

ELIPSS

Étude Longitudinale par Internet Pour les Sciences Sociales

1 L'équipement DIME-SHS

Le projet DIME-SHS (Données Infrastructures et Méthodes d'Enquête en Sciences Humaines et Sociales), coordonné par Laurent Lesnard, directeur du Centre de Données Socio-Politiques (CDSP), porté par Sciences Po, en partenariat avec le GENES, l'INED, l'Université Paris 5, Telecom-ParisTech, EDF R&D et le GIS Quetelet, fait partie des lauréats de l'appel à projets "Investissements d'avenir - Équipement d'excellence" 2010 mis en place par le Ministère de l'Enseignement supérieur et de la recherche.

L'équipement DIME-SHS vise à combler le retard accumulé par les sciences humaines et sociales françaises en matière de méthodologie d'enquêtes afin de répondre aux plus hautes exigences de compétitivité internationale en ce domaine. Cet équipement consiste à doter la France d'une structure intermédiaire novatrice en matière de collecte, d'enrichissement et de diffusion des données utilisables par la recherche en SHS, autant quantitative que qualitative. Son originalité tient d'abord à ce qu'il envisage de manière unifiée ces trois phases des protocoles de la recherche, en tirant parti des innovations qu'offrent l'Internet et les NTIC pour non seulement prendre en charge et capitaliser les données existantes, mais aussi pour créer de nouveaux types de données pour la recherche.

Ainsi, DIME-SHS s'organise autour de trois grands modules :

- la collecte de données quantitatives, assortie de protocoles innovants (panel web mobile et centre d'appel téléphonique)
- la constitution d'une banque de données qualitatives
- la collecte et analyse des expressions spontanées sur le web

2 ELIPSS, le panel web mobile

2.1 Intitulé du projet

Étude Longitudinale par Internet Pour les Sciences Sociales (ELIPSS). ELIPSS est un dispositif d'enquête par internet reposant sur un échantillon représentatif de la population française métropolitaine qui permettra à la communauté scientifique de mener des enquêtes transversales ou longitudinales sélectionnées par un comité scientifique. Le panel sera constitué selon un plan de sondage probabiliste. Les personnes sélectionnées pour faire partie du panel se verront proposer une tablette informatique et un abonnement internet mobile en échange de leur participation régulière à des enquêtes.

2.2 Historique

Le premier module de DIME-SHS, pensé sur le modèle des Survey Research Centers (SRC) existant aux Etats-Unis ou dans certains pays d'Europe, vise à approfondir les partenariats, encore insuffisamment développés, entre les universités et les instituts responsable ou participant aux grandes opérations de la statistique publique (INSEE et INED). Il vise également à affranchir en partie la recherche publique des intermédiaires du privé pour la réalisation d'enquêtes par questionnaire en s'alignant d'abord sur les normes méthodologiques internationales tout en diminuant globalement les coûts investis dans les opérations de collecte. A côté d'un dispositif classique (plateforme téléphonique), la création d'un panel de

répondants sollicité via les NTIC ou encore la réalisation d'expérimentations de collecte multi-mode permettront à la France de se repositionner à l'avant-garde de la recherche quantitative en SHS.

Le panel web mobile ELIPSS qui est présenté au CNIS s'inspire du Longitudinal Internet Studies for the Social sciences (LISS panel) de l'institut de recherche néerlandais CentERdata (situé à Tilburg) et du Knowledge panel de Knowledge Networks aux Etats-Unis. Le LISS panel est un panel internet représentatif de la population néerlandaise constitué selon un plan de sondage probabiliste réalisé par le Centraal Bureau voor de Statistiek (CBS, Institut national de statistique des Pays-Bas). Il repose sur la mise à disposition d'un ordinateur simplifié et d'une connexion à Internet aux répondants qui en sont dépourvus. Le Knowledge panel repose sur les mêmes principes. Cette procédure permet d'éviter les principaux biais des enquêtes par internet classiques qui ne sont menées qu'auprès de personnes volontaires et disposant d'une connexion. S'il reprend le principe d'un panel internet représentatif, ELIPSS se distingue par le choix de l'internet mobile comme mode de collecte principal. Une tablette de type iPad et un abonnement internet mobile seront en effet proposés à tous les panélistes en échange de leur participation au panel.

Les tablettes offrent de nombreux avantages par rapport aux enquêtes par internet sur ordinateur. Plus proches des téléphones portables, les tablettes sont mobiles et mieux intégrées à la vie quotidienne. Parce que leurs interfaces sont beaucoup plus intuitives, elles permettent d'offrir un accès à internet simplifié aux personnes peu familières des nouvelles technologies qui ne disposent pas d'équipements informatiques et de connexion internet à leur domicile. L'accès internet mobile qui leur est associé devrait donner aux panélistes plus de souplesse pour répondre aux enquêtes (moment de la journée, lieu). Ces terminaux mobiles permettent également d'envisager des enquêtes qui tirent profit de leurs caractéristiques : étude des déplacements (géolocalisation des tablettes), étude de l'emploi du temps (carnet d'emploi du temps simplifié ou méthode alternative comme le day reconstruction method), enquêtes budget, etc. Par ailleurs, ce dispositif évite l'installation d'équipements à domicile.

Le panel ELIPSS de l'équipement DIME-SHS est le premier du genre en France et dans sa dimension web mobile, dans le monde. Le présent document expose les principes et le programme de test qui sera mis en œuvre sur la période 2012-2014.

2.3 Service producteur

Le CDSP et le Service des enquêtes et sondages de l'Ined sont les maîtres d'ouvrage et assurent la collecte des données par internet. Le recrutement des panélistes sera assuré par le CDSP et l'Ined¹.

2.4 Les thèmes abordés

L'objectif du panel ELIPSS est de mettre à disposition un outil de collecte de données par questionnaire pour des recherches à buts scientifiques. Les thèmes abordés seront ceux des projets qui seront retenus par le comité de sélection scientifique. Ces enquêtes auront en commun un ensemble d'informations socio-démographiques collectées chaque année et actualisables à chaque interrogation. Avant même la mise en place du panel, il n'est évidemment pas possible de préciser quels seront les principaux thèmes abordés par les enquêtes proposées par les équipes de recherche et sélectionnées par le comité. Aucune thématique n'est a priori écartée. La finalité strictement scientifique constitue le critère impératif d'éligibilité des enquêtes, aucune démarche commerciale n'y sera associée. Des thèmes sensibles seront possibles.

La combinaison des données collectées constituera une originalité dans le système statistique français. Le suivi des panélistes dans le temps assurera d'importantes économies de moyens et la constitution d'une variété de données impossibles à collecter autrement et favorisera la réalisation d'enquêtes longitudinales. Toutefois l'appariement des données individuelles issues d'enquêtes successives sera strictement encadré (voir le point sur la diffusion des données).

¹ Une expérimentation à petite échelle sera conduite en 2012 avec les enquêteurs d'un prestataire de service externe, sélectionné après appel d'offres, pour évaluer l'effet de ce mode de contact et de recrutement sur le taux de réponse (voir le point sur le mode de collecte).

L'appariement des informations du panel avec des données extérieures (fiscales, santé, etc.) est en revanche exclu.

D'une manière générale les thèmes abordés relèveront autant de la statistique publique que de la recherche. Pendant la phase de test de deux ans, il est prévu :

- de répliquer (pour ce qui concerne le questionnement, en partie au moins) un certain nombre d'enquêtes de la statistique publique figurant au programme du CNIS comme l'enquête Famille et logements de 2011, l'enquête sur les technologies de l'information et de la communication TIC 2011, l'enquête sur la transmission intergénérationnelle des désavantages sociaux² ;
- de répliquer une partie de la 6e vague de l'enquête sociale européenne³, de répliquer l'enquête Fecond de l'Ined (sur la santé sexuelle et reproductive), d'administrer une enquête de sociologie électorale⁴ et une enquête sociologique sur l'empreinte carbone des styles de vie.

2.5 Moyens humains et financiers mobilisés pour le projet

Les moyens humains directement mobilisés par le panel web mobile sont :

- 2 équivalents temps plein de chercheur en méthodes d'enquêtes
- 2 équivalents temps plein de gestionnaire de panel
- 2 équivalents temps plein de documentation et création de questionnaires
- 0,5 équivalent temps plein de statisticien
- 2,5 équivalents temps plein d'informaticien entre 2011 et 2013, 1,5 au delà

Les moyens financiers prévisionnels hors salaires et charges afférentes sont de 260 000 euros pour 2011, 750 000 euros pour 2012, 1 410 000 euros pour 2013 et 1 060 000 euros pour 2014⁵.

3 Méthodologie

3.1 Cible de l'enquête

La cible de l'enquête est l'individu, la taille d'échantillon final espéré étant d'environ 6 000 personnes.

3.2 Champ de l'enquête

Le champ de l'enquête est composé de l'ensemble des individus résidant en ménage ordinaire en France métropolitaine, âgés de 15 ans ou plus, lisant suffisamment le français pour répondre à des questionnaires auto-administrés et acceptant de faire partie du panel. L'inclusion des mineurs sera soumise à l'autorisation parentale. La participation de l'individu sera actée par la signature d'un contrat avec DIME-SHS réglant les droits et devoirs du panéliste valant consentement libre et éclairé, ainsi que la signature d'un contrat avec un fournisseur d'accès Internet pour l'équipement en tablette et en abonnement 3G permettant l'accès gratuit à Internet durant la durée du contrat.

3.3 Champ géographique de l'enquête

Le champ géographique couvert par cette demande est la France métropolitaine.

3.4 Base de sondage, échantillonnage

Idéalement, l'échantillon serait tiré dans le fichier d'imposition des personnes (FIP) qui contient des données personnelles (nom, prénom, adresse, date et lieu de naissance) pour chaque redevable et pour les

² Module secondaire de SRCV 2011, du Panel sur les ressources et les conditions de vie des ménages, du dispositif des statistiques sur les ressources et les conditions de vie (SRCV).

³ Principaux thèmes : la confiance dans les institutions, l'engagement politique, les valeurs morales, politiques et sociales, l'exclusion sociale, les identités collectives, le bien-être...

⁴ Comparaison dans le temps des mécanismes de changement électoral, appuyée sur les enquêtes du Cevipof (1978-2007), rolling cross-sectional survey permettant de saisir les dynamiques des enjeux de la campagne, les logiques de l'offre, les effets des nouveaux médias.

⁵ Par ailleurs, des agents du service Facteur Plus de La Poste et des techniciens de l'opérateur de téléphonie mobile seront mobilisés dans le cadre des expérimentations initiales portant sur la mise en place du panel.

personnes à charge. Les avantages du FIP pour le panel sont de deux ordres : il permet de tirer au sort directement des individus et il est adapté au mode de recrutement des panélistes (par courrier principalement). À défaut, l'échantillon serait tiré dans le recensement de la population. L'intérêt de disposer de telles données pour la base de sondage de l'échantillonnage est de fournir quelques informations sur les non répondants : localisation, tranche de revenu, par exemple.

Afin de préserver la représentativité transversale de l'échantillon, le tirage d'un échantillon complémentaire chaque année par l'Insee est envisagé. Un système de pondérations sera calculé afin de permettre des estimations transversales et longitudinales pour les enquêtes qui reposent sur des interrogations multiples.

3.5 Procédure d'inclusion

Le recrutement des panélistes étant une phase déterminante pour la qualité des données collectées par la suite, les tests prévus au cours de la première année seront principalement consacrés à la non-réponse au moment de l'inclusion. Le recrutement des 1500 premiers panélistes en 2012 sera l'occasion de tester différents modes de contact et d'évaluer leurs effets sur le taux d'inclusion :

- recrutement par courrier uniquement,
- recrutement par courrier complété par un recrutement téléphonique pour les individus dont le numéro est disponible dans l'annuaire,
- recrutement par un agent du service Facteur plus de la Poste,
- recrutement par courrier complété par un recrutement téléphonique pour les individus dont le numéro est disponible dans l'annuaire et par un recrutement par un agent du service Facteur plus de la Poste pour ceux dont le numéro n'a pu être trouvé,
- recrutement en face-à-face par un enquêteur professionnel.

