

SEMAINE DATA-SHS

Organisée par la PUD-AMU

Du 05 au 09 décembre 2022

MMSH, Aix en Provence

Organisée simultanément par toutes les Plateformes Universitaires de Données du réseau PROGEDO, cette semaine propose une série de présentations et d'ateliers, afin de sensibiliser la communauté scientifique à la culture des données quantitatives en sciences humaines et sociales.

Cela se déroulera intégralement en présentiel.
Chaque journée sera précédée d'un accueil café à 9h30

Lundi : Présentation des services autour des données à AMU et exemples d'utilisation de données de la statistique publique

Mardi : Initiation à l'analyse de données avec JASP et JAMOVI puis R

Mercredi : Le cycle de vie de la donnée

Jedi : Journée spéciale Egypte : infrastructures, données et travaux

Vendredi : L'analyse textuelle : exemples et démonstrations

Lieu : Salle Germaine Tillion, MMSH, 5 Rue du château de l'horloge, Aix-en-Provence.

Inscriptions : Obligatoire - Avant le 01 décembre 2022, à l'adresse suivante :
<https://evento.renater.fr/survey/semaine-data-shs-gn1ohxol>

Informations et programme : <http://pud.mmsh.univ-aix.fr> ou QR code ci-dessous

Contact : pud-amu@univ-amu.fr ~ 04 42 52 49 19



10h-12h

Présentation des services autour des données à l'université d'Aix-Marseille

- Introduction --> **Isabelle Blöss-Widmer et Clément De Belsunce**
- Présentation de la PUD-AMU --> **Clément De Belsunce**
- Présentation de CEDRE --> **Trardi Youssef**
- Présentation du SCD --> **Isabelle Gras et Léa Mosnier**
- Présentation du Collectif Data Provence
--> **Marie-Laure Trémélo, Nathalie Leménager et Clément De Belsunce**

Alors que la question des données est au centre de la recherche, cette matinée est l'occasion de présenter les principales structures qui s'y rapportent au sein de l'Université d'Aix-Marseille.

Que ce soit au sein de la MMSH ou non, plusieurs services sont au cœur des laboratoires de recherche pour permettre aux chercheurs de gagner en efficacité et en compétence.

Enfin, trois ingénieurs vous présenteront leur initiative de lancement d'un collectif autour des données.

14h-16h

Utilisation de données issues de la statistique publique

- Présentation du nouveau portail Quetelet-Progedo --> **Clément De Belsunce**

Dans la perspective de toujours améliorer le service rendu aux communautés de recherche, PROGEDO a soutenu la rénovation du portail de données Quetelet-Progedo-Diffusion. Après 2 ans de travail, le nouveau catalogue unifié est en ligne. Le portail de commande reste pour l'instant distinct du catalogue. Il sera intégré d'ici sous peu sous forme de panier validable, pour rendre l'expérience utilisateur toujours plus fluide.

- Collecte de données en économie de la santé à partir d'un panel de médecin généraliste --> **Bruno Ventelou**

La présentation se fera en deux temps, nous présenterons d'abord un outil d'observation des pratiques en médecine générale, né en PACA à l'initiative d'acteurs régionaux en 2002, et devenu un outil national depuis 2008, mis en œuvre par le ministère de la santé et déposé sur Quetelet. Dans un second temps, nous présenterons des travaux de recherches en économie de la santé et en sciences sociales qui utilisent ces données.

- Un exemple de modélisation de la décroissance démographique en France avec l'utilisation des données de produit sur mesure --> **Dieynaba Ndour**

A partir de sources statistiques non disponibles dans l'offre standard, l'Insee propose de réaliser un produit sur mesure (PSM) à la demande des chercheurs. Cette communication sera l'occasion de montrer un exemple depuis la "Base historique des données harmonisées des recensements de la population 1968-2017" et montrer comment les données obtenues vont être utilisées dans le cadre d'une thèse.

