

<b>Acronyme du projet/ Acronym of the project</b>	<b>M.I.D.</b>	
<b>Titre du projet en français</b>	Méthodologie – Infrastructure - Données	
<b>Project title in English</b>	Methodology – Infrastructure - Data	
<b>Coordinateur du projet/Coordinator of the project</b>	Nom / Name : Laurent Lesnard Etablissement / Institution : Sciences Po Laboratoire / laboratory : CDSP – Centre de données socio-politiques / Centre for socio-political data Numéro d'unité/unit number : UMS 828	
<b>Aide demandée/ Requested funding</b>	<b>Tranche 1/Phase 1</b> €	<b>Tranche 2/Phase 2</b> €
<b>Champs disciplinaires / disciplinary field</b>	<input type="checkbox"/> Santé, bien-être, alimentation et biotechnologies / health, well-being, nutrition and biotechnologies <input type="checkbox"/> Urgence environnementale et écotechnologies / environmental urgency, ecotechnologies <input type="checkbox"/> Information, communication et nanotechnologies / information, communication and nanotechnologies <input type="checkbox"/> Sciences humaines et sociales / social sciences <input type="checkbox"/> Autre champ disciplinaire / other disciplinary scope compléter	
<b>Domaines scientifiques/ scientific area</b>	compléter	

**Affiliation(s) du partenaire coordinateur de projet/ Organization of the coordinating partner**

<b>Laboratoire(s)/Etablissement(s) Laboratory/Institution(s)</b>	<b>Numéro(s) d'unité/ Unit number</b>	<b>Tutelle(s) /Research organization reference</b>
CDSP	UMS 828	Sciences Po / CNRS

**Affiliations des partenaires au projet/Organization of the partner(s)**

<b>Laboratoire(s)/Etablissement(s) Laboratory/Institution(s)</b>	<b>Numéro(s) d'unité/ Unit number</b>	<b>Tutelle(s)/Research organization reference</b>
compléter		
<b>Entreprise(s) / company</b>	<b>Secteur(s) d'activité/activity field</b>	<b>Effectif/ Staff size</b>
compléter		

1.	RESUME / SUMMARY .....	3
2.	ENVIRONNEMENT SCIENTIFIQUE ET POSITIONNEMENT DU PROJET D'EQUIPEMENT / SCIENTIFIC ENVIRONNEMENT AND POSITIONING OF THE EQUIPMENT PROJECT .....	4
3.	DESCRIPTION SCIENTIFIQUE ET TECHNIQUE / TECHNICAL AND SCIENTIFIC DESCRIPTION OF THE ACTIVITIES .....	4
<b>3.1.</b>	<b>Originalité et caractère novateur du projet d'équipement/Originality and innovative features of the equipment project 9</b>	
<b>3.2.</b>	<b>Description du projet / Description of the project.....</b>	<b>12</b>
3.2.1	Présentation scientifique du projet/Scientific programme	12
3.2.2	Structure et composition de l'équipement /Structure and building of the equipment	19
3.2.3	Environnement technique / Technical environment	22
4.	STRATEGIE DE VALORISATION DES RESULTATS/ DISSEMINATION AND EXPLOITATION OF RESULTS .....	24
5.	MANAGEMENT DU PROJET / PROJECT MANAGEMENT.....	27
<b>5.1.</b>	<b>Aspects organisationnels / Management .....</b>	<b>27</b>
5.1.1	Qualification du coordinateur de projet /Relevant experience of the project coordinator	27
5.1.2	Modalités de coordination/ Coordination modalities	27
<b>5.2.</b>	<b>Organisation du partenariat / Collaboration organization.....</b>	<b>30</b>
5.2.1	Description, adéquation et complémentarité des partenaires/Partners description & relevance, complementarity	30
5.2.2	Qualification, rôle et implication des partenaires / Qualification, role and involvement of individual partners	30
6.	EVALUATION FINANCIERE DU PROJET/ FINANCIAL ASSESSMENT .....	31
7.	ANNEXES / APPENDICES .....	31
<b>7.1.</b>	<b>Références bibliographiques de l'état de l'art/State of art references .....</b>	<b>31</b>
<b>7.2.</b>	<b>Références bibliographiques des partenaires/Partners' references .....</b>	<b>31</b>
<b>7.3.</b>	<b>Estimate for the equipment.....</b>	<b>35</b>

**APPEL A PROJETS EQUIPEX /  
CALL FOR PROPOSALS**

**2010**

**Acronyme du projet /  
Acronym**

**M.I.D**

**DOCUMENT SCIENTIFIQUE B /  
SCIENTIFIC SUBMISSION FORM B**

## **SUMMARY**

*(2 pages maximum)*

## 1. SCIENTIFIC ENVIRONMENT AND POSITIONING OF THE EQUIPEMENT PROJECT

Dix ans après le rapport Silberman sur les sciences sociales et leurs données, les universités françaises accusent toujours un retard important en matière de collecte, de capitalisation, de diffusion de données et de recherche et d'enseignement en méthodologies d'enquête, qu'elles soient qualitatives ou quantitatives. L'excellence méthodologique, indispensable dans la compétition internationale, s'acquiert aujourd'hui plus que jamais par la possibilité d'élaborer des dispositifs innovants et expérimentaux, ce qui suppose en retour la capacité de mener des enquêtes et des analyses secondaires avec une maîtrise de toutes les étapes dont elles se composent. Aux enquêtes par questionnaire et par méthodes qualitatives s'ajoutent également les nouveaux terrains d'investigations apparus avec Internet (notamment le web social).

### Collecte de données par questionnaires

Bien qu'il ait été mis en évidence dans le rapport Silberman il y a dix ans, la France accuse toujours un très grand retard en matière d'enquêtes universitaires par questionnaire. Les raisons de ce retard, leurs conséquences et les solutions pour y remédier sont connues. En l'absence d'une plateforme universitaire française d'enquêtes par questionnaire, la collecte de données doit être sous-traitée aux entreprises de sondages privées qui présentent trois inconvénients liés à leur modèle économique (Pages et al. 2006, Caveng 2009). Le premier est leur coût très élevé occasionné par les marges bénéficiaires appliquées. Le second est la faible exigence méthodologique ainsi que la mauvaise qualité du recueil d'informations occasionnées par les conditions d'emploi (contrats courts, pas d'incitations à la professionnalisation, turn over, voir Duclos et al. 1996 p. 82-85 et Caveng 2007), les conditions de travail (cadences et pression élevées) et la faible formation du personnel d'enquête. Enfin, ces entreprises ont peu d'expérience dans le traitement d'échantillons aléatoires ou de grande taille et leur personnel est peu familiarisé aux problématiques des SHS (l'essentiel de l'activité du secteur consiste en études de marché et, pour les entreprises réalisant des enquêtes d'opinion, la part de ces dernières est généralement comprise entre 10% et 20% de leur chiffre d'affaire).

Dès lors, les enquêtes sont en France très coûteuses (pour l'enquête RAPFI auprès des Français d'origine maghrébine, africaine et turque, c'est presque 90% du budget qui a été ainsi dépensé), donc difficiles à financer. Peu nombreuses en raison de leurs coûts, ces enquêtes accusent souvent du retard du point de vue des normes méthodologiques internationales qui reposent d'abord sur des échantillons probabilistes, et non sur la méthode des quotas qui est actuellement la seule méthode abordable pour la recherche universitaire française, mais aussi et de plus en plus sur le suivi au cours du temps d'une même population (enquêtes par panel), méthode rarement utilisée en France. Peu formés et le plus souvent payés au questionnaire, les enquêteurs des instituts de sondage privés ignorent aussi fréquemment les consignes et règles de passation des questionnaires (notamment les relances) pour limiter le temps passé pour chaque interview (Caveng 2009), réduisant à néant les efforts méthodologiques des concepteurs des questionnaires. Peu propice à la réalisation d'enquêtes par questionnaire obéissant aux normes internationales, le contexte français favorise moins encore la recherche en méthodologies d'enquête qui repose par définition sur la réalisation d'expérimentations selon des protocoles stricts. Privée de moyens d'enquêtes propres, la communauté universitaire française peine à rivaliser à l'international en matière d'enquêtes et de méthodologie d'enquête.

Les solutions à ces problèmes sont bien connues et ont été mises en œuvre depuis longtemps en Amérique du Nord et en Europe dans le cadre de *Survey Research Center* (SRC). Présentes dans

presque toutes les universités de premier rang en sciences sociales (Princeton, Berkeley, par exemple), ces structures universitaires intègrent toute la chaîne des opérations d'une enquête par questionnaire et donnent ainsi accès aux chercheurs à des moyens d'enquêtes pour un coût moindre et un haut niveau méthodologique qui découle notamment de l'expertise accumulée par la recherche méthodologique menée dans les SRC. C'est par exemple le SRC de l'Université du Michigan qui réalise la *National Election studies*, une enquête académique conduite tous les deux ans depuis 1948, qui est considérée par la communauté scientifique internationale comme la référence tant en termes de questions de recherche que de qualité méthodologique (avec une vague pré- et une vague post-électorale ainsi que des panels qui s'étendent parfois sur quatre années). En France, seul l'IEP de Grenoble est doté d'une structure d'enquête par téléphone dédiée, associée à un Master spécialisé dans les études d'opinion et de marché (PROGIS), mais cette structure est de taille limitée (10 postes) et n'est pas ouverte à la communauté scientifique nationale. L'Ined est la seule institution française, hors Insee, qui a su mener des enquêtes au plus haut niveau méthodologique et développer une recherche en méthodologies d'enquête. En l'absence de moyens propres de collecte de données, l'Ined dépend cependant de l'Insee et des instituts de sondage.

Toutefois, si dix ans après le rapport Silberman le retard français subsiste, le contexte international a considérablement changé. La recherche empirique quantitative en SHS est en effet en train d'être profondément renouvelée par le développement des nouvelles technologies de l'information et de la communication : diffusion de l'internet, de nouveaux dispositifs de communication et de mesure comme les téléphones portables, les smartphones, les tablettes (ipad), le GPS, des outils de mesure de biomarqueurs (Raento 2009). Ces nouveaux outils offrent de nouveaux moyens de collecte de données plus riches, plus rapides et à des coûts plus faibles que les moyens classiques d'enquête. La possibilité d'utiliser des aides visuelles (Toepel et al. 2009), des vidéos, des applications interactives, de géo-localiser les répondants (avec leur consentement) ouvre de nouveaux horizons pour la recherche en SHS et plus largement pour la communauté scientifique dans son ensemble (épidémiologie, ergonomie, informatique, etc.). Contrairement aux autres modes d'enquête (face-à-face, téléphone, auto-administré), il n'y a pas de limite au nombre de personnes interrogées simultanément ce qui permet de mener des enquêtes et d'avoir accès aux données très rapidement. L'absence d'intermédiaires permet également d'abaisser le coût de collecte et de mieux contrôler les règles et les conditions de passation des questionnaires (relances, aides, interruptions et reprises du questionnaire, etc.).

Ces nouveaux outils redistribuent les cartes de la recherche empirique en SHS. Parce qu'ils ont su saisir cette opportunité, les Pays-Bas, dépourvus comme la France de SRC ancienne génération, sont dotés avec le LISS panel de l'Université de Tilburg d'un panel internet représentatif de la population néerlandaise qui les positionnent à l'avant garde de la recherche quantitative en SHS. Novateur, le LISS panel est également au plus haut niveau méthodologique, notamment dans la constitution de l'échantillon qui est constitué de manière à représenter la population néerlandaise et qui repose sur la mise à disposition de moyens de connexion à Internet aux répondants qui en sont dépourvus. Cette procédure permet d'éviter les nombreux biais des enquêtes par internet classiques qui ne sont menées que sur les personnes qui disposent d'une connexion internet (Cooper 2000).

### **Archiver et diffuser - pour les réutiliser - les enquêtes qualitatives : un impératif pour le développement et l'innovation des SHS**

Alors même que les SHS françaises se caractérisent, comparativement à ce qui se fait à l'étranger (notamment aux Etats-Unis), par un recours massif aux méthodes qualitatives (en science politique par exemple, voir Billordo 2005), la contribution française en termes d'expertise et d'innovation

méthodologiques reste limitée, ce qui s'explique notamment par l'absence complète d'archivage en matière d'enquêtes qualitatives. Pourtant les inconvénients liés à la non capitalisation de ces enquêtes, soulignés dans le rapport Cribier il y a déjà plusieurs années de cela (Cribier 2003), restent toujours d'actualité. Pire, la plupart des données qualitatives (entretiens, notes d'observations, etc.) recueillies depuis la grande période de développement que les sciences sociales ont connue à partir des années soixante ont disparu ; celles qui restent sont conservées dans de mauvaises conditions et risquent fortement de disparaître à leur tour si rien n'est fait rapidement. Cette situation s'explique, outre la méconnaissance ou l'insensibilité de la communauté des chercheurs pour ces questions, par le faible intérêt des archives publiques pour ce type de matériaux (Braibant, 1996 ; Duclert, 2002) et les faibles moyens d'archivage des institutions d'enseignement supérieur et de recherche. Malgré quelques initiatives localisées et partielles, notamment en histoire orale (Duclert, 2002), il n'y a pas, à l'heure actuelle, de vraie politique d'archivage scientifique des données d'enquêtes qualitatives. En sociologie ou science politique, les rares expériences pionnières en matière d'archivage et de partage des données qualitatives sont de plus venues de la recherche appliquée<sup>1</sup>. Ce retard est d'autant plus paradoxal que la création d'une banque centralisant les enquêtes qualitatives en SHS au niveau national correspond à un souhait régulièrement affiché par les institutions publiques, qui ont financé plusieurs rapports en ce sens (Braibant, 1996 ; Silberman, 1999 ; Dubar *et al.*, 2001 ; Cribier, 2003).

