Data, infrastructures and survey methods in the humanities and social sciences

Assessment and outlook - 2014

INVESTISSEMENTS D'AVENIR

CONTENTS

Introduction

Data, infrastructures and survey methods in the humanities and social sciences (DIME-SHS) is a project that aims to close the development gap in French humanities and social sciences with respect to survey methodology and data collection[1]. DIME-SHS is a survey research centre that takes advantage of new technologies to offer tools for the social science community to collect and disseminate data. DIME-SHS is structured around three instruments:

- DIME Quali: establishment of a qualitative survey database to enhance secondary analysis;
- DIME Quanti: quantitative data collection through questionnaires complemented by innovative protocols (mobile Internet panel and call centre);
- DIME Web: collection and analysis of spontaneous expressions on the web.

DIME-SHS was selected in 2011 in the context of the Forward-looking investment programme's first "equipment of excellence" (équipex) call for projects. Coordinated by Sciences Po, DIME-SHS draws on the expertise of a consortium of research and higher education institutions: Paris Descartes University, National institute for demographic studies (Ined), School for advanced studies in the social sciences (EHESS), Telecom ParisTech engineering school, Grouping of national economics and statistics schools (Genes), Quetelet network (Very large research infrastructure Progedo) and EDF research and development (EDF R&D). DIME-SHS has received 10.4 million euros in funding for the 2011-2019 period.

DIME-SHS has three governing bodies. The project's Executive Coordinator is responsible for ensuring that the project proceeds smoothly. It consists of the project coordinator, Laurent Lesnard; the head of the CDSP's information technology department, Geneviève Michaud; the CDSP's executive secretary; and the heads of the three instruments: Guillaume Garcia (quali), Anne Cornilleau and Anne-Sophie Cousteaux (quanti) and Mathieu Jacomy (web). The Steering Committee, which brings together the directors or presidents of the consortium's institutions, makes decisions on the overall direction of the project. Finally, the Scientific Council, composed of internationally recognized independent experts, provides scientific guidance to the Coordinator and Steering Committee (for more details, see p. 2). At least three of the Scientific Council's members must be experts on ethics.

[1] Silberman, Roxane. 1999. *Les sciences sociales et leurs données [The social sciences and their data]*, http://www.education.gouv.fr/cid1925/les-sciences-sociales-et-leurs-donnees.html

DIME Quali

Presentation of the instrument

The creation of this instrument was motivated by the need to provide France with a qualitative survey database in political science and sociology. This need was underscored in several reports at the end of the 1990s and beginning of the 2000s. The primary goal was to bring France to the same level as other European countries, especially the English-speaking ones, in terms of compiling data from surveys conducted with qualitative methodologies (interviews, observations, etc.). From the beginning, there were many objectives behind this creation. The provision of data is meant to give the scientific community the means to more completely tap into the wealth of data from surveys, which are often under-analysed, and to make it easier to compare (across time, space and social groups). Another goal is to improve research training via the teaching of methods based on real and proven data, in line with the "data in the classroom" method. Finally, survey sharing has an epistemological purpose: fostering transparency in field practices and in the implementation of methods enhances the scientific rigor of the qualitative approach and makes a positive contribution to the history of the social sciences.

These objectives are hardly achievable without adequate equipment, and developing the instrument will directly address this need. The priority objectives are to encourage secondary analysis (or reuse) of archived surveys and to support the training of future researchers. Indeed, these objectives will have the most direct and immediate impact on research practices and the general academic landscape. Objectives with a more epistemological or historical purpose are certainly important, but will only be achieved in the longer term. For this reason, we built and configured the instrument around short-term objectives, leading us to make some particular choices about the instrument. We deliberately chose to create a scientific tool that prioritises making surveys available to enable quick reuse. In other words, we are not positioning ourselves as an archive service devoted to the history or epistemology of the social sciences. The instrument is therefore organised by survey only – that is, laboratory and researcher documents will not be archived, but only the actual surveys.

At the scientific level, our approach seeks to build and offer tools to secondary researchers allowing them to understand the context of carrying out surveys. The idea is to give researchers the means to reuse the surveys in a valuable way. This concern is reflected in the implementation of two guiding principles that we will elaborate later. First, we collect and provide raw data. Then we attempt to gather and integrate as much of the documentation as possible to shed light on the research process; it is in this sense that we disseminate surveys. Next, we create and provide access to a "survey of the survey" for secondary researchers. The survey of the survey aims to help retrace the original survey process in order to reduce the risk of decontextualized use of the data. In addition, each survey is organized and presented as a mini website. The idea behind the "study-site" is to facilitate navigation through the various archived documents and foster familiarity with the material. Thus, the instrument is intended as a toolbox to help users become familiar with the survey before they begin working with the raw data.

Finally, we have also worked to reach a wide audience of secondary researchers, teachers and students. From the very beginning we aimed to link the instrument to the Quetelet network's data dissemination portal; the goal here was both to avoid reinforcing the quantitative/qualitative divide and to ensure broad and secure dissemination of survey documents.

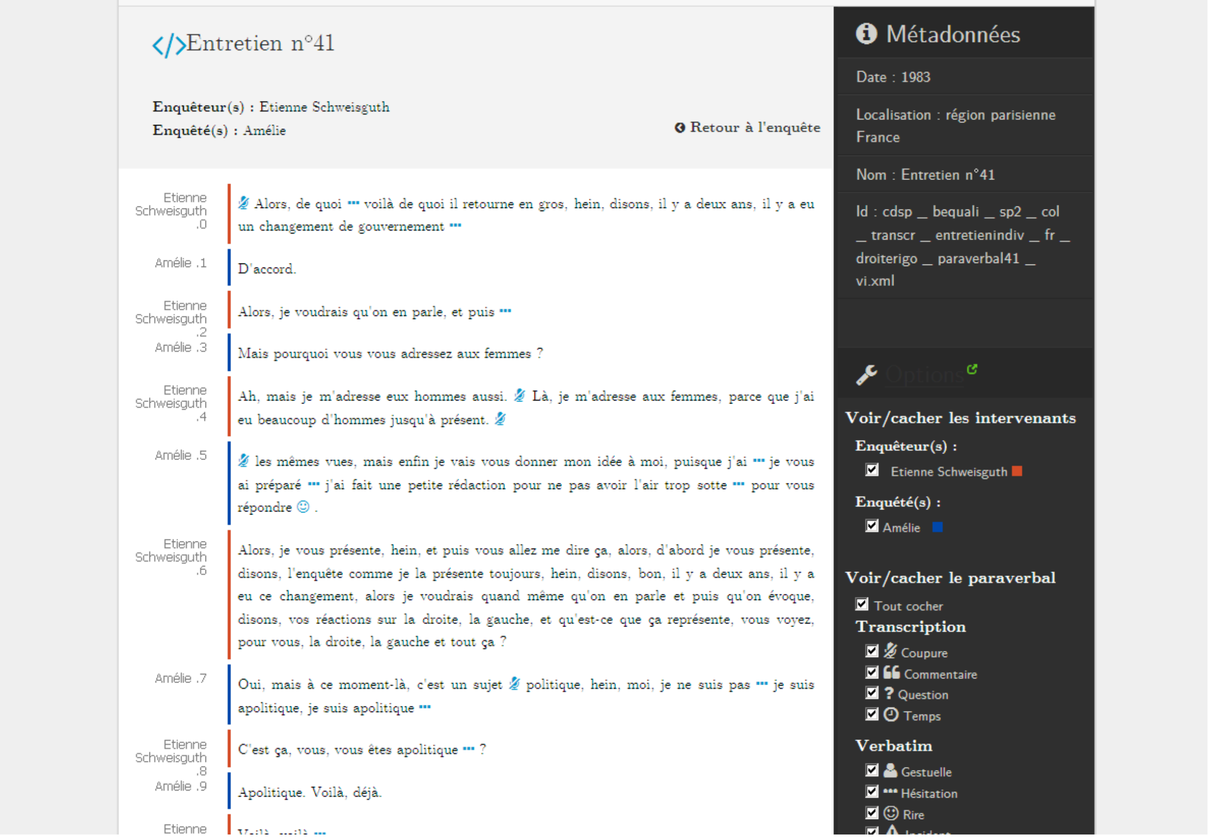
The project had been designed several years before the 2010 Equipex call for projects. There had been previous discussion in France on the secondary analysis of qualitative surveys in connection with questions about the equipment needed to capitalize on this type of data. A first step was taken in November 2005 at a symposium organized in Grenoble, of which the proceedings have not been published to date. Several years later, the CDSP conducted tests on archiving qualitative data in cooperation with researchers at Cevipof, one of Sciences Po's laboratories. Beginning in 2009, a more general feasibility study was completed to take stock of the situation across all the fields involved (scientific, ethical and legal, technical and strategic). Sciences Po, in partnership with the CDSP, ran a first experiment on the basis of 3 surveys selected for the diversity of the issues they raised in terms of archiving and documentation. These are the three first surveys that were made available via the instrument (see infra). In 2010 the ANR provided funding for the réAnalyse project that allowed us to begin conceptualising – and building – a prototype website in line with the aforementioned objectives. It should be noted that at this stage the goal was not yet to create lasting equipment for the whole scientific community. Rather, it was to test the usefulness of a number of tools through several secondary analysis projects, using an experimental approach; each of these projects raised specific issues allowing us to identify archiving and availability conditions necessary for a convincing reuse of data. The DIME-SHS quali instrument was derived from all of these tests and experiments, and the discussion that ensued. However, it only includes some of the initially imagined or tested options and solutions, as described below.

Status of the project

The first development phase of the equipment started (at the very beginning of 2012) as the prototype website created for the ANR réAnalyse project was being completed. The work therefore initially consisted of ensuring the prototype website's development, from delivering the application to debugging. An attempt was made to use the prototype and move to a production version; it revealed that the website needed a major overhaul. However, we had to wait about one year before we were able to recruit a dedicated developer. During this period we had to make up for the lack of a developer by using two successive developers on short assignments. By the end of this period we were only able to reach some of the goals: to improve the survey overview page, the survey of the survey module, the faceted search engine for exploring survey documents, and the tools to edit and administer the website. The developer hired in February 2013 then had to take the time to master his predecessors' work in order to continue adapting the website, called *enQuêtes* [*SurVeys*]. He also worked on the viewers for posting survey documents (i.e. files in PDF, TEI & CSV format). While waiting for the Quetelet monitoring application to be delivered (it was in the spring of 2014) to enable the integration of qualitative surveys, we had to develop a temporary application to serve as a registration interface for the website and a control and file download interface for users. These constraints played a large role in delays to making the public website live.

They also explain why we had to drop the implementation of several functions that we had tested in the prototype: specifically, what we had referred to as the “visualisation” function, as well as the “stakeholders” page. In the prototype version these two interfaces were designed to enable online exploratory analysis of the interview transcripts and of the respondents’ socio-graphical characteristics. More generally, these visualisation tools were meant to help researchers elaborate first hypotheses, according to the principles of exploratory data analysis.

Besides this first set of computing operations, we also had to develop other dimensions involved in building the equipment. Thus a second set of operations focused on archiving issues. The goal first consisted of connecting the instrument to the archive world by establishing connections between the actors in research and higher education institutions. Specifically, the idea was to support the creation of an archiving service at Sciences Po, closely tied to the CDSP, and to forge ties with leading research archives (particularly the EHESS archive) or professional networks (particularly the Aurore section of the Association of French archivists). We worked to create favourable conditions for partnering with this network, to share insights on our practices, to improve our procedures, and to anticipate the collection of new surveys. The next step was to put in place concrete procedures to preserve the documents gathered. A classification plan was developed to organize survey documents into a coherent package that would enable comparisons among surveys; it should be noted that this classification plan is used in the website’s faceted search engine (see infra). We also made efforts to facilitate the sustainable preservation of digital files, with respect to both naming rules and preferred electronic formats (text files, audio, spread-sheets, images, video, etc.). In addition, we anticipated the issue of long term survey preservation by forming a partnership with the Huma-num Very Large Research Infrastructure; Huma-num thus serves as an interface between, on the one hand, research and higher education institutions or bodies such as Equipex, and on the other, CINES – the body that is officially responsible for the long term preservation of digital files for higher education and research.



</>Entretien n°41

Enquêteur(s) : Etienne Schweisguth
Enquêté(s) : Amélie

Retour à l'enquête

Etienne Schweisguth .0 — ✂ Alors, de quoi ••• voilà de quoi il retourne en gros, hein, disons, il y a deux ans, il y a eu un changement de gouvernement •••

Amélie .1 — D'accord.

Etienne Schweisguth .2 — Alors, je voudrais qu'on en parle, et puis •••

Amélie .3 — Mais pourquoi vous vous adressez aux femmes ?

Etienne Schweisguth .4 — Ah, mais je m'adresse eux hommes aussi. ✂ Là, je m'adresse aux femmes, parce que j'ai eu beaucoup d'hommes jusqu'à présent. ✂

Amélie .5 — ✂ les mêmes vues, mais enfin je vais vous donner mon idée à moi, puisque j'ai ••• je vous ai préparé ••• j'ai fait une petite rédaction pour ne pas avoir l'air trop sotte ••• pour vous répondre ☺ .

Etienne Schweisguth .6 — Alors, je vous présente, hein, et puis vous allez me dire ça, alors, d'abord je vous présente, disons, l'enquête comme je la présente toujours, hein, disons, bon, il y a deux ans, il y a eu ce changement, alors je voudrais quand même qu'on en parle et puis qu'on évoque, disons, vos réactions sur la droite, la gauche, et qu'est-ce que ça représente, vous voyez, pour vous, la droite, la gauche et tout ça ?

Amélie .7 — Oui, mais à ce moment-là, c'est un sujet ✂ politique, hein, moi, je ne suis pas ••• je suis apolitique, je suis apolitique •••

Etienne Schweisguth .8 — C'est ça, vous, vous êtes apolitique ••• ?

Amélie .9 — Apolitique. Voilà, déjà.

Etienne — Voilà, voilà •••

Métadonnées

Date : 1983

Localisation : région parisienne France

Nom : Entretien n°41

Id : cdsp _ bequali _ sp2 _ col _ transcr _ entretienindiv _ fr _ droiterigo _ paraverbal41 _ vi.xml

Options

Voir/cacher les intervenants

Enquêteur(s) :

☑ Etienne Schweisguth 🟥

Enquêté(s) :

