



# PLATEFORME UNIVERSITAIRE DE DONNEES DE L'UNIVERSITE DE POITIERS (PUD-UP)

## ❖ Programme annuel d'activités

### Table des matières

I.	Introduction.....	2
1.	Qu'est-ce qu'une PUD?.....	2
2.	Qu'est-ce que PROGEDO?.....	2
3.	La PUD-UP.....	3
4.	Les principales missions de la PUD-UP.....	3
II.	Programme annuel d'activités de la PUD-UP (proposition).....	5
1.	À la rencontre des utilisateurs.....	5
2.	Formation.....	5
a.	Accompagnement personnalisé.....	6
b.	Ateliers destinés aux étudiants en Master 2 et en Doctorat.....	6
i.	Ateliers d'analyse quantitative de données.....	6
ii.	Ateliers curiosité et découverte.....	7
c.	Séminaires destinés aux étudiants et aux enseignants-chercheurs.....	8
i.	La journée RGD.....	8
ii.	La journée PGD (Plan de gestion de données).....	8
iii.	Les journées des grandes enquêtes.....	9
3.	Semaine data SHS.....	9
4.	Semaine de l'intelligence artificielle.....	10

**Responsable scientifique de la DUP-UP :** Liliane BONNAL

**Ingénieur en charge de l'PUD-UP :** Germano VERA RUZ



## I. INTRODUCTION

### 1. Qu'est-ce qu'une PUD ?

- **Composante régionale** de la Très Grande Infrastructure de Recherche (TGIR) en Production et Gestion de Données en Sciences Sociales (PROGEDO).
- Plateforme de **soutien** aux recherches quantitatives en sciences sociales.

### Répartition des PUD au niveau national



### 2. Qu'est-ce que PROGEDO ?

- **Objectifs** ?
  - ◇ Promouvoir la culture des données pour la recherche en SHS,
  - ◇ Mettre en place la politique publique des données en SHS.
- **Rôle** ?
  - ◇ Garantir et faciliter à la communauté scientifique :

- # L'accès à des données fiables et de qualité,
- # Une meilleure exploitation de ces données.

### 3. La PUD-UP

- Portée et hébergée par la **MSHS-Poitiers** (institutionnellement rattachée à l'UP & CNRS)
- Sous la responsabilité scientifique de **Liliane Bonnal** (professeur des Universités en Sciences Économiques) et animée par **Germano Vera Cruz** (ingénieur en production, traitement et analyse de données).
- Public visé : Enseignants, chercheurs, doctorants, étudiants en Master
- Domaine disciplinaire : Sciences Humaines et Sociales

### 4. Les principales missions de la PUD-UP

- Aider les personnes souhaitant travailler sur des données en :
  - ◇ **Identifiant les données pertinentes** pour leur problématique : source, type, temporalité, complexité, localisation
  - ◇ **Récupérant ces données** auprès des organismes collecteurs ou diffuseurs
  - ◇ **Fournissant l'aide à la compréhension** des données récupérées : métadonnées, échantillonnage, etc.
- Aider les personnes à travailler sur ces données en :
  - ◇ Fournissant un **appui pour l'utilisation des méthodes statistiques et des logiciels d'analyse**,
  - ◇ **Organisant des formations** (de groupe ou personnalisées) sur les méthodes statistiques, le traitement et l'analyse des données.

- Informer en :
  - ◇ Faisant une veille sur les **formations promouvant l'utilisation des données quantitatives** (écoles d'été, journées d'études, etc.),
  - ◇ Organisant les événements scientifiques pouvant **faire connaître les enquêtes quantitatives et/ou les résultats de recherche** basés sur les données quantitatives secondaires ou collectées soi-même.

## **II. PROPOSITION D'UN PROGRAMME ANNUEL D'ACTIVITES**

### **1. A la rencontre des utilisateurs**

#### **Description :**

Il s'agit, au début de chaque année universitaire, d'aller à la rencontre des potentiels utilisateurs des services fournis par la DUP-UP / PROGEDO et leurs présenter les activités prévues dans l'année universitaire en cours. Nous rencontrerons les étudiants de chaque Master 2 en sciences humaines et sociales (dans leurs salles de cours respectives), les étudiants en doctorat et les enseignant-chercheurs en sciences humaines et sociales (dans leurs laboratoires respectifs).

#### **Dates :**

A la fin du mois de septembre et au début du mois d'octobre.

### **2. Formation**

Nous proposons trois catégories de formations : (1) l'accompagnement personnalisée ; (2) des ateliers spécifiquement destinés aux étudiants en Master 2 et en Doctorat ; (3) des séminaires destinés à tous les étudiants, enseignants-chercheurs et chercheurs intéressés.

Pour la deuxième catégorie, les formations dureront 5h (Matin :9h-12h ; Après-midi :14h-16h) ; la capacité sera limitée à 10-15 personnes maximum. Pour la troisième catégorie, les formations dureront 3h (de 9h à 12h) ; le nombre de places peuvent être limité.

