



plateforme universitaire
de données de Caen

La PUDC, plateforme universitaire de données à Caen

2014-2015

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

RESPONSABLE : CLOTILDE LEMARCHANT

INGÉNIEUR : ALPHA DIOP

M R S H

NORMANDIE - CAEN
Maison de la Recherche en Sciences Humaines
CNRS - UNIVERSITÉ DE CAEN

Sommaire

Objet et fonctionnement de la PUDC	5
Objet	5
Un contexte national.....	5
Financement et fonctionnement : un soutien complet de l'université de Caen UCBN	6
Un service de soutien à la recherche quantitative aujourd'hui bien identifié	7
Bilan au sujet des utilisateurs.....	8
Les utilisateurs depuis mars 2014.....	8
Principales activités de la PUDC	12
Accès aux données.....	13
Commande de données auprès du réseau quetelet.....	15
Analyse statistiques de données brutes.....	16
Initiation et perfectionnement en statistiques pour la recherche	16
Travail en réseau entre la PUDC, PROGEDO et les autres plates-formes	17
Séminaire de formation	17
Avec PROGEDO	17
Adossement à la recherche.....	18
Suivi du projet ANR InEduc.....	18
Suivi du projet «Inégalités liées au genre dans la fonction publique»	18
Adossement à l'enseignement supérieur	20
Adossement à des missions internes à l'université	21
Participation à l'observatoire genre et formation à l'UCBN.....	21
Rappel du Contexte : une demande de la présidence de l'université	21
Evolution des taux de féminisation des filières à l'UCBN entre 1996 et 2013	22
Evolutions des taux de féminisation des filières entre 1996 et 2013	22
Les primo-entrants à L'UCBN néo-bacheliers, par filière et par sexe en 2013	24
Les filières éloignées de la parité.....	24
Lien avec l'Observatoire de l'université	25
Aide au service de scolarité de l'UFR de Sciences de l'Homme.....	26
Projets, perspectives et questions.....	27

Bilan des publications et travaux en lien avec la PUDC	28
Bilan financier.....	29
Frais de fonctionnement	29
Glossaire.....	30
Liste des tableaux	30
Liste des figures	30
Annexes.....	31

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 Basse 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 recherches, 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, 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. D'autres sont en cours de création, à Nantes, Dijon, etc.

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'un TGIR PROGEDO, une très grande infrastructure de recherche « Production et gestion de données », dont Pascal Buléon est le président.

Financement et fonctionnement : un soutien complet de l'université de Caen UCBN

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 UCBN 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'UCBN. L'année 2014 a été marquée par le départ d'Emilie Herbrecht, qui avait succédé à Cyril Coinaud, et l'arrivée d'Alpha-Oumar Diop au mois de septembre.

La PUDC est localisée à la Maison de la recherche en sciences humaines (MRSH) de l'université de Caen Basse-Normandie, au bureau n° 109 (1^{er} étage). Un site internet est disponible à l'adresse suivante : <http://www.unicaen.fr/recherche/mrsh/pudc>

Alpha Oumar Diop est ingénieur d'études, Tél. : 02 31 56 61 43 ; Courriel : pudc@unicaen.fr et alpha-oumar.diop@unicaen.fr

Clotilde Lemarchant, est la responsable de la plate-forme ; Courriel : clotilde.lemarchant@unicaen.fr

La MRSH 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 des membres 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 MRSH 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'ingénieure de la plate-forme

Un service de soutien à la recherche quantitative aujourd'hui bien identifié

Après quatre 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 Géographie du 21 octobre 2014, du département de sociologie le 8 décembre 2014, ou de l'école doctorale de Normandie du 19 novembre 2014, la PUDC a présenté ses différentes activités, retraçant 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. Cette année, une attention particulière a été 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 MRSH et le mode d'information par le bouche-à-oreille, à la cafétéria par exemple, complètent le mode de communication du service.

Bilan au sujet des utilisateurs

Pour la période allant du 25 juillet 2014 au 15 septembre 2014, il n'y a eu personne à la PUDC, le temps que soit recruté le successeur de l'ingénieure.

Une constante fréquentation

Après quatre années d'existence, une sorte de vitesse de croisière semble se dessiner. Chaque année, environ soixante utilisateurs différents ont recours aux services de la PUDC.

Cet extrait du précédent rapport d'activité le montre, pour une période de 7 mois (octobre 2013-mars 2014).

Discipline sur la période oct. 2013- mars 2014	Effectif
Sciences de l'éducation	4
Sociologie	6
STAPS	4
Psychologie	2
Mathématiques Appliquées et Sciences Sociales	1
Géographie	5
Géographie sociale	2
Sciences économiques	3
Littérature et civilisation	1
Médecine	1
Aménagement	1
Mathématiques	2
Politiques sociales	5
TOTAL	37

Tableau 1 : extrait du rapport d'activités octobre 2013-mars 2014

Les utilisateurs depuis mars 2014

Le profil des utilisateurs de la Plate-forme Universitaire de Données de Caen est assez varié. Ce public peut être classé par statut, par discipline d'études ou selon l'institution de rattachement. Actuellement, les différents statuts retenus sont essentiellement les étudiants en Master, les doctorants, les enseignants-chercheurs, les chercheurs et les chargés d'études.

Cette année, sur la période située entre mars 2014 et mars 2015, soixante et un (61) utilisateurs différents ont adressé des demandes à la Plate-forme Universitaire de données de Caen. Parmi eux, quinze (17) sollicitent très régulièrement ses services, notamment depuis septembre 2015, période de la reprise des activités. Ce sont des abonnés ponctuels dont les demandes portent le plus souvent sur les analyses de données, la formation à un logiciel statistique, la commande de données ou des tâches liées à l'administration d'une enquête en ligne.

Répartition des utilisateurs suivant leur statut

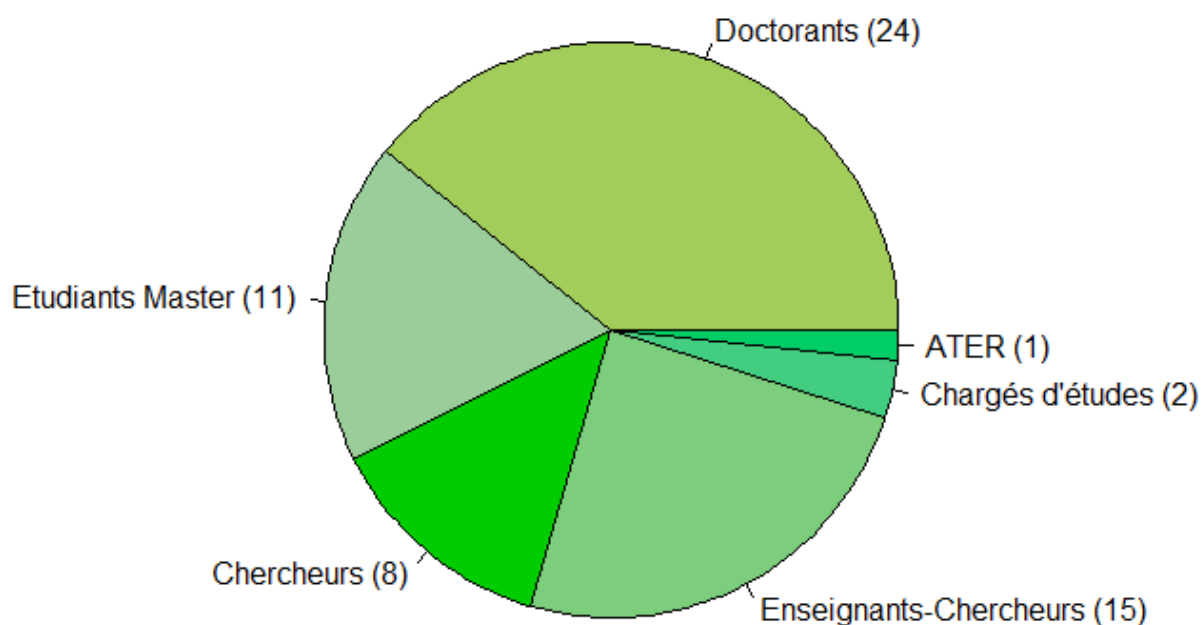


Figure 1 : Répartition des utilisateurs suivant leur statut

Lecture : 24 doctorants différents ont adressés des demandes à la PUDC sur la période située entre Mars 2014 et Mars 2015.

