

EDITO



Françoise Thibault
Déléguée générale
de l'Alliance Athéna

PROGEDO : des initiatives de chercheurs à la maturité institutionnelle

PROGEDO a vu le jour en 2008 en même temps qu'était reconnu aux sciences humaines et sociales le besoin de disposer de très grandes infrastructures de recherche (TGIR). Cette reconnaissance officielle par le ministère français en charge de la recherche, qui a pris la forme d'une mention dans la liste publique des TGIR, s'inscrivait dans le sillage d'une initiative prise en 2006 par le forum stratégique européen pour les infrastructures de recherche (ESFRI) qui a décidé de publier la première feuille de route européenne des infrastructures de recherche.

Il n'a pas échappé aux observateurs des politiques de recherche de l'époque que cet ensemble de chercheurs confirmés issus d'un large spectre disciplinaire, réunis par la Commission européenne pour dessiner de grandes voies stratégiques étaient allés au-delà de la commande initiale. La liste détaillée d'infrastructures jugées par eux essentielles pour l'avenir de la science européenne n'était pas attendue tout comme le spectre disciplinaire qu'ils avaient retenu et qui n'était plus limité à toutes les filières de la physique et de l'astronomie. Il s'ouvrait à la biologie, à l'informatique et aux sciences humaines et sociales.

En France, PROGEDO a été imaginé après la réalisation d'un état de l'art des enquêtes et dispositifs d'accès aux données de la statistique publique et après le constat renouvelé des besoins financiers nécessaires pour réaliser les grandes enquêtes scientifiques. Définie comme « un outil établi en vue de mener une recherche propre d'importance et pouvant assurer une mission de service pour une ou plusieurs communautés scientifiques de grande taille », les

TGIR de 2008 sont caractérisées par leur coût de construction et d'exploitation. Ces derniers critères étaient inapplicables à PROGEDO comme à bien d'autres infrastructures inscrites dans la feuille de route. Toutes ont des coûts de fonctionnement beaucoup plus significatifs que leurs coûts de construction. L'énonciation de ces critères a été purement rhétorique et n'a pas écarté les disciplines entrantes pour lesquelles les TGIR étaient devenues une nécessité.

Présente maintenant dans le paysage institutionnel, la TGIR PROGEDO a connu de profondes évolutions depuis quelques années. Elle s'est structurée en départements dont le nombre croît comme en témoigne la présente lettre. Elle a renforcé sa gouvernance avec l'appui de l'InSHS. Les plateformes universitaires de données (PUD) se sont déployées sur le territoire grâce au financement obtenu dans le cadre du Plan Vidal. L'articulation effective avec les MSH a permis d'intervenir plus près du monde de la recherche et de faciliter le dialogue entre les personnels qui produisent, utilisent et gèrent les données. En janvier 2018, la transformation de l'unité de service Progedo, en Unité de Service et de Recherche (USR) a constitué une preuve de la confiance institutionnelle désormais acquise.

Si cette évolution est réjouissante pour qui est persuadé de l'importance de l'usage scientifique des données et de la réflexion épistémologique qui doit l'accompagner, elle exige que de nouveaux efforts financiers soient faits pour que PROGEDO soit en mesure de faire face aux défis de formation et au financement des grandes enquêtes européennes indispensables à la connaissance des sociétés contemporaines.

STRUCTURATION DE LA RECHERCHE

L'ACCÈS AUX DONNÉES PAR LE PORTAIL QUETELET-PROGEDO-DIFFUSION.

Ouverture des données et protection des individus

En 2018, la création du Comité pour la Science Ouverte (<https://www.ouvrirlascience.fr>) marque en France un jalon dans l'organisation institutionnelle de la science et des données scientifiques. PROGEDO était déjà largement engagé dans cette démarche qui consiste, notamment, à « ouvrir les données autant que possible et à les fermer autant que nécessaire ».

En effet, la mission de PROGEDO est bien de développer les accès aux données d'enquêtes et aux bases de données quantitatives en SHS, ce qui s'exprime à travers la production d'enquêtes et leur mise à disposition ; le financement de l'accès à des bases de données nationales et internationales ; l'accompagnement des communautés de recherche dans l'utilisation de ces données.

