

**BILAN DU PILOTE DE L'ÉTUDE LONGITUDINALE PAR INTERNET POUR LES SCIENCES SOCIALES  
(ELIPSS) : 2012-2015**

Réalisé pour la Commission Démographie et questions sociales du CNIS

Anne Cornilleau, Anne - Sophie Cousteaux, Stéphane Legleye,  
Nirintsoa Razakamanana et l'équipe ELIPSS



Septembre 2015

## Table des matières

Introduction.....	4
1 Un panel internet pour la recherche.....	5
1.1 Les enquêtes par internet en population générale.....	5
1.2 Les expériences étrangères.....	5
1.2.1 Le Longitudinal Internet Studies for the Social Sciences (LISS Panel) de l'institut de recherche néerlandais CentERdata (Université de Tilburg).....	5
1.2.2 Le KnowledgePanel de Knowledge Networks aux Etats-Unis.....	6
1.2.3 Le German Internet Panel (GIP) de l'Université de Mannheim.....	6
1.2.4 Des panels mixtes.....	6
1.2.5 Trois panels internet en Europe du Nord.....	6
1.3 Le dispositif ELIPSS.....	6
1.3.1 Le contexte français.....	6
1.3.2 La collecte par internet mobile.....	7
1.3.3 Le calendrier et l'équipe.....	7
1.3.4 Sécurité et confidentialité des données.....	8
2 Le recrutement du pilote.....	9
2.1 La base de sondage.....	10
2.1.1 L'échantillon fourni par l'INSEE.....	10
2.1.2 La préparation de la base de sondage.....	10
2.2 Le déroulement du terrain.....	11
2.2.1 Les campagnes d'envoi des lettres et de rappel téléphonique.....	11
2.2.2 Le recrutement par les enquêteurs de TNS Sofres.....	11
2.2.3 L'expérimentation des bons cadeaux.....	13
2.3 L'entrée des panélistes.....	13
2.3.1 La convention.....	13
2.3.2 La prise en main de la tablette.....	14
2.3.3 La première enquête.....	14
2.4 Le bilan du recrutement.....	15
2.4.1 Un taux de recrutement de 27%.....	15
2.4.2 L'effet des bons cadeaux.....	16
2.4.3 La représentativité socio-démographique initiale.....	17
2.4.4 Les individus qui n'avaient pas d'accès internet avant leur entrée dans le panel.....	18
3 Les enquêtes : des projets de recherche à la diffusion des données.....	19
3.1 La sélection des enquêtes.....	19
3.1.1 Le comité scientifique et technique DIME quanti.....	19

3.1.2	L'évaluation des projets .....	19
3.1.3	Les projets soumis .....	20
3.2	La production des enquêtes .....	21
3.2.1	L'élaboration des questionnaires .....	21
3.2.2	La tablette comme mode de collecte.....	22
3.2.3	Les enquêtes administrées au panel .....	23
4	La qualité statistique du panel .....	24
4.1	Redressement.....	25
4.1.1	Données et méthode utilisées.....	25
4.1.2	Résultats .....	25
4.2	Attrition et représentativité au fil des enquêtes.....	27
4.2.1	La participation aux enquêtes .....	27
4.2.2	L'attrition .....	28
4.2.3	Les profils de participation .....	29
4.2.4	Evolution de la représentativité au fil des enquêtes.....	31
4.2.5	Influence des caractéristiques individuelles et des attributs d'enquête sur la participation .....	33
4.3	Analyse des premières impressions des participants au panel.....	35
4.4	Qualité des réponses aux enquêtes annuelles.....	39
5	Les premiers résultats du pilote .....	40
5.1	La diffusion des données .....	40
5.2	La valorisation du dispositif.....	41
5.2.1	Publications .....	41
5.2.2	Communications.....	42
5.3	Les collaborations internationales .....	43
	Conclusion .....	43
	Bibliographie .....	45

## Introduction

Le Centre de données socio-politiques (CDSP) de Sciences Po est une unité mixte de service du CNRS et fait partie du réseau Quêtelet (réseau français des archives de données pour la recherche en SHS). Dirigé par Laurent Lesnard (directeur de recherche CNRS), le CDSP a pour missions principales le soutien à la production, la documentation et la diffusion de données en sciences sociales auprès de la communauté académique française ou internationale.

Depuis mars 2011, le CDSP est fortement impliqué dans la mise en place de l'équipement d'excellence DIME-SHS, lauréat de l'appel à projets 2010 des "investissements d'avenir" (ANR-10-EQPX-19-01). Cet équipement, (coordonné par Laurent Lesnard et porté par Sciences Po, en partenariat avec le GENES, l'INED, l'Université Paris 5, Telecom-ParisTech, EDF R&D et le GIS Quetelet) vise à doter la France d'une nouvelle structure en matière de collecte, d'enrichissement et de diffusion des données quantitatives, qualitatives et web. En particulier, il s'agit de mettre en place le panel internet ELIPSS (Étude longitudinale par internet pour les sciences sociales) constitué à partir d'un échantillon aléatoire de la population résidant en France. Le CDSP assure également la réalisation de l'enquête sociale européenne (ESS), dispositif d'enquête comparative et répétée sur les attitudes des Européens réalisée tous les deux ans auprès d'échantillons nationaux représentatifs de près de trente pays.

Le panel ELIPSS (Étude Longitudinale par Internet Pour les Sciences Sociales) est un dispositif d'enquêtes par internet destiné à la communauté scientifique. Il vise à combler l'absence de moyens d'enquête par questionnaire propres aux chercheurs français en sciences humaines et sociales. Par son échantillon probabiliste et sa finalité scientifique, le panel internet de l'équipement d'excellence DIME-SHS est en effet le premier du genre en France et par sa dimension web mobile, en Europe.

Il s'agit d'un panel internet représentatif de la population âgée de 18 à 75 ans résidant en France métropolitaine. Une tablette tactile et un abonnement internet mobile sont fournis aux panélistes sélectionnés aléatoirement à partir du recensement afin qu'ils participent aux enquêtes mensuelles. Ces dernières sont élaborées par des chercheurs et sélectionnées par un comité scientifique et technique.

La première partie est consacrée à la présentation générale du dispositif, notamment des contextes français et étrangers en matière d'enquêtes par internet en population générale. La deuxième partie présente le recrutement du pilote du panel ELIPSS, depuis la base de sondage fournie par l'INSEE jusqu'à la signature d'une convention de participation par les panélistes et la première enquête lancée en décembre 2012, ainsi que le bilan du recrutement en termes de taux de recrutement et de représentativité initiale du panel. Les différents aspects opérationnels du dispositif sont ensuite exposés dans la troisième partie, allant de la sélection des enquêtes à la diffusion des données en passant par la production des questionnaires et la participation mensuelle des panélistes. La quatrième partie vise à évaluer la qualité statistique du panel en détaillant la méthode de redressement, l'évolution de l'attrition, la représentativité au fil des enquêtes et la qualité des réponses aux enquêtes annuelles de 2013, 2014 et 2015. La cinquième partie présente les premiers résultats du pilote en matière de diffusion des données à des fins de recherche, de valorisation du dispositif (publications et communications) et de collaborations internationales avec les panels européens de même type. Enfin, la conclusion souligne les enjeux du développement du panel en 2016.

# 1 Un panel internet pour la recherche

## 1.1 Les enquêtes par internet en population générale

Si les enquêtes par internet présentent des avantages certains, leur utilisation en population générale revêt plusieurs difficultés. D'une part, on retrouve les avantages des enquêtes par questionnaires auto-administrés. Les coûts de collecte sont réduits, essentiellement par l'absence d'enquêteurs ; les enquêtés peuvent répondre au moment qui leur convient le mieux ; l'absence d'enquêteurs permet également d'aborder des questions plus personnelles (santé, sexualité...). D'autre part, le mode de passation par internet présente des avantages spécifiques. Il permet d'envisager de nouveaux modes de questionnement qui intègrent des vidéos, du son ou des applications interactives. Par ailleurs, la période de collecte peut être réduite puisqu'il n'y a pas (ou presque) de limite au nombre de personnes interrogées simultanément. De plus, les réponses sont sauvegardées au fur et à mesure de leur collecte.

Pour autant, l'utilisation d'internet pour enquêter en population générale se heurte à plusieurs difficultés qui mettent en cause la représentativité de l'échantillon et l'extrapolation des résultats. Premièrement, les enquêtes par internet sont réalisées à partir d'échantillons de personnes volontaires, c'est-à-dire d'échantillons non probabilistes. Deuxièmement, les personnes n'ayant pas accès à internet sont de fait exclues. Or en France en 2012, une personne sur cinq n'avait pas d'accès internet à son domicile (Gombault, 2013).

Afin de dépasser ces biais, une possibilité est de construire un panel internet à partir d'un échantillon aléatoire de la population. Ceci suppose de recourir à un mode de recrutement hors-ligne, d'inclure les personnes qui n'ont pas accès à internet et de les équiper d'une connexion le cas échéant (Das, Ester and Kaczmirek, 2011). Dans le cas du panel ELIPSS, ceci est réalisé par le recours à un échantillon aléatoire d'adresses tiré par l'INSEE et par la mise à disposition de tablettes tactiles connectées en 3G à tous les panélistes.

## 1.2 Les expériences étrangères

Deux expériences étrangères ont inspiré le projet de panel ELIPSS (le LISS panel et le KnowledgePanel) et deux initiatives allemandes (le GIP et le GESIS panel) se sont développées en même temps que le pilote d'ELIPSS. Des discussions sont également en cours dans plusieurs pays pour mettre en place des projets similaires, notamment au Royaume-Uni<sup>1</sup> ou encore en Europe du Sud.

### 1.2.1 Le Longitudinal Internet Studies for the Social Sciences (LISS Panel) de l'institut de recherche néerlandais CentERdata (Université de Tilburg)

Ce panel internet représentatif des ménages néerlandais a été constitué en 2007 selon un plan de sondage probabiliste réalisé en collaboration avec 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 ménages qui en sont dépourvus. Le LISS Panel est constitué de 5 000 ménages, soit 8 000 individus de 16 ans et plus. Ce dispositif est gratuit et exclusivement dédié à des opérations de recherche.

---

<sup>1</sup> <http://www.natcenweb.co.uk/genpopweb/>

## **1.2.2 Le KnowledgePanel de Knowledge Networks aux Etats-Unis**

Créé en 1999 par deux universitaires américains, ce dispositif repose sur un échantillonnage aléatoire et sur l'installation d'une connexion à internet chez les répondants qui n'en disposent pas à leur domicile. L'échantillon, tiré aléatoirement à partir d'une base d'adresses de logements, est composé de 50 000 personnes de 18 ans et plus. Contrairement au Liss Panel, ce dispositif est également ouvert aux études commerciales.

## **1.2.3 Le German Internet Panel (GIP) de l'Université de Mannheim**

Ce panel représentatif de la population âgée de 16 à 75 ans résidant en Allemagne repose sur un échantillonnage aléatoire de 1 500 personnes. Les membres du panel sont interrogés tous les 2 mois. À l'instar du LISS Panel, un ordinateur simplifié et une connexion internet sont mis à disposition des personnes qui en sont dépourvues. Ce dispositif est réservé aux chercheurs de l'Université de Mannheim et les questionnaires sont centrés sur les réformes politiques.

## **1.2.4 Des panels mixtes**

Contrairement aux panels précédents qui utilisent uniquement internet comme mode de collecte, le GESIS panel du GESIS Leibniz Institute for the Social Sciences de Mannheim se caractérise par le recours à deux modes d'interrogation auto-administrés. Les personnes disposant d'un accès à internet depuis leur domicile peuvent répondre sur le web, tandis que les personnes ne disposant pas d'un accès à internet ou ne souhaitant pas répondre sur le web peuvent répondre sur papier aux questionnaires envoyés par voie postale. L'échantillon tiré aléatoirement à partir des registres municipaux est constitué de 4 000 individus âgés de 18 à 70 ans et résidant en Allemagne. Les enquêtes bimestrielles sont proposées par des équipes de recherche en sciences sociales dans le cadre d'appels à projets et ne peuvent aucunement servir un intérêt commercial.

L'utilisation de deux modes de collecte, internet et téléphone, est aussi la solution adoptée par le panel Amerispeak de 10 000 ménages qui vient d'être créé par le centre de recherche NORC de l'université de Chicago.

## **1.2.5 Trois panels internet en Europe du Nord**

Dans les pays où le taux de pénétration d'Internet est très élevé, supérieur à 95 % de la population, le biais de couverture est réduit. C'est pourquoi le Social Science Research Institute Online Panel créé en Islande en 2010, une partie du Citizen Panel créé en Suède en 2011 et le Norwegian Citizen Panel créé en Norvège en 2013 ont fait le choix de recruter les panélistes à partir d'un échantillon aléatoire mais de ne pas équiper et donc de ne pas inclure les personnes ne disposant pas d'un accès à internet à domicile.

# **1.3 Le dispositif ELIPSS**

## **1.3.1 Le contexte français**

En France, il n'existe pas de service de production d'enquêtes par questionnaire à vocation nationale dédié à la recherche en sciences humaines et sociales.

Pour pallier cette absence, les chercheurs français ont deux solutions. Ils peuvent produire leurs enquêtes par le recours à un institut de sondage, mais cette solution est onéreuse, d'autant plus lorsqu'il s'agit de réaliser des enquêtes à partir d'un échantillonnage aléatoire. Ils peuvent également

réutiliser des enquêtes statistiquement représentatives, par exemple les enquêtes de la statistique publique, mais celles-ci n'abordent pas l'ensemble des thèmes et des questions susceptibles d'intéresser les chercheurs.

Par ailleurs, on constate une diminution des taux de participation aux enquêtes, liée essentiellement aux refus de répondre et aux difficultés croissantes pour joindre les personnes à interroger. Or ceci fragilise la qualité statistique des enquêtes.

Ainsi, la mise en place du panel ELIPSS répond à deux objectifs principaux :

- permettre aux chercheurs de mener des enquêtes sur des thèmes qui ne sont pas traités par la statistique publique française,
- affranchir la recherche publique des intermédiaires du privé pour la réalisation d'enquêtes par questionnaire à partir d'un échantillonnage aléatoire tout en diminuant les coûts et le temps de collecte (par la passation auto-administrée sur internet).

### **1.3.2 La collecte par internet mobile**

Le panel ELIPSS se distingue des dispositifs similaires à l'étranger par le choix de l'internet mobile comme mode de collecte principal. Lors du pilote, une tablette tactile 7 pouces et un abonnement illimité 3G sont fournis à l'ensemble des panélistes en échange de leur participation. Ainsi, ils peuvent répondre aux questionnaires même s'ils ne disposent pas de connexion internet.

Le choix de cette nouvelle technologie s'explique par des raisons à la fois méthodologiques et pratiques. Il s'agit de tirer parti des possibilités offertes par internet (images, vidéos...) et par la mobilité pour renouveler certaines techniques d'enquête (étude des déplacements, carnets d'enquêtes budget-temps, etc.). Les tablettes offrent aussi de nombreux avantages par rapport aux enquêtes sur ordinateur. Parce que leurs interfaces sont plus intuitives, elles offrent un accès internet simplifié aux personnes peu familières des nouvelles technologies. L'accès web mobile donne également aux panélistes plus de souplesse pour répondre aux enquêtes (choix du moment et du lieu). De plus, l'outil de collecte est le même pour tous les panélistes, via une application spécifique installée sur la tablette, ce qui minimise les biais de mesure.

Outre les avantages de la tablette comme mode de collecte, ce choix a également été guidé par l'effet incitatif attendu par la mise à disposition d'un tel équipement. Le taux de pénétration était faible en France au moment du recrutement du pilote (9% des ménages étaient équipés en 2012<sup>2</sup>). Il faut toutefois noter la très forte progression en seulement deux ans puisque un ménage sur trois était équipé en 2014.

### **1.3.3 Le calendrier et l'équipe**

L'étude pilote a débuté en 2012 afin de définir la procédure de recrutement, d'affiner la méthodologie, de mettre au point les procédures de gestion de panel et de production d'enquêtes et de développer les outils informatiques. À partir de 2016, le panel ELIPSS sera constitué de 3500 individus.

La construction de ce dispositif repose sur une équipe aux compétences variées qui s'est étoffée au fur et mesure du pilote. Cette équipe est localisée dans deux institutions. Le Centre de données socio-politiques de Sciences Po assure la coordination du projet, l'ensemble de la production et la diffusion des enquêtes, les développements et l'infrastructure informatiques. Pour gagner en

---

<sup>2</sup> Enquête GfK / Médiamétrie – Référence des Equipements Multimédia

efficacité, la gestion de panel initialement assurée par l'Institut national d'études démographiques (INED) a rejoint l'équipe du CDSP en janvier 2015. Le suivi de la qualité statistique du panel reste pris en charge par le service des enquêtes de l'INED.

**Tableau 1 : l'équipe ELIPSS et son évolution depuis 2012**

	2012				2013				2014				2015			
	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Coordination	Anne Cornilleau															
	Anne-Sophie Cousteaux															
	Stéphane Legleye															
Gestion de panel	Carmen Calandra															
	Gabrielle Bouchet															
	Elodie Pétorin (40%)															
	Marc Sigaud															
	Kevin Boudelle															
	Patricia Sossa															
Production et documentation des enquêtes	Emmanuelle Duwez															
	Matthieu Olivier															
	Alexandre Mairot															
	Alina Danciu															
	Djery Sow															
Statistiques	Nirintsoa Razakamanana															
	Stéphane Legleye															
Développement et déploiement d'applications	Geneviève Michaud*															
	Jérémy Richard*															
	Adrien Ferreira															
	Daniele Guido*															
	Alexandre Aazzouz*															
	Alexandre Chevallier*															
	Anthony Jacques*															
Raphaël Paoloni*																

\*: personnes travaillant sur les trois volets de DIME-SHS

#### 1.3.4 Sécurité et confidentialité des données

La déclaration du dispositif ELIPSS a été inscrite au registre CIL du CNRS sous le numéro 2-12030 avec pour finalité du traitement la réalisation d'enquêtes à des fins de recherche. Cette déclaration décrit précisément le système d'information mis en œuvre, les moyens utilisés pour assurer la sécurité et la confidentialité des données, la procédure de gestion des enquêtes (assurée par le CDSP) et la gestion des contacts avec les panélistes (assurée à l'INED jusqu'en janvier 2015).

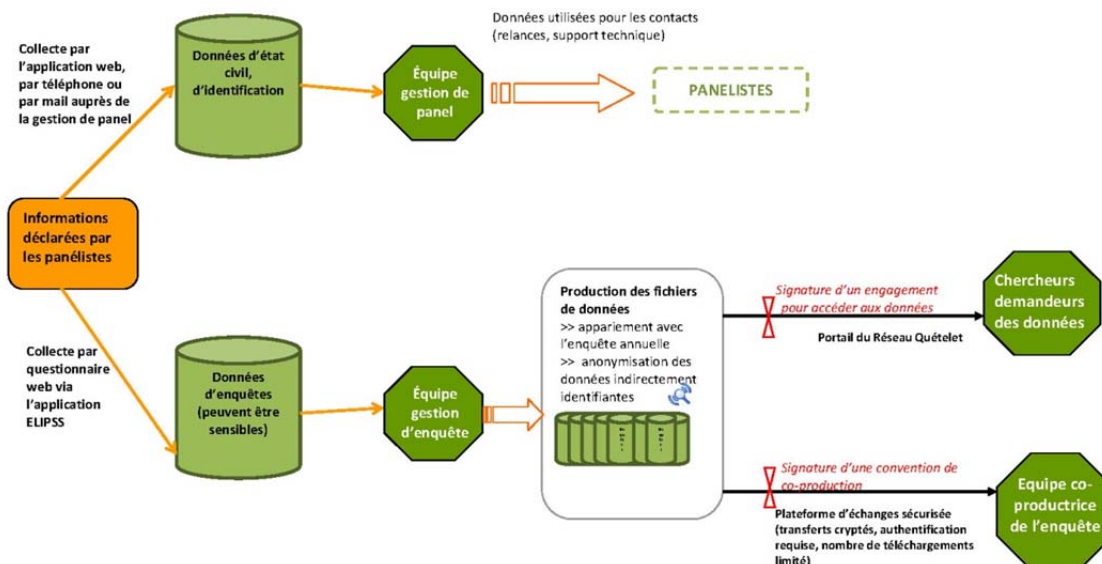


La sécurité de l'information collectée à partir du panel ELIPSS est primordiale. Les données nominatives et les données d'enquêtes sont stockées dans deux systèmes d'information différents. D'un point de vue technique, le cryptage des données et les restrictions mises en place pour contrôler l'appariement des données garantissent également la confidentialité.

Par ailleurs, les données sont diffusées d'une part aux équipes de recherche productrices de l'enquête et d'autre part à l'ensemble de la communauté après une période d'exclusivité de douze mois maximum. Dans les deux cas, les fichiers sont préalablement anonymisés<sup>3</sup> de façon à rendre l'identification directe et indirecte impossible. Les fichiers de données sont diffusés aux équipes de recherche productrices par une plateforme d'accès à distance sécurisée<sup>4</sup> et ceux destinés à la communauté scientifique sont diffusés par le portail du Réseau Quetelet.

La Figure 1 résume le système d'information mis en œuvre dans le dispositif ELIPSS.

**Figure 1: Système d'information du dispositif ELIPSS**



Le traitement pour les données nominatives du panel a été enregistré au registre CIL de l'INED sous le numéro 2012-CIL-0012.

## 2 Le recrutement du pilote

Le recrutement du pilote s'est déroulé en deux temps. L'exploitation de l'échantillon principal (3500 adresses) a eu lieu de juin 2012 à février 2013. Trois modes de contact ont été mis en œuvre. Une invitation à participer a été envoyée par courrier postal. Ce premier contact a fait l'objet d'une expérimentation : la moitié des courriers a été envoyée avec un bon cadeau de 10 euros et l'autre moitié sans bon cadeau. Des relances postales et téléphoniques ont ensuite été réalisées. À l'issue d'un appel d'offres, l'institut de sondage TNS Sofres a été retenu pour exploiter en face-à-face les

<sup>3</sup> Cette anonymisation est assurée en appliquant les méthodes habituellement mises en œuvre pour respecter le secret statistique (variables identifiantes retirées, recodages de modalités à faible effectif, etc.)

<sup>4</sup> La sécurité est assurée par cryptage des fichiers système et par un accès en https. Par ailleurs, les journaux des opérations sont enregistrés quotidiennement, les liens de téléchargement sont à durée de vie courte, les dossiers à téléchargés sont protégés par mot de passe et le nombre de téléchargements autorisés est limité à un par destinataire.

adresses pour lesquelles il n’y avait pas eu de contact et pour tenter de convaincre certains des ménages qui avaient refusé de participer. De janvier à avril 2013, les 1000 adresses de réserve ont été exploitées par téléphone et en face-à-face par TNS Sofres. Au terme de cette procédure, 1039 personnes ont accepté de participer au panel.

## 2.1 La base de sondage

### 2.1.1 L’échantillon fourni par l’INSEE

Le panel ELIPSS est un panel aléatoire d’individus résidant en ménages ordinaires au sens de l’Institut national des statistiques et des études économiques (INSEE). Sont donc exclues les personnes sans domicile ou vivant dans des habitations précaires, les personnes résidant dans des collectivités (prisons, maisons de retraite, résidences étudiantes, etc.), et également celles ne maîtrisant pas suffisamment la langue française pour répondre aux questionnaires auto-administrés.

L’objectif initial était de recruter 1500 personnes.

La base de sondage est constituée des logements recensés en 2011<sup>5</sup>. A partir de cette base de sondage, un échantillon de 4500 logements a été tiré par l’INSEE<sup>6</sup> par sondage à deux degrés stratifié par région et type de commune (urbain/rural). Le premier degré est un tirage d’unités primaires (UP, également appelé à l’INSEE, ZAE – zones d’action enquêteur – correspondant à des communes ou des groupes de communes) avec un taux de sondage de 1/3<sup>7</sup>. Le second degré est un tirage de logements au sein des 241 PSU. Les 4500 adresses ont été divisées en trois sous-échantillons : un échantillon principal de 3500 adresses et deux échantillons de réserve de 300 et 700 adresses à utiliser au cas où l’objectif des 1500 panélistes n’était pas atteint.

Au sein de chaque logement sélectionné par l’INSEE, une seule personne était ensuite tirée au sort. Cette personne devait être âgée de 18 à 75 ans au moment de l’inclusion et être en mesure de répondre à des questionnaires auto-administrés en français.

Le poids de sondage est obtenu à partir de la probabilité de sélection du logement (fournie par l’INSEE) et de la probabilité de sélection d’un individu au sein de son logement (fournie au cours du protocole de recrutement).

### 2.1.2 La préparation de la base de sondage

La phase préliminaire du travail a consisté à saisir les fiches adresses fournies par l’INSEE et sur lesquelles les noms et adresses des chefs de ménage étaient manuscrites. Une vérification de la cohérence et de la vraisemblance des adresses saisies a été faite manuellement par Internet par les deux gestionnaires de panel de l’INED avec l’aide d’un vacataire pour deux mois. Elle a été complétée par l’institut de sondage pour les adresses des deux échantillons de réserve. Le format des adresses a ensuite été vérifié automatiquement par la Poste afin de les confronter au registre des adresses et de mettre celles-ci au format normalisé La Poste en vue du publipostage. La base a également été enrichie par la recherche des numéros de téléphone.

Au total, 76 adresses dans l’échantillon principal et 13 dans l’échantillon de réserve ont été considérées comme inexploitable. Par ailleurs, environ 10% des lettres sont revenues avec la

---

<sup>5</sup> 3% des logements ont été recensés en 2009 ou en 2010.

<sup>6</sup> Cet échantillon a été fourni gracieusement par l’INSEE à ELIPSS à des fins expérimentales.

<sup>7</sup> L’Échantillon-Maître de l’INSEE est constitué de 567 UP. La taille effective de l’échantillon du pilote ELIPSS n’était pas compatible avec la mobilisation de l’ensemble des UP définies par l’INSEE.

mention NPAI (n'habite pas à l'adresse indiquée), ces adresses ont directement reçu la visite d'un enquêteur.

## 2.2 Le déroulement du terrain

### 2.2.1 Les campagnes d'envoi des lettres et de rappel téléphonique

La première phase de recrutement s'est déroulée comme suit : une première lettre d'annonce (L1) a été envoyée par courrier le 08 juin 2012 auprès de l'échantillon principal (soit 3424 adresses). Cette lettre expliquait le panel ELIPSS en général et son fonctionnement avec la tablette, ainsi que la convention et la participation aux enquêtes. Dans cette lettre se trouvait également une sollicitation de la personne de référence du ménage à donner l'âge et le sexe de chaque membre du foyer afin de pouvoir tirer un individu parmi les éligibles (c'est-à-dire. francophone et âgé entre 18 et 75ans). La lettre d'invitation offrait trois modes de réponse aux ménagés sélectionnés. Il leur était possible de renvoyer un coupon-réponse par courrier postal, d'utiliser un accès personnel sur le site internet elipss.fr ou de téléphoner au numéro indiqué sur la lettre. Ce premier envoi a recueilli 7% d'acceptations, 2% de refus et un taux de non réponse de 92%.

Une première lettre de relance (R1) a été envoyée le 21 juin 2012 aux non-répondants à L1 (soit 2884 adresses). Cette première relance (cumulée avec les résultats de la L1) a permis d'avoir un taux d'acceptation de 12%, contre un taux de refus de 6% et un taux de non réponses de 83%.

Enfin une dernière relance postale a été envoyée aux non répondants de L1 et R1 pour lesquels nous n'avions pas trouvé de numéro de téléphone, et une relance téléphonique pour ceux pour lesquels nous disposions d'un numéro. La relance postale a été envoyée le 05 juillet 2012 à destination de 1566 adresses, et la relance téléphonique a été réalisée auprès de 1324 adresses entre la période du 30 juin 2012 au 3 juillet 2012.

Le bilan du 4 septembre 2012, qui regroupe tous les résultats de la lettre d'annonce L1 et des relances R1/R2, montre que cette première phase de recrutement assuré par l'équipe ELIPSS a permis d'obtenir 15% d'acceptation (530 adresses, et plus tard 541 adresses avec les acceptations tardives), 20% de refus (677 adresses) et 65% de non-réponses (2217 adresses dont 257 adresses en retours postaux ). Il est à noter que deux tiers de ces refus ont été obtenu pendant la relance téléphonique R2 (451 refus et 79 acceptations dans cette relance).

### 2.2.2 Le recrutement par les enquêteurs de TNS Sofres

La Sofres a d'abord eu à ré-exploiter les refus « modérés »<sup>8</sup> et les non-contacts de l'échantillon principal puis les adresses des échantillons de réserve, soit au total 3655 adresses. Le recrutement en face-à-face était privilégié mais le téléphone pouvait aussi être utilisé pour prendre contact avec le ménage et recruter le panéliste. Les enquêteurs devaient décrire avec soin leurs tentatives de contact des ménages et des individus sélectionnés à l'aide d'une fiche de contact détaillée, puis présenter le projet à l'aide d'un argumentaire, faire une démonstration succincte de la tablette, présenter les termes de la convention de participation au panel et la faire signer, le cas échéant.

---

<sup>8</sup> Parmi les refus obtenus par courrier ou téléphone, 542 ont été réexploités par l'institut de sondage. En particulier, les refus liés à des questions de confidentialité ou d'anonymat n'ont pas été réexploités. Les raisons évoquées pour les refus à réexploiter sont relativement classiques dans les enquêtes par questionnaire comme le manque d'intérêt pour le panel, le manque de temps pour répondre auxquels se sont ajoutés des refus de personnes déclarant ne pas être à l'aise avec les nouvelles technologie et de personnes ne souhaitant pas avoir un accès à internet ou une tablette. A l'issue de la réexploitation de ces refus, 49 ont donné lieu à une acceptation à participer.

En tout, 110 enquêteurs de TNS Sofres ont suivi une demi-journée de formation approfondie, dont le contenu et le déroulement ont été préparés conjointement par l'équipe ELIPSS et l'institut de sondage. La formation comportait une partie générale de présentation du projet et des objectifs de l'opération ainsi qu'un module spécifique dédié au protocole de contact et à la nécessité du remplissage des fiches contacts. En effet, contrairement aux opérations ordinaires assurées par ces enquêteurs, le recrutement ELIPSS repose sur une méthodologie aléatoire stricte, interdisant le remplacement des non-répondants. Des informations précises étaient de ce fait exigées de la part des enquêteurs.

Enfin, une partie importante de la formation était consacrée aux spécificités ELIPSS : présentation de la convention à faire signer par l'individu acceptant, et démonstration de la tablette et de ses fonctionnalités. En effet, afin de faciliter l'acceptation et de rendre concrète la participation au panel, les enquêteurs étaient équipés d'une tablette du même modèle que celui des panélistes et sur laquelle un questionnaire de type « ELIPSS » était présenté.

### Bilan quantitatif du recrutement

Le recrutement s'est effectué en quatre vagues selon la provenance des adresses et les modes de contact utilisés par la Sofres. La vague 1 reprend l'échantillon des non-contacts et des refus légers issu du recrutement INED, ce qui constituait au final 2668 adresses de l'échantillon principal. Cette première vague s'est effectuée en face-à-face entre le 24 octobre 2012 et le 25 mars 2013.

Pour les échantillons de réserve, les 375 adresses pour lesquelles aucun numéro de téléphone n'avait pu être retrouvé ont été approchées en face-à-face dans la vague 2 entre le 10 février 2013 et le 2 avril 2013. Tandis que les adresses avec un numéro de téléphone ont été recrutées en CATI dans la vague 3 du 7 février 2013 au 19 février 2013. Une partie seulement de ces adresses a été recrutée par téléphone, l'autre partie a été recrutée en face-à-face (vague 3bis).

Le Tableau 2 montre les résultats du recrutement Sofres selon les différentes vagues.

**Tableau 2 : Bilan général du recrutement par la Sofres**

Echantillon	Vague	Mode recrutement	Adresses exploitées	Acceptation			
				Ménage		Individu	
Principal	Vague 1	Réexploitation en face-à-face	2668	510	19%	395	15%
Réserves	Vague 2	Face-à-face uniquement	375	124	33%	107	29%
	Vague 3	Téléphone	612	177	29%	96	16%
	<i>dont : Vague 3bis</i>	<i>Face-à-face après téléphone</i>	<i>366</i>	<i>77</i>	<i>21%</i>	<i>54</i>	<i>15%</i>

La vague 1 obtient le plus faible taux de participation (19%), rappelons que cet échantillon contient essentiellement les non-réponses et les refus léger du recrutement courrier. De fait, même en face-à-face, ces ménages sont plus difficiles à recruter.

C'est la vague 2 qui a le plus fort taux d'acceptation (33%). Ces ménages ont été recrutés directement en face-à-face sans tentative de contact par courrier ni par téléphone. La vague 3 a obtenu un taux d'acceptation de 29% pour un recrutement par téléphone. La vague 3bis est moins performante (21%). Comme la vague 1, cette vague est une réexploitation des non-répondants de la vague 3.

C'est donc le face-à-face en premier contact qui semble le plus performant pour le recrutement du pilote et qui sera donc être le mode de recrutement choisi pour l'élargissement du panel en 2016.

## **Bilan qualitatif de la prestation Sofres**

La prestation assurée par la Sofres est satisfaisante, mais des problèmes de gestion des enquêteurs, de respect des délais et certains défauts de qualité des données sont à signaler.

Précisons en premier lieu, que le type de travail demandé aux enquêteurs en a dérouté plus d'un, malgré le soin apporté à leur formation. Certains se sont révélés mal à l'aise dans le maniement de la tablette tandis que d'autres n'ont pas pris suffisamment de soin à remplir les fiches contact. Enfin, leur travail consistait plus à prendre contact et à convaincre les membres du ménage et *in fine* une personne sélectionnée aléatoirement de participer à un panel plus qu'à administrer un questionnaire. Le travail d'approche et de persuasion était donc très important relativement au temps passé à remplir le questionnaire de contact (réduit à quelques questions portant sur le sexe, l'âge et l'équipement Internet du ménage).

Il a aussi été relevé un certain nombre de difficultés dans la gestion des enquêteurs : les bilans hebdomadaires permettaient de vérifier l'avancement des travaux et de repérer les enquêteurs en difficulté ou ne respectant pas le cahier des charges. Si très peu étaient concernés, il a été difficile de pouvoir faire intervenir rapidement les superviseurs Sofres pour leur venir en aide ou les recadrer. D'une part, la communication passait par des intermédiaires entre l'équipe ELIPSS et les enquêteurs, et d'autre part, le type inhabituel de prestation demandée aux enquêteurs suscitait quelques hésitations sur les critères qualité à appliquer et sur la marche à suivre.

La collecte a duré trois mois de plus qu'initialement prévu à cause notamment du retard pris au démarrage du terrain. La première vague de recrutement qui devait initialement commencer en septembre n'a démarré qu'en novembre et s'est terminée pendant les fêtes de fin d'année. Cette période a été peu propice aux recrutements d'autant plus qu'elle correspondait aussi à une période de forte charge pour les enquêteurs.

Enfin, de nombreuses incohérences entre le document du bilan de collecte et les fichiers de recrutement (tous fournis par la Sofres) ont imposé un travail important de vérification. Dans le présent document, les analyses sont fondées sur les données détaillées contenues dans les bases de données et non sur les bilans de collecte agrégés.

### **2.2.3 L'expérimentation des bons cadeaux**

La littérature montrant un effet positif maximal des incitations financières non-conditionnelles (c'est-à-dire indépendantes de la participation effective) dans les enquêtes postales et internet (Edwards, Roberts et al. 2009 ; Scherpenzeel, 2009), nous avons décidé de procéder à une expérimentation similaire pour le recrutement dans ELIPSS, sur l'échantillon principal. Cette incitation consistait en un bon cadeau de 10 euros. 2000 bons cadeaux ont été joints à la première lettre d'invitation envoyée aux logements issus de l'échantillon principal.

Cette expérimentation est à notre connaissance, une première en France. Elle a donné des résultats très positifs qui sont détaillés ci-après.

## **2.3 L'entrée des panélistes**

### **2.3.1 La convention**

Pour faire partie du panel ELIPSS, les personnes sélectionnées doivent signer une convention qui régit les conditions de participation aux enquêtes et d'utilisation de la tablette.

En signant cette convention, les panélistes prennent trois engagements. Ils s'engagent à répondre personnellement et régulièrement aux enquêtes. Ils s'engagent aussi à prendre soin de la tablette et informer l'équipe ELIPSS en cas de casse, de perte ou de vol. Ils s'engagent enfin à restituer la

tablette à la fin de leur participation au panel. En échange, ils peuvent librement utiliser la tablette et internet dans le respect de la législation en vigueur. A tout moment, ils peuvent suspendre ou mettre fin à leur participation.

En cas de non-réponse prolongée aux questionnaires ou d'utilisation frauduleuse de la tablette, les panélistes peuvent être exclus du panel.

Outre la mise à disposition gratuite de la tablette et de l'abonnement internet 3G, la convention interdit l'utilisation commerciale des enquêtes et garantit l'anonymat des réponses aux questionnaires.

### **2.3.2 La prise en main de la tablette**

Une fois la convention signée et renvoyée, les panélistes reçoivent une tablette à leur domicile. En page d'accueil, est notamment installée l'application qui permet de remplir les questionnaires.

Les panélistes se sont vus proposer une formation téléphonique pour prendre en main la tablette, découvrir l'application ELIPSS et, le cas échéant, paramétrer le wifi. Ces formations ont été assurées par un prestataire extérieur à partir d'un script défini par l'équipe ELIPSS. Deux tiers des panélistes ont été formés, 30% n'ayant pu être joints par téléphone<sup>9</sup> et 5% ayant refusé la formation. A l'issue de la formation, les formateurs devaient évaluer le niveau d'aisance avec les nouvelles technologies. Ils ont estimé que 20 % des panélistes formés n'étaient pas à l'aise avec la tablette.

### **2.3.3 La première enquête**

Lancée en décembre 2012, la première enquête comportait deux parties :

- Le « didacticiel » devait permettre aux panélistes de se familiariser avec différents types de question et le design des enquêtes ELIPSS.
- Le module « Enquêtes et internet » visait à mesurer l'accès à internet, les pratiques numériques et la participation aux enquêtes avant l'entrée dans le panel ELIPSS.

Administrée jusqu'à la fin du recrutement en avril 2013, 90% des panélistes ont répondu à cette première enquête.

D'après cette enquête, environ 90% des panélistes disposent d'un accès internet à domicile. Ils sont 79% à utiliser internet tous les jours et 6% à se connecter moins d'une fois par semaine. Cette enquête nous apprend également que 13% des panélistes affirment avoir toujours refusé de participer à une enquête avant le panel ELIPSS. Elle confirme enfin que la mise à disposition d'une tablette tactile a été la principale motivation pour accepter de participer au panel ELIPSS (citée par 62% des panélistes), suivie de la confiance dans les institutions impliquées dans le projet (pour 46% des panélistes). Viennent ensuite l'originalité du projet (37%), l'intérêt pour la recherche (32%) et la mise à disposition d'un abonnement internet (13%). Une analyse plus détaillée des motifs déclarés par les panélistes est présentée au paragraphe 4.3.

---

<sup>9</sup> Dans la majorité des cas, ces personnes n'étaient pas relancées pour suivre la formation dans le cas où elles avaient répondu à la première enquête.

## 2.4 Le bilan du recrutement

### 2.4.1 Un taux de recrutement de 27%

Rappelons qu'au terme de l'exploitation des 4500 adresses de l'échantillon fourni par l'INSEE, 1039 individus ont accepté de faire partie du panel et ont signé la convention. On peut distinguer ici deux types d'accord à participer (initial consent - IC) tel que défini par Callegaro et DiSogra (2008). En effet, dans la procédure spécifique aux adresses de l'échantillon principal, une première étape a consisté à demander l'accord du ménage pour décrire la composition des membres du logement<sup>10</sup>, puis le second accord intervient au niveau individuel au moment de la signature de la convention de participation au panel. A la différence de la procédure décrite par Callegaro et DiSogra (2008, et 2015), il n'y a pas eu de « profil/connection stage », ou plus précisément, elle est confondue avec l'étape dite d' « initial consent » au niveau de l'individu.

Les résultats finaux de la procédure de recrutement du panel ELIPSS sont reproduits dans le Tableau 3 ci-dessous. Ils ont été calculés à partir des formules de l'*American Association for Public Opinion Research* (AAPOR), en considérant, comme le proposent Callegaro et DiSogra (2008, p.1018), le RR3 comme taux de recrutement (RECR), tel que définit ci-dessous :

$$\text{Recruitment rate (RECR)} = \frac{IC}{IC + (R + NC + O) + e(UH + UO)}$$

where

*IC* = initial consent

*R* = cases directly and actively refusing

*NC* = noncontacts

*O* = other cases

*UH* = unknown if household is occupied

*UO* = unknown other

*e* = estimated proportion of cases of unknown eligibility that are eligible.

Notons que l'éligibilité est parfois délicate voire impossible à déterminer quand aucun contact avec le ménage n'a pu avoir lieu. Pour tous ces cas, nous avons appliqué le taux d'éligibilité (*e*) observé sur les unités pour lesquelles l'éligibilité est connue.

---

<sup>10</sup> Ceci est particulièrement vrai dans le cas de la première phase de recrutement de l'échantillon principal puisque les ménages devaient avoir une démarche volontaire pour renvoyer le coupon-réponse, téléphoner aux gestionnaires de panel ou remplir le formulaire en ligne pour décrire des habitants du logement.

**Tableau 3 : Taux de réponse des ménages et taux de recrutement des individus**

		Ménages	Individus	Echantillon principal (individus)	Echantillons de réserve (individus)
Eligibles	Accord (IC)	1349	1039	797	242
	Refus (R)	1496	1687	1279	408
	Non contacts (NC)	0	2	2	0
	Autres cas (O)	27	144	137	7
Inéligibles		485	485	381	104
Eligibilité inconnue	Adresse non localisée (UH)	350	350	265	85
	Ménage injoignable (UO)	793	793	639	154
	e	0.856	0.856	0.853	0.863
RR3 = RECR		35%	27%	27%	28%

Les adresses traitées directement en face-à-face par la Sofres (n=546) ont donné un taux de recrutement de 29% contre un taux de 27% pour l'ensemble du recrutement. Le face-à-face a donc été plus efficace d'autant plus que les recrutements ont été obtenus nettement plus rapidement. En effet, le face-à-face permet une démonstration de la tablette et une signature immédiate de la convention le cas échéant. Le processus de recrutement s'en trouve donc considérablement accéléré et la gestion simplifiée. C'est pourquoi le recrutement des 2700 panélistes supplémentaires en 2016 sera entièrement confié à un institut de sondage et se fera directement en face-à-face après l'envoi d'une lettre-avis.

#### 2.4.2 L'effet des bons cadeaux

Les 2000 bons furent envoyés aléatoirement aux 3424 adresses exploitables de l'échantillon principal. Ainsi, à la fin du recrutement, 797 panélistes furent recrutés au sein de cet échantillon, 509 (parmi les 2000 adresses concernées) en ayant reçu un bon cadeau et 288 (parmi les 1424 adresses restantes) sans bon cadeau préalable. Dans l'échantillon principal, les bons sont ainsi associés à une hausse significative du recrutement : 20.2% sans bon cadeau contre 25.5% avec bon cadeau ( $p=0.0004$ , OR brut=1.35).

Un modèle multivarié permet d'explorer plus précisément l'effet du bon cadeau et de montrer que seuls l'âge du chef de ménage, son niveau de diplôme, la taille du logement et le fait qu'il soit en HLM ou non ont de l'importance, toutes choses égales par ailleurs, sur la probabilité de devenir panéliste. On peut par ailleurs montrer qu'il n'existe pas d'interaction significative entre l'envoi d'un bon cadeau et : l'âge de la personne de référence du ménage, le diplôme de la personne de référence du ménage, le fait de résider en HLM et le nombre de pièces du logement.

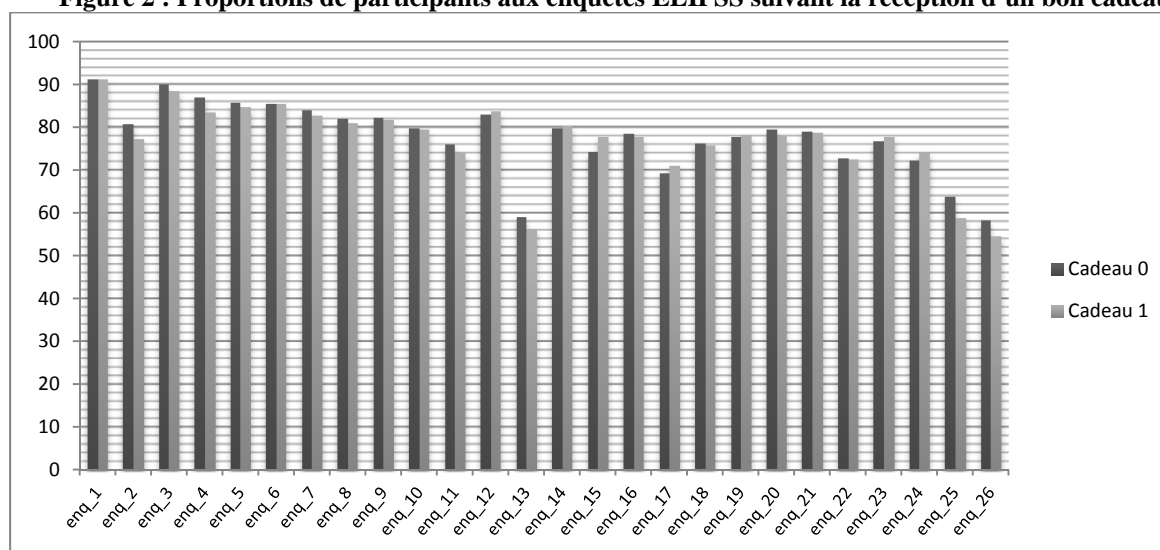
Enfin, l'envoi de bons cadeaux ne modifie globalement pas la participation au panel. Sur les 20 premières enquêtes, le nombre moyen de réponses vaut 15.96 parmi les récipiendaires contre 16.11 parmi les autres ( $p=0.73$ ).

Le détail des participations aux enquêtes des panélistes de l'échantillon principal (n=797) suivant l'envoi initial d'un bon cadeau (n=509) ou non (n=288) est représenté sur la Figure 2. Elle montre



clairement que la réception initiale d'un bon cadeau n'affecte en rien la participation aux enquêtes successives.

**Figure 2 : Proportions de participants aux enquêtes ELIPSS suivant la réception d'un bon cadeau**



Comme l'envoi des bons cadeaux augmente le taux de recrutement sans affecter la participation ultérieure aux enquêtes, il pourrait être intéressant de cibler ces incitations matérielles sur des populations qui sont sous-représentées, en particulier les ménages dont la personne de référence a plus de 55 ans. Pour des raisons budgétaires, le recours à des bons-cadeaux pour le recrutement de 2700 nouveaux panélistes en 2016 n'a pas encore été décidé.

### 2.4.3 La représentativité socio-démographique initiale

Le Tableau 4 montre la distribution de quelques variables socio-démographiques (utilisées pour le calage) dans l'échantillon de l'enquête et dans la population (à partir des données du recensement de la population en 2011). Ces distributions sont pondérées par le poids de sondage pour l'échantillon. Les personnes sous-représentées sont les femmes, les personnes âgées de 55 ans et plus, de nationalité étrangère, peu diplômées (niveau inférieur au bac) et habitant la région parisienne, le bassin parisien, le Nord et l'Est de la France. A l'inverse, les personnes sur-représentées sont les hommes, les personnes âgées de 35 à 54 ans, les français de naissance, les personnes les plus diplômées (bac et plus) et qui habitent dans les grandes régions suivantes : Centre, Ouest et Méditerranée.

**Tableau 4 : Structure de l'échantillon de répondants à l'enquête annuelle d'avril 2013 (EA1 2013)**

			Poids non corrigé non calé	Poids corrigé non calé
Variable	Modalité	Population (RP2011 %)	Echantillon EA1 2013 (%)	Echantillon EA1 2013 (%)
Age	18-24ans	11.86	10.8	9.59
	25-34ans	17.67	18.4	17.05
	35-44ans	19.74	24.6	22.01
	45-54ans	19.65	21.48	19.57
	55-64ans	18.31	16.61	20.05
	65-75ans	12.77	8.12	11.74
Sexe	Homme	48.66	51.07	48.92
	Femme	51.34	48.93	51.08
Nationalité	Français (naissance)	88.1	90.43	89.47
	Français (acquisition)	5.45	6.15	6.33
	Etranger	6.45	3.42	4.2
Diplôme	Aucun/CEP/BEPC	27.48	15.5	19.88
	CAP/BEP	24.77	20.49	22.95
	Bac_bac+2	32.92	40.01	37.05
	Bac+3 et plus	14.83	23.99	20.12
ZEAT	Région parisienne	19.16	14.79	15.77
	Bassin parisien	17.01	12.34	12.64
	Nord	6.39	4.97	3.25
	Est	8.66	6.74	5.81
	Ouest	13.53	16.21	17.69
	Sud-ouest	11.03	13.45	12.63
	Centre-Est	12.14	15.54	15.37
	Méditerranée	12.08	15.95	16.84

Note : les proportions calculées dans l'échantillon sont pondérées par le poids de sondage, et ensuite par le poids corrigé avant calage.

Les résultats après correction de la non-réponse totale sont présentés dans la dernière colonne du tableau ci-dessus. Une amélioration significative de la représentativité a été apportée pour l'âge, la nationalité, et le sexe. Il y a également une légère amélioration pour le diplôme mais toujours avec une faible proportion des moins diplômés. Par contre, nous constatons une détérioration non négligeable pour la région de résidence (Nord et Ouest notamment).

#### **2.4.4 Les individus qui n'avaient pas d'accès internet avant leur entrée dans le panel**

D'après leurs réponses lors du recrutement et lors de la première enquête sur les pratiques numériques, environ 10% des panélistes n'avaient pas d'accès internet à domicile avant d'intégrer le panel ELIPSS. Cette proportion est inférieure à celle mesurée dans la population générale par l'enquête TIC de l'INSEE en 2013 (14,8%).

Un article que nous venons de publier avec Mélanie Révilla sur cette question dans *Social Science Computer Review* souligne que l'apport des individus non équipés d'Internet à la représentativité du panel ELIPSS est limité à l'exception notable de quelques variables sociodémographiques importantes comme le diplôme, la vie en couple et la nationalité (Révilla et al. 2015). L'équipement uniforme des panélistes, mobile qui plus est, présente toutefois d'autres avantages. Il uniformise

l’affichage et les conditions de passation, ce qui permet de s’affranchir des problèmes de comparabilité de mesure liés aux réponses sur des équipements internet différents.

### **3 Les enquêtes : des projets de recherche à la diffusion des données**

#### **3.1 La sélection des enquêtes**

Les propositions d'enquête se font lors d’appels à projets. Jusqu'en 2013, dans le cadre de l'étude pilote, ces appels à projets étaient réservés aux équipes de recherche membres d'une institution partenaire de l'équipement d'excellence DIME-SHS.

##### **3.1.1 Le comité scientifique et technique DIME quanti**

Ce sont les membres du Comité scientifique et technique (CST) qui sont chargés de sélectionner les enquêtes administrées aux panélistes. Le CST est actuellement composé des experts suivants :

- Michel Bozon, sociologue, directeur de recherche à l'INED
- Rémy Caveng, maître de conférences en sociologie à l'Université de Picardie Jules Verne (Centre Universitaire de Recherches sur l'Action Publique et le Politique, Centre européen de sociologie et de science politique et Laboratoire de Sociologie Quantitative)
- Jérôme Cubillé, chercheur à EDF R&D
- Bernard Denni, professeur de science politique à Sciences Po Grenoble (PACTE)
- François Denord, chargé de recherche en sociologie au CNRS (Centre européen de sociologie et de science politique)
- Céline Goffette, chargée d'études au département des Entrées et Evolutions dans la Vie Active au CEREQ
- Anne Jadot, maître de conférences en science politique à l'Université Nancy 2
- Cécile Lefèvre, professeur de sociologie à l'Université Paris Descartes (Centre de recherche sur les liens sociaux)
- Muriel Letrait, ingénieur d'études à l'Université Paris Descartes (Centre de recherche sur les liens sociaux)
- Pierre Mercklé (président), maître de conférences en sociologie à l'ENS de Lyon (Centre Max Weber)
- Gaël de Peretti (représentant de l'Insee), administrateur INSEE, chef de la division Recueil et traitement de l'information au sein du Département des méthodes statistiques
- Thierry Rocher, ingénieur de recherche, adjoint au chef du bureau de l'évaluation des élèves, Direction de l'évaluation, de la prospective et de la performance - DEPP, Ministère de l'éducation nationale
- Élise Tenret, maître de conférences en sociologie à l'Université Paris Dauphine
- Loup Wolff, administrateur INSEE

Les membres du CST apportent une expertise méthodologique et disciplinaire. Un représentant de l'Insee y siège. Actuellement, il s'agit de Gaël de Peretti.

##### **3.1.2 L'évaluation des projets**

L'évaluation d'une enquête s'appuie sur au moins deux rapports établis par des membres du CST. Lorsque la compétence disciplinaire n'est pas représentée dans le CST, nous avons recours à des

rapports d'experts extérieurs. La finalité exclusivement scientifique constitue le critère impératif d'éligibilité des enquêtes et l'utilisation commerciale est exclue.

Les projets sont évalués selon les critères suivants :

- la qualité scientifique de la proposition : objectifs, état de l'art, originalité de la proposition, pertinence des échelles ou des indicateurs retenus, résultats attendus ;
- la pertinence de la collecte par le panel ELIPSS : faisabilité (taille de l'échantillon, adéquation des formats des questions à la tablette...), concision du questionnaire (notamment par le recours aux questions de l'enquête annuelle ELIPSS ou d'autres enquêtes ELIPSS déjà réalisées), dimension longitudinale, exploitation des possibilités techniques liées à internet et à la tablette ;
- l'intérêt méthodologique de la proposition : comparabilité des résultats avec d'autres sources, innovations méthodologiques...

Des règles concernant la diffusion des données, le temps d'enquête et les liens avec les panélistes sont décrites dans les appels à projets. Les projets retenus font l'objet d'une convention établissant la copropriété des données entre l'équipe de recherche porteuse du projet et DIME-SHS. Cette convention prévoit le dépôt des données au CDSP et autorise la diffusion des données à la communauté scientifique après une période d'exclusivité d'un an maximum pour l'équipe de recherche porteuse du projet. Compte tenu du temps limité d'interrogation mis à la disposition de la communauté scientifique, les propositions d'enquête longue (plus de 30 minutes) ou d'enquête longitudinale doivent être particulièrement justifiées. La durée annuelle cumulée ne peut excéder 60 minutes. Par ailleurs, aucune incitation financière ne peut être versée aux panélistes et les panélistes ne peuvent pas interagir les uns avec les autres, pour participer à des jeux interactifs par exemple.

Outre la sélection des projets par le CST, les questionnaires proposés sont également examinés par l'équipe Elipss. Ces questionnaires ayant pour objectif une recherche particulière, il revient à l'équipe Elipss de considérer ceux-ci d'un point de vue longitudinal. Ainsi, l'équipe prend soin d'économiser les répétitions de questions injustifiées, d'optimiser les modes de questionnement et d'éviter les questions qui pourraient indisposer les répondants.

Par ailleurs, inscrit au registre du Correspondant informatique et libertés (CIL) du CNRS (n°2-12030), le dispositif ELIPSS bénéficie de l'expertise d'Emilie Masson, juriste dans l'équipe CIL du CNRS, qui est de plus en plus associée aux projets d'enquête. Afin de veiller au plus strict respect de la loi Informatique et libertés, les projets déposés sont donc également discutés d'un point de vue juridique en amont de leur sélection par le CST.