Chercheurs chevronnés ou débutants ont recours aux services de la PUDC mais les doctorants sont les principaux utilisateurs de la PUDC, ce qui correspond au but recherché.

Répartition des utilisateurs suivant leur discipline

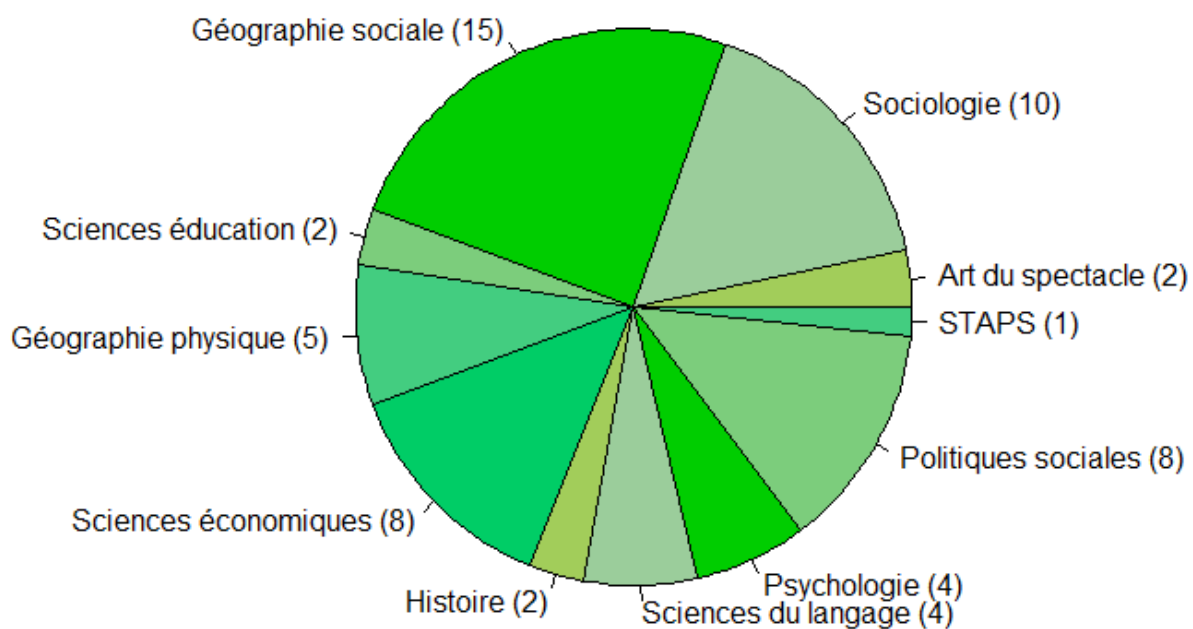


Figure 2 : Répartition des utilisateurs suivant leur discipline

Lecture : Parmi les utilisateurs des services de la PUDC, 15 viennent de Géographie sociale

Les utilisateurs de la PUDC représentent onze disciplines différentes. A noter cette année la présence de littéraires parmi les utilisateurs, dans la spécialité « sciences du langage », notamment pour le traitement de données qualitatives et d'analyses textuelles.

Les utilisateurs par institutions, Septembre 2014-Mars 2015

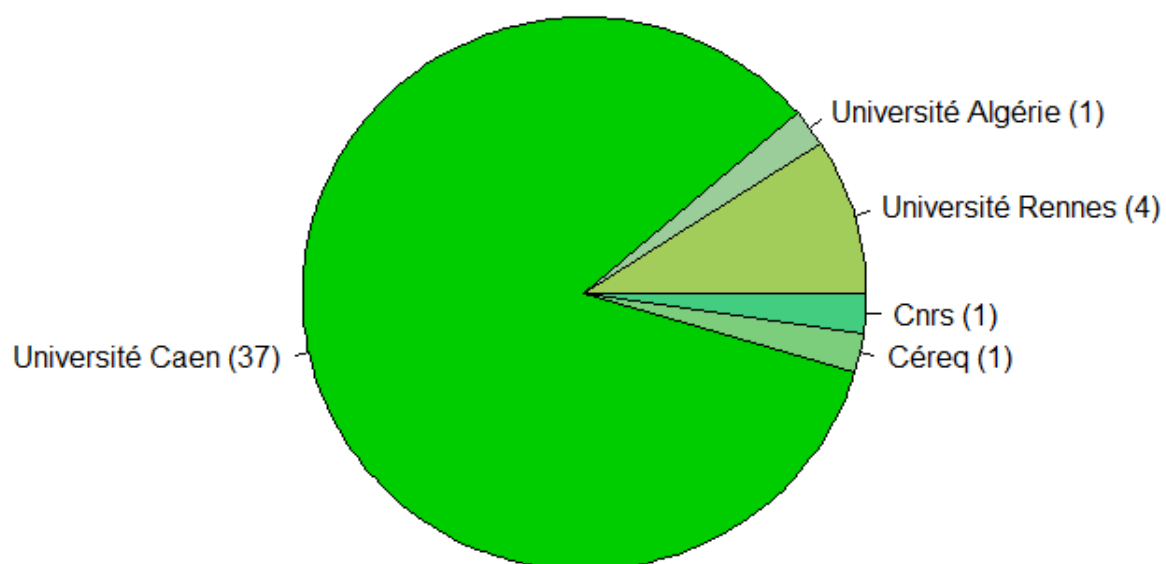


Figure 3 : Répartition des utilisateurs suivant leur institution

Lecture : 37 utilisateurs de la PUDC sont rattachés à l'Université Caen Basse Normandie.

Principales activités de la PUDC

Les utilisateurs ont recours aux services de la PUDC pour trois sortes de besoins : l'aide à la recherche et la commande de données, à l'analyse de ces données et les formations en matière de statistiques. La commande de données consiste à trouver les données adaptées très précisément au thème de la recherche auprès du réseau Quételet/ PROGEDO. A défaut de l'existence de ces données, la mission consiste à accompagner la production des données par questionnaire. Quant à l'analyse, elle couvre l'ensemble des services liés au traitement statistique des données, l'interprétation et la présentation des résultats. Enfin, la formation fait référence aux différentes notions de statistiques ou à l'utilisation des logiciels R, SAS ou Excel, à tous niveaux, initiation ou perfectionnement. Très récemment, la PUDC a acheté une nouvelle licence du logiciel statistique «statistica» qui a fait l'objet de plusieurs demandes et fera bientôt partie de la liste des outils techniques d'initiation proposés aux utilisateurs. Il s'agit en effet toujours de s'adapter aux besoins et demandes nouvelles.

Répartition des demandes selon les besoins des utilisateurs de la PUDC

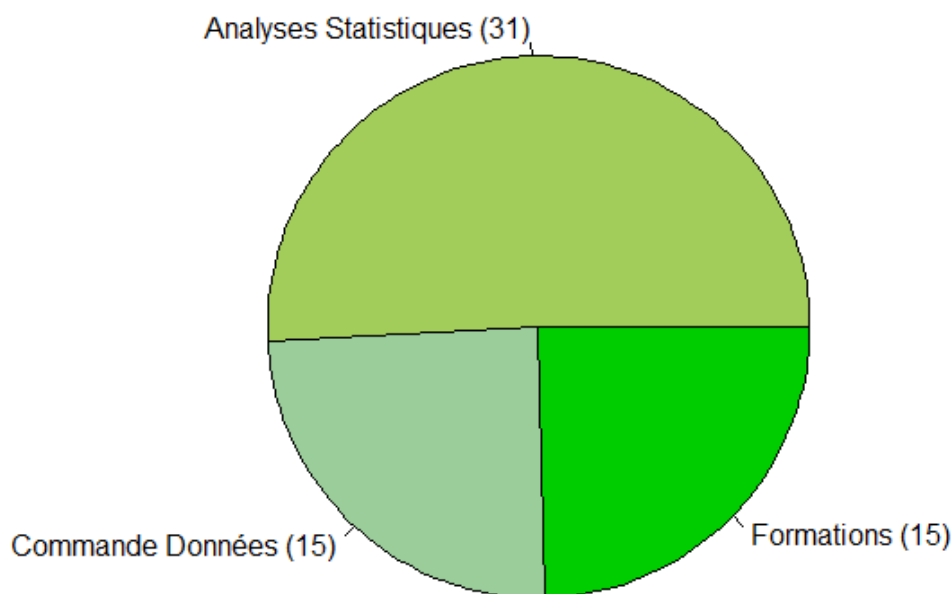


Figure 4 : Répartition des utilisateurs suivant les besoins

Lecture : Parmi les 61 utilisateurs de la PUDC, 31, soit la moitié, ont formulé des demandes en lien direct avec l'analyse statistique des données.

Temps de travail en heures de Septembre 2014-Mars 2015

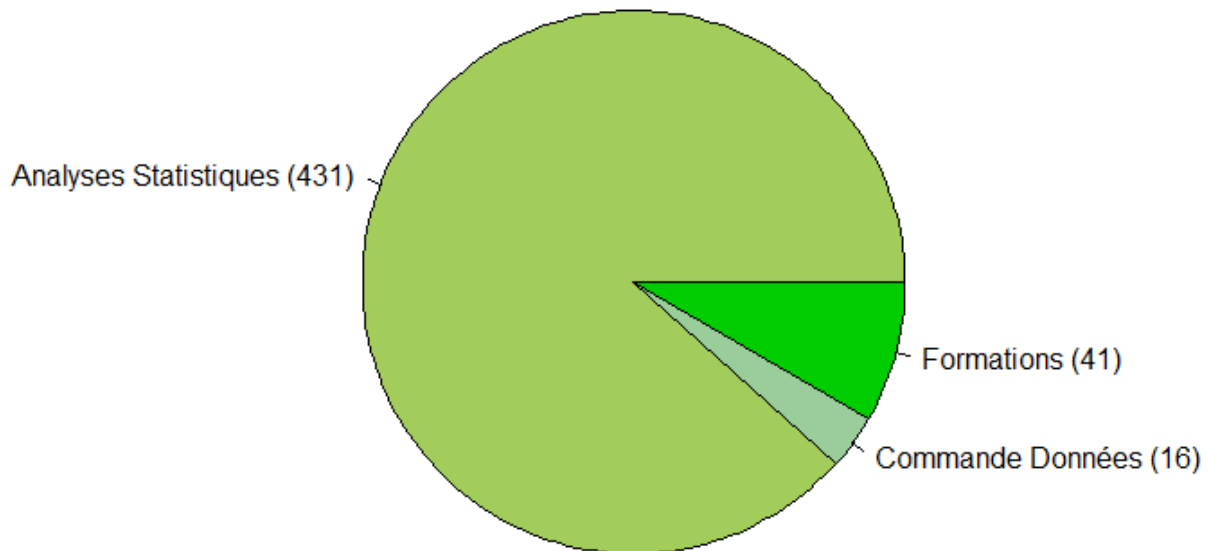


Figure 5 : Temps de travail présents en heures

Lecture : Entre septembre 2014 à Mars 2015, 417 heures ont été consacrées à l'échange avec les utilisateurs sur les objectifs de leur projet de recherche ou à l'interprétation des résultats obtenus suite à une analyse réalisée.

Accès aux données

La difficulté est que certains utilisateurs ne disposent pas des connaissances portant sur les données recherchées. Une part du travail à la PUDC consiste à aider à formuler la demande, à préciser les besoins, à clarifier les thèmes pour les traduire en variables, à orienter vers les enquêtes adaptées. Il s'agit de permettre d'aider à s'orienter dans cette vaste bibliothèque que représentent ces listes d'enquêtes. Des échanges avec la plate-forme universitaire de Lille sont utiles à cette étape.

Dans tous les cas, ces genres de situation rendent les démarches beaucoup plus délicates pour la PUDC, auprès du réseau de production de données. Elle recueille auprès de l'utilisateur concerné, le maximum d'informations nécessaires sur le projet de recherche, afin de contacter le personnel du réseau apte à apporter des solutions. Dans le cas où les données recherchées ne sont pas disponibles auprès du réseau, une possibilité de production de données par un questionnaire se pose. C'est le cas du projet « ANR INEDUC » (voir plus loin).

A la réception des données auprès du réseau ou par les procédures de production via un questionnaire, beaucoup d'utilisateurs envisagent de traiter eux même leurs données. Ils sollicitent tout de même, les services de la PUDC pour transformer ou installer les fichiers des données en d'autres formats sur leur poste de travail. Les statistiques montrent que le profil de cette catégorie d'utilisateur est en nette progression. Sur la période considérée, Mars 2014 à Mars 2015, la proportion des demandes qui concernent la transformation de formats de fichiers de données vers d'autres formats, Excel à l'occurrence est de vingt-deux pour cent (22%).