☑ Amélie 🟦

Voir/cacher le paraverbal

☑ Tout cocher

Transcription

☑ ✂ Coupure
☑ ❝ Commentaire
☑ ? Question
☑ 🕘 Temps

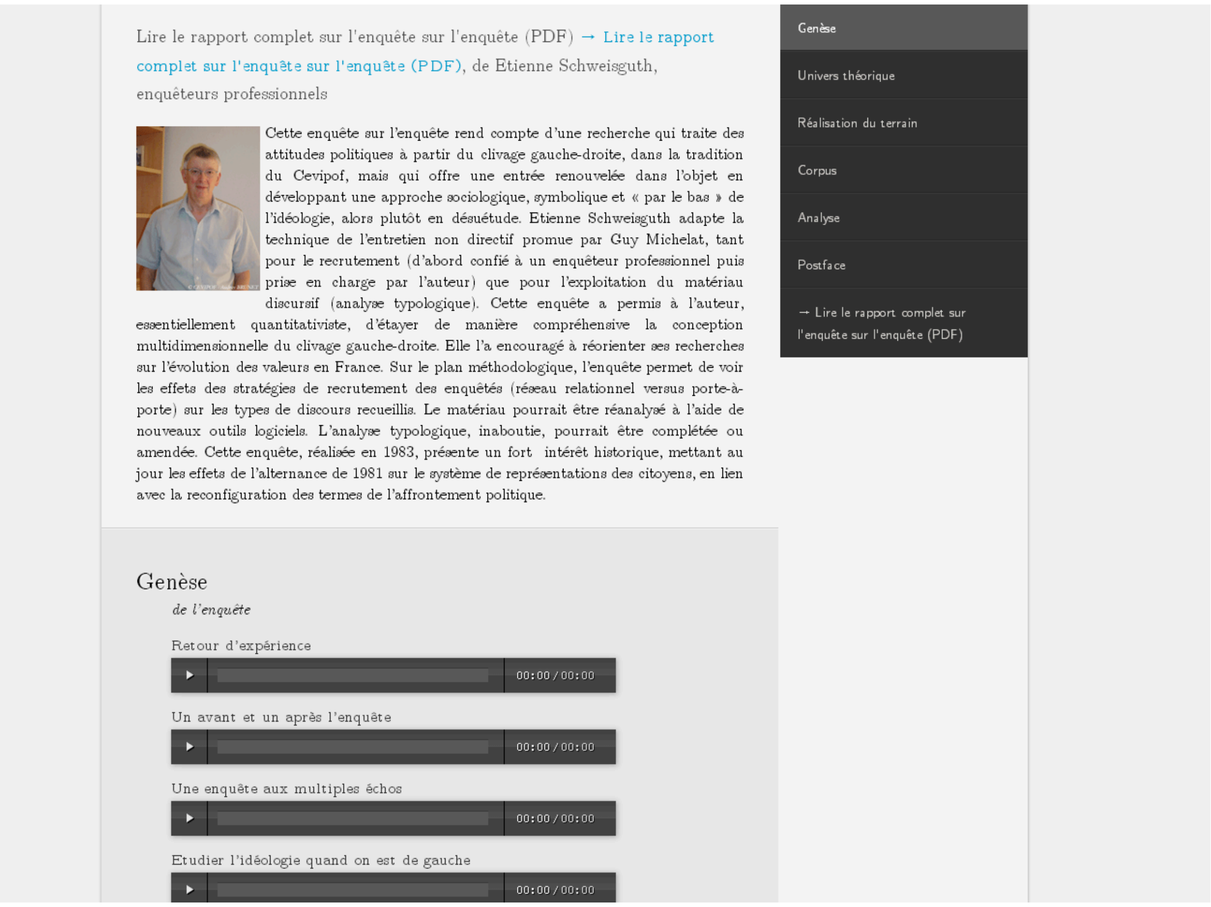
Verbatim

☑ Gestuelle
☑ ••• Hésitation
☑ ☺ Rire
☑ ⚠ Incident

In parallel, we worked to organize the processing of documents collected for the purpose of making them available. Thus, we have developed a whole protocol for encoding the interview transcripts into TEI. The structuring of verbal and paraverbal levels not only ensures the long-term preservation of textual documents encoded this way, but also increases opportunities to interactively explore the documents online thanks to the display tools developed on the website. This necessitated the development of a dictionary of TEI tags adapted to interview transcripts, as well as a text encoding protocol. We also had to develop adapted display tools (viewers). This work was done for the réAnalyse prototype on the basis of test surveys; it then had to be adapted to the production website (enQuêtes [SurVeys]). Additionally, we developed a protocol to digitise paper documents. Given our resources and competences, we chose to externalise this task. To do this we drew up specifications and conducted a digitisation test with a service provider. This work was made possible by collaboration with the Sciences Po library, which provided key financial and human resources for the project.

At the same time, we engaged in a strictly scientific reflection to refine our goals. In doing so we sought to both improve our efficiency and better define the steps to processing surveys and making them available. This is how we ensured we were producing something that was both useful and scientifically valid for the instrument's users.

This effort was first conducted with respect to what we call the survey of the survey. Our goals have significantly changed since we ran the first prototype tests; specifically, we have abandoned the idea of developing any socio-historical analysis whatsoever of the survey, or even of carrying out an analysis of the survey's contribution in relation to the state of the art at the time. The survey of the survey essentially aims to reconstruct as well as possible researchers' knowledge and thoughts about their survey, and to provide the researchers' framework of understanding their work. This entails familiarisation with the survey's documents and some reading to learn about both the scientific context at the time and the research area involved. But this work is primarily intended to prepare for the interview(s) with the original researcher. The goal is therefore not to know more than the researchers do, but rather to know what is necessary to effectively discuss their survey with them, i.e. to best support researchers in their efforts to remember how things happened. In concrete terms, the survey of the survey is presented in two forms: in the form of a long report that can be downloaded as a PDF (20 to 30 p.) and in the form of audio chapters that can be accessed online (these chapters are an assemblage of the interview(s) previously conducted with the original researcher). The principle distinction between these two forms is that the text is drafted as a summary, while the oral assemblage draws on the liveliest and most illustrative moments from discussions with the researcher. The oral version has a strong pedagogical dimension, especially for students, that is, audiences that are least familiar with field surveys. The two versions – written and oral – are organized into the same six major sections, which make it possible to distinguish and articulate the information necessary to understanding the survey: its genesis; its theoretical universe; the fieldwork; the collected and preserved corpus; the analysis of the materials; and finally, in an afterword, the survey's legacy and potentialities.



Lire le rapport complet sur l'enquête sur l'enquête (PDF) → Lire le rapport complet sur l'enquête sur l'enquête (PDF), de Etienne Schweisguth, enquêteurs professionnels

Cette enquête sur l'enquête rend compte d'une recherche qui traite des attitudes politiques à partir du clivage gauche-droite, dans la tradition du Cevipof, mais qui offre une entrée renouvelée dans l'objet en développant une approche sociologique, symbolique et « par le bas » de l'idéologie, alors plutôt en désuétude. Etienne Schweisguth adapte la technique de l'entretien non directif promue par Guy Michelat, tant pour le recrutement (d'abord confié à un enquêteur professionnel puis prise en charge par l'auteur) que pour l'exploitation du matériau discursif (analyse typologique). Cette enquête a permis à l'auteur, essentiellement quantitativiste, d'étayer de manière compréhensive la conception multidimensionnelle du clivage gauche-droite. Elle l'a encouragé à réorienter ses recherches sur l'évolution des valeurs en France. Sur le plan méthodologique, l'enquête permet de voir les effets des stratégies de recrutement des enquêtés (réseau relationnel versus porte-à-porte) sur les types de discours recueillis. Le matériau pourrait être réanalysé à l'aide de nouveaux outils logiciels. L'analyse typologique, inaboutie, pourrait être complétée ou amendée. Cette enquête, réalisée en 1983, présente un fort intérêt historique, mettant au jour les effets de l'alternance de 1981 sur le système de représentations des citoyens, en lien avec la reconfiguration des termes de l'affrontement politique.

Genèse

de l'enquête

Retour d'expérience

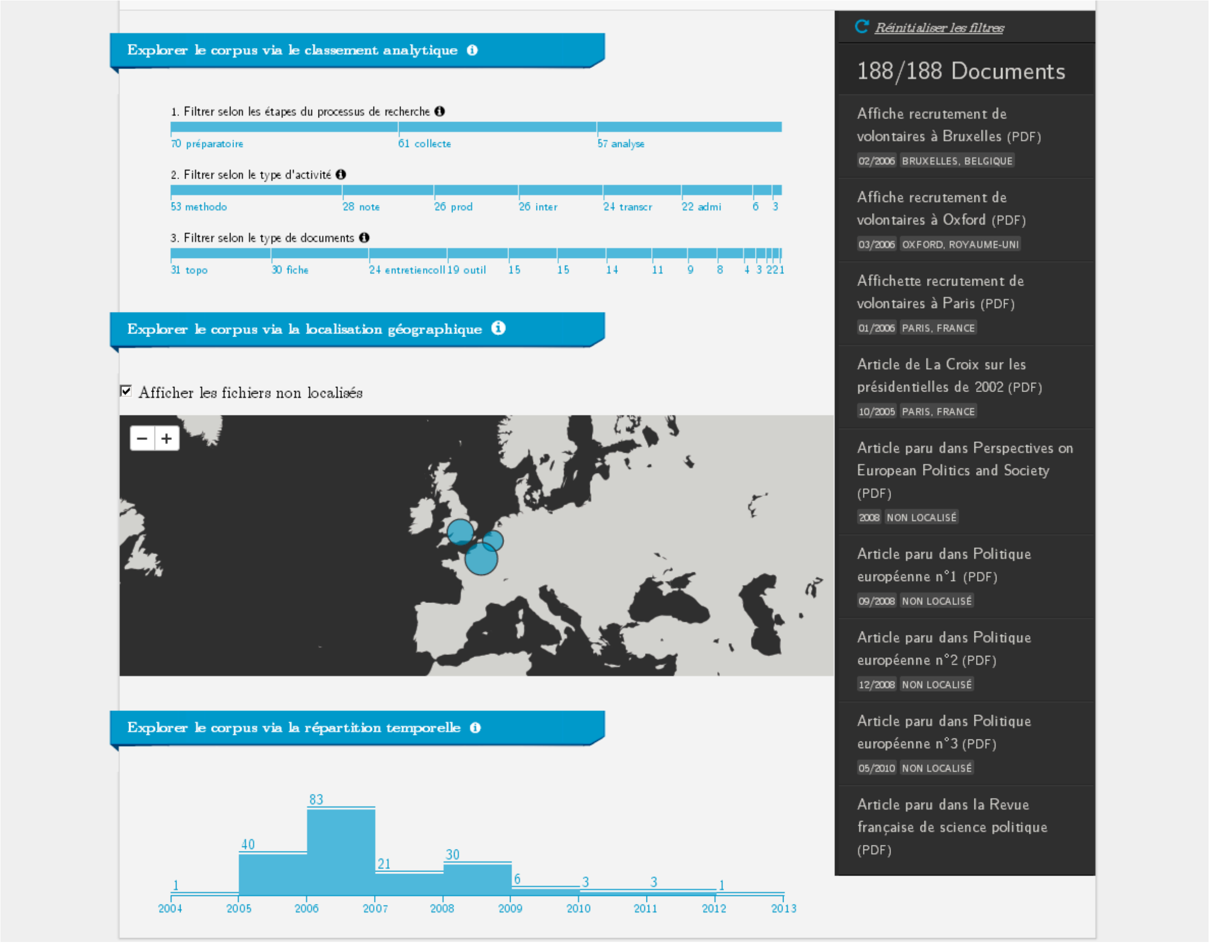
Un avant et un après l'enquête

Une enquête aux multiples échos

Etudier l'idéologie quand on est de gauche

An effort to define the documentary scope of the survey came next. It allowed us to determine the types of documents we deemed necessary to post with the survey in line with our objectives. In order to reconstruct the context of the survey's production, we favoured a deliberately broad vision of the documentary scope, which ranges from documents drafted in preparation for the research (administrative, methodological…), to the analysis (rough drafts, intermediate and final versions) and of course the collected materials. Our approach to gathering and posting the documents does not aim to be exhaustive, however: the goal is to create a curated collection including only that which will allow other researchers to make valid use of the offered materials, i.e. by having the information needed to interpret them. This list (reproduced in the appendix p. 55) will be updated as the archiving proceeds and new surveys are made available.

In connection to this, we have worked to develop navigation tools for the survey's body of documents. Specifically, each document is indexed according to multiple criteria (time, geographical, analytical…). The survey can therefore be visually represented in different ways: documents placed on a timeline can show the survey's chronology; documents placed on a map can show the geographical dimension of the research process; navigation through categories used in the classification plan provides a better grasp of the structure of the body of documents. These tools allow for sorting and selections across the body on the basis of these criteria, or a combination of them if necessary. These navigation tools are meant to facilitate familiarisation with surveys consisting of numerous and diverse documents and a complex structure. The idea is to go beyond simply going through a rigid classification plan that pushes the secondary researcher into a forest of files. This exploration was therefore designed to help the user grasp the survey and to give the user the tools to understand the composition of the body of documents gathered without predetermining a path of discovery. Of course, all of this occurs before the analysis, which secondary researchers will conduct as they see fit, and with their own tools, after they have downloaded the documents.



Explorer le corpus via le classement analytique

1. Filtrer selon les étapes du processus de recherche

2. Filtrer selon le type d'activité

3. Filtrer selon le type de documents

Explorer le corpus via la localisation géographique

☑ Afficher les fichiers non localisés

Explorer le corpus via la répartition temporelle

Réinitialiser les filtres

188/188 Documents

Affiche recrutement de volontaires à Bruxelles (PDF)
02/2006 BRUXELLES, BELGIQUE

Affiche recrutement de volontaires à Oxford (PDF)
03/2006 OXFORD, ROYAUME-UNI

Affichette recrutement de volontaires à Paris (PDF)
01/2006 PARIS, FRANCE

Article de La Croix sur les présidentielles de 2002 (PDF)
10/2005 PARIS, FRANCE

Article paru dans Perspectives on European Politics and Society (PDF)
2008 NON LOCALISÉ

Article paru dans Politique européenne n°1 (PDF)
09/2008 NON LOCALISÉ

Article paru dans Politique européenne n°2 (PDF)
12/2008 NON LOCALISÉ

Article paru dans Politique européenne n°3 (PDF)
05/2010 NON LOCALISÉ

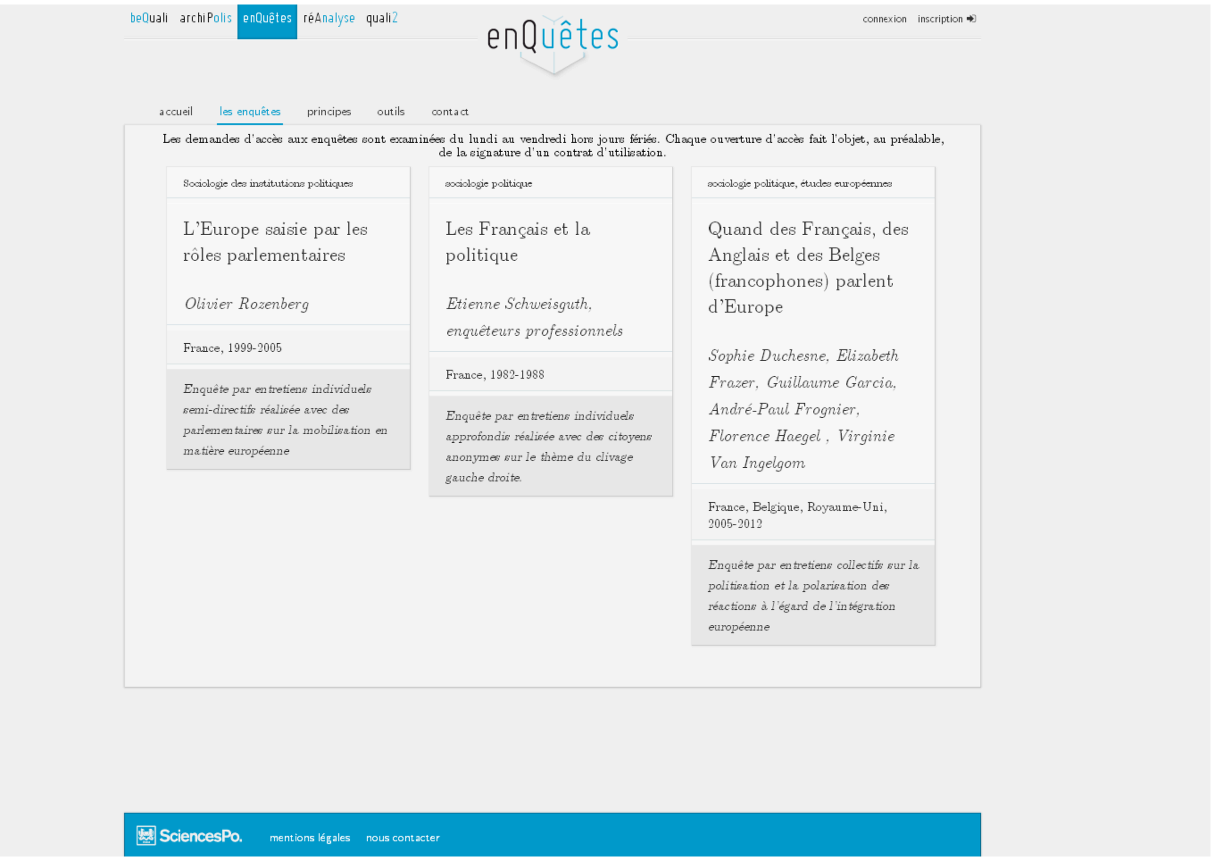
Article paru dans la Revue française de science politique (PDF)

All of these activities have been pursued simultaneously since the beginning of 2012. At the time of writing of this report, we have fully processed 3 surveys: the first survey was published at the same time as the website was launched (http://bequali.fr/app), in July

2013; the latest one was published in February 2014. This phase was followed by an effort to finalise the instrument – or at least produce a first operational version. The goal is to stabilise both the software part (fixing the last bugs that appeared when the three surveys were published) and our own work procedures to improve our efficiency in processing future surveys.

Our current catalogue includes three surveys:

- **survey 1**: *When the French, English and (French speaking) Belgians talk about Europe* (ed. Sophie Duchesne), 188 documents to explore online, 871 documents available for download (including 527 questionnaires);
- **survey 2**: *The French and politics* (Etienne Schweisguth), 73 documents to explore online, 76 documents available for download;
- **survey 3**: *Europe understood through parliamentary roles* (Olivier Rozenberg), 142 documents to explore online, 169 available for download;



Les demandes d'accès aux enquêtes sont examinées du lundi au vendredi hors jours fériés. Chaque ouverture d'accès fait l'objet, au préalable, de la signature d'un contrat d'utilisation.