### **3.1.3 Les projets soumis**

Depuis le premier appel à projets en 2011, 32 propositions d'enquêtes ont été reçues (cf. Tableau 5). Les trois premiers appels à projets étaient réservés aux équipes de recherche membres d'une institution partenaire de l'équipement d'excellence DIME-SHS. Le dernier appel, ouvert sur le pilote à toute la communauté scientifique, a permis de recueillir davantage de propositions venant d'institutions et de disciplines diverses.

Plusieurs types de décision peuvent être pris par les membres du CST : rejet du projet d'enquête ; rejet du projet d'enquête en l'état et demande de révision et de re-soumission ; acceptation sous réserve de modifications ; acceptation en l'état, sans modification. Sur les trois premiers appels à projets, le taux moyen de sélection a été de 65 % (13 projets sur 20), le plus souvent après demande de modifications ou de révision.

**Tableau 5 : Décisions du CST quant aux projets soumis en 2011, 2012, 2013 et 2015**

	2011	2012	2013	2015
<b>Nombre de propositions</b>	<b>5</b>	<b>8</b>	<b>7</b>	<b>12</b>
Nombre de projets acceptés sans modification	2	3	1	0
Nombre de projets acceptés après modifications	1	1	1	4
Nombre de projets acceptés après révision et re-soumission	1	1	2	0
Nombre de projets refusés	1	3	3	6
Nombre de projets en cours d'évaluation	0	0	0	2

Plusieurs projets comparatifs ont été acceptés par le comité scientifique et technique en dehors du cadre des appels à projets.

Le premier a été administré au panel ELIPSS en avril 2014. Il s'agit d'un projet porté par Jon Krosnick (Université de Stanford) qui a pour objet de répliquer des expérimentations classiques aux Etats-Unis sur plusieurs panels internet à travers le monde. Les 18 questions visent à étudier la formulation des questions, la tendance à l'acquiescement, les options de non-réponse, l'ordre des questions et l'ordre des modalités de réponse. L'analyse des données est actuellement en cours.

Le second a été administré sur ELIPSS en mai 2014 en même temps que sur les autres panels internet probabilistes en Europe : le GESIS panel, le GIP et le LISS panel. Ce questionnaire construit en commun emprunte des questions à de grandes enquêtes comparatives (ESS – European Social Survey, SHARE – Survey of Health, Ageing and Retirement in Europe, PIAAC – Programme for the International Assessment of Adult Competencies, EES – European Election Study). Outre la production de données comparatives au moment des élections européennes, l'objectif était surtout de réussir à organiser cette collecte simultanée sur les quatre panels de manière à pouvoir envisager d'autres projets comparatifs à l'avenir. Cette collaboration avec les autres panels probabilistes européens a été reproduite en mai 2015 à partir d'un questionnaire sur les réformes du système social proposé par des chercheurs de l'Université de Mannheim.

## 3.2 La production des enquêtes

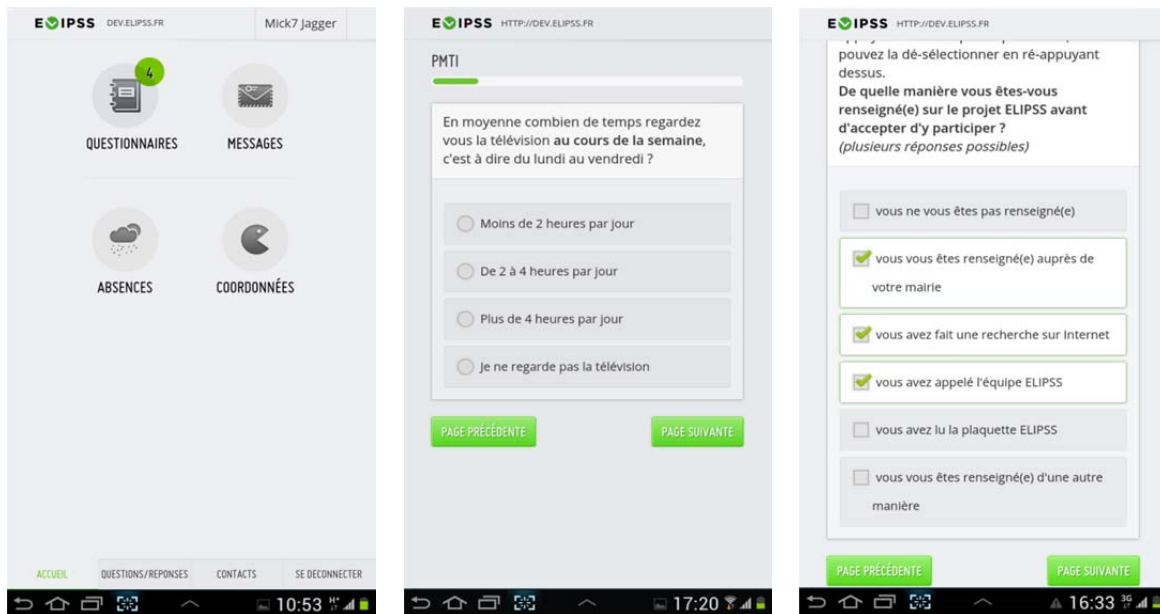
### 3.2.1 L'élaboration des questionnaires

Une fois le projet d'enquête retenu par le CST, des discussions méthodologiques et techniques s'engagent entre l'équipe de recherche et l'équipe ELIPSS. Jusqu'à la mise en production via l'application ELIPSS, le questionnaire évolue en fonction des recommandations du CST, des spécifications techniques, des développements informatiques nécessaires, des options choisies pour la non-réponse et des pré-tests réalisés par l'équipe de recherche et l'équipe ELIPSS.

En règle générale, le terrain démarre le jeudi et dure 5 semaines. Le terrain peut être plus court. Cela a eu lieu pour les enquêtes pré-électorales et post-électorales de 2014 et de 2015 pour lesquelles les dates de terrain dépendent des dates des élections. Le terrain peut aussi durer plus longtemps. Cela est le cas deux fois par an pour tenir compte des fêtes de Noël et des vacances d'été. Par ailleurs, certains terrains ont été allongés de quelques jours suite à des problèmes techniques ou pour permettre d'augmenter le taux de réponse de certaines enquêtes.

### 3.2.2 La tablette comme mode de collecte

Les panélistes répondent aux questionnaires via une application développée en interne et pré-installée en page d'accueil de la tablette (cf. copies d'écran ci-dessous). La conception des questionnaires et la collecte des données en ligne reposent sur le logiciel Blaise. Développée par Statistics Netherlands, cette solution est destinée à la réalisation d'enquêtes de la statistique publique et est utilisée par la plupart des instituts nationaux de statistiques, y compris l'INSEE. Pour publier l'enquête en ligne avec Blaise IS, il a été nécessaire de développer une feuille de style adaptée à l'écran tactile de la tablette.



Certains projets d'enquête ont été l'occasion de développer des fonctionnalités rendues possibles par la mise à disposition du même équipement mobile, connecté et tactile, à tous les panélistes. Par exemple, les tablettes étant équipées de micro, les panélistes ont pu être invités à enregistrer oralement leurs réponses à des questions ouvertes sur l'environnement. Pour utiliser l'auto-enregistrement dans l'enquête d'octobre 2013, il a fallu préalablement avertir les panélistes dont l'espace de stockage apparaissait insuffisant sur notre logiciel de gestion de flotte de tablettes (Mobile Device Manager). La fonctionnalité de glisser-déposer (drag-and-drop) a été spécifiquement développée pour une enquête sur les catégorisations et la connaissance du monde social administrée en juillet-août 2014. Ainsi, les panélistes ont pu de manière intuitive faire glisser des étiquettes de professions pour créer leurs propres groupes sociaux.

Si l'homogénéité des équipements internet fournis aux panélistes est un atout indéniable d'un point de vue méthodologique et d'un point de vue technique, le choix des tablettes implique aussi certaines limites qui tiennent aux problèmes de connexion et à la taille de l'écran. Par exemple, la fonctionnalité d'auto-complétion qui permet de rechercher sa réponse dans une liste (de pays, de communes...) requiert une bonne connexion internet. De même, pour éviter d'avoir à faire défiler le texte verticalement (scroll), la longueur du texte des questions et le nombre de modalités de réponse doivent être limités, notamment pour les batteries de questions.

### 3.2.3 Les enquêtes administrées au panel

Depuis le début de l'année 2013, 30 enquêtes ont été réalisées sur des sujets très divers et avec des taux de réponse élevés (calculés sur le nombre de panélistes invités à répondre).

**Tableau 6 : Liste des enquêtes administrées depuis 2013**

Dates	Titres	Taux de réponse (%)
janvier - mars 2013	Internet et enquêtes	99,5
avril 2013	Enquête annuelle 2013 - Module signalétique	91,4
mai 2013	Enquête annuelle 2013 - Module barométrique	87,4
juin 2013	Pratiques culturelles, médias et technologies de l'information – vague 1	88,2
juillet - août 2013	Fécondité, contraception, dysfonctions sexuelles	87,9
septembre 2013	Dynamique de mobilisation : comprendre la formation des choix électoraux - vague 1	86,9
octobre 2013	Enquête sur les valeurs, l'environnement et l'énergie	84,7
novembre 2013	Situation de couple, intentions de fécondité et opinions sur la famille	87,9
décembre 2013 – janvier 2014	Santé, travail et environnement. Enquête sur les expositions aux poussières inorganiques / Dynamique de mobilisation - vague 2	83,1
février 2014	Solidarité entre générations (soligene) et tronc commun biographique	78,3
mars 2014	Enquête annuelle 2014	90,8
mars 2014	Dynamique de mobilisation – vague 3	64,3
avril 2014	Cross-National Replication of Question Design Experiments	88,4
avril 2014	Dynamique de mobilisation - vague 4	84,2
mai 2014	Etude comparative des panels européens / Pratiques numériques 2014	85,5
mai 2014	Dynamique de mobilisation - vague 5	76,6
juin 2014	Dynamique de mobilisation - vague 6	84,2
juin 2014	Attitudes towards Alternative Partnership Arrangements	86,4

Dates	Titres	Taux de réponse (%)
juillet – août 2014	Catégorisations et connaissances ordinaires de la société (jeu des métiers)	88,6
septembre-octobre 2014	Goûts, morales, groupes sociaux	89,5
novembre 2014	Catégorisations et connaissances ordinaires de la société (jeu des portraits)	84,0
décembre 2014 – janvier 2015	Extrait de l'Enquête sociale européenne / Dynamique de mobilisation - vague 7	90,4
février 2015	Diversité des Modes de vie et Pointes de Consommation d'Énergie	86,1
mars 2015	Dynamique de mobilisation – vague 8	75,5
avril 2015	Enquête annuelle 2015	90,4
avril 2015	Dynamique de mobilisation – vague 9	82,3
mai 2015	Etude comparative des panels européens 2015 / Pratiques numériques 2015	86,9
juin 2015	Dynamique de mobilisation – vague 10	78,5
juillet – septembre 2015	Pratiques culturelles, médias et technologies de l'information – vague 2	86,0

Parmi ses enquêtes, l'enquête dite annuelle a un statut particulier. Il s'agit une fois par an de disposer de nombreuses variables sociodémographiques (état civil, travail et formation, description du ménage, logement et quartier, revenus et patrimoine) ainsi que de variables de croisement et d'indicateurs fréquemment utilisés en sciences humaines et sociales (liens sociaux, loisirs et pratiques culturelles, religion, politique, santé et habitudes de vie). Ainsi, le questionnaire a été construit en collaboration avec plusieurs chercheurs spécialistes des thèmes abordés et avec les membres du comité scientifique et technique. En outre, les questions retenues sont très largement issues d'enquêtes existantes et une grande partie est répétée chaque année depuis 2013. La plupart des variables sociodémographiques sont systématiquement appariées aux fichiers de données diffusés.

#### 4 La qualité statistique du panel

Cette partie présente les différentes méthodes testées et mises en œuvre au cours du pilote afin de contrôler et de mesurer la qualité du panel ELIPSS. D'une part, il a fallu définir la méthode de redressement adaptée à ELIPSS et aux données disponibles. D'autre part, l'attrition ainsi que l'évolution de la représentativité a fait l'objet d'analyses approfondies et continuent d'être suivies au cours des enquêtes. Enfin, la qualité des données est évaluée régulièrement par la comparaison des réponses aux questions répliquées dans les enquêtes annuelles.



## 4.1 Redressement

### 4.1.1 Données et méthode utilisées

Le redressement a été réalisé à partir de données individuelles collectées lors des enquêtes annuelles ELIPSS (âge, sexe, niveau de diplôme, nationalité et ZEAT) à partir des données disponibles sur le logement fournies par l'échantillon initial INSEE ayant servi au recrutement (nombre de pièces, type de logement, statut d'habitation à loyer modéré (HLM), appartenance à une zone urbaine sensible, zone d'étude et d'aménagement du territoire (ZEAT), etc.).

Le calcul des variables de pondération a été effectué en deux phases : une correction de la non-réponse totale et un redressement par calage sur les marges du recensement pour chacune des enquêtes, car cette méthode permet un gain par rapport à un calage direct sur les données individuelles (Haziza and Lesage 2015)<sup>11</sup>.

Pour traiter le potentiel problème lié à la non-réponse totale (notamment la présence de biais), il est nécessaire de corriger le poids des répondants en prenant en compte les non-répondants. Cette correction de la non-réponse totale a été réalisée au niveau du ménage par groupes de réponses homogènes et en plusieurs étapes : d'abord une modélisation de la participation au niveau du ménage par arbre de décision binaire avec la technique du boosting<sup>12</sup>, ensuite une classification hiérarchique ascendante sur la probabilité prédite de participation. C'est en effet au niveau de l'acceptation du ménage que les distorsions de l'échantillon par rapport à la base de sondage sont les plus importantes.

Enfin, un redressement avec un calage sur marges<sup>13</sup> a été appliqué à chaque enquête ELIPSS, en utilisant comme poids initial le poids de sondage corrigé de la non-réponse totale<sup>14</sup>.

Pour réaliser le calage, en raison de la non-réponse de certains panélistes aux enquêtes annuelles de 2013 et 2014, des imputations ont été faites pour obtenir le diplôme et la nationalité. Initialement, des valeurs étaient en effet manquantes pour 80 panélistes parmi les 1039 recrutés. Pour 58 panélistes, ces imputations ont été faites par cold-deck en identifiant ces individus au sein de leur logement dans la base de sondage INSEE à partir de leur sexe et de leur date de naissance. Pour les 22 individus qui n'ont pas pu être retrouvés, les imputations ont été faites par hot-deck en sélectionnant dans l'échantillon de l'INSEE un individu ayant les mêmes caractéristiques géographiques et socio-démographiques que le non-répondant. Ainsi, tous les répondants aux enquêtes ont un poids, même s'ils ont pu ne pas répondre aux enquêtes annuelles fournissant les variables de calage.

### 4.1.2 Résultats

Il existe finalement plusieurs poids de sondage obtenus aux différentes étapes du redressement.

Le poids de tirage est celui fourni avec l'échantillon initial tiré par l'INSEE.

Le poids de sondage est égal au produit du poids de tirage de l'échantillon initial fourni par l'INSEE par l'inverse de la probabilité de sélection d'un individu éligible dans le ménage.

---

<sup>11</sup> Les variables d'intérêt sont évidemment très nombreuses et variables d'une enquête à l'autre. Néanmoins, elles sont très certainement liées aux variables de type socioéconomique et cadre de vie de niveau individuel et collectif utilisées dans le traitement de la non-réponse.

<sup>12</sup> Ce type de modélisation, qui est une méthode itérative, est plus efficace qu'une régression logistique classique et moins sensible à la multicollinéarité qui peut entacher cette dernière d'erreurs lorsque les prédicteurs sont nombreux. Elle permet par ailleurs une prise en compte d'interactions complexes.

<sup>13</sup> Les variables utilisées pendant le calage sont l'âge, le sexe, le diplôme, la nationalité et la ZEAT.

<sup>14</sup> La méthode du raking-ratio a été utilisée sans troncature des poids initiaux. Cette méthode consiste à faire une repondération des poids de sondage en ajustant certaines caractéristiques de l'échantillon sur celles des variables auxiliaires de sources extérieures. Il s'agit ici du recensement de la population 2011.

Le poids corrigé avant calage est obtenu de la façon suivante :

$$\text{Poids corrigé} = \text{Poids de sondage} * \text{Coefficient correcteur}$$

où le coefficient correcteur est égal à l'inverse du taux de participation pour chaque classe obtenue lors de la classification hiérarchique ascendante.

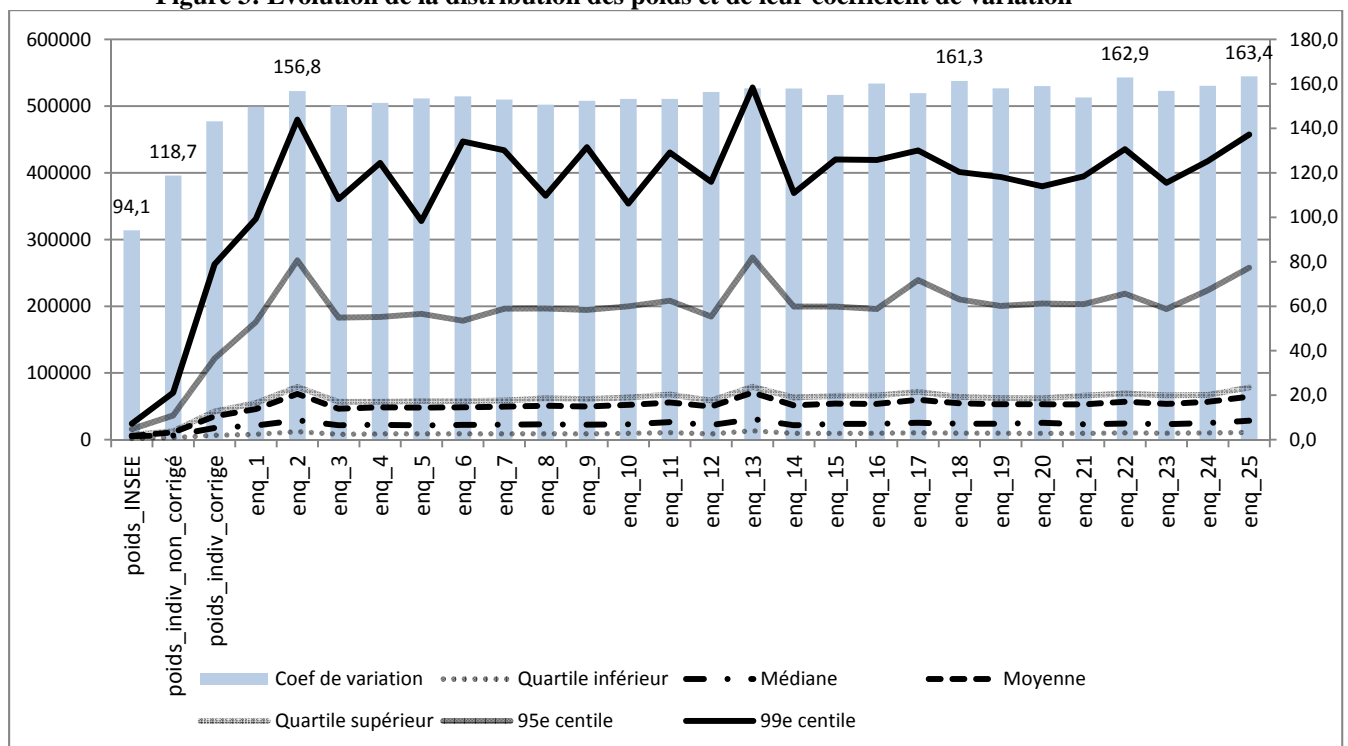
Le tableau ci-dessous présente la distribution des différents poids de sondage. Le poids corrigé est disponible pour 1349 observations qui correspondent aux ménages qui avaient accepté de répondre pour communiquer la composition de leur ménage (ces informations étaient un préalable nécessaire à la sélection d'un individu pouvant devenir panéliste).

**Tableau 7: Caractéristiques des poids avant calage**

Variable	Minimum	Maximum	Somme	N	Coef de variation
Poids de tirage	309.75	38241.39	26461817	4500	94.14
Poids de sondage	310.60	114724.16	14702783	1349	118.75
Poids corrigé	335.18	499346.33	47300532	1349	143.20

La distribution des poids aux différentes étapes, initial (poids de tirage), corrigé de la non-réponse et après calage (pour chaque enquête) est fournie ci-dessous.

**Figure 3: Évolution de la distribution des poids et de leur coefficient de variation**



Note : l'échelle de gauche correspond à la distribution des poids et celle de droite aux coefficients de variation.

On observe le plus grand écart entre le coefficient de variation du poids de tirage (« poids INSEE ») et le poids de sondage corrigé de la non-réponse totale (« poids\_indiv\_corrige »). Il apparaît ensuite que le coefficient de variation se stabilise au fil des enquêtes. Néanmoins, quelques fluctuations sont visibles notamment à cause des enquêtes atypiques (enquêtes à période de collecte très courte...).

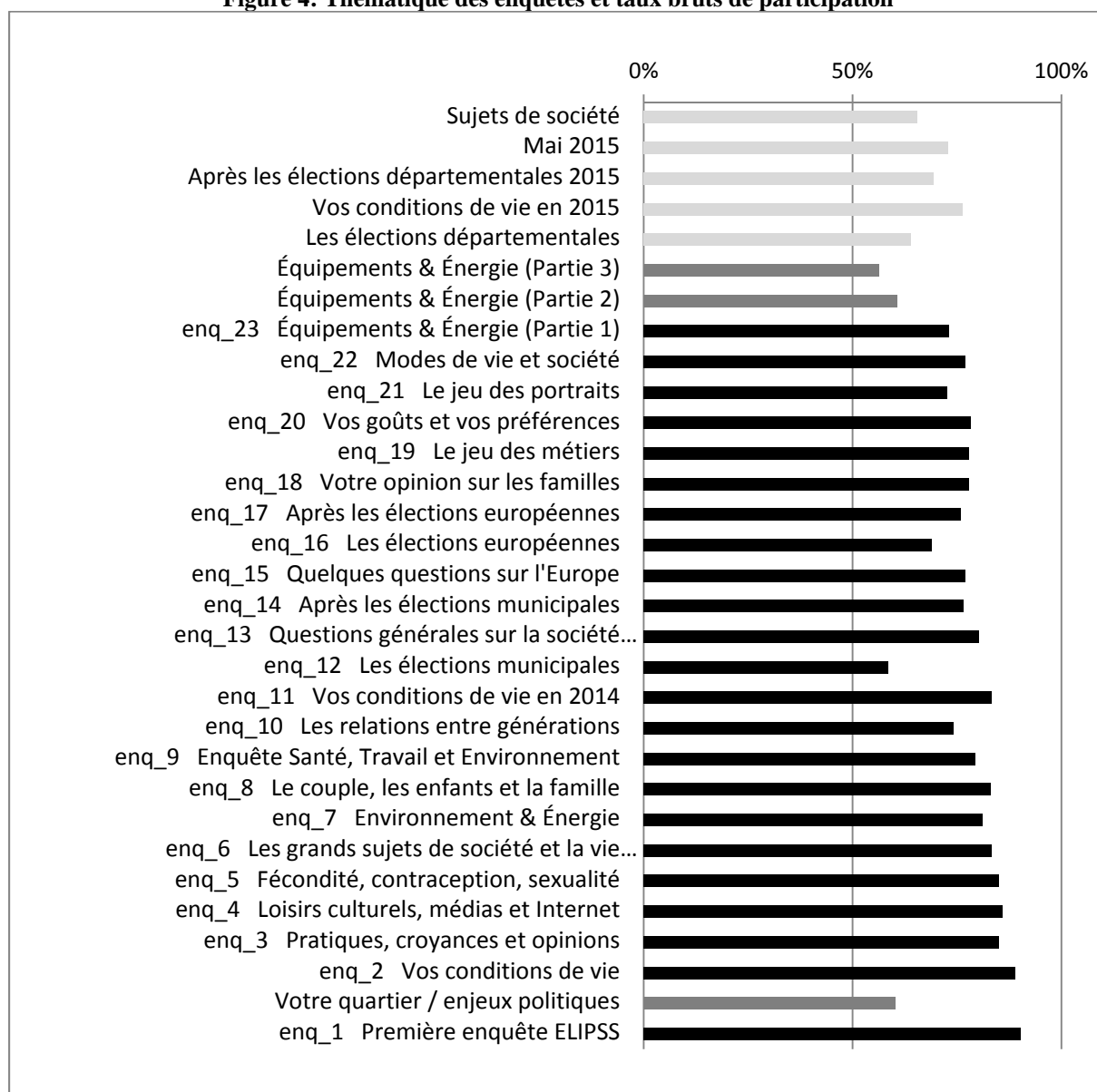
## 4.2 Attrition et représentativité au fil des enquêtes

### 4.2.1 La participation aux enquêtes

La participation au fil des enquêtes est un indicateur important de la qualité du dispositif ELIPSS. Il s'agit ici de s'intéresser au taux brut de participation, c'est-à-dire au ratio du nombre de répondants à chaque enquête sur le nombre de panélistes initial (1039 personnes).

La figure ci-dessous présente les taux bruts de participation aux enquêtes administrées jusqu'en juin 2015.

Figure 4: Thématique des enquêtes et taux bruts de participation



Les 23 enquêtes utilisées pour la plupart des analyses de participation ci-après apparaissent en noir, les enquêtes exclues de ces analyses apparaissent en gris moyen<sup>15</sup> et les enquêtes en cours d'analyse sont indiquées en gris clair.

<sup>15</sup> L'enquête Votre quartier/enjeux politiques, qui fut une enquête « technique » administrée afin de tester certaines fonctionnalités de la tablette et adressées à une fraction des panélistes ; la partie 2 de l'enquête

Si le taux brut de participation diminue au cours des enquêtes, il est également variable d'une enquête à l'autre. Plusieurs éléments pourraient expliquer cette variation. D'abord la durée de collecte des enquêtes qui est parfois réduite, notamment dans le cas des enquêtes électorales ; ensuite, un effet de la longueur du questionnaire telle qu'annoncée dans l'invitation à participer ; enfin, la sensibilité de la thématique n'est pas à exclure. En effet, nous avons procédé à un examen des libellés des enquêtes tels que présentés aux panélistes lors des invitations mensuelles, c'est-à-dire tels que figurant dans la Figure 4. Si l'on considère que les panélistes peuvent être réticents à participer aux enquêtes dont le titre évoque la politique, les conditions ou les modes de vie, la santé, alors un peu plus de la moitié des 23 enquêtes analysées peuvent être classées comme sensibles du point de vue de leur thématique (56%).

Un tri croisé permet de vérifier que la sensibilité de la thématique n'est pas associée à la participation (le taux de participation brut est 78,4% pour les enquêtes dont la thématique peut paraître sensible et 79,4% pour les autres,  $p=0.69$ ). En revanche, la proportion de participants est plus élevée à mesure que la période de collecte s'élève (elle vaut 75% lors d'une collecte de 12 à 23 jours, de 80% lors d'une collecte entre 26 et 31 jours et 83% lors d'une collecte de 32 à 100 jours<sup>16</sup>) et il en va de même avec la durée du questionnaire : 75% pour une durée entre 5 et 15 minutes, 80% pour une durée entre 20 et 25 minutes et 83% pour 30 minutes<sup>17</sup>). Ce dernier effet est paradoxal et pourrait s'expliquer par une attente forte des panélistes au regard du caractère important et scientifique ou sérieux des enquêtes : une enquête de quelques minutes pourrait être considérée comme peu utile ou importante et ne valant pas la peine d'être répondue. De plus, ce sont souvent les enquêtes électorales qui ne durent que quelques minutes. Il est donc également possible que la thématique des enquêtes joue un rôle ou que ces associations brutes soient biaisées par le fait que les variables de sensibilité, durée et période ne sont pas indépendantes.

Une analyse de variance-covariance montre que dans un modèle contrôlant pour les durées de collecte et de questionnaire ainsi que le caractère sensible, la période de collecte jouent positivement sur la participation moyenne : une enquête dont la durée de collecte est supérieure à 32 jours a en moyenne une participation augmentée de 7.4% relativement aux enquêtes dont la durée de collecte est comprise entre 12 et 23 jours ; une durée de questionnaire de 30 minutes augmente la participation de 6.8% relativement à une durée comprise entre 5 et 15 minutes. En revanche, comme dans le cas bivarié, le caractère sensible de l'enquête ne joue pas significativement dans cette modélisation toutes choses égales par ailleurs ( $p=0.31$ ).

#### 4.2.2 L'attrition

A l'issue d'une collecte, cinq états sont possibles en réponse à l'invitation à participer adressée aux panélistes :

- le questionnaire a été fini ;
- le questionnaire a été commencé mais non terminé (soit car il a été abandonné en cours de route, soit toutes les questions ont été complétées mais le questionnaire n'a pas été validé après la dernière question) ;
- le panéliste n'a pas été invité (soit car il a signalé son absence à l'équipe ELIPSS, soit car sa tablette est cassée ou perdue) ;
- le panéliste n'a pas répondu ;
- le panéliste est définitivement sorti du panel à la date de l'enquête (attrition).

---

Équipements & Énergie, qui n'a été proposée qu'aux répondants de la partie 1 ; de même la partie 3, qui n'a été proposée qu'aux répondants de la partie 2.

<sup>16</sup> Chaque classe de période de collecte inclut un tiers des enquêtes analysées.

<sup>17</sup> Comme pour les périodes de collecte, chaque classe de durée de questionnaire comprend un tiers des enquêtes analysées.

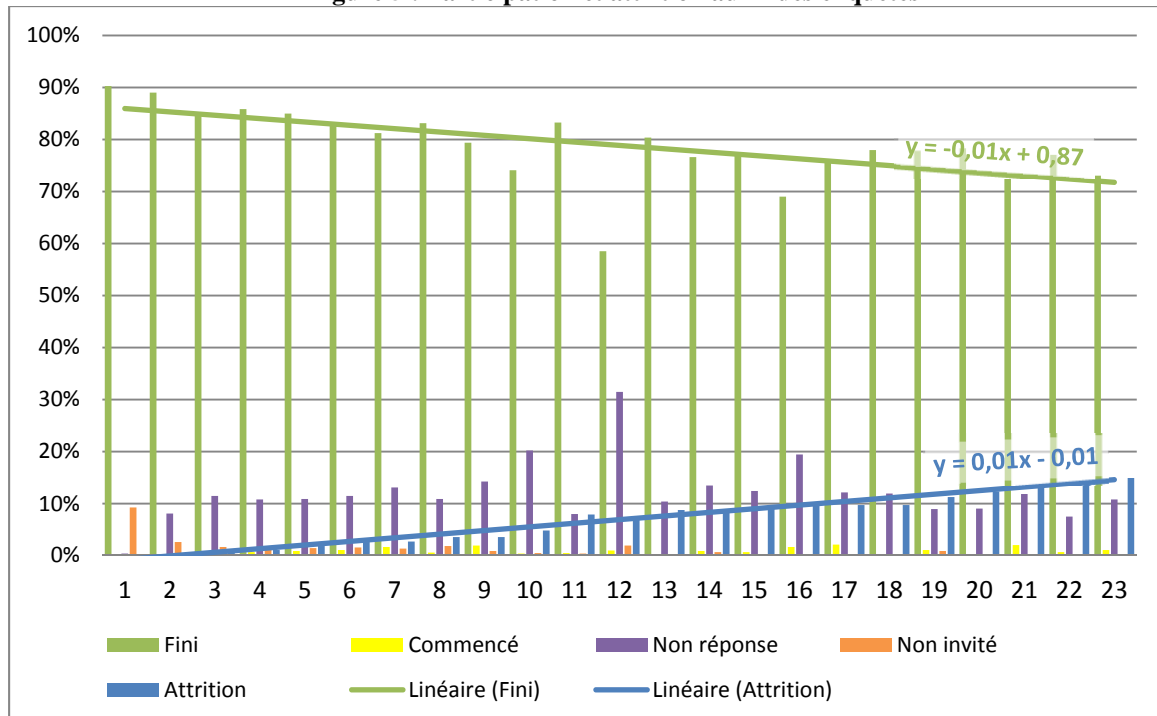
La Figure 4 montre que la participation (au sens de la proportion de répondants ayant fini leur questionnaire parmi les 1039 panélistes initiaux) diminue lentement mais régulièrement au cours du temps, d'environ 1% par enquête, avec toutefois des variations importantes suivant les enquêtes.

La proportion d'enquêtés n'ayant pas validé leur questionnaire ou l'ayant abandonné en cours (série « commencé ») est très faible, de même que celle des personnes non invitées. En revanche, la proportion de non-réponses est plus élevée (environ 10% des panélistes) et relativement stable dans le temps sur le long terme.

L'attrition, c'est-à-dire la sortie du panel<sup>18</sup>, est quant à elle croissante. Au final, 170 panélistes sont sortis du panel après 28 enquêtes, soit un taux d'environ 16% en deux ans et demi. Il s'agit dans 55% des cas d'exclusion (n=94) et dans 45% des cas de départs volontaires (n=71) ou de décès (n=5).

Le taux de participation brut (calculé sur la base des 1039 panélistes initiaux) est de 78,4% sur les 23 premières enquêtes (en excluant les enquêtes techniques initiales proposées avant recrutement de tous les panélistes et les deux derniers volets de l'enquête Équipements & Énergie), tandis que le taux de réponse moyen effectif (calculé sur la base des seuls invités, cf. Tableau 6) est de 85,2%.

**Figure 5 : Participation et attrition au fil des enquêtes**



### 4.2.3 Les profils de participation

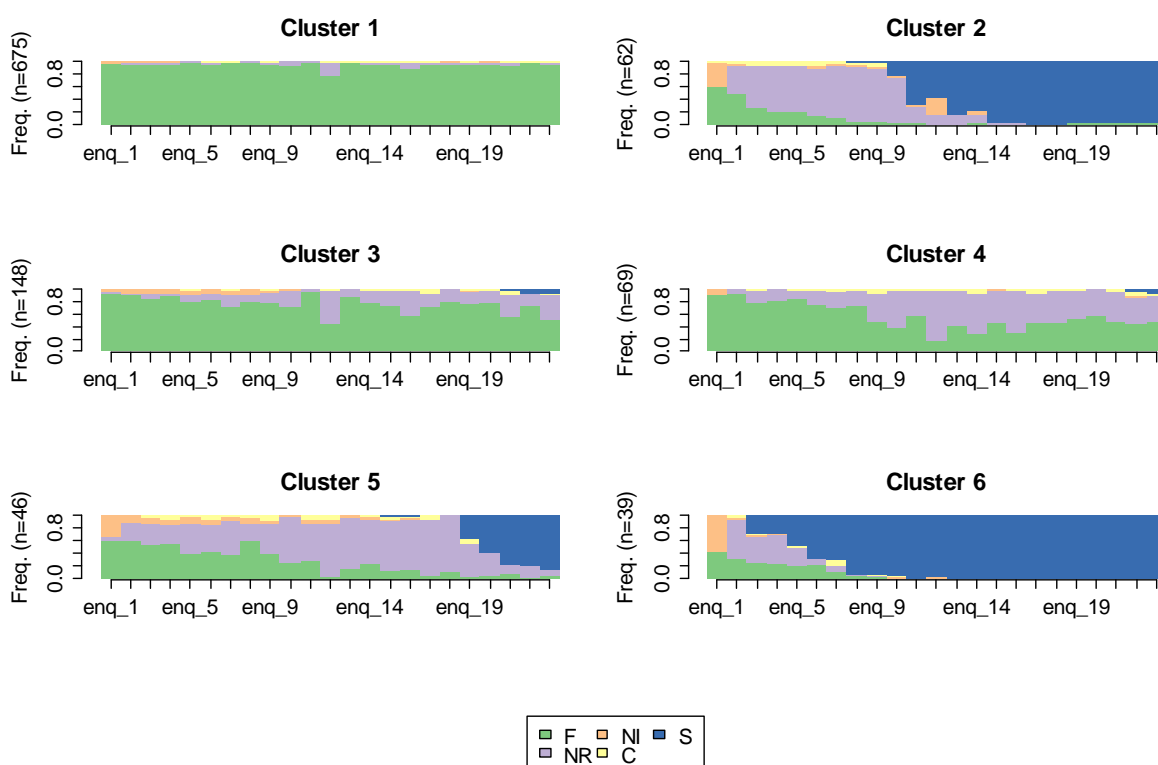
En distinguant les 5 états possibles de réponse à une enquête (questionnaire fini –F-, questionnaire commencé mais non validé –C-, non invité –NI-, non répondant –NR- et sorti du panel S), nous pouvons étudier les trajectoires de participation des panélistes et tenter de les regrouper en classes homogènes afin de décrire efficacement la diversité des comportements tout au long des enquêtes. Pour cela, la méthode de l'optimal matching a été mise en œuvre<sup>19</sup>.

Si nous considérons les 23 enquêtes analysées (cf. supra § 4.4), nous pouvons distinguer 6 groupes (ou clusters) de trajectoires de participation (Figure 6).

<sup>18</sup> La sortie de panel peut être à l'initiative du panéliste (sortie volontaire) ou à l'initiative de l'équipe ELIPSS (exclusion) dans le cas de non-réponse aux enquêtes pendant 5 mois consécutifs.