Commande de données auprès du réseau quetelet

Adresse professionnelle	Titre de l'enquête
Doctorant	Aires urbaines (définition 1990) : indicateurs statistiques - 1990
Doctorant	Inventaire communal : équipements et attraction des communes - 1998
Doctorant	Démographie des entreprises et des établissements - 2013
19 rue Claude Bloch	Emploi - 1992
19 rue Claude Bloch	Emploi - 1998
19 rue Claude Bloch	Emploi (en continu) - 2004
19 rue Claude Bloch	Emploi (en continu) - 2010
19 rue Claude Bloch	Formation, qualification professionnelle (FQP) - 1970
19 rue Claude Bloch	Formation, qualification professionnelle (FQP) - 1977
19 rue Claude Bloch	Formation, qualification professionnelle (FQP) - 1985
19 rue Claude Bloch	Formation, qualification professionnelle (FQP) - 1993
19 rue Claude Bloch	Formation, qualification professionnelle (FQP) (version FPR) - 2003
IUT CHERBOURG MANCHE	International social survey programme (ISSP) Enquête France - Environnement III - 2010
IUT CHERBOURG MANCHE	International social survey programme (ISSP) Enquête France - Environnement III - 2010
IUT CHERBOURG MANCHE	EPCV (partie variable les pratiques environnementales et sensibilité aux problèmes d'environnement - janvier 1998
IUT CHERBOURG MANCHE	EPCV (partie variable) les pratiques environnementales et sensibilité aux problèmes d'environnement - janvier 1998

Tableau 2 : Tableau récapitulatif des demandes de fichiers de données auprès de quetelet

Analyse statistiques de données brutes

Plus de la moitié des utilisateurs adressent à la PUDC des demandes liées au traitement statistique des données et à l'interprétation des résultats. Les différentes tâches liées à cette demande consistent à préparer les données à l'analyse, puis appliquer des méthodes statistiques. La phase préparatoire consiste à détecter les valeurs qui paraissent incohérentes (aberrantes) pouvant profondément modifier les analyses, effectuer des tris à plats ou des tris croisés, recoder des variables ou regrouper des modalités du fichier de données puis gérer des valeurs manquantes. Les techniques statistiques proposées sont les régressions linéaires simples, l'analyse en composante principale, l'analyse factorielle de correspondance, la classification des données.

Initiation et perfectionnement en statistiques pour la recherche

La mission de la Plate-forme Universitaire de données de Caen est de faire faire et non de faire à la place... Les contacts avec les écoles doctorales Homme-Sociétés-Risques-Territoire (ED 556) et Histoire-Mémoire-Patrimoine-langage (ED 558) se sont densifiés cette année en vue de proposer aux doctorants des formations complémentaires. Depuis la date du 03 décembre 2014, dix doctorants de différentes disciplines, bénéficient chaque mercredi deux séances de formation sur des notions statistiques et sur l'utilisation du logiciel R. Parmi eux, quatre doctorants viennent très régulièrement assister aux différentes séances. Les autres prennent souvent des contacts avec la plate-forme pour traiter des questions statistiques spécifiques liées à leurs domaines de recherche.

Le programme est axé essentiellement aux statistiques pour la recherche avec des exemples concrets, traités avec le logiciel libre R. Le contenu du cours consiste à collecter les informations chiffrées ou textuelles, réaliser les traitements nécessaires et présenter les résultats à travers des tableaux, des diagrammes ou éventuellement des courbes. Dans certains cas, avec des données géographiques, les résultats sont présentés en cartographie sur le fond de la carte de France, par département et par région.

Les bénéficiaires sont sensibilisés aux notions d'échantillonnage et à la nécessité d'avoir des échantillons représentatifs pour les études statistiques à visée inférentielle. La partie analyse porte sur le traitement univarié, bivarié et multivariés des données et sur les tests mettant en évidence des éventuelles liaisons entre les différentes variables.

Pour les utilisateurs provenant des sciences sociales, le 11 février 2015 a été présenté IRAMUTEQ, un logiciel de traitement de données qualitatives. La séance a été présentée par Amadou Diaw, doctorant en Sociologie ayant une expérience assez importante d'IRAMUTEQ, auprès de sept doctorants de la MRSB avec des profils assez variés.

Travail en réseau entre la PUDC, PROGEDO et les autres plates-formes

La gestion des données nécessite un réseau de partage d'informations et d'expériences avec les différents partenaires du réseau Quetelet, devenu PROGEDO : d'une part, les organismes de production et de diffusion de données, le CASD, le CDSP (Sciences Po), l'ADISP du CMH, l'INED, d'autre part, le CASD, enfin les autres plates-formes de données universitaires. Des échanges d'information d'ordre technique et méthodologique ont souvent lieu entre la Plate-forme Universitaire de Caen et celles de Lyon et de Lille.

Séminaire de formation

Le domaine des statistiques en sciences humaines est en constante évolution et nécessite des efforts de formation. Les organismes de productions de données proposent régulièrement des séminaires de formation dont a bénéficié l'ingénieur statisticien de la PUDC. Notamment, il a participé à deux séances de formation organisées par l'INED et animées par Claire Tissot, qui ont respectivement eu lieu, Jeudi 18 décembre 2014 et Jeudi 19 mars 2015, de 9h00 à 11h00, à l'EHESS, (avenue de France) à Paris, portant sur l'utilisation du logiciel libre R dans les sciences sociales. «Text mining et recherche de thèmes dans un ensemble de documents : une application de R au traitement de données d'accidents de salariés nouvellement embauchés dans une entreprise ».

Avec PROGEDO

La PUDC s'insère dans un dispositif national aujourd'hui architecturé autour de la TGIR PROGEDO et de CESSDA France, dont l'objectif principal est de soutenir et développer l'usage de deux grandes enquêtes européennes : SHARE, une enquête européenne sur la santé et le vieillissement et l'ESS, enquête sur les opinions des Européens.

Une réunion du comité de pilotage de la TGIR PROGEDO s'est tenue à Paris le 30 septembre à l'EHESS (avenue de France), dans la continuité d'une réunion de bilan du CESSDA. La prochaine réunion PROGEDO aura lieu le 13 mai 2015 au ministère de la Recherche à Paris.

Une réunion centrée sur les activités et le mode de fonctionnement des plates-formes existantes à Lille, Lyon et Caen s'est tenue le mardi 10 mars 2015 à l'EHESS à Paris. La rencontre a été animée par Pascal Buléon, directeur de la MRSH de Caen. L'enjeu de ces échanges était de favoriser l'ouverture de nouvelles plates-formes de données, notamment à Nantes, Dijon et Besançon. Un tour de France des MSH organisé en partenariat avec l'Alliance Athéna et le Réseau National des MSH sera amorcé à l'occasion de l'ouverture de nouvelles plates-formes universitaires de données pour promouvoir auprès des chercheurs et des étudiants les services d'accès aux données dans les universités.

Adossement à la recherche

Suivi du projet ANR InEduc

La Plate-forme Universitaire de Données de Caen a pris part, dès le début, au projet de l'ANR InEduc, coordonnée par le laboratoire ESO, qui vise à identifier les inégalités éducatives liées aux contextes et espaces de vie des jeunes de 11 à 15 ans (période scolaire du collège) en France. L'objectif est d'analyser les différents facteurs qui influent sur les parcours des jeunes dans le cadre de l'institution scolaire, des loisirs sportifs et culturels avec une observation particulière portée sur les environnements numériques. La dimension comparative de ce projet, qui traite, sur le plan quantitatif, de trois territoires régionaux (Aquitaine, Basse-Normandie et Bretagne) et sur le plan qualitatif, d'espaces locaux sélectionnés dans ces régions, vise à enrichir la connaissance de la diversité des politiques d'éducation mises en œuvre et des effets de lieu sur les parcours éducatifs des jeunes. La dimension longitudinale de ces parcours est en cours d'exploration dans les trois contextes éducatifs choisis.

La Plate-forme Universitaire de Données de Caen continue à s'impliquer dans la partie quantitative de la recherche dirigée à Caen par Patrice Caro, du laboratoire ESO-Caen et directeur du CAR Céreq de Caen. Trois mille cinq cent questionnaires ont été administrés dans les collèges des trois régions concernées, auprès des parents et des jeunes collégiens. La Plate-forme Universitaire de Données de Caen a accompagné les diverses étapes du protocole d'enquête : élaboration et maquettage des deux questionnaires, codage et nettoyage des données après la phase d'administration des questionnaires, saisie des questionnaires au moyen du logiciel Data Scan, apurement des bases de données obtenues, formatage des variables et des modalités de réponses de la base de données, participation à la pondération des données, réalisation de multiples tris à plat et tris croisés en vue de la rédaction de l'analyse descriptive, de tests et analyses factorielles. Sept enseignants-chercheurs et chercheurs formulent des demandes régulières à ce sujet. De septembre 2014 à mars 2015, le volume horaire consacré aux différentes tâches du projet s'élève à 260 heures.

