



Maison Méditerranéenne  
des Sciences de l'Homme  
Aix-Marseille Université

# Plateforme Universitaire de Données PUD-AMU



[pud-amu@univ-amu.fr](mailto:pud-amu@univ-amu.fr)



<https://pud.mmsh.univ-aix.fr/>

# PUD-AMU

(1<sup>er</sup> avril 2018)

## Référente scientifique

Isabelle Blöss-Widmer

## Ingénieur d'études

Clément De Belsunce

Mohammed Zahaf



## Localisation

Maison Méditerranéenne des Sciences de l'Homme  
Bureau 107, bâtiment C.

# TGIR PROGEDO

## Produire

Grandes enquêtes  
Bases de données

## Gérer

Archiver  
Diffuser

## Promouvoir

Faire connaître  
Accompagner

- Assurer la production des volets français des grandes enquêtes internationales
- Assurer la gestion des enquêtes quantitatives en SHS
- Assurer l'archivage et la diffusion des données issues de la statistique publique
- Assurer l'accès aux bases de données quantitatives en SHS
- Assurer la promotion des données quantitatives en SHS pour développer leur utilisation



# SHARE

Survey of Health, Ageing  
and Retirement in Europe  
50+ in Europe



European  
Social  
Survey



ISSP



European  
*Values*  
Study



GGP

## Les enquêtes internationales

---

L'enquête européenne SHARE (Survey of Health, Ageing and Retirement in Europe) construit une base de données longitudinale, multidisciplinaire et internationale concernant plus de 120 000 Européens âgés de 50 ans et plus et ayant pour objectif d'analyser les problèmes économiques et sociaux liés au vieillissement de la population.

8 vagues d'enquêtes depuis 2012

L'European Social Survey (ESS) est un dispositif d'enquête comparative et répétée sur les attitudes des Européens. Cette enquête est réalisée à partir de la passation tous les deux ans d'un questionnaire auprès d'échantillons nationaux représentatifs dans près de trente pays.

La France participe au dispositif « European Social Survey » depuis la première édition de ce programme de recherche en 2002.



L'international Social Survey Programme (ISSP) est un programme d'enquête sociale internationale permettant des comparaisons transnationales et dans le temps. Non thématisée, cette enquête permet de questionner chaque année un sujet important pour la recherche en sciences sociales, en complément d'une série de questions sur les caractéristiques sociodémographiques des enquêtés.

La France participe depuis 1996.

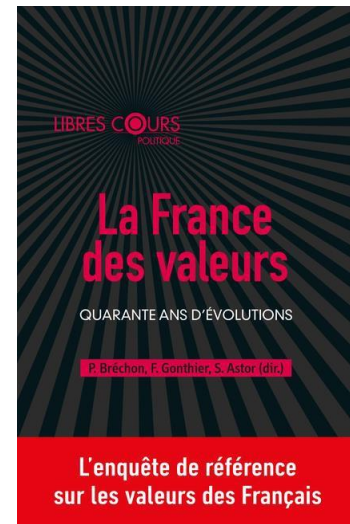


<http://www.issp-france.fr/>



L'European Values Study est une enquête à grande échelle, internationale et longitudinale sur les comportements, opinions et valeurs des européens. La European Values Study (EVS) constitue aujourd'hui le programme de recherche le plus complet sur l'évolution des valeurs.

Elle est devenue une référence en matière de comparaison internationale. C'est l'enquête qui existe depuis le plus longtemps : tous les neuf ans depuis 1981.





Initié en 2001 par les Nations Unies, le programme a été mis en place pour éclairer les principaux défis sociodémographiques et économiques des pays européens (et les politiques publiques qui tentent d'y répondre).

Sur la base de panels représentatifs de la population âgée de 18 à 79 ans, l'enquête GGP se déroule sur trois vagues répétées à intervalle de trois ans.

En France, l'enquête ERFI (Etude des relations familiales et intergénérationnelles) a collecté les données françaises de GGP.

