

# Journées dataSHS 2018

Traiter et analyser ses données en Sciences Humaines et Sociales

**Du 10 au 14 décembre 2018**

**5 journées-ateliers de formation**

**Lieu : Maison Européenne des Sciences de l'Homme et de la Société,  
2 Rue des Canoniers, Lille - Métro Lille Flandres**

*Ouvert aux étudiants (niveau Master 2), aux doctorants, chercheurs, enseignants-chercheurs et ingénieurs des laboratoires affiliés à la MESHS. Les ateliers sont basés sur des logiciels libres et de prise en main accessible.*

## **LUNDI 10 DECEMBRE : NETTOYAGE ET HARMONISATION DE BASES DE DONNEES**

**Intervenant :** Julien BOELAERT, MCF en science politique, CERAPS, Univ. Lille

**Description :** L'une des premières difficultés que peuvent présenter les bases de données, en particulier en présence de données texte (noms, phrases, ...), sont les petites scories qu'elles contiennent : fautes d'orthographe, conventions de notations différentes... Avant tout traitement statistique, il faut alors nettoyer ou harmoniser la base. L'objectif de cette journée est précisément de familiariser les participants aux outils de nettoyage de données. Nous traiterons d'une part les expressions régulières (regex), un langage de requêtes qui permet de faire des recherches complexes, d'extraire des informations et de nettoyer des données de texte. Nous passerons pour cela par des fonctions R, mais les regex sont un langage commun à beaucoup d'autres programmes de traitement de données. D'autre part, nous emploierons le logiciel OpenRefine, qui permet d'harmoniser de façon efficace des données texte.

## **MARDI 11 DECEMBRE : CADRE JURIDIQUE ET GESTION DES DONNEES DE LA RECHERCHE**

Matin : La réglementation relative aux données à caractère personnel en sciences sociales

**Intervenant :** Thomas SOUBIRAN, Ingénieur d'études en analyse de données, CERAPS, Univ. Lille

**Description :** Comment mettre en conformité un projet de recherche en sciences humaines et sociales au regard de la réglementation applicable aux données personnelles ? Cet atelier rappellera les notions clés et les grands principes de la réglementation. Il en développera ensuite les implications pratiques à partir d'exemples concrets correspondant aux différentes étapes de la recherche (collecte, analyse, conservation).

Après-midi : Principes et enjeux de la gestion des données de la recherche

**Intervenante :** Marie PUREN, Chercheuse post-doctorale, LARHRA

**Description :** L'atelier vise à présenter les principes et les enjeux de la gestion des données de la recherche pour ensuite permettre aux participant(e)s de prendre en main les problématiques liées à la création d'un plan de gestion des données à l'aide de l'outil DMP Opidor. Des jeux de données leur seront fournis, mais les participant(e)s qui le souhaitent sont vivement incité(e)s à travailler sur leurs propres jeux de données.

## MERCREDI 12 DECEMBRE : ANALYSE DE DONNEES TEXTUELLES (TRANSCRIPTIONS D'ENTRETIENS)

**Intervenant :** Alex ALBER, MCF en sociologie, Univ. Tours

**Description :** L'atelier visera à familiariser les participants avec le maniement des corpus audio-textuels dans le logiciel Sonal et présentera divers outils d'analyse, notamment lexicométriques. La matinée sera consacrée à l'organisation des données (importation des fichiers texte/son/vidéo, ajout de méta-données, retranscription, thématisation). L'après-midi proposera différents exercices d'extraction et d'analyse à partir des matériaux.

## JEUDI 13 DECEMBRE : ANALYSE DE RESEAUX EN SCIENCES HUMAINES ET SOCIALES

**Matin :** *Introduction à la théorie de l'analyse de réseaux et à la théorie des graphes*

**Intervenant :** Fabien ELOIRE, MCF en sociologie, CLERSE, Univ. Lille

**Après-midi :** *Atelier basé sur Cytoscape, logiciel pour l'analyse et la visualisation de réseaux*

**Intervenant :** Guillaume BRYBAERT, Ingénieur de recherches en bioinformatique, UGSF, Univ. Lille

**Description :** L'atelier permettra aux participants de découvrir et pratiquer les fonctionnalités du logiciel Cytoscape. Les participants auront l'occasion de travailler sur l'import et le filtrage des données, la visualisation et l'analyse de réseaux (mesures de centralité, etc.), l'export de données. L'atelier sera conclu par une discussion autour des données des participants.

Liens utiles : [tutoriel Cytoscape en français](#) ; [tutoriel Cytoscape en anglais](#).

## VENDREDI 14 DECEMBRE : CARTOGRAPHIE (AVEC MAGRIT)

**Intervenants :** Nicolas LAMBERT, Ingénieur en cartographie thématique et développement web, RIATE ; Ronan YSEBAERT, Ingénieur en traitement et analyse de bases de données, RIATE

**Description :** Magrit est une application de cartographie en ligne qui permet réalisation de représentations cartographiques usuelles telles que les cartes choroplèthes ou les cartes en symboles proportionnels. D'autres types de cartes sont aussi proposés: lissages, carroyages, discontinuités, anamorphoses, etc. Magrit entend couvrir dans un même environnement logiciel, l'ensemble de la chaîne de traitement, de la donnée géographique à la carte d'édition. Des jeux de données seront fournis par les formateurs. Toutefois, les participant(e)s peuvent venir avec leurs propres données et travailler dessus.

Participation libre dans la limite de 15 places par atelier

 **[S'inscrire aux ateliers : pudl@univ-lille1.fr](mailto:pudl@univ-lille1.fr)**  
Merci de préciser les ateliers qui vous intéressent

Les journées dataSHS sont organisées par l'**axe Humanités numériques de la MESHS** et la **Plateforme universitaire de données de Lille**. Cette action de formation bénéficie du soutien financier du Conseil régional des Hauts-de-France dans le cadre du contrat de plan État-Région ISI-MESHS.