Sociologie des institutions politiques

L'Europe saisie par les rôles parlementaires

Olivier Rozenberg

France, 1999-2005

Enquête par entretiens individuels semi-directifs réalisée avec des parlementaires sur la mobilisation en matière européenne

sociologie politique

Les Français et la politique

Etienne Schweisguth, enquêteurs professionnels

France, 1982-1988

Enquête par entretiens individuels approfondis réalisée avec des citoyens anonymes sur le thème du clivage gauche droite.

sociologie politique, études européennes

Quand des Français, des Anglais et des Belges (francophones) parlent d'Europe

Sophie Duchesne, Elizabeth Frazer, Guillaume Garcia, André-Paul Frognier, Florence Haegel , Virginie Van Ingelgom

France, Belgique, Royaume-Uni, 2005-2012

Enquête par entretiens collectifs sur la politisation et la polarisation des réactions à l'égard de l'intégration européenne

Two other surveys are currently being processed and should be ready by the end of the first semester of 2014: these include Nonna Mayer's *The boutique against the left* survey and the *Representation of the social field, political attitudes and socioeconomic changes* survey by Guy Michelat, Michel Simon and Jean-Marie Donégani. We have also developed contacts with a handful of other researchers to work on their surveys in the second semester of 2014 (including, for instance, Agnès Zanten's *Choosing one's school* survey, or the *Formation of couples* survey by Michel Bozon and François Héran).

Thus, the instrument can currently be used on the three surveys in the catalogue. At the end of 2013 the system's usage statistics were as follow:

- 34 people signed up as users

- we received about ten requests for access to the surveys for research purposes (without counting team members or members of our larger network)
- we received 2 requests for access to the surveys for method teaching purposes. The first request was linked to a qualitative methodology course at Sciences Po (21 enrolled). A second request was linked to a training on text analysis software provided at the Pacte laboratory in Grenoble (8 enrolled).

Given that the first survey was only published in the summer of 2013 and that secondary analysis takes time, no reanalysis project has been completed yet. We have not proactively promoted the project. Nonetheless, the instrument has started to be used by both réAnalyse project partners and researchers not involved in the project. Once the two other pending surveys are published we intend to actively promote the database to researchers and teachers, laboratories, and research bodies and networks.

However, we have also already started to promote the project in its current form. We have thus advanced on the publication side. Four articles written by team members were produced or accepted for publication in 2013 (they are scheduled to be published in 2014); they present a reflection on the equipment and its implications for transformation in social science research practices (see the list reproduced in the appendices):

- Anne Both, Guillaume Garcia, « Le chercheur, l'archiviste et le webmaster : la polyphonie patrimoniale ? Le cas de beQuali, banque d'enquêtes qualitatives en sciences sociales. », in Bernadette Dufrêne (ed.), *Patrimoines et humanités numériques : quelles formations ?*, Berlin, Lit Verlag, 2014, forthcoming.
- Sophie Duchesne, Guillaume Garcia, « beQuali : une archive qualitative au service des sciences sociales », in Marie Cornu, Jérôme Fromageau (ed.), *Les archives de la recherche, pratique des acteurs et enjeux juridiques*, Paris, Editions L'Harmattan, collection droit du patrimoine culturel et naturel, 2014, forthcoming.
- Sophie Duchesne, Guillaume Garcia, Anne Both et Sarah Cadorel, « Retour vers le futur : la numérisation des enquêtes qualitatives de sciences sociales entre patrimonialisation et transformation des pratiques scientifiques. », published online on 20/02/2014: http://humanum.hypotheses.org/147
- Sophie Duchesne, Guillaume Garcia, « Partager les enquêtes en sciences sociales : la révolution numerique *behind closed doors* », *Socio*, special issue on "Les sciences humaines et sociales à l'ère du numérique : approches critiques", paper submitted on 28/02/2014.

In addition, we have made a great number of public presentations about the project at seminars, symposiums, conferences, etc. These specifically include:

- the organisation of a seminar on the archiving of qualitative surveys in Europe (Paris, November 2011)
- participation in the activities of the DDI qualitative working group (Göteborg, December 2011)
- a presentation at a symposium on research archives (Paris, January 2012)
- a presentation at the IASSIST's annual conference (Washington, June 2012)

- a presentation at the EDDI13 – 5th Annual European DDI User Conference, (Paris, December 2013)

Moreover, at the beginning of 2014 the CDSP hired two postdoctoral fellows (15 and 18 months) with funding from Idex Sorbonne Paris Cité. Each one will re-analyse a survey to be archived and disseminated via beQuali, and will contribute to the methodology.

Prospects

Prospects are tied to three focal points: to increase and diversify the survey catalogue, to improve the tool providing data, and to further methodological research on the tool.

1-Our primary objective in 2014 is to increase and diversify the available survey catalogue

Achieving this target will depend on our ability to process more surveys. It will also depend on how the system is set up to inventory and collect surveys. This issue was not directly foreseen in the initial project as tabled in 2010; its attendant tasks were therefore not funded as a part of Equipex. We will have to actively explore and create a network of counterparts in the archives and laboratories to ensure a continuous flow of surveys. An important challenge is that this process will unfold in an academic environment in the social sciences that lacks a culture of sharing or granting visibility to qualitative data on the one hand, and that suffers from underfunding of activities linked to the preservation of scientific archives on the other. However, Equipex is already structurally connected to the network of political science laboratories that are part of the archiPolis consortium (itself part of the Huma-Num TGIR). These connections need to be strengthened without remaining limited to this network. It will be especially important to develop connections with the political science and sociology laboratories that are part of Equipex. Another goal will be to create closer ties with the network of higher education and research archives. It should be underscored that nothing can be accomplished without the active participation of researchers themselves. Indeed, survey documents are most often in the possession of researchers. Their active contribution is also essential to producing the survey of the survey. We are not seeking to reproduce a system modelled on the English one, which requires publicly funded surveys to be filed. To the contrary, we support researchers' freedom to decide whether or not they wish to file their surveys. However, we hope that our project will be supported and relayed by professional bodies – graduate schools and associations such as the AFS, AFSP or ANR – to heighten the scientific community's awareness about archiving and reuse challenges. Ideally, the filing of a survey should be equivalent to a major publication in the career evaluation of researchers, in order to encourage them to share their data. We also hope that researchers will be motivated by the greater visibility they would gain from their survey being reused and cited.

The DIME-SHS/Quali team will also seek to distribute a good practices guide to researchers and teachers to facilitate the archiving of past surveys and especially of future surveys, as in available elsewhere (for example, see the Handbook on *Managing and Sharing Research Data* recently produced by UKDA).

We will also ensure that the surveys we provide are diverse, both in terms of the subjects and themes studied, and the methods used. We have thus far focused on subjects linked to politics in a very broad sense, using a multidisciplinary approach.

Eventually though, we would like to create different thematic clusters corresponding to coherent fields of study: European studies, sociology of labour movements, sociology of militancy and parties, sociology of administrations or of elites, sociology of institutions, sociology of the media, sociology of the family, sociology of education, etc. Within these clusters we will endeavour to represent the diversity of trends and approaches, meaning that we will have to best represent the range of researchers and of research teams who produce these surveys

DIME-SHS/Quali aims to be a selective instrument: the CST will be responsible for selecting the surveys that should be prioritised for archiving. So long as we have very limited resources, the CST will not only have to assess the value of making a survey available, but also determine an order of priority. In order to fulfil this objective, we wanted CST members, who are mostly experts outside of the consortium, to reflect the diversity of the elements we would like this project to encompass: diversity of scientific and computing competences, disciplinary diversity and methodological and epistemological diversity. Doing so will require jointly establishing a series of transparent evaluation criteria and procedures on the basis of several principles, such as in particular: the potential to reuse the surveys for both new research and the teaching of methods; the ethical risks involved in making the surveys available; the objective of representing various theoretical and methodological currents of sociology and of political science. In the two meetings (June 2012 and May 2013) the CST has held, the experts present deemed a specific review of this issue premature. The CST determined that the DIME Quali team should continue to make progress on building the instrument by working on surveys allowing it to explore issues of relevance to the wide variety of qualitative surveys in the social sciences; only after this would the CST have the ability and duty to make a selection. Note that a CST will be organised in the month of June 2014, after this report will have been submitted. The issue of survey selection criteria might be raised again then.

2- A second important objective will be to improve the tool providing data.

In the immediate future, the goal will be to integrate the Quetelet portal via the application of commands that were recently established (March 2014), in order to benefit from functions enabling the referencing of surveys, creation and management of user accounts, and file ordering and downloading. We are currently in a test phase. We anticipate integrating the portal by the summer of 2014.

Second, we will improve the website's functionality for making surveys available. Adaptation of the prototype website with a view to creating a production website allowed the team to take stock of the discrepancy between the ambitions reflected in the prototype and the possibilities for large-scale implementation. It will take more time to develop the features that were initially planned, taking into account user feedback, before we start testing new tools.

Another goal is to build on an effort that began in 2013 to thoroughly review the metadata. Indeed, the various types of work involved in the inventorying, posting and long-term archiving of surveys have led to inconsistent use of descriptive information. For the systems linked to these different goals to communicate, it will be essential that the information flowing from one to another be structured. Moreover, our various partnerships require us to guarantee interoperability with their infrastructures. This is the case, for instance, for the archives with which we must be able to communicate via the use of EAD, for the Quetelet network in which we must be referenced using DDI, with the CINES for which we must be able to generate a SIP (Submission Information Package), etc. We have therefore set out to create a correspondence table between the different standards and metadata that are relevant to us in order to be able to pass information between systems. With this data model, our internal system should be able to import and export information following these different standards. Furthermore, the effort has allowed us to correct and standardize the metadata used throughout our survey processing.

3- Finally, we also aim to further methodological research on the tool.

A first project will consist of analysing user feedback to gain a better understanding of whether the survey of the survey system is effective in its current form – that is, is it really useful in helping secondary researchers, teachers or students understand the dynamics of research that they did not experience themselves? Does it really add value compared to survey documents that are already available? If necessary, the current model will have to be adapted according to issues raised as we process new types of surveys.

It will also be necessary to integrate the needs of newly collected surveys, particularly surveys using ethnographic methods. Indeed, the repository website was built upon interview-based surveys. It needs to be adapted to surveys that draw on observations, note-taking, and documentary and visual material. Other areas that will probably be tested include our de-identification protocols and protocols for determining the documentary scope and for visually representing the survey, etc.

Specific issues that arise from making the quanti/quali surveys available will also have to be addressed. We have a prime test case for this: Michel Bozon and François Héran's survey on couple formation, which will be covered by one of the two post-doctoral fellows we have just hired.

Finally, we will have to explore potential uses of provided surveys for the teaching of methods. This might take the form of a consultation mechanism tailored to students and teachers, of adapted pedagogical kits, or of secondary analysis training, through seminars or summer schools for instance.

Ethical principles and de-identification for data dissemination

In order to remove uncertainties about the legal framework for collecting, processing and disseminating surveys, we had to engage the services of a specialised firm beginning in autumn 2013. The contracts drawn legally cover the actions undertaken by DIME-Quali, particularly in terms of responsibility. In addition, they seek to implement a principle of double protection: for the researchers filing surveys and for the respondents.

With regard to de-identification, we operate on the basis two scenarios. First scenario: surveys consisting of individual interviews with individuals who are not acquainted and who do not have a public profile; in this case, eliminating personal information suffices to drastically reduce the risk of recognition. Second scenario: surveys involving respondents with a public profile such that it would be absurd to de-identify them; in this case the survey can only be posted if the respondents' consent is secured, possibly ex post. Between the two scenarios is a grey area that will require input from DIME-SHS/Quali's scientific committee, or even the scientific committee. At stake here is the need to share thoughts and to adapt the ideal level of de-identification for each survey. Indeed, short of sanitizing surveys to the point where they become sociologically useless, it is often difficult to eliminate everything that might trace back to the respondents, especially in the case of local surveys or surveys on a well-defined milieu.

To protect respondents, we de-identify the data ourselves: we do not just request that researchers do it (even if they previously did so, we need to double-check). In all cases, we create a table of equivalence (that the researcher gives us or that we create ourselves), preserved separately in a secure area on the DIME-SHS server, apart from the files uploaded on the website.

We are currently using a basic de-identification policy that consists of:

- systematically de-identifying the elements that enable the direct identification of respondents (last name, first name, telephone number, address…)
- to the greatest extent possible, de-identifying the elements that would enable easy identification of respondents (for example, mayor of a small village at a specific date, etc.)

On this basis, we recommend adopting the following de-identification conventions:

- replace direct identifiers with a superordinate (such as [person's name], [person's telephone number], etc.), to show what type of information was deleted
- when the element that needs to be de-identified is complex (typical case: mayor of such a village), keep what is most significant from a sociological standpoint ((mayor) and de-identify the name of the village [name of the village])

- when the identity is not masked, create a sociologically coherent pseudonym with respect to social, geographical and generational characteristics (replace Marcel with Robert and not Jean-Edouard).

These conventions were established on the basis of non-ethnographic surveys; they will most likely be amended when we start de-identifying ethnographic surveys, which raise more complex issues.

We believe it is essential to keep the rest of the information; the data cannot be too impoverished lest they lose their potential for reuse. Furthermore, surveys will only be made available to researchers (irrespective of their status). By this we mean official researchers (from the CNRS or other public research institutions), teacher-researchers, research or design engineers, post-doctoral and doctoral students, as well as research masters students (under the supervision of their research director) or undergraduate students (under the teacher's supervision). When these researchers sign the reuse contract to gain access to a survey, they commit to respecting both the respondents' anonymity and the survey author's reputation. It is also, in principle, a form of guarantee for respondents – and also the authors of the surveys entrusted to us – since the activity of the re-users is presumably subject to professional ethics, to know-how, and even to forms of peer oversight.

The conventions for reuse commit re-users to accept a number of constraints: to respect witnesses in analyses produced from the materials, to not seek to lift the anonymity during the analysis and to protect it upon publication, to not transmit the materials to third parties, etc. Beyond the protection of witnesses, what is at stake is ensuring the future participation of researchers in expanding the catalogue of surveys that are archived and made available. This implies formalising the recognition owed to the original researcher: the reuse contract includes a "scientific civility" principle that must be respected (to avoid the settling of scores, for example), but also stipulates that the primary researcher must be cited (in addition to the principal resource publication) in any scientific production pursued on the basis of the survey's reuse. As has occurred in other countries that have already established this type of system, we hope to create a positive incentive for researchers to file and document their surveys. This type of exposure might be considered a legitimate quid pro quo for the sharing of data.

DIME Quanti

The ELIPSS panel (Longitudinal Internet Studies for the Social Sciences) is an online survey system for the scientific community. It aims to address the lack of questionnaire survey tools specifically deigned for French researchers in the humanities and social sciences. Its probability sample and scientific purpose make the online panel of the DIME-SHS equipment of excellence the first of its kind in France, and its mobile web dimension makes it the first of its kind in Europe.

ELIPSS includes an online panel that is representative of the population living in metropolitan France. Panellists are randomly selected from the census and provided with a touchscreen tablet and mobile Internet subscription so that they can participate in monthly surveys. The surveys are created by researchers and selected by a scientific and technical committee.

The first part of this section is devoted to a general presentation of the system, focusing on the situation in France and abroad with regard to general population-based online surveys. The second part presents the results of the pilot ELIPPS panel, specifically preliminary results on fieldwork and on the panel's representativeness. Next, various operational aspects of the system are discussed, ranging from survey selection to data dissemination, and including the production of questionnaires and monthly participation of panel members. Finally, the conclusion offers a preliminary assessment of the pilot, highlighting panel development challenges for 2015.

An online research panel

Online general population surveys

While online surveys have distinct advantages, their use in the general population presents several difficulties.

On the one hand, they share the advantages of self-administered questionnaire surveys. Collection costs are lower essentially because no interviewers are needed; respondents can fill the questionnaires whenever it is most convenient for them; the absence of interviewers also allows for more personal questions (health, sexuality…).

Online surveys also have their own particular advantages. They enable new question formats that integrate video, sound and interactive applications. Moreover, the collection period can be reduced since there are no (or almost no) limits on the number of people who can be interviewed at the same time. In addition, answers are saved as they are being collected.