« Quetelet-PROGEDO-Diffusion » : un portail commun d'accès aux données

Le portail d'accès aux données « Quetelet-Progedo-Diffusion » est un outil essentiel qui repose sur la collaboration de plusieurs équipes (Service des enquêtes de l'INED, Centre de Données de Sciences Po, Centre d'Accès Sécurisé aux Données, Archives de Données Issues de la Statistique Publique). Il s'agit d'offrir à l'ensemble de la communauté de recherche un accès documenté aux sources disponibles, mais aussi de rendre ces sources accessibles. Ainsi, l'équipe de l'Adisp met à disposition les données de la statistique publique produites par l'Insee et les services statistiques ministériels et publics ; le service des enquêtes de l'Ined diffuse les productions de l'institut et le CDSP un corpus de données d'enquêtes socio-politiques issues de la recherche académique ainsi que l'ensemble des résultats électoraux de la V^e République. Ce catalogue de plus de 1 600 références, essentiellement composé de données quantitatives, couvre un large spectre thématique : population et démographie, éducation et formation, travail et emploi, salaires et revenus, conditions de vie et société, santé et protection sociale, politique et société... En parallèle, le CASD propose une infrastructure d'accès sécurisé aux données confidentielles que les producteurs acceptent de mettre à disposition des chercheurs.

Les Fichiers Production Recherche (FPR) et l'accès aux données confidentielles

Comme annoncé dans la lettre PROGEDO Actu n°15, la procédure de diffusion des fichiers de production et de recherche (FPR) a changé en juillet 2019, suite aux changements de législation en France (loi pour une république numérique de 2016) et en Europe (Règlement Général sur la Protection des Données en 2018). Pour rappel, les fichiers FPR sont constitués par les acteurs de la statistique publique depuis 2007, à la demande initiale de chercheurs en sciences humaines et sociale, et donc spécifiquement dédiés à la communauté des chercheurs et étudiants. Ces

fichiers offrent un niveau intermédiaire d'information entre les données standards anonymisées (plus agrégées) qui peuvent être diffusées en open data, et les données confidentielles très détaillées, accessibles sous le contrôle du CASD ; niveau intermédiaire qui permet la réalisation de nombreuses recherches qui n'auraient pas pu être effectuées avec des données standards. L'ensemble du catalogue de l'Ined ainsi que la majorité des données diffusées par l'Adisp sont des fichiers FPR ; inversement, aucune donnée du CDSP n'est estampillée « FPR » (les données étant toutes anonymisées).

Comme auparavant, l'accès aux fichiers FPR nécessite que l'utilisateur s'inscrive et en fasse la demande sur le portail de Quetelet-Progedo-Diffusion. Deux conditions pour obtenir des données : être chercheur, enseignant-chercheur, docteur ou étudiant ; et être membre d'une unité de recherche ou d'enseignement (EPST, universités, écoles, ...). En signant le formulaire associé à sa demande, le demandeur s'engage à respecter les conditions légales d'utilisation des fichiers diffusés (non transmission des données, citation des sources et du diffuseur, information sur les publications produites).

Un accès désormais simplifié pour les chercheurs

Depuis juillet 2019, et conformément à l'avis du 14 décembre 2018 du Comité du secret statistique (CSS), la nouvelle procédure d'accès aux sources FPR impose que tout utilisateur soit « habilité » par le Comité. Concrètement, l'utilisateur doit donc signer et renvoyer un deuxième formulaire : l'engagement de confidentialité CSS. La procédure d'habilitation, intégralement prise en charge par les équipes de Quetelet-Progedo-Diffusion, nécessite la signature du CSS puis des Archives de France, le délai d'obtention des données est donc de 3 à 4 semaines. Néanmoins, une fois obtenue, l'habilitation individuelle reste valable sans limite de temps (sous réserve que l'utilisateur soit toujours rattaché à un établissement de recherche ou d'enseignement) : le délai cité ne concerne donc que la première demande, les suivantes étant généralement traitées en 2 à 3 jours ouvrés (après réception des formulaires).

