



plate-forme universitaire
de données de Caen

La PUDC, plate-forme universitaire de données de Caen

2016-2017

*Maison de la Recherche en Sciences Humaines
CNRS - Université de Caen Normandie*

RESPONSABLE : CLOTILDE LEMARCHANT

INGÉNIEUR : ALSENY NIARE

M R S H | N O R M A N D I E - C A E N
Maison de la Recherche en Sciences Humaines
C N R S - U N I V E R S I T É D E C A E N

Sommaire

Objet et fonctionnement de la PUDC.....	5
Objet	5
Un contexte national	5
Le Réseau Quetelet, la TGIR PROGEDO et le contexte européen	6
Financement et fonctionnement : un soutien de l'université de Caen Normandie	7
Un service de soutien à la recherche quantitative bien identifié	8
Le bilan au sujet des utilisateurs	9
Un travail en réseau.....	17
Des échanges avec les partenaires	17
Des contours géographiques souples	19
L'archivage de données.....	20
Bilan des publications et travaux en lien avec la PUDC	21
Annexes	22
Compte rendu du comité de pilotage de la PUDC	23
Tableau récapitulatif des dépenses 06/2016 à 05/ 2017	25

Objet et fonctionnement de la PUDC

Objet

La plate-forme Universitaire de données de Caen, située au bureau SH 109 de la Maison de Recherche en Sciences Humaines de Caen, vise à faciliter l'accès aux grandes enquêtes et données statistiques intéressant les sciences humaines et leur utilisation. Elle soutient la recherche quantitative au sein de l'Université de Caen Normandie en s'impliquant dans le déroulement de ses différentes étapes, qu'il s'agisse d'enquêtes et de données créées par les chercheurs caennais, doctorants ou chercheurs confirmés, ou du traitement secondaire de grandes enquêtes régionales, nationales ou internationales.

Elle offre ses services dans la recherche de fichiers adaptés aux projets de recherche, les démarches administratives avec le CASD (Centre d'accès sécurisé aux données) et le réseau Quetelet/PROGEDO, et la récupération et l'installation des bases de données. Elle propose un accompagnement méthodologique et des formations adaptées, de la conception des questionnaires jusqu'au traitement statistique. La PUDC permet de connaître les organismes, les grandes enquêtes les plus appropriées à une recherche, les démarches pour obtenir les fichiers, la déontologie à respecter dans l'utilisation de ces fichiers.

Elle propose un accompagnement personnalisé en traitement de données et analyse statistique : interprétation, description d'une distribution, tests statistiques, méthodes d'analyse, mais aussi gestion de fichiers, navigation sur le WEB, utilisation des logiciels statistiques, transfert des données d'un logiciel à un autre. Elle a enfin pour fonction d'archiver de façon pérenne les données des recherches réalisées dans le cadre de l'université de Caen, quantitatives ou qualitatives.

Un contexte national

Les plates-formes universitaires de diffusion des données en sciences humaines et sociales font partie du dispositif national de mise à disposition de la recherche des grands fichiers d'enquête produits par des organismes publics ou parapublics (INSEE, INED, Ministères, CNRS, Centres de recherches universitaires ou publics, etc.).

Avec le Comité de concertation pour les données en sciences humaines et sociales et le Réseau Quetelet/ PROGEDO, les plates-formes de données forment le troisième élément d'une infrastructure nationale visant à développer l'utilisation des enquêtes et des grands fichiers de données en sciences humaines et sociales. La première plate-forme de ce type a été créée à Lille en 2002 (PUDL), la seconde à Lyon en 2008, la troisième à Caen en 2011.

En septembre et novembre 2015, les plates-formes universitaires de données de Nantes et de Dijon ont respectivement ouvert leur service, s'inscrivant dans la même logique d'accompagnement et de soutien de la recherche quantitative en sciences humaines et sociales. Les journées d'ouverture ont mis en valeur l'existence d'un réseau national de plates-formes universitaires de données. En 2016, la plate-forme de données de Strasbourg a ouvert ses portes et d'autres sont en cours de création à Toulouse, Aix ou Saclay.

Le Réseau Quetelet, la TGIR PROGEDO et le contexte européen

La PUDC a été d'abord partenaire du GIS Réseau Quetelet, Groupement d'intérêt scientifique, dispositif national d'archivage, de valorisation et de diffusion des données pour la recherche qui met à disposition gratuitement des données à destination de la recherche. Désormais, cette architecture s'inscrit dans le cadre d'une TGIR PROGEDO, une très grande infrastructure de recherche « Production et gestion de données », dont Pascal Buléon est le président. Elle réunit tous les organismes de production des enquêtes quantitatives pour une politique de production et de mise à disposition de données en sciences humaines et sociales.

La TGIR PROGEDO (production et gestion de données) est impliquée dans trois consortiums européens :

1. ERIC (Consortium européen pour les Infrastructures de Recherche)
2. CESSDA (Consortium of European Social Data Archive / Réseau Quetelet)
3. Enquêtes européennes : ESS et SHARE.

L'enquête européenne SHARE (Survey on Health, Ageing and Retirement in Europe) est une base de données longitudinale, multidisciplinaire et internationale concernant plus de 80 000 Européens âgés de 50 ans et plus, pour environ 19 pays différents. L'enquête ESS (European Social Survey) porte sur le bien-être social et personnel des Européens. L'équipe de

coordination française de l'enquête sociale européenne est basée au Centre de données socio-politique de Sciences Po à Paris.