However, using the Internet to survey the general population also raises issues about the representativeness of the sample and the extrapolation of results:

- Online surveys are conducted on the basis of volunteer samples, that is, nonprobability surveys.
- People without access to the Internet are effectively excluded. Yet in France in 2012, one out of every five people did not have Internet access at home (Gombault, 2013).

One way of overcoming this bias is to build an online panel from a random sample of the population. This means recruiting people offline, including people who do not have Internet access, and equipping them with a connection if needed (Das, Ester and Kaczmirek, 2011). For the ELIPSS panel this is achieved by using a random sample of addresses drawn by the national institute of statistics and by providing touchscreen tablets with a 3G connection to all the panel members.

Foreign initiatives

Two foreign initiatives inspired the ELIPSS panel project, and two German initiatives were developed at the same time as the ELIPSS pilot. Discussions are also currently underway in several countries to create similar projects, including in Norway, the United Kingdom[2], and Southern Europe (a consortium involving Cyprus, Spain, Italy, Greece, Portugal and Turkey could be formed in the coming years).

The Longitudinal Internet Studies for the Social Sciences (LISS Panel) of the CentERdata Dutch research institute (Tilburg University)

This online panel representing Dutch households was formed in 2007 on the basis of a probability survey developed in cooperation with the Centraal Bureau voor de Statistiek (CBS, Statistics Netherlands). A simplified computer and Internet connection are provided to households without them. The LISS Panel includes 5,000 households, that is, 8,000 individuals aged 16 and over. This system is free and exclusively devoted to research.

The KnowledgePanel from Knowledge Networks in the United States

Founded in 1999 by two American academics, this system involves random sampling and provision of an Internet connection for respondents who lack one at their home. The sample is randomly drawn from a database of home addresses and involves 50,000 people aged 19 and over. Unlike the LISS Panel, this system is also open to commercial studies.

The German Internet Panel (GIP) at the University of Manheim

This panel represents the population aged 16 to 75 and living in Germany, and is based on a random sampling of 1,500 people. Panel members are surveyed every 2 months. As with the LISS Panel, a simplified computer and Internet connection are provided to people who lack them.

This system is reserved for researchers at Manheim University, and the questionnaires focus on political reform.

The GESIS Panel at the GESIS Leibniz Institute for the Social Sciences in Mannheim

This system uses two types of self-administered questionnaires:

- People who have home Internet access can answer online,
- People who do not have Internet access or do not want to respond online can respond on paper to questionnaires sent by post.

[2] http://www.natcenweb.co.uk/genpopweb/

The sample is randomly drawn from municipal record and includes 4,000 individuals aged 18 to 70 and living in Germany. The bimonthly surveys are provided by social science research teams through calls for projects, and they cannot serve any commercial interest.

The ELIPSS system

The French context

France has no service producing national questionnaire surveys in the humanities and social sciences.

French researchers have two options to fill this gap. They can produce surveys by using a survey organization, but this is a costly solution, especially when the survey requires a random sample. They can also reuse statistically representative surveys, such as national statistical surveys, but these do not cover all the subjects and questions that are of interest to researchers.

In addition, survey participation rates have decreased, essentially due to people's refusals to respond and to the growing difficulties in reaching people to interview. This jeopardizes the statistical quality of surveys.

Thus, implementation of the ELIPSS panel has two main objectives:

- Allow researchers to conduct surveys on subjects that are not covered by French public statistics,
- Free public research of the need for private middlemen to produce questionnaire surveys based on random sampling, while diminishing costs and collection times (by using self-administered Internet surveys).

Collection via mobile Internet

The ELIPSS panel differs from similar systems used abroad by using mobile Internet as a principal method of collection. A touchscreen tablet and unlimited 3G plan are provided to all the panel members in exchange for their participation. Thus, they can answer the questionnaires even if they do not have an Internet connection.

This new technology was selected for both methodological and practical reasons:

- The idea is to take advantage of the possibilities offered by the Internet (images, video…) and mobility to update certain survey techniques (travel study, budget-time logs, etc.);
- The tablets offer many advantages over computer surveys. Because their interface is more intuitive, they provide simplified Internet access to people unfamiliar with new technologies. Mobile web access also gives panel members more flexibility in completing surveys (they can choose the time and place).
- The collection tool is the same for all panel members, via a specific application installed on the tablet.

Besides its advantages as a collection method, the tablet was also chosen for its expected incentive effect. The penetration rate for such equipment was low in France at the time of the pilot recruitment (9% of households had a tablet in 2012[3]). It should, however, be noted that adoption has spread rapidly in just two years, since close to one out of every three households had one at the end of 2013.

Timetable and team

The pilot study began in 2012 to define the recruitment process, refine the methodology, establish procedures for managing the panel and producing surveys, and develop software tools (see timetable in the appendix 3 p. 57). During this test phase, use of the ELIPSS survey production service has been restricted to research teams in the DIME-SHS consortium responding to calls for projects. Beginning in 2015, the ELIPSS panel is projected to include 5,000 individuals, and the calls for survey projects will be open to the whole scientific community.

The development of this system draws on a diversely skilled team that has grown throughout the pilot (see table 1). This team is based at two institutions. The Sciences Po's Centre for Socio-Political Data is responsible for coordinating the project, the entire production and dissemination of surveys, and software development and infrastructure. The National Institute for Demographic Studies (INED) is in charge of managing the panel and monitoring the statistical quality of the panel.

[3] GfK / Médiamétrie Survey – Multimédia Equipment Reference

Table 1: the ELIPSS team and its development since 2012

		2012				2013				2014			
Role	Location	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
Coordination	Sciences Po	Anne Cornilleau	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
		Anne-Sophie Cousteaux	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
											to be recruited	▨	▨
Panel management (including technical support)	INED		Carmen Calandra	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
			Gabrielle Bouchet	▨	▨	▨	▨	Marc Sigaud	▨	▨	Patricia Sossa	▨	▨
							Elodie Pétorin (40%)	▨	▨	▨	▨	▨	▨
											Kevin Boudelle	▨	
Survey management	Sciences Po	Emmanuelle Duwez	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
				Matthieu Olivier	▨	▨	▨	▨	▨	▨	▨	▨	▨
										Alexandre Mairot	▨	▨	▨
											to be recruited	▨	▨
Statistics	INED									Nirintsoa Razakamanana	▨	▨	▨
ICT	Sciences Po	Adrien Ferreira	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
		*Daniele Guido	▨	▨	▨								
		*Geneviève Michaud	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
		*Jérémy Richard (50%)	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨	▨
											to be recruited	▨	▨

*: staff on the 3 instruments of DIME-SHS

Pilot recruitment

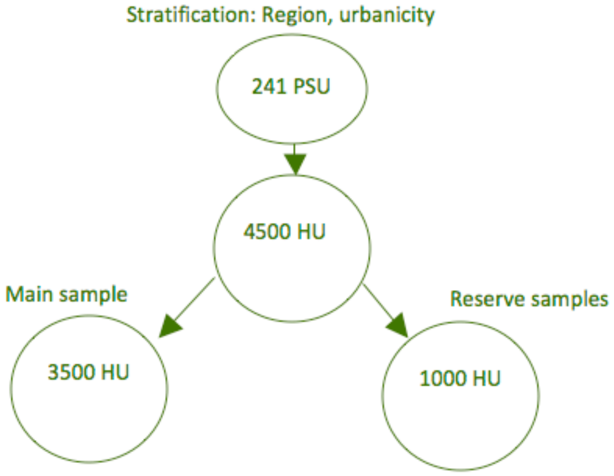
The sampling process

The sampling frame

The ELIPSS panel is a random panel of individuals living in ordinary households as defined by the National Institute for Statistics and Economic Studies (INSEE). This excludes people who are homeless or living in precarious housing, people living as groups (prisons, retirement homes, student housing, etc.) and people who do not have sufficient command of the French language to answer self-administered questionnaires.

The sampling frame includes homes surveyed in 2011[4]. From this frame, INSEE drew a sampling of 4,500 homes[5] using two-stage sampling stratified by region and type of municipality (urban/rural). The first stage is a sample of primary units (PSU corresponding to municipalities or groups of municipalities) with a sampling rate of 1/3[6]. The second stage is a sample of homes from the 241 PSU. The 4,500 addresses were divided into three sub-samples: a main sample of 3,500 addresses and two reserve samples of 300 and 700 addresses to be used in the event that the target of 1,500 panel members was not met.

Figure 1: Sampling Design



One person is then randomly chosen from each home selected by INSEE.

Preparation of the base

The preliminary phase of the work has consisted of entering the address records provided by INSEE; these include the handwritten names and addresses of heads of households. INED's panel managers manually verified the coherence and plausibility of the entered addresses on the Internet. The survey organization completed the work for reserve sample addresses. In addition, the postal office automatically checked the address formatting. The base was also supplemented by a search for telephone numbers.

A total of 76 addresses in the main sample and 13 in the reserve sample were considered unusable. Furthermore, around 10% of letters came back labelled NATA (not at this address); these addresses received a direct visit from a interviewer.

The fieldwork

Two different procedures for using the addresses from the main sample and the reserve samples were implemented. The principal sample was used to study the effect of

[4] 3% of homes were surveyed in 2009 or in 2010.

[5] INSEE kindly provided this sample to ELIPSS for experimental purposes.

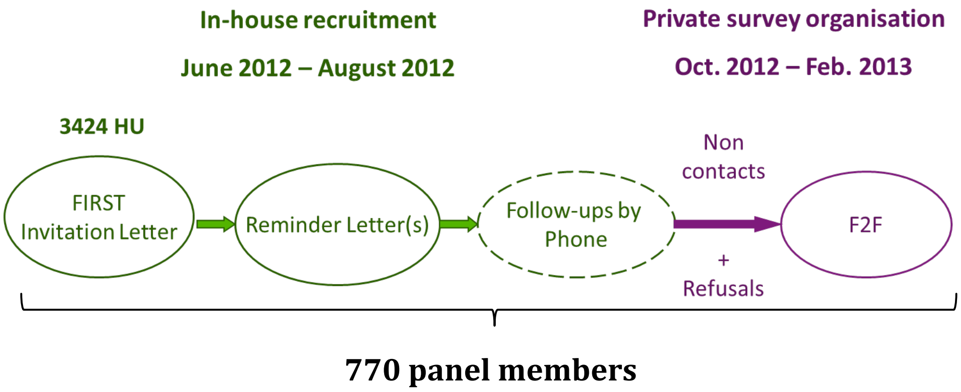
[6] INSEE's Master Sample includes 567 PSU. The effective sample size of the ELIPSS pilot was not compatible with the mobilisation of all the PSU defined by INSEE.

different forms of contact (mail, followed by a phone call, and finally a face-to-face visit), while the reserve sample was exclusively processed by interviewers (phone call and face-to-face visit).

The main sample

For the main sample including 3,424 addresses, the recruitment process involved three successive forms of contact. First, the ELIPSS team sent out an invitation to participate by postal mail in June 2012, and a reminder letter 15 days later. Through mid-July 2012 the team followed up by phone with households whose number was found (around 50% of the main sample's addresses). Face-to-face recruitment proceeded from October 2012 to March 2013 for households that had not yet been contacted (unresponsive and NATA) and for most of the households that had refused to participate (refusals account for 17.5% of addresses that were reused).

Figure 2: Main Sample Recruitment Design

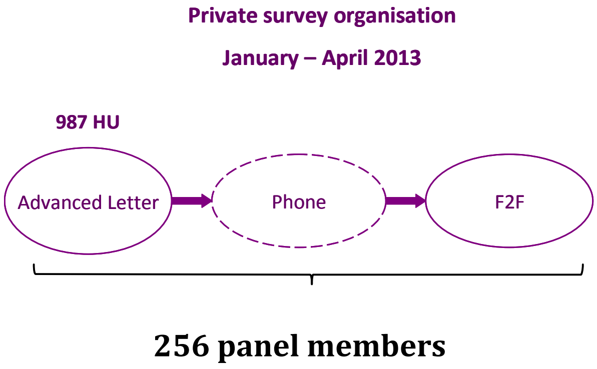


The reserve samples

The main sample yielded 770 recruits. In order to get closer to the target of 1,500 panel members, in January 2013 the team decided to entrust the survey organization with the 987 usable addresses from the reserve samples. The first attempts at contact were made by phone whenever possible[7]. If 12 attempts to call were unsuccessful, or if a member of the household refused to participate, then a interviewer would visit the address in person. Homes without a telephone number only received a face-to-face visit. Use of the reserve samples enabled the recruitment of 256 additional panel members.

[7] A phone number was retrieved for 612 homes.

Figure 3: Reserve Samples Recruitment Design



Thus, the recruitment process yielded a panel of 1,026 individuals.

Actions to improve the recruitment rate

Several actions were taken with a view to improving the recruitment rate:

- Vary means of responding to the invitation to participate

The invitation letters provided three ways for the selected households to respond. Thus, 421 people sent back the reply card by mail, 248 used their access key on the elipss.fr website, and 28 phoned the number included in the letter. Half of the people who responded by mail accepted to participate, while virtually all those who responded online accepted (95%).

- Offer gift vouchers

The pilot recruitment provided an opportunity to conduct an experiment by including 2,000 gift vouchers among the first invitation letters sent to homes from the main sample. Drawing on results from the international literature, and particularly the experiments conducted by the LISS panel (Scherpenzeel, 2009), financial incentives worth 10 euros were randomly and unconditionally distributed among the invitations to participate.

Households that received a gift voucher responded to the invitation more frequently (to accept or refuse to participate) and were also more likely to agree to participate (see table 2).

Table 2: Response to the panel invitation according to receipt of a gift voucher

	No answer	**Answered**				**Total**
		Refusal	*Agreement*	*Ineligible*	**All**	
No incentives (n =1424)	67,8	*18,9*	***11,2***	*2,1*	**32,2**	100
Incentives (n = 2000)	62,1	*16,6*	***18,8***	*2,5*	**37,9**	100
Total (n = 3424)	64,5	*17,6*	*15,7*	*2,3*	35,5	100

Note: the chi-2 test is significant (p-value <0.001)

The positive effect of financial incentives on the rates of response and acceptance is not a novel result, but it was worth confirming in France, where this practice is not common in scientific research. Furthermore, receiving a gift voucher led to a twofold-increased chance of accepting to participate, all else being equal. The modelling of agreement to participate in the panel shows that following the gift voucher, having a phone available was the second most decisive variable. To a lesser extent, the head of household's education level and age also played a role.

- Training the interviewers

The 110 interviewers had a half-day of in-depth training; the ELIPSS team and survey organization jointly prepared its content and format. The trainings, in which ELIPSS team members took part, focused on presenting the ELIPSS panel, on arguments and techniques to convince people to participate, on explaining the terms of the agreement that people must sign (see sections 2.3.1), on filling in the contact form, and on the interviewers' handling of the digital tablets. Indeed, interviewers were equipped with tablets so that they would be able to make demonstrations to the selected individuals.

- Refusal conversions

Of the refusals received by mail or phone, 542 were reworked by the survey organization. Refusals linked to confidentiality or anonymity were not reworked. The reasons cited for the refusals that were reworked are relatively common in questionnaire surveys, such as lack of interest in the panel and lack of time to answer; other refusals involved people who were uncomfortable with new technologies and people who did not want to have access to the Internet or to a tablet. After a reworking of these refusals, 49 subsequently agreed to participate.

Panellist involvement

The agreement

In order to be part of the ELIPSS panel, selected individuals signed an agreement governing the conditions of participation in the surveys and of use of the tablet.

By signing this agreement, panel members make three commitments. They commit to personally and regularly complete the surveys. They also commit to take care of the tablet and to inform the ELIPSS team if the tablet is broken, lost, or stolen. Finally, they commit to return the tablet at the end of their participation in the panel. In exchange, they are free to use the tablet and Internet in accordance with the law in force. They may suspend or end their participation at any time.

In the event of a long period of unresponsiveness or fraudulent use of the tablet, panel members may be excluded from the panel.

In addition to providing the tablet and 3G Internet connection free of charge, the agreement prohibits the commercial use of surveys and guarantees the anonymity of questionnaire responses.

Handling the tablet

Once the agreement was signed and returned, a tablet was sent to the panel member's home. The home screen includes an application to complete the questionnaires.

Panellists were offered training by phone on handling the tablet, exploring the ELIPSS application, and if needed, setting up the wifi. An external contractor led theses trainings using a script drafted by the ELIPSS team. Two thirds of the panel members were trained, 30% were not reachable by phone[8] and 5% refused training. At the end of the training, trainers had to evaluate panel members' comfort level with the new technologies. They reported that 20% of trained panel members were not comfortable with the tablet.