<sup>19</sup> Cela a été fait avec le logiciel R et la technique de l'optimal matching (package TramineR)

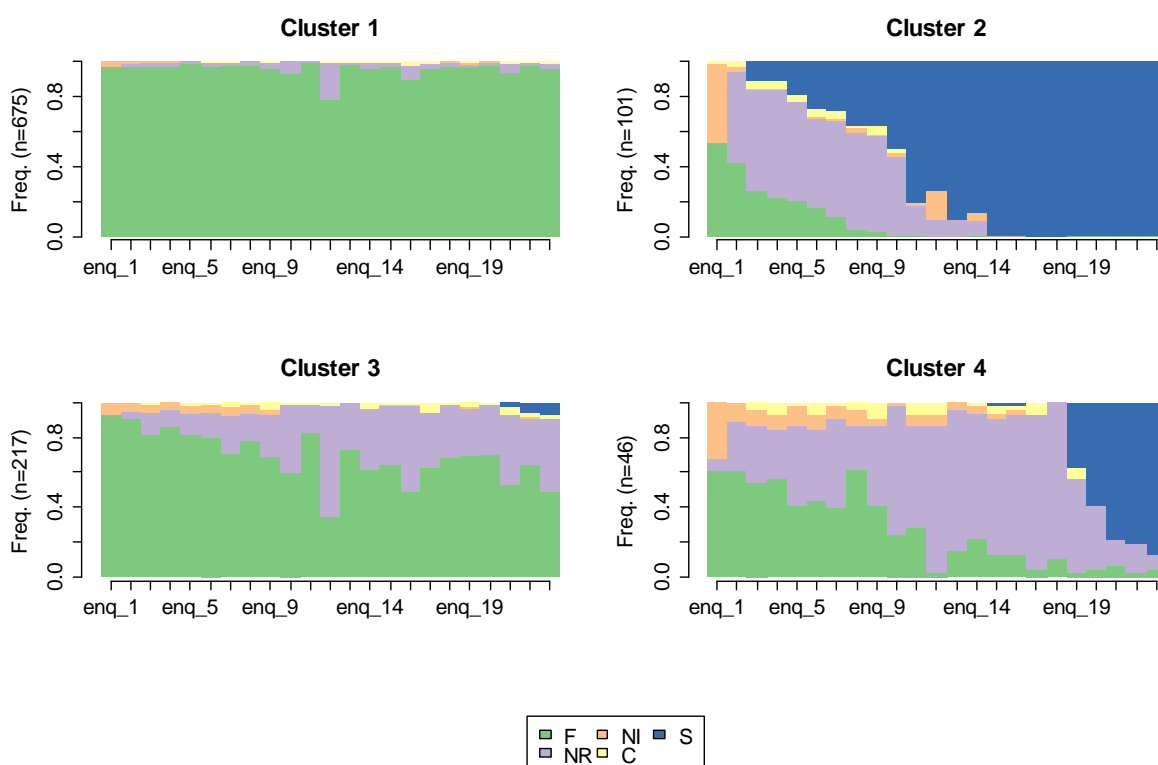
**Figure 6 : Groupes homogènes de trajectoires de participation à 23 enquêtes ELIPSS en 6 classes**



Le premier groupe (cluster 1) est composé de 65% des 1039 panélistes initiaux. Ce sont les fidèles : tous ont répondu à toutes les enquêtes ou presque. Les clusters 3 et 4 regroupent respectivement 14% et 7% des panélistes initiaux ; ce sont des panélistes à la participation non systématique mais plutôt élevée (cluster 3) ou moyenne (cluster 4), et dont la tendance est clairement à la baisse. Des sorties de panel commencent à apparaître dans ces deux groupes vers la 20<sup>ème</sup> enquête. Le cluster 2 (6% des panélistes), le cluster 5 (4%) et le cluster 6 (4%) distinguent trois groupes de panélistes sortis du panel, suivant la date de leur sortie. Le groupe 6 est celui des sorties précoces, le groupe 2 des sorties intermédiaires (dès la 10<sup>ème</sup> enquête environ) et le groupe 5 celui des sorties tardives (dès l'enquête 18 environ).

On peut résumer ce résultat en regroupant les clusters les plus proches. On obtient alors 4 groupes (Figure 7) : le cluster des fidèles, intact, les clusters 2 et 4 qui regroupent les sorties suivant la précocité de celle-ci et le cluster 3 qui regroupe les panélistes à la participation moyenne et déclinante. Les analyses suivantes seront réalisées à partir du résultat en 4 classes.

**Figure 7 : Groupes homogènes de trajectoires de participation à 23 enquêtes ELIPSS en 4 classes**



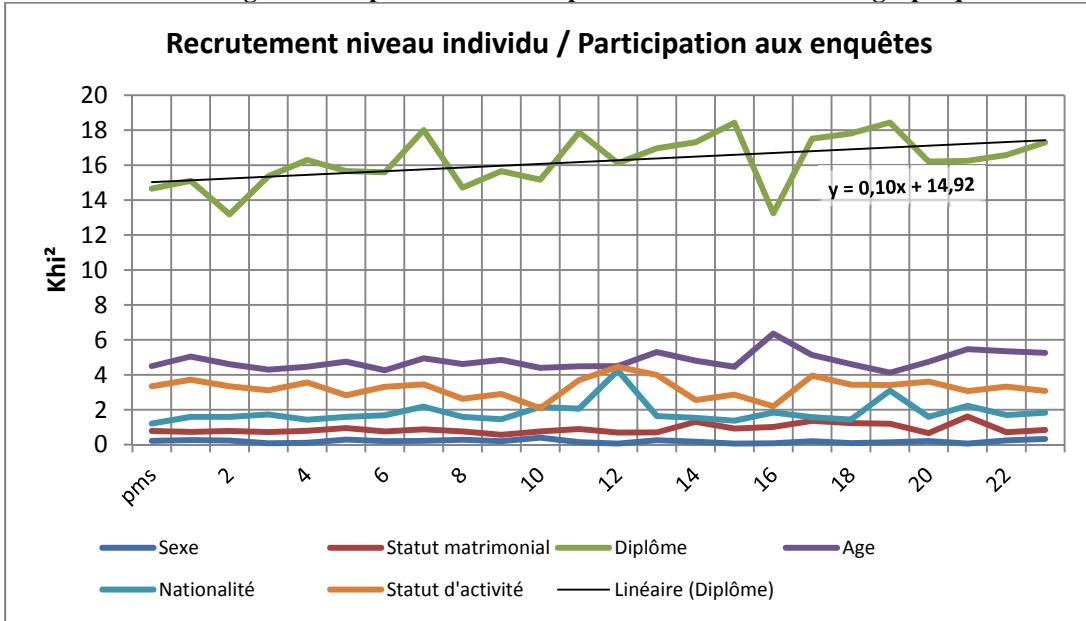
Une analyse des caractéristiques sociodémographiques des panélistes fidèles (cluster 1) montre que les panélistes encore présents à la dernière enquête (n=675) sont plus souvent des personnes de nationalité française que le reste des panélistes (92% vs 87%,  $p=0.007$ ), possèdent plus souvent un diplôme supérieur au Bac (64% vs 57%,  $p=0.032$ ), sont plus souvent mariées (46% vs 36%,  $p=0.001$ ) et sont aussi plus souvent âgés de 35 à 64 ans (69% vs 54%,  $p=0.001$ ). Par ailleurs, ces panélistes fidèles résident moins souvent en HLM (13% vs 21%,  $p=0.001$ ) ou en zone urbaine sensible –ZUS– (4% vs 8%,  $p=0.014$ ), et plus souvent dans des logements d’au moins 4 pièces (68% vs 62%,  $p=0.028$ ) ou 5 pièces (46% vs 35%,  $p=0.001$ ).

Les panélistes fidèles présentent donc un profil socioéconomique relativement favorisé par rapport aux autres.

#### **4.2.4 Evolution de la représentativité au fil des enquêtes**

Un aspect important de la qualité statistique du panel est la représentativité de l’échantillon des répondants au fil des enquêtes et de l’attrition. Il apparait que la représentativité sur le long terme est peu altérée par les sorties du panel : les écarts de distribution de quelques variables sociodémographiques par rapport au recensement sont plutôt stables. Seule la distribution des diplômes des panélistes évolue négativement, comme le montre la Figure 8.

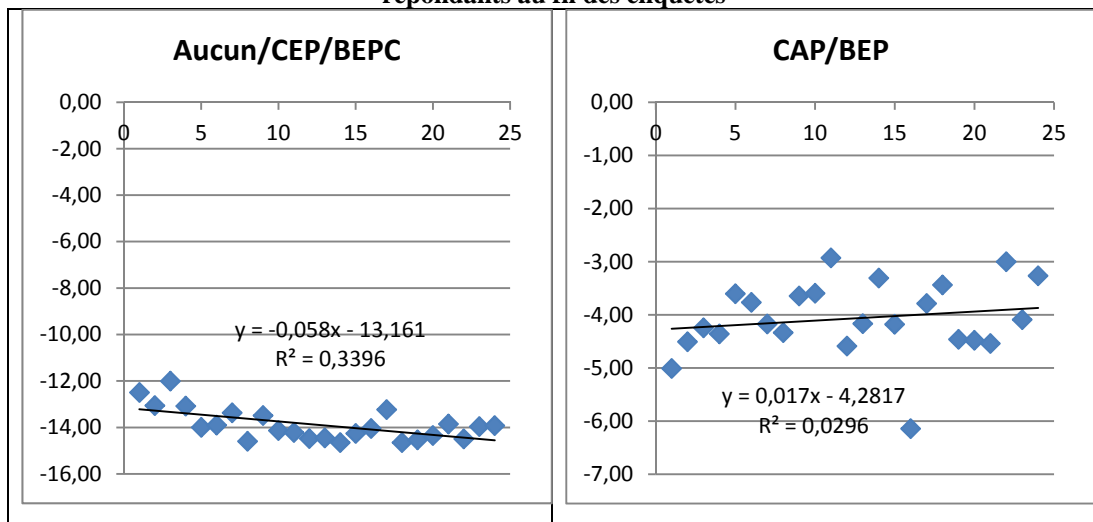
**Figure 8 : Représentativité du panel en termes sociodémographiques**



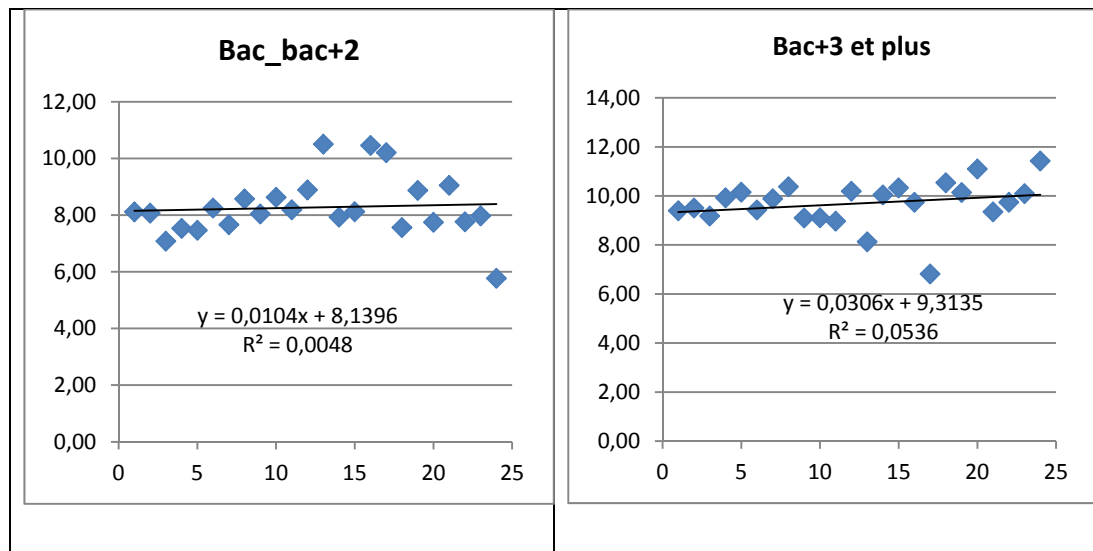
Note de lecture : PMS=Panel management system est l’outil de gestion du panel ; l’entrée définitive dans le panel débute par l’inscription sur ce outil. En abscisse, les enquêtes successives ; en ordonnée, la distance du KHI² normalisée par rapport au recensement, pour chaque variable, calculée pour les participants à chaque enquête.

Une analyse plus fine de la distribution des niveaux de diplôme indique que l’attrition et les non-réponses induisent surtout une perte des peu diplômés comme on peut l’observer à partir des graphiques de la Figure 9. Ces graphiques représentent l’écart en pourcentage des niveaux de diplôme des panélistes par rapport au recensement au fil des enquêtes.

**Figure 9 : Ecart des niveaux de diplôme entre le recensement de la population et les panélistes répondants au fil des enquêtes**







#### 4.2.5 Influence des caractéristiques individuelles et des attributs d'enquête sur la participation

Les analyses précédentes montrent à la fois l'influence des caractéristiques sociales des individus (âge, diplôme, etc.) et celle des caractéristiques des enquêtes telles qu'annoncées aux enquêtés (durée du questionnaire, thématique, durée de collecte) sur le taux de participation. Est-il possible de démêler l'influence de ces déterminants de niveau individuel et de niveau enquête ?

Pour cela, il faut procéder à une modélisation de type multiniveau. En effet, ELIPSS étant une cohorte, les panélistes sont invités à participer à toutes les enquêtes et les participations successives d'un même individu sont corrélées entre elles et donc non indépendantes. Une telle analyse permettra, en plus de fournir des éléments prédictifs de la participation des panélistes au long du panel, de mettre en place une gestion du panel plus efficace, par exemple en aidant à la définition de protocoles de relances des panélistes les plus susceptibles d'abandonner.

Au vu des écarts de comportements observés entre les groupes homogènes de trajectoires de participation représentés sur la Figure 7, il apparaît utile de distinguer les groupes avant d'entreprendre l'analyse. En effet, pour les panélistes fidèles, l'effet du temps est quasi-nul puisque leur participation est quasi-systématique. Pour les panélistes dont la participation est moins systématique, celle-ci diminue linéairement avec le temps. Ainsi, une mesure de l'effet du temps sur la participation calculée sur l'ensemble des panélistes serait tout à fait trompeuse et insuffisante pour décrire des comportements de participation très contrastés. Nous distinguerons ainsi 3 groupes : le groupe des fidèles (n=675), le groupe 3 des participants moyens et déclinants (n=217) et les groupes 2 et 4 des panélistes sortis suivant leur date de sortie du panel (n=101 et n=46).

Nous pouvons raisonnablement formuler les hypothèses suivantes quant aux résultats :

1. Les effets des variables sociodémographiques individuelles et des caractéristiques des enquêtes sont plus réduits pour le groupe des fidèles
2. Les effets des caractéristiques des enquêtes sont plus marqués pour les autres groupes

L'hypothèse 1 repose sur le fait que la trajectoire moyenne des panélistes fidèles est presque exclusivement étalée : aucune variation importante n'y est perceptible (voir le cluster 1 de la Figure 7). Par conséquent, tous les panélistes participent systématiquement et leurs caractéristiques individuelles, de même que les caractéristiques des enquêtes ne jouent pas. L'hypothèse 2 repose sur la même approche appliquée aux autres groupes : la participation n'étant pas systématique, il est

fort probable que les caractéristiques des individus et des enquêtes jouent davantage sur la participation.

Nous retiendrons l'âge et le diplôme comme caractéristiques sociodémographiques individuelles<sup>20</sup> ; l'appartenance au groupe des fidèles à la cinquième enquête, la durée du questionnaire, la durée de la collecte et le caractère sensible des enquêtes comme attributs de niveau enquête.

Dans ces modèles, l'unité statistique est l'enquête, et la variable modélisée est la participation à l'enquête. Or, les participations d'un même individu aux enquêtes successives ne sont pas indépendantes les unes des autres, puisqu'elles sont liées au profil et aux motivations de l'individu. Autrement dit, les participations successives des individus sont corrélées entre elles. Le modèle multiniveau permet de tenir compte de cette non-indépendance des observations au fil du temps. Les variables dites à effets fixes sont les caractéristiques des enquêtes (durée, période, thématique sensible) et les caractéristiques individuelles des panélistes (âge, diplôme). Chaque individu est considéré comme ayant un niveau moyen de participation sur les 23 enquêtes qui lui est propre, mais la participation varie avec le temps. Le temps est pris en compte par l'introduction d'une variable de compteur d'enquête (de 1 à 23) dont l'effet sur la participation est linéaire (le taux de participation décline de façon continue avec le nombre d'enquêtes) mais varie d'un individu à un autre. La moyenne de participation et le temps sont des variables aléatoires<sup>21</sup>.

Les résultats pour les trois groupes d'individus sont présentés dans le Tableau 8 ci-dessous.

**Tableau 8 : Modélisations multiniveaux de la participation aux 23 enquêtes suivant le groupe de trajectoire d'appartenance (OR et significativité)**

Effect	Fidèles n=675		Participation déclinante n=217		Sortie progressive n=147	
	OR	Pr >  t	OR	Pr >  t	OR	Pr >  t
Temps (+1 enquête)	1.00	0.8076	<b>0.90</b>	<.0001	<b>0.52</b>	<.0001
Age (18-24)	0.64	0.1315	<b>0.46</b>	0.0036	2.41	0.1665
Age (25-34)	0.75	0.244	<b>0.54</b>	0.0119	<b>4.90</b>	0.0086
Age (35-44)	1.01	0.9503	<b>0.58</b>	0.0234	<b>3.32</b>	0.0442
Age (45-54)	1.01	0.9524	0.69	0.1289	1.83	0.3208
Age (55-64)	1.28	0.3007	0.63	0.1056	1.14	0.8413
Age 65-75 (ref.)	1.00	.	1.00	.	1.00	.
Dipl (<BEP)	1.16	0.4557	0.85	0.3255	0.54	0.1995
Dipl (<Bac)	1.36	0.0822	1.15	0.3076	<b>0.37</b>	0.0552
Dipl (Bac-Bac+2)	1.27	0.1064	<b>1.27</b>	0.0579	0.63	0.3422
Dipl (≥Bac+3) (ref.)	1.00	.	1.00	.	1.00	.
Thème sensible	1.06	0.5858	0.91	0.2607	1.08	0.696
Non fidèle au temps 5	<b>0.09</b>	<.0001	<b>0.63</b>	0.0144	<b>0.01</b>	<.0001
Durée du questionnaire						
3 <sup>ème</sup> tiers =30 mn	<b>2.16</b>	<.0001	0.89	0.3183	1.60	0.0757

<sup>20</sup> D'autres variables ont été testées : sexe, nationalité, mode de recrutement, logement en HLM. Leur effet est très rarement significatif, leur présence diminue la qualité du modèle et n'apporte que très peu d'information.

<sup>21</sup> La modélisation a été menée avec la procédure GLIMMIX de SAS V9.4.