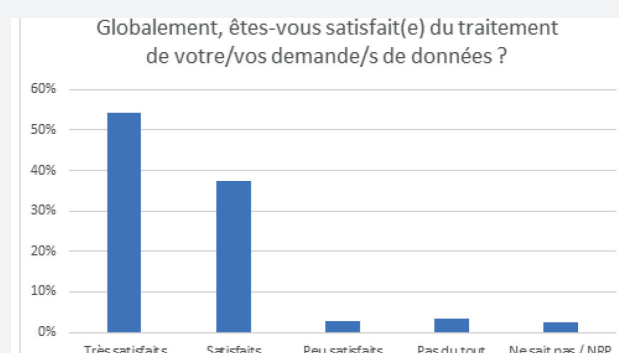
Autres nouveautés de la nouvelle procédure qui simplifient les démarches de l'utilisateur : les formulaires ne requièrent plus que la signature de l'intéressé (et non plus celle de son « responsable ») et peuvent être directement renvoyés en format PDF via le portail de commande (et non par voie postale comme auparavant).

Enquête de satisfaction

L'enquête de satisfaction coordonnée par Alina Danciu (du Centre de Données Socio-Politiques), en collaboration avec les équipes de Quetelet-Progedo-Diffusion montre un taux de satisfaction dépassant 90% de satisfaits (<http://www.progedo.fr/app/uploads/2019/07/EnqueteQueteletProgedoDiffusion2019.pdf>). On espère qu'il augmentera encore avec ces nouvelles modalités d'accès.

SATISFACTION GÉNÉRALE

À la question « Globalement, êtes-vous satisfait(e) du traitement de votre/vos demande/s de données ? », les utilisateurs ont massivement répondu oui : 91% des 362 répondants. On notera d'ailleurs que le taux de satisfaction augmente légèrement si l'on a obtenu des données (97% des 336 personnes ayant obtenu des données).



ACTUALITÉ DES DÉPARTEMENTS

Trust Working Group Cessda

Les 11 et 12 juin 2019, PROGEDO a accueilli à Paris la réunion du Trust Working Group du CESSDA qui vise à permettre aux différentes archives de données européennes de répondre aux exigences du Core Trust Seal (<https://www.core-trustseal.org/>) et obtenir ainsi leur certification. Il s'agit de certifier un ensemble de bonnes pratiques de gestion des données en recherche depuis leur dépôt et leur documentation jusqu'à leur diffusion et sa durabilité, seuls 150 entrepôts de données dans le monde sont certifiés à ce jour.



Réunion du trust Working Group du Cessda à Paris

Le positionnement du portail Quetelet-Progedo-Diffusion est original car il demande une certification commune pour l'ensemble des entrepôts du réseau (ADISP, CDSP, INED). Ce type de configuration est amené à se développer et le CESSDA a donc souhaité nous accompagner, en tant que projet pilote, pour obtenir la certification. Hervé L'Hours (membre de UK Data Archive et du Vice-Chair de Core Trust Seal) avait réuni une dizaine d'autres experts européens à cette occasion.

Création d'un département « données historiques, économiques et financières »

Le rôle des départements de PROGEDO est d'animer les communautés de recherche autour de grands thèmes transversaux pour mieux faire connaître les ressources disponibles. Il en existait 3 jusqu'à présent : « Méthodes et statistiques publiques » autour de l'activité du portail Quetelet-Progedo-Diffusion et du consortium européen CESSDA ; « Opinion et enquêtes sociopolitiques » autour notamment des enquêtes ESS, EVS, ISSP, etc. ; « Santé » qui valorise l'enquête SHARE et construit des ponts hors des SHS.

Les 27 et 28 novembre 2019, profitant du colloque « divorcer autrefois » de la société de démographie historique (<http://www.societededemographiehistorique.fr/>), accueilli au sein de la MSH-LSE à Lyon, Sébastien Oliveau (directeur de PROGEDO) a annoncé le démarrage d'un département « données historiques, économiques et financières » dont l'animation est confiée à Sandra Brée (chargée de recherches au CNRS). On vous en reparle dans la prochaine lettre.