#### **◇ L'accompagnement personnalisée**

#### **Descriptions :**

Accompagnement/formations personnalisées, sur demande. Par exemple, en cas de sollicitation par un étudiant pour traiter d'un sujet qui concerne l'acquisition et le traitement de données quantitatives, nous pouvons lui proposer un rendez-vous pour travailler avec lui.

**Dates :**

Tout le long de l'année universitaire

**Accompagnateur :**

Germano Vera Cruz

◇ **Les ateliers**

≠ **Ateliers d'analyse quantitative de données destinés aux étudiants en Master 2 et en Doctorat**

**Description :**

Formations collectives sur les différentes méthodes d'analyse quantitative de données. L'utilisation des données du catalogue Quetelet-Progedo-Diffusion sera privilégiée. Ces formations ont été mises en place en collaboration avec les écoles doctorales pour permettre aux doctorants qui le souhaitent de valider des crédits équivalents à 5h de formation.

**Dates, lieu, créneau horaire, thématiques :**

Les dates seront fixées en coordinations avec les écoles doctorales et les laboratoires hébergés à la MSHS. Pour chaque atelier, une annonce avec la date, le lieu, la thématique sera envoyée aux intéressés potentiels à travers les écoles doctorales, les labos et les responsables des Masters 2. Chaque atelier durera 5h (Matin :9h-12h ; Après-midi :14h-16h) et la capacité limitée à 10-15 personnes maximum.

**La liste non exhaustive des formations :**

1. Construction et mise en ligne d'un questionnaire avec le logiciel Sphinx ;
2. Analyse des données avec Excel : Prise en main du logiciel, les bases du langage VBA, les graphiques, analyse descriptive, tableaux croisés dynamiques, etc. ;
3. Initiation à l'analyse statistique avec SPSS ;
4. Initiation à l'analyse statistique avec R ;
5. Analyses statistiques avancées avec R : Analyses factorielles (ACP, ACM, AFM), Multilevel linear models (MLM), Analyses de survie, Analyses de séquences, Analyses spatiales.

**Formateur :**

Germano Vera Cruz et d'autres intervenants.

### ≠ **Ateliers curiosité et découverte**

#### **Description :**

Il s'agit de faire découvrir (une présentation et une démonstration des fonctionnalités) aux participants un outil (existant sous une forme gratuite) de recueil, traitement ou d'analyse de données.

#### **Dates, lieu, outil, créneau horaire :**

Ces ateliers auront lieu une fois par mois. Les dates précises seront fixées en coordinations avec les écoles doctorales et les laboratoires hébergés à la MSHS. Pour chaque atelier, une annonce avec la date, le lieu, l'outil à découvrir sera envoyée aux intéressés potentiels à travers les écoles doctorales, les labos et les responsables des Masters 2. Chaque atelier durera 5h (Matin :9h-12h ; Après-midi :14h-16h) et la capacité limitée à 10-15 personnes maximum.

#### **Formateur/animateur :**

Germano Vera Cruz.

#### **Liste non exhaustive d'outils à faire découvrir :**

1. Limesurvey (pour construire et mettre en place des questionnaires en ligne) ;
2. Knime (un workflow pour faire du machine learning) ;
3. Tableau software (pour résumer et représenter les données sous une forme graphique et/ou cartographique) ;
4. Microsoft Power BI (pour résumer et représenter les données sous une forme graphique et/ou cartographique) ;
5. Google Studio (pour résumer et représenter les données sous une forme graphique et/ou cartographique) ;
6. Trello (pour faire une planification d'activités dans le cadre d'un projet professionnel).
7. Google Analytics (pour l'analyse de la performance d'un site-web) ;
8. MySQLWorkBench (pour la construction d'une base de données relationnelle) ;
9. Here (pour faire du géocodage de données) ;
10. Beautifulsoup (un outil python pour le web-scraping).

### ◇ **Les séminaires**

### ≠ La journée RGPD

#### **Description :**

Pendant la première heure, il sera proposé aux participants une intervention récapitulant les principes généraux véhiculés par le Règlement Général de Protection de Données (RGPD). Ensuite, un atelier permettra aux participants de se familiariser avec le RGPD à travers des cas concrets d'application. À cet effet, le site web de la CNIL sera utilisé (<https://atelier-rgpd.cnil.fr/>)

#### **Dates, lieu, créneau horaire :**

Cette journée aura lieu une fois par année universitaire, au mois d'octobre. La date précise sera fixée en coordinations avec les écoles doctorales et les laboratoires hébergés à la MSHS. Pour chaque journée, une annonce avec la date et le lieu sera envoyée aux intéressés potentiels à travers les écoles doctorales, les labos et les responsables des Masters 2. Chaque journée durera 3h (9h-12h).

#### **Intervenants :**

Germano Vera Cruz et un autre intervenant spécialiste du RGPD.