Les plates-formes de données et Progedo travaillent en lien avec le centre d'accès sécurisé aux donnés (CASD). Créé en 2009 par le Groupe des Ecoles Nationales d'Economie et Statistique (Genes), il est le garant du respect des règles éthiques de confidentialité, des lois et des exigences de la CNIL (Commission nationale de l'informatique et des libertés), et des exigences des autorités européennes en matière de protection des données personnelles.

Financement et fonctionnement : un soutien de l'université de Caen Normandie

D'abord financée dans le cadre d'un PPF du précédent quadriennal qui a pris fin en décembre 2011, la PUDC est désormais entièrement financée et soutenue par l'Université de Caen Normandie depuis le 1^{er} janvier 2012.

Elle a débuté en janvier 2011, avec un ingénieur d'étude, Cyril Coinaud, financé pour un an sur la fin d'un financement PPF, et une responsable, Clotilde Lemarchant, MCF de sociologie à l'Université de Caen Normandie. L'année 2014 a été marquée par le départ d'Emilie Herbrecht, qui avait succédé à Cyril Coinaud, et l'arrivée d'Alpha Diop au mois de septembre. L'année 2016 a vu l'arrivée de Alsény Niare suite au départ de Alpha Diop.

La PUDC est localisée à la Maison de la recherche en sciences humaines (MRSIH) de l'université de Caen Basse-Normandie, au bureau n° 109 (1er étage). Un site internet est disponible à l'adresse suivante : <http://www.unicaen.fr/recherche/mrsh/pudc>. Alpha Diop est l'ingénieur chargé des études de recherche auprès des utilisateurs. Depuis avril 2016, Alsény Niare a pris le relais. Il est joignable à partir de la ligne fixe suivante : 02 31 56 61 43, ou par courriel à alseny.niare@unicaen.fr. Clotilde Lemarchant, MCF à l'Université de Caen Normandie est la responsable de la plate-forme universitaire de données, elle est joignable également à cette adresse électronique, clotilde.lemarchant@unicaen.fr.

La MRSIH de Caen, lieu et moyen de mutualisation de la recherche en sciences humaines et sociales notamment au travers de l'ingénierie, des centres de ressources et d'équipements, a soutenu la PUDC dès sa création en lui procurant des locaux, finançant ses premiers frais, et lui laissant libre accès à ses services mutualisés, notamment de communication. Par

définition, la PUDC est au service de la recherche pluridisciplinaire et participe, modestement, à la mutualisation de ces services.

Un comité de pilotage a été constitué, composé d'une quinzaine de membres :

- La secrétaire générale du Comité de Concertation pour les Données en SHS
- Un-e représentant-e du réseau Quetelet/ TGIR Progedo
- Un-e représentant-e de la présidence de l'université de Caen
- Un-e représentant-e du CNRS en Région
- Le représentant de la MRSN de Caen
- Les représentant-e-s des laboratoires de recherches en SHS de Basse-Normandie
- Les représentant-e-s des responsables de masters
- Les représentant-e-s des écoles doctorales
- La responsable et l'ingénierie de la plate-forme

Un service de soutien à la recherche quantitative bien identifié

Après six ans de fonctionnement, la PUDC est aujourd’hui bien identifiée par bon nombre de chercheur-e-s et d’étudiant-e-s. Il lui faut toutefois constamment renouveler ces efforts pour faire connaître les services qu’elle propose. Sa visibilité n’est jamais acquise.

Ainsi, lors des réunions d’UFR, de département, de laboratoire ou d’écoles doctorales, la PUDC a régulièrement présenté ses différentes activités, retracant l’ensemble des tâches réalisées au sein du service, présentant des exemples d’enquêtes ciblées sur la discipline concernée, distribuant sa plaquette de présentation ainsi que les coordonnées du service. Chaque année, une attention particulière est apportée aux échanges avec les deux écoles doctorales des sciences humaines et des messages d’information ont régulièrement transité par ce canal. Un site internet hébergé à la MRSN et le mode d’information par le bouche-à-oreille, à la cafétéria par exemple, complètent le mode de communication du service.

Le bilan au sujet des utilisateurs

Après six années d’existence, une sorte de vitesse de croisière semble se dessiner autour des utilisateurs et utilisatrices de la PUDC. Chaque année, environ soixante personnes différentes ont recours aux services de la PUDC de façon permanente ou occasionnelle. Cette année : exactement cinquante et une personnes de disciplines différentes. Cette légère diminution s’explique par l’interruption du fonctionnement de la plate-forme liée au changement de personnel.

Ces usagers se répartissent entre abonnés réguliers, très majoritaires et usagers occasionnels. Dans tous les cas, les doctorant.e.s constituent le public le plus important, ce qui correspond au but recherché. La parité est respectée : 45% des usagers sont des femmes.

La pluridisciplinarité constitue le fondement des activités de la PUDC. L’année 2016-2017 ne faillit pas à la règle, avec douze laboratoires représentés. Géographie, sociologie, sciences de l’éducation et psychologie réunissent les usagers qui fréquentent le plus la PUDC.

Méthodologie :

Les graphiques et les tableaux ci-dessous ont été réalisés à partir du logiciel Excel (tris-à-plat, tableaux croisés). La base de données utilisée est la base « des utilisateurs de la Plateforme Universitaire de Données de Caen (PUDC) avril 2016-avril 2017 ».

NB : Plateforme Universitaire de Données de Caen (PUDC)

Les doctorants sont les utilisateurs majoritaires de la PUDC

Parmi les utilisateurs de la PUDC, plus de la moitié sont des doctorants soit 53%.