On voit bien, pourtant, tout ce que l'archivage et la mise à disposition de ces données pourrait apporter aux SHS : préserver leur histoire mais aussi asseoir leur scientificité, et leur permettre de progresser. Les publications tirées d'enquêtes qualitatives n'offrent en effet généralement aucun accès aux données sur lesquelles portent l'analyse et les protocoles d'enquêtes y sont décrits de façon limitée (Moravcsik 2010), ce qui ne permet pas aux lecteurs de comparer directement les stratégies méthodologiques mises en œuvre et leur impact sur les résultats obtenus. Plus généralement le bricolage est en fait souvent la règle, tout comme le conformisme envers les principes consacrés ; notamment l'innovation méthodologique est bridée puisqu'elle ne peut pas se nourrir de la connaissance fine des expériences antérieures. Ici, c'est l'amélioration des protocoles de recherche qui se trouve entravée, et la cumulativité des recherches, sur le plan méthodologique, reste limitée.

Enfin, au-delà des questions de méthodes, l'absence d'archivage rend également impossible la ré-analyse des enquêtes. Or publier, en qualitatif comme ailleurs, suppose d'abord et avant tout de faire des choix, de ne retenir que certains aspects des observations qu'on a pu faire, ceux qui prennent sens dans le cadre théorique qu'on a élaboré. Pourtant une enquête est toujours susceptible d'analyses multiples ; surtout si elle est mise en perspective, voire directement comparée à d'autres enquêtes, d'autres données : l'absence d'archivage prive notamment les SHS de toute possibilité de comparaison temporelle.

Ces différents arguments<sup>2</sup> ont été largement mis en avant par les créateurs de la banque de données qualitative pionnière que constitue ESDS Qualidata, un service du UK data archive<sup>3</sup>. Les concepteurs et responsables de Qualidata ont ouvert un large débat<sup>4</sup>, (voir la liste des références

---

<sup>1</sup>C'est le cas de EDF R& D, avec la base « Verbatim » ; développée depuis 1998 sur le modèle de Qualidata, elle comportait 72 études et 778 entretiens de 1994 à 2007 (pour un total de 160 Mo d'informations) ; elle est accessible uniquement aux chercheurs de EDF R&D. Pour une description de cette initiative, voir Le Roux, Vidal, 2000a, Le Roux, Vidal, 2000b Dargentas, Le Roux, 2005 ; Le Roux, 2009b ; Le Roux, Vidal, 2000.

<sup>2</sup>L'archivage des données est aussi un puissant moyen d'encourager la collaboration entre chercheurs (McDermott 2010)

<sup>3</sup>Localisé à l'Université d'Essex, le UK data archive, qui mêle donc archives quantitatives et qualitatives de sciences sociales, est une institution publique qui date de 1967. Les deux archives, qualitative et quantitative, ont été intégrées en 2001 dans le Economic and Social Data Service (ESDS). <http://www.esds.ac.uk/qualidata/about/introduction.asp>

<sup>4</sup>Cf. par exemple le forum Qualitative social research (<http://www.qualitative-research.net>)

bibliographiques compilée par Qualidata, qui comprend plus d'une cinquantaine de textes: <http://www.esds.ac.uk/qualidata/support/reusearticles.asp>) dont les chercheurs français sont - malheureusement - singulièrement resté à l'écart jusqu'à présent<sup>5</sup>, sur la justification de l'archivage et les conditions méthodologiques de l'analyse secondaire des données qualitatives. Leurs arguments, d'orientation positiviste, rencontrent le scepticisme de chercheurs porteurs d'une épistémologie constructiviste et critique pour qui les informations recueillies ou les données construites ne sont interprétables et analysables que pour celui qui a directement participé à l'enquête, tout autre usage des données étant considéré comme fallacieux (Duchesne, à paraître). Nous nous plaçons entre ces deux pôles, au sens où nous pensons qu'il est possible d'archiver pour qu'elles soient réutilisées par d'autres les données d'enquêtes qualitatives, en posant comme condition que l'enquête soit comprise que comme un tout et que les contextes et les métadonnées soient renseignés le plus possible. Malgré la résistance d'une proportion non négligeable de la communauté scientifique, l'exemple de Qualidata a pourtant fini par s'imposer ; la plupart des pays européens mettent en place une base de données de même type, en intégrant dans des structures communes les enquêtes qualitatives et quantitatives et en s'efforçant de centraliser l'archivage, pour des raisons d'accessibilité mais aussi de compétence des personnels chargés de ce travail<sup>6</sup>. Face à cela, la France n'est pourtant pas en reste puisque les événements scientifiques portant sur les questions d'analyse qualitative des données y ont foisonné ces dernières années, témoignant d'une plus grande sensibilité de la communauté à ces problématiques<sup>7</sup>. Surtout une initiative fondatrice a récemment été prise, en 2009, en partenariat entre Sciences Po et le CNRS (via le TGE-Adonis), qui ont financé une mission exploratoire pour créer une banque d'enquêtes qualitatives en SHS, en s'appuyant sur le développement de l'Internet (grâce à l'expertise du médialab de Sciences Po), de manière à répondre concrètement aux arguments des chercheurs sceptiques quant aux bienfaits de l'archivage et de l'analyse secondaire<sup>8</sup>.

### Le web comme corpus

Les usages du web se sont diversifiés depuis les années 1990 où la forme dominante était le site web. Avec l'essor des connexions haut-débits, de l'accès mobile à Internet et des technologies d'interactions riches (réunies autour du terme web 2.0), le web est devenu un média incontournable où se tissent de nombreuses interactions sociales (Presner & Schnapp, 2009). Or le web est un média numérique où toute donnée est aisément traçable. Un nombre croissant de nouvelles données issues du web sont désormais disponibles mais celles-ci souffrent toujours d'une mauvaise réputation quant à leur qualité (Shirky, 2005). L'environnement numérique apporte un nouveau paradigme où la versatilité et la facilité de production de l'information implique une perte de contrôle sur l'authenticité et la qualité des données disponibles. Pourtant la quantité et surtout la diversité de ces traces laissées par les interactions sociales représentent des opportunités inédites pour les sciences humaines et sociales.

---

<sup>5</sup>A l'exception des contributions de l'équipe de Verbatim et notamment de Dargentas M., Brugidou M., Le Roux D., Salomon A.-C. (dir.), *L'analyse secondaire en recherche qualitative : une nouvelle pratique en sciences humaines et sociales*, Paris, Lavoisier, à paraître

<sup>6</sup>Finlande, Irlande, encore en Suisse.

<sup>7</sup>Exemple : journées d'études organisées par le groupe CAPAS à Grenoble sur l'éthique de la recherche, journée d'études de l'axe DIM (données, infrastructure, méthodes) de Sciences Po sur la mise en forme et la restitution des entretiens en sciences sociales (décembre 2007) et l'analyse textuelle assistée par ordinateur (juin 2009), la récente journée d'étude organisée par le GRHISPO à Amiens sur les usages des entretiens dans les sciences sociales et historiques (juin 2009), le colloque prochainement organisé à l'Université de Limoges sur le thème « Droit d'enquêter / Droits des enquêtés » (septembre 2009), etc.

<sup>8</sup>Pour une présentation des résultats de cette mission exploratoire, voir <http://blogs.sciences-po.fr/recherche-donnees-qualitatives/>

Si ces données peuvent paraître sales et non conformes aux exigences méthodologiques, c'est qu'elles sont abordées d'un point de vue méthodologique traditionnel en rapport aux techniques de récolte classique (non numérique). Les données du web resteront inexploitées ou mal-exploitées tant que des méthodes spécifiquement numériques ne seront pas mises au point. L'enjeu qu'elles représentent d'un point de vue scientifique justifie que des investissements spécifiques soit mobilisés pour le développement de méthode de récolte et d'analyse adaptés au numérique (Lazer et al., 2009). Le médialab de Sciences Po a d'ailleurs été créé dans cette optique.

La qualité de la structuration des données est très hétérogène. Des plateformes établies comme wikipedia ou les différents réseaux sociaux récents ont pris soin de structurer finement la production des données. Cela explique que les premiers champs de recherche de sociologie sur le web se sont d'abord intéressés à ces corpus (wikipedia studies, online social network analysis) (Marres, 2009). Mais une grande partie des données disponibles réside dans les nombreux sites non uniformément structurés que sont les forums, les blogs et autres sites Internet qu'ils soient personnels ou institutionnels. Le champ de recherche des web studies a contribué à exploiter ce corpus en s'intéressant notamment aux questions de topologie (liens entre sites web) et de mémétique (diffusion de mème). Cependant la question de l'exploitation de l'expression spontanée reste trop peu explorée hormis les initiatives privées orientées vers le marketing (émergences des activités de mesures d'influence et de sentiment analysis sur des corpus web).

## 2. TECHNICAL AND SCIENTIFIC DESCRIPTION OF THE ACTIVITIES

### 2.1. ORIGINALITY AND INNOVATIVE FEATURE OF THE EQUIPEMENT PROJECT

Le projet M.I.D. vise à doter la France d'une structure intermédiaire novatrice de collecte, d'enrichissement et de diffusion de données destinée non seulement à rattraper ces retards mais aussi à se positionner comme un interlocuteur répondant aux plus hautes exigences de compétitivité internationale. Son originalité tient à ce qu'il considère de manière unifiée la collecte, la documentation et la diffusion de trois des principales données des SHS et qu'il tire parti des possibilités qu'offre l'Internet et les nouvelles technologies, notamment en termes de coûts, mais aussi de possibilité d'exploration et de diffusion des données classiques, ainsi que de création de données nouvelles. Considérer conjointement les différentes données et les différents modes de leur collecte permettra aux chercheurs de mener des enquêtes novatrices (par exemple suivi dans le temps des comportements alimentaires, des modes de vie et du Body Mass Index et analyse des forums du web, notamment ceux qui portent sur les troubles du comportement alimentaire) et de développer des recherches multi-méthodes qui permettront de replacer la France au meilleur niveau international et d'offrir aux chercheurs qui auront recours aux services de M.I.D. le plus haut niveau d'exigence scientifique. Réunir ces trois activités permettra également d'alléger leur coût de fonctionnement par les nombreuses synergies et mutualisations rendues possible par la place croissante de l'informatique et d'Internet tant dans la collecte que l'enrichissement et la diffusion des données en SHS.

#### **Collecte de données par questionnaires**

Le premier dispositif de M.I.D. s'inscrit dans le contexte international d'excellence méthodologique et d'innovation évoqué dans la 1<sup>re</sup> section. La composante SRC de M.I.D. vise à mettre en place un panel représentatif de la population française interrogé par internet via un terminal mobile complété par un dispositif d'enquête par téléphone et en ligne. Outre la réduction des coûts permise par la collecte de données directe via le web et le respect des critères méthodologiques les plus exigeants en matière d'échantillonnage, l'utilisation de terminaux mobiles offre un certain nombre d'avantages : enquêtes innovantes qui tirent profit des smartphones, facilité d'utilisation par rapport à un ordinateur (taux de pénétration des téléphones mobiles de 95 %, 2009 ARCEP), possibilités de participation aux enquêtes proposées plus étendues que via un ordinateur (mobilité), objet plus personnel qu'un ordinateur. En outre, cette méthode d'enquête ouvre des perspectives de suivi de longue durée des répondants (attrition limitée du fait des possibilités de suivi par le numéro de téléphone portable).

#### **Archivage et diffusion des enquêtes qualitatives: innover pour mieux respecter la nature de ce type de données**

Le module BEQuali de M.I.D. a vocation à prolonger l'initiative Française en finalisant l'intégration de la banque à un centre de ressource plus large au niveau national. Le dispositif que nous développons est innovant à double titre. Il repose sur *l'enquête* plutôt que sur les *données*, c'est-à-dire qu'il vise à limiter les risques de traitement décontextualisé des données. Une "enquête sur l'enquête" (entretien approfondi avec le chercheur déposant ou un membre de l'équipe de recherche qui a réalisé l'enquête) sera réalisée pour donner sens aux documents archivés et reconstituer *ex post* la dynamique de la recherche. Le deuxième élément novateur de BEQuali consiste à utiliser plus complètement les potentialités du web pour valoriser ces données au travers de mini-site webs par enquête qui (ré)tablissent tous les liens possibles entre les documents proposés et proposent des outils de navigation et d'exploration.

### Corpus web

Le module données du web de M.I.D sera le premier équipement en France à proposer aux chercheurs un service de collecte de données issues du web, service aujourd'hui uniquement disponible dans le secteur privé. Ce service apportera des opportunités nouvelles dans la diversité des données, dans les fréquences de récolte et dans l'ampleur des corpus disponibles. De plus la prise en compte du corpus web dans un équipement proposant également des méthodes d'enquêtes par questionnaire et qualitative permettra d'offrir une véritable complémentarité méthodologique aux projets de recherche. Les méthodes webs pourront permettre de préciser le reflet d'une problématique avant de monter une enquête par questionnaire ou inversement venir compléter une étude classique par une analyse quali-quantitative de verbatims webs.

### Partenaires

M.I.D. repose sur la complémentarité des partenaires réunis autour de ce projet qui apportent leurs expertises en recherche méthodologique reconnues au niveau international. La réussite d'un tel équipement novateur dépend en effet de la mise en commun et de la complémentarité d'expertises méthodologiques variées, que l'équipement, en retour, permettra de renforcer et de développer. Les recherches méthodologiques qui pourront être menées par les partenaires de M.I.D. permettront de maintenir la compétitivité de l'équipement et la qualité des recherches qu'il permettra de mener.

La collecte et diffusion de données quantitatives de M.I.D. s'appuient sur les expériences de réalisation d'enquêtes sur des sujets très divers réalisées par les laboratoires du PRES Sorbonne Paris Cité (CDSP<sup>9</sup>, CEE<sup>10</sup>, CERLIS<sup>11</sup>, CEVIPOF<sup>12</sup> et OSC<sup>13</sup>), par ceux du GENES (dont la composante recherche est le Centre de recherche en économie et statistique). Il dispose également avec le CDSP d'un centre de données innovant qui fait partie du Réseau Quetelet. L'Ined, dont le service des enquêtes est également membre du Réseau Quetelet, a une solide expérience en matière de conduites d'enquêtes sur des sujets sensibles (sexualité, santé reproductive, etc.) et avec des protocoles innovants (mixed mode surveys, internet, papier, téléphone etc.). Par ailleurs, cette activité bénéficiera des compétences du GRETS d'EDF R&D (analyse statistique des questions ouvertes) et de Telecom ParisTech (méthodes d'enquêtes).

L'acquisition, l'enrichissement et la diffusion des enquêtes qualitatives bénéficie de l'expérience du GRETS (appartenant à EDF R&D) qui dispose d'une base de données d'entretiens depuis 1998 et mené des recherches sur l'analyse secondaire avec le laboratoire PACTE (Grenoble). Les différents laboratoires de Sciences Po impliqués dans M.I.D. (CDSP, CEE, Cevipof mais aussi OSC, CSO<sup>14</sup>, Centre d'Histoire, ainsi que le médialab) ont contribué à la mission exploratoire, soutenue par sa direction scientifique et le TGE-Adonis du CNRS, pour l'établissement d'une base d'enquêtes qualitatives, et dont les conclusions vont servir de fondements au dispositif que nous entendons créer. Cette activité bénéficiera également de l'expérience du département de sciences économiques et sociales de Telecom ParisTech qui a développé une expertise dans le domaine de l'ethnographie de l'activité.

La collecte, l'analyse et la visualisation des données du web s'appuiera sur l'expertise du PRES Sorbonne Paris cité qui dispose d'une structure, le médialab ayant pour vocation de développer une expertise méthodologique pour l'utilisation des technologies numériques dans le cadre de recherche en SHS. De nature interdisciplinaire par nature, l'équipe du médialab regroupe les compétences de

<sup>9</sup>Centre de données socio politiques, UMS Sciences Po et CNRS, évalué A+ par l'AERES.

<sup>10</sup>Centre d'études européennes, Sciences Po, équipe d'accueil conventionnée avec le CNRS (EA 4459).

<sup>11</sup>Centre de recherche sur les liens sociaux, UMR CNRS, Paris Descartes et Paris III, Sorbonne Nouvelle, évalué A+ par l'AERES.

<sup>12</sup>Centre de recherche politique de Sciences Po, évalué A+ par l'AERES.

<sup>13</sup>Observatoire sociologique du changement, UMR CNRS et Sciences Po, évalué A+ par l'AERES.

<sup>14</sup>Centre de sociologie des organisations, UMR CNRS et Sciences Po, évalué A+ par l'AERES.

chercheurs en sciences sociales, d'ingénieurs experts des données numériques et de designer de l'information. Ces activités bénéficieront également de l'expertise de Telecom ParisTech qui a été impliqué dans la conception de dispositifs de recueil et d'analyse de corpus du Web à partir d'échantillons représentatifs et d'outils de mesure des usages de l'internet. EDF R&D apportera également sa compétence reconnue dans l'analyse des données du web notamment sur l'analyse des données hétérogènes et l'analyse audio mining (analyse des émotions dans la voix)<sup>15</sup>.

Cumulées, les compétences et expertises des différents partenaires permettent de proposer un dispositif de collecte, d'enrichissement et de diffusion de données en SHS qui répond aux trois enjeux majeurs de la recherche en SHS en tirant tout le profit (innovation et coût) des ruptures technologiques et sociales du web, du web mobile et des smartphones.

---

<sup>15</sup> Ses équipent participant notamment au projet collaboratif DOXA sur l'analyse des opinions et des sentiments dans les textes et notamment dans des données WEB (blogs, forum), voir <http://www.capdigital.com/xwiki/bin/view/Projet/ProjectDoXa> et au projet collaboratif Vox factory sur l'analyse des interactions émotionnelles dans les centres d'appel, voir sur le web : <http://www.capdigital.com/vox-factory>.

## 2.2. DESCRIPTION OF THE PROJECT

### 2.2.1 SCIENTIFIC PROGRAMME

#### Collecte de données

La création des Survey Research Centers en Amérique du Nord et en Europe visait à dépasser les limites (échantillonnage par quota insuffisant au niveau international, faibles possibilités d'innovation méthodologique) et abaisser les coûts des enquêtes académiques sous-traitées à des entreprises privées de sondage. Si le succès durable des SRC montre que la collecte de données par des universitaires pour un coût moindre et une meilleure qualité est possible, il n'est toutefois plus possible d'appliquer en 2010 les solutions imaginées dans les années 1960. Si les limites et les coûts de la sous-traitance d'enquêtes par questionnaire aux instituts de sondage demeurent et deviennent de plus en plus handicapant pour la recherche française, deux autres paramètres doivent être pris en compte. Le premier est la baisse des taux de réponse aux enquêtes, qu'elles soient académiques ou commerciales. Le second est la rupture technologique de l'Internet et des dispositifs mobiles connectés.

La composante SRC de M.I.D. vise à doter la France de l'équipement de collecte de données compétitif à l'international qui fait défaut aux chercheurs en SHS. Il repose sur deux dispositifs complémentaires : un panel web et une salle d'appel téléphonique. Le panel web s'inspire des expériences américaines et néerlandaises et a pour principe de combiner le plus haut degré de rigueur et de qualité statistiques avec les avantages offerts par les évolutions technologiques récentes. Comme le LISS panel de l'université de Tilburg, le panel web reposera sur un échantillon représentatif de la population française tiré selon un plan de sondage aléatoire tiré par l'Insee. Chaque personne sélectionnée pour faire partie du panel recevra en échange de sa participation au dispositif d'enquête un dispositif mobile connecté (un smartphone Android par exemple) et un abonnement correspondant. Le recours à des smartphones permet de cumuler les avantages des enquêtes internet et de dispositifs mobiles.

Les enquêtes par internet permettent de réduire substantiellement le coût des enquêtes puisque par définition elles se déroulent sans le recours à des enquêteurs. Ce type d'enquête permet également de mener des enquêtes beaucoup plus rapidement puisqu'il n'y pas de phénomène de liste d'attente (capacité d'une salle d'appel téléphonique ou nombre d'enquêteurs pour des enquêtes face-à-face). Par ailleurs, les personnes interrogées peuvent répondre aux enquêtes par internet lorsqu'elles le souhaitent et non quand elles sont sollicitées par des enquêteurs par téléphone ou à leur domicile. Elles permettent également de réaliser des enquêtes plus riches (multimedia), de poser des questions différemment (son, texte et image), mais aussi de créer une relation différente avec les répondants (association des participants aux résultats, possibilité de se situer dans l'univers des réponses). À ces avantages propres aux enquêtes internet, s'ajoutent ceux qui découlent du choix d'un terminal mobile, originalité de M.I.D. par rapport aux projets équivalents aux États-Unis et en Europe. Les terminaux mobiles présentent trois sortes d'avantages par rapport aux ordinateurs classiques.

En premier lieu, le téléphone portable est totalement intégré dans la vie quotidienne, c'est même un des objets technologiques qui a été le plus rapidement et massivement adopté par les Français (taux de pénétration des téléphones mobiles de 95 % en 2009 d'après l'ARCEP). Objet de la vie quotidienne, il est plus facile d'utilisation qu'un ordinateur en particulier ceux des dernières générations équipés de systèmes d'exploitation (iOS ou Android) simples et intuitifs à utiliser, très loin des premiers smartphones et des PDA manipulables à l'aide de stylets et d'interfaces très proche des ordinateurs traditionnels. Ces terminaux rendent accessibles l'informatique connectée au plus grand nombre et apparaissent idéaux pour la collecte de données. En outre, parce qu'ils sont mobiles, ils équipent le plus souvent en permanence les personnes interrogées. Celles-ci peuvent donc répondre aux enquêtes quand et où elles le souhaitent et non pas seulement à leur

domicile. Mais de tels terminaux permettent également d'envisager des enquêtes beaucoup plus originales qui tirent profit de cette caractéristique et de la technologie embarquée dans les smartphones : étude des déplacements (géolocalisation des téléphones), étude de l'emploi du temps (carnet d'emploi du temps simplifié ou méthode alternative comme le *day reconstruction method*), suivi d'activités par récupération de marqueurs ultrasonores ou de tags présents dans l'environnement, suivi de biomarqueurs (auto-prélèvement de sang ou de salive, voir Finch, Vaupel et Kinsella 2001). Enfin, la mise à disposition de terminaux mobiles et d'abonnements devrait permettre la réduction de la non-réponse et de l'attrition (perte de répondants initialement dans un panel). En premier lieu tout simplement parce que le numéro de téléphone portable est connu, mais d'autre part par l'effet d'incitation que représente la mise à disposition d'un smartphone et le paiement d'un abonnement téléphonique. Les *incentives* réduisent la non-réponse et ce d'autant plus que les cadeaux offerts sont inconditionnels (i.e. donnés au début et non après comme une récompense) et d'une certaine valeur (Mack et al. 1998, Singer and Kulka 2002).

À ces avantages s'ajoutent ceux d'une enquête par panel qui permet de suivre dans le temps les mêmes individus.

EXEMPLE EN SCIENCE POLITIQUE (will be added later). Mise en relation de plusieurs thèmes

D'un point de vue méthodologique, le suivi longitudinal d'une même population permet d'améliorer la qualité des informations collectées (répétition de certaines questions, vérification de la cohérence, possibilité de modification) et de mieux prendre en compte au moment de l'analyse les spécificités individuelles inobservées.

Bien qu'innovant, le panel web doit être complété d'un centre d'enquête téléphonique. Ce centre aura notamment pour vocation de traiter les enquêtes transversales ponctuelles pour lesquelles soit un échantillon différent de celui du panel est souhaitable, soit l'assistance d'un enquêteur paraît nécessaire à la bonne compréhension des questions, ainsi que celles pouvant difficilement s'inscrire dans le calendrier du panel, mais aussi pour conduire des enquêtes sur le panel (mixed mode). Par ailleurs, il servira à maintenir le contact avec les enquêtés, effectuer certaines relances pour des enquêtes de panel etc. Des outils pour mener des enquêtes en ligne seront également mis à disposition. Ils seront accompagnés d'une assistance méthodologique pour une utilisation rigoureuse de ce mode d'enquête, autant dans les modes de constitution des échantillons que dans les méthodes de redressement. Les recherches méthodologiques sont encore embryonnaires sur les biais liés à ces méthodes, mais pourraient faire l'objet de recherches avec les autres outils de M.I.D.

La réalisation du centre d'appel téléphonique et du panel web mobile soulève plusieurs problèmes juridiques, techniques et économiques comme tout grand projet structurant. M.I.D. veillera à s'associer aux meilleurs juristes sur les questions de confidentialité des données personnelles. M.I.D. devra recruter des ingénieurs en informatique et nouer un partenariat avec un opérateur de télécom de façon à lever tous les problèmes techniques et à offrir les services qui sont annoncés pour le panel. Les plateformes de développement d'applications sur téléphone portable sont désormais accessibles au travers d'interfaces de programmation (API) mais il n'est pas toujours possible d'en contrôler l'usage selon les fournisseurs de matériel voire selon les opérateurs. Un accord d'intéressement mutuel doit pouvoir être trouvé qui préserve l'indépendance des objectifs scientifiques et la déontologie de la collecte de données personnalisées. L'opérateur de téléphonie mobile partenaire pourrait être Orange dont la structure de recherche et développement (Orange Labs), qui a une longue tradition d'enquêtes sociologiques appliquées, est partenaire du projet.

### Archivage et diffusion des enquêtes qualitatives

Le module BEQuali apparaît innovant au niveau national puisque une seule initiative proche, sans être strictement comparable, a vu le jour dans la période récente : le programme ARSHS,

financé d'abord par une ACI réseau MSH, puis par une ANR à l'initiative du réseau des MSH (et notamment de la MSH de Dijon). Ce programme a initialement consisté à coordonner une enquête nationale en forme d'état des lieux des archives en SHS, par questionnaire, dont les résultats sont mitigés (*Genèses*, 2006). L'équipe s'achemine actuellement vers la constitution d'une plate-forme archivistique pour les fonds de sciences sociales spécialisés sur les mondes ruraux avec laquelle BEquali travaillera de manière à mutualiser efforts et expériences.

Pour construire une base d'enquêtes qualitatives pérenne, les difficultés à surmonter sont multiples et bien identifiées par les dispositifs similaires en Europe<sup>16</sup>. Nous nous inspirerons donc des solutions qu'ils ont adoptées en les adaptant au contexte - institutionnel, juridique, scientifique - français, tout en innovant. Pour ce faire nous disposons déjà d'une première expérience puisque le module BEquali s'inscrit dans le prolongement de la mission exploratoire menée depuis 2009 et qui a déjà permis d'identifier les conditions d'archivage nécessaires à une réutilisation convaincante des données, et de commencer à y apporter des solutions. Cette expérience, qui en est à une phase de test d'un prototype composé de trois enquêtes choisies pour la diversité des problèmes qu'elles posent en termes d'archivage, montre qu'il est possible de documenter suffisamment *ex post* les données pour qu'elles conservent leur validité. En parallèle à la constitution de la banque, le projet REANALYSE, financé dans le cadre du programme blanc 2010, qui doit démarrer à l'automne 2010 et dont Sciences Po est aussi porteur, permettra de convaincre nos collègues de la faisabilité - au sens épistémologique autant que méthodologique - de l'analyse secondaire et des apports potentiels d'un tel dispositif pour la recherche.

L'option, conforme aux expériences menées ailleurs, en Europe notamment, consistant à intégrer le module BEquali à un centre de ressources déjà existant, le CDSP nous permettra de bénéficier de l'expérience préalablement accumulée dans le domaine quantitatif et de réaliser des économies d'échelle une fois le dispositif mis en place et rôdé. Cependant l'archivage, la diffusion et la réutilisation des données d'enquêtes qualitatives posent certaines difficultés spécifiques - particularité des matériaux, essentiellement textuels, en cause ; particularité des procédures de documentation et d'enrichissement des données - qui ne peuvent être surmontées qu'en mobilisant un personnel spécialisé. Par ailleurs, la réalisation d'une base "nouvelle génération" reposant sur des sites-enquête dotés d'outils de navigation permettant l'exploration et l'exploitation des données en ligne implique des développements informatiques spécifiques. Ici les verrous technologiques résident essentiellement dans la lourdeur du développement informatique. Nous partirons pour ce faire du prototype déjà imaginé avec le médialab<sup>17</sup> et bénéficierons des moyens informatiques mutualisés entre les trois composantes de M.I.D.. Nous devons également tenir compte des usages qui seront concrètement faits du dispositif, ce qui suppose de s'adapter aux attentes et pratiques effectives des utilisateurs.

La collecte des enquêtes au niveau national est un défi puisqu'il n'existe pas, comme en Suisse ou au Royaume-Uni, d'annuaire des chercheurs ou d'inventaire des projets de recherche en cours. Les chercheurs apparaissent titulaires de la propriété intellectuelle de leurs enquêtes (Cornu, 2003 ; Mallet-Poujol, 2004 ; Descamps, 2007, etc.) et seul le cas des chercheurs fonctionnaires nécessite d'obtenir l'accord du CNRS (ou d'autres instances comme l'INRA, etc.). Pour démarrer, nous disposons déjà d'un potentiel conséquent à travers les (nombreux) chercheurs des laboratoires du Pres Sorbonne Cité. Nous escomptons que la publicité accumulée par le dispositif permettra ensuite de le populariser et de le rendre attractif au sein de la communauté nationale, de façon à lever les réticences, nombreuses, des collègues sur le plan épistémologique, éthique ou académique<sup>18</sup>. C'est pourquoi nous porterons une attention réelle aux modalités d'association des

<sup>16</sup>.cit .op .,semblées dans Dargentas et alSur ce point voir les contributions ras

<sup>17</sup>lien en annexe vers la présentation de la maquette

<sup>18</sup>Les mêmes motifs freinant l'acceptabilité de tels centres d'archivage (et, avec elle, de l'analyse secondaire) reviennent dans les enquêtes d'acceptabilité menées dans plusieurs pays (Opitz, Mauer, 2005 ; Medjedović, à paraître)

chercheurs déposants au travail d'archivage et de diffusion, à travers la garantie de l'anonymisation des données, la réalisation de l'enquête sur l'enquête ou la possibilité pour les déposants de suivre les réutilisations effectives de leurs données et d'en bénéficier (par le jeu des citations par exemple). Nous escomptons qu'à terme, comme au Royaume-Uni ou aux Etats-Unis, le dépôt d'une enquête finira par être considéré comme une forme de reconnaissance de sa qualité (au même titre qu'une publication majeure par exemple), et les réanalyses auxquelles il donnerait lieu constitueraient un gage de publicité et d'approfondissement du travail original.

Un des bénéfices attendus du projet est d'améliorer les méthodes et les protocoles d'analyse de la recherche qualitative, à travers notamment le développement de l'analyse secondaire. En effet la possibilité non seulement de disposer de jeux de données beaucoup plus étendus et diversifiés qu'en l'état actuel, mais aussi d'avoir accès à des enquêtes configurés de manière holistique, devrait constituer un atout majeur pour un enseignement des méthodes dès le niveau master. De plus, la capitalisation des données qualitatives permettra de repousser certaines limites inhérentes à la démarche qualitative, en aidant d'abord les (jeunes) chercheurs à mieux préparer de nouvelles enquêtes mais surtout à produire de nouvelles connaissances en faisant "du neuf avec du vieux", et ceci à un moindre coût. En ce sens elle contribuerait à raccourcir de manière significative la durée d'une partie des thèses en SHS, en réduisant la phase, toujours chronophage, de terrain. Elle incitera les chercheurs à s'investir davantage dans l'analyse de données déjà existantes, mais insuffisamment exploitées, que dans celle de la production de nouvelles données. En permettant également d'accroître le nombre de cas analysés, elle pourrait même permettre de dépasser les critiques faites aux recherches qualitatives en termes notamment de généralisation de leurs résultats (King, Kehoane & Verba, 1994). En ce sens BEQuali contribuera à élargir le champ des questions que les chercheurs en SHS peuvent habituellement traiter – restreints qu'ils sont par leur capacités limités à produire, le plus souvent seuls ou en comités réduits, de nouvelles données qu'ils exploitent ensuite a minima, faute de temps notamment. Un tel dispositif favorisera ainsi la réalisation d'analyses diachroniques ou comparatives, et accompagnera la structuration de la recherche en SHS au niveau national, en facilitant l'intégration de divers domaines disciplinaires ou encore en encourageant au dépassement du clivage quali / quanti par une meilleure collaboration entre chercheurs spécialisés. A terme, une amélioration du travail scientifique pourrait donc s'opérer tout en réduisant les coûts de financement public de la recherche. La mise en œuvre du dispositif requiert essentiellement les moyens de rassembler les compétences nécessaires : les verrous sont bien identifiés et les solutions à portée de main.

#### **La récolte d'expression spontanée issues du web**

Le volet web de l'équipement M.I.D. aura pour mission de mettre au point les techniques et les méthodes de récolte des données web qui permettront leur exploitation dans un cadre méthodologique nouveau respectant les plus hautes exigences scientifiques. Il s'agit d'offrir un service de repérage et de cartographie des espaces de discussions en ligne pertinents suivant le sujet de la recherche puis une méthode d'extraction et de nettoyage semi-automatique de verbatims. Toute activité de récolte de donnée webs doit s'accompagner d'une réflexion sur les usages du web, sur la spécificité des interactions sociales qui se déploient sur des territoires variés (blogs, commentaires, messages de forums, messages sur des réseaux sociaux). L'ensemble de la chaîne (identification des données, constitution du corpus puis interprétation) doit être pensée en accord avec les usages du média web. Les données du web sont très hétérogènes de par leur nature et leur structuration. Nous distinguerons entre les données générées par une activité web (topologie des liens entre les sites, les requêtes envoyés aux moteurs de recherche, activités de collaboration sur wikipedia...) et les données issues de l'expression spontanée sur les différentes plateformes d'expression web (blogs, forums...).

La seule initiative se rapprochant de cela en France se situe dans le privé au sein de la société Linkfluence qui offre un service d'analyse de la présence de marques ou de communication institutionnelle sur le web. Le partenariat de recherche signé entre le médialab et Linkfluence en 2010 a permis le développement récent de projets de recherche exploitant les données de Linkfluence. Il est à ce jour trop tôt pour commenter ces initiatives mais la première des difficultés rencontrées concerne la non maîtrise du corpus disponible. Les conceptions méthodologiques de l'équipement M.I.D. et de Linkfluence s'opposent sur ce sujet. Linkfluence a bati un corpus de l'ensemble des communautés de blogs pour plusieurs pays grâce à une analyse topologique fine. Ce corpus est récolté et indexé régulièrement et sert de base pour des analyses variées pour chacun de leurs clients. À l'inverse l'équipement M.I.D. construira un corpus par projet de recherche. Les réservoirs de données d'expression spontanée pourront en partie se recouper mais d'un projet à l'autre la définition du corpus changera. Les méthodes d'identification et d'extraction des données seront elles évidemment mutualisées et réutilisées par chaque projet. L'équipement M.I.D. permettrait le développement par les chercheurs de leur propre méthodes de collecte pour leur redonner la maîtrise de la constitution du corpus analysé.

Un des enjeux sera d'améliorer la qualification de l'identité des auteurs des données web. Déterminer "qui parle ?" sur des corpus de données web est une problématique encore trop peu explorée. Adresser ce problème revient également à traiter le problème de la représentativité qui se pose en tout autre termes dans ce contexte. D'un côté, des investigations récentes ont montré que l'on pouvait retracer précisément le profil et les expériences récentes d'un individu par le seul fait de recouper des traces trouvées sur les nombreuses plateformes Internet ([article du Tigre](#)). De l'autre il est très difficile d'identifier les très nombreux rédacteurs des différents forums et autres blogs cachés derrière des pseudonymes. Le problème est plus ou moins complexe en fonction des plateformes, notamment celles des réseaux sociaux, mais dans tous les cas nous sommes loin des profils socio-démographiques des enquêtes classiques alors que d'un autre côté des éléments nouveaux de profils (appartenances diverses, goûts, activités) peuvent devenir porteurs d'une autre classification des auteurs. L'équipe de récolte de données web devra s'intéresser au paradoxe de l'identité sur Internet : peut-on et comment qualifier le profil des auteurs des données web ?

Cette question est à adresser d'une perspective à mi-chemin entre qualitatif et quantitatif. D'une part nous avons à faire à des verbatims "qualitatif" (d'une nature particulière car spontanée), de l'autre les possibilités de récolte de ces verbatims laisse imaginer un nombre très important poussant vers un contexte quantitatif. De faite l'immensité des données du web amène un renversement de perspective sur l'approche des corpus. La difficulté et le coût que représentait jusqu'ici la récolte de données posait la constitution de corpus dans un contexte de rareté des données et donc un objectif d'exhaustivité. Le contexte numérique connaît une toute autre réalité. L'abondance et non la rareté des données tenant à la facilité de la récolte pousse à développer des stratégies de filtrage (et de nettoyage) afin de limiter l'analyse aux données essentielles et non à l'ensemble d'un corpus trop large. Les corpus numériques ont à faire à des problématiques d'échantillonnage dans un contexte de lois de puissance où 20% des données rassemble 80% de la quantité d'information.

Ce passage à l'échelle quali-quantitatif exigé par les corpus web change également les méthodes d'analyse des contenus d'expression spontanée issues du web. Si les méthodes d'analyses des données qualitatives sont connues et maîtrisées, l'analyse des corpus d'expression spontanée reste un domaine encore vierge de méthodes établies. La première difficulté concerne l'analyse de la parole par des méthodes semi-automatique. l'enjeu est ici d'assister le chercheur en le dotant de moyens de recherche et d'exploration de contenus. Il s'agit de s'adjoindre les compétences de linguistique informatique spécialiste du traitement automatique du langage. La matière de l'expression spontanée étant particulièrement difficile car non formatée et non cadrée. Sur le point de l'analyse, l'équipement s'appuiera sur un autre projet d'équipement d'excellence MUSCA qui assure le développement de techniques d'analyses de grand corpus textuel. Le dernier verrou méthodologique lié à la récolte de données concerne la complémentarité des méthodes liées au

web avec les méthodes classiques et l'intégration des données issues des différents moyens de récolte. L'équipement proposé ici sera un lieu unique pour adresser ce défi.

D'un point de vue technique la récolte de données d'expression spontanée sur le web est un faux problème simple. Le numérique assure une simplicité certaine dans l'accès et la duplication de l'information mais la versatilité et la diversité des contenus compliquent considérablement la tâche. Il s'agira tout d'abord de lever le verrou du nettoyage des contenus webs afin d'extraire les verbatims des pages. Le problème se pose en terme différent suivant la plateforme d'expression à laquelle on s'intéresse. Si des initiatives existent pour la recherche de méthode automatique de nettoyage la grande variété des contenus web semblent s'opposer à cette approche. Une adaptation des méthodes d'extraction au plateforme assure en revanche une bien meilleure qualité d'extraction. Cette méthode permet notamment d'ajouter de nombreuses informations au verbatims comme sa date de publication, des caractéristiques à propos de son auteur... L'équipement devra investir tant dans la constitution de méthodes de nettoyage automatique que dans le développement d'heuristiques de nettoyage dédiées aux plateformes ciblées par la campagne de récolte. Une des valeur de l'équipement résidera pour la communauté dans son expertise et son capital d'heuristique afin d'assurer des critères de qualité suffisant aux données récoltées.

La très grande versatilité du web pose le défis de l'obsolescence des données d'une part et des méthodes d'extraction de l'autre. Ce verrou est un problème majeur des initiatives ponctuelles d'exploitation des données webs utilisées par la recherche à ce jour. Seul un équipement dédié à la tâche peut atteindre la masse critique suffisante à assurer un maintien des corpus et des méthodes d'extractions. Il s'agit tout d'abord de maintenir les corpus dans le temps en assurant une récolte régulière des contenus. Nous nous inspirerons des solutions imaginées par les deux acteurs en france du dépôt légal que sont la BNF et l'INA et par le consortium international IIPC<sup>19</sup>. Les méthodes de récolte devront être contrôlées et mise à jour en fonction de l'évolution des plateformes. L'équipe technique dédiée à l'équipement pourra assurer ce maintien "logiciel" de la même manière que tout grand équipement scientifique nécessite de la maintenance technique.

Les aspects juridiques de la gestion des données collectées par l'équipement sera une clef de son succès. Le cadre juridique de la constitution d'une base de données de verbatims d'expression spontanée extraits du web reste flou mais proche des problématiques d'archivage du web qui sont actuellement discutées en France à l'approche de la publication du décret des missions de dépôt légal du web par la BNF et l'INA. De plus comme toute base de données qualitatives, leur gestion impose une politique rigoureuse de respect de la vie privée. La culture et les expériences de l'ensemble des partenaires permettront de concevoir dans cet équipement une telle politique respectueuse des lois en vigueur.

Les différents services proposés par l'équipement M.I.D. devraient permettre aux chercheurs français de mener des enquêtes innovantes, au plus haut niveau scientifique, mono- ou multi-méthodes (enquêtes par questionnaire, ré-analyse d'enquêtes qualitatives, enquête sur le web). Ils permettront aussi aux chercheurs français de développer une expertise en méthodologie d'enquêtes construites sur des expérimentations et des recherches menées dans d'autres institutions de recherche. L'équipement M.I.D. permettrait également de mener des études comparatives avec des universités dotées d'instruments comparables comme l'Université de Tilburg aux Pays-Bas pour le panel Internet.

