

SEMAINE DATA-SHS

Traiter et analyser les données en Sciences Humaines et Sociales

Semaine du 6 au 10 Décembre 2021

**Centre de Données Socio-Politiques (Sciences Po / CNRS)
Plateforme Universitaire de Données
des Grands Moulins (Université de Paris)**

Le Centre de données socio-politiques (CDSP) et la Plateforme Universitaire de données (PUD) des Grands Moulins s'unissent pour organiser une session de conférences et d'ateliers s'inscrivant dans le cadre de la semaine Data-SHS, un événement national coordonné par la TGIR PROGEDO.

Le Centre de données socio-politiques (Sciences Po/CNRS)

Le Centre de données socio-politiques est une unité mixte de service dépendant de Sciences Po et du CNRS, localisée à Paris. Il a pour missions principales la production, la documentation et la mise à disposition d'enquêtes en sciences sociales à destination de la communauté scientifique. Le CDSP contribue à la recherche en termes de méthodes, qu'il s'agisse de données quantitatives, des enquêtes qualitatives ou des méthodes mixtes.

Par ses missions et ses pratiques, il s'inscrit pleinement dans le cadre du développement de la science ouverte. Le CDSP s'inscrit également dans le périmètre d'action de la TGIR Progodo : il est l'un des trois centres diffuseurs de données sur le portail Quetelet-PROGEDO Diffusion et, depuis fin 2020, héberge une plateforme universitaire de données (PUD).

La PUD des Grands Moulins, une nouvelle plateforme de services (Université de Paris)

La PUD des Grands Moulins fait partie intégrante du *Centre des Humanités Numériques*, plateforme de services créée par la Direction générale déléguée des bibliothèques et musées (DGDBM) d'Université de Paris, à destination des publics de la Faculté *Sociétés & Humanités*. La mission principale de la PUD des Grands Moulins est d'accompagner les communautés de recherche (chercheur.e.s, enseignant.e.s-chercheur.e.s, doctorant.e.s, ingénieur.e.s, étudiant.e.s en master...) du site géographique parisien dans l'accès et l'utilisation de données quantitatives de la statistique publique (enquêtes nationales ou internationales, recensements...) qui offrent de larges possibilités d'analyse en sciences humaines et sociales.

Des données pour la recherche

La semaine Data SHS propose un ensemble d'interventions sur la production de données en sciences humaines et sociales et les méthodes de traitement et d'analyse de ces données :

- La présentation de données d'enquêtes publiques utiles pour la communauté de recherche en SHS
- Des travaux mobilisant des données quantitatives, portés par les chercheur.e.s, enseignant.e.s-chercheur.e.s, doctorant.e.s et étudiant.e.s de Sciences-Po et Université de Paris, ou d'autres structures d'Île de France
- Des méthodes de description et des pratiques de valorisation de différents types de données
- Des ateliers autour d'outils permettant de traiter les données : atelier d'analyse textuelle, atelier d'anonymisation...

Une organisation multi-site et mixte

Cette année, le déroulé de la semaine alternera des séances en présence ou à distance (visio-conférence, streaming), sur plusieurs sites parisiens. Pour les séances en présentiel, nous encourageons les publics qui le peuvent à se rendre sur place pour plus de convivialité et pour faciliter les échanges.

La participation est gratuite mais l'inscription obligatoire. En raison des mesures en vigueur, le « pass sanitaire » est par ailleurs nécessaire pour assister en présentiel aux événements.

Les inscriptions sont ouvertes jusqu'au 3 décembre : [Formulaire d'inscription](#)

Envie de participer à la semaine Data-SHS lilloise, lyonnaise, marseillaise, grenobloise... ?

La semaine Data-SHS est inscrite dans le plan national de la TGIR PROGEDO : elle est proposée au même moment dans chacune des [14 Plateformes universitaires de Données](#) (dites PUD). La simultanéité de cet événement dans les différents sites universitaires souligne l'inscription et la vocation locales des activités de chaque PUD.

Lundi 6 Décembre

Accès aux données en SHS – via les diffuseurs de Quetelet-Progedo-Diffusion

UNIQUEMENT EN VISIO-CONFERENCE (Lien communiqué après inscription)

10h00 – 11h30

Introduction | Alina **DANCIU** (CDSP)

Le portail Quetelet PROGEDO Diffusion : les données disponibles

Valentin **BRUNEL** (CDSP)
Ariana **CAPORALI** (INED, DataLab)
Quentin **GALLIS** (CDSP)
Erik **ZOLOTOUKHINE** (ADISP, Progedo)

Panel ELIPSS

UNIQUEMENT EN VISIO-CONFERENCE (Lien communiqué après inscription)

14h00- 15h30

Introduction | Lucie **MARIE** (CDSP)

Le Panel ELIPSS : catalogue, évolution et fonctionnement actuel

Marion **ELIE** (CDSP)
Mathieu **OLIVIER** (CDSP)

Mardi 7 Décembre

Réutilisation des données de l'histoire orale (sous réserve)

UNIQUEMENT EN PRESENTIEL (Sciences Po, Salle 900, Bâtiment C, 9 rue de la chaise, 75007 Paris)

10h00 – 12h00

Introduction | Guillaume **GARCIA** (CDSP)

L'analyse secondaire des entretiens en sociologie et en science politique, une situation radicalement différente de l'histoire orale ?