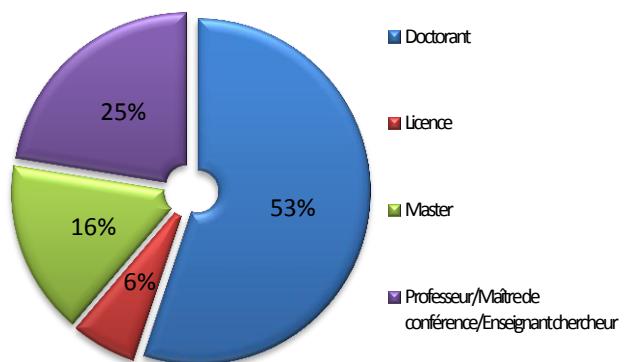
Néanmoins nous constatons que les utilisateurs sont variés, 25% ont un statut de Professeur/Maître de conférence/Enseignant chercheur, les étudiants de master représentent 16% des utilisateurs et les étudiants de licence 6%.

La durée moyenne d'accompagnement à la PUDC est de 4,7 heures avec une fréquence moyenne de passage de 2,1 fois par personne.

Sur l'ensemble des utilisateurs de la PUDC, 55% sont des hommes et 45% sont des femmes.

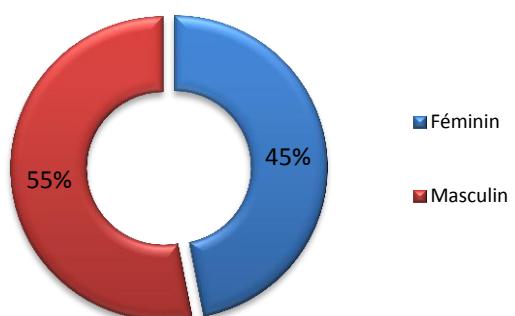
Sexe	Effectifs
Féminin	23
Masculin	28
Total général	51

Répartition des utilisateurs de la PUDC selon le statut



Durée moyenne d'accompagnement en heures	Fréquence moyenne de passage des utilisateurs
4,7 h	2,1 fois

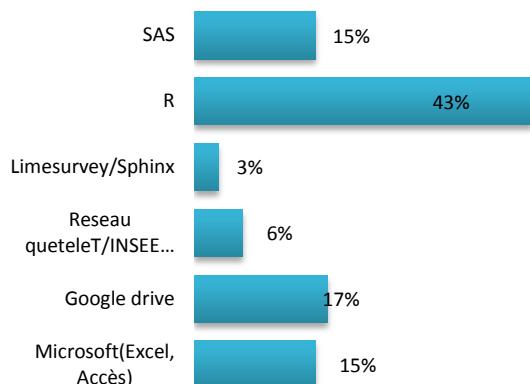
Répartition des utilisateurs de la PUDC selon le sexe



Le logiciel R est le plus utilisé à la PUDC

Plusieurs outils sont à la disposition des utilisateurs de la PUDC. Parmi ces outils, **le logiciel R est le plus utilisé soit 43%**. Le logiciel Google drive est souvent utilisé pour la réalisation des questionnaires en ligne soit 17%. Le logiciel SAS et les logiciels Microsoft (Excel, Accès) représentent un tiers de l'ensemble des outils utilisés à la PUDC soient 30%.

Proportion d'utilisation des outils ou des logiciels à la PUDC en 2016



Des formations par petits groupes de quatre personnes sont organisées en lien avec les écoles doctorales en sciences humaines et sociales. **La formation au logiciel R et au logiciel SAS représente respectivement 45% et 20%**. Il s'agit d'apprendre aux utilisateurs à effectuer des traitements statistiques sur des logiciels.

La formation au logiciel Excel (tableaux croisés, test de khi 2...) représente 15% et 15% des formations sont consacrées à la méthodologie/recherche de données.

Formations proposées %

Formations proposées	%
Logiciel R	45%
Logiciel SAS	20%
Google drive	5%
Excel	15%
Méthodologie/Recherche de données	15%
Total	100%

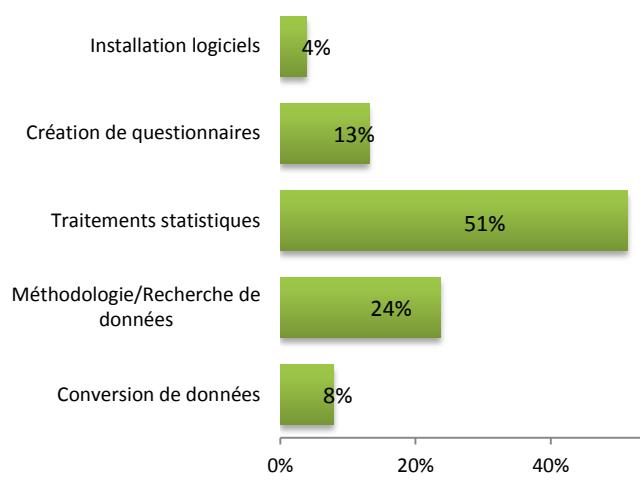
Les demandes des utilisateurs

Plus de la moitié des demandes de la PUDC sont liées à l'**analyse et aux traitements statistiques, elles représentent 51% des demandes**, un service souvent sollicité par les doctorants.

La méthodologie/recherche de données représente 24% des demandes. Il s'agit concrètement d'accompagner les utilisateurs dans leurs recherches de données, où et comment obtenir les données ?

Souvent les données font l'objet de conversion dans un format exploitable par les utilisateurs soit 8% des demandes. La création de questionnaires représente 13%.

Répartition des utilisateurs de la PUDC selon les types de demandes



Des utilisateurs provenant de divers établissements

Les utilisateurs de la PUDC proviennent de divers laboratoires ou établissements. Les laboratoires CeRRev (sociologie et psychologie), CERSE (sciences de l'éducation) et ESO (géographie sociale) sont les principaux destinataires des services de la PUDC. Ils représentent 53% des usagers.