### Conditions d'accès

L'équipement M.I.D. sera accessible à l'ensemble de la communauté scientifique nationale et internationale (enquêtes comparatives) mais aussi aux acteurs publics (administrations,

---

<sup>19</sup>netpreserve.org

ministères) et privés (associations, entreprises). De vocation nationale, M.I.D. pourra être mobilisé par l'ensemble de la communauté académique en SHS : démographes, économistes, juristes, géographes, historiens, linguistes, politistes, sociologues. Mais les SHS sont elles-mêmes partenaires d'autres disciplines scientifiques et en conséquences, M.I.D. pourra contribuer au développement de la recherche en d'autres domaines (santé, analyse des risques, environnement par exemple) et travailler avec d'autres acteurs non universitaires : services de l'État, collectivités territoriales, ONG, entreprises. Un conseil scientifique définira les règles de la plateforme Web SHS, les différents types de participation et sélectionnera les projets. D'une façon générale, l'accès aux moyens d'enquête sera conditionné à une participation financière aux coûts de fonctionnement et à l'amortissement de l'équipement M.I.D. Celle-ci sera déterminée en fonction du projet et du statut des utilisateurs. Ainsi les étudiants auront un accès libre aux services de M.I.D., conformément à sa vocation d'appui à l'enseignement et au développement des compétences méthodologiques des nouvelles générations. Pour ce qui est de l'usage des données collectées par M.I.D. pour des utilisateurs professionnels, scientifiques ou non, la contribution individuelle en fonction du projet pourra être atténuée par la création d'un système d'adhésions institutionnelles, universitaires ou autres (conformément par exemple au fonctionnement de l'ECPR - European Consortium of Political Science): les membres des institutions partenaires se verront accorder des modalités d'accès privilégiées aux services de M.I.D..

### **Positionnement national et international**

Au niveau national, M.I.D. vient compléter la Très Grande Infrastructure de Recherche PROGEDO et le Très Grand Équipement du CNRS pour les sciences humaines et sociales Adonis. En effet, PROGEDO participe à des missions de collecte, de diffusion, de promotion, d'aide à la production et de préservation d'un vaste ensemble de données quantitatives, permet la production d'enquêtes européennes identifiées sur la feuille de route ESFRI (ESS/european social survey, SHARE/ Survey of health, ageing and retirement in Europe) et intervient en appui à la politique définie par le comité de concertation pour les données en sciences humaines et sociales (CCDHS). Centré sur des enquêtes innovantes et réactives, des procédures de collecte de données par et sur le web, le développement de l'archivage et de la diffusion des données qualitatives, M.I.D. propose des activités complémentaires du GIS PROGEDO-Réseau Quetelet en cours de création centré sur l'archivage de données et le soutien aux grandes enquêtes interanationales. Par ailleurs les modes de diffusion des données qualitatives, quantitatives et quali-quantitatives produites par M.I.D. seront conçus pour être interopérables avec les autres dispositifs proposés par Adonis.

M.I.D. apparaît également complémentaire à d'autres projets d'excellence. En premier lieu, il s'articule étroitement avec les autres projets d'excellence du PRES Sorbonne Paris Cité. M.I.D. viendra en appui aux deux projets de laboratoire d'excellence et tout particulièrement au projet "Santé Globale et Politiques de Santé", qui vise à construire un pôle pluridisciplinaire de référence pour l'analyse des politiques de santé et de la santé, notamment dans ses déterminants sociaux. Le panel Internet permettrait aux chercheurs de ce laboratoire de mener des enquêtes innovantes qui lie les comportements quotidiens à des conséquences pour la santé de plus long terme (par exemple suivi des pratiques alimentaires et des habitudes de vie et obésité). M.I.D. pourra proposer au projet de laboratoire d'excellence "Monde en Mutation" son expertise multi-méthodes (quali et web notamment) et l'archivage, documentation et diffusion sur le web des enquêtes qui pourraient être réalisées dans ce cadre. Par ailleurs, M.I.D. apparaît fortement complémentaire avec le projet d'équipement d'excellence GeoMedia Mapper dont l'objectif est de développer des outils de visualisation de données géographiques et de constitution de bases de données. Notamment, les données géolocalisées qui seront collectées dans le cadre du panel pourront être analysées avec les outils qui seront développés dans Terr'Lab. Elles pourront également être enrichies par leur mise en relation avec d'autres données géographiques au moyen des outils et standard qui seront développés dans le projet d'équipement d'excellence Centre Belgrand soutenu

par le PRES Paris Est. Enfin, M.I.D. est complémentaire du projet de laboratoire d'excellence Ined, qui vise à développer la recherche en démographie et disciplines apparentées en articulant et intégrant la recherche méthodologique d'enquêtes à leur production pour mieux en apprécier la qualité.

Le projet M.I.D. est aussi fortement complémentaire d'autres projets d'équipement d'excellence. Le médialab est partenaire du projet d'équipement d'excellence MUSCA (MULTI-SCALE collaborative mapping of large digital corpora, porté par l'École Polytechnique avec CAMS, CERSA, CSI, INIST, INRA, INSERM, Institut Curie, Univers Sciences) qui vise à doter la France d'un équipement de collecte massive de données et de calcul intensifs. M.I.D est également complémentaire avec deux autres projets d'équipement d'excellence déposés par d'autres membres du GIS Réseau Quetelet : le projet de Centre d'accès sécurisé aux données du PRES Paris Sud (mise à disposition hautement sécurisée de données de la statistique publique pour la recherche) et SHARE. De fortes synergies sont également attendues avec le projet d'équipement d'excellence data.lab du PRES Université de Grenoble qui vise notamment à se doter de moyens d'enquêtes par téléphone étendus (nombre de postes et capacité d'enregistrement). La plateforme téléphonique de M.I.D. sera équipée avec des matériels et des logiciels compatibles de manière à pouvoir mobiliser conjointement les deux dispositifs pour mener des enquêtes par téléphone de plus grande ampleur ou plus rapidement qu'il ne serait possible de faire avec une seule d'entre elles.

## 2.2.2 STRUCTURE AND BUILDING OF THE EQUIPMENT

### *(2 pages maximum)*

M.I.D. est un équipement d'excellence qui vise à collecter des données, à les documenter, à les préserver et à les diffuser en tirant profit des nouvelles possibilités et des coûts réduits que les nouvelles technologies de l'information et de la communication permettent.

### **Élément 1 - SRC (panel web et centre d'appel téléphonique)**

Il n'existe pas de structure équivalente en France. En Europe, seul CenterData, de l'Université de Tilburg aux Pays-Bas, qui gère le LISS panel s'approche de la composante SRC de M.I.D.. Marcel Das, directeur de CenterData, assistera en tant que consultant la mise en place pratique de l'équipement M.I.D.. Le SRC s'appuie principalement sur le Centre de données socio-politiques (CDSP), l'Ined et le GENES. Il sera localisé sur les trois sites de manière à renforcer les liens et les synergies entre les trois partenaires. La construction du SRC débutera par une phase d'expérimentation qui visera à tester et affiner en grandeur nature les options retenues pour le panel Web. Cette phase sera suivie du lancement du panel Web. Le lancement du centre l'appel téléphonique viendra dans un troisième temps lorsque les locaux dont disposera le GENES sur le campus de Sacaly seront livrés. Cette solution permettra de disposer de locaux modernes et destinés dès le départ à une utilisation comme plateforme d'enquête téléphonique sans que les coûts immobiliers soient supportés par le projet M.I.D..

### **Collecte de données**

Durant la phase d'expérimentation et de montée en charge du panel (durée approximative d'un an), passage progressif de 0 à 5000 panélistes.

Matériel : serveurs, terminaux mobiles, postes d'appel téléphoniques

Personnel : permanent à recruter: Un statisticien responsable de panel; deux gestionnaires de panel, un statisticien spécialiste des méthodes d'enquête, un ingénieur en informatique, un ingénieur d'études pour l'intégration des questionnaires. Ces personnels seront placés à l'Ined.

Personnel non permanent: 1 consultant (6 mois).

Dans la phase de production:

Au personnel précédent il faut ajouter 5 gestionnaires de panel, deux ingénieurs informaticiens, deux statisticiens spécialistes des méthodes d'enquêtes et un statisticien spécialiste des panels, un ingénieur d'études.

Deux statisticiens seront présents au CDSP, ainsi qu'un ingénieur d'études et informaticien. Ce plan d'occupation pourra varier suivant les calendriers de réalisation du Campus Condorcet, du centre de Saclay et des éventuels investissements immobiliers de Sciences Po. L'idéal serait de conserver toute la gestion du panel sur un site, avec les informaticiens et les statisticiens nécessaires.

Pour le centre d'appels téléphonique, il sera placé au Genes, à Saclay :

Personnel permanent à recruter : 1 responsable du plateau téléphonique, 20 enquêteurs, deux ingénieurs informaticiens, 2 gestionnaires superviseurs de collecte, deux statisticiens spécialistes des sondages, un gestionnaire du matériel.

### **Documentation de données**

La documentation des données collectées doit respecter la norme internationale DDI (Data Documentation Initiative). Elle nécessite des compétences en statistique et sciences sociales et une expérience d'analyse statistique de grandes enquêtes. Encadrement par le CDSP

Personnel : 2 chargés d'études

Localisation : Sciences Po (CDSP)

### **Diffusion de données**

Prise en charge par le CDSP

### **Élément 2 - BEquali**

L'archivage d'enquêtes qualitatives suppose un peu de matériel (serveurs, scanners, matériel informatique courant) mais surtout, beaucoup de temps et de compétence pour acquérir, préparer, documenter et diffuser les données. Qualidata, l'infrastructure modèle en matière de banque d'enquête qualitative, fonctionne avec 6 personnes dont trois sont titulaires d'un PhD; la toute jeune Irish qualitative data archive, officiellement ouverte en juin 2010, comprend 5 personnes, toutes docteur(e)s en sciences sociales. Ces effectifs sont tout à fait cohérents avec les estimations que nous avons pu faire en réalisant un prototype de base (<http://blogs.sciences-po.fr/recherche-donnees-qualitatives/>). Cette expérimentation nous a enseigné que la collecte, la documentation et le référencement des données d'enquêtes qualitatives constituent des tâches sensiblement différentes - et plus compliquées - qu'en matière quantitative ; surtout l'enrichissement des données apparaît comme une étape incontournable pour garantir un potentiel de réutilisation future.

### **Acquisition**

Le travail d'acquisition demande les compétences nécessaires à la fois pour intéresser et convaincre les collègues de mutualiser leurs enquêtes, ce qui suppose une réputation scientifique établie de la ou des personne(s) en charge de cette mission, et pour en évaluer la qualité. Même si in fine la décision d'archivage sera prise par un comité scientifique, un rapport préalable sur les enquêtes devra être préparé par un membre de BEQuali, faisant apparaître non seulement la qualité des travaux reposant sur les données, mais la qualité intrinsèque de l'enquête et son potentiel de réutilisation. Ce même rapport servira de base à l'enquête complémentaire ("enquête sur l'enquête") qui sera menée et archivée avec l'enquête pour renforcer la contextualisation des données, et prendra place dans la négociation avec le ou les auteurs de l'enquête.

Personnel: 2 ingénieurs de recherche

Matériel: postes de travail équipés et enregistreurs numériques

Moyens: frais de mission (visite des laboratoires pour inventaire, entretiens auprès des chercheurs donateurs, participation aux événements scientifiques permettant de sensibiliser les collègues aux enjeux de l'archivage et de l'analyse secondaire)

### **Préparation et documentation**

Le travail à faire sur les données varie fortement suivant leur ancienneté et leur état de conservation. Pour les enquêtes récentes, l'essentiel des données est déjà numérisé et suppose surtout l'anonymisation et la mise en forme de l'enquête en site web. Pour les enquêtes plus anciennes, conservées parfois uniquement sous format papier et potentiellement en mauvais état, ou conservées sous des formats informatiques obsolètes, la numérisation s'ajoute à la préparation des données ou à leur restauration. La nécessité de procéder à l'anonymisation de données complexes, car textuelles, oblige également à un travail d'immersion en profondeur dans les données.

Le travail archivistique de documentation des données exige une reconstitution minutieuse du contexte de l'enquête originelle, qui doit être si possible effectuée en partenariat avec le chercheur primaire. Il nécessite également d'extraire des métadonnées à partir directement des données, ce qui suppose un fort investissement dans l'appropriation des données de l'enquête et une bonne connaissance des méthodologies de la recherche en SHS.

L'intégration des données dans des "sites-enquêtes" exigent enfin une structuration poussée des données et donc de valoriser leurs relations, très probablement en XML (Muhr, 2000 ; Carmichael, 2002). La préservation des enquêtes archivées suppose également un formatage des données sous des formats portables et pérennes, de manière à assurer des capacités de conservation et d'échange sur le long terme et leur réutilisation sous d'autres applications.

Personnel: 2 ingénieurs d'étude; expert information (mission) pour la finalisation du modèle de site-enquête

Matériel: postes de travail équipés et scanner

Moyens: frais de mission (observation des protocoles mis en place dans les banques existantes en Europe ou aux Etats-Unis)

### **Diffusion**

La diffusion des enquêtes sera assurée par le CDSP, selon des procédures et des moyens adaptés aux spécificités des enquêtes qualitatives, dont l'anonymisation ne peut être que partielle, ce qui nécessite, pour des raisons tant éthiques que légales, des procédures plus strictes d'accès aux données que ce n'est le cas pour les enquêtes quantitatives.

En termes de calendrier, la phase de test du prototype n'étant pas encore finalisée, nous prévoyons de commencer par recruter un développeur informatique ainsi qu'un premier ingénieur de recherche ; ils achèveront ensemble de construire l'architecture de la banque tandis que l'ingénieur de recherche publicisera le dispositif tout en initiant le recensement et la collecte des enquêtes susceptibles d'être archivées. Il seront secondés par un ingénieur d'étude pour la partie logistique et archivistique des opérations. Une fois la banque structurée et éprouvée, à horizon d'un an, nous prévoyons de recruter un second ingénieur de recherche ainsi qu'un second ingénieur d'étude, qui seront formés par leurs homologues ; leur recrutement permettra de faire face à l'afflux d'activité généré par la mise en route du dispositif.