2 <sup>ème</sup> tiers =20-25 mn	<b>1.89</b>	<.0001	<b>1.41</b>	<.0001	1.05	0.8456
1 <sup>er</sup> tiers =5-15 mn (ref.)	1.00	.	1.00	.	1.00	.
Période de collecte						
3 <sup>ème</sup> tiers =32-100 jours	<b>3.93</b>	<.0001	<b>2.53</b>	<.0001	0.94	0.7898
2 <sup>ème</sup> tiers= 24-31 jours	<b>4.23</b>	<.0001	<b>2.19</b>	<.0001	1.06	0.8167
1 <sup>er</sup> tiers =12-23 jours (ref.)	1.00	.	1.00	.	1.00	.

Note : OR=odds ratio. Pour les trois modèles, les tests de différences de moyenne de participation, d'évolution de la participation avec le temps sont significatifs ( $p < 0.0001$ ) ; les groupes sont repérables sur la Figure 7 : cluster 1, cluster 3 et clusters 2&4.

Les résultats montrent que globalement, l'hypothèse 1 est vérifiée : l'âge est significatif pour les groupes de non-fidèles, mais pas pour le groupe des fidèles. Pour les panélistes à la participation déclinante, relativement aux personnes âgées de 65-75 ans, les jeunes participent en moyenne moins ( $OR < 1$ ). Pour les panélistes sortants, les panélistes de 25-44 ans participent au contraire un peu plus que les 65-75 ans ( $OR > 0$ ). En revanche, le résultat pour le diplôme est moins convaincant, bien que la tendance soit similaire : l'effet du diplôme est plus proche du seuil de significativité dans les groupes qui ne sont pas fidèles (la tendance étant à une moindre participation parmi les moins diplômés) alors qu'il n'y a pas d'effet significatif du diplôme parmi le groupe des fidèles.

L'hypothèse 2 est finalement réfutée : la durée du questionnaire et la longueur de la période de collecte jouent plus souvent et plus fortement au sein du groupe de fidèles. L'effet est même nul dans le groupe des sortants, ce qui dans leur cas peut sans doute s'expliquer par leur comportement général d'indifférence au panel et à ses enquêtes.

La longueur de la période de collecte a un effet supérieur à celui de la longueur du questionnaire, ce qui ne surprend pas, dans la mesure où l'allongement de la période de collecte est plus important et offre plus d'opportunités de participer. En détail, on retrouve qu'une durée de questionnaire élevée accroît les chances de participer parmi les fidèles ( $OR > 2$  pour une durée de 30 minutes vs. 5-15 minutes) ; parmi les panélistes à la participation déclinante, l'effet est sensible pour un temps de réponse entre 20 et 25 minutes relativement à une durée inférieure à 15 minutes ( $OR = 1.4$ ). Parmi les sortants, une durée extrême a tendance à augmenter les chances de participer relativement à une durée courte ( $OR = 1.6$ ), mais l'effet est non significatif.

Enfin, le caractère sensible de la thématique des enquêtes ne joue quant à lui clairement jamais : les OR ne sont jamais significativement différents de 1.

Parallèlement, l'effet du temps (la succession des enquêtes) a un effet prononcé dans le groupe des participants moyens dont la participation décline ( $OR = 0.90$ ) et nettement plus marqué dans le groupe des sortants progressifs ( $OR = 0.52$ ), alors qu'il est nul dans le groupe des fidèles ( $OR = 1.00$ ), ce qui était attendu compte tenu des trajectoires de participation de ces groupes de panélistes (cf. Figure 7).

### 4.3 Analyse des premières impressions des participants au panel

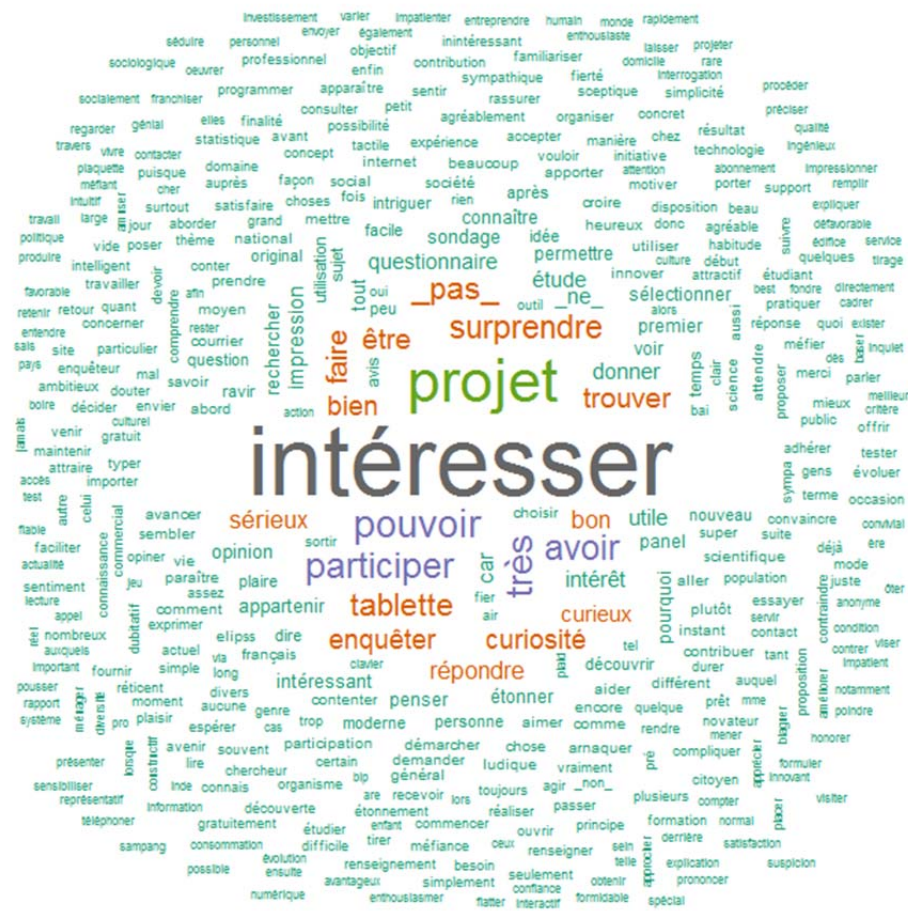
La première enquête a été consacrée au didacticiel et aux pratiques numériques des panélistes. Cette enquête a permis, entre autres, de collecter des informations sur la façon dont le projet a été perçu par les personnes sélectionnées, grâce notamment à la question ouverte suivante : « Pourriez-vous, en quelques mots, décrire vos premières impressions quand vous avez découvert le projet ELIPSS ? ».

Nous proposons dans cette section, d’abord, une analyse textuelle de cette question en supposant qu’elle reflète la motivation première des gens à faire partie du panel, ensuite, de tester et de décrire le cas échéant la relation entre cette motivation et la participation au fil des enquête.

Environ 90% des 1039 panélistes ont participé à cette enquête (N=938). Seuls trois panélistes parmi les répondants n’ont pas répondu à cette question. Un important travail d’apurement et de correction orthographique et syntaxique a été nécessaire avant l’analyse, de même qu’une lemmatisation. Initialement, les commentaires varient entre 2 caractères et 255 caractères (espaces compris, dont 49 mots contenus dans 2 phrases), et en moyenne les réponses contiennent 53 caractères. Les commentaires les plus courts sont les « OK », les « bon », et même des émoticônes.

Voici un aperçu général des commentaires laissés par les panélistes<sup>22</sup>.

**Figure 10: Nuage des mots cités par les panélistes suivant leur fréquence d’apparition (données lemmatisées non pondérées)**



La couleur et la taille des mots renseignent sur la fréquence de leurs occurrences dans la totalité des réponses. Le mot « intéresser » est le plus fréquent (310 occurrences) suivi du mot « projet ». Ces deux lemmes sont par ailleurs redoublés de lemmes sémantiquement proches, comme « intéressant », « intérêt », « étude », « panel », « enquêter »... Ce classement reflète une approche positive au projet et de sa démarche dans son ensemble. Il y a évidemment un biais de sélection très fort, puisque seuls les participants au panel ont répondu, donc les personnes ayant accepté et

<sup>22</sup> Les négations « ne » et « pas » sont conservées sous la forme « \_ne\_ » et « \_pas\_ » afin de contribuer à la bonne interprétation du sens des propos dans les analyses suivantes.

compris le projet, mais il est tout aussi clair que la dimension novatrice du projet ainsi que le matériel et la technologie utilisés comme support d'enquête ont eu un effet important, et notamment provoqué la surprise des panélistes. On notera que le mot « tablette » est parmi les plus fréquents : ce dispositif technique a pu séduire et surprendre par son originalité et sa modernité comme support de collecte mais aussi plus prosaïquement comme incitation matérielle et financière à participer. La dimension scientifique est également présente, mais elle est diluée dans un plus grand nombre de lemmes : « chercheur », « scientifique », « connaissance », « étudier », « panel », « étude », « science ».

La Figure 10 ci-dessus montre un résumé très général des réponses mais ne permet pas de juger le sens et la liaison entre les mots. Une analyse plus poussée est nécessaire. Pour ce faire, les mots rares (dont les occurrences sont peu nombreuses dans le corpus analysé, et dans notre cas le seuil est fixé avec un taux<sup>23</sup> inférieur ou égal à 99%) sont supprimés pour des questions de simplicité. La diversité lemmatique étant très grande, 10% des mots seulement ont été conservés. La Figure 11 montre le positionnement multidimensionnel (ou « multidimensional scaling » (MDS) en anglais) de ces mots les plus fréquents sur un plan à deux dimensions selon leur similarité. Pour faciliter l'interprétation, nous avons ajouté les résultats d'une classification en Kmeans des termes sur la figure, ces résultats sont présentés en 6 couleurs ce qui correspond aux 6 classes de la classification.

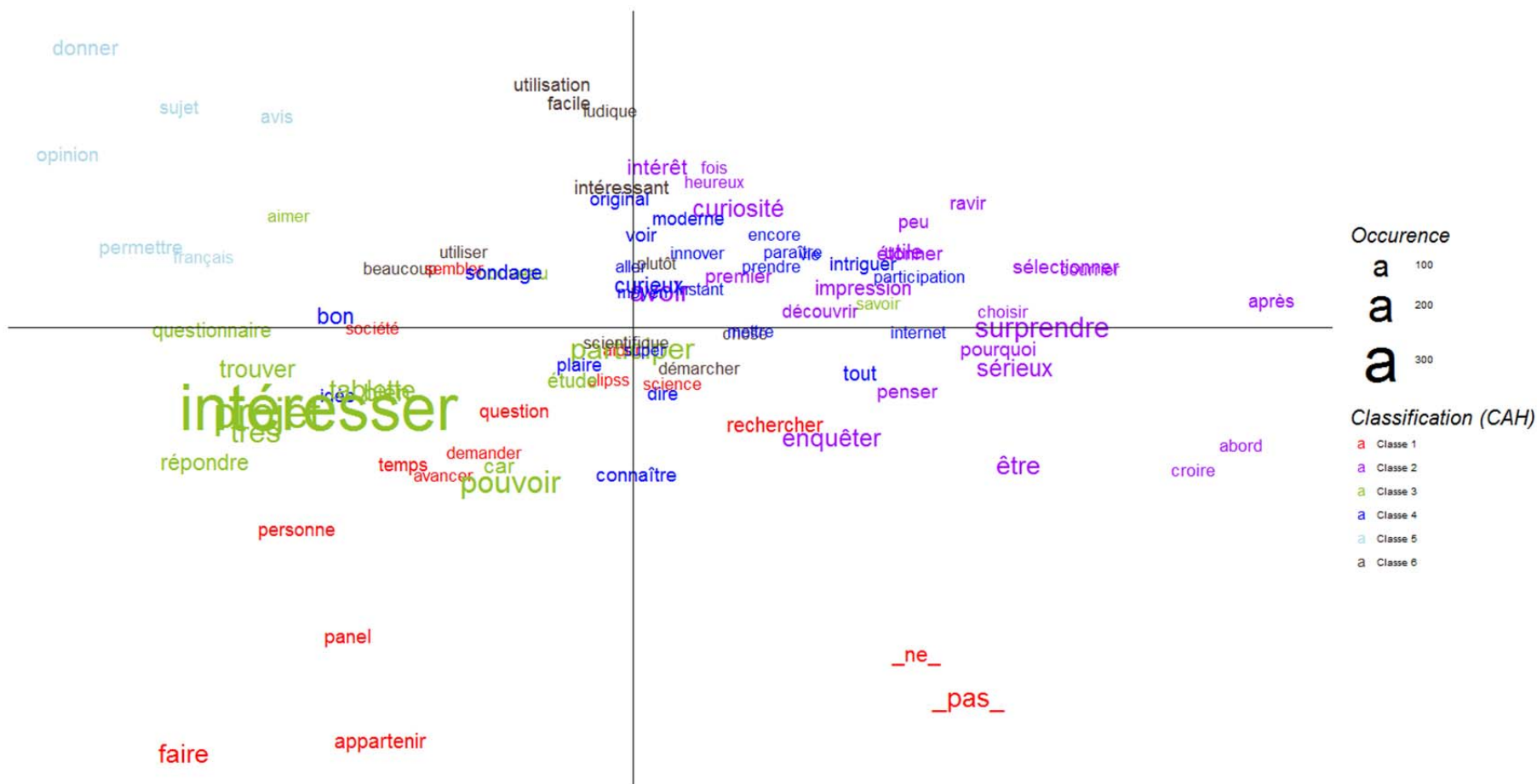
L'axe horizontal est la première dimension du MDS (2% de l'inertie totale). C'est celui opposant les personnes valorisant fortement le projet à celles qui ont été agréablement surprises par la démarche. Il sépare, d'un côté (à gauche), les gens qui sont intéressés par le panel et sont positifs vis-à-vis du projet, et de l'autre (à droite), les gens ravis d'être sélectionnés, ceux qui étaient plus hésitants et intrigués mais que le sérieux du projet a convaincu. L'axe vertical représente la seconde dimension (2% de l'inertie totale), celle de la curiosité et de la valorisation de l'utilisation de la tablette comme objet ludique et technique. Cet axe décrit moins un regard porté sur la dimension scientifique et originale du projet qu'un intérêt plus matériel pour la tablette et ses utilisations. En haut, on trouve les personnes qui trouvent de l'intérêt à participer parce que la tablette leur semble un outil moderne et ludique qui a suscité leur curiosité. En bas, on trouve les personnes se déclarant contentes de faire partie du panel ou qui pensent que participer au panel leur offrira la possibilité de faire ou de dire des choses qui pourraient être utiles ; ces personnes n'ont pas vraiment exprimé de façon précise leur intérêt ou leur point de vue vis-à-vis du projet, à part faire partie du panel. Leur motivation est moins grande que celle des autres, tant sur le plan du matériel que des objectifs scientifiques d'ELIPSS<sup>24</sup>.

---

<sup>23</sup> Ce taux est appelé « sparse percentage ». Dans l'analyse textuelle, il est défini comme étant l'inverse de l'espace occupé (ou densité) d'un mot dans une matrice terme-document, par rapport aux autres mots. Plus la fréquence du mot est faible, plus ce taux est grand pour le mot.

<sup>24</sup> Une analyse plus poussée sur ces commentaires est en cours pour étudier le lien entre ces premières impressions et l'évolution de la participation des panélistes au fil du temps. Cette analyse utilise les résultats de l'«optimal matching» en 4 classes et ceux de l'analyse textuelle.

Figure 11: Positionnement multidimensionnel des lemmes les plus courants du corpus cité par les panélistes



Note de lecture : Les couleurs représentent les classes de la classification Kmeans et la taille des mots représente leur occurrence.

#### 4.4 Qualité des réponses aux enquêtes annuelles

Le contrôle de la qualité des réponses a été réalisé sur trois variables présentes dans les enquêtes annuelles et utilisées lors du redressement des enquêtes : la nationalité, le diplôme et la date de naissance.

La définition des incohérences est propre à chaque caractéristique. Dans tous les cas, les raisons des incohérences peuvent être diverses, il peut y avoir un changement de situation, une mauvaise déclaration, comme un changement de répondant.

Pour la nationalité, il y a incohérence si les situations suivantes apparaissent: (1) si la personne avait déclarée qu'elle est de nationalité étrangère ou avait acquis la nationalité française, puis déclare après qu'elle est française de naissance ; (2) si la personne avait déclarée qu'elle est française de naissance, et après qu'elle est de nationalité étrangère ou de nationalité française par acquisition ; (3) si la personne avait déclarée qu'elle a acquis la nationalité française et après elle déclare qu'elle est de nationalité étrangère.

Pour le diplôme, l'incohérence est définie comme le fait de déclarer un diplôme antérieur plus élevé par rapport au diplôme actuellement déclaré. En effet, la question posée porte sur le diplôme le plus élevé obtenu. Dans ce cas, même si le dernier diplôme obtenu est moins élevé que celui d'avant, c'est toujours le diplôme le plus élevé qui doit être déclaré. Un autre cas d'incohérence, rare mais possible, peut être lié au délai d'obtention entre deux diplômes, par exemple, si la personne déclare avoir le BEPC une année et déclare un bac+3 l'année suivante.

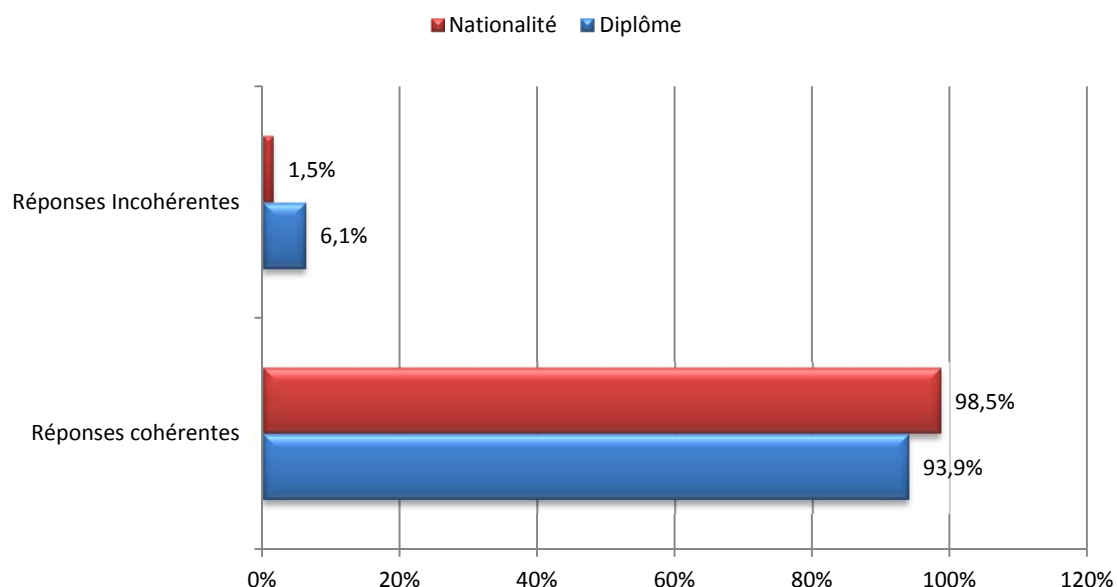
Pour la date de naissance, l'incohérence est simple à définir puisque cette information ne peut pas changer d'une année sur l'autre. Tout changement de date de naissance peut donc être interprété comme une réponse incohérente.