Sandra Brée et Sébastien Oliveau lors de la présentation de PROGEDO au colloque de la société de démographie historique

ACTUALITÉ DES PLATEFORMES UNIVERSITAIRES DE DONNÉES

Tour des données de la recherche - Grenoble

Le 13 septembre 2019, la PUD-GA a organisé à Grenoble une étape du Tour national des données de la recherche en SHS. Chercheurs et étudiants grenoblois ont largement répondu présents pour cette journée présentant l'ensemble du dispositif de PROGEDO (grandes enquêtes, données statistiques nationales et internationales) et insistant sur le rôle des plateformes universitaires de données dans l'accompagnement des communautés locales de recherche.

Journées d'étude sur les enquêtes longitudinales et les comparaisons internationales - Strasbourg

La PUD-S a organisé les 14 et 15 octobre 2019 deux journées d'études sur les dispositifs d'enquêtes longitudinales et internationales. Outre les enquêtes soutenues par PROGEDO telles que SHARE, EVS, GGP ou encore ELIPSS, ce fut l'occasion de présenter l'enquête SRCV (volet français du dispositif européen SILC). Les producteurs des enquêtes et leurs utilisateurs étaient présents pour échanger et montrer à l'assistance composée de collègues et étudiants la richesse de ces sources.



Réunion nationale des PUD

Profitant des journées organisées par la PUD de Strasbourg autour de la comparaison internationale des données longitudinales, PROGEDO a réuni l'ensemble des ingénieurs et référents scientifiques des PUD pour une journée d'échanges, de formation et d'information le 16 octobre 2019. Nous avons profité de la présence en France de Rachad Antonius, professeur de sociologie à l'UQAM (Québec), qui a présenté son travail sur la pédagogie de l'enseignement des statistiques à un public non spécialiste.



Tour des données de la recherche - Aix-en-Provence

Le 12 novembre, la PUD-AMU a organisé à la Maison Méditerranéenne des Sciences de l'Homme d'Aix-en-Provence une étape du Tour national des données de la recherche en SHS. Plus de 100 étudiants, enseignants, chercheurs, et personnels d'appui à la recherche étaient présents pour cette journée de présentation des activités de PROGEDO.

Colloque sur les archives numériques dans le monde arabe

PROGEDO a soutenu l'organisation du colloque organisé par Hala Bayoumi (UMIFRE CEDEJ) à Abou Dhabi du 26 au 28 octobre 2019. Ce fut l'occasion pour son directeur, Sébastien Oliveau, d'y rappeler l'importance des normes et standards dans la création et la gestion des archives. L'occasion pour lui aussi de présenter, avec Clément De Belsunce, ingénieur de la PUD-AMU, les dispositifs d'accompagnement que sont les plateformes universitaires de données en s'appuyant sur l'expérience de l'Université d'Aix-Marseille.

Journée sur la nomenclature des professions et catégories socioprofessionnelles - Nantes

PROGEDO-Loire, la Plateforme Universitaire de Données de Nantes a organisé une journée d'études autour des enjeux de construction, de diffusion et d'utilisation de la nomenclature PCS 2020. Elle a réuni 80 chercheurs et étudiants de toute la France.

Il s'agissait, autour des acteurs de sa rénovation, de présenter le travail initié par le Conseil national de l'information statistique (CNIS) pour actualiser cette nomenclature statistique de classement des métiers qui se trouve au cœur de la production de statistiques et d'enquêtes en France. Cette nomenclature est produite par l'INSEE et les données qui en sont issues sont diffusées par les Archives de Données Issues de la Statistique Publique (ADISP-PROGEDO). La nouvelle nomenclature bénéficiera d'un site internet dédié, co-construit par l'INSEE et PROGEDO, qui permettra de mieux comprendre et utiliser cette nomenclature, grâce à des tutoriels, des illustrations et des outils d'exploration.

