

Journées data-shs 11-15 décembre 2017, MESHS, Lille.

Les journées data-shs: traiter et analyser ses données en sciences humaines et sociales proposent un choix d'ateliers d'initiation aux méthodes de traitement et d'analyse de données numériques en sciences humaines et sociales. Plus particulièrement, les ateliers ont deux objectifs:

- 1) faire acquérir aux participants des notions théoriques sur des méthodes d'analyse et de représentation de données numériques en sciences humaines et sociales afin que ceux-ci puissent les adapter à leurs propres travaux;
- 2) initier les participants via des travaux pratiques au traitement, à la structuration et à l'analyse de données.

Les ateliers sont ouverts aux étudiants (niveau Master 2 et doctorat), aux jeunes chercheurs, aux enseignants-chercheurs et ingénieurs des laboratoires affiliés à la MESHS. La participation est libre mais se fait sur inscription, dans la limite de 15 places par atelier.

Les ateliers sont basés sur des logiciels libres et de prise en main accessible.

Programme provisoire

Lundi 11 décembre : préparer et préserver ses données de recherche

Cette première journée est consacrée au travail préalable permettant d'analyser ses données dans de bonnes conditions. Elle s'organise autour de trois interventions complémentaires.

Comment organiser ses fichiers de recherche, les nommer, les structurer ?

Intervenant : Sofia Papastamkou (CNRS, MESHS) (sous réserve)

Comment préparer ses données avant le traitement et l'analyse ?

Intervenant : Pierre Mounier (INHA)

Atelier basé sur le logiciel Openrefine (<http://openrefine.org/>)

OpenRefine (anciennement Google Refine) est un outil puissant pour manipuler de grandes quantités de données, permettant des opérations de nettoyage en masse, l'exploration rapide de jeux de données, la conversion d'un format à l'autre, l'application de transformations cellulaires de base et avancées, le filtrage et le partitionnement des données avec des expressions régulières, l'extraction d'entités nommées sur des champs de texte intégral. Son usage, en tant qu'outil libre et gratuit, s'impose donc rapidement et cet atelier offrira un aperçu large de ses possibilités à l'appui de cas concrets.

Introduction pratique aux obligations déontologiques et juridiques en matière de collecte et de traitement de données en SHS

Intervenant : Thomas Soubiran (CERAPS, Université Lille 2)

Comment intégrer les contraintes de la loi relative à l'informatique, aux fichiers et aux libertés (lil) dans un projet de recherche en sciences humaines et sociales ? Cet atelier rappellera les notions clefs et les grands principes d'application de la loi. Il en développera ensuite les implications pratiques à partir d'exemples concrets correspondant aux différentes étapes de la recherche (collecte, analyse, conservation).

Mardi 12 décembre : analyse statistique de données textuelles avec TXM

Intervenant : Serge Heiden (ENS, Lyon)

Il s'agit d'apprendre à utiliser TXM dans ses principales fonctions d'analyse textométrique. Cela comprend une familiarisation avec l'organisation générale et le fonctionnement de l'interface, ainsi qu'avec l'approche textométrique. On se focalise sur les fonctionnalités les plus utilisées (Lexique, Index, Concordance, Spécificités, Cooccurrence), avec une présentation à la fois technique et méthodologique : à quoi sert cette fonctionnalité, comment préparer et paramétrer le calcul, comment interpréter les résultats, comment récupérer les tableaux et graphiques pour publication. L'atelier est finalement très complet et peut intéresser à la fois des "vrais" débutants comme des débutants "avancés" et même des autodidactes débrouillés, d'autant que son contenu peut s'adapter aux participants présents en fonction des informations qu'ils nous auront transmises au moment de l'inscription.

<http://textometrie.ens-lyon.fr/>

Mercredi 13 décembre : cartographie avec Magrit

Intervenants : Matthieu Viry et Nicolas Lambert (RIATE, Paris)

Magrit est une application de cartographie en ligne. Magrit propose la 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.

<http://riate.cnrs.fr/?p=5698>

Jeudi 14 décembre : analyse de réseaux en shs

Intervenants : Fabien Eloire (CLERSE, Lille 1), Guillaume Brysbaert (UGSF, Lille 1)

La journée proposera une introduction à la théorie de l'analyse de réseaux et à la théorie des graphes.

Elle se poursuivra par un atelier basé sur Cytoscape, logiciel pour l'analyse et visualisation de réseaux. 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.

<http://www.cytoscape.org/>

Vendredi 15 décembre : cartographie du web avec Hyphe

Intervenant : Benjamin Ooghe-Tabanou et Maxime Crépel (médialab, Sciences Po)

Publié sous la forme d'un logiciel libre, Hyphe a été conçu pour offrir aux chercheurs et étudiants un outil de création et nettoyage de corpus web reposant sur un crawler orienté pour la recherche. Les utilisateurs sont accompagnés par une méthodologie pour construire leur corpus web de manière à la fois granulaire et flexible avec des principes de curation simples.

<http://hyphe.medialab.sciences-po.fr/>

La séance proposera tout d'abord une présentation générale de la méthode et de l'outil, ainsi qu'un cas d'usage de Hyphe.

Le cas présenté sera la création d'un "datascape" dans le cadre d'un projet réalisé au médialab entre 2015 et 2017 ayant consisté en la réalisation d'une interface d'exploration d'un corpus Hyphe pour l'analyse quali-quantitative de controverses autour des enjeux de data privacy.

<http://tools.medialab.sciences-po.fr/privacy/>

Dans un second temps il sera proposé aux participants d'essayer Hyphe concrètement lors d'un atelier pratique. Les participants sont invités à venir équipés d'un ordinateur portable pour pouvoir prendre en main le logiciel durant cet atelier.