### ≠ La journée PGD (Plan de gestion de données)

#### **Description :**

Cette journée inclura les interventions suivantes : (a) Plan de gestion de données : qu'est-ce que c'est et pour quoi faire ? ; (b) présentation de la plateforme DMP-Opidor (il s'agit de la plateforme CNRS fournissant de l'aide et des conseils aux chercheurs pour l'élaboration des PGD) et des services fournis par celle-ci ; (c) présentation d'un plan de gestion de données effectué par un enseignant-chercheur dans le cadre d'une étude en SHS financée par un organisme public français tel que la ANR ; (d) présentation d'un plan de gestion de données effectué par un enseignant-chercheur dans le cadre d'une étude en SHS financée par un organisme européen.

#### **Dates, lieu, créneau horaire :**

Cette journée aura lieu une fois par année universitaire, au mois de janvier/février. La date précise sera fixée en coordinations avec les écoles doctorales et les laboratoires hébergés à la MSHS. Une annonce avec la date et le lieu sera envoyée aux intéressés potentiels à travers les écoles doctorales, les labos et les listes de diffusion de l'UP. Chaque journée durera 3h (9h-12h).



**Intervenants :**

Des intervenants internes ou externes à l'UP.

**≠ Les journées les grandes enquêtes****Description :**

Ces journées visent à mettre en lumière les grandes enquêtes nationales, européennes et internationales réalisées (avec des fonds publics) dans le domaine des sciences humaines et sociales. Chaque journée correspondra à une enquête précise (e.g., l'enquête SHARE, l'enquête ESS, etc.). De 9h à 10h, un intervenant présentera l'enquête (les questionnaires associés, la population cible, les méthodes de recueil des données, les principales variables). De 10h à 11h, d'autres intervenants présenteront les études réalisées à partir des données de l'enquête en question. De 11h à 12h, un intervenant utilisera le logiciel de traitement et d'analyse de données SPSS pour montrer les données et faire quelques analyses descriptives. Le but est de montrer aux participants les possibilités d'exploitation des données de l'enquête à l'ordre du jour et de leur donner envie d'explorer ou d'exploiter eux-mêmes les mêmes données. Si possible, ces journées seront coorganisées avec des plateformes universitaires de données / PROGEDO localisées dans d'autres universités.

**Dates, lieux, nom de l'enquête, créneau horaire :**

Ces journées auront lieu une fois par semestre. Les dates précises seront fixées en coordinations avec les écoles doctorales et les laboratoires hébergés à la MSHS. Pour chaque journée, une annonce avec la date, le lieu et le nom de l'enquête à l'ordre du jour sera envoyée aux intéressés potentiels à travers les écoles doctorales, les labos et les responsables des Masters 2. Chaque journée durera 3h (9h-12h).

**Intervenants :**

Des chercheurs internes et/ou externes à l'UP.

**3. Semaine data SHS****Description :**

Dans le cadre de la semaine de la data, pendant cinq jours successifs, nous proposons d'organiser des conférences / des retours d'expériences autour de sujets liés à la production et

à l'exploitation de données. L'enjeu de cette semaine est d'attirer l'attention des participants (étudiants, enseignant-chercheurs, chercheurs) sur les données des grandes enquêtes (e.g., ESS, Share, GGP, ISSP, EVS, Cohorte Constance, etc.), sur les données du catalogue Quetelet-Progedo-Diffusion, sur des dispositifs de collecte, centralisation et partage de données tels qu'Elipss, DFIH, ethmigsurveydatahub.eu, etc., pour qu'ils puissent savoir où trouver les données susceptibles de les intéresser et comment les exploiter dans leur travail de recherche.

**Date :**

La deuxième semaine de décembre.

**Intervenants :**

Des invités internes et externes à l'UP (étudiants en doctorat, enseignant-chercheurs, autres professionnels du secteur public ou privé).

#### **4. Semaine de l'intelligence artificielle**

**Description :**

Il s'agit, pendant 5 jours consécutifs, de faire découvrir (présentation et démonstration) aux participants les outils, les méthodes et les modèles d'intelligence artificielle (en langage python) utilisés pour recueillir, traiter et analyser des données. Ici, aussi, nous proposerons l'utilisation de données du catalogue PROGEDO.

**Présentateur :**

Germano Vera Cruz.

**En particulier, nous montrerons :**

Modèles :

- a) Machine learning (l'apprentissage automatique) ;
- b) Deep learning (l'apprentissage profond) ;
- c) Reinforcement learning (l'apprentissage par renforcement) ;
- d) Supervised learning (l'apprentissage supervisé) ;

e) Unsupervised learning (l'apprentissage non supervisé).

Outils :

- a) Artificiel Neural Network (ANN) - les réseaux de neurones artificiels classiques ;
- b) Convolutional Neural Network (CNN) – les réseaux de neurones convolutifs ;
- c) Recurrent Neural Network (RNN) – les réseaux de neurones récurrents.

Libraries python :

- a) Nltk ;
- b) Scikit-learn ;
- c) Tensorflow ;
- d) Keras ;
- e) PyTorch.

**Dates :**

Première semaine d'avril.

**FIN**