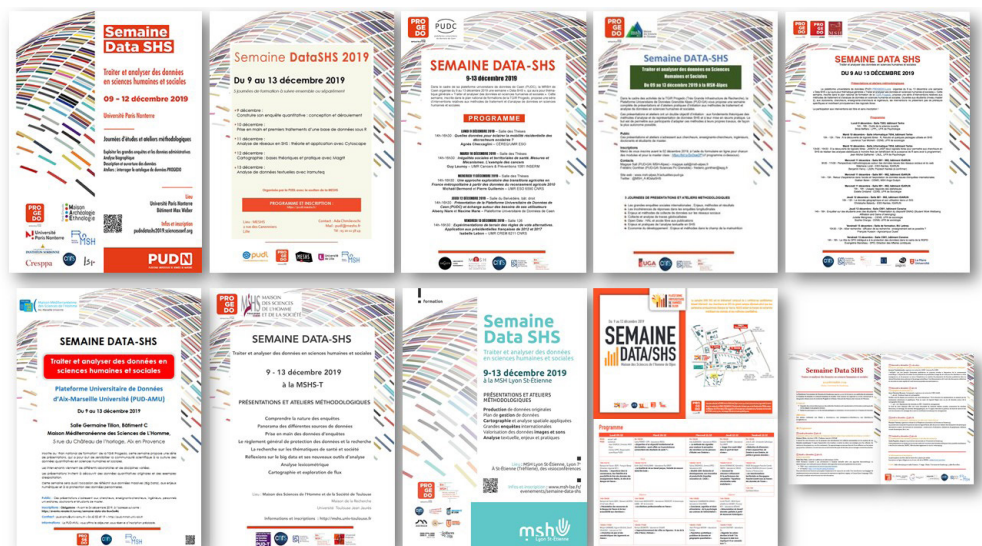


Table ronde sur la diffusion des données et leurs usages lors de la journée consacrée aux PCS à Nantes

Semaine Data-SHS

L'année 2019 s'achève sur la semaine nationale DATA SHS organisée avec le Réseau national des MSH du 09 au 13 décembre. Il s'agit de mobiliser durant une semaine les communautés locales d'utilisateurs pour mettre en valeur les ressources qu'offrent en données quantitatives en sciences humaines et sociales.

10 PUD organisent ainsi dans les MSH des présentations, formations, séminaires sur les données d'enquêtes et les bases de données quantitatives disponibles pour la recherche en SHS.



DONNÉES – GRANDES ENQUÊTES

EVS – Publication

La France des valeurs

Issu de l'enquête EVS (European Value Study) qui collecte depuis 1981 des données sur les valeurs et les croyances des Européens, cet ouvrage présente les résultats pour la France. L'enquête a été conduite en France par une vingtaine de chercheurs de plusieurs laboratoires de sciences sociales, réunis dans l'Association pour la recherche sur les systèmes de valeurs (Arval <http://www.valeurs-france.fr/>). Le pilotage en est assuré par le laboratoire Pacte (Sciences Po Grenoble, CNRS, UGA). L'échantillon principal aléatoire comporte 1 870 personnes, résidant en France, âgées de 18 ans et plus, auxquelles s'ajoute un échantillon complémentaire de 721 jeunes de 18 ans à 29 ans, sélectionnés par quotas.

La France des valeurs a été publié sous la direction de Pierre Bréchon, Frédéric Gonthier et Sandrine Astor aux Presses universitaires de Grenoble.



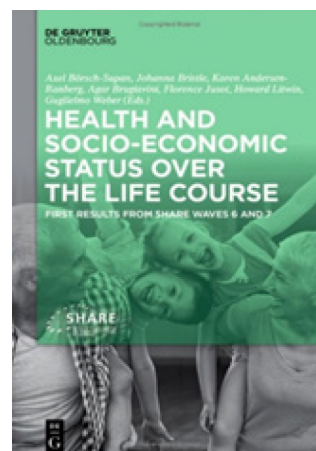
SHARE – Publication

Health and socio-economic status over the life course

Issu de l'enquête longitudinale SHARE (<https://share.dauphine.fr/>), un nouvel ouvrage présente les résultats des vagues 6 et 7. Publié par De Gruyter, il est édité par le collectif de chercheurs en charge de cette grande enquête européenne : Axel Börsch-Supan, Johanna Bristle, Karen Andersen-Ranberg, Agar Brugiavini, Florence Jusot, Howard Litwin et Guglielmo Weber.

L'enquête SHARE concerne plus de 80 000 européens âgés de 50 ans et plus. Elle est réalisée tous les deux ans depuis 2004 et se déroule désormais dans vingt-sept pays européens. Les données recueillies grâce à SHARE portent sur l'état de santé des répondants et de leurs proches ainsi que sur leur situation sociale (famille, entraide, réseaux sociaux) et économique (emploi, retraite, patrimoine).