**E**  **IPSS**

beQ<sup>u</sup>ali

**D** **FIH**  
Données Financières Historiques

Des dispositifs nationaux

---

# TGIR PROGEDO

<b>Produire</b>	<b>Gérer</b>	<b>Promouvoir</b>
Grandes enquêtes Bases de données	<b>Archiver</b> <b>Diffuser</b>	Faire connaître Accompagner

- Assurer la production des volets français des grandes enquêtes internationales
- Assurer la gestion des enquêtes quantitatives en SHS
- Assurer l'archivage et la diffusion des données issues de la statistique publique
- Assurer l'accès aux bases de données quantitatives en SHS
- Assurer la promotion des données quantitatives en SHS pour développer leur utilisation



L'ICPSR conserve et diffuse les données de plus de **9.500 enquêtes** en sciences sociales et comportementales. Il héberge 21 collections spécialisées de données dans les domaines de l'éducation, du vieillissement, de la justice, etc.



Le LIS acquiert des ensembles de données sur le revenu, la richesse, l'emploi et les données démographiques de nombreux pays à revenu élevé et intermédiaire, les harmonise pour permettre des comparaisons transnationales et les rend accessibles au public dans deux bases de données : la **Luxembourg Income Study Database (LIS)** et la **Luxembourg Wealth Study Database (LWS)**.



L'ADISP diffuse des enquêtes et bases de données produites par l'INSEE, plusieurs services statistiques ministériels et des institutions de recherche publiques. Les données diffusées sont accessibles gratuitement à l'ensemble de la communauté scientifique.

- Le catalogue compte aujourd'hui plus de 1.300 références.
- L'ADISP permet aussi d'obtenir des **produits sur mesure** (tableaux construits à la demande à partir d'une vingtaine de sources de l'INSEE (recensement, état civil, revenus, emploi, entreprises, etc.).
- L'ADISP donne accès aux **projections démographiques** de l'INSEE (Omphale).
- L'ADISP donne accès au **distancier** METRIC (MEsure des TRajets Inter-Communes / Carreaux).

<http://quetelet.progedo.fr/>

archives de données issues  
de la statistique publique  
**disp**

**QUETELET**  
**PROGEDO**  
DIFFUSION



**SciencesPo**  
CENTRE DE DONNÉES SOCIO-POLITQUES

**ined**   
INSTITUT NATIONAL  
D'ÉTUDES DÉMOGRAPHIQUES

CASD 



Coordonner et harmoniser  
les pratiques au niveau européen



# TGIR PROGEDO

## Produire

Grandes enquêtes  
Bases de données

## Gérer

Archiver  
Diffuser

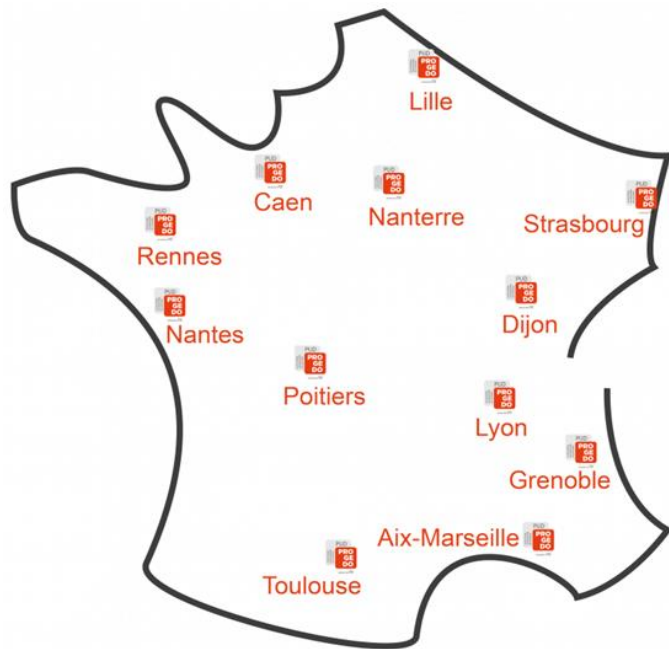
## Promouvoir

Faire connaître  
Accompagner

- Assurer la production des volets français des grandes enquêtes internationales
- Assurer la gestion des enquêtes quantitatives en SHS
- Assurer l'archivage et la diffusion des données issues de la statistique publique
- Assurer l'accès aux bases de données quantitatives en SHS
- Assurer la promotion des données quantitatives en SHS pour développer leur utilisation

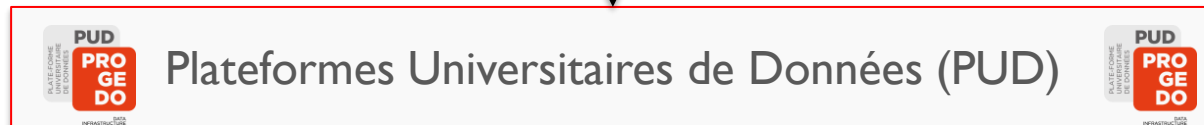


# TGIR PROGEDO



**PRO**duction  
et **GE**stion  
des **DO**nnées

DATA  
INFRASTRUCTURE



# La PUD-AMU au service de quels publics ?

CHERCHEURS  
ENSEIGNANTS-CHERCHEURS

Articles, montage de projets ...  
Fournir des jeux de données  
quantitatives supports aux cours ...