Suivi du projet «Inégalités liées au genre dans la fonction publique»

Isabelle Lebon, du laboratoire CREM (Centre de Recherche en Economie et Management) a sollicité les services de la plate-forme de données. La demande portait sur l'exploitation des données des enquêtes « Formation et Qualification Professionnelle » de l'INSEE, des années 1970, 1977, 1985, 1993 et 2003.

La population de la base de données est constituée des agents de la fonction publique et du secteur privé. Elle est également caractérisée par le niveau du diplôme le plus élevé obtenu, le temps de travail, le salaire et d'autres éléments spécifiques. La PUDC s'est impliquée considérablement dans l'analyse statistique des données de ce projet, dont la présentation des résultats aux syndicats est en cours. Le temps consacré aux différents échanges sur les objectifs et à l'interprétation des résultats pour ce projet est de 62 heures.

La phase préparatoire était centrée essentiellement sur le recodage et le redressement des variables, notamment celles liées au salaire. Le nettoyage des données a nécessité également la gestion des valeurs manquantes et les valeurs aberrantes afin d'éviter les biais sur les analyses statistiques. Les principales tâches consistaient à décomposer les salaires suivant le secteur, public ou privé, et suivant l'âge et le sexe, avec le logiciel statistique SAS. Cette étape a nécessité l'ajout de cinq nouvelles variables suivant l'algorithme de décomposition proposé par Isabelle Lebon, responsable du projet. Le calcul des indices économiques (Gini, Theil, Atkinson...) est appliqué sur chaque variable nouvelle obtenue lors de cette décomposition.

La suite des analyses statistiques sur le projet « Inégalités liées aux genres dans la fonction publique » restait dans la même logique de décomposition des salaires que celle décrite ci-dessus. Le seul changement enregistré à ce niveau, est l'introduction du diplôme des agents comme étant la nouvelle variable d'opération de décomposition. Le calcul des indices économiques de comparaison, Gini, Theil et Atkinson est réalisé sur chaque sous-population du jeu de données.

Les résultats obtenus sur les données l'année 2003 étaient surprenants, les indices obtenus ne permettaient pas de dégager des conclusions sur les inégalités liées aux salaires des agents. Une exploration plus avancée des données a conduit à des constats très particuliers. En 2002, certains pays européens ont connu l'arrivée de la monnaie européenne, en remplacement du franc en France par exemple. Ainsi, certains agents avaient déclaré, lors de cette enquête, des revenus en francs en lieu et place de l'euro, celle recommandée dans l'étude. Du fait de cette confusion sur l'unité monétaire, un individu sans diplôme et du secteur public se retrouvait avec un salaire nettement supérieur aux salaires normaux.

Adossement à l'enseignement supérieur

Les compétences développées à la PUDC sont très variées ainsi que l'esprit du travail en équipe. Depuis le 26 février 2015, l'ingénieur d'études à la PUDC assure des travaux dirigés en techniques d'enquête pour les étudiants en deuxième année de licence « Sociologie ». Le cours magistral de technique d'enquête est assuré par Clément Pouteau, doctorant en Sociologie. Les travaux pratiques consistent à présenter aux étudiants les bases statistiques nécessaires à la mise en application de la technique de redressement d'une enquête sous Excel et le traitement des questions ouvertes sous R.

Pendant six semaines, les étudiants ont eu droit à la présentation de certaines méthodes statistiques permettant de tester la représentativité d'un échantillon suivant un certain nombre de variables d'intérêt. Les travaux pratiques réalisés sous Excel et R ont permis aux étudiants d'acquérir des connaissances sur la nécessité ou non de recourir aux opérations de redressement et aux traitements des questions ouvertes d'une enquête avec Iramuteq (une interface de R).

Adossement à des missions internes à l'université

Participation à l'observatoire genre et formation à l'UCBN

La PUDC est parfois sollicitée par la présidence de l'université pour un accompagnement ponctuel d'une action d'un de ses services (notamment la DRH) nécessitant le recours à des données quantitatives. Le plus souvent, la mission porte sur la réalisation d'un document régulier sur la situation des parcours d'étudiants et d'étudiantes, et l'organisation d'une sorte d'observatoire privilégiant la grille de lecture du genre. Un document réalisé avec Emilie Herbrecht a été remis au président Pierre Sineux puis au service de communication en juillet 2014. En voici un extrait.

Rappel du Contexte : une demande de la présidence de l'université

En avril 2011, la présidente de l'université de Caen, Josette Travert, puis le président Pierre Sineux ont demandé à Clotilde Lemarchant, MCF de sociologie et responsable de la plate-forme universitaire de données (PUDC), de devenir co-référente égalité hommes-femmes à l'université de Caen avec Patrice Georget, et de constituer une sorte d'observatoire « genre, trajectoires de formation et ressources humaines à l'université de Caen Basse-Normandie ». Ce dossier engage les co-référents égalité des chances à l'UCBN, l'Observatoire de l'université de Caen / SUOIP, la PUDC, le service inscriptions DEVE, le service ressources humaines, le service de la base Apogée au CRISI, la cellule d'aide au pilotage, le correspondant Informatique et Liberté.

Il s'agit de porter un regard sur les réalités sociales de notre établissement en fonction d'une problématique de genre et d'intégrer davantage ces réalités au projet d'établissement et au bilan social. Les réflexions en termes d'égalité professionnelle sont déclinées en deux domaines de recherche et d'action :

- Ressources humaines : à cet égard, un premier travail a été effectué en 2011 avec le mémoire de Lucie Hémerly, master 2 psychologie sociale, encadré par Patrice Georget, MCF. Les deux co-référents égalité hommes-femmes restent en lien avec la DRH afin de définir de nouveaux thèmes privilégiés d'observation (situations et trajectoires des BIATOSS, etc).

-Formation : il s'agit d'observer les trajectoires des étudiant-e-s en privilégiant la grille de lecture du genre (orientation vers les diverses spécialités des filles et des garçons, rythme des formations). Un

premier travail a été réalisé par la PUDC, en 2013¹, en lien avec l'observatoire de l'université et le Céreq-Caen. Le CEVU s'est dit très intéressé par un tableau de bord sur des données sexuées interprétées. Faire un état des lieux annuel sur les formations servira de levier éventuel pour le CA de l'UCBN. Ce projet d'étude a été inscrit dans le contrat d'établissement.

Evolution des taux de féminisation des filières à l'UCBN entre 1996 et 2013

Le graphique ci-dessous présente les taux de féminisation des filières de formation proposées à l'UCBN de 1996 à 2013. La stabilité des proportions de jeunes femmes et hommes par filière de formation est frappante, surtout dans les filières les plus éloignées de la parité, Les années 2008 et 2009, marquées par les dernières réformes des universités, bousculent les chiffres sans en transformer vraiment le sens : les chiffres oscillent autour des valeurs habituelles et les proportions restent en général les mêmes.