### **Élément 3 - Données web**

L'équipement de collecte de données web viendra compléter le médialab créé par Sciences Po en mai 2009. Si le médialab a pour objectif de développer les méthodes numériques pour la recherche en Sciences humaines et sociales, il ne dispose pas aujourd'hui des moyens suffisants pour assurer la collecte des données d'expression spontanée sur le web.

### **Acquisition**

La collecte de données d'expression spontanée du web demande la mise en place d'une équipe regroupant les compétences suivantes :

- sociologie des communautés web et de l'usage des médias sociaux
- technologies d'extraction et nettoyage des données web
- technologies et stratégie de crawl
- linguistique informatique
- base de données et datamining
- architecture logicielle de systèmes distribués
- administration système

Le système de collecte exploitera des ressources informatiques mutualisées avec les autres éléments de l'équipement dans une ferme de serveur et un centre de données. L'architecture logicielle sera composée d'éléments de récolte et d'extraction de données capable de se déployer sur une ferme de serveur assurant la parallélisation des calculs. Ce même nœud de calcul accueillera les algorithmes de traitement de données nécessaire au maintien des corpus et au fonctionnement de l'équipement. Les analyses plus profondes des corpus pourront profiter de collaborations avec d'autres équipements comme ceux existant au médialab de Sciences Po et plus particulièrement l'équipement d'excellence MUSCA. Le stockage et la préservation des données sera assuré par le centre de données dédié. Cet équipement sera situé dans le même bâtiment que l'actuel médialab à Sciences Po.

### **Documentation**

La notion de corpus web est en cours de normalisation dans le cadre des initiatives d'archivage du web. Le consortium IIPS a mis au point le format WARC<sup>20</sup>. Les corpus d'expression spontanée sur le web sont à mi-chemin entre les corpus d'archive web et les verbatims des données qualitatives. Il n'existe pas à ce jour de format de référence dans le domaine de l'exploitation du web en Sciences Humaines et Sociales. La création de l'équipement serait une initiative d'envergure qui permettrait d'adresser ce manque et de contribuer à l'élaboration d'un format commun avec les partenaires internationaux du domaine (GOVCOM de l'université d'Amsterdam notamment).

### **Diffusion**

Partageant les problématiques de respect de la vie privée et du droit d'auteur que le volet BEQuali de l'équipement, la diffusion des données sera assurée au cas par cas par le CDSP suivant les procédures établies.

## **2.2.3 ENVIRONNEMENT TECHNIQUE / TECHNICAL ENVIRONMENT**

**(3 pages maximum)**

### **Élément 1 - SRC (panel web et centre d'appel téléphonique)**

---

<sup>20</sup><http://netpreserve.org/press/pr20090601.php>

L'équipe panel sera accueillie à l'Ined: le responsable, 2 statisticiens, 2 informaticiens et les 7 gestionnaires du panel et 1 ingénieur d'études.

La salle d'appel téléphonique sera localisée au Genes, à Saclay avec le personnel et l'équipement nécessaires.

Deux statisticiens, un informaticien et un ingénieur d'études seront à Sciences Po.

La salle d'appel téléphonique

Accueil de l'équipe : surfaces nécessaires à Sciences Po, GENES et Ined ?

### Élément 2 - BEquali

Accueil de l'équipe : Il semble essentiel que BEQuali soit hébergé dans des locaux relevant du CDSP: toutes les archives qualitatives existantes sont développées au sein des archives quantitatives qui leur préexistent généralement. Cette connexion répond au développement que les méthodes des sciences sociales connaissent au niveau international, qui encourage la mutualisation ou le croisement des questionnements épistémologiques et techniques entre spécialistes des enquêtes qualitatives et quantitatives, et valorisent tout particulièrement l'usage combiné des deux (pratique dite de mixed-methods). Il importe que la France reste en phase avec cette évolution et ne reproduise pas plus avant l'opposition traditionnelle entre quantitativistes et qualitativistes.

### Élément 3 - collecte de données web

L'équipe de collecte de données web sera accueillie au médialab de Sciences Po. Elle permettra de bénéficier d'un contexte idéal entre technologie numérique et recherche en Sciences Humaines et Sociales. L'équipe du médialab participera au développement et la supervision de l'équipement y apportant son expérience et son réseau de partenaires.

### Élément 4 - Centre de données et ferme de serveurs

Les serveurs seront accueillis dans les locaux de la Direction des Systèmes d'Information de Sciences Po. Sciences Po dispose déjà d'une architecture réseau de pointe suite notamment aux investissements réalisés pour la création du médialab (nouveau coeur de réseau, connection des différents sites par fibre noire, mise en place de SAN).

La nouvelle ferme de serveur dédié à l'équipement sera composée d'un équipement de serveurs LAME permettant un encombrement minimum et meilleure efficacité énergétique. Cette ferme sera connectée au réseau Internet par une connection dédiée bénéficiant de plusieurs adresses IP publiques. Bien que le réseau de Sciences Po soit déjà très performant, une connexion dédiée est indispensable pour assurer un cloisonnement de la consommation de bande passante entre l'équipement et le reste du réseau.

Le centre de données sera réalisé par un SAN (Storage Area Network) permettant de mutualiser les espaces de stockage dans une architecture technique virtuelle assurant un haut débit d'accès et une sécurisation des données par réplication. Cette technologie est déjà utilisée et maîtrisée par la DSI de Sciences Po.

Les liens déjà très fort entre l'équipe du médialab et la DSI assureront la bonne intégration de l'équipement au sein de Sciences Po.

*Un engagement du(des) responsable(s) du(des) site(s) d'accueil confirmera que l'équipement pourra être installé et mis en service à sa livraison, et que les moyens nécessaires à l'accompagnement de son fonctionnement, y compris en termes de personnel, seront effectivement mis en place.*

### 3. STRATEGIE DE VALORISATION DES RESULTATS/ DISSEMINATION AND EXPLOITATION OF RESULTS

#### Diffusion des données et des résultats

Les nouvelles technologies de l'information et de la communication se trouvent au centre du projet M.I.D., pour collecter des données mais aussi pour les mettre à disposition de la communauté scientifique. Les données (panel, téléphone, qualitatives, web) collectées seront documentées selon les standard internationaux et mises à disposition sur le web avec des outils et interfaces innovantes. Le modèle de diffusion des données s'inscrit dans le mouvement de la recherche libre de droits (open science), encore relativement nouveau en sciences humaines et sociales mais en fort développement en économie, en science politique et en sociologie (voir le numéro spécial de la revue américain *Sociological Methods & Research* paru en 2007 sur la réplication et l'accès aux données et en particulier Freese 2007 et King 2007). Selon cette perspective, les données et procédures (par exemples programmes statistiques) font partie intégrante des résultats publiés dans des articles ou des livres et à ce titre doivent être documentées et accessibles, éventuellement selon certaines conditions (après période d'exclusivité, pour réplication seulement, etc.). Les mêmes questions sont posées pour les méthodes qualitatives, où l'économie des citations et l'impossibilité de consulter les données analysées est de plus en plus mise en cause (Moravcsik 2010). Un portail web présentant l'équipement d'excellence permettra d'accéder à l'ensemble des données collectées et archivées par M.I.D.. Des outils en ligne permettront de mettre en valeur les différents types de données gérés par M.I.D..

Après leur documentation selon le standard xml DDI (Data Documentation Initiative), les données d'enquêtes par questionnaire (panel web et téléphone) seront diffusées par le CDSP sur une [plateforme NESSTAR](#) qui permet d'accéder aux informations sur les enquêtes (méta-données), aux questions posées mais aussi de réaliser des premières analyses (recodage, tris à plat, tableaux croisés, graphiques) et de télécharger les données dans un des formats acceptés par les principaux logiciels de statistique (Stata, SAS, SPSS, etc.). Les micro-données d'une enquête pourront être téléchargées gratuitement une fois passée la période d'exclusivité décidée avec les auteurs de l'enquête et après enregistrement auprès du CDSP.

La coordination externe du module BEQuali, à savoir le développement et l'entretien de liens entre les partenaires de l'équipe et la communauté scientifique dans son ensemble, fera l'objet de tous nos soins, le succès du projet dépendant aussi de notre capacité à le faire connaître et accepter par nos collègues. Convaincre la communauté des avantages à attendre de la mutualisation des enquêtes ne pourra se faire que dans la durée, en s'accompagnant d'un fort effort de valorisation, notamment par la participation à ou l'organisation de nombreux événements scientifiques, des publications et le développement de modules d'accompagnement à l'enseignement des méthodes qualitatives.

Nous sommes actuellement inscrits dans une synergie au niveau national, avec notamment le projet REANALYSE qui servira à démontrer l'intérêt de l'analyse secondaire en qualitatif.

Nous avons également démarré la création d'une synergie au niveau international, à travers le CESSDA dont est membre le CDSP, mais aussi par le développement de contacts avec Qualidata ou FORS, de manière à développer de nouvelles solutions de mise en valeur des enquêtes (diffusion d'un modèle de sites enquêtes) ou d'établissement de standards de documentation et de référencement. Dans le cas de l'archivage des enquêtes qualitatives, aucune spécification comparable à la norme DDI pour les données quantitatives n'a en effet réellement été développée, qui permette de répondre aux caractéristiques propres à ces données. Des collaborations avec les autres centres de ressources partenaires du CESSDA (notamment Qualidata) sont à envisager pour mettre au point une telle spécification, ou du moins tester sa faisabilité, sur la base de la Text Encoding Initiative (TEI) et d'une initiative existante de Qualidata (projet DeXT, pour Data Exchange Tools ; <http://www.data-archive.ac.uk/dext/about/introduction.asp>)

## **Valorisation**

De manière générale, M.I.D. permettra une meilleure utilisation des fonds publics (ANR ou PCRD) qui sont investis dans les enquêtes en SHS puisque ces enquêtes seront moins chères, plus innovantes, au plus haut niveau d'excellence méthodologique, réutilisables pour différentes recherches, et que l'argent investi consolidera l'équipement d'excellence des partenaires du projets (notamment la recherche méthodologique). Les travaux déjà programmés par les différents partenaires devraient gagner en qualité et certains devraient voir diminuer leurs coûts dès lors que le centre d'appel ou l'internet seront mobilisés. À titre d'exemple, le projet de panel électoral du Centre d'études européennes de Sciences Po (enquête initiale de 20 minutes suivie de 12 réinterrogations de 10 minutes) pourrait être mené sur une plus longue durée, être réalisé dans de meilleures conditions méthodologiques, et pour un coût bien moindre que les 150 000 euros requis par un institut de sondage privé. À niveau de financement public stable, il sera donc possible de mener plus d'enquêtes et de meilleure qualité méthodologique. De la même manière, l'archivage, la documentation et la diffusion d'enquêtes qualitatives permettront de développer leur ré-analyse et, outre les économies réalisées du fait de l'utilisation plus extensive de ces données, elle permettra d'envisager des études comparatives dans le temps ou dans des contextes sociaux et culturels différents. Un exemple très récent est le livre du sociologue britannique Mike Savage (2010) qui a ré-analysé des enquêtes qualitatives menées par des sociologues anglais entre les années 1930 et 1960 pour saisir les transformations de la société britannique après la Seconde Guerre mondiale. Par ailleurs, la mise au point d'outils et de méthodes d'analyse du web et leur mise à disposition pour la communauté scientifique permettra une utilisation plus parcimonieuse des budgets de la recherche en limitant les outils redondants.

Les premières études d'envergure mobilisant les composantes de production d'enquête (panel et webcorpus) auront lieu dans la deuxième année de la mise en place du projet global. Elles pourraient être centrées sur deux thèmes assurant la synergie des moyens de collecte classique, web et panel. Le premier thème est l'étude des troubles du comportement alimentaire, centrées sur la recherche des facteurs associés de l'anorexie, boulimie, obésité et maigreur. Le suivi dans le temps des sujets, leurs interrogations répétées sur leurs pratiques alimentaires, leur budget alimentaire, la perception de leurs troubles, leurs difficultés scolaires ou professionnelles, etc. sera complété par l'étude des forums internet traitant de ces thématiques offrant un matériel qualitatif complémentaire original. Cette synergie des moyens de collecte et le caractère longitudinal de l'enquête fourniront des données inédites. Une étude similaire portant sur les consommations de substances psychoactives dans la population (alcool, tabac, cannabis et autres drogues illicites) offrira des informations inédites sur les trajectoires de consommation (rétrospectives et prospectives), leur lien avec les événements de vie (études, vie amoureuse et familiale, professionnelle, etc.). Un des axes de recherche de ces deux études sera l'exploration fine du gradient social ordinairement observé sur ces problématiques (les individus les moins favorisés socialement et économiquement présentant les troubles les plus importants) et ses anomalies, l'autre étant la mise au jour de facteurs de vulnérabilité (gravité des dommages induits par les troubles) en plus des facteurs d'exposition au risque (présence de troubles chez soi ou dans l'entourage).

Les recherches et avancées méthodologiques qui seront menées grâce à l'équipement M.I.D. profiteront également à l'ensemble des acteurs intéressés par la collecte de données et en premier lieu les instituts de sondage. La branche néerlandaise de TNS a ainsi fait évoluer certaines de ses procédures suite aux expériences menées dans le cadre du LISS panel. En effet, la recherche, notamment méthodologique, du LISS panel est disponible en libre accès sur le web. M.I.D. diffusera également par internet les résultats des expérimentations méthodologiques et pourra ainsi contribuer au débat entre acteurs publics et privés sur les méthodes de collecte et d'analyse de données. L'équipement d'excellence M.I.D. contribuera également à la compétitivité des entreprises du privé au travers des formations de haut niveau méthodologique qui pourront prendre place en son sein.