PERSONNELS ADMINISTRATIFS  
ET TECHNIQUES

Formations ...

DOCTORANTS

Thèse, articles, formations ...

ETUDIANTS DE MASTER

Mémoires, projet de thèse ...

# Les missions de la PUD-AMU

- **Accompagnement individuel**
- Accompagnement collectif
- Promotions des bonnes pratiques autour des données en SHS
- Organisation de manifestations scientifiques

# Accompagnement individuel

## Pour connaître les données existantes

- Identification des bases : choix des sources, des instituts diffuseurs/producteurs
- Données existantes en relation avec un thème de recherche
- Comprendre les métadonnées



## Pour accéder aux données



## Pour traiter les données



## Pour interpréter les résultats



# Plateforme Universitaire de Données - AMU



# Où et comment chercher ?



Documents ▾



Fr ☰

## Mobilité professionnelle Fr

### Fiche du sujet

Référentiel

Rameau

Identifiant

http://data.bnf.fr/ark:/12148/cb11950828g

### Synonyme

Occupational mobility Job mobility

Mobility Occupational Movilidad laboral

Movilidad de mano de obra

Movilidad ocupacional Movilidad profesional

Trabajadores-Movilidad

### Discipline

Sociologie (139) - Economies et finances (96) - Histoire (43) - Droit (35) - Education (26) - Géographie (22) - Gestion et management (16) - Psychologie (8) - Science politique (6) - Anthropologie sociale et ethnologie (6) - Musique, musicologie et arts de la scène (3) - Linguistique (3) - Sciences de l'information et de la communication (3) - Architecture, aménagement de l'espace (2) - Démographie (1)

### Type de document

Articles (250) - Données d'enquêtes (141) - Livres et chapitres d'ouvrages (95) - Textes imprimés (68) - Mémoires, Thèses et HDR (50) - Colloques et conférences (37) - Autres (24) - Documents audiovisuels (12) - Billets de blog (12) - Recensions (10) - Rapports (9) - Page Web (4) - Prépublication (2) - Séminaires (1) - Présentations (1) - Périodiques (1) - Photos et images (1)

### Auteur

Institut National de la Statistique et des Études

672 documents

- Les pulsations de la mobilité en entreprise. Entre reconfiguration d'un marché interne et constitution d'un régime d'épreuve à la mobilité professionnelle**  
 Christophe Claisse (3 janv. 2) Textes imprimés
- Mobilité sociale : enquête formation qualification professionnelle de 1985 / Ministère de l'économie, des finances et de la privatisation, Institut national de la statistique et des études économiques ; [réd. par] Michel Gollac, Pierre Laulhé, Jeanine Soleilhavoup**  
 Institut national de la statistique et des études économiques (France). Auteur du texte (1988) Textes imprimés
- Enquête Emploi en continu - 2016**  
 Institut National de la Statistique et des Études Économiques (10 avr. 2018) Données d'enquêtes
- L'effet de l'âge sur la distribution des salaires**  
 Sarah Le Duigou (2012) Articles
- L'adaptation de la relation de travail pour motif économique**  
 Pierre Lopes (9 déc. 2017) Mémoires, Thèses et HDR
- Développer l'employabilité et favoriser la mobilité professionnelle par l'étude et la prise en compte du contexte de travail**  
 Anne Pignault et al. (janv. 2005) Colloques et conférences
- Enquête Emploi - 1968**  
 Institut National de la Statistique et des Études Économiques (9 avr. 2018) Données d'enquêtes
- Navette domicile-travail à grande vitesse : situation d'exception, arrangement traditionnel**  
 Xavier Lanéelle (2006) Articles
- Se rencontrer à la frontière : marchands et Inquisition dans l'Italie**  
Livres et chapitres d'ouvrages