A l'issue de ces tests, un mode de contact sera privilégié en fonction de son taux d'inclusion, de la représentativité de l'échantillon ainsi obtenu, mais aussi de son coût. En raison des contraintes budgétaires, il est d'ores et déjà exclu de recourir à un institut de sondage et à son réseau d'enquêteurs professionnels pour recruter l'ensemble des 6000 panélistes. Le test de ce mode de contact a seulement pour objet de mesurer la perte éventuelle induite par le non-recours à un enquêteur.

Pour limiter les refus liés à la méconnaissance des nouvelles technologies, une expérimentation évaluera l'utilité du recours à des agents du service Facteur Plus de La Poste ou des techniciens de l'opérateur de téléphonie mobile partenaire du projet pour mettre en route la tablette 3G après acceptation du contrat.

3.6 Modes de collecte, date et périodicité des enquêtes

Une fois inclus, les panélistes auront la possibilité de répondre aux enquêtes par internet depuis leur tablette ou depuis un ordinateur (avec possibilité de passer de l'Internet mobile à l'Internet fixe entre deux enquêtes et au sein d'une même enquête si l'individu dispose d'un accès personnel)⁶.

A priori, dix interrogations par an sont prévues, au rythme d'une par mois sauf en juillet-août, d'une durée maximale de 30 à 45 minutes par mois⁷. Il convient de raisonner en temps d'enquête disponible pour la recherche plutôt qu'en nombre d'enquêtes. En l'occurrence, 30 minutes mensuelles peuvent correspondre à une seule enquête, comme à deux (par exemple l'une de 20 minutes, l'autre de 10), comme à la moitié d'une longue enquête d'une heure qui serait alors posée sur deux mois. En dehors d'un tronc commun qui sera répété annuellement, chaque enquête sera différente ou dans le cas d'enquêtes longitudinales, pourra être répétée partiellement. Les deux premières années, les enquêtes seront menées exclusivement par les équipes de recherche du consortium DIME-SHS à des fins d'expérimentation. Durant cette période,

⁶ Des enquêtes auto-administrées par voie postale (et retour par enveloppe T) sont également envisagées pour les individus qui préféreraient ce mode de passation et des enquêtes simples ne comportant que peu de filtres.

⁷ Il sera néanmoins possible de constituer des sous-panels ou tout au moins de sélectionner certaines catégories de personnes pour certaines enquêtes (selon l'âge, le sexe etc.).

méthodologie, charge pour l'enquête, utilisation des données etc. seront notamment testées et les processus envisagés révisés, le cas échéant.

D'autres tests, en plus de l'évaluation des modes de contact et de recrutement, sont envisagés pendant la période 2012-2014 :

- L'interface graphique, c'est-à-dire la mise en forme des questionnaires, la présentation des modalités de réponse, le recours à des images ou des vidéos, feront l'objet d'une attention particulière.
- La durée maximale des interrogations, fixée entre 30 et 45 minutes par mois, sera évaluée en fonction des taux de non-réponse et d'abandon.
- En attendant l'ouverture de la plateforme téléphonique de DIME-SHS (prévue en 2014), l'effet mode de collecte par internet pourra être testé en répliquant certaines enquêtes réalisées en même temps en face-à-face ou par téléphone (par exemple, celles menées en population générale par l'Insee et l'Ined, ou la prochaine vague de l'enquête sociale européenne).
- La professionnalisation éventuelle des panélistes pourra être évaluée au moment de l'entrée des 4500 nouveaux panélistes dont les réponses pourront être comparées à celles des 1500 premiers panélistes qui auront déjà l'expérience de plusieurs enquêtes ainsi que par la comparaison des résultats des enquêtes du panel avec d'autres enquêtes comparables menées en population générale par d'autres moyens.
- Les moyens à mettre en œuvre pour limiter l'attrition (nombre de relances, mode des relances, connaissance des raisons de la non-participation, incitation spécifique pour remobiliser les non-répondants comme offrir une application pour tablette...) devront eux aussi être évalués.