The first survey

Launched in December 2012, the first survey included two parts:

- The "tutorial" was intended to help panel members familiarize themselves with different types of questions and the design of ELIPSS surveys.
- The "Surveys and Internet" module aimed to measure Internet access, digital practices, and participation in surveys before joining the ELIPSS panel.

This first survey was administered until the recruitment process was completed in April 2013, and 90% of panel members[9] responded to it.

According to this survey, 91% of panel members had Internet access at home[10]; 79% used the Internet everyday and 6% connected less than once a week. The survey also showed that 13% of panel members claimed they had always refused to participate in a survey before the ELIPSS panel. Finally, it confirmed that the touchscreen tablet was the primary motivation for agreeing to participate in the ELIPSS panel (cited by 62% of panel members), followed by trust in the institutions involved in the project (for 46% of panel members). Next came the project's originality (37%), interest in research (32%) and the provision of Internet service (13%).

The panel's representativeness

It is not only important to know how many individuals finally agreed to be a part of the ELIPSS panel and to regularly participate in surveys, but also to know who these individuals are, and their characteristics. In other words, to what extent is the panel representative of the general population? In this section, we will draw various comparisons between the profile of the panel drawn from the annual survey and national statistics.

[8] In most cases, these people were not re-contacted about the training if they had completed the first survey.

[9] Only 943 panel members were invited to participate in this survey because the last people to be recruited joined the panel in April 2013 as the 2013 annual survey was being administered. If only the people invited to respond to the "Surveys and Internet" survey are considered, the completion rate was 99% (see table 7).

[10] These data are not weighted.

The annual survey

The ELIPSS annual survey aims to identify many socio-demographic variables (identification module[11]) as well as cross variables and indicators that are frequently used in the humanities and social sciences (barometric module[12]). The questionnaire was built together with several researchers specializing in the subjects covered, and with members of the scientific and technical council (see section 3.1.1). Chosen questions were overwhelmingly derived from existing national and international surveys. Most of the variables derived from the identification module are systematically matched with available data files.

The first ELIPSS annual survey was administered in 2013 after the panel's recruitment process was completed – in April for the identification module and in May for the barometric module. In March 2014 the two modules were administered together: the identification module was entirely reproduced (the questions were adapted to measure changes since the first questionnaire) and selected questions from the barometric module were asked again, excluding the section on behaviour and political views[13].

Comparison of the panel's profile with national statistics

While the goal of providing touchscreen tablets to ELIPSS panel members was to address coverage issues linked to Internet surveys, differences nonetheless appeared between the characteristics of the panel members and those of the target population. Table 3 compares[14] the distribution of several socio-demographic variables in the panel (based on the 2013 annual survey) and in the population aged 18-75 and living in metropolitan France (based on data from the 2010 census). The distortions are similar to those observed at the end of the recruitment process for the LISS panel (Leenheer and Scherpenzeel, 2013; Knoef, de Vos, 2009) in terms of age and level of education. As expected, the highest age groups were underrepresented, as were those aged 18-25 – a group that is always difficult to reach in surveys in France. By contrast, one-person households were overrepresented, as were highly educated and employed people. The male-female distribution, however, was very close to that of the population. Multivariate analyses are currently underway to complete these results.

[11] This module covers the following issues: civil status, work and training, socio-demographic description of the household, housing and neighbourhood, income and wealth.

[12] This module focuses on social connections, leisure and cultural practices, beliefs and religious practices, political behaviour and opinions, health status, health behaviours and lifestyles.

[13] The panel is regularly asked these questions for the longitudinal Dynamob project.

[14] The gaps between the census data and the ELIPSS panel data were chi-squared; only significant statistical differences are noted.

Tableau 3: Distribution of several socio-demographic variables from the sample of respondents and comparison with the population

Variable	Modality	Population statistics (2010) in %	ELIPSS Panel members (April 2013) in %	
		18-75 years	Non weighted	Weighted
Gender*	Men	48.6	48.0	48.6
	Women	51.4	52.0	51.4
Age*	18-24 years	12.0	8.2	12.0
	25-34 years	17.7	19.1	17.7
	35-44 years	20.0	25.7	20.0
	45-54 years	19.6	21.8	19.6
	55-64 years	18.1	16.5	18.1
	65-75 years	12.6	8.8	12.6
Household size	1	16.7	25.9	19.5
	2	34.9	25.8	25.9
	3	20.1	17.9	22.3
	4	18.0	20.6	23.7
	5+	10.4	9.9	8.7
Marital status	Married/in couple	51.1	44.9	49.6
	Single	36.8	40.1	38.2
	Other	12.1	15.1	12.2
Nationality*	French by birth	88.2	90.9	88.2
	French by naturalization	5.4	5.2	5.4
	Foreigner	6.4	3.9	6.4
Professional activity status	Employed	59.7	64.9	55.8
	Student or intern	4.6	6.0	8.7
	Unemployed	7.8	7.5	8.3
	Retired	20.1	15.2	19.6
	Other situation	8.0	6.5	7.6
Housing status	Owner (or co-owner)	60.5	58.5	59.0
	Tenant or subtenant	37.2	35.3	32.6
	Occupant free of charge	2.3	4.8	6.5
	NSP or NVPR	0.	1.4	1.8
Education level*	None/CEP/BEPC	28.4	17.6	28.4
	CAP/BEP	24.9	20.9	24.9
	Bac to bac+2	32.3	36.5	32.3
	Bac+3 and more	14.4	24.9	14.4

Note: the variables indicated with a * were used to carry out the marginal calibration. The area of residence is an additional variable.

An important issue in the implementation of a project like ELIPSS is providing Internet access to those who did not have access before participating in the panel (*offliners*). A comparison with data from the survey on information technologies and electronic communication and commerce conducted by INSEE in 2012 (table 4) shows that the panel members living in households already equipped with Internet access were overrepresented. The LISS Panel and GIP experienced a similar situation (Blom, Gathmann, Krieger, 2015; Leenheer and Scherpenzeel, 2013).

Table 4: Home Internet access in the population and in ELIPSS

	2012 INSEE Survey (population 18-75)	ELIPSS Panel members
Equipped with home Internet access before ELIPSS	83%	91%
Without home Internet access before ELIPSS	17%	9%

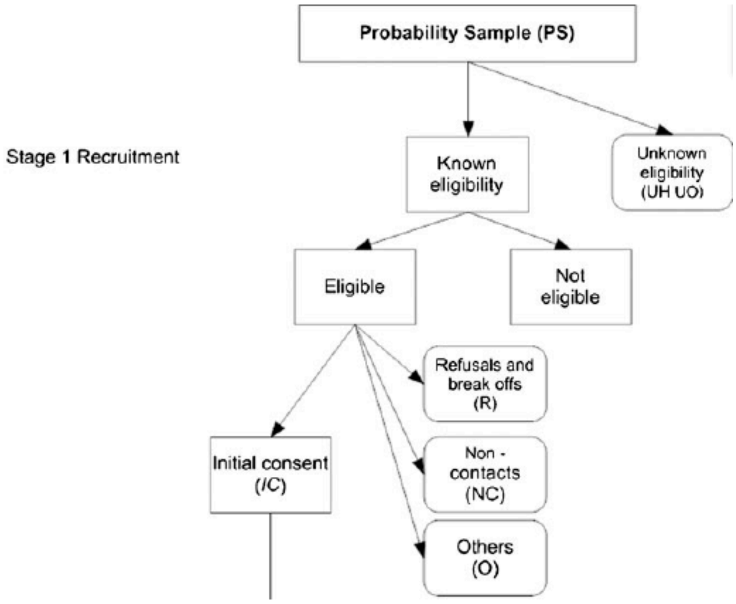
Analyses on the profile of *offliners* will be carried out in collaboration with Mélanie Revilla (Pompeu Fabra University, Barcelona) and Pablo de Pedraza García (Salamanca University) beginning in June 2014.

General results

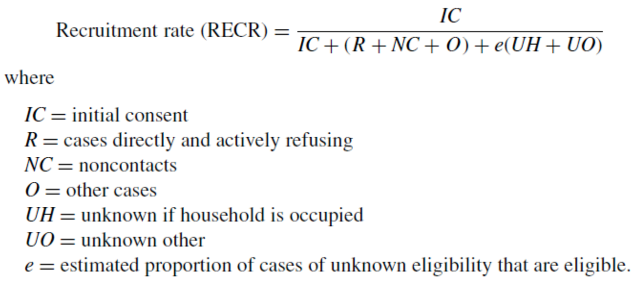
Calculation of the recruitment rate is the pilot's first element of assessment. Two types of initial consent can be distinguished in the recruitment for the pilot, as defined by Callegaro and DiSogra (2008). Indeed, in the procedure specific to addresses from the main sample, a first step consisted of asking for the household's consent to describe the members of the home[15], and the second initial consent occurred at the individual level when the agreement to participate in the panel was signed. Unlike the procedure described by Callegaro and DiSogra (2008) though, there was no profile/connection stage, or to be more precise, this stage was combined with the initial consent at the individual level.

[15] This is particularly true for the main sample's first phase of recruitment since households had to engage in a voluntary process when sending the reply card, phoning the panel manager or filling out the online form to describe the household's inhabitants.

Figure 4: Recruitment Strategy Stage 1 (Callegaro, DiSogra, 2008, p.1012)



The final results of the ELIPSS panel recruitment process are reproduced in table 5 below. They were calculated using the AAPOR formula by considering, as Callegaro and DiSorga suggest (2008, p.1018), the RR3 as a recruitment rate (RECR) defined in the following terms:



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

where

IC = initial consent
R = cases directly and actively refusing
NC = noncontacts
O = other cases
UH = unknown if household is occupied
UO = unknown other
e = estimated proportion of cases of unknown eligibility that are eligible.

Table 5: Household response rate and recruitment rate for individuals

		ALL households	ALL individuals	*Main sample (individuals)*	*Reserve samples (individuals)*
Eligible	Initial consent (IC)	1352	1026	*770*	*256*
	Refusals (R)	1481	1695	*1318*	*377*
	Non contacts (NC)	0	2	*2*	*0*
	Other cases (O)	65	175	*163*	*12*
Not eligible		570	570	*426*	*144*
Unknown eligibility	Unknown household (UH)[16]	172	172	*89*	*83*
	Unknown other (UO)[17]	860	860	*732*	*128*
	e	0,836	0,836	*0,841*	*0,817*
RR3 = RECR		**0,36**	**0,27**	***0,26***	***0,31***

This yields a recruitment rate of 27% at the individual level, with a response rate of 36% (RR3) at the household level. The reserve samples yielded a slightly higher recruitment rate in comparison to the main sample. This difference can be attributed to a greater number of unknown eligibility cases in the main samples, given that the refusal rates were similar (45% in the main sample and 46% in the reserve samples). This can largely be explained by the more in-depth correction and verification of addresses carried out for the reserve samples, leading to a lower number of non-contacts.

From draft survey to data dissemination

The selection of surveys

Survey proposals are made during calls for projects. Since 2011 there has been one call per year. Through 2013 calls for projects in the context of the pilot study were reserved for research teams that were members of an institution partnered with the DIME-SHS

[16] The address could not be located.

[17] The household could not be reached.

equipment of excellence. The next call in 2014 will be open to the whole French and international scientific community.

DIME Quanti's scientific and technical committee

DIME Quanti's scientific and technical committee (CST) selects survey projects submitted by researchers in response to calls for projects, and reviews data matching requests. It includes 15 members, of which at least half are affiliated with institutions that are not part of the DIME-SHS consortium. The CST includes researchers from various social science disciplines (sociology, political science, demography, epidemiology, etc.) specialised in questionnaire survey methods. It also includes an INSEE representative given that the institute draws the sample and that the National Council for Statistical Information (CNIS) validates the surveys.

Evaluation criteria

In order to be eligible, surveys must have an exclusively scientific purpose; commercial use is not allowed.

The projects are evaluated according to the following criteria:

- The scientific quality of the proposal: objectives, state of the art, originality of the proposal, relevance of the chosen scales or indicators, expected results;
- The relevance of the data collection through the ELIPSS panel: feasibility (sample size, adequate formatting of the questions for the tablet…), concision of the questionnaire (particularly by using questions from the ELIPSS annual survey or other existing ELIPSS surveys), longitudinal dimension, exploitation of the technical opportunities linked to the Internet and to the tablet;
- The proposal's methodological interest: comparability of the results with other sources, methodological innovation…

Rules on data dissemination, survey time, and panel member relations are detailed in the call for proposals, particularly:

- Selected projects shall be the subject of an agreement establishing co-ownership of the data between the project research team and DIME-SHS. This agreement provides for the submission of the data to the CDSP and authorises the dissemination of data to the scientific community after a one-year period of exclusive use for the project research team.
- Given the limited query time available to the scientific community, long survey proposals (more than 30 minutes) and longitudinal surveys particularly need to be justified. The cumulative annual duration cannot exceed sixty minutes.
- No financial incentive can be offered to panel members. In addition, panel members cannot interact with one another.

Submitted projects

Since the fist call for projects in 2011, 20 survey proposals have been received, of which 11 were accepted by DIME Quanti's scientific and technical committee.

Table 6: Survey proposals in response to calls for projects

	2011	2012	2013
Number of proposals	5	8	7
Number of surveys selected	4	5	2
Number of surveys in the process of being evaluated	0	0	2

From December 2012 to April 2014, twelve surveys were administered to the ELIPSS panel. In addition to the survey on digital practices that was administered upon entry into the panel, and the ELIPSS annual survey, other surveys have addressed a variety of topics such as cultural practices, contraception, political values and opinions, the environment, the couple, the family and intergenerational relations, and health and occupational exposure (see table 7).

The scientific and technical committee has also accepted two comparative projects outside of the call for projects.

The first one was administered to the ELIPSS panel in April 2014. It was a project carried out by Jon Krosnick to replicate classical U.S. experiments on several online panels across the world. The 18 questions aimed to study the formulation of questions, the tendency to acquiesce, non-answer options, the order of questions and procedures for answering.

The second one will be administered on ELIPSS in May 2014 at the same time as the other probability online panels in Europe: the GESIS panel, the GIP and the LISS panel. This jointly built questionnaire uses questions from major comparative surveys (ESS – European Social Survey, SHARE – Survey of Health, Ageing and Retirement in Europe, PIAAC – Programme for the International Assessment of Adult Competencies, EES – European Election Study). Besides the production of comparative data at the time of the European elections, the primary goal is to successfully organise this simultaneous collection from the four panels so that other comparative projects can be considered in the future[18].

Survey production

Questionnaire development

Following the draft survey selection, methodological and technical discussions are held between the research team and ELIPSS team. Before it is launched in the ELIPSS application, the questionnaire is edited to take into account the CST's recommendations,

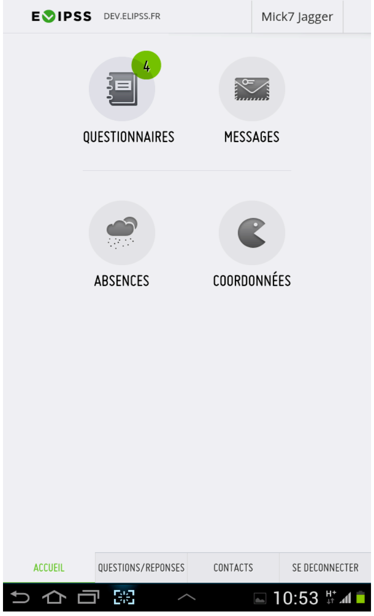
[18] This collaboration between the GESIS panel, the GIP, ELIPSS and the LISS panel has also led to the submission of a joint article in March 2014, entitled " A Comparison of Four Probability-based Online and Mixed-Mode panels in Europe" for a special issue of *Social Science Computer Review.*

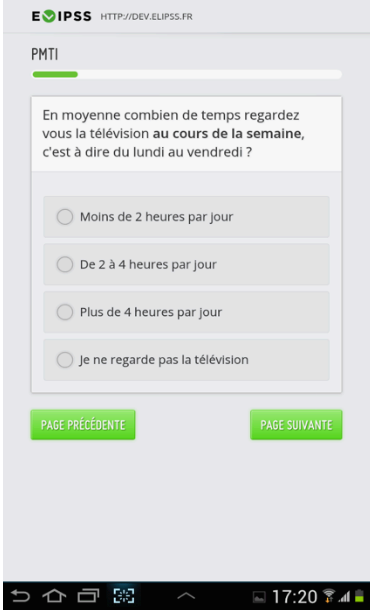
technical specifications, necessary software developments, the non-response options selected, and pre-testing by the research team and ELIPSS team.