# Accompagnement individuel

## Pour connaître les données existantes

- Identification des bases : choix des sources, des instituts diffuseurs/producteurs
- Données existantes en relation avec un thème de recherche
- Comprendre les métadonnées



## Pour accéder aux données

- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès



## Pour traiter les données



## Pour interpréter les résultats



# Comment accéder aux données ?



ADISP Commandier des données

DESCRIPTION TABLEAU ANALYSE

ADISP

- Éducation et formation
- Travail et emploi
  - Enquêtes Emploi
    - Série 1962-1967
    - Série 1968-1974
    - Série 1975-1981
    - Série 1982-1989
    - Série 1990-2002
    - Enquête Emploi en continu 2003-2016 - version standard
    - Enquête Emploi en continu 2003-2018 - version FPR
      - Enquête Emploi en continu (version FPR) - 2003
      - Enquête Emploi en continu (version FPR) - 2004
      - Enquête Emploi en continu (version FPR) - 2005
      - Enquête Emploi en continu (version FPR) - 2006
      - Enquête Emploi en continu (version FPR) - 2007
      - Enquête Emploi en continu (version FPR) - 2008
      - Enquête Emploi en continu (version FPR) - 2009
      - Enquête Emploi en continu (version FPR) - 2010
      - Enquête Emploi en continu (version FPR) - 2011
      - Enquête Emploi en continu (version FPR) - 2012
      - Enquête Emploi en continu (version FPR) - 2013
      - Enquête Emploi en continu (version FPR) - 2014
      - Enquête Emploi en continu (version FPR) - 2015
      - Enquête Emploi en continu (version FPR) - 2016
      - Enquête Emploi en continu (version FPR) - 2017
      - Enquête Emploi en continu (version FPR) - 2018
    - Métadonnées
    - Description des variables
      - Base individus
        - Caractéristiques individuelles
          - Age détaillé en années révolues au 31 décembre
          - Age en années révolues au 31 décembre (3 postes, premier type de regroupement)
          - Age en années révolues au 31 décembre (3 postes, second type de regroupement)

Jeux de données: Enquête Emploi en continu (version FPR) - 2018

**Variable AG: Age détaillé en années révolues au 31 décembre**

Valeurs	Catégories	N	
100	100 ans	29	0,0%
101	101 ans	30	0,0%
102	102 ans	20	0,0%
103	103 ans	20	0,0%
104	104 ans	11	0,0%
105	105 ans	4	0,0%
106	106 ans	2	0,0%
107		4	0,0%
15	15 ans	3553	0,8%
16	16 ans	6788	1,6%
17	17 ans	6920	1,6%
18	18 ans	7034	1,7%
19	19 ans	6267	1,5%
20	20 ans	6035	1,4%
21	21 ans	5516	1,3%
22	22 ans	5072	1,2%
23	23 ans	4779	1,1%
24	24 ans	4463	1,1%
25	25 ans	4198	1,0%
26	26 ans	4476	1,1%
27	27 ans	4470	1,1%
28	28 ans	4645	1,1%
29	29 ans	4623	1,1%
30	30 ans	4611	1,1%
31	31 ans	4915	1,2%
32	32 ans	5338	1,3%
33	33 ans	5248	1,2%



# Comment accéder aux données ?

Online Codebook & Analysis

User Guide

DESCRIPTION TABULATION ANALYSIS

Dataset: Generations and Gender Survey France Wave 1 & Wave 2

Opinion: Marriage... Institution (W2): Categories  Sex respondent (W2): Categories

Marital status respondent (W2): Categories  Type: Column percentage

Sex respondent (W2)	male					female					Total
	Never Married	Married	Divorced	Widowed	In Divorce Process	Never Married	Married	Divorced	Widowed	In Divorce Process	
strongly agree	13.6	10.8	21.5	9.5	53.8	10.7	8.0	15.7	12.0	10.5	11.5
agree	13.5	8.1	13.5	17.6	0.0	11.0	8.6	14.1	13.7	10.5	10.4
neither agree nor disagree	24.4	11.3	21.3	4.1	23.1	21.0	13.2	19.3	14.7	10.5	16.0
disagree	16.2	17.6	13.3	16.2	7.7	19.9	17.2	15.4	16.1	26.3	17.2
strongly disagree	32.3	52.2	30.4	52.7	15.4	37.4	53.0	35.5	43.5	42.1	44.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N=	668	1,633	362	74	13	963	1,852	597	292	19	6,473