Guillaume **GARCIA** (CDSP)

Panorama de l'archivage et de la réutilisation des témoignages oraux en histoire

Florence **DESCAMPS** (EPHE)

Mémoires d'un ministère : une analyse secondaire de l'enquête orale du Service d'histoire de l'éducation

Bénédicte **GIRAULT** (Héritages)

Des métadonnées ouvertes pour une bibliométrie au service de la recherche

PRINCIPALEMENT EN PRESENTIEL (Salles des Conférences – Campus Saint-Germain-des-Prés, 45 rue des Saints-Pères, 75006 Paris) ; Possibilité de suivre à distance (lien streaming communiqué après inscription)

14h00

Bibliométrie et science ouverte : enjeux en jeu et en BD

Introduction | Christophe **PION** (DGDBM, Université de Paris)

Cyril **HEUDE** (DRIS, Sciences Po)

14h30

Baromètre Science ouverte d'Université de Paris : lumière sur les SHS

Maxence **LARRIERU** (DGDBM, Université de Paris)

15h00

Identifier « sans identifiants » l'utilisation des enquêtes de la statistique publique par les chercheurs français

Antoine **LAVACQUERY** (Université de Lyon III)

Mercredi 8 Décembre

Panels et cohortes : une approche méthodologique

UNIQUEMENT EN PRESENTIEL (Sciences Po, Salle 900, Bâtiment C, 9 rue de la chaise, 75007 Paris)

10h00 – 12h00

Introduction | Mathieu **OLIVIER** (CDSP)

Le panel ELIPSS

Blazej **PALAT** (CDSP)

Le panel de Caen

Claire **BIDART** (LEST)

L'enquête européenne SHARE

Louis **ARNAULT** (LEDa-LEGOS, PSL)

Données de santé pour la recherche en SHS : le cas des cohortes épidémiologiques

PRINCIPALEMENT EN PRESENTIEL (Salles des Conférences – Campus Saint-Germain-des-Prés, 45 rue des Saints-Pères, 75006 Paris) ; Possibilité de suivre à distance (lien streaming communiqué après inscription)

Introduction | Charlotte **DION**
(PUD des Grands Moulins)

14h00

Constances, une cohorte épidémiologique ouverte à la communauté des chercheur·e·s

Marie **ZINS** (INSERM)

14h15

Elfe, une cohorte interdisciplinaire de 18 000 enfants, portée par l'INED et l'INSERM

Xavier **THIERRY** (INED)

14h30

Faire la sociologie des changements de pratiques avec une cohorte épidémiologique : l'apport de Constances

Marie **PLESSZ** (INRAE, CMH)

14h45

Variations sociales et temporelles dans les représentations de l'épidémie de Covid19.

Alexis **SPIRE** (CNRS, IRIS, EHESS)

15h – 16h30

Table ronde : Quels atouts et potentiels, quels freins et limites à l'utilisation des données de cohortes de santé pour les chercheur·e·s en sciences humaines et sociales ?

Marie **PLESSZ** (INRAE, CMH)
Alexis **SPIRE** (CNRS, IRIS, EHESS)
Xavier **THIERRY** (INED)
Marie **ZINS** (INSERM)

Jeudi 9 Décembre

UNE JOURNEE EN DEUX VOLETS

Volet 1. La [PUD de Nanterre](#) et la PUD des Grands Moulins s'associent pour organiser un séminaire et un atelier dédiés à l'analyse textuelle.

Volet 2. Le CDSP organise un atelier sur l'anonymisation des données et un séminaire sur l'ouverture du code.

Volet 1.