As a general rule, questionnaires are introduced on the 1st Thursday of the month and run 3 to 5 weeks. The run can be shorter. This will be the case in 2014 for pre-electoral and post-electoral surveys, whose runs depend on election dates. The run can also be longer. This is the case twice a year to accommodate the Christmas holidays and summer vacation. In addition, some runs were extended by several days following technical problems.

The tablet as a collection device

Panellists respond to questionnaires via an internally developed application that is pre-installed on the tablet's homepage (see the screenshots below)[19]. The questionnaire design and online data collection are based on Blaise software. Developed by Statistics Netherlands, this product is designed for national statistical surveys and is used by most national statistical institutes, including INSEE. In order to publish the survey online with Blaise IS, we had to develop a style sheet adapted to the tablet's touchscreen.





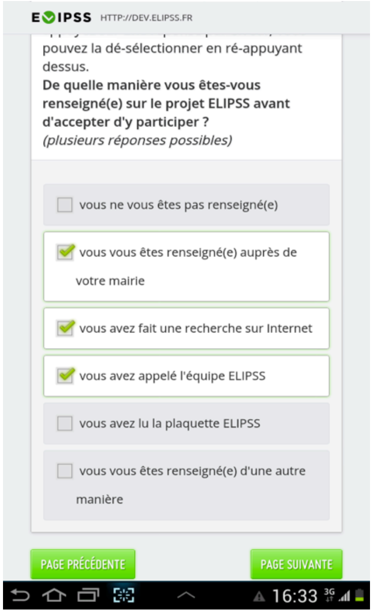
En moyenne combien de temps regardez vous la télévision **au cours de la semaine**, c'est à dire du lundi au vendredi ?

○ Moins de 2 heures par jour

○ De 2 à 4 heures par jour

○ Plus de 4 heures par jour

○ Je ne regarde pas la télévision



pouvez la dé-sélectionner en ré-appuyant dessus.
De quelle manière vous êtes-vous renseigné(e) sur le projet ELIPSS avant d'accepter d'y participer ?
(plusieurs réponses possibles)

□ vous ne vous êtes pas renseigné(e)

☑ vous vous êtes renseigné(e) auprès de votre mairie

☑ vous avez fait une recherche sur Internet

☑ vous avez appelé l'équipe ELIPSS

□ vous avez lu la plaquette ELIPSS

□ vous vous êtes renseigné(e) d'une autre manière

Some draft surveys have provided an opportunity to develop functions, which were made possible because all the panel members have the same connected and touchscreen mobile equipment. For example, since the tablets have a microphone, we were able to invite panel members to record their answers to open questions on the environment. To use self-recording in the October 2013 survey, we first had to inform the panel members whose storage space appeared insufficient on our Mobile Device Manager. The drag-and-drop functionality was specifically developed for the survey on categorisations and

[19] This application also allows panel members to send messages to panel managers, indicate periods of unavailability, change their contact information and find answers to frequently asked questions.

knowledge of the social world scheduled for July-August 2014. Panellists will be able to intuitively drag profession labels to create their own social groups.

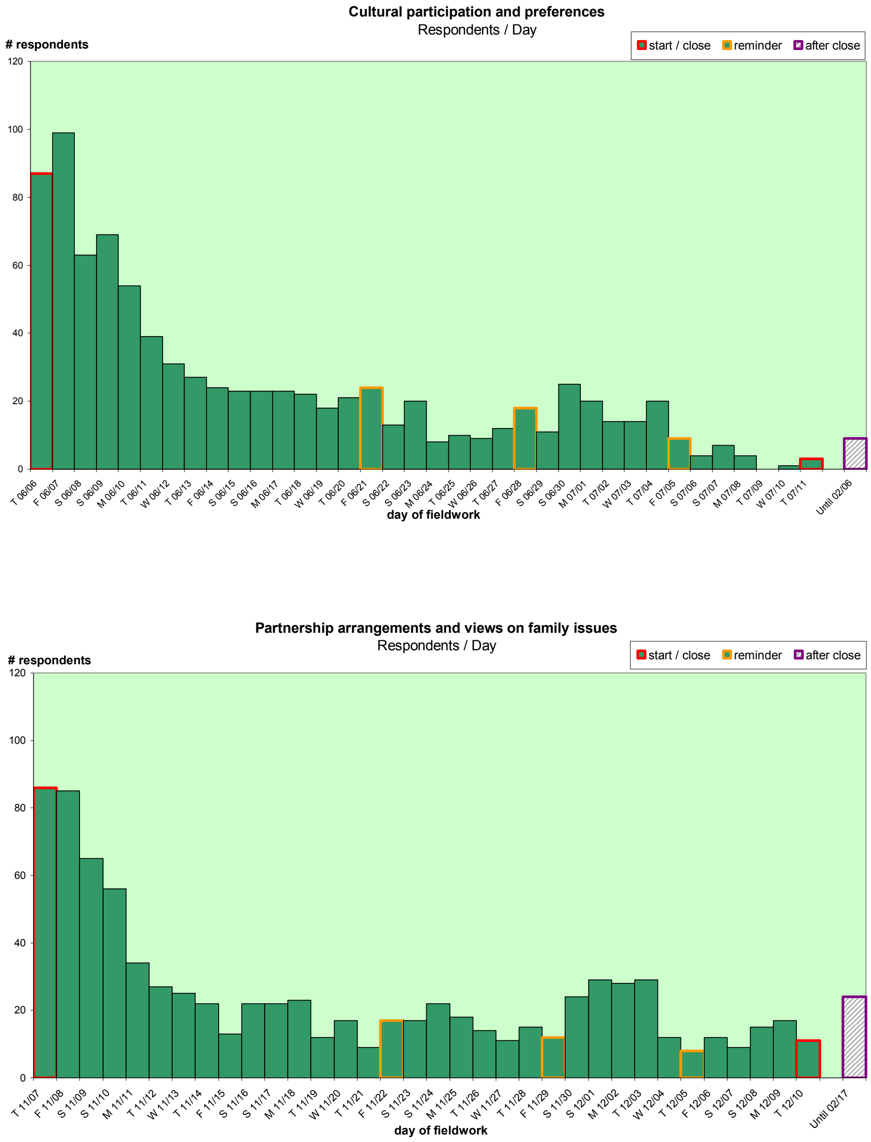
While the homogeneity of the Internet equipment provided to panel members is an undeniable advantage from a methodological (Callegaro, 2010) and technical perspective, the tablets also have limitations linked to connectivity problems and screen size. For example, the autocomplete function that allows people to search for their answer in a list (of countries, municipalities…) requires a good Internet connection. Similarly, to avoid having to scroll, the text length for questions and the number of response options must be limited, especially for question sets. Indeed, given the screen size the question set must be limited to five response options. This limitation was particularly vexing in the comparative project with other European panels. In order to keep some of the European Election Studies questions in the form of a set, the number of response options had to be limited to 6 for the ELIPSS, while the GIP and the LISS panel were able to use the original versions offering 11 response options. The use of question sets is discussed in the literature, particularly for web surveys (Couper, 2008). As Mick Couper suggested when he came to Paris in October 2013, the ELIPSS panel offers a unique opportunity to study the way people respond to question sets on a medium-sized screen, between that of a computer and a smartphone. During the November 2013 survey on couples and families we were able to conduct an experiment with the consent of the research team: half of the panel members answered some questions presented as a set, while the other half answered these questions as standalones (one per page).

Participation in the surveys

Maintaining participation

Reminders for panel members to participate in the monthly surveys are issued in several ways. Following an invitation to the panel on the first day of fieldwork through a message on the tablet (and by email when the email address is available), automatic reminder messages are sent to the tablets of the non-responding on Fridays at the start of the second week of fieldwork. The charts below show the fluctuations in the number of respondents by day of fieldwork for the May 2013 survey (the first survey with automatic reminders) and the November 2013 survey. For both surveys, a first wave of answers came in after the invitation. For the second survey, there were visible peaks on the weekends. In the latter case it is hard to know yet whether this phenomenon can be associated with an effect from the Friday reminders.

Figure 5: Daily participation in the surveys of May and November 2013



Participation in the surveys also involves regular contact between the panel members and panel managers. In the first months, the contact primarily consisted of addressing technical problems linked to the tablets and the 3G connection, and incidents in which the tablet was broken, lost or stolen. This allowed us to establish different processes to manage these situations. Reminder procedures were put in place to follow up with panel members who do not respond to several surveys. The "sleepers" (no response to both the survey underway and the previous one) and the "invisibles" (no response to at least two surveys) receive a personalised follow-up by phone and postal mail each week.

Those who do not respond to the survey underway or respondents who started but did not finish it[20] receive follow-ups depending on the panel manager's workload and the non-response rate in these two groups.

[20] Panellists who start a questionnaire are granted an additional one-month extension to complete it.

During these personalised reminders, panel members often cite technical problems and theft or loss of the tablet as reasons for not responding. If panel members prove technically unable to respond to the surveys, they are removed from the panel.

In order to facilitate tracking of the panel members, the ELIPSS team's developer created an online panel management tool[21]. This enables panel managers to keep a record of all their interactions with panel members.

Monthly participation and attrition

Table 7 details panel members' participation in surveys administered since they joined the panel. Through November 2013 the participation rate (this rate is the COMR described by Callegaro, Disogra, 2008) was above 85% and then started to decline, falling to 77% in the February 2014 survey[22]. While this was a significant drop in participation, it is important to note that the latter survey was the only one to have a four-week run without an extension of the deadline; also, due to technical problems reminder emails were not sent. The March 2014 survey experienced one of the highest response rates, reaching 90% after a reminder campaign and run extension by several days. A major effort was made because this was the annual survey, which is used to update panel members' socio-demographic characteristics.

The attrition rate started to increase at the end of 2013 and reached 8% in March 2014. This reflected the first deactivations of "invisible" panel members that took place following the systematization of reminders to this group.

Table 7: List of surveys administered to the ELIPSS panel since December 2012

Run	Survey	Response rate	Number invited	Attrition
Dec.2012-March 2013	Surveys and Internet	99%	943	0%
April 2013	2013 annual survey (identification module)	91%	1012	0%
May 2013	2013 annual survey (barometric module)	87%	1011	0%
June 2013	Cultural practices, the medias and information technologies	88%	1011	0%
July 2013-August 2013	Fertility, contraception, sexual dysfunctions	87%	1005	2%

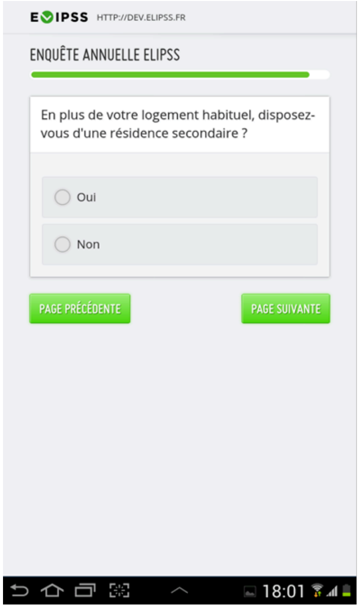
[21] This tool benefited from the experience of the Dutch CentERdata research institute, which is responsible for the LISS panel. The institute allowed us to use their own panel member management system.

[22] Dynamob's pre-municipal wave in March 2014 experienced an even lower rate of 64%, which can mainly be attributed to a run of only 2 weeks that was unprecedented for both the panel members and the ELIPSS team.

Run	Survey	Response rate	Number invited	Attrition
September 2013	Mobilisation dynamic - wave 1	87%	996	2%
October 2013	Survey on values, the environment and energy	85%	997	3%
November 2013	Couple status, fertility intentions and opinions on the family	88%	983	3%
Dec. 2013-Jan. 2014	Health, work and environment. Survey on exposures to inorganic dust/ Mobilisation dynamic – wave 2	83%	993	3%
February 2014	Intergenerational relations through the lens of solidarity and social justice norms	77%	984	4%
March 2014 (2 weeks)	Mobilisation dynamic - wave 3 (pre-municipal)	64%	945	8%
March 2014	2014 annual survey	90%	945	8%

Partial non-responses

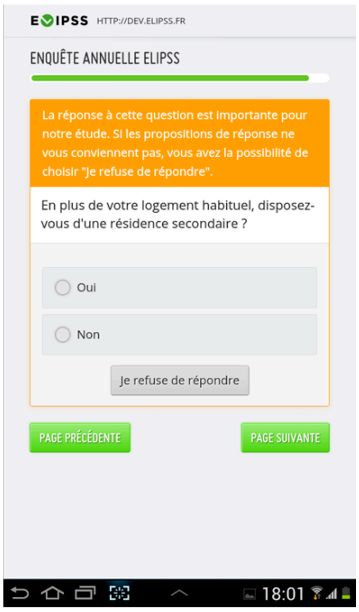
Partial non-responses to surveys also were a particular focus in questionnaire design. Several non-response options are available and are selected for each question when the questionnaires are drafted. A non-response can either be included as a substantive response category or can be distinct through the display of specific buttons (refusal and/or do not know). A reminder message can be posted after the panel member tries to move on to the next question and can be paired or not with specific non-response buttons (see the screenshot below).



En plus de votre logement habituel, disposez-vous d'une résidence secondaire ?

○ Oui

○ Non



La réponse à cette question est importante pour notre étude. Si les propositions de réponse ne vous conviennent pas, vous avez la possibilité de choisir "Je refuse de répondre".

En plus de votre logement habituel, disposez-vous d'une résidence secondaire ?

○ Oui

○ Non

Je refuse de répondre

Data dissemination

Access to the data

The Centre for Socio-Political Data, which is one of the three French data centres for the social sciences, is responsible for documenting and disseminating the data produced through the panel. The surveys are documented according to the Data Documentation Initiative international norm and are published online.

Once the one-year exclusivity period for the teams that co-produced the surveys lapses, the latter will be filed in the survey catalogue of the Quetelet Network portal (French network of social science data centres). Beginning in the fourth quarter of 2014, the first data files will be freely accessible to French and foreign researchers, PhD students, post-doctoral fellows, and Masters students, exclusively for research projects.

To obtain the data files, requesters will have to sign a user agreement in which they commit, in particular, to respect the confidentiality of respondents, to not re-disseminate the data to a third party, and to cite the data source in publications.

Matching

To guarantee data confidentiality, it will be impossible to match all the individual data from the panel. Only the identification module of the ELIPSS annual survey will be systematically matched with each survey file. Any request to match data drawn from several surveys (outside of the longitudinal surveys) will be strictly monitored and submitted to the scientific and technical committee for review.

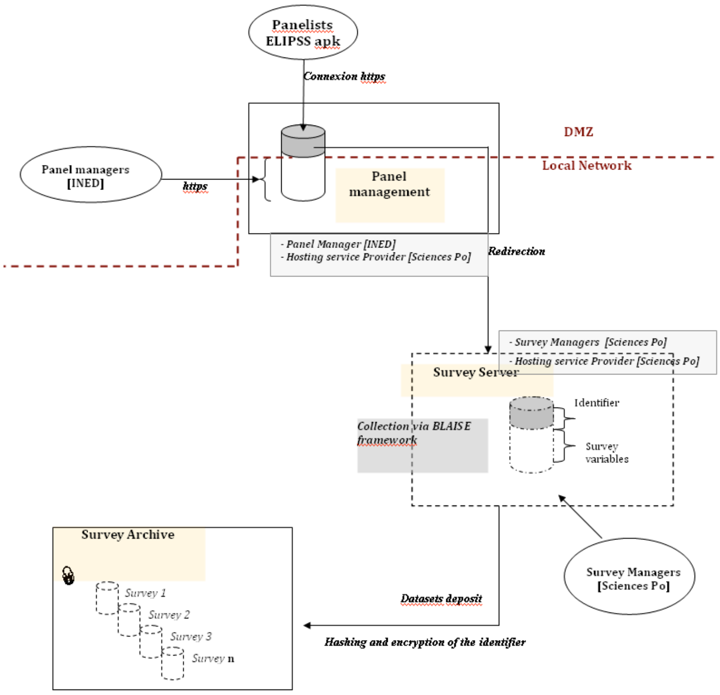
In addition, the matching of panel information with external data (tax, health, etc.) is precluded.