# Accompagnement individuel

## Pour connaître les données existantes

- Identification des bases : choix des sources, des instituts diffuseurs/producteurs
- Données existantes en relation avec un thème de recherche
- Comprendre les métadonnées



## Pour accéder aux données

- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès

## Pour traiter les données

- Choix des méthodes adaptées
- Utilisation des logiciels informatiques pour le traitement des données



## Pour interpréter les résultats

# Quelle méthode choisir ?

## Choisir une méthode d'analyse multivariée

Les différentes méthodes d'analyse multivariée permettent de répondre à des problématiques variées. Le choix d'une méthode dépend de l'objectif initial, des types de variables manipulées mais aussi, de la forme des résultats obtenus qui peuvent être plus ou moins faciles à présenter et à expliquer.

<b>OBJECTIF</b>	Résumer l'information en minimisant la déperdition et repérer des dimensions "cachées"	Variables Numériques et/ou ordinales	→	<b>Analyse en Composantes Principales (ACP)</b>
		Deux variables qualitatives	→	<b>Analyse Factorielle des Correspondances (AFC)</b>
		Trois variables qualitatives ou plus	→	<b>Analyse des Correspondances Multiples (ACM)</b>
	Constituer des groupes d'individus aussi similaires que possible	Le nombre de groupes est fixé	→	<b>Analyse typologique</b>
		Le nombre de groupes n'est pas fixé	→	<b>Classification hiérarchique</b>
	Expliquer une variable par plusieurs autres	Variable à expliquer numérique ou ordinale	→	<b>Régression multiple</b>
		Variable à expliquer qualitative et variables explicatives numériques	→	<b>Analyse discriminante</b>
		Variable à expliquer qualitative et variables explicatives qualitatives	→	<b>Segmentation</b>

# Quelles fonctions utiliser ?

```

1 install.packages("tidyverse")
2 install.packages("haven")
3 install.packages("FactoMineR")
4 install.packages("missmda")
5 install.packages("Factoshiny")
6 install.packages("FactoInvestigate")
7 install.packages("openxlsx")
8
9 library(tidyverse)
10 library(haven)
11 library(FactoMineR)
12 library(missmda)
13 library(Factoshiny)
14 library(FactoInvestigate)
15 library(openxlsx)
16
17 setwd("C:/Users/debe/sunce/Downloads")
18
19 donnees1 <- read.xlsx("milieu.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
20                     rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
21                     skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
22                     namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)
23
24 res.mca1 = MCA(donnees1)
25
26 Investigate(res.mca1)
27
28 ## le fichier origine ne marche pas##
29
30 donnees2 <- read.xlsx("origine2.xlsx")
31 res.mca2 = MCA(donnees2)
32 plot.MCA(res.mca2)
33 Investigate(res.mca2)
34
35 #####
36
37
38 donnees3 <- read.xlsx("ambiance.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
39                     rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
40                     skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
41                     namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)
42
43 res.mca3 = MCA(donnees3)
44
45 Investigate(res.mca3)

```

2423 | (Top Level) | R Script

Console Terminal Jobs

```

C:/Users/debe/sunce/Downloads/
> setwd("C:/Users/debe/sunce/Downloads")
>
+ donnees1 <- read.xlsx("milieu.xlsx", sheet = 1, startRow = 1, colNames = TRUE,
+                     rowNames = FALSE, detectDates = FALSE, skipEmptyRows = TRUE,
+                     skipEmptyCols = TRUE, rows = NULL, cols = NULL, check.names = FALSE,
+                     namedRegion = NULL, na.strings = "NA", fillMergedCells = FALSE)
+
+ res.mca1 = MCA(donnees1)
+ |

```

# Accompagnement individuel

## Pour connaître les données existantes

- Identification des bases : choix des sources, des instituts diffuseurs/producteurs
- Données existantes en relation avec un thème de recherche
- Comprendre les métadonnées



## Pour accéder aux données

- Exploiter les nouveaux outils numériques de recherche de données
- Démarches/procédures d'accès

## Pour traiter les données

- Choix des méthodes adaptées
- Utilisation des logiciels informatiques pour le traitement des données



## Pour interpréter les résultats

- Interprétations des résultats pour répondre aux questions de recherche
- Valorisation graphique et mise en forme des résultats

# Comment interpréter ses résultats ?

## 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **13.66%** de l'inertie totale du jeu de données ; cela signifie que 13.66% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage très faible, et le premier plan ne représente donc seulement qu'une part de la variabilité contenue dans l'ensemble du jeu de données actif. Cette valeur est supérieure à la valeur référence de **10.45%**, la variabilité expliquée par ce plan est donc significative (cette inertie de référence est le quantile 0.95 de la distribution des pourcentages d'inertie obtenus en simulant 680 jeux de données aléatoires de dimensions comparables sur la base d'une distribution uniforme).