- Les enquêtes mobilité certifiées Cerema : Usages, potentiels et exploitations de la donnée publique de mobilité quotidienne
--> **Lionel Kieffer, Constance Bouard et Sarah Joubaire**

L'EMC2 est la principale base de données publique sur les déplacements quotidiens en France. Or, la demande de données sur les déplacements des populations n'a jamais été aussi forte. L'objectif de cette communication sera d'exposer la diversité d'informations des EMC2, leurs applications opérationnelles mais aussi leurs limites et contraintes d'une collecte de données quantitatives sur de grands territoires.

Initiation à l'analyse de données à l'aide d'un logiciel

10h-12h

- Initiation à l'analyse de données avec JASP et JAMOVI
--> **Bruno Davier**

Cet atelier a pour but de présenter JASP et JAMOVI. Ce sont deux logiciels libres d'analyse des données offrant une interface utilisateur très accessible. Ils permettent de réaliser les analyses principalement utilisées en sciences humaines. Sur la base de données illustratives, les analyses descriptives courantes seront présentées suivies d'analyses inférentielles classiques (comparaison de moyennes, corrélations, régression, analyses factorielles...). Enfin, les analyses plus avancées comme l'analyse en réseau disponible sur JASP ou les modèles linéaires généraux accessibles par le biais de JAMOVI seront abordés.

14h-16h

- Initiation à l'analyse de données avec R --> **Clément De Belsunce**

Cet atelier est une introduction à l'utilisation du logiciel de traitement statistique R. Elle s'adresse donc aux personnes n'ayant aucune expérience dans ce domaine. Les principales fonctionnalités seront présentées notamment l'importation et la visualisation d'une base de données.

Le but étant de s'appropriier le logiciel et de traiter des données de manière superficielle. Selon l'intérêt du public, une formation plus complète sera proposée prochainement.

Ces ateliers sont limités à 12 personnes. Ce sont donc les premiers inscrits qui pourront assister à cette session.

Merci de vous inscrire si vous avez la certitude d'assister à ces sessions.

Le cycle de vie de la donnée

10h-12h

- La donnée AMU - au cœur de la Science Ouverte --> **Fabien Borget**

Au cours de cette présentation, je présenterai la politique et la stratégie science ouverte d'AMU en contextualisant dans l'espace national, européen et mondial. L'aspect données de la Recherche sera ensuite développé et une partie de la présentation se focalisera sur les enjeux de ces données et l'ensemble des services qu'Aix-Marseille Université est en train de développer pour l'ensemble de ses communautés.

- Du laboratoire à l'Université : l'enseignant-chercheur face à la Science Ouverte --> **Clara Galliano**

La Science Ouverte s'insère de plus en plus dans les pratiques de recherche et d'évaluation dans toutes les disciplines scientifiques. Les chercheurs sont amenés à modifier leurs pratiques en intégrant certaines possibilités : publication dans des revues en accès libre, dépôt de la production dans une archive ouverte, dépôt de données dans des entrepôts de données... Notre communication propose de revenir sur ce sujet en présentant plusieurs méthodologies utilisées pour étudier l'enseignant-chercheur face au mouvement de la SO, en croisant notamment plusieurs terrains.

- À la découverte du Plan de Gestion de données : présentation et utilisation --> **Léa Mosnier**

Bien qu'il soit devenu un document incontournable dans le paysage scientifique, le PGD peut poser beaucoup de questions lorsque l'on est amené à l'utiliser pour un projet de recherche : Qu'est-ce que c'est ? A quoi sert-il ? Quels sont les bénéfices ? Quel contenu ? Autant de questions qui trouveront leurs réponses durant cette intervention.

14h-16h

- Données FAIR en archéologie : exemples dans les projets numériques du Centre Camille Jullian --> **Stéphanie Satre et Bruno Baudoin**

A travers les différents projets réalisés avec des équipes de recherche du laboratoire (bases de données de recherche, bibliothèque numérique) nous présenterons les principes mis en œuvre pour répondre aux exigences de la science ouverte en mettant en lumière les partenariats qui nous ont sensibilisés à ces bonnes pratiques et accompagnés dans cette démarche d'ouverture et de pérennisation des données.

