

EDITO



Didier Breton,

Responsable scientifique de la plate-forme universitaire de données de Strasbourg

Sous l'impulsion de PROGEDO, le réseau des plateformes universitaires de données (PUD) se densifie et se structure peu à peu. Ces dispositifs, nés de l'expérience lilloise, sont des relais locaux, jouant le rôle de facilitateurs d'utilisation de données d'enquêtes existantes auprès de la communauté des chercheurs en Sciences Humaines et Sociales. Leur ancrage dans les MSH prend alors tout son sens. Le travail quotidien des PUD est d'apporter un soutien, le plus souvent personnalisé, aux chercheurs et étudiants dans deux cas de figure. Le premier est celui de chercheurs ou étudiants en quête de jeux de données pour une recherche ou l'écriture d'un projet de recherche. Le deuxième est celui de chercheurs ou étudiants ayant identifié des données mais souhaitant un accompagnement pour leur exploitation. A chaque fois les PUD réalisent un travail sur-mesure, une aide personnalisée, c'est leur plus-value. Certes ces structures ont un coût, mais si les établissements décident de les financer c'est bien qu'elles sont au cœur de la production de la recherche des prochaines années en SHS.

Mais au-delà du travail des PUD, le développement d'une pratique des données d'enquête en Sciences Humaines et Sociales ne sera possible qu'à condition d'un enrichissement de l'enseignement des méthodes indispensables à cela dès les premiers cycles, qu'il soit obligatoire ou optionnel. Si des enseignements en statistiques et enquête existent, celui de langages de programmation et

d'algorithmique l'est beaucoup moins. Une maîtrise dans ce domaine est pourtant la condition pour l'acquisition d'une autonomie et d'une liberté par des futurs chercheurs qui auront à leur disposition tant de données, qu'elles soient structurées ou non, qu'elles soient ou non massives.

Différentes expériences pédagogiques prouvent aujourd'hui la facilité avec laquelle des étudiants, souvent issus de filières non scientifiques, adhèrent à la logique algorithmique. C'est le cas à Strasbourg à travers la mise en place d'une unité d'enseignement, financée par l'Idex, intitulée « Pour une pratique et une culture du numérique en SHS », ouverte à des étudiants de Licence de sociologie, d'archéologie, de géographie, de démographie et sciences politiques. Cette expérience pédagogique, mettant la donnée au centre, participe alors au décroisement disciplinaire si souvent souhaité.

Ici encore PROGEDO et ses relais locaux que sont les PUD devront jouer un rôle important. Une fois acquise l'autonomie souhaitée, la mise à disposition de données d'enquête « réelle », par exemple sous forme de fichiers pédagogiques, permettrait aux étudiants dès le premier cycle de reproduire relativement facilement des résultats publiés dans des revues de référence et, par la suite, en deuxième et troisième cycle, d'en imaginer des prolongements dans le cadre de travaux inédits, instaurant ainsi une dynamique de recherche par et grâce aux données d'enquête.

Bilan du colloque QuantiSES

Les 4 et 5 septembre dernier s'est tenu à Paris le Colloque QuantiSES organisé par le Conseil Scientifique de PROGEDO à l'Université Paris Diderot autour des questions liées à l'enseignement des méthodes quantitatives en sciences économiques et sociales en France.

Au fil de ces trois demi-journées, une trentaine d'intervenants se sont succédés pour témoigner des nouveaux métiers, des dispositifs d'accès aux données, des logiciels de traitement et de l'état de leurs enseignements en sciences économiques et sociales.

Retrouvez le bilan de ces journées dans **la lettre de l'INSHS de septembre 2017.**

Les données de santé en France

Dans le cadre du projet européen CESSDA SaW, PROGEDO a pour mission la rédaction d'un rapport sur les données de santé en France (Health data). Pour préparer ce rapport, PROGEDO a rencontré trois experts en données de santé qui participent à la mise en place d'enquêtes aux niveaux national et européen :

- **Catherine Sermet**, Directrice adjointe de l'IRDES (Institut de Recherche et Documentation en économie de la santé) et membre du Conseil Scientifique de PROGEDO,
- **Emmanuelle Cambois**, Directrice de recherche à l'Ined (Institut National d'études démographiques) et membre du Conseil Scientifique de PROGEDO,
- **Florence Jusot**, Professeur de sciences économiques à l'Université Paris-Dauphine et Co-country team leader du projet SHARE-France.

PROGEDO actu en profite pour faire un point sur l'état de la gestion des données de santé en France.

On parle beaucoup des « données de santé », pourriez-vous nous aider à définir ce qu'elles recouvrent en France ?

Florence Jusot : « Les données de santé sont explorées en France depuis une trentaine d'année. Sur les données génériques, il y a eu deux grands piliers qui ont permis de suivre les tendances sur le long cours : l'enquête ESPS (enquête sur la santé et la protection sociale, 1988-2014) menée tous les deux ans par l'IRDES et l'enquête décennale de santé menée par l'Insee (1960-2003). Il y a aussi d'autres études menées sur des thématiques plus spécifiques: les enquêtes HID (handicaps-incapacités-dépendance) puis Handicap-Santé pour analyser les situations de handicap, ou encore l'enquête CARE (Capacités, Aides et Ressources des seniors) qui interroge les personnes âgées. On compte également de nombreuses enquêtes ponctuelles de l'IRDES, l'Ined ou la DRESS (Direction de la recherche, des études, de l'évaluation et des statistiques, gérée par les ministères sanitaires et sociaux). Les baromètres de santé qui cherchent à étudier les comportements à risque (alcool, tabac, drogues illégales, etc.) sont une autre source de données très intéressante.»

Emmanuelle Cambois : « Les « données de santé », c'est un concept assez large. On y inclut d'une part des données administratives concernant l'offre de soins, avec par exemple la Statistique annuelle des établissements de santé (SAE) ou les données de démographie médicale (DREES, Assurance maladie ou CNOM, le Conseil national de l'Ordre des médecins), ou concernant la demande de soins, avec en particulier les données de l'Assurance Maladie (SNIRAM), des hôpitaux (PMSI) et les causes médicales de décès. Ces bases de données médico-administratives comportent généralement peu d'information sur les situations sociodémographiques des personnes. Mais les données de santé peuvent aussi couvrir les données d'enquêtes sur l'état de santé de la population. À partir de d'informations recueillies auprès d'échantillons représentatifs de la population vivant dans des ménages ordinaires, ce type d'enquête permet de connaître différents aspects de la santé et des conditions de vie des enquêtés : il y a par exemple l'enquête décennale de l'Insee, les enquêtes européennes, ou encore les enquêtes sur la santé et la protection sociale de l'IRDES. Ces données sont généralement déclaratives et couvrent des échantillons relativement grands ; elles permettent de balayer toutes sortes d'informations sur la santé (symptômes, maladies, incapacités) et de les étudier au regard de la sphère sociale, professionnelle ou familiale.

Le troisième corpus, ce sont les données épidémiologiques avec des suivis de groupes de population, par exemple des patients d'un service hospitalier ou des personnes atteintes d'une maladie. Ces données sont généralement précises en termes de diagnostic et permettent de mesurer l'évolution de l'état de santé ; elles sont toutefois ciblées sur des groupes particuliers. Au croisement de ces trois types de données - administratives, déclaratives

et épidémiologiques, le dispositif CONSTANCES est une cohorte de grande échelle basée sur les informations de personnes volontaires qui sont invitées à se présenter au centre d'examen de santé de la sécurité sociale pour avoir un examen santé et remplir un questionnaire spécifique; ces personnes sont alors suivies dans le temps. Cette cohorte fournit donc des données de différents types, en lien avec des examens de santé.»

Comment l'accès à ces données de santé est-il géré en France?

Catherine Sermet : « Les bases de données de santé médico administratives sont désormais plus facilement accessibles grâce au Système national de données de santé (SNDS), qui comporte les données de l'Assurance Maladie (SNIRAM), c'est-à-dire les prestations remboursées pour les consultations chez le médecin, le recours aux soins d'un infirmier ou d'un kinésithérapeute, les médicaments, etc., les données des hôpitaux (PMSI), et les causes médicales de décès (base CépiDC de l'Inserm). Ces données vont être progressivement enrichies avec les données relatives au handicap en provenance des Maisons Départementales des Personnes Handicapées (MDPH) ainsi que, pour un échantillon d'assurés, les données des remboursements de l'Assurance maladie complémentaire. L'Institut national de données de santé (INDS) assure la mise à disposition de ces données en vue de recherches, études et évaluations présentant un intérêt public. La procédure d'accès comprend deux étapes : la validation de la méthodologie et des finalités de l'étude par le Comité d'Expertise pour les Recherches, les Etudes et les Evaluations dans le domaine de la Santé (CEREES) suivie d'une autorisation de la Commission Nationale Informatique et Libertés (CNIL). A noter qu'un nombre limité d'institutions et d'organismes ont un accès permanent à cette base de données.

Par ailleurs, concernant les enquêtes santé, outre un accès direct auprès des producteurs, elles sont mises à disposition des chercheurs par l'intermédiaire de Quételet Progedo diffusion dont les partenaires sont les Archives de données issues de la statistique publique, le CDSP, Centre de données socio-politiques, le Service des Enquêtes de l'INED, Institut National d'Etudes Démographiques et le CASD, Centre d'Accès Sécurisé aux Données.»

Y a-t-il des orientations qui se dessinent pour l'utilisation des données de santé ?

Florence Jusot : « La tendance va à l'appariement de données, notamment avec des sources administratives car celles-ci sont fiables et n'ont pas le biais de réponses. Par exemple, si je demande aux gens combien de fois ils sont allés chez le médecin dans l'année, ils en auront une vague idée et me répondront de manière approximative, alors que je peux trouver l'information exacte dans les données administratives. Le bémol des données administratives, c'est que leur contenu est décidé par l'administration qui

les collecte et non par la recherche scientifique. Dans le cadre d'une enquête, le chercheur choisit les questions en fonction de sa problématique scientifique, mais lorsqu'il utilise les données administratives, il travaille à partir des informations pertinentes pour l'administration qui ne correspondent pas nécessairement à toutes les dimensions qu'il souhaiterait pouvoir étudier. Par contre, ces données administratives ont l'avantage d'être extrêmement représentatives car elles ne sont pas soumises à l'acceptation du répondant. Je pense que c'est pour toutes ces raisons que l'on va vers la complémentarité des sources.»

DONNÉES

Nouveautés 2^{ème} trimestre 2017

CDSP - nouvelles données ELIPSS disponibles

Le Centre de données socio-politiques (CDSP) continue l'ouverture des données ELIPSS et met à disposition deux enquêtes qui étudient la place du numérique et des technologies de l'information :

la 1^{ère} vague des Pratiques numériques mesure l'évolution des comportements liés aux technologies de l'information et de la communication. Cette série d'enquêtes s'inscrit dans la tradition des enquêtes TIC (Technologie de l'information et de la communication de l'INSEE), et décrit l'équipement des ménages (ordinateur, accès internet, etc.), ainsi que les usages numériques des répondants (téléphonie, lecture en ligne, etc.). Dans un panel où les membres ont été équipés d'une tablette tactile connectée, la première vague des Pratiques numériques est une source riche de connaissances ; en plus d'être un référentiel à la mesure de l'évolution des pratiques, elle permettra d'évaluer l'impact en population générale d'un équipement connecté à Internet.

Enquête proposée aux membres du panel ELIPSS dès leur entrée dans le dispositif en 2013, elle est depuis administrée annuellement.

7^{ème} appel à projets d'enquêtes ELIPSS

Le CDSP s'est engagé en juillet dernier son 7^e appel à projets d'enquêtes ELIPSS ouvert jusqu'au 20 novembre 2017.

Le panel ELIPSS (Étude Longitudinale par Internet Pour les Sciences Sociales) est un dispositif d'enquêtes par Internet destiné à la communauté scientifique. Original et innovant, ELIPSS, donne une opportunité unique aux chercheurs en SHS de réaliser gratuitement des enquêtes par questionnaire, collectées par l'Internet mobile, auprès d'un échantillon aléatoire de 2 900 individus de la population française métropolitaine. Chaque mois, un questionnaire de 30 minutes maximum est adressé via une application dédiée à ce panel.

Les enquêtes soumises au panel sont élaborées par des chercheurs et sélectionnées par un comité scientifique et technique. Sous réserve de respecter une finalité scientifique, toute équipe de recherche peut soumettre un projet dans le cadre des appels à projets d'enquêtes ELIPSS.

Les enquêtes sélectionnées lors de l'appel en cours seront administrées aux panélistes dès la fin d'année 2018.

Rendez-vous sur <http://quanti.dime-shs.sciences-po.fr> pour télécharger le texte de l'appel à projets.

Catherine Sermet : « Je pense que dans un premier temps, le développement des travaux de recherche sur les données du Système National de Données de Santé va prendre de l'ampleur, d'autant plus qu'à moyen terme, elles sont appelées à s'enrichir. Malgré tout, ces données administratives manquent d'informations précises sur les dimensions sociales, les caractéristiques d'état de santé ou les facteurs de risque. Ces éléments ne peuvent être étudiés qu'à partir des enquêtes santé dont nous aurons donc toujours besoin. Les appariements entre données d'enquête et données médico-administratives se développent de plus en plus permettant de compléter les informations de part et d'autre.»



ESS - Défi ESS, un concours d'articles pour étudiants

Le Défi ESS, un concours d'articles de recherche francophones à destination des étudiants de 2^{ème} cycle (en master ou récemment diplômés) est en cours. Il vise à récompenser des papiers de recherche basée sur l'analyse quantitative d'un des modules rotatifs de la 7^{ème} édition de l'European Social Survey (ESS) dédiés :

- d'une part aux inégalités sociales de santé,
- et d'autre part à l'immigration.

Organisé par PROGEDO en lien avec les équipes de coordination nationale française, belge et suisse, le défi ESS est une occasion unique pour les étudiants de faire connaître leurs travaux et de se voir récompensés pour leur qualité.

Les trois meilleurs papiers seront publiés sur les sites des équipes de coordination nationale organisatrices et sur le site de l'ERIC ESS. Des bourses financières seront également attribuées aux deux premiers articles sélectionnés (à hauteur de 300 et 100).