Du fait de ces observations, il serait alors certainement nécessaire de considérer également les dimensions supérieures ou égales à la troisième dans l'analyse.

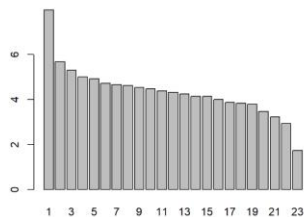


Figure 2 - Decomposition of the total inertia on the components of the analyse

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 5 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95 de distributions aléatoires (28.86% contre 25.19%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.

## 3. Description du plan 1:2

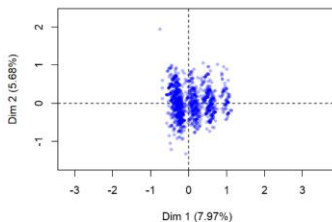
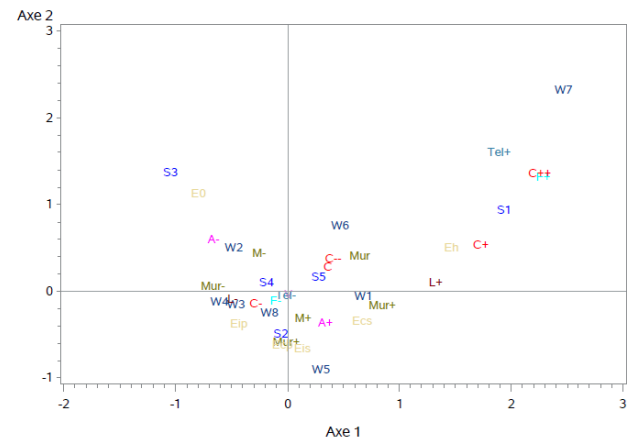


Figure 3.1 - Graphe des individus (ACM) Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan.



# Les missions de la PUD-AMU

- Accompagnement individuel
- **Accompagnement collectif**
- Promotions des bonnes pratiques autour des données en SHS
- Organisation de manifestations scientifiques

# Accompagnement collectif

- **Intervention dans les Masters :**
  - Outils d'aide à la recherche de données
  - Initiation au traitement statistique
  - Initiation à la gestion d'une base de données
- **Formations :**
  - Logiciel de traitement statistique R (3 niveaux)
  - Logiciel tableur Excel (2 niveaux)

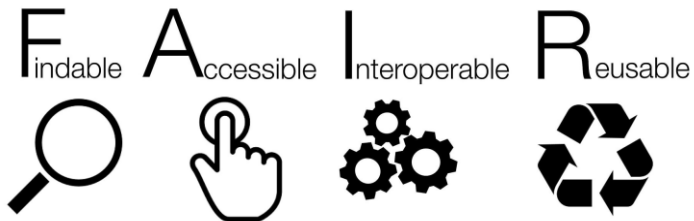


# Les missions de la PUD-AMU

- Accompagnement individuel
- Accompagnement collectif
- **Promotions des bonnes pratiques autour des données en SHS**
- Organisation de manifestations scientifiques

# Promotions des bonnes pratiques autour des données en SHS

- Ouvrir la science
- Relais informatif dans le domaine des données
- Sensibilisation à la culture des données
- Aiguillage vers les services traitant des sujets précis (PGD, archivage, stockage...)



# Les missions de la PUD-AMU

- Accompagnement individuel
- Accompagnement collectif
- Promotions des bonnes pratiques autour des données en SHS
- **Organisation de manifestations scientifiques**

# Organisation de manifestations scientifiques

- **Conférence + Master Class**
- **Montre-moi ta donnée**
- **Tour national des données de la recherche**
- **Semaine Data SHS : 07 au 11 décembre 2020**



Maison Méditerranéenne  
des Sciences de l'Homme

Aix★Marseille Université

**Merci de votre attention !**

**[pud-amu@univ-amu.fr](mailto:pud-amu@univ-amu.fr)**

**<http://pud.mmsh.univ-aix.fr>**