Evolutions des taux de féminisation des filières entre 1996 et 2013

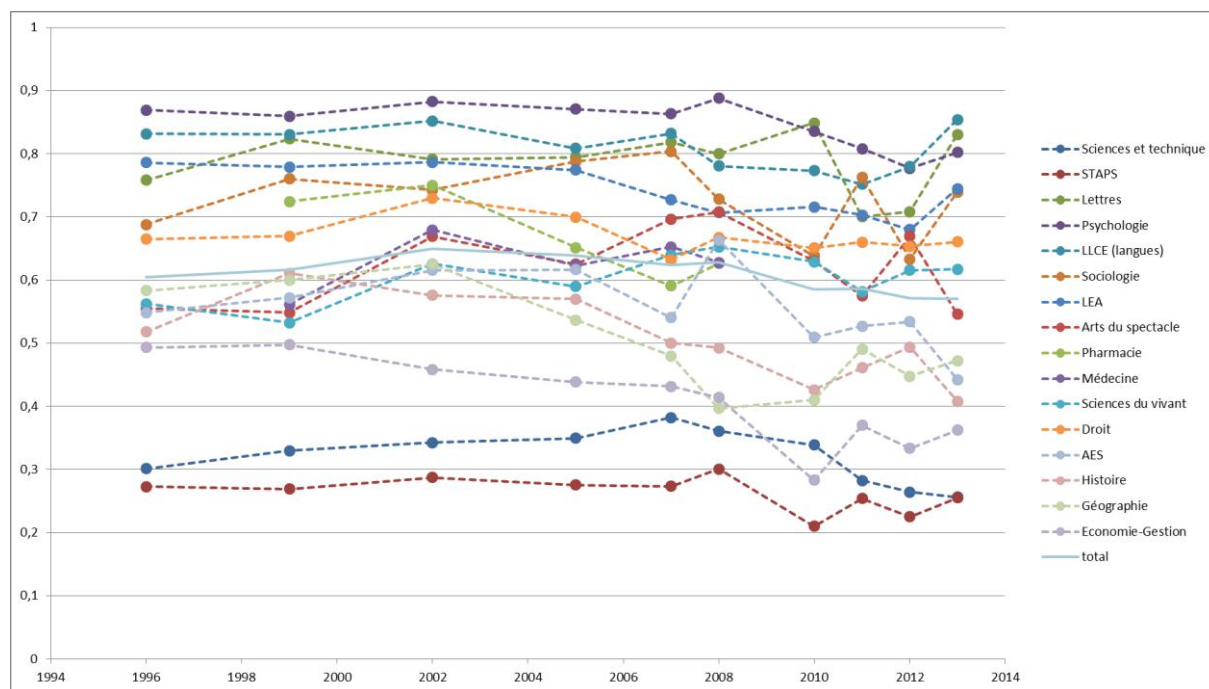


Figure 6 : Evolutions des taux de féminisation des filières entre 1996 et 2013

¹ Clotilde Lemarchant, Cyril Coinaud, *Parcours des étudiants et étudiantes de l'UCBN. Mixité, parité, égalité : une lecture des trajectoires par le genre*, UCBN, 2013.

Sources :

Pour 1996, « Les entrants à l'université de Caen Basse-Normandie », ORFS, Mars 2001.

Pour 1999, « Subanor 1999 : Les bacheliers 1999 - parcours scolaires, projets, situations à l'automne 1999 », ORFS, Juin 2000.

Pour 2002, « Subanor 2002...Après une première année dans l'enseignement supérieur », ORFS, Février 2005.

Pour 2005, « Subanor 2005 ... Qui sont les entrants en ? », ORFS, Juillet 2006.

Pour 2007, APOGEE 2007, extraction par les services de l'UCBN et calculs par nos soins.

Pour 2008, « Subanor 2008 ... Après une première année dans l'enseignement supérieur en Basse-Normandie », ORFS, Janvier 2011.

Pour 2010, 2011,2012 et 2013, extraction des données Apogée par les services de l'UCBN, calcul par nos soins (hors pharmacie et médecine).

Les primo-entrants à L'UCBN néo-bacheliers, par filière et par sexe en 2013

Ce second point présente de façon plus détaillée les choix de filière de formation universitaire de l'ensemble des étudiants et étudiantes entrant à l'UCBN, nouvellement diplômé-e-s du baccalauréat en 2013. A été reprise la présentation et la méthode d'analyse du rapport de l'an dernier pour qu'une comparaison entre 2007 et 2013 puisse être faite par les lecteurs.

En raison des effectifs trop faibles, nous avons regroupé EEA (électronique, électrotechnique et automatique) avec Physique ; Lettres classiques avec lettres modernes ;

Les filières éloignées de la parité

Le tableau 1 ci-dessous présente les filières éloignées de la parité (seuil < 40%).

Filière	Femmes	Hommes	Total
Informatique	4 (5,1%)	75 (95%)	79
EEA & Physique	3 (13%)	20 (87%)	23
STAPS	101 (25,5%)	295 (74,5%)	396
Mathématiques	8 (34,8%)	15 (65,2%)	23
Economie-Gestion	46 (36,2%)	81 (63,8%)	127
Lettres classiques & Lettres modernes	49 (83,1%)	10 (16,9%)	59
Psychologie	271 (80,2%)	67 (19,8%)	338
LLCE (langues)	181 (85,4%)	31 (14,6%)	212
Sociologie	51 (73,9%)	18 (26,1%)	69
LEA	175 (74,5%)	60 (25,5%)	235
Arts du spectacle	78 (54,5%)	65 (45,5%)	143
Sciences du vivant	95 (61,7%)	59 (38,3%)	154
Droit	344 (66%)	177 (34%)	521

Tableau 3 : Étudiant-e-s en Licence 1 par filière et par sexe en 2013 (primo-entrants)

Il y a plus de filières statistiquement féminines que de filières statistiquement masculines. Les filières « masculines » (en bleu) le sont de façon plus marquée. Ce résultat est observable à d'autres niveaux de formation (niveaux IV et V, c'est-à-dire bac et en deçà)².

² Clotilde Lemarchant (2010), «Etre rare, un avantage dans le métier ? Les filles dans les filières techniques masculines et inversement », *Vie et Santé mentale (VST)*, Revue du champ social et de la santé mentale, n°106,

Les filières mathématiques et économie-gestion étaient classées parmi les filières proches de la parité dans le rapport de l'an dernier, mais s'en sont éloignées récemment. Il reste à voir si cette tendance se renforce dans les années à venir.

On peut aussi considérer une variété de seuils qui ont un sens différent : des travaux de sociologie du travail et études de genre ont montré que les seuils de 20% et 40% ne présentent pas les mêmes réalités dans les façons de vivre sa situation de minoritaire.

Les filières proches de la parité (écart maximal de 20%)

Filière	Femmes	Hommes	Total
AES	46 (44,2%)	58 (55,8%)	104
Philosophie	9 (52,9%)	8 (47,1%)	17
Histoire	64 (40,8%)	93 (59,2%)	157
Géographie	17 (47,2%)	19 (52,8%)	36
Chimie	17 (48,6%)	18 (51,4%)	35
MASS	13 (54,2%)	11 (45,8%)	24

Tableau 4 : Étudiant-e-s en Licence 1 par filière et par sexe en 2013 (primo-entrants)

La philosophie, l'histoire, la géographie, la chimie, les MASS (Mathématiques Appliquées et Sciences Sociales) et l'AES (Administration Economique et Sociale) sont des filières qui se situent près de la parité, entre 40% et 60% d'hommes et de femmes. Entre 2007 et 2013, les filières mathématiques et économie-gestion sont passées du statut de filières proches de la parité à des filières masculines.