Les pré-inscriptions au défi ESS sont ouvertes du 1^{er} octobre au 20 novembre 2017 ; le retour des papiers définitifs est attendu pour le samedi 16 juin 2018, 18h, heure de Paris.

L'appel à candidature détaillé est disponible sur le site www.progedo.fr

ESS - Appel pour les modules de la 10ème édition

L'appel à projet pour les modules de la 10ème édition de l'Enquête Sociale Européenne est en cours. Les équipes de recherche multi-nationales en sciences sociales ayant un attrait pour les démarches comparatives sont invitées à soumettre, au plus tard le 18 janvier 2018, les projets de modules de questions qu'elles souhaiteraient voir intégrer à la 10ème édition de l'ESS (terrain prévu pour septembre 2020, et données accessibles à partir de fin 2021).

Le dossier de candidatures se constitue d'un court résumé du projet soulignant la pertinence du module proposé pour des recherches transnationales, le calendrier de travail, et les qualifications des membres de l'équipe soumissionnaire concernant le design de questionnaire.

L'appel à proposition complet est disponible sur le site de l'ERIC ESS : www.europeansocialsurvey.org/

ÉVÉNEMENTS

GGP - mises en avant automnales

GGP - Journée scientifique « Les comparaisons internationales avec les données du projet Générations et Genre »

Dans le cadre du Pôle Perspectives Internationales, l'Institut National d'études démographiques (Ined) organise une journée scientifique le 24 Novembre 2017 (Ined, salle Sauvy, 9h30-17h00), sur le projet Generations and Gender Programme (GGP).

Cette journée vise à promouvoir l'utilisation des données de GGP dans des analyses comparatives. Dans ce but, la matinée sera consacrée à la présentation du projet en général, l'historique, les données et la méthodologie, en incluant également les développements actuellement en cours. L'après-midi verra la présentation de travaux comparatifs de quatre chercheurs français et étrangers. Le programme de la journée sera diffusé prochainement sur le site de l'Ined.

Informations pratiques : Salle Sauvy, Ined, 133, boulevard Davout, 75020 Paris

GGP - 28ème Congrès international de la population

GGP participe du 29 octobre au 4 novembre 2017 au 28ème Congrès international de la population à Cape Town (Afrique du Sud).

Notamment le mardi 31 octobre de 19h00 à 21h00, une réunion parallèle abordera la thématique « Quelles sont les innovations dans les méthodes de collecte des données et de « data linking » que GGP est en train de mettre en place afin de faire face aux enjeux posés par les enquêtes longitudinales ? Comment pouvez-vous devenir un point focal de GGP dans votre pays? ».

De plus, du 30 octobre au 3 novembre, un stand d'exposition GGP accueillera les utilisateurs et toute personne intéressée par GGP. L'équipe de coordination et des conseillers techniques de GGP seront présents pour répondre à toutes les demandes d'information.

Toutes les informations sur le site du congrès: <http://ipc2017lecap.iussp.org>

Les actualisations seront communiquées via les comptes Twitter et Facebook de GGP.

ACTUALITÉS DES DÉPARTEMENTS

PUD - Nouvelle responsable scientifique à la PUD de Lille

Suite au départ de Frédérique Cornuau, Héloïse Petit est devenue en juillet 2017 la nouvelle responsable de la Plateforme Universitaire de Données de Lille (PUDL). Professeur en économie, Héloïse Petit est à l'Université de Lille 1 depuis 2015 après avoir été en poste à l'Université Paris I, puis en détachement au Centre d'Etudes de l'Emploi (CEE). Economiste du travail, elle est aujourd'hui chercheuse au Clersé, et associée au CEET. Ses travaux portent sur la relation salariale et la diversité des formes de gestion de l'emploi. Le recours aux données est une constante, tant de son travail de recherche que d'enseignement. Elle a ainsi une grande pratique des modalités de mise à disposition des données en tant qu'utilisatrice mais aussi des voies de leur construction via la participation à des comités de pilotage d'enquête ensuite expert national Open Science à la Commission Européenne pour la Direction Générale de la Recherche et de l'Innovation avant de rejoindre CESSDA au 1er Mars 2017.



GGP - un site web plus proche des utilisateurs

Le programme Generations and Gender s'est doté d'un nouveau site web. Axé sur la communication avec les utilisateurs, le site <http://www.ggp-i.org> permet d'obtenir plus d'informations sur le projet et ses plans pour l'avenir. C'est également un point privilégié pour garder contact avec le programme, accéder à sa newsletter et à ses comptes sur Twitter et Facebook

Les commentaires sur le nouveau site peuvent être transmis aux gestionnaires du projet via le formulaire de la rubrique contact: <http://www.ggp-i.org/about/contact/>.