Data confidentiality

The ELIPSS panel has been declared to the National commission for data protection and privacy (CNIL) and is registered with the CNRS under the number 2-12030. An authorisation request from CNIL might be necessary for surveys selected by the scientific and technical committee if they include one or several sensitive questions (ethnic origin, political, philosophical or religious opinions, union membership and sex life).

The security of the information collected from the ELIPSS panel is paramount. The personal data and the survey data are stored in two different information systems. From a technical standpoint, data encryption and restrictions implemented to monitor data matching are further guarantees of confidentiality.

Figure 6 - Infrastructure informatique



Assessment and prospects

The construction of a system like the ELIPSS panel involves a wide range of activities that reflect different stages in the lifecycle of the data, from the draft survey to the posting of the data produced. The pilot, which aimed to test the feasibility of such a panel in France, consisted of inventing something that had not previously existed. First, the recruitment of panel members entailed a relatively complicated process to help us determine the best approach to developing the panel in 2015. Next, ELIPSS' innovative model of providing a touchscreen and 3G subscription to each panel member required a contract with a mobile telephone operator, which had technical, operational, legal and financial implications. We also had to internally develop software tools since no existing tool fulfilled all of the project's needs. Several of these operations, such as the contract with the telephone operator, were essential prerequisites to the construction of a panel, and are now set for the transition to 5,000 panel members.

The pilot ELIPSS panel uncovered difficulties to take into consideration, and ways to integrate improvements, in the development of the 2015 panel. First, using an address database to recruit panel members has disadvantages. INSEE provides it in a form that requires substantial correction and verification of addresses. Furthermore, a description of all household members is needed before a person can be selected and invited to participate in the panel. All surveys conducted on the basis of an address sample face these constraints, but recruitment for the pilot faced additional difficulties which were specific to the selected process and the timing of the project: the fieldwork that partly took place during the summer and the delay that resulted from contracting with the operator; the mail invitations and telephone reminders that turned out not be as effective as expected; the signing of the agreement, which added a step to convince selected people and made the commitment to participate official. While the network coverage was better than expected (no eligible person was excluded for this reason), the tablet had the foreseen effect of encouraging people to participate. Another of its advantages is that all the panel members have the one and same device to answer questionnaires, enabling control over how they are displayed.

This assessment should also note the intensive work involved in producing one 30-minute questionnaire per month, in addition to panel and storage management, software development and service provision for the academic community.

In conclusion, we need to draw lessons from the pilot to continue building and sustaining the ELIPSS panel with 5,000 individuals until October 2017.

We tested different contact options, primarily in an attempt to minimize costs, but it is clear that face-to-face contact is the most effective strategy for recruiting panel members. It would therefore be difficult to do without the services of a survey organization for the recruitment scheduled for the beginning of 2015. The size and quality of the sample are at stake.

We can consider a two-staged protocol at most: a letter inviting interested people to sign up directly on the dedicated website and otherwise announcing a visit from a interviewer, followed by face-to-face recruitment of those who fail to respond. Indeed, the online sign-up option allowed us to easily collect almost 15% of positive responses at the household level. For methodological and technical reasons, we aim to continue equipping panel members with tablets all sharing the same format, operating system, and pre-installed ELIPSS application. However, these recruitment and Internet equipment choices are not currently funded; complementary funding from the Ile de France region has been requested.

In addition, the recruitment of new panel members in 2015 should be better prepared. We will obviously draw from the experience we acquired during the pilot. The process will also be simpler and will be entirely handled by the survey organization. Moreover, we intend to strengthen the team by hiring a person who will specifically be responsible for preparing and monitoring the face-to-face fieldwork.

Finally, this period will provide a unique opportunity to conduct experiments and to study, for example, a possible professionalization by comparing the response behaviour of old and new panel members. During the pilot, methodological work was generally of secondary importance in relation to establishing the panel and producing monthly surveys, because the project's workforce was much too small. Given the planned strengthening of the team, methodological research, which is essential to monitoring and maintaining the quality of the system, should play a larger role in the project.

References

Blom A., Gathmann C., Krieger U., The German Internet Panel: Method and Results, 2015

Callegaro M., 2010 – Do You Know Which Device Your Respondent Has Used to Take Your Online Survey?, *Survey Pratice*, vol.3, n°6, (http://www.surveypractice.org/index.php/SurveyPractice/article/view/250/html)

Callegaro M., DiSogra C., 2008 – « Computing Response Metrics for Online Panels », *Public Opinion Quaterly*, vol.75, n°5, p.1008-1032

Couper M., 2008 – *Designing effective web surveys*, New York: Cambridge University Press

Das M., Ester P., Kaczmirek L.(eds.), 2011 – *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies*, Boca Raton: Taylor & Francis

Gombault V., 2013 – « L'Internet de plus en plus prisé, l'internaute de plus en plus mobile », *INSEE première*, n°1452, juin

Knoef M., de Vos K., 2009 – The representativeness of the LISS, an online probability panel, Tilburg, CentERdata, 29p.
(http://www.lisspanel.nl/assets/uploaded/representativeness_LISS_panel.pdf)

Leenheer J., Scherpenzeel A.C.„ 2013 – Does it pay off to include non-Internet households in an Internet panel?, *International Journal of Interent Science*, vol. 8, n°1, p.17-29

Scherpenzeel A., 2009 – Start of the LISS panel: Sample and recruitement of a probability-based Internet panel, Tilburg, CentERdata, 9 p.

(http://www.lissdata.nl/assets/uploaded/Sample_and_Recruitment.pdf)

DIME Web

Presentation of the instrument

Dime Web helps researchers study digital traces, and the web in particular, through a variety of means known as digital methods. Researchers in the humanities have the opportunity to access data that can be considered a new social trove, but that are difficult to exploit because of their size, dynamism or complexity. Dime Web offers tools, training and methodological support to researchers who need help including digital fields in the scope of their research.

The team includes two engineers with research experience. Mathieu Jacomy is in charge of the project, upgrades to it, and the development of its interfaces and tools. Benjamin Ooghe-Tabanou handles server processing and storage, as well as the scripts capturing digital traces. Both provide support to researchers depending on the projects. They can also draw on the médialab's resources to complement their skills (designers, developers and researchers).

Challenges

Dime Web is dedicated to providing researchers with *adapted means*, which include making tools available as well as transmitting the skills needed to use them.

Available tools are not always adapted to researchers in the humanities. Most software for processing digital traces was developed for other uses (monitoring, management…) and in line with other quality criteria (performance, profitability…). Researchers have a long history of eschewing these tools, not without some major drawbacks. The first difficulty is to translate research subjects into tool operations (and vice versa). The second difficulty is to access documentation on the processing operations carried out (algorithms, conversions and approximations); this documentation is often non-existent. Finally, the last and most common difficulty is the high level of technical expertise needed (lines of code, regular expressions…). Dime Web helps researchers overcome these problems through support, documentation, or the creation of new tools.

The cost of software solutions is another barrier to the democratisation of digital methods in the social sciences. Researchers are not a strategic economic market for software companies, and as a result the technology gap has widened in these disciplines. However, it is possible to open access to new tools in cases where the technology exists but the interfaces and documentation are missing. This is particularly true for the collection of formatted data on the web, called "scraping": it is possible to deploy this type of technique in several lines of code, but no graphical interface was available until very recently (beginning of 2014). In such cases there is an economic opportunity, since little effort is required to provide the service. Dime Web is contributing to the democratisation of digital methods by developing the missing pieces of software when the cost/benefit ratio is favourable. Our developments are always shared, through the deployment of online tools accessible to all or through open source publishing of the code. Thus, our efforts complement those of other teams who share the same goals, and with whom we coordinate (particularly Richard Rogers' team in Amsterdam – Digital Methods Initiative, and Paolo Ciuccarelli's team in Milan – Density Design Lab).

Operation

Researchers are supported on a project basis. They are required to apply, and a Scientific and Technical Committee specific to Web activities makes the selection on the basis of scientific relevance and technical feasibility criteria. Each project has its own research themes, problems, methods and needs. We try to create a personalised approach to each project by adapting to each one's methodological requirements. In practice, technical needs, while the need for support and transfer of know-how remains constant. Our support will take a more stable form that we will develop as we exit the equipment' construction phase.

Alongside support for researchers, we are developing the *Hyphe* crawler, which enables the creation of a personalised body of web documents. Its aim is to remove a major methodological hurdle in the study of online communities. A later section of this report presents it in greater detail.

Status of the project

Progress on selected projects

OpenMarriage – Dynamic analysis of controversies over "marriage for all", with Eric Dagiral (CERLIS). Autumn 2012 call for projects.

A collection of comments on selected articles from LeMonde.fr, with open source publishing of the scraping code.

A launch meeting in March 2013, a session to collect the Le Monde article comments (2 days, reusable), and a meeting in December 2013 (in addition to online exchanges). Health issues have slowed collaboration, which is currently on hold.

SitPol – Exploratory mapping of the political web, with Dinah Galligo at the Sciences Po Library. Autumn 2012 call for projects.

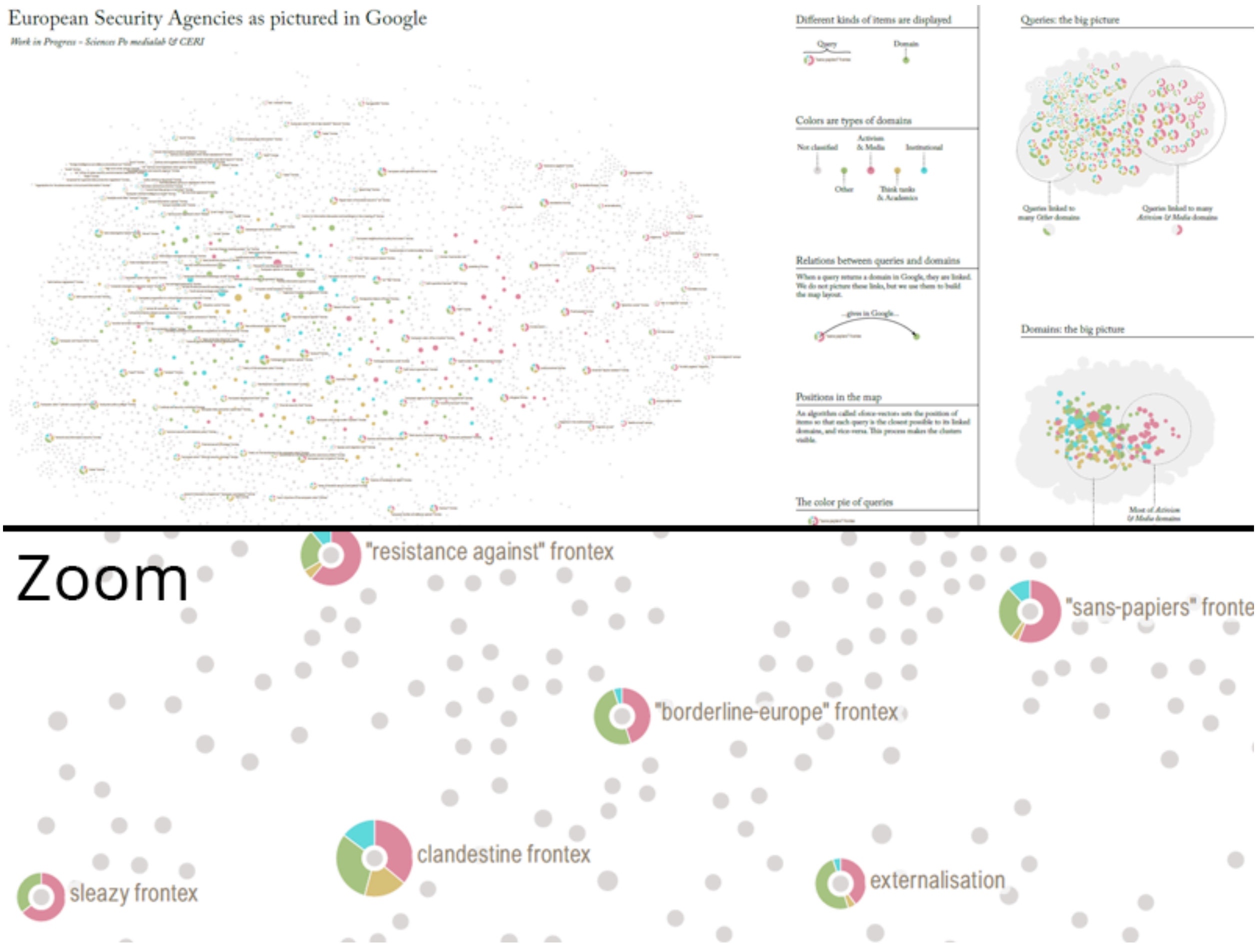
We have trained the librarians on using Hyphe, and we have crawled their extensive list of websites. Hyphe's functions for tailoring definitions of web features have largely been deployed. The network of hypertext links has been produced and is featured in another project – "online sitothèque" [website library] (in progress and directly managed by Sciences Po's library).

A launch meeting in March 2013, a Hyphe training at the end of April 2013, a two-week session on specific Hyphe developments in July 2013, followed by the launch of crawls, and a closing meeting in October 2013.

Aesif Web – The online perception of European interior security and border management agencies, with Didier Bigo, CERI. Autumn 2012 call for projects.

We drew a list of 172 Google searches and collected the top 500 results for each one. We mined the textual content of these pages (n-grams) and created different visualisations (see the screenshot below).

An inaugural 2-day spurt in April 2013, two weeks of specific developments in mid-April and the beginning of August, a meeting and a technical spurt in October 2013. Closing meeting at the beginning of 2014.


European Security Agencies as pictured in Google
Work in Progress - Sciences Po medialab & CERI
Different kinds of items are displayed
Query
Domain
Colors are types of domains
Not classified
Activism & Media
Institutional
Other
Think tanks & Academics
Relations between queries and domains
...gives in Google...
Positions in the map
The color pie of queries
Queries: the big picture
Queries linked to many Other domains
Queries linked to many Activism & Media domains
Domains: the big picture
Most of Activism & Media domains
Zoom
"resistance against" frontex
"sans-papiers" fronte
"borderline-europe" frontex
clandestine frontex
sleazy frontex
externalisation

When a query returns a domain in Google, they are linked. We do not picture these links, but we use them to build the map layout.

An algorithm called «force-vector» sets the position of items so that each query is the closest possible to its linked domains, and vice-versa. This process makes the clusters visible.

Figure 7 - Google searches on Frontex for AESIF Web

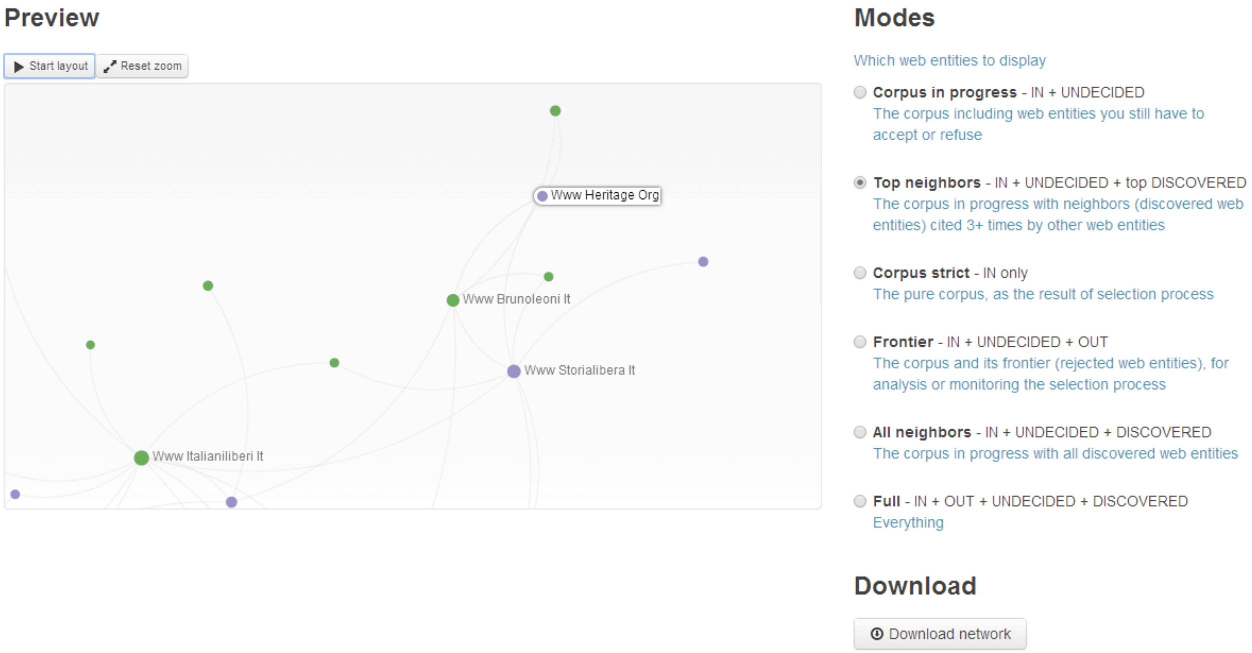
Amour2Pré – Love is in the field and elsewhere, a sociology of belief in love through the articulation of married life and the reception of cultural productions. Christophe Giraud, CERLIS. Spring 2013 call for projects.