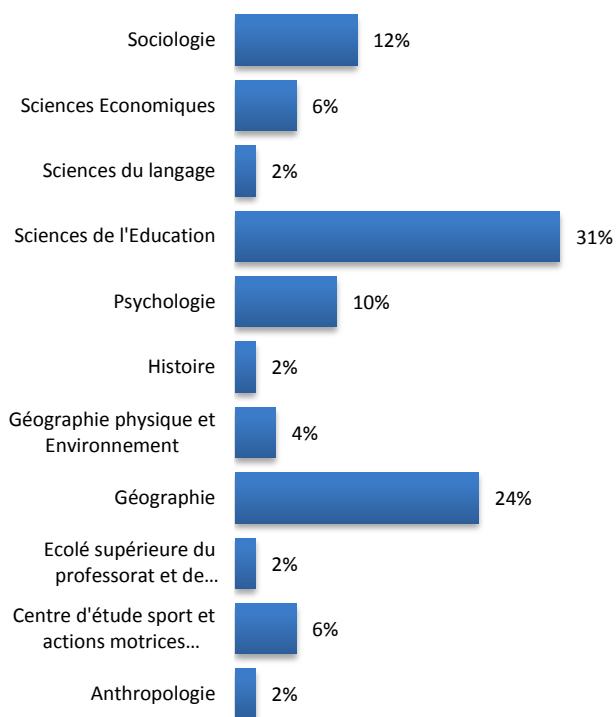
Répartition des utilisateurs par laboratoire ou établissement en 2016	
Laboratoire et institution	Effectifs
CeRReV	10
CERSE	8
CeSamS (STAPS)	3
CIRNEF	1
CRAHAM	1
CREM	2
CRISCO	1
ESO - Caen	8
ESO-Angers	1
Géophen	2
IDEES	1
PALM	2
Université Caen	10
Université Rennes	1
Total général	51

Répartition des utilisateurs par discipline

Les utilisateurs de la PUDC proviennent de disciplines variées, la pluridisciplinarité restant au fondement de ses activités. Les usagers en provenance des sciences de l'éducation, de géographie sociale ou physique, de la sociologie-anthropologie et de la psychologie sont les plus fréquents : ils représentent 83% de l'ensemble des utilisateurs.

Au total onze disciplines sont représentées, les disciplines des sciences sociales et humaines étant très majoritaires.

Répartition des utilisateurs de la PUDC par discipline en 2016

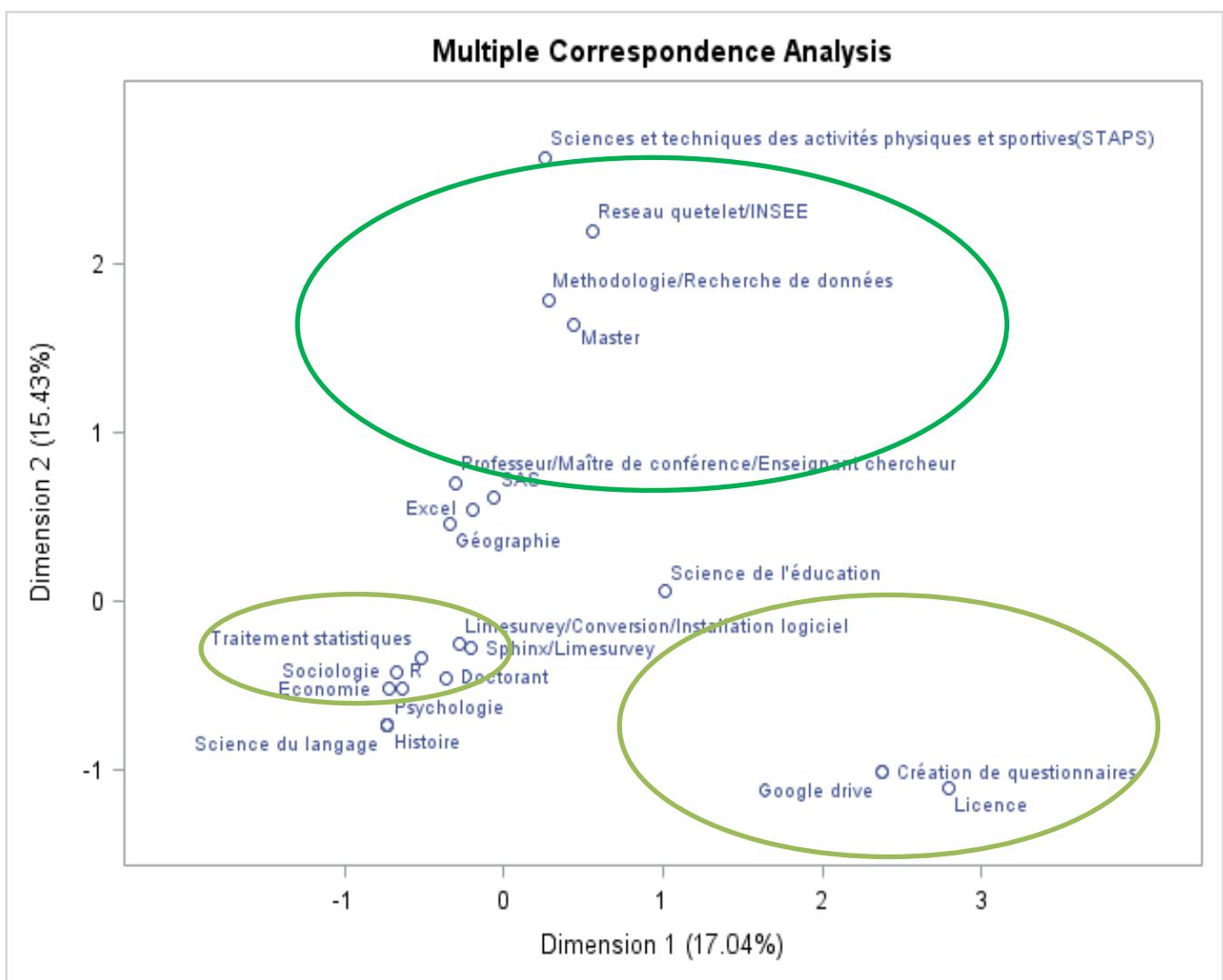


Méthodologie :

Le graphique 1 a été réalisé à partir du logiciel SAS. Le graphique 2 et le tableau 1 ont été réalisés à partir du logiciel NVIVO. La base de données utilisée est la base « des utilisateurs de la Plateforme Universitaire de Données de Caen (PUDC) avril 2016-avril 2017 ».

Des demandes différentes selon les utilisateurs

Graphique 1 : Analyse factorielle des correspondances multiples (ACM) sur les données recueillies auprès des utilisateurs de la PUDC



Graphique 1

L'axe 1 met en opposition les traitements statistiques et la création de questionnaires.

Nous constatons qu'au niveau des utilisateurs de la PUDC, **les doctorants ont des demandes plus orientées vers l'analyse et les traitements statistiques avec l'utilisation du logiciel R**. Quant aux étudiants de licence, ils viennent le plus souvent pour être accompagnés dans la création de leurs questionnaires en ligne.

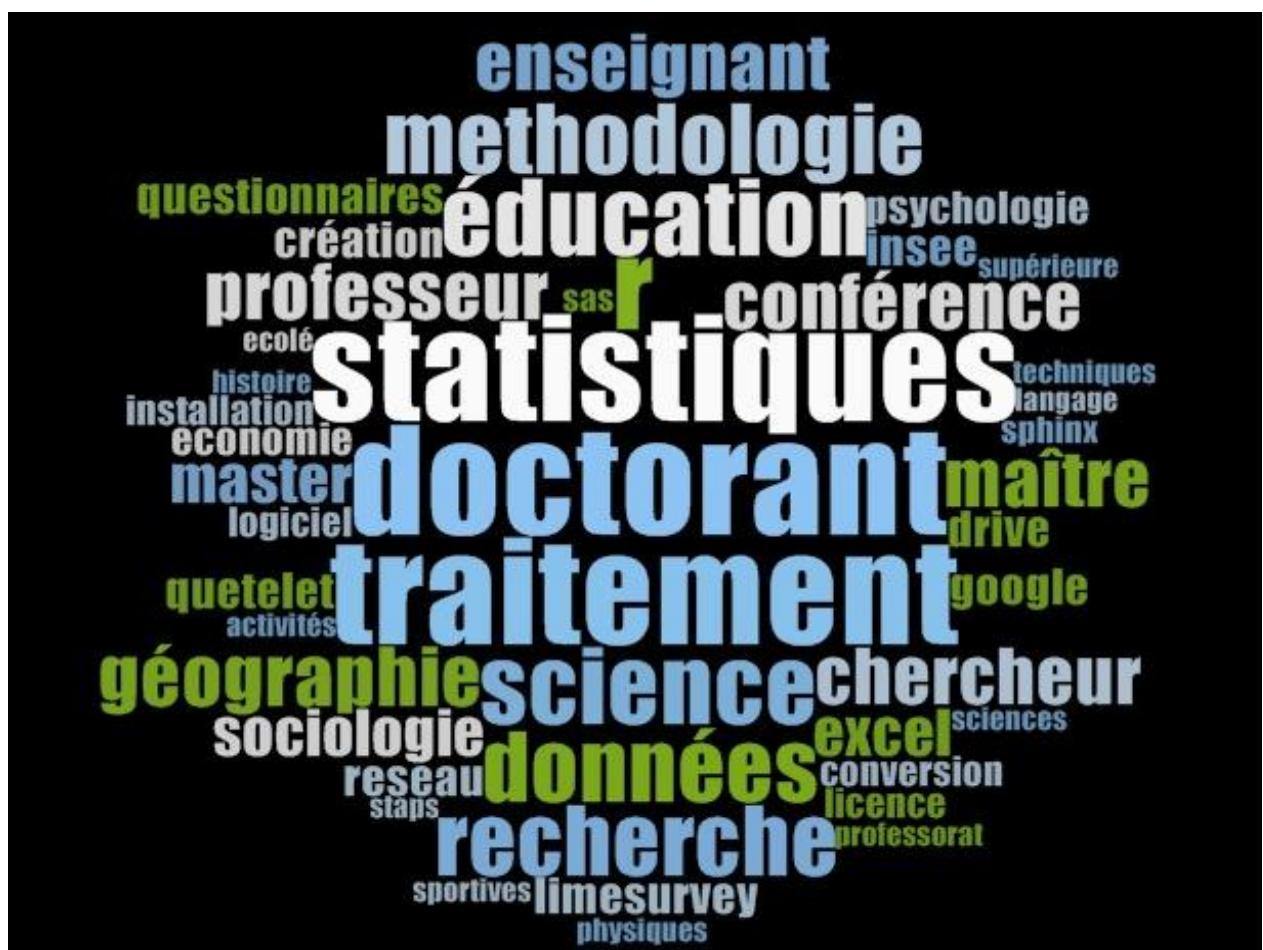
L'axe 2 est l'axe de la méthodologie/Recherche de données.