Lien avec l'Observatoire de l'université

Des liens réguliers avec l'Observatoire de l'université permettent des échanges de services mutuels. Le vendredi 03 octobre 2014 s'est tenue une réunion à l'Observatoire de l'université de Caen en présence de représentants de divers services « études » du Conseil régional de Basse-Normandie, du

p.57-63 ; Clotilde Lemarchant (2007), «La mixité inachevée : garçons et filles minoritaires dans les filières techniques», *Travail, genre et sociétés*, n°18, p.47-64.

rectorat, du SAIO, du Cereq, de la plate-forme universitaire de données... La séance a débuté par la présentation des activités de l'observatoire.

L'observatoire de Caen évalue périodiquement la situation scolaire des étudiants, entrants et sortants, et souhaite mettre en œuvre une étude sur les nouveaux bacheliers à N+1 visant à identifier le profil des nouveaux bacheliers décrocheurs ainsi que ceux qui changent en cours de formation au sein de l'établissement. L'étude permettra aux rectorats, aux autorités universitaires et aux collectivités régionales d'améliorer le dispositif éducatif de lutte contre le décrochage.

Ainsi, à travers un questionnaire envoyé par mail, les étudiants sont invités à répondre à un certain nombre de questions portant sur leur scolarité. Après les analyses de données, certains résultats statistiques d'enquêtes réalisées par l'observatoire sont disponibles sur le site de l'Université de Caen. L'objectif de cette rencontre, animée par Laurent RAOUL, responsable de l'Observatoire, était d'inciter les participants à discuter le cadrage et le contenu du questionnaire.

Aide au service de scolarité de l'UFR de Sciences de l'Homme

Cette année, le service de la PUDC a pour mission au sein de l'université, le traitement statistique de données du projet «Amélioration Qualité Etablissement» et la présentation des résultats. Sous la responsabilité de Noëlle Gambey, du service de scolarité des sciences de l'homme, une rencontre est prévue très prochainement pour redéfinir les objectifs recherchés.

Projets, perspectives et questions

La PUDC compte mettre en place un programme d'archivage de données afin de permettre la réutilisation d'un certain nombre d'enquêtes locales pour la recherche. C'est le cas de l'enquête réalisée par le Cereq dans le cadre du projet INEDUC et du panel de jeunes caennais, coordonné par Claire Bidart, Directrice de recherche au CNRS.

Le monde des grandes enquêtes est en perpétuel renouvellement et des questions nouvelles se posent à la PUDC : notamment, le rôle croissant des procédés de sécurisation des données. Il est en effet de plus en plus difficile de disposer sur son propre ordinateur portable d'enquêtes conséquentes jusque-là aisément transposables. De nombreux collègues s'interrogent sur les freins et le coût du recours au CASD, centre d'accès sécurisé aux données, et de la politique tarifaire de ce dernier.

Bilan des publications et travaux en lien avec la PUDC

BOUDESSEUL Gérard, CARO Patrice, THEMINES Jean-François, Pole FETE, MRSH, 2014, L'école à l'épreuve des pratiques des élèves (journée autour d'Anne Barrère) : Discussion avec des chercheurs de l'ANR INéduc sur les loisirs des 12-15 ans

P. CARO, O. DAVID, C. VIVENT, 2014, « Caractéristiques socio-spatiales de résidence et pratiques de temps libre des jeunes », Colloque "L'espace en partage", UMR ESO, Rennes, 9 avril 2014.

CARO Patrice, 2015, « présentation des 1ers résultats de l'ANR Inéduc », Rencontre sur l'école rurale avec l'Observatoire éducation et territoires (OET), pole FETE, MRSH, Université de Caen, 28 janvier 2015.

DANIC Isabelle, Teenage girls and leisure in different French contexts, XVIII ISA World Congress of Sociology, organisé par l'International Sociological Association, Yokohama, 13-19 juillet 2014.

DANIC Isabelle, «La construction de l'adolescence féminine par les contextes socio-spatiaux», Colloque «L'espace en partage», organisé par ESO, Rennes 9-11 avril 2014.

CHANTREUIL Frédéric, LEBON Isabelle, « Gender contribution to income inequality », december 11-2014.

Bilan financier

Voici le bilan financier de juin 2014 à mars 2015, les dépenses du 1er semestre 2014 ayant été fournies pour le précédent bilan (voir rapport d'activité PUDC 2013-2014). Le principal poste budgétaire porte en grande partie sur le salaire de l'ingénieur statisticien. Le budget des salaires est entièrement pris en charge par l'université de Caen. Pour la période allant du 25 juillet 2014 au 15 septembre 2014, il n'y a eu personne à la PUDC.

Salaires juin 2014-mars 2015

Pour Emilie Herbrecht :

Juillet, aout 2014 : 5 092,20 euros

Pour Alpha Oumar Diop :

15 septembre 2014-mars 2015 : 22 886,60 euros

total : 27 978 euros

Frais de fonctionnement

Les dépenses en frais de fonctionnement sont entièrement prises en charge par la MRSH de Caen.

Frais de restauration	66,00 €
Frais de déplacements	603,34 €
Logiciels SAS	518,40 €
Logiciels Statistica	196,00 €
Frais téléphoniques	90,00
Frais reprographie	140,00
Total	1 613,74

Tableau 5 : Tableau récapitulatif des frais liés au fonctionnement de la PUDC

La PUDC s'oriente de plus en plus vers l'utilisation du logiciel R, gratuit. Il n'y a pas eu de frais de déplacement pour le séminaire IASSIST en 2014 (International Association for Social Science Information Service and Technology). Il en est prévu un du 2 au 5 juin 2015 à Minneapolis (USA). Alpha Oumar Diop s'y est inscrit.

Glossaire

UCBN : Université de Caen Basse Normandie

MCF : Maître de Conférences à l'UCBN

PUDC : Plate-forme Universitaire de Données de Caen

INSEE : Institut National de la Statistique et des études économiques

MRSH : Maison de Recherche en Sciences Humaines et Sociales

TGIR : Très grandes infrastructures de recherche

PROGEDO : Production et Gestion des Données

CESSDA : Council of European Social Science Data Archives

CASD : Centre d'Accès Sécurisé Distant aux données

CDSP : Centre de Données Socio-Politiques

CMH-ADISP : Centre Maurice Halbwachs - Archives de Données Issues de la Statistique Publique

INED : Institut National d'Etudes Démographiques

UFR : Unités de Formation et de Recherche

INRS : Institut National de Recherche et de Sécurité

IASSIST : International Association for Social Science Information Service & Technology

Liste des tableaux

Tableau 1 : extrait du rapport d'activités octobre 2013-mars 2014.....	8
Tableau 2 : Tableau récapitulatif des demandes de fichiers de données auprès de quetelet.....	15
Tableau 3 : Étudiant-e-s en Licence 1 par filière et par sexe en 2013 (primo-entrants)	24
Tableau 4 : Étudiant-e-s en Licence 1 par filière et par sexe en 2013 (primo-entrants)	25
Tableau 5 : Tableau récapitulatif des frais liés au fonctionnement de la PUDC.....	29

Liste des figures

Figure 1 : Répartition des utilisateurs suivant leur statut	9
Figure 2 : Répartition des utilisateurs suivant leur discipline	10
Figure 3 : Répartition des utilisateurs suivant leur institution.....	11
Figure 4 : Répartition des utilisateurs suivant les besoins	12
Figure 5 : Temps de travail présentiels en heures.....	13
Figure 6 : Evolutions des taux de féminisation des filières entre 1996 et 2013.....	22

Annexes

Comité de pilotage de la PUDC (20/06/2014)

Pascal Buléon ouvre la séance du comité de pilotage et cède la parole à Clotilde Lemarchant qui remercie les personnes présentes et cite les personnes excusées.

Un petit tour de table est fait en guise de présentation.

- Alain Hugon représente l'histoire
- Anne Pelissier les sciences de l'éducation
- Jean-Marc Fournier, directeur d'Eso
- Patrice Caro, Céreq
- Denis Jacquet, MRSH
- Pascal Buléon, MRSH, Progedo
- Roxane Silberman, GIS Quetelet, CESSDA
- Daniel Delahaye, Vice-président UCBN
- Christophe Leconte, direction de la recherche UCBN
- Dominique Beynier, Sociologie
- Anne-Julia Rollet, Geophen

Clotilde Lemarchant et Emilie Herbrecht présentent le bilan et le rapport d'activité pour l'année 2013-2014. Un document est remis à chaque participant.

Dominique Beynier souligne le fait que la PUDC a joué un rôle de support lors de l'installation du logiciel SAS notamment avec les étudiants de master MESS. Il insiste sur les problèmes de gestion informatique au sein de l'université. Bon retour des étudiants sur la PUDC. Il souligne également l'importance de l'archivage et de la pérennité du poste.

Patrice Caro est satisfait de l'aide et du travail fournis par la plateforme. Il insiste sur les problèmes de carence du Céreq (départ à la retraite d'Yvette Grelet). La pérennité du poste de statisticien serait également une bonne chose.

Jean-Marc Fournier explique que la plateforme est plus qu'essentielle pour son labo et remercie Clotilde et Emilie pour la qualité du travail.

Anne Pelissier précise que les doctorants et étudiants de Sciences de l'Education sont très intéressés et satisfaits de la PUDC. Elle espère que la pérennité du poste et la possibilité d'archiver via la plateforme puisse se faire.

Alain Hugon explique qu'il n'est pas utilisateur de la plateforme, mais qu'il vient pour y adresser son soutien, car pour lui c'est un outil important.

Anne-Julia va parler de la plateforme aux doctorants et à ses collègues et viendra suivre une formation sur R et SAS.

Roxane Silberman souligne le fait que le nombre d'utilisateurs et le listing des publications sont très importants.

Daniel Delahaye félicite le fonctionnement de la plateforme qui est un exemple. Il apprécie la forme du rapport d'activité très propre et réutilisable. La plateforme tourne et l'UCBN continuera à financer le CDD, mais il restera probablement de courte durée.

Denis Jacquet pense qu'il faudrait contacter la cellule archivage de la MRSH. Il insiste tout comme Dominique Beynier sur les ennuis informatiques au sein de l'université. Une coordination serait utile entre la PUDC et la DSI en ce qui concerne les formations sur certains logiciels informatiques (R, SAS...).

Clotilde Lemarchant clôture le tour de table et donne la parole à Roxane Silberman.

Roxane Silberman se réjouit de la pérennité de la PUDC et souligne la différence intéressante entre la plateforme de Lille et celle de Caen. Elle demande si les commandes de données répertoriées dans le tableau à la page 12 du rapport d'activité ont toutes été passées via la plateforme. Clotilde lui répond que non. Roxane demande également si les utilisateurs de la plateforme se renseignent sur le CASD et souligne le fait qu'il serait important de faire une présentation du CASD aux chercheurs, car il est fort probable que dans les années à venir il faille passer par le CASD pour commander des données. L'idéal visé est un espace de communication virtuelle qui permettrait d'éviter les locaux contrôlés.

Dominique Beynier explique qu'il existe des problèmes dans le cadre du CASD pour l'obtention des données (délai d'attente très long 18 mois dans son cas).

Roxane Silberman explique que les réunions au CASD sont peu nombreuses (4 par an) et que normalement, il devrait y en avoir plus dans les années à venir. Dans une dizaine d'années, des boîtiers disposant des données vont se développer. Il sera alors possible de travailler à distance (Espagne, Allemagne,...). La PUDC doit jouer un rôle de documentation. Pour l'archivage, il faudrait établir des règles sur la mise à disposition des données. Il faudrait également mettre les données à disposition via le site du réseau Quetelet.

FICHE PROFIL DE POSTE

BAP : D

CORPS : Ingénieur-e d'études

NATURE : Contractuel-le 3 ans

ETABLISSEMENT : Université de Caen Basse-Normandie

UFR ou SERVICE : Maison de la Recherche en Sciences Humaines, plate-forme universitaire des données de Caen (PUDC)

DESCRIPTION DU POSTE

Contribuer à la diffusion, auprès du personnel de recherche et d'étudiant-e-s normand-e-s, de données d'enquêtes et de bases de données dans le domaine des sciences humaines. Aider les chercheur-e-s dans la recherche et l'utilisation de ces données.

ACTIVITES

Suivre les évolutions des programmes d'enquêtes de la statistique publique et veiller au calendrier de leur diffusion.

Accompagner des chercheur-e-s et étudiant-e-s dans leur recherche et acquisition de fichiers issus de grandes enquêtes : choix des fichiers adaptés et des organismes, déontologie à respecter dans l'utilisation de ces fichiers, démarches administratives, mise en œuvre de la production de données (questionnaires, échantillonnage), aide à l'interprétation des analyses statistiques (analyse de variance, classification, régressions, chi2...), mise en œuvre de la documentation de travail en vue de l'archivage (questionnaire, plan de sondage, liste et contenu des variables...).

Organiser des formations en matière de statistique (interprétation des tableaux de contingence, description d'une distribution statistique, tests, méthodes d'analyse ou d'exploration des données) et d'informatique (gestion de fichiers, navigation sur le web, logiciels statistiques, notamment SAS et/ou R).

COMPETENCES ET CONNAISSANCES SOUHAITEES

Maîtrise avancée des procédures de gestion et d'analyse statistique d'un logiciel d'analyse de données (SAS, R).

Maîtrise du système d'exploitation Windows et d'Access.

Maîtrise des outils bureautiques usuels.

Bonne culture en sciences humaines et sciences sociales (sociologie, économie, histoire, géographie...) et notamment en matière de production de données d'enquêtes par sondage et de base de données.

Connaissance approfondie des normes internationales et nationales en matière de nomenclatures usuelles dans les SHS (emploi, éducation, géographiques...). Connaissance approfondies des règles déontologiques en matière de grandes enquêtes.

Maîtrise du traitement de données statistiques en SHS.

Contribuer à la réflexion sur les techniques et méthodes en Sciences Humaines et Sociales

Aisance en anglais, au moins écrit.

QUALITES REQUISES

Aptitude au travail en équipe dans un contexte de projet, aisance relationnelle.

CONTEXTE : La plate-forme universitaire des données de Caen est partenaire du réseau Quételet/PROGEDO. Elle est amenée à nouer des relations privilégiées avec l'ADISP (Archives de données issues de la statistique publique) du Centre Maurice Halbwachs, le service des enquêtes de l'INED et le CDSP (Centre de données socio-politiques) de Sciences Po Paris. Enfin, elle est amenée à échanger avec les deux autres plates-formes existantes de Lille et de Lyon.