The project is awaiting broadcast of the Love is in the Field show. Tweets including selected hashtags referencing the show are continuously archived (live stream and api searches) and currently total 30,000 tweets.

A preparatory meeting was held the end of 2013.

HateWeb – Representing the landscape of islamophobic expression in Italy: structure, evolution, and intellectual leaders. Tommaso Vitale, CEE Sciences Po. Spring 2013 call for projects.

The project essentially includes a collection of websites with Hyphe and text content processing. Hyphe is currently deployed, training for researchers was held and the corpus is partially compiled (see the Hyphe screenshot below). A follow-up meeting was also held in April 2014.



Modes

Which web entities to display

○ **Corpus in progress** - IN + UNDECIDED
The corpus including web entities you still have to accept or refuse

● **Top neighbors** - IN + UNDECIDED + top DISCOVERED
The corpus in progress with neighbors (discovered web entities) cited 3+ times by other web entities

○ **Corpus strict** - IN only
The pure corpus, as the result of selection process

○ **Frontier** - IN + UNDECIDED + OUT
The corpus and its frontier (rejected web entities), for analysis or monitoring the selection process

○ **All neighbors** - IN + UNDECIDED + DISCOVERED
The corpus in progress with all discovered web entities

○ **Full** - IN + OUT + UNDECIDED + DISCOVERED
Everything

Download

Figure 8- HateWeb corpus in the process of being compiled in Hyphe

Projects at the margin – We have deployed Hyphe on numerous occasions to gather user feedback outside of calls for projects: for a mapping of Sciences Po's institutional websites with Ève Demazières, for an analysis of the controversy on family planning, followed by one on caesareans, for an analysis of the controversy over adapting to climate change, and for student projects.

Other operational progress

Development of Hyphe - Two versions published in 2013 that are easy to install on different Linux distributions. Contribution to the creation of the ScrapyD generic package for CentOS. With respect to the interface, the integration of Domino.js, a CSV diagnostic interface, the exporting of graphs and a corpus overview that is 50% complete. Hyphe is described in the appendix 4, p. 59.

Other developments - ScienceScape, a tool to easily produce graphs from scientometric data. It is used by students and during Gephi training to obtain data that is easy to interpret.

Dissemination - We have given presentations on Dime Web, Hyphe and/or digital methods 11 times in 2013 (King's College in London, University of Amsterdam, University of Göteborg, CEREQ in Marseille…).

Project cycle - Our current tasks include drafting two tenders, creating a roadmap, organizing 2 CSTs, establishing a mechanism to report work time, and proposing a more detailed product offering. We have also held a series of interviews with various researchers interested in the project (Yann Algan, Fabrice Epelboin, Dana Diminescu, Lionel Villard, Tommaso Vitale, Cécile Brousse…)

Academic publications - An article is planned for the AESIF Web project. The operational team has submitted a scientific article to PlosOne journal on a network spatialisation algorithm that we use in Hyphe and Gephi. A first revision of the article was re-submitted in March 2014. A publication presenting Hyphe software challenges is currently being drafted.

Prospects

Tender process – The process of issuing calls for projects two times a year is unsatisfactory for several reasons. Researchers in need of our support often have a hazy understanding of the opportunities that digital methods offer, as well as of their constraints and limitations: putting together an application is therefore a difficult, and even dissuasive, task. Three of the nine applications we have received over two tenders were so inadequately justified that the CST was simply not able to evaluate them. Moreover, some of the informal requests we receive require quick, one-off forms of support that the schedule of calls for projects does not allow us to pursue because it is too slow, and because the application is too burdensome in relation to the degree of support requested. We are therefore in the process of modifying the tender to take into account small and large projects, with a view to encouraging researchers to discover digital methods on small projects and propose more ambitious ones in a second phase.

Business plan - The self-financing constraint, which is one of the Equipex rules, raises many issues. To manage this aspect we are also currently drafting a business plan including the aforementioned thoughts on the tender process. Dime Web could raise funds by offering services for a fee: methodological support, hosting a corpus on a dedicated infrastructure, training, specific IT development… The test phase has enabled use to consider this option in view of the needs mentioned thus far. While we believe there is a market for such services, many questions remain: price and budgetary balance, size of the demand, types of contracts, and legal form of the entity managing the service…

These issues will be our main challenge in coming years.

Ethical rules

Our discussions with Sciences Po's Information Technology and Civil Liberties contact allowed us to identify the right procedure to protect personal data. The Dime Web team commits to ensuring that rules protecting personal data are respected (declaration of the database and de-identification if necessary). In the vast majority of cases Dime Web's activities consist of gathering public data on the web without creating databases of names or profiles, even if they appear in the data. Thus, most of our activities do not raise ethical concerns.

The team also commits to alert and advise the researchers with whom it collaborates on these issues.

Furthermore, we respect restrictions on the use of data gathered from private providers – typically Twitter's prohibition on redistributing raw data.

Finally, our software code is open source and freely available, primarily on médialab's GitHub account.

Appendices

Appendix 1 - Excerpt from the consortium agreement

6.4 THE SCIENTIFIC COUNCIL

This consultative body is informed of the use of the DIME-SHS equipment of excellence by the COORDINATOR, who is a non-voting participant in the SCIENTIFIC COUNCIL's sessions.

6.4.1 Composition of the SCIENTIFIC COUNCIL

The SCIENTIFIC COUNCIL consists of 12 internationally recognised experts in the field of social science methods, nominated by the STEERING COMMITTEE. At least half of these people are experts in their field other than the PARTNERS.

SCIENTIFIC COUNCIL members serve as independent experts and in no way represent the institution(s) to which they belong, be it for professional or other reasons.

The SCIENTIFIC COUNCIL ensures the balanced representation of DIME-SHS' three instruments.

The SCIENTIFIC COUNCIL includes 3 experts in ethics.

The President of the SCIENTIFIC COUNCIL is appointed by the STEERING COMMITTEE and is responsible for convening the SCIENTIFIC COUNCIL's meetings, drafting reports, and distributing them to members of the SCIENTIFIC COUNCIL, the STEERING COMMITTEE and the COORDINATOR.

In order to ensure the fulfilment of his/her PROJECT duties, the COORDINATOR is a permanent guest of the SCIENTIFIC COMMITTEE and can place items on the meeting's agenda. S/he receives all the reports, minutes, and documents produced by the SCIENTIFIC COUNCIL.

The STEERING COMMITTEE determines term limits (3 or 4 years for example) and renewal terms for members of the SCIENTIFIC COUNCIL.

6.4.2 SCIENTIFIC COUNCIL meetings

The SCIENTIFIC COUNCIL meets at least once a year at the invitation of its president. The use of collaborative processes (teleconference, video-conference) is an option. The president of the SCIENTIFIC COUNCIL can convene extraordinary meetings in the event of an emergency upon the written and reasoned request of the COORDINATOR, one or several PARTNERS or members of the SCIENTIFIC COUNCIL.

Unless there is an emergency, the president sends the agenda to the members of the SCIENTIFIC COUNCIL at least fifteen (15) days before the meeting.

6.4.3 Decision-making rules within the SCIENTIFIC COUNCIL

The SCIENTIFIC COUNCIL's meetings are valid if three fourths (3/4) of its members are present or represented. If the quorum is not reached, the SCIENTIFIC COUNICL must be reconvened no later than 4 weeks from the date of the initial meeting. After this second attempt, the SCIENTIFIC COUNCIL's meeting is valid if 1/4 of its members are present or represented.

SCIENTIFIC COUNCIL members can appoint another member as a proxy for a meeting. A member can only serve as a proxy of one member per meeting. All the members of the SCIENTIFIC COUNCIL have one vote.

6.4.4 Role of the SCIENTIFIC COUNCIL

The SCIENTIFIC COUNCIL has the following responsibilities:

- Set scientific guidelines for the COORDINATOR and the STEERING COMMITTEE, and if needed make proposals to amend the scientific project to the COORDINATOR and the STEERING COMMITTEE;
- Provide a scientific perspective on future needs in terms of data for projects, and suggest priorities in the development of databases and linkages;
- Give advice with respect to the operation of the DIME-SHS equipment of excellence from the perspective of both French and foreign users;
- Ensure technological, methodological, legal and ethical oversight on access to confidential data in line with international developments;
- Make proposals on the scientific activities of the DIME-SHS equipment of excellence;
- Maintain oversight on de-identification/anonymity issues, with the occasional assistance of external experts and/or the National Social Science Data Committee (CCDSHS);
- Provide oversight and make proposals on partnerships with other centres providing access at the national level, in order to promote the harmonisation of procedures and standards, and good synergy;
- Ensure that the DIME-SHS equipment of excellence is involved in projects and infrastructures developed at the European and international level.

Appendix 2 – Outline of documents for a qualitative survey

1. (quasi-) Essential documents	2. Complementary documents	3. Archival documents
Research project(s)	Parts of researchers' project assessment report	Researcher's preparatory handwritten project notes
Successful requests for funding	Failed requests for funding	
Total budget	Final statement of project expenditures	Bills, travel orders, expense claims
Paper correspondence between research team members (and for theses, with the thesis supervisor)	Possible selection of electronic correspondence between team members or with the thesis supervisor	Non-selected correspondence documents
Paper correspondence with respondents / the "field"	Possible selection of electronic field correspondence	Non-selected correspondence documents
Team meeting summaries or minutes	*When the minutes are unavailable*: handwritten or electronic notes from team meetings	*Otherwise*: handwritten or electronic notes from team meetings
Survey's preliminary bibliography	Survey's preliminary grey literature – if the documents are not available elsewhere	Grey literature and possibly a press review of the survey when the documents are available elsewhere – especially if they are annotated by the researcher
Selection criteria for interviewees, letters and announcements used for recruiting	All other documents related to recruiting interviewees (if applicable)	

All the final fieldwork documents: observation and interview checklist, projective material, questionnaires, etc.	Preliminary fieldwork documents: draft and test versions	Notes on these documents
Collection: notebooks, recordings, transcripts, body of documentation	Annotated copies of some transcripts or documents	
All documents on the analytical approach, its design and/or implementation	Extracts or examples of how the analysis was conducted	Handwritten notes and correspondence
Unavailable scientific output: contributions to symposiums, non-published articles…	Pre-prints of articles that are only available for a fee	Rough drafts, preliminary versions, reprints

Appendix 3 – ELIPSS

Timetable

		Recruitment of panel members	Telephone operator	Software tools	Surveys
2012	March-12	INSEE draws the sample		Beginning of development of panel member applications, panel and survey management	
	April-12		Operator tender		
	May-12	Preparation of panel member documents		Launch of the elipss.fr site and recruitment application	Blaise training
	June-12	Dispatch of invitation and follow-up letters Polling institute tender	Selection of Bouygues Télécom		ELIPSS annual survey working group
	July-12	Selection of TNS Sofres Follow-up by phone	Contract negotiations		
	Aug.-12				
	Sept.-12	Fieldwork preparation by TNS Sofres		1st version of the panel member application	DIME Quanti CST
	Oct.-12	Dispatch of participation agreements Recruitment by TNS SOFRES	Signing of the contract	2nd version of the panel and survey management application	2012 calls for projects
	Nov.-12		Dispatch of tablets Activation of subscriptions Training by phone	Programming, design and testing of the 1st survey	
	Dec.-12				1st survey on digital practices
2013	Jan.-13				
	Feb.-13				
	March-13				

Key steps 2014-2017

		Recruitment of panel members	Telephone operator	Software tools	Surveys
2014	Jan.-14				
	Feb.-14				
	March-14			Improvement of the panel management tool Development of a storage management tool	
	April-14				
	May-14	Assessment of the pilot Decision on pursuing the project			2014 call for projects open to the whole academic community
	June-14				
	July-14				
	Aug.-14				
	Sept.-14	Polling institute tender			
	Oct.-14				
	Nov.-14	Recruitment of new panel members			
	Dec.-14				
2015	Jan.-15		Dispatch of tables Subscription activation		
	Feb.-15				
	March-15				
	April-15				
2016					
2017	Oct.-17		End of the contract		

Appendix 4 - Presentation of the Hyphe crawler

Researchers in the humanities and social sciences use the web as a source of information and knowledge, and a forum for social exchange. They need to gather evidence of activity to feed their research hypotheses, and to this end seek to create bodies of documentation. Yet documents can come in many forms. The diversity of the sources that can be gathered (tweets, webpages, websites, downloads) greatly complicates the creation of a solid body and raises the issue of body unity.

We developed a methodological solution to this problem in the form of the Hyphe crawler, which collects data through the definition of web entities provided by the researcher. These web entities specify the detail level of each collection and guarantee the unity of each body. This idea sets Hyphe apart from other crawlers and creates a bridge between research concepts (Actors, Arguments, Influence, Networks, Dissemination) and the technical structure of the web (URLs, …). Hyphe is a crawler built to support research problems.

The other crawlers (Heritrix, Issue Crawler…) used by researchers do not actually meet a primary need: the indexing of information. These tools are not made to build bodies, but rather to collect websites or webpages and then index them to make them easier to use. These crawlers copy and reproduce the technical pathways of web content, but the problems underlying the creation of a body for research purposes are entirely different. Researchers need to be able to retrieve their research objects, which are not organized according to the web's technical divisions. For example, a person has several blogs, while a blog has several authors. If the researcher would like to study the traces of actors, s/he will need to crawl different sources at a specific level of detail that we call "web entities".

The principle of web entities is based on a morphological analysis of URLs. Hyphe is able to understand the structure of URLs to reach and differentiate the contents by exploiting their "terms" (a term between each slash "/"). A web entity can be seen as referring to one or several of the terms of one or several URLs. By defining their own web entities, researchers can choose not only to guide the semi-automatic collection of information, but also to label them. In effect, web entities translate research objects. Researchers choose the entities that correspond to their research questions. They can do it a priori (when the research is launched) or most importantly a posteriori, after the webpages are collected. Hyphe allows researchers to maintain control over the creation of their body of research and avoids the black box effect of a great number of collection and building tools.

We will now illustrate our method and the use of Hyphe by describing a concrete research problem. A researcher would like to study the dynamics of marriage for all in the public realm, and especially on the web. His starting point consists of a list of previously identified blogs. He uses Hyphe to explore, based on the assumption that these blogs are gateways to other relevant resources. Like most crawlers, Hyphe recognizes hypertext links and can follow them to expand the body (if the researcher so decides). Our researcher discovers that his initial blogs lead to very different things: blogs, Monde.fr articles, institutional websites… After manually going through the content of these resources, he realizes that only some of these articles are relevant to his research problem (those that address marriage for all). He can then select these articles and turn them into new web entities. Hyphe allows for any grouping of pages to be declared a web entity, regardless of its level of detail: a page, a section of a website, a domain name, or combinations of these. Thus, this process does not require that each Monde.fr article be declared a web entity: only those that discuss marriage for all become web entities. The rest of the Monde.fr website can remain a single web entity that does not include the extracted articles. The researcher builds his web body such that it contains actors (blogs, institutions) and press articles (Monde.fr articles). He can export this body to study it with other tools, for example to analyse which authors cite which articles.

Since research is always an iterative process in which the researcher is constantly redefining the scope of his objects of study, it is often after the collection that researchers would like to redefine their web entities. Other crawlers generally cannot do this, unless the researcher is willing to redo everything, beginning with the collection. The blockage usually occurs at the indexing stage, a costly operation in terms of time and computing power that transforms the collected data into a more accessible and computable form. We developed a different and original indexing engine for Hype that allows for web entities to be redefined without redoing the indexing. We use a different strategy from other crawlers that enables a dynamic declaration of each web entity, that is, the entities can be declared at any point in the process. This feature gives researchers the ability to change the scope of any web entity at the time they see fit to do so.

In conclusion, Hyphe is a tool that was specifically developed for social science researchers. Its specialised configuration makes it more adapted to the methodological needs of digital humanities. While Hyphe is still being developed, its code is accessible online as open source software.