SciencesPo
CENTER FOR SOCIO-POLITICAL DATA

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

Assessment and outlook –2014-2017

DIME-SHS Scientific Advisory Board meeting 26 June 2017

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 a mobile Internet panel;
- DIME Web: collection and analysis of spontaneous expressions on the web.

We brought together these three instruments for two reasons. One the one hand, each instrument relies heavily on digital tools and methods, therefore we wanted to develop synergies between them. On the other hand, one of our overall goals was to make it possible for our users to use simultaneously these instruments, when necessary. Developing these instrument in a common equipment could have fostered multi-methods research. Indeed, at the beginning there were a few proposals using jointly DIME Quanti and Web, but each time they were inadequate for one of the two instruments. For the moment, maybe because of the difficulties we experienced within each instrument, we were not able to promote effectively this possibility. However, the development of multi-methods research based on the instruments of DIME-SHS remains a goal for the years to come.

DIME-SHS was selected in 2011 in the context of the Forward-looking investment programme's first "equipment of excellence" (équipex) call for projects. It is funded until the end of 2019. There is a possibility that this funding might be extended. The Ministry of higher education and research made this suggestion to the body in charge of allocating this kind of funding. This year we had to write an intermediate report for the years 2011-2016. There will be an audition on 28 June 2017. The report and the audition might condition the possibility to get extra funding. If our funding is not extended, we could try to get a new one by submitting a new project to the forthcoming call for "innovative equipment".

In any event, we will do our best to carry on with the development of these three instruments.

[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

Brief recap of the instrument and its purposes

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 in the late 1990s and early 2000s. The primary goal was to bring France to the same level as other European countries, especially the Northern Europe ones, in terms of compiling data from surveys conducted with qualitative methodologies (interviews, observations, etc.). From the beginning, there were multiple objectives behind the creation of the instrument. The provision of data is intended to give the scientific community the means to tap more fully into the wealth of data from surveys, which are often under-analysed, and to make those data 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 rigour 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, which prompted us to make some specific choices. 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, so only actual surveys will be archived, not documents produced by laboratories and researchers.

At the scientific level, our approach seeks to build tools for secondary researchers in order to help them understand the context in which the surveys are conducted, and to give them 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 collate and integrate as much of the documentation as possible in order 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 purpose of the survey of the survey is to help retrace the original survey process in order to reduce the risk of decontextualised use of the data. In addition, each survey is organised and presented as a mini website. The idea behind the "study-site" is to facilitate navigation through the various archived documents and to foster familiarity with the material. The instrument is therefore intended as a toolbox to help familiarise users 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 philosophy, guiding principles and ambitions of the instrument remain the same. Nonetheless, there have been a few adjustments in the instrument's positioning since spring 2014, which we will cover in detail later on (see Appendix 1). In summary:

1- the disciplinary range of the surveys has been specified and somewhat widened.

While the core target is still sociology and political science, multidisciplinary surveys or surveys that are anchored in allied disciplines, such as ethnology or social geography, are eligible.

2- the instrument's development strategy has been refined, in two respects:

- Since spring 2015, a substantial effort has been made to develop both the network for the collection of new surveys and tools for raising awareness on the deposit of surveys or on the procedures for selecting surveys.
- Since 2016, a similar effort has been undertaken to develop the network of survey users, in two directions:
  - ✓ developing a proactive communication strategy addressed to laboratories, educational institutions and professional networks; this strategy also entails producing more publications to promote the instrument;
  - ✓ placing greater emphasis on the pedagogical uses of the instrument, which would seem to constitute a major direction for development and one that is also easier to manage than secondary analysis in the strict sense.

Progress Report (May 2014-May 2017)

We have moved forward on the three lines of development announced in spring 2014.

Expansion and diversification of the survey catalogue

Since 2015, survey deposit proposals have been organised through national calls for proposals (one or two a year), coupled with corresponding laboratory visits. To this end, a formal survey deposit and selection procedure was established in 2015, in concert with the beQuali Technical and Scientific Committee (CTS). A method and support materials have been developed: proposal guide, deposit form (completed by survey depositors, with the help of the BeQuali team if necessary), selection form (completed by the members of the CTS), together with presentation kits designed for laboratory visits. Two national calls for survey deposit proposals were issued at the end of 2015 and in mid-2016; a third call is currently in preparation for autumn 2017.

The CTS itself will be renewed in the coming months, following an early renewal at the end of 2014, when it consisted of 12 members and when a new chairman was elected. At that time, it was joined by 2 representatives from the archiPolis consortium (see below), increasing its membership to 14. After several resignations over time, the CTS has been

reduced to 8 people. New experts will be recruited, maintaining the diversity of backgrounds and profiles which is the basis of the CTS's structure.

Between 2011 and the end of 2015, before the formal procedure for depositing and selecting surveys could be established, surveys were collected directly by means of interpersonal or professional networks, at Sciences Po or beyond (INED, archiPolis network). By this means, the beQuali team collected 10 surveys by researchers who, at least at the time of the production of the surveys, were members of the consortium's laboratories (Sciences Po, INED or EHESS). The processing of one of these surveys had to be stopped after the producing researcher was dismissed.

Since the end of 2015, calls for deposit proposals have been open to the whole scientific community in France and abroad. A deposit proposal must relate to a survey that is part of a research project conducted in a social science discipline (in particular sociology, political science or allied disciplines) using primarily qualitative methods; the materials stored must be sufficiently comprehensive and diverse to offer high potential for reuse, and not pose excessive difficulties regarding the confidentiality, sensitivity, ownership or legibility of the data; the languages used in the survey must be within the capacity of the team; the source researchers must undertake to participate in certain operations (validation of the limits of the material to be retained and of the classification plan, definition of an anonymisation plan, scientific contextualisation). In all, some 15 proposals were submitted, and 9 surveys were selected by the CTS. As of now, half the surveys in the catalogue (already available or in preparation) have been selected by the formalised procedure entailing an open call for deposit proposals and selection by the CTS on formal criteria.

In all, 8 surveys are currently available, and 10 others are in preparation, 3 of them at an advanced stage. At the beginning of 2018, therefore, the beQuali catalogue will consist of 11 surveys.

A significant effort has been made to diversify the catalogue in disciplinary terms (sociology surveys now outnumber political science surveys) and also in institutional terms (surveys from Sciences Po laboratories are now outnumbered by those from other university laboratories). The surveys already online tackle a variety of topics, such as: electoral behaviours, political opinions and institutions; educational and residential choices; the family, the couple and sexuality; gender; the professions; immigration. The new surveys tackle new themes such as heritage, hospitality, social movements, parenthood, territorial governance, far right-wing activism, poverty and exclusion.

The guide to good practices designed to facilitate the process of archiving past and future surveys, announced in the previous report, could not be produced. It seemed to us too costly to undertake such a project, especially as a dynamic has developed at the same time in France around the management of archives and research data, involving very large numbers of people at local or national level, which made the task even more complex. We judged it best to put this project on standby.

Improving the system for making surveys available

Working procedures, methods and techniques have been improved.

The collection and usage contracts, as well as the consent form (designed to obtain the retrospective written consent of survey respondents to disseminate their personal information, when their input cannot be anonymised), were made legally watertight with

the help of a specialist lawyer and with validation by Science Po's legal department. The usage contract is now built into the Quetelet ordering application. In addition, the anonymisation procedure was formalised in consultation with the CNRS data protection lawyer and an ad hoc working group drawn from the CTS. One particular result of this was the preparation of an article that includes a guide to good practices, both ethical and legal, for the dissemination of social science data. This guide is published by a working group called "Ethics and Law", and labelled by DARIAH. The article written by the beQuali team consists of feedback on the anonymisation practices implemented in the handling of surveys.

An internal team seminar was organised in order to stabilise the model and method of this survey on the survey (open access written report and audio chaptering available on the website). The result of this was a first document-related document for internal use, which will be improved in the future.

The outlines of a survey corpus were specified (definition of the documents necessary, or indeed essential, to making a survey available) and the classification and naming plan improved, in order to adapt to the specificities of the new surveys collected and to the need to raise awareness amongst researchers on the survey deposit process.

A protocol for the digitisation of paper (and also sound) documents was established; for reasons of resources and skills, it was decided that this task should be outsourced, other than in exceptional cases. To this purpose, specifications were drawn up, and 6 surveys were digitised by this means.

Professional standard procedures were established for the storage of collected documents during the processing of archives, since the latter are not kept at the CDSP, but returned to the producers or redirected to the appropriate archive departments once processing is complete. In this way, survey archives are reintegrated into the life cycle of archives. The issue of long-term archiving of digital files in CINES was anticipated, via a collaboration with the Huma-num Very Large Research Infrastructure, which acts as an interface between the Equipex and CINES (National computer Centre for higher education). The protocol for CINES deposits is currently being finalised. More generally, a case-by-case review of metadata was undertaken, in order to maintain interoperability between beQuali and other systems or entities with which it needs to communicate (Quetelet portal, new CDSP website, CINES, archive software such as LIGEO, Dataverse archiPolis, etc.).

In the IT field, an overhaul of the beQuali site was completed in 2015. The architecture and drafting of the pages on the editorial site were simplified and made more informative. The tools used to consult surveys were improved: posting of a more accurate and manageable document inventory in Excel format; implementation of secure survey downloading from the Quetelet portal. The back office was also improved to facilitate the team's work: new content management tools; improved ergonomics; improved survey loading; English language interface.

Since 2015, IT developments on the application for making surveys available have been frozen, for three reasons: first, weakness and uncertainties regarding the availability of developers in the IT section; second, the decision to focus on the development by CDSP, on behalf of the archiPolis consortium, of a web application that enables the standardised input of descriptive metadata on surveys stored in that network's member laboratories, as well as their consultation by the whole scientific community; this application, put into production at the end of 2016, took the form of notices published under the

Dataverse application; and finally, the choice to undertake an overhaul of the CDSP website, which went live at the beginning of 2017.

In spring 2017, following the completion of the IT work for archiPolis and the news CDSP website, and the reinforcement of the development team in the IT section, the decision was made to reschedule IT developments on the beQuali website for 2017-2018. These are essentially relatively basic functional improvements to the beQuali site. We are not yet at the stage of developing new survey mining tools, which would require a procedure to consult with the user community and to structure its needs (see below).

Lines of development for 2017-2019

The lines of development for the period from spring 2017 to the end of 2019 are structured around three points, the first two of which maintain and extend the avenues already announced at the time of the previous consultation of the scientific advisory board in 2014.

Continue to develop, diversify and promote the survey catalogue

This will entail the launch of 2 calls for survey deposit proposals (end 2017 and end 2018/start 2019) with the aim of achieving, by the end of the Equipex process, a collection of around 25 surveys in preparation or available for use.

A special effort will be made to further increase the diversity of these new surveys in terms of their disciplinary, institutional or methodological base, or in terms of their subject matter and research topics.

- While it is impossible at this stage to anticipate what deposit proposals will actually be submitted, or which ones the CTS will select, we will make strong efforts to extend the collection sites to more research institutions in France, which will require a continued emphasis on awareness raising through an increase in the numbers of local laboratory visits.
- We will also work with the CTS on further widening the methodological scope of the selected surveys, so as to offer users an increasingly diverse range of different qualitative methods that can be implemented by researchers.

One quanti-quali survey has already been collected and processed since the end of 2014; the quantitative part of the survey had already been made available by the producing institution, INED (member of the Quetelet network), so that it was only necessary to process the qualitative part. We tried to test the processing of a quanti-quali survey (a survey on Parisian retirees by the french sociologist and gerontologist Françoise Cribier) by carrying out the whole processing chain on both parts of the survey; however, it was not possible to support the producers of the research through to the deposit proposal. In this respect, one of the objectives may be – once again without making assumptions about deposit proposals or the CTS's selections – to tackle the problems posed by this type of survey, in collaboration with our colleagues at the CDSP who specialise in the documentation of quantitative data.

Continue to improve the system for making surveys available

The foundations of the instrument were laid down through three test surveys, up to 2014. Since then, the need to manage new surveys, more diverse in many respects – types,

numbers and formats of documents, issues of personal data confidentiality, management of relations with the researchers for survey processing and contextualisation, etc. – confronted us with new problems. These problems helped us to refine possible solutions between 2014 and 2017.

Nevertheless, we are a long way from having experienced all the challenges that may arise in these different domains; when we encounter new surveys that raise new problems, some of these processes will need to be extended and adapted, for example to cover: the anonymisation of photos or audio or video files, or of ethnographic materials; the contextualisation of new survey methods or protocols, or methodologically mixed surveys (quanti-quali) etc. In particular, we plan to finalise the survey on survey production guide, as well as a guide to good practices for potential survey depositors.

On the IT side, we can develop new survey mining tools, which will entail establishing a procedure for consulting the user community and structuring its needs; this procedure could involve the setting up of a user committee, or another formal solution that will perform the same function of elucidating the needs of users.

In the shorter term, we notably plan to finalise the procedure for the long-term archiving of digital files at CINES, and to finalise the allocation of digital object identifiers (DOI) to surveys, as well as to introduce the OAI-PMH harvesting protocol.

Continue to develop the uses of surveys, especially in education

This third goal replaces, while including certain of its aspects, the objective of focusing on methodological research announced in 2014.

With regard to the reuses of available surveys up to now, we think that it is too soon to draw firm conclusions, although the outlines of certain key elements are already emerging; reuse for the purpose of teaching, of methodological research or of secondary analysis would seem, at this stage, to predominate over other possible forms of reuse (reanalysis for verification, revisiting a survey field, history of sciences, etc.). Since 2015, three cases of the instrument being used for doctoral theses have shown that its use in research projects is not confined to established researchers, and that the reuse of surveys at PhD level is possible and should perhaps be further encouraged.

The biggest priority today seems to be to put in place a proactive strategy to promote use of the system, in particular by developing the use of surveys in education, which would seem to be the optimum way to attract the maximum number of users, with the aim that students, or at least some of them, should in their turn become future users as researchers or educators. At this stage, therefore, we plan to continue the strategy pursued since 2016, in several areas (which may be linked):

- continue to develop teaching (in particular on method) and to provide it ourselves using the catalogues in the survey;
- continuing to develop pedagogical kits based on the surveys in the catalogue;
- developing a network of educators interested in using survey archives as new pedagogical resources, in at least two ways:
  - ✓ expanding the activities of the working group, which could notably take the form of a one-day seminar in 2018 or 2019;

- ✓ increasing the number of visits to laboratories for pedagogical training, in different French universities, in particular training focused on methods (we have already initiated contact with educational groups interested in our approach at the University of Grenoble and at University Paris II), together with doctoral training programmes;

- pursuing the networking initiatives already underway at European level, for example within academic frameworks such as ECPR or in concert with equivalent instruments in Europe (for example, we recently began talking to FORS in Switzerland on this subject).

This line of development will mean reinforcing our communication strategy around the project, in particular with professional networks, notably groupings and communities in different academic disciplines. In concrete terms, the aim will be to develop our communication, starting in particular with the beQuali website. Here, one method will be to produce demonstration videos on the possible uses of the beQuali system, or videos showing testimonies by researchers who have deposited or used surveys, in order to inform potential users about the benefits of the instrument. More generally, the aim will be to communicate more, for example via a newsletter or other channels, on the new surveys available in the catalogue, or being added, or on examples of actual uses (publications or papers based on the re-use of surveys in the catalogue, etc.). The likely inclusion of an additional person in the CDSP team, responsible for the laboratory's communication, should make this objective more achievable and tangible. This objective will also mean producing more publications that highlight the different aspects of the project, including in partnership with users of the instrument (whