?

Au terme d'un travail collectif, il apparaît que les analyses automatisées sont insuffisantes pour associer les résultats de la décision à des déterminants légaux ou a-légaux, l'extraction de données fiables tant peu satisfaisante. La diffusion de tels outils et leur utilisation par les professionnels du droit impose par conséquent de s'interroger sur le contrôle qualité que le droit devrait ou pourrait prévoir.

15h-16h30 : **Les données juridiques dans le cadre d'une démarche de justice algorithmique : le projet JADE (Justice algorithmique des élections)**

La contribution fait état des problématiques liées aux données juridiques dans le cadre du projet JADE, Justice algorithmique des élections, qui est un projet de recherche IRGA/IDEX visant à l'élaborer une justice algorithmique des élections politiques. Ces questionnements rejoignent ceux que l'on peut avoir dans le cadre d'une réflexion plus générale sur la justice algorithmique en tant que méthode interdisciplinaire appliquée.

La contribution suivra les étapes suivantes :

- 1) Présentation générale du projet JADE (Romain Rambaud)
- 2) L'élaboration technique des bases de données du projet JADE (Caroline Bligny)
- 3) Les problématiques juridiques des bases de données du projet JADE : données de recherche vs. données sensibles (Alya Hafsaoui)
- 4) L'enrichissement des données : la question de l'annotation (Thomas Lebarbé)
- 5) L'utilisation mathématique des données : description voire prédiction (Marie-Jo Martinez, Frédérique Letué étant je crois indisponible)
- 6) L'apprentissage à partir des données : deep learning et/ou explicabilité (Gilles Bisson et Annette Casagrande).

Horaires : 14h-16h30– **Lieu** : MSH, Lyon 7^e (salle André Bollier, rdc)
[retransmission en duplex entre la MSH Lyon St-Étienne et la MSH Alpes]

Inscription > formulaire en ligne : <https://enquetes.msh-lse.fr/index.php/681695?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, Mardi, Mercredi, Vendredi : MSH Lyon St-Etienne, 14 av. Berthelot, Lyon 7^e (salle André Bollier)

Jeudi : Univ. Jean Monnet, Campus Tréfilerie, Maison de l'université, 10 rue Tréfilerie, St-Etienne (salle du conseil)

Inscriptions

Merci de vous inscrire, avant le 5 décembre 2023, 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-2023

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