Traiter des données issues de la presse écrite

Conférences et atelier : analyses textuelles en SHS et Humanités Numériques

UNIQUEMENT EN PRESENTIEL (MSH Mondes – Université Paris Nanterre, Bâtiment Max Weber, 8 allée de l'université, 92000 Nanterre)

Introduction | Brian **CHAUVEL** (PUD de Nanterre)
Charlotte **DION** (PUD des Grands Moulins)

9h30

ISOCO : Produire des séries statistiques à partir de données textuelles

Patrice **BAUBEAU**
(Université Paris Nanterre – IDHES)

9h50

Typologie d'outils d'analyses textuelles en SHS

Mathilde **DE SAINT-LÉGER**
(Université Paris Nanterre – MoDyCo)

10h10

Visualiser les données textuelles en Humanités numériques

Suzanne **MPOULI**
(Centre des Humanités Numériques (CHN),
Université de Paris)

10h30

RetroNews : une base enrichie de métadonnées pour constituer des corpus en SHS

Philipp **LEU** (Responsable des opérations, BnF
– Partenariats)

10h50 - 11h20

Discussion

11h20 - 14h30

Pause repas

14h30 - 17h30 **Atelier UNIQUEMENT EN PRESENTIEL** (BnF, Salle de l'Aquarium, Hall Est, Paris 75013) : **inscription obligatoire**

Découverte et maniement d'outils d'analyse textuelle : Calliope, IramuteQ, TXM, Gargantext

Patrice **BAUBEAU**,
Brian **CHAUVEL**, Charlotte **DION**, Philipp **LEU**,
Daphné **MATHELIER**, Suzanne **MPOULI**,
Mathilde **DE SAINT-LEGER**

Volet 2.

Atelier : Anonymisation d'un jeu de données

UNIQUEMENT EN PRESENTIEL (Sciences Po, Salle 921, Bâtiment C, 9 rue de la chaise, 75007 Paris) : **inscription obligatoire**

Introduction | Quentin **GALLIS**, Valentin **BRUNEL** (CDSP)

10h00 – 11h30

Atelier d'anonymisation des données quantitatives et qualitatives :
Bring your data !

Valentin **BRUNEL** (CDSP)
Paul **COLIN** (CDSP)
Alina **DIANCIU** (CDSP)
Guillaume **GARCIA** (CDSP)
Quentin **GALLIS** (CDSP)
Nawale **LAMRINI** (DPO, Sciences Po)
Marion **LEHMANS** (DPO, Sciences Po)

En quoi l'ouverture du code participe-t-elle à la science ouverte ?

UNIQUEMENT EN VISIO-CONFERENCE (*Lien communiqué après inscription*)

Introduction | Paul **COLIN** (CDSP)

14h00 -15h30

WPSS, une participation du CDSP à SSHOC

Quentin **AGREN** (CDSP)
Geneviève **MICHAUD** (CDSP)

L'humain derrière le software

Tam Kien **DUONG** (Etalab)

Table ronde : les dimensions du code ouvert

Quentin **AGREN** (CDSP)
Tam Kien **DUONG** (Etalab)
El Hassan **GARGEM** (CDSP)
Geneviève **MICHAUD** (CDSP)



Vendredi 10 Décembre

Comment « se mettre au quanti » ? Retours d'expériences

PRINCIPALEMENT EN PRESENTIEL (Salle 479F du bâtiment la Halle aux Farines, Esplanade Pierre Vidal-Naquet, 5 rue Thomas Mann, 75013 Paris) ; Possibilité de suivre à distance (lien zoom communiqué après inscription)

Introduction | Pierre **PISTRE** (*PUD des Grands Moulins*)

10h00 – 11h30

« Se mettre au quanti » pour répondre à une question de recherche

Irène **BERTHONNET**
(*LADYSS, Université de Paris*)

Initier la communauté de recherche à l'analyse d'enquêtes avec R et RStudio : retours d'expériences d'analyse-R et des webin-R

Joseph **LARMARANGE**
(*IRD, Ceped, Université de Paris*)

Des bases de données « simplifiées » à destination des étudiants

Quentin **GALLIS** (*CDSP*)

UN APRES-MIDI EN DEUX VOLETS

Volet 1. La PUD des Grands Moulins organise un atelier dédié aux débutants sur la prise en main d'une enquête publique avec des outils simples.

Volet 2. Le CDSP organise une session dédiée à l'European Social Survey.

Volet 1.

Atelier : Prendre en main facilement une enquête nationale INSEE

UNIQUEMENT EN PRESENTIEL (Salle informatique de la Bibliothèque des Grands-Moulins, 2^{ème} étage, 5 rue Thomas Mann, 75013 Paris) : **inscription obligatoire**

14h00 – 17h00

Explorer les données en accès libre de l'Enquête Emploi avec des outils simples (tableur, application « click-bouton »)

Pierre **PISTRE** (*PUD des Grands Moulins*)
Charlotte **DION** (*PUD des Grands Moulins*)

Volet 2.

Session : L'enquête European Social Survey

UNIQUEMENT EN VISIO-CONFERENCE (*Lien communiqué après inscription*)

14h00 – 15h30

Introduction | Quentin **GALLIS**, Valentin **BRUNEL** (*CDSP*)

L'enquête European Social Survey

Nicolas **SAUGER** (*CDSP*)
Agnalys **MICHAUD** (*CSDP*)