Pour résumer, sont considérées comme cohérentes les réponses qui n'ont pas changé ou qui ont évolué de façon logique au cours d'au moins deux des trois enquêtes. Sont considérées comme incohérentes les réponses qui présentent au moins une différence non logique entre les trois enquêtes annuelles. Cette analyse de la qualité des réponses ne peut évidemment pas être menée pour les 15% du panel n'ayant répondu à moins d'une enquête annuelle sur les trois administrées depuis 2013. Les résultats suivants sont donc établis sur la base des panélistes ayant complété au moins deux enquêtes annuelles sur trois.

Les déclarations sont plus souvent incohérentes pour le diplôme que pour la nationalité (cf. Figure 12). Deux raisons peuvent l'expliquer, d'abord parce que la question sur la nationalité française ou étrangère est plus simple avec seulement 3 ou 6 modalités contre 11 modalités pour le diplôme. Ensuite, la compréhension de la question sur le diplôme peut donner lieu à des interprétations plus diverses. Des répondants peuvent par exemple anticiper l'obtention d'un diplôme en cours de préparation, qu'ils n'obtiendront finalement pas ; d'autres peuvent considérer le niveau maximum atteint même s'ils n'ont pas obtenu le diplôme correspondant ; d'autres enfin peuvent répondre pour le dernier diplôme obtenu au lieu de celui ayant le niveau le plus élevé.

Il est clair que les réponses aux enquêtes annuelles sont de bonne qualité avec 94% de réponses cohérentes pour le diplôme et même 98% pour la nationalité pour les panélistes ayant répondu à au moins deux enquêtes annuelles. En outre, d'après les réponses en 2015, il semble que 70% des incohérences observées entre 2013 et 2014 sont en réalité des mauvaises déclarations en 2013.

**Figure 12 : Qualité des réponses dans les enquêtes annuelles 2013/2014/2015 pour la nationalité et le diplôme**



Dans chaque enquête annuelle, la date de naissance du panéliste est demandée comme un moyen de vérifier l'identité du répondant. Elle peut ainsi permettre un contrôle de cohérence des réponses des panélistes et fournir un indicateur général de la qualité des réponses des panélistes. Dans l'enquête annuelle 2014, il y a 19 cas d'anomalies relatives à la date de naissance. Dans tous ces cas, c'est la date du jour du remplissage du questionnaire qui a été enregistrée par défaut. Globalement, 88,2% des déclarations de date de naissance (n=850) sont cohérentes (c'est-à-dire identiques dans toutes les enquêtes complétées) tandis que 2,7% des réponses (n = 26) présentent au moins une différence entre les dates de naissance renseignées, sachant que les 9,1% restants (n=88) sont des réponses uniques (c'est-à-dire à une seule enquête annuelle sur les trois). Pour ces dernières, on peut évaluer la qualité des réponses par rapport à la date de naissance déclarée au moment du recrutement et dans 99% des cas, la date déclarée dans l'enquête annuelle est bien celle enregistrée lors de l'entrée dans le panel. Les données sur la date de naissance sont donc de bonne qualité, ce qui est rassurant sur l'identité des répondants.

## 5 Les premiers résultats du pilote

L'objectif principal du panel ELIPSS est la collecte de données et leur mise à disposition, à destination des équipes de recherche conceptrices des enquêtes et plus largement de la communauté scientifique. Les résultats consistent en partie dans la documentation des données et leur diffusion sur le portail du Réseau Quetelet. Par ailleurs, outre la valorisation du dispositif réalisée par l'équipe en charge du panel ELIPSS, les principales publications seront celles des équipes de recherche à l'origine des enquêtes administrées au panel ELIPSS et des travaux menés par les chercheurs habilités à réutiliser les données dans le cadre d'une demande par le portail du Réseau Quetelet.

### 5.1 La diffusion des données

Le Centre de données socio-politiques de Sciences Po, qui est l'un des trois centres de données français en sciences sociales, est responsable de la documentation et de la diffusion des données produites dans le cadre du panel. Les enquêtes sont documentées selon la norme internationale Data Documentation Initiative et diffusées par internet sur le portail sécurisé du réseau Quetelet.



Une fois la période d'exclusivité d'un an pour les équipes co-productrices des enquêtes, celles-ci sont répertoriées dans le catalogue d'enquêtes du portail du Réseau Quetelet (réseau français des centres de données en sciences sociales). Ainsi, les fichiers de données sont accessibles gratuitement pour les chercheurs français et étrangers, les doctorants, les post-doctorants et les étudiants de master dans le cadre exclusif d'un projet de recherche.

Pour obtenir les fichiers de données, les demandeurs doivent signer une convention d'utilisation dans laquelle il s'engage notamment à respecter la confidentialité des répondants, à ne pas rediffuser des données à un tiers et à citer la source des données dans les publications.

Le premier jeu de données issu du panel ELIPSS (issues de l'enquête annuelle 2013) a été mis à disposition sur le portail du réseau Quetelet en mars 2015.

Pour garantir la confidentialité des données, il est impossible d'apparier toutes les données individuelles issues du panel entre elles. Seul le module signalétique de l'enquête annuelle ELIPSS est apparié de manière systématique à chaque fichier d'enquête. Toute demande d'appariement de données provenant de plusieurs enquêtes (en dehors des enquêtes longitudinales) est strictement encadrée et est soumise à l'examen du comité scientifique et technique et du CIL du CNRS.

Par ailleurs, l'appariement des informations du panel avec des données extérieures (fiscales, santé, etc.) est exclu.

## 5.2 La valorisation du dispositif

Plusieurs travaux ont été menés par le CDSP et l'INED sur le pilote du panel ELIPSS, essentiellement sous forme de rapports, de présentations lors de séminaires (Société française de statistique, INED, université de Lausanne) ou de conférences (European Survey Research Association, European Congress of Methodology, WAPOR, colloque francophone des sondages, journées de méthodologie statistique de l'INSEE, Webdatanet) et de deux articles dans *Social Science Computer Review* (voir listes ci-dessous). La plupart des travaux menés par ou en collaboration avec l'équipe ELIPSS sont essentiellement de nature méthodologique, par exemple sur la représentativité de l'échantillon, l'évolution de la participation des panélistes au fil des enquêtes ou le design d'enquêtes sur tablettes. Ces travaux ont vocation à être poursuivis et approfondis lors du développement du panel à partir de 2016.

### 5.2.1 Publications

Annelies Blom, Michael Bosnjak, Anne Cornilleau, Anne-Sophie Cousteaux, Marcel Das, Salima Douhou, Ulrich Krieger, " 2015 – A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe", *Social Science Computer Review* (online)

Mélanie Revilla, Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Pablo de Pedreza, 2015 – "What is the gain in a probability based online panel of providing internet access to sampling units who previously had no access?", *Social Science Computer Review* (online)

Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Nirintsoa Razakamanana et l'équipe ELIPSS, 2014 – *Le recrutement du panel ELIPSS*, Bilan du projet réalisé avec le soutien financier de l'INPES dans le cadre de la subvention n°060/12-DAS, juin 2014, 38 p.

## 5.2.2 Communications

Emmanuelle Duwez, Matthieu Olivier, 2015 – "Adapting survey questionnaires to touch-screen tablets. The ELIPSS Panel example", 6th Conference of the European Survey Research Association (ESRA), Reykjavik, Islande, 13-17 juillet

Mélanie Revilla, Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Pablo de Pedreza, 2015 – "What is the gain in a probability-based online panel to provide Internet access to sampling units that did not have access before?", 6th Conference of the European Survey Research Association (ESRA), Reykjavik, Islande, 13-17 juillet

Stéphane Legleye, Nirintsoa Razakamanana, Anne Cornilleau, Anne-Sophie Cousteaux, 2015 – "What drives the participation in a monthly research web panel? The experience of ELIPSS, a French random web panel in the general population", 6th Conference of the European Survey Research Association (ESRA), Reykjavik, Islande, 13-17 juillet

Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Nirintsoa Razakamanana, 2015 – "Setting-up a probability-based web panel. Lessons learned from the ELIPSS Pilot Study", 6th Conference of the European Survey Research Association (ESRA), Reykjavik, Islande, 13-17 juillet

Anne Cornilleau, Anne-Sophie Cousteaux, 2015 – "La mise en place d'un panel internet pour la recherche en sciences sociales", 6e Congrès de l'Association française de sociologie, 29 juin-2 juillet

Anne Cornilleau, Anne-Sophie Cousteaux, 2015 – "Challenges and opportunities in a probability-based online panel", Webdatanet Conference, Salamanque, Espagne, 26-28 mai

Stéphane Legleye, Anne Cornilleau, Anne-Sophie Cousteaux, Nirintsoa Razakamanana, 2015 – "La mise en place d'un panel représentatif connecté à Internet mobile : bilan du pilote Elipss", Journée de méthodologie statistique de l'INSEE, 31 mars-2 avril

Anne Cornilleau, Anne-Sophie Cousteaux, 2015 – "Recruiting and maintaining a probability-based internet panel in France : the ELIPSS pilot study", Rencontres méthodes et recherche, FORS-UNIL, Lausanne, Suisse, 24 février

Stéphane Legleye, Anne Cornilleau, Anne-Sophie Cousteaux, Nirintsoa Razakamanana, 2014 – "ELIPSS : un dispositif longitudinal représentatif pour la recherche en sciences sociales", Lundis de l'INED, Paris, 1er décembre

Stéphane Legleye, Anne Cornilleau, Anne-Sophie Cousteaux, Nirintsoa Razakamanana, 2014 – "ELIPSS : un dispositif longitudinal représentatif pour la recherche en sciences sociales", 8e colloque francophone sur les sondages, Dijon, 18-20 novembre

Anne Cornilleau, Anne-Sophie Cousteaux, 2014 – "Challenges in using tablet in a probability-based online panel", 67th Annual Conference of the World Association for Public Opinion Research (WAPOR), Nice, 4-6 september

Anne Cornilleau, Anne-Sophie Cousteaux, 2014 – "Challenges in using tablet in a probability-based online panel", 6th European Congress of Methodology, Utrecht, Netherlands, 23-25 July

Anne Cornilleau, Florent Gougou, 2013 – "Le dispositif pilote du panel ELIPSS et l'enquête DYNAMOB", séminaire SFDS-AFSP, 21 novembre

Anne Cornilleau, 2013 – "Recruitment of a Probability-based Internet Panel in France. The ELIPSS Pilot Study", 5th Conference of the European Survey Research Association (ESRA), Ljubjana, Slovenia, 15-19 July

Anne Cornilleau, 2013 – "ELIPSS. The French Example", Workshop Longitudinal Research in Internet Panel, Mannheim, Germany, 23-24 May

Laurent Lesnard, 2011 – "ELIPSS: A New Mobile Web Panel for Social Scientists", MESS Workshop, Oisterwijk, Netherlands, August

### 5.3 Les collaborations internationales

Dès le début, le projet a bénéficié d'échanges fructueux avec l'institut de recherche néerlandais CentERdata, responsable du LISS panel, notamment lors de trois workshop à Paris et à Tilburg en 2011 et 2012. En mettant à notre disposition leur système de gestion des panélistes, leur feuille de style et leur logiciel d'accès à distance aux données, cette collaboration a joué un rôle déterminant dans le démarrage du projet dès le début de l'année 2012.

Depuis, cette collaboration s'est élargie aux autres panels européens. D'abord avec les deux panels allemands : le GIP et le GESIS panel. Nous avons organisé un workshop à Paris en 2013 pour préparer le questionnaire d'une enquête conjointe. Cette collaboration entre le GESIS panel, le GIP, ELIPSS et le LISS panel s'est donc traduite par deux enquêtes comparatives en mai 2014 et en mai 2015 et par un article commun comparant la méthodologie des quatre panels intitulé « A Comparison of Four Probability-based Online and Mixed-Mode panels in Europe » et publié en 2015 dans un numéro spécial de *Social Science Computer Review*. En octobre 2015, nous organisons un workshop avec ces mêmes partenaires et des représentants des panels norvégiens et islandais dont l'objectif est précisément de consolider cette collaboration européenne par l'harmonisation et l'analyse conjointe des données comparatives produites en 2014 et 2015 et par du temps réservé sur les différents panels pour administrer régulièrement des enquêtes comparatives.

### Conclusion

La construction d'un dispositif tel que le panel ELIPSS recouvre des activités très diverses qui correspondent aux différentes étapes du cycle de vie des données, allant du projet d'enquête à la mise à disposition des données produites. En particulier, le pilote, destiné à tester la faisabilité d'un tel panel en France, a nécessité d'inventer l'inexistant. En premier lieu, le recrutement des panélistes a fait l'objet d'une procédure relativement complexe afin de déterminer la meilleure approche pour le développement du panel en 2016. Par ailleurs, l'originalité d'ELIPSS de fournir une tablette tactile et un abonnement 3G à chaque panéliste a nécessité la contractualisation avec un opérateur de téléphonie mobile, ce qui impliquait des aspects techniques, opérationnels, juridiques et financiers. Aussi, il a fallu développer en interne les outils informatiques puisqu'aucun outil existant ne remplissait l'ensemble des besoins du projet. Plusieurs de ces opérations, notamment le contrat avec l'opérateur de téléphonie, étaient des préalables indispensables à la construction du panel et sont maintenant définis pour le passage à 3500 panélistes.

Le pilote du panel ELIPSS a mis au jour des difficultés à prendre en compte et des pistes d'amélioration en vue du développement du panel en 2016. En premier lieu, l'utilisation d'une base de logements pour recruter les panélistes présente des inconvénients. Elle nécessite un important travail de nettoyage et de vérification des adresses. Elle implique aussi de décrire préalablement tous les membres du logement avant de sélectionner une personne pour l'inviter à participer au panel. Si ces contraintes sont communes à toutes les enquêtes menées à partir d'un échantillon d'adresses, le recrutement du pilote a rencontré des difficultés spécifiques à la procédure retenue et au timing du projet : le terrain qui s'est en partie déroulé pendant l'été ainsi que le retard pris du fait de la contractualisation avec l'opérateur ; l'invitation par courrier et les relances par téléphone qui ne se

sont pas avérées aussi efficaces qu'escompté ; la signature de la convention de participation qui ajoute une étape pour convaincre les personnes sélectionnées et formalise l'engagement à participer. Si la couverture réseau a été meilleure qu'attendue (aucune personne éligible n'a été exclue pour cette raison), la tablette a eu l'effet prévu en incitant les personnes à participer. Elle a aussi l'avantage de fournir un seul et même outil aux panélistes pour répondre aux questionnaires et dont l'aspect visuel est contrôlé.

Dans ce bilan, il convient également de souligner le rythme de travail soutenu que suppose la production d'une interrogation de 30 minutes par mois, en plus des activités de gestion de panel, de gestion de stock, des développements informatiques et de l'organisation du service auprès de la communauté académique.

En conclusion, il reste à tirer les enseignements du pilote pour continuer à construire et faire vivre le panel ELIPSS avec 3500 individus jusqu'en janvier 2019.

Nous avons testé différents modes de contact, notamment par souci financier, mais force est de constater que le face-à-face est la stratégie la plus efficace pour recruter des panélistes. Il semble donc difficile de faire l'économie du recours à un institut de sondage pour le recrutement prévu au début de l'année 2016. Il en va de la taille et de la qualité de l'échantillon. Il a donc été décidé de réaliser le recrutement en face-à-face uniquement. Afin de maximiser les chances de contact avec les ménages de l'échantillon, plusieurs méthodes seront mises en œuvre par l'institut, comme l'envoi de lettres d'annonce de la visite, de courriers aux mairies des communes présentes dans l'échantillon, les enquêteurs devront se rendre jusqu'à cinq fois à l'adresse, etc.

Ce nouveau recrutement bénéficiera évidemment de l'expérience acquise lors du pilote et d'une procédure plus simple et entièrement prise en charge par l'institut de sondage. De plus, nous avons renforcé l'équipe avec une personne chargée précisément de préparer et de suivre le terrain en face-à-face. Ainsi, il est possible que le taux de recrutement soit amélioré même si l'effet incitatif de la tablette a peut-être évolué depuis les trois dernières années.

## Bibliographie

- Blom A., Bosnjak M., Cornilleau A., Cousteaux A-S., Das M., Douhou S., Krieger U., " 2015 – A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe", *Social Science Computer Review* (online)
- Blom A., Gathmann C., Krieger U., *The German Internet Panel: Method and Results*, 2013
- Callegaro M., 2010 – Do You Know Which Device Your Respondent Has Used to Take Your Online Survey?, *Survey Practice*, vol.3, n°6  
(<http://www.surveypactice.org/index.php/SurveyPractice/article/view/250/html>)
- Callegaro M., DiSogra C., 2008 – Computing Response Metrics for Online Panels, *Public Opinion Quarterly*, vol.75, n°5, p.1008-1032
- Couper M., 2008 – *Designing effective web surveys*, New York: Cambridge University Press
- Das M., Ester P., Kaczmirek L. (eds.), 2011 – *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, Boca Raton: Taylor & Francis
- Edwards P.J., Roberts I., Clarke, M.J., DiGiuseppi C., Wentz R., Kwan I., Cooper R., Felix L.M., Pratap S., 2009 – Methods to increase response to postal and electronic questionnaires, *Cochrane database of systematic review*, n°02/2009
- Gombault V., 2013 – L'internet de plus en plus prisé, l'internaute de plus en plus mobile INSEE première, n°1452, juin
- Haziza, D., Lesage E., 2015 – "A discussion of weighting procedures for unit nonresponse.", *Journal of Official Statistics* (to be published).
- Knoef M., de Vos K., 2009 – The representativeness of the LISS, an online probability panel, Tilburg, CentERdata, 29p.  
([http://www.lisspanel.nl/assets/uploaded/representativeness\\_LISS\\_panel.pdf](http://www.lisspanel.nl/assets/uploaded/representativeness_LISS_panel.pdf))
- Leenheer J., Scherpenzeel A.C., 2013 – Does it pay off to include non-internet households in an internet panel?, *International Journal of Internet Science*, vol. 8, n°1, p.17-29
- Revilla M., Cornilleau A., Cousteaux A-S., Legleye S., de Pedreza P., 2015 – "What is the gain in a probability based online panel of providing internet access to sampling units who previously had no access?", *Social Science Computer Review* (online)
- Scherpenzeel A., 2009 – Start of the LISS panel: Sample and recruitment of a probability-based internet panel, Tilburg, CentERdata, 9 p.  
([http://www.lissdata.nl/assets/uploaded/Sample\\_and\\_Recruitment.pdf](http://www.lissdata.nl/assets/uploaded/Sample_and_Recruitment.pdf))