Outre ces vérifications générales et incontournables, des aspects plus spécifiques pourront faire l'objet de tests, comme par exemple l'effet sur les réponses aux questions sensibles d'une sécurisation informatique accrue (avec saisie d'un identifiant et d'un mot de passe personnels pour une enquête donnée).

Une partie du panel (aléatoire et renouvelée pour chaque enquête) pourra être réservée à la conduite de tests techniques pour l'implémentation des enquêtes à venir.

4 Engagements respectifs

4.1 Contraintes imposées à l'enquête

Pour les panélistes, la charge de dix interrogations par an d'une durée de 30 à 45 minutes maximum dont le contenu n'est pas connu à l'avance et dont les thèmes sont très divers, est importante, sans compter que l'accumulation de données personnelles au fil du temps peut susciter des craintes. Néanmoins, l'expérience des cohortes similaires à l'étranger (Liss Panel et Knowledge Panel) montre que la constitution d'un panel de ce type n'est pas impossible.

L'inscription au panel implique le consentement du panéliste à l'obligation de participer à un nombre minimal d'interrogations (7 ou 8) au cours de l'année⁸. Le non-respect de cet engagement aura pour conséquence l'exclusion du panéliste qui prendra effet dès la réception de la lettre envoyée par le consortium. L'abonnement 3G du panéliste sera alors résilié. La tablette tactile sera en revanche conservée par l'ex-panéliste⁹.

⁸ Dans toute enquête, le panéliste est libre de ne pas répondre à un certain nombre de questions, à sa convenance. Toutefois, un questionnaire dont plus de 30% des questions sont indûment restées sans réponses (après analyse des cohérences internes au sein du questionnaire) sera considéré comme non valide et il sera convenu que le panéliste n'a pas participé à l'interrogation.

⁹ Les exclusions devront être motivées et proposées par les gestionnaires du panel. Les responsables du panel décideront de leur effectivité. Les décisions et motivations seront enregistrées par les gestionnaires.

À chaque enquête pour laquelle il est sollicité, le panéliste est réputé accepter de participer (acceptation tacite)¹⁰. En cas d'acceptation tacite, le panéliste dispose d'un mois pour répondre à compter de la sollicitation. Au terme de cette période, les panélistes ne pourront plus répondre et seront considérés comme non participants à cette enquête.

Afin d'alléger la charge des répondants, on permettra, comme indiqué précédemment, que les questionnaires soient complétés en plusieurs fois et sur des supports différents. Un tronc commun comportant les principales caractéristiques socio-démographiques sera administré annuellement pour éviter la répétition de ces questions à chaque enquête. A chaque interrogation, les panélistes auront accès à leur profil socio-démographique pour faire état des changements éventuels de leur situation personnelle.

4.2 Déontologie et obligation du consortium

Le contrat passé avec les panélistes à leur inclusion sera simple, clair et transparent. Les enquêtes seront à but strictement scientifique : aucune démarche commerciale ne devra y être associée. Le consentement des panélistes est révocable à tout moment.

L'essentiel de l'information sur le panel, les recherches en cours et les résultats des différentes enquêtes (après embargo pour les délais de publication et sous condition d'accord des chercheurs concernés) sera visible sur le site internet de DIME-SHS. La bibliographie sera régulièrement mise à jour. Sur le site du panel ELIPSS, des informations seront en permanence présentées sur le panel : données sur la participation, sur l'attrition, etc. L'historique des enquêtes administrées et un programme prévisionnel des enquêtes à venir seront également présents.

Les responsables de DIME-SHS s'engagent à être totalement transparents sur les commanditaires des enquêtes et les objectifs généraux de celles-ci lors des sollicitations adressées aux panélistes. Les procédures techniques de sécurisation des données et de protection de l'anonymat et de la confidentialité seront décrites et accessibles sur le site internet de DIME-SHS. De même, les responsables s'engagent à donner de l'information complémentaire sur les enquêtes aux panélistes qui le demandent. Une section du site Internet sera consacrée à l'aide technique pour l'usage des tablettes 3G. Une section du site sera consacrée aux démarches administratives et au rappel du contrat.