La valorisation des résultats sera opérée par plusieurs biais. D'abord par la publication d'articles scientifiques et la participation à des colloques spécialisés et par la création d'une plateforme Internet dédiée qui comprendra tous les matériels utiles à l'information de la communauté scientifique. Techniquement, des indicateurs de qualité des enquêtes seront produits et développés (mesures des biais, erreur totale, qualité des déclarations, etc.) qui documenteront à l'avenir les enquêtes produites et auront pour objectif de servir de standard pour la recherche en SHS. Les outils informatiques de gestion des données de collecte seront également mis à disposition de la communauté scientifique après un délai de test et de validation. Le dépôt de brevets est possible. Le développement de la formation universitaire est aussi une des retombées naturelles du projet. En effet, l'équipement et l'expertise méthodologique accumulée permettront également de mieux former les étudiants aux méthodes d'enquêtes en SHS et à leur renouvellement par les innovations technologiques récentes. Les étudiants et doctorants des partenaires du projet M.I.D. pourront être formés aux méthodes d'enquêtes par questionnaire par la pratique puisqu'une enquête du panel par an sera réservée à l'enseignement. La formation aux enquêtes par questionnaire bénéficiera également de la salle d'appel téléphonique qui sera doté de capacités d'enregistrement. L'archivage, la documentation, et la mise à disposition d'enquêtes qualitatives permettront d'approfondir les réflexions sur la validité du travail scientifique en méthodologie qualitative. Ils permettront aussi de favoriser le développement de l'analyse secondaire, une pratique qui reste encore marginale en France, et apporteront un soutien notable à l'enseignement de la méthodologie des entretiens, notamment dans la phase d'analyse et de traitement des données recueillies.

Mieux formés méthodologiquement et armés de l'expérience acquise par l'utilisation des services de l'équipement, ces étudiants seront plus compétitifs sur le marché du travail et pourront mener des thèses dans de meilleures conditions. Les expériences variées et à la pointe de l'innovation permettront aux étudiants de s'insérer facilement dans les entreprises de sondage, les cabinets d'études, etc. mais également de lancer leur start up. Les étudiants qui poursuivent des études de thèse mettront à profit ces enseignements pratiques pour mener des projets de recherche novateurs et au meilleur niveau méthodologique. Par ailleurs, un certain nombre d'enquêtes du panel web sera réservé à des projets doctoraux. Dans les deux cas, l'équipement M.I.D. permettra de mieux former les étudiants en thèse et à accroître plus encore leur compétitivité à l'international, aussi bien pour leurs publications que pour leur employabilité.

Pour EDF R&D le développement d'un équipement comme WeBSHS présente l'avantage de disposer d'un accès privilégié à une plateforme d'enquête répondant aux critères de la recherche scientifique (échantillonnage aléatoire, meilleur contrôle des conditions de passation, dispositif d'enquête expérimentaux etc.) à des coûts raisonnables. Il permet également de travailler en collaboration avec des équipes universitaires et industrielles sur des recherches méthodologiques liés à cet équipement (réseaux sociaux sur le web, analyse des données web, analyse des données textuelles etc.).

### **Stratégie de développement**

Le développement inévitable des enquêtes par et sur internet et de l'archivage et diffusion de données qualitatives par le web rend nécessaire le développement de plateformes comme M.I.D.. Ce nouveau type d'infrastructure devra nécessairement prendre place dans le cadre du European Strategy Forum on Research Infrastructures (ESFRI) qui structure, coordonne et finance en partie la mise en place d'infrastructures européennes de recherche. À l'heure actuelle, les deux principales infrastructures en SHS sont le réseau européen des centres de données (CESSDA) et l'enquête sociale européenne (ESS). Fondé dans les années 1970, le CESSDA coordonne au niveau européen les centres qui archivent, documentent et diffusent des données d'enquêtes par questionnaire. Plus récente (créée en 2001), l'enquête sociale européenne est menée tous les deux

ans en Europe selon une méthodologie très stricte, excellence qui lui a valu en 2005 le prix Descartes qui couronne les Outstanding Scientific and Technological Achievements Resulting from European Collaborative Research.

Avec le panel web de M.I.D., la France prend une position stratégique pour la 2e génération d'infrastructure de recherche européenne qui est en train d'émerger. Autour du LISS panel de l'université de Tilburg est en train d'apparaître un embryon de réseau européen de panel web (Université de Mannheim et GESIS en Allemagne, Université Catholique de Louvain en Belgique, Université de Lausanne et FORS en Suisse). Ce réseau permettrait de mener des études comparatives en Europe pour un coût beaucoup plus faible que l'ESS et avec une méthodologie tout aussi rigoureuse. L'exemple de l'ESS est aussi intéressant puisqu'il montre qu'un projet de collecte de données innovant, au plus haut niveau méthodologique et récent peut dans le cadre d'ESFRI prétendre à des financements européens en tant qu'infrastructure de recherche. Bien que les modalités de co-financement des infrastructures de recherche européennes (European Research Infrastructure Consortium) ne soient pas encore arrêtées, il devrait prendre la forme d'une participation au financement des équipements nationaux et de la structure de coordination. Les financements qui pourraient être ainsi obtenus permettraient à M.I.D. d'assurer son développement et de se maintenir au meilleur niveau scientifique, mais aussi d'être compétitif pour devenir la structure de coordination européenne. Les dix années de financement permettraient ainsi à la France de prendre un positionnement stratégique au niveau européen en la dotant d'un équipement innovant qui permet de mener des recherches, notamment méthodologiques, qui lui donnerait toute la légitimité scientifique nécessaire pour être le leader européen.

Enfin, l'expertise méthodologique qui va être développée au travers de M.I.D. pourra être mobilisée dans le cadre de projets européens collaboratifs ou individuels du 7e PCRD. Outre l'accueil de doctorants et post-doctorants Marie Curie, M.I.D. pourrait être associé à des projets de recherche comparatifs en Europe ou répondre à des appels d'offres européens (DG, ESF, etc.). Financé par une ERC Starting Independent Researcher Grant (2010-2014 1,5 millions d'euros), le projet Trust de Yann Algan sur la coopération et les mécanismes de confiance mobilise et combine enquêtes par questionnaire, entretiens qualitatifs et analyse des discours du web. Dans les deux cas (financement en tant qu'infrastructure ou partenaire de projets européens), les financements européens pourraient être des relais de croissance considérables pour M.I.D. à moyen et long terme.

## **4. PROJECT MANAGEMENT**

### **4.1. MANAGEMENT**

#### **4.1.1 RELEVANT EXPERIENCE OF THE PROJECT COORDINATOR**

*(1 page maximum)*

Given the size of the consortium, the multisite organization and the ambition of the project, M.I.D. has to implement an efficient management structure that enables decision making and flexibility in order to ensure the best collaboration possible between the different activities and teams.

The management structure is based on the experience of an institution and the expertise of a scientific director.

Sciences Po will be in charge of the administrative, legal and financial management. Involved in more than 35 European projects and 18 ANR projects, this top-ranking University of research in the social sciences and humanities in France has developed a solid and fruitful experience in projects' management.

The scientific director will be Laurent Lesnard, Director of the Centre de données socio-politiques (CDSP), Senior Research Fellow at the National Scientific Research Centre (CNRS) working at the Observatoire sociologique du changement (OSC) at Sciences Po. *He is director of several research or projects linked to XXX. Indiquer tes « fonctions significatives », tes responsabilités dans des projets ou des comités importants etc. Attention aucun site internet te mentionnant n'est à jour. Il faut absolument que les évaluateurs puissent s'y référer et qu'on tes nouvelles responsabilités apparaissent partout.*

Laurent Lesnard will manage the scientific issues of M.I.D. and will chair the *Steering Committee*.

He will be helped by an executive director to be recruited in charge of the daily follow up of the project, in particular the administrative, legal and financial issues.

In addition to the scientific director and the executive director, the **Management Team** will be strengthened by:

- A lawyer responsible for the elaboration and the implementation of a consortium agreement, the negotiation of grant agreements and contracts, the monitoring of all legal aspects related to the collection, dissemination and exploitation of data;
- A marketing manager, responsible for business development of M.I.D. in France and abroad.

An administrative and financial management will enhance the team once the equipment is set up.

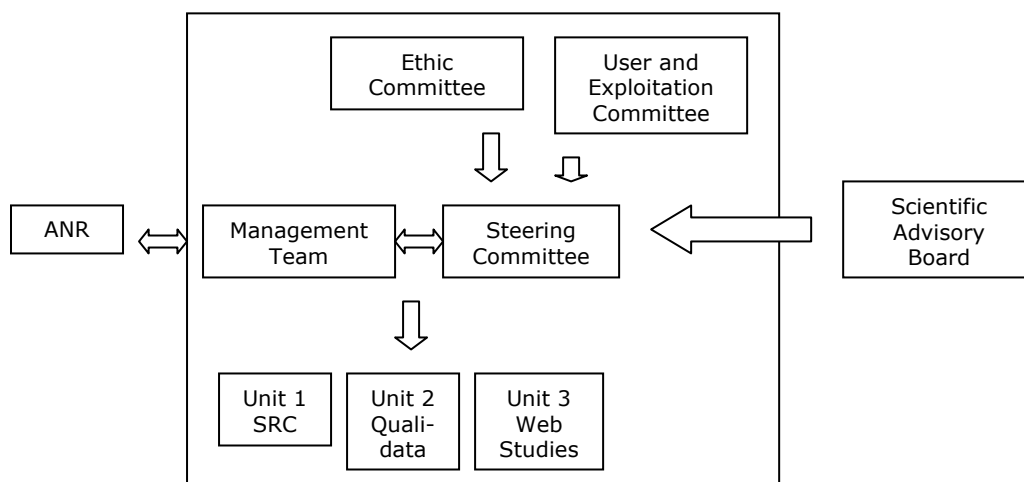
The Management Team will ensure the overall management of the project:

- follow up of the administrative, legal and financial management;
- administration and control of the budget (a contact by partner will be identified);
- overall responsibility for the relationships between the partners of the project and the ANR;
- project supervision and reinforcement of progress according to the time schedule;
- updating of the working plan and related budget with the agreement of the *Steering Committee*,
- preparation of the annual reports and final report;
- preparation of the meetings (decisions, agenda), draft the minutes and implementation of decisions taken;
- ensuring exchange between the partners of relevant information received in relation to the project, as well as information regarding the results achieved within it;
- development of the activity of M.I.D.

#### 4.1.2 COORDINATION MODALITIES

(1 page maximum)

The management structure of M.I.D. will be composed as follows:



### **Management Team**

The composition and role of the Management Team is described above.

### **Steering Committee**

The Steering Committee is the decision-making body of M.I.D.

He supervises the methodological orientations as well as the functioning of the equipment. He assesses the collection requests and the research projects, modifies the work plan and the budget on the basis of the UEC and SAB's opinions.

The Steering Committee is composed of: one representative of each partner, one representative of the *Comité de concertation pour les données en SHS* (CCDSHS), one representative of the Insee and the leaders of the units.

### **Units**

The three units, one for each component of the equipment, are in charge of the proper implementation of the project.

### **Scientific Advisory Board**

The Scientific Advisory Board (SAB) advises the Steering Committee and the Management Team about the methodological orientations, the functioning of M.I.D., the projects of future development and its scientific results. The SAB is composed of experts who have international reputation.

### **Ethic Committee**

The Ethic Committee ensures that the ethic rules and the anonymity of the data collected and spread by M.I.D. are respected. Keeping the anonymity of the participants to the web panel is a very important issue and will be the subject of a detailed follow up by the Ethic Committee in relation with the CNIL. The fact that the qualitative inquiries will be anonymised raises many questions of which answers will have to be validated by the Ethic Committee.

### **User and Exploitation Committee**

The User and exploitation Committee will be in charge of estimating the effects of the inquiries related to data collection and of the projects proposed to M.I.D. The analysis will rely on a study of their potential in terms of exploitation of the results (see §3 below) and feedback of M.I.D.'s users. It will welcome, among others, representatives of firms sensible to these questions.

### **Rules for use of equipment**

The services proposed by the equipment are available to any person, team, administration, local authority, firm, association etc whatever its geographical localisation provided that the request has a scientific interest as well as an interest in terms of exploitation of results and that it is adapted to the equipment. The use requests of M.I.D. will be selected by the Steering Committee, from the recommendations of the SAB and the UEC, regardless of the affiliation of the holder (public or private).

The fee of these services will be determined in the equipment test phase. It should cover the inquiry expenses, the equipment functioning and depreciation expenses. The activities, services and selection criteria of the projects will be presented in the website of M.I.D.

Part of the inquiry phase of the panel and the call centre will be reserved to partners, for their research project, their methodological projects as well as for the teaching of methods to their students.

A consortium agreement governing the shared rights and obligations of the partners will be implemented during the first stage of the project. The issues of governance, management of the budget or intellectual propriety will be detailed.

## 4.2. COLLABORATION ORGANIZATION

### 4.2.1 PARTNERS DESCRIPTION , RELEVANCE AND COMPLEMENTARITY

**(0,5 page maximum par partenaire)**

*Le reste Sera complété directement en anglais*

#### EDF R&D

Le département ICAME est l'un des 16 départements de la R&D du groupe EDF. Sa vocation est d'accompagner les directions opérationnelles en proposant de nouvelles offres de services et de fourniture, et cela pour l'ensemble des clients qu'ils soient particuliers, professionnels, industriels ou des collectivités locales.

Les nouvelles offres sont conçues d'après l'observation et la compréhension des clients sous leurs différentes facettes, qu'ils soient consommateurs ou citoyens.

Pour remplir au mieux cette mission, le Département ICAME s'est doté des compétences suivantes :

- Des compétences en sciences humaines et sociales avec la présence de sociologues, d'anthropologues, de sémiologues. C'est essentiellement là que se concentre la connaissance client autour des tendances de la société, des liens entre les consommateurs et l'entreprise, des observations sur les problématiques environnementales.
- Des compétences techniques pour la conception, l'élaboration d'offres de services dans un objectif d'appropriation des différentes technologies qui serviront de support pour porter le service jusqu'au consommateur.
- Des compétences d'intégration qui puisent dans la connaissance client et dans les technologies disponibles pour assembler les offres tant de services que de prix.
- L'innovation est également mise en exergue avec des compétences dédiées à élargir la capacité à innover par le biais de méthodes de créativité et par la connaissance de différentes sources innovantes.
- Et enfin, des compétences en statistique et en système d'informations dont la mission est d'étudier les outils et les méthodes novatrices qui faciliteront demain l'accès et la connaissance des données client dans les systèmes d'information.