On peut y accéder en ligne en open-access : <https://www.degruyter.com/viewbooktoc/product/509241>



ÉVÉNEMENTS

Journées Mate-SHS

3 et 4 juin 2019 : Les 6^{es} journées du réseau Mate-shs se sont déroulées à la MSH de Strasbourg (MISHA). Organisées par le groupe local des ingénieur(e)s en sciences sociales de Strasbourg (GLISSS), elles ont portées sur le thème « Être ingénieur(e) en SHS : état des lieux et perspectives » (<http://mate-shs.cnrs.fr/?JA2019-Strasbourg>).

Journée sur la protection des données personnelles

Le 04 juin 2019 a eu lieu à Paris la présentation du guide sur la protection des données personnelles mis en place par Gaëlle Bujan pour les SHS du CNRS. Muriel Roger, chargée du département « méthodes et statistiques publiques » de la TGIR PROGEDO et Fabrice Boudjaaba, Directeur adjoint Scientifique de l'InSHS ont participé à l'écriture de ce guide que l'on peut retrouver en ligne : https://www.inshs.cnrs.fr/sites/institut_inshs/files/pdf/guide-rgpd_2.pdf



6^e Conférence utilisateurs SHARE

L'enquête SHARE a tenu sa sixième conférence utilisateurs conjointement avec le «Budapest Pension Seminar» les 19 et 20 septembre à Budapest, autour du thème «Understanding Ageing Societies».

La conférence visait à réunir des utilisateurs de SHARE provenant de différentes disciplines et pays pour discuter des projets de recherche en cours.

Plus d'informations disponibles ici : <https://conference.tcs.allamkincstar.gov.hu/en/budapest-pension-seminar/invitation-2019.html>



Actions Nationales de Formation de Mate-shs

Du 06 au 11 octobre 2019, le réseau de métier Mate-shs a organisé une Action Nationale de Formation sur l'analyse des données qualitatives d'enquête, soutenue par le CNRS et PROGEDO. Une douzaine d'experts ont été mobilisés pour accompagner les 40 ingénieurs et chercheurs participant à cette formation.

Mate-shs est un réseau métier initié et porté par des ingénieurs qui travaillent à la production, au traitement, à l'analyse et à la représentation des données dans la recherche en Sciences Humaines et Sociales (SHS). Ce réseau est reconnu par l'InSHS (Institut national des

Sciences Humaines et Sociales) du CNRS. Il rassemble à la fois des agents du CNRS, mais aussi des Universités et d'autres EPST.



5^e conférence des utilisateurs de GGP



GGP, en coopération avec l'Institut national d'études démographiques (Ined), a organisé à la Paris School of Economics la 5^e conférence des utilisateurs de GGP du 23 au 25 octobre 2019. Cette

conférence internationale avait pour objectif de réunir les chercheurs travaillant avec les données des enquêtes

GGP et les invitait à présenter et discuter leurs approches méthodologiques et leurs résultats empiriques les plus récents.

Elle a été précédée d'un atelier sur la visualisation des données les 22 et 23 octobre.

Informations : <https://www.ggp-i.org/event/5th-ggp-user-conference-paris-october-23-25-2019/>

11^e conférence des utilisateurs de DDI



EDDI, la conférence des utilisateurs européens de DDI, s'est tenue à l'université de Tampere (Finlande) les 03 et 04 décembre 2019. Data Documentation Initiative (DDI) est un projet international initié en 1995 et destiné à créer et maintenir un standard de documentation technique pour décrire et conserver les informations statistiques et plus globalement les informations et données d'enquêtes en sciences humaines et sociales.

La dynamique de PROGEDO était relayée par le Centre de Données Socio-Politiques qui y présentait plusieurs communications : «challenges and benefits of implementing the CESSDA Metadata Model in an existing environment», par Alina Danciu et Alexandre Mairot; «What do you hear when I say data curation? A study on data curation perceptions and practices in France» par Alina Danciu et Nicolas Sauger ; « The process of creating an in-house interface for converting Blaise output into DDI metadata » par Alexandre Mairot; « Let's take a look at the DDI-documented paradata collected by the information system of a longitudinal panel! » par Geneviève Michaud et Baptiste Rouxel.