- Représentations animales depuis la Préhistoire : un modèle d'édition collaborative pour diffuser et valoriser des données de recherche --> **Stéphane Renault**

Cette présentation s'attachera à expliciter notre démarche qui conduit à proposer un modèle d'édition collaborative et évolutive à travers une plateforme de publication créée pour différents projets de représentations animales durant la Préhistoire. Reposant sur le CMS Omeka S, ce modèle participe à la FAIRisation des données de l'archéologie qui tient, au même titre que d'autres initiatives, à lier étroitement publications et données en s'inscrivant dans une démarche de science ouverte.

- Archiver ses codes-sources et logiciels : SoftWare Heritage --> **Julien Caugant**

Les logiciels et les codes sont aujourd'hui des outils utilisés pour le traitement, l'analyse et la visualisation des données, ou sont également l'objet de recherche dans des domaines spécialisés comme le Machine et le Deep Learning. Dans le cadre de la Science Ouverte et de la reproductibilité de la recherche, il apparaît essentiel de pouvoir conserver, référencer et partager les codes et logiciels que nous produisons. SoftWare Heritage a pour ambition de relever ce défi.

- ORRSO - Outil de Recherche de Ressources sur la Science Ouverte - pour promouvoir et encourager les pratiques de science ouverte dans la communauté du DRIIHM --> **Marie-Laure Trémélo, Corinne Pardo, Mathieu Massaviol**

Informier et former la communauté scientifique aux bénéfices du partage et de l'ouverture des données pour tendre vers un changement progressif des pratiques de gestion et de diffusion de la donnée sont des objectifs portés par le groupe Data-Driihm (Dispositif de Recherches Interdisciplinaires sur les Interactions Hommes-Milieus). Dans cette démarche, il est important de prendre en compte aussi bien les chercheurs qui croulent sous les informations de leurs établissements voire au-delà, que ceux isolés qui ne bénéficient d'aucun accompagnement sur cette thématique. Ainsi, Data-Driihm se positionne comme "aiguilleur de la science ouverte" dans le cadre d'un projet ANR SO-DRIIHM. À ce titre, ORRSO - Outil de Recherche de Ressources en Science Ouverte - a été imaginé pour faciliter l'accès aux ressources pédagogiques sur la science ouverte développées par les universités et organismes de recherches.

Ce sont cet outil et ses objectifs à court et moyen termes que nous vous présenterons dans cette intervention.

Journée spéciale Egypte : infrastructures, données et travaux

10h-12h

- Introduction et contextualisation --> **Isabelle Blöss-Widmer et Elena Ambrosetti**
- De la collecte à l'exploitation des données en Egypte : retour d'expériences du CEDEJ autour du recensement et du Covid. (visio) --> **Hala Bayoumi**

Le CEDEJ, Centre français de recherche en Egypte, œuvre à la collecte, l'analyse et la valorisation de données sur l'Egypte. Cette présentation se propose de revenir sur ce qui illustrent à la fois les potentiels de recherche offerts et les réalisations aussi bien en termes de collecte (sur les données liées au COVID), que d'analyse (autour de la réalisation d'un atlas) que de valorisation (par la mise en disposition de données censitaires géolocalisées).

- CartoMundi : plateforme de valorisation du patrimoine cartographique --> **Romain Suarez**

Dédié à la valorisation en ligne du patrimoine cartographique, CartoMundi est à la fois un site web ouvert à tous les publics et une plateforme collaborative destinée aux professionnels qui conservent des documents cartographiques.

CartoMundi s'adresse à tous les établissements qui souhaitent valoriser leur collection et proposer l'accès à la visualisation de leurs documents cartographique. Ce service est placé sous l'égide d'Aix-Marseille Université et piloté par la MMSH - à Aix-en-Provence.

14h-16h

- EGYLandscape : cartographie historique de l'Égypte rurale pré-moderne --> **Nicolas Michel**

Le projet ANR-DFG EGYLandscape Land and Landscapes in Mamluk and Ottoman Egypt, 13th-189th Centuries, élabore un webSIG de l'Égypte rurale pré-moderne. Cet outil permettra de localiser sur les cartes de l'époque contemporaine (de 1800 à nos jours) les toponymes issus des cadastres médiévaux.