**(1 page maximum)**

*idem*

### 4.2.2 QUALIFICATION, ROLE AND INVOLVEMENT OF INDIVIDUAL PARTNERS

*idem*

Partenaire/Partner	Nom/Surname	Prénom/First name	Poste/Position	Discipline/Domain	Organisme de rattachement ou entreprise/Organization or company	Rôle dans le projet (4 lignes max.) / Contribution in the project (4 lines max)
Exemple	LATIFI	Fatima	Professeur		CNRS	Gestion de

						l'installation de l'équipement
Coordinateur						

## 5. FINANCIAL ASSESSMENT

Sera complété ensuite directement en anglais

## 6. APPENDICES

### 6.1. STATE OF ART REFERENCES

Quel ordre ? Alphabétique, chronologique, ordre de citation dans le texte ?

Billordo, Libia. 2005 'Publishing in French Political Science Journals: an Inventory of Methods and Sub-fields', *French Politics* 3(2): 178-186.

Braibant, Guy, 1996. *Les archives de France*. Paris: La Documentation française.

Caveng, Rémy. 2009. "Inversement des positions et ré-enchantement de l'interaction". *Actes de la recherche en sciences sociales* 178:88-97.

Caveng, Rémy. 2007. "La "liberté" des vacataires en instituts de sondages". *Interrogations*, 4:59-77.

Cornu, Marie. 2003. « Droit des biens culturels et des archives » ([www.educnet.education.fr/chrgt/bienculturels.pdf](http://www.educnet.education.fr/chrgt/bienculturels.pdf)).

Couper, Mick P. 2000. "Web surveys: A review of issues and approaches." *Public opinion quarterly*. 64 (4):464.

Cribier, Françoise (avec Feller, Elise). 2003. *Projet de conservation des données qualitatives des sciences sociales recueillies en France auprès de la « société civile »*, Rapport présenté au ministère délégué à la Recherche et aux nouvelles technologies, Paris, CNRS-EHESS, LASMAS.

Dargentas, Magdalena, Brugidou, Mathieu, Le Roux Dominique, Salomon Annie-Claude (eds.). forthcoming. *L'analyse secondaire en recherche qualitative : une nouvelle pratique en sciences humaines et sociales*. Paris: Lavoisier.

Descamps, Florence. 2007. « L'entretien de recherche en histoire : statut juridique, contraintes et règles d'utilisation ». *Histoire@politique. Politique, culture et société*. 3.

Demazière, Didier, Dubar, Claude, Leimdorfer, François, Reinert, Max, Wald, Paul. 2001. *Rapport sur la faisabilité d'une banque de données qualitatives en sciences humaines et sociales*, Laboratoire Printemp, Université Versailles Saint-Quentin.

Duclert, Vincent. 2002. « Archives orales et recherche contemporaine. Une histoire en cours », *Sociétés et représentations*. 13: 69-86.

Duclos, Denis and Hélène Y. Meynaud. 1996. *Les sondages d'opinion*. Paris: La Découverte.

Finch, Caleb E., James W. Vaupel, and Kevin Kinsella. 2001. *Cells and Surveys: Should Biological Measures Be Included in Social Science Research?* Washington: National Academies Press.

Freese, Jeremy. 2007. "Replication Standards for Quantitative Social Science: Why Not Sociology?" *Sociological Methods & Research*. 36 (2):153.

« Sciences sociales : archives de la recherche ». 2006. *Genèses*. 63.

Kaufmann, Jean-Claude. 1996. *L'entretien compréhensif*. Paris, Nathan, coll.128.

King, Gary. 2007. "An introduction to the Dataverse Network as an infrastructure for data sharing." *Sociological Methods & Research*. 36 (2):173.

King, Gary, Keohane, Robert, Verba, Sydney. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Mack, Stephen, Vicki Huggins, Donald Keathley, and Madhi Sundukchi. 1998. *Do Monetary Incentives Improve Response Rates in the Survey of Income and Program Participation?* Presentation. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Mallet-Poujol, Nathalie. 2004. « Protection de la vie privée et des données personnelles ». *Legamedia*. ([www.educnet.education.fr/chrgrt/guideViePrivee.pdf](http://www.educnet.education.fr/chrgrt/guideViePrivee.pdf)).

Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research". *Political science & politics*. 43(1): 29-35

Pagès, Fabienne, Marianne Tribel, and Alexis Bonis-Charancle. 2006. "Indélicatesse et manque de rigueur dans les sondages. Enquête auprès des professionnels du sondage." *Bulletin de Méthodologie Sociologique*. 89:31-48.

Raento, Mika, Antti Oulasvirta, and Nathan Eagle. 2009. "Smartphones: An Emerging Tool for Social Scientists." *Sociological Methods & Research*. 37 (3):426-454.

Savage, Mike. 2010. *Identities and Social Change in Britain since 1940: The Politics of Method*. Oxford: Oxford University Press.

Silberman, Roxane. 1999. *Les sciences sociales et leurs données*, <http://www.education.gouv.fr/cid1925/les-sciences-sociales-et-leurs-donnees.html>.

Singer, Eleanor and Richard A. Kulka. 2002. "Paying respondents for survey participation." Pp. 105-127 in *Studies of welfare population: data collection and research issues, Studies of welfare populations: data collection and research issues*, edited by Michele Ver Ploeg, Robert A. Moffit, and Constance F. Citro. Washington: National Academies Press.

Strauss, Anselm. 1992. *La trame de la négociation. Sociologie qualitative et interactionnisme*. Paris, L'Harmattan.

Toepoel, Vera, Corrie Vis, Marcel Das, and Arthur Van Soest. 2009. "Design of Web Questionnaires: An Information-Processing Perspective for the Effect of Response Categories." *Sociological Methods & Research*. 37 (3):371-392.

Beauchemin, C., C. Hamel, et al. (2010). "Les discriminations: une question de minorités visibles. Population et Sociétés." *Population et sociétés*(466): 1-4.

Insee (2006). *Sans-domicile*. Paris: Institut national de la statistique et des études économiques.

Laflamme, F. and C. Mohl (2007). *Research and Responsive Design Options for Survey Data Collection at Statistics Canada: American Statistical Association*, .

Laporte, A. and P. Chauvin (2010). *La santé mentale et les addictions chez les personnes sans logement d'Île-de-France*. Paris, Observatoire du samu social.

Lefèvre, C. and A. Nourry, Eds. (2006). *Enquêter en sociologie et en démographie : questions de déontologie, de méthodes et de terrain, Bibliographie critique*. Population.

Lefèvre, C. and J.-M. Rohrbasser (2009). *Ethical and demographic research: general reflexions and situation in France*. XXVI International Population Conference, Marrakech, .

Lelièvre, E. and G. Vivier (2001). "Evaluation d'une collecte à la croisée du quantitatif et du qualitatif : l'enquête « Biographies et entourage »." *Population* 56(6): 1043-1074.

Massey, J. T., C. Wolter, et al. (1996). *Optimum Calling Patterns for Random Digit Dialed Telephone surveys: American Statistical Association*.

Mayet, A., S. Legleye, et al. (2009). *A multi-state modelling of transitions between tobacco and cannabis uses among adolescents in France: progression from onset to daily use*. Annual Meeting of the International Society of Addiction Medicine (ISAM). Calgary.

Mayet, A., S. Legleye, et al. (2010). "The mediation role of licit drugs in the influence of socializing on cannabis use among adolescents: A quantitative approach." *Addict Behav* 35(10): 890-5.

Quaglia, M. and G. Vivier (2006). *Enquêter des populations difficiles à joindre : principes méthodologiques et adaptations pratiques. Méthodes d'enquêtes et Sondages. Pratiques européennes et nord-américaines*. P. Lavallée and J.-P. Rivest. Paris: 140-146.

Quaglia, M. and G. Vivier (2010). "Construction and field application of an indirect sampling method (Time-Location Sampling): an example of surveys carried out on homeless persons and drug users in France." *Methodological Innovations Online* 5(2): 16-25.

Razafindratsima, N., M. Mazuy, et al. (2006). *La déperdition dans l'enquête sur les "Intentions de fécondité"*. Méthodes d'enquêtes et sondages : pratiques européenne et nord-américaine. P. Lavallée and L.-P. Rivest Paris, Dunod: 23-27.

Reedman, L. and M. Robinson (1997). *An Improved Call-Scheduling Method Using Call History and Frame Information: American Statistical Association*.

Rzafindratsima, N. and M. Marpsat (2010). "N° spécial sur l'échantillonnage auprès de populations difficiles à joindre (TLS, RDS, Capture-Recapture)." *Methodological Innovations Online* 5(2).

Strauss, Anselm. 1992. « Une perspective en termes de monde social », in Baszanger, Isabelle (ed.). *La trame de la négociation. Sociologie qualitative et interactionnisme*. Paris: L'Harmattan

Toulemon, L. and N. Razafindratsima (2008). *L'apport d'un échantillon de possesseurs exclusifs de téléphones mobiles dans l'enquête "Contexte de la sexualité en France" (2005-2006)*. Paris, Dunod: 249-253.

Vivier, G., P. Issenhuth, et al. (2010). Peut-on suivre, en longitudinal, une population jeune et mobile ? Les apports de l'enquête ELAP. Actes du Colloque Sondages de la Société Française de Statistiques, Tanger, Mars 2010., Paris, SFDS (to be published).

Benvegna N., Brugidou M., « Prendre la parole sur Internet : des dispositifs socio-techniques aux grammaires de la discussion », *Réseaux*, septembre 2008, 51-82.

Brugidou M. et Moine M. « Normes émergentes et stigmatisation- Une analyse comparative à partir des deux questions ouvertes sur les raisons de ne pas trier les déchets et de ne pas faire d'économie d'énergie » *JADT 2010 Rome*, 10 juin 2010

Brugidou M. et Moine M., « Le public en campagne. Analyse du discours des publics de la campagne électorale présidentielle à travers une séquence de question ouverte et fermée », *JADT 2008- Lyon*, 13 mars 2008.

Brugidou M., *L'opinion et ses publics*, Presses de Sciences Po, collection référence, (2008)

Brugidou M. et Escoffier C., « Questions ouvertes et opinion publiques discursives », in Marc X. et Tcherna J.-F. (dir.), *Etudier l'opinion*, PUG, Grenoble, 2007, pp. 91-111.

Brugidou M., « Vers une représentation discursive de l'opinion publique », *Corpus*, 4, 2005

Brugidou M. « Argumentation and Values : an analysis of ordinary political competence via an open-ended question », *International Journal of Public Opinion Research*, V15 n°4, winter 2003.

Brugidou M. , Cautrès B. , « Compte-rendu de la journée d'études du cidsp, les nouvelles méthodes d'analyse des entretiens- Analyse assistée par ordinateur et capitalisation des entretiens non-directifs de recherche 9 MARS 2001 - Maison des sciences de l'homme - Grenoble. *BMS*, 2001.

Brugidou, M. , « La combinaison des inférences statistiques, linguistiques et sociologiques dans l'analyse d'une question ouverte », *Journal de la Société Française de Statistique*, , 2001, 4.

Brugidou M., Labbé D., « Le discours syndical français contemporain (CFDT, CGT, FO en 1996-98) », *CERAT*, 2000.

Brugidou M., Escoffier C., Folch H., Le Roux D., Morin-Andréani P., Piat G., « Les facteurs de choix et d'utilisation de logiciels d'analyse des données textuelles », *JADT*, Lausanne, 2000.

Brugidou M., « Epitaphes, l'image de François Mitterrand à travers l'analyse d'une question ouverte posée à sa mort », *Revue Française de Science Politique* , vol. 48, n°1, 1998 p.97-120.

Clavel C. et Richard G., *Reconnaissance acoustique des émotions*, in *Systèmes d'Interaction Emotionnelle*, 2010, chap. 5, Hermès.

Dargentas, M., & Le Roux, D. (2005). Potentials and Limits of Secondary Analysis in a Specific Applied Context: The Case of EDF-Verbatim. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6 (1). Article en ligne : <http://www.qualitative-research.net/fqs-texte/1-05/05-1-40-e.htm>

Kuznik .L, Guénet A.-L , Anne Peradotto et Chloé Clavel. L'apport des concepts métiers pour la classification des questions ouvertes d'enquête, TALN 2010

Le Roux, D. (1998). *Verbatim, capitalisation des entretiens au GRETS* (N° Rapport interne H-N51-1998-0003) Clamart : EDF R&D.

Le Roux, D., & Vidal, J. (2000a). *Verbatim: Qualitative Data Archiving and Secondary Analysis in a French Company*. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1 (3). Article en ligne : <http://www.qualitative-research.net/fqs-texte/3-00/3-00rouxvidal-e.htm>

Le Roux, D., & Vidal, J. (2000b). *Verbatim, une expérience de capitalisation d'entretiens qualitatifs*. *Bulletin de Méthodologie Sociologique*, 65, 58-67.

Le Roux D., Brugidou M., Dargentas M., Developing Computer-Aided Secondary Analysis : consequences of such an Innovation for Sociologists in an Industrial Context, Impacts of software use on methodology and analysis, *CAQDAS 2007 Conference "Advanced in Qualitative Computing"*,

**APPEL A PROJETS EQUIPEX /  
CALL FOR PROPOSALS**

**2010**

**Acronyme du projet /  
Acronym**

**M.I.D**

**DOCUMENT SCIENTIFIQUE B /  
SCIENTIFIC SUBMISSION FORM B**

Royal Holloway University of London, 19, 20 avril 2007.  
<http://caqdas.soc.surrey.ac.uk/conference/abstracts.htm>

**6.2. PARTNERS' REFERENCES**

*Sera complété ensuite*

**6.3. ESTIMATE FOR THE EQUIPEMENT**

*Idem*