Les étudiants de master viennent à la PUDC pour des demandes liées à la méthodologie et la recherche de données.

Néanmoins, nous avons également les Professeurs/ Maîtres de conférence/ Enseignants chercheurs qui y viennent pour la même demande. Afin de satisfaire leur demande, nous leur proposons des formations pour mieux utiliser le réseau Quetelet/INSEE.

Les enseignants et les étudiants de master utilisent en général les logiciels SAS et Excel soit pour les traitements statistiques soit pour la conversion de leurs données.

Graphique 2 : Nuage de mots sur les données recueillies auprès des utilisateurs de la PUDC



Le graphique 2 ci-dessus nous montre que plus la représentation visuelle d'un mot est grande et plus il a une fréquence d'apparition importante dans l'ensemble des mots.

Comme nous pouvons le constater dans le tableau ci-dessous, le mot « doctorant » est celui qui revient le plus souvent chez les utilisateurs de la PUDC, 7,9% de l'ensemble, suivi respectivement par les mots « traitement » 7,6% et « statistiques » 7,6%.

Le mot « R » représente 4,9% de l'ensemble des mots, il s'agit de l'utilisation du logiciel R. Nous pouvons également noter que les mots « science », « éducation », « recherche », « méthodologie » et « données » sont souvent utilisés par les utilisateurs.

Tableau 1 : Les fréquences de quelques mots qui apparaissent le plus souvent auprès des utilisateurs de la PUDC

Mot	Longueur	Nombre	Pourcentage pondéré (%)
doctorant	9	27	7,85
traitement	10	26	7,56
statistiques	12	26	7,56
R	1	20	5,81
science	7	17	4,94
éducation	9	17	4,94
recherche	9	15	4,36
méthodologie	12	15	4,36
données	7	15	4,36
géographie	10	12	3,49
professeur	10	11	3,20
maître	6	11	3,20
enseignant	10	11	3,20
conférence	10	11	3,20
chercheur	9	11	3,20
sociologie	10	8	2,33
master	6	8	2,33
excel	5	8	2,33
réseau	6	5	1,45
quetelet	8	5	1,45
questionnaires	14	5	1,45
limesurvey	10	5	1,45

insee	5	5	1,45
google	6	5	1,45
drive	5	5	1,45
création	8	5	1,45
psychologie	11	4	1,16
économie	8	4	1,16
logiciel	8	3	0,87
licence	7	3	0,87
installation	12	3	0,87
conversion	10	3	0,87
sphinx	6	2	0,58
sas	3	2	0,58
techniques	10	1	0,29
supérieure	10	1	0,29
staps	5	1	0,29
sportives	9	1	0,29
sciences	8	1	0,29

Bilan des utilisateurs

- Cette année, cinquante une personnes ont fait appel aux services de la PUDC.
- Parmi les utilisateurs de la PUDC, plus de la moitié sont des doctorants soit 53%.
- La durée moyenne d'accompagnement à la PUDC est de 4,7 heures avec une fréquence moyenne de passage de 2,1 fois par personne.
- La formation au logiciel R et au logiciel SAS représente respectivement 45% et 20%.
- Plus de la moitié des demandes de la PUDC sont liées à l'analyse et aux traitements statistiques, elles représentent 51% des demandes.
- Les usagers en provenance des sciences de l'éducation, de la géographie sociale et de la sociologie sont les plus fréquents, ils représentent 67% de l'ensemble des utilisateurs.

Un travail en réseau

La gestion des données nécessite un réseau de partage d'informations et d'expériences avec les différents partenaires du réseau Quetelet/Progedo : d'une part les organismes de production et de diffusion de données, l'ADISP (Archives de Données Issues de la Statistique Publique, Centre Maurice Halbwachs), le CDSP (Centre de Données Socio-Politiques), l'INED (Institut National des Etudes Démographique, le CASD (Centre d'accès sécurisé aux données), d'autre part auprès des autres plate-formes universitaires de données.

Des échanges avec les partenaires

En 2016, la Plateforme Universitaire de Données de Caen a participé à des réunions et à des colloques avec ses partenaires. Ce fut des moments d'échanges, de rencontres et de formations enrichissants.

Vendredi 07/10/2016 : Conférence téléphonique PUD

- Point à date sur les activités de PROGEDO/ Pascal Buléon
- Tour de table et activités de chaque PUD
- Journées d'études dans les PUD
- Formation des nouveaux ingénieurs

Jeudi 13/10/2016 et vendredi 14/10/2016 : Journées de rencontre et d'information à la PUD de Lille

Ces deux journées d'informations et de formations ont été réalisées par Marie Cross, ingénierie d'étude à la PUDL.

L'objectif de cette rencontre était d'échanger sur les actions possibles à mettre en place au sein des PUD, expliquer le fonctionnement de la PUD de Lille, apprendre à rechercher les données, donner les pistes pour permettre de commencer l'activité à la PUD de Nantes.

Mardi 18/10/2016 : Présentation de la TGIR Progedo

Présentation de l'ADISP (Archives de Données Issues de la Statistique Publique)

L'ingénieur d'étude nous a présenté la structure. Il nous a parlé des différentes missions que l'ADISP propose : la collecte, la documentation et la diffusion des données.

Il a fait une présentation du site internet avec une description générale des données (méthodologie, producteur, date...).

Présentation du CDSP (Centre de Données Socio-Politiques).

Mardi 29/11/2016 : L'enquête sociale européenne (ESS) : recherche, analyse, méthodologie

Journée d'étude à Lyon

PROGEDO : organisation, mission, bases de données

Pascal Buléon, président de la TGIR PROGEDO a présenté le nouveau site de PROGEDO et ses différentes missions.

Pendant cette journée d'étude, les thématiques suivantes ont été élaborées par les intervenants:

L'enquête sociale européenne / European Social Survey

Simon Le Corgne (CDSP, SciencesPo Paris) et *Nicolas Sauger (LIEPP, SciencesPo Paris)* ont fait une présentation générale de l'enquête ESS (objectif, missions principales et méthodologie).

Deux cas d'utilisation de l'enquête ESS dans les travaux de recherche

Caractéristiques et valeurs des musulmans en Europe à partir des données cumulées de l'enquête European Social Survey (ESS) *Vincent Tournier (PACTE, SciencesPo Grenoble)*

Recherches méthodologiques dans le cadre de l'ESS : regard sur les expérimentations nationales et internationales réalisées par l'équipe Suisse *Michèle Ernst Stähli (FORS – Swiss Centre of Expertise in the Social Sciences, Université de Lausanne)*

Ateliers de prise en main des données

Atelier 2- Un regard comparatif sur le rapport au politique en Europe avec l'Enquête sociale européenne (ESS) *Anja Durovic (CEE/LIEPP, SciencesPo Paris)*

Cet atelier était axé sur les différents aspects du rapport au politique (participation politique institutionnelle et non- institutionnelle, engagement cognitif des individus dans la politique et attitudes politiques) afin de permettre aux participants de se familiariser à la fois avec la structure des questionnaires des différentes vagues de l'ESS, mais aussi avec la base de données elle-même (codage des variables, pondérations, pays participants...).

Des contours géographiques souples

La PUDC offre ses services principalement aux chercheurs, débutants ou confirmés, de l'université de Caen mais ses contours géographiques sont souples et prennent en considération les partenariats et les conventions passées avec d'autres chercheurs. Ainsi, la PUDC a reçu Nourédine Koudria de l'université d'Annaba (Algérie) le 28 septembre 2016, économiste très intéressé par les données sur les territoires. La question qui se pose alors est : Comment travailler à distance ? Il arrive aussi d'être sollicité par des collègues rennais ou rouennais.

L'archivage de données

L'archivage de données fait partie des objectifs et des compétences de la PUDC. Lors du dernier comité de pilotage du 14 juin 2016 (voir compte rendu en annexes), il a été abondamment question de cet aspect : d'une part, de nombreuses enquêtes réalisées par des collègues caennais sont terminées et les matériaux s'abiment alors qu'ils pourraient faire l'objet de traitement secondaire ou de recherches comparatives ; d'autre part, des institutions telles que l'HCERES demandent de plus en plus que les recherches rendent visibles leurs données.

Patrice Caro propose la mise à disposition de l'enquête de l'ANR INEDUC (inégalités éducatives), du moins pour commencer la partie quantitative de l'enquête réalisée auprès de 3356 adolescents en classe de 4^{ème} et de 1100 parents d'élèves. Les chercheurs et chercheuses participant à cette ANR se sont mis d'accord sur le principe. La PUDC connaît bien cette enquête qu'elle a accompagnée dès ses débuts (voir rapports d'activités antérieurs). Plusieurs réunions (17 juin 2016, 19 mai 2017) ont eu lieu à la PUDC pour envisager les étapes de la procédure d'archivage. De nombreuses questions se posent, notamment juridiques (garantir l'anonymat des personnes interviewées, formaliser la mise à disposition auprès de chercheurs) et techniques (comment documenter les données, nommer les variables, etc.). Ce travail est en cours, en lien avec le Centre associé au Cereq de Normandie.

Se pose également la question de l'archivage d'enquêtes qualitatives, ce qui pose de plus amples problèmes déontologiques, notamment en matière d'anonymisation des données. A cet égard, Clotilde Lemarchant a participé à l'atelier « Be-quali » organisé à l'occasion des 10 ans du CDSP, le 13 décembre 2016, à Sciences Po Paris. Le CDSP est en effet novateur en matière d'archivage de données qualitatives.

Bilan des publications et travaux en lien avec la PUDC

Frédéric Chantreuil, Frédéric Gavrel, Jean-Pascal Guironnet et Isabelle Lebon, « La contribution des écarts de rémunération entre les femmes et les hommes à l'inégalité des rémunérations dans la fonction publique : *une approche par la décomposition des inégalités* », Normandie Université, Caen, CREM-CNRS et TEPP-CNRS, ÉCONOMIE ET STATISTIQUE N° 488-489, 2016.

Isabelle Lebon, Antoinette Baujard, Frédéric Gavrel, Herrade Iggersheim, Jean-François Laslier, « Ce que le vote par approbation révèle des préférences des électeurs français », *Revue économique* – vol. 68 [article en prépublication].

Clotilde Lemarchant, « *De la parenté par alliance aux formations techniques : figures du genre* », Mémoire inédit en vue de l'obtention de l'Habilitation à diriger des recherches, Université de Poitiers, 2017.

Clotilde Lemarchant, *Unique en son genre, Filles et garçons atypiques dans les formations techniques*, Paris, PUF, 2017.

Annexes

Compte rendu du comité de pilotage de la PUDC

Université de Caen, MRS, 14 juin 2016

Présent-e-s : Pascal Buléon, Denis Jacquet, Annie Laurent, Daniel Delahaye, Bénédicte André, Jean-Marc Fournier, Vincent Merlin, Dominique Beynier, Alseny Niare, Clotilde Lemarchant.

Excusé-e-s :

Patrice Caro, Christophe Leconte, Joël Brée, Pierre Bauduin, Anne Pellissier, Olivier Maquaire, le représentant du CNRS en région.

Clotilde Lemarchant remercie les participant-e-s de leur disponibilité et leur intérêt pour la PUDC et présente le rapport d'activité de l'année 2015-2016. Elle présente aussi Alseny Niare le nouvel ingénieur statisticien de la plate-forme. En effet, l'année écoulée a été marquée par un changement de personnel, Alpha Diop ayant quitté la PUDC, partant pour Paris et un CDI.

La PUDC, malgré ces moments d'interruption lié au temps de recrutement du successeur, a néanmoins maintenu son activité : soixante-trois utilisateurs différents l'ont fréquentée cette année, répartis en douze disciplines différentes. Elle affirme sa dimension transversale, touchant chercheurs confirmés ou débutants, de différents statuts, les doctorants représentant les plus grands utilisateurs, ce qui correspond aux objectifs de la PUDC.

Alseny Niare se présente : il est titulaire d'un master 2 de démographie et d'un DU de statistiques. Il présente les nouvelles formations sur le logiciel R qu'il a mises en place depuis deux mois, en lien avec les deux écoles doctorales SHS présentes à la MRS.

Clotilde Lemarchant commente les différentes fonctions de la PUDC détaillées dans le rapport d'activité distribué à chaque participant (soutien à la recherche, à l'enseignement en statistique, adossement à des missions internes à l'université) et conclut en présentant quelques questions :

-Les chercheurs se plaignent des délais et des tarifs pratiqués par le CASD.

-L'archivage d'enquête est un projet qui se précise : l'équipe de l'ANR INEDUC vient de voter la mise à disposition de leur enquête sur le thème de l'éducation à partir de 2017. L'archivage pourrait être une dimension à mettre en œuvre mais il faudrait un second ingénieur.

-La PUDC a connu quatre ingénieurs en cinq ans : la stabilisation de la PUDC est liée à celle du poste de l'ingénieur actuel, qui est un CDD d'un an. Proposer un CDI permettrait cette stabilisation nécessaire au fonctionnement de la PUDC.

Daniel Delahaye dit qu'il entend bien la demande, mais que compte tenu du contexte, il préfère montrer la bouteille à moitié pleine c'est-à-dire l'effort de l'université pour financer le poste.

Pascal Buléon précise la nouvelle architecture du dispositif national de production et diffusion des données. La TGIR PROGEDO (production et gestion des données) a pris le relais de la structure du réseau Quetelet. Les plates-formes universitaires de données sont amenées à se multiplier et sont le ressort régional d'une organisation charpentée au niveau international, national et régional. Celles de Dijon et Nantes ont ouvert. D'autres suivront (Strasbourg, Aix-en-Provence).

Vincent Merlin précise que plusieurs économistes sont utilisateurs de la PUDC et qu'un nouveau thésard viendra prochainement. L'HCERES fait aux économistes le reproche du manque de diffusion des leurs données. Il y a à l'heure actuelle un gros travail de diffusion à réaliser. Il faudrait aller plus vers les sciences de gestion qui réalisent de nombreux travaux de terrain mais ne stockent pas leurs données.

Il est abondamment question d'archivage de données :

Patrice Caro propose la mise à disposition de l'enquête de l'ANR INEDUC (inégalités éducatives), du moins pour commencer la partie quantitative de l'enquête réalisée auprès de 3356 adolescents en classe de 4^{ème} et de 1100 parents d'élèves.

Dominique Beynier propose la mise à disposition de l'enquête sur la santé des bas-normands (3000 questionnaires).

Jean-Marc Fournier insiste sur l'importance de l'archivage de données : on a beaucoup de données anciennes et il faut capitaliser les connaissances, les transmettre aux jeunes générations.

Denis Jacquet donne l'exemple d'enquêtes à archiver sur les addictions, en psychologie.

Bénédicte André donne trois informations : tout d'abord elle souligne la qualité des exposés réalisés par Alpha Diop à Nantes et Dijon ; ensuite elle souligne l'importance du réseau des PUD et de Progedo ; enfin, elle précise qu'une PUD va ouvrir à Saclay en lien avec l'ENS Cachan et Youssef Safri en sera l'ingénieur d'études.

Clotilde Lemarchant prend note de ces propositions d'archivage et pose le problème du nombre et du statut du personnel de la PUDC. Il faudrait un CDI.

Daniel Delahaye entend le message et propose de faire apparaître en 1^{er} le poste de l'IE de la PUDC, pour une demande en CDI, lors du Dialogue de gestion de la MRS. Pascal Buléon en prend bien note.

Tableau récapitulatif des dépenses 06/2016 à 05/ 2017

date	BC /OM	Fournisseur	Date de la facture	Objet de la facture	Montant HT	pièce sifac SF
		NIARE Alseny		Salaire 06/2016 à 05/2017	30 532,21	
	4500241836	Université de Bordeaux		Logiciel SAS	216,00	5000344226
	4500243430	Travel planet		Mr Niare Alseny Caen Paris le 01/06	55,00	5000347213
	4500242118	BOUVET Laura		Gratification de stage	907,20	5000343280
	om 2628	NIARE Alseny		Rbt déplacement Paris le 01/06	19,05	
07/10/2017	4500230383	Université de Bordeaux		logiciel sas lemarchant clotide	216,00	
12/10/2017	OM 63484	NIARE ALSENY		Rbt Seminaire Plateforme Lille	40,15	
28/11/2017	om 64810	NIARE ALSENY		Rbt déplacement conférence étude sociale	334,76	
04/10/2017	4500229857	TRAVEL PLANET		Niare Alseny Lille séminaire plateforme	189,80	
04/10/2017	4500229860	IALBATROS		Niare Alseny Lille séminaire plateforme	179,16	
				TOTAL	32 689,33	