- L'observatoire démographique de la Méditerranée : les enjeux de la comparaison transnationale à échelle fine --> **Isabelle Blöss-Widmer et Sébastien Oliveau**

Les comparaisons internationales restent peu nombreuses dans les sciences sociales françaises en regard de la diversité des sources disponibles. On pourrait s'attarder sur les raisons de cette situation, mais nous souhaitons ici plutôt mettre en lumière l'exigence comparative à l'origine de Demomed, l'observatoire démographique de la Méditerranée, et de son outil de diffusion et de visualisation des données. Nous présenterons plusieurs résultats scientifiques qui illustrent la richesse de ces approches ainsi que les enjeux portés par la mise à disposition en ligne des données sous-jacentes.

- La diffusion spatiale de la baisse de la fécondité en Egypte (1950-2006) --> **Yoann Doignon**

L'Égypte comptait près de 27 millions d'habitants en 1960. Depuis, sa population a été quasiment multipliée par 4 (presque 100 millions d'habitants en 2018). Les gouvernements égyptiens ont depuis longtemps considéré cette forte croissance démographique (et la surpopulation associée) comme un défi, dont ils identifient la fécondité comme la principale cause. L'objectif est de tracer l'évolution de la distribution spatiale de la fécondité en Egypte entre 1960 et 2006, pendant le déclin de la fécondité, en utilisant les données du recensement.

- Les défis de la cartographie à des niveaux fins en Egypte : l'apport de la base de données IPUMS-international --> **Elise Leveque**

Cette présentation porte sur le processus de création d'un fond de carte inédit de l'Égypte à un niveau fin. Nous révélerons les dessous de cette démarche d'harmonisation, qui a permis de mettre en valeur des données de recensement de la base IPUMS-International.

L'analyse textuelle : exemples et démonstrations

10h-12h30

- Introduction à la textométrie avec le logiciel Iramuteq
--> **Natacha Souillard**

En vue de fournir des éléments d'introduction généraux à la textométrie, la communication sera scindée en deux temps. Nous commencerons par un exposé des principes et concepts fondamentaux de l'analyse statistique de données textuelle. Puis, à partir de corpus de recherche, nous proposerons une présentation de quelques fonctionnalités et analyses permises par le logiciel Iramuteq (LERASS, Université de Toulouse) notamment la Classification Hiérarchique Descendante (CHD) selon la méthode décrite par M.Reinert (1983,1986,1991).

14h-16h

- Analyse lexicométrique d'un corpus de presse francophone sur l'entrée des saisonniers agricoles étrangers en Europe, en contexte de pandémie --> **Béatrice Mesini et Sylvain Bausson**

Dans le cadre du Projet FINCA « Frontière, Immobilisation et Néolibéralisme en temps de Covid dans l'Agriculture nous avons mis en place une recherche booléenne portant sur divers items. Un premier corpus constitué entre mars et mai 2020, comprend 562 articles de presse qui détaillent, par le biais des « nuages de mots » et de l'outil Voyant Tools, les dynamiques d'un secteur agricole galvanisé pour éviter toute rupture des chaînes de production alimentaire. Un deuxième corpus étendu de mars 2020 à mars 2021 composé de 1 300 articles, a fait l'objet d'un traitement avec Iramuteq, permettant une analyse lexicométrique globale (type de presse, pays, flux...) mais aussi focale, centrée sur les items saillants de nos publications collectives (discrimination, vulnérabilité...).

- Quelques principes de base sur l'utilisation de N-Vivo
--> **Nathalie Boutin**

Il s'agira de présenter rapidement l'application N-Vivo, et, à partir d'un retour d'expérience en tant qu'utilisatrice (pour ses recherches), et d'accompagnement de la recherche (formation sur N-Vivo d'équipes) il s'agira d'expliquer en quoi N-Vivo se différencie des autres applications (ex. Iramuteq), quels sont les grands principes à retenir pour une meilleure utilisation, et à partir de ses travaux de recherche, des exemples de traitements seront présentés pour illustrer l'utilisation de N-Vivo en rapport des principes évoqués.