Au centre des instances de concertation de l'équipement DIME-SHS se trouve le comité de pilotage qui réunit les représentants des institutions membres du consortium. Le comité de pilotage supervise l'ensemble des activités de DIME-SHS et ses décisions sont informées par trois comités : un comité scientifique qui veillera aux orientations et à la qualité scientifiques du projet ; un comité d'éthique qui s'assurera du respect des plus hautes exigences en matière d'éthique et d'anonymat ; un comité d'utilisateurs qui sera consulté sur l'intérêt des enquêtes, la facilité de leur accès, l'impact des résultats obtenus, etc. Les données du panel utilisées pour la recherche seront encadrées par le comité d'accès aux données dont les principales règles de décision seront rendues publiques. Les demandes d'accès aux données seront publiques.

5 Diffusion des données et des résultats

5.1 Mise à disposition des données et protection de la personne

La sécurité de l'information collectée dans DIME-SHS est primordiale. Nous distinguons le stockage et la fourniture des données aux chercheurs d'une part et la collecte des données d'autre part. Cette distinction conduira à la création de deux systèmes d'information différents. Le premier devra fonctionner pendant toute la durée de DIME-SHS. Le second est propre à chaque opération de collecte.

¹⁰ Il peut toutefois manifester son refus de participer afin de n'être pas relancé. Un court questionnaire de refus lui sera alors proposé, le cas échéant.

Une déclaration à la CNIL du dispositif est en cours. Selon le Correspondant Informatique et Libertés du CNRS, ce régime est adapté à des enquêtes qui ne comportent pas de questions sensibles. Dans une telle éventualité, une demande d'autorisation serait déposée.

Pour la gestion des risques et plus généralement pour la gestion de la sécurité de l'information, nous avons choisi d'utiliser la famille de norme ISO 2700x relative aux systèmes de management de la sécurité de l'information avec un objectif de certification ISO 27001 du système d'information de stockage et de fourniture des données. Quel que soit le système d'information concerné, les mesures techniques et organisationnelles élémentaires sont employées, à savoir : chiffrement et signature électronique des échanges, sécurisation des développements applicatifs, utilisation de pare-feu pour la sécurisation des réseaux, sauvegardes, mise à jour régulière des systèmes, redondance des équipements, sécurisation des locaux (incendie, inondation, effraction) et surtout prévention, formation et sensibilisation des personnels.

5.2 Exploitation et diffusion des résultats

Une fois anonymisées, les données seront mises à la disposition de la communauté internationale au fur et à mesure de leur collecte et dans les délais les plus brefs possibles (après un délai d'exclusivité pour les chercheurs ayant conçu l'enquête) par le Centre de données socio-politiques. Le CDSP est l'un des trois centres d'archivage et de diffusion de données qui constituent le Réseau Quetelet. Les enquêtes seront documentées selon le standard Data Documentation Initiative (DDI 3) et diffusées par internet. Elles apparaîtront dans le catalogue d'enquêtes du portail du Réseau Quetelet. Outre les équipes de recherche à l'origine des enquêtes qui seront menées avec le panel, c'est donc l'ensemble des utilisateurs du Réseau Quetelet qui auront accès aux données.

Il sera impossible d'apparier toutes les données individuelles issues du panel entre elles. Seules les données de signalétiques non indirectement nominatives seront communes aux enquêtes et accessibles aux chercheurs sans restriction. Toute demande d'appariement de données (en tout ou partie) de deux enquêtes en plus de la signalétique sera soumise à examen.

De plus, les gestionnaires et responsables du panel, ainsi que les chercheurs qui demanderont des données se verront dans l'obligation de signer un engagement strict de confidentialité qui les rendra potentiellement responsables personnellement des utilisations illégales ou bien contraires aux règles de l'art en matière de secret statistique et de la protection des libertés individuelles.

L'équipe du panel ELIPSS s'engage aussi à diffuser le plus largement possible les résultats des enquêtes sur le site internet du panel web.

Les analyses statistiques sur les enquêtes seront à la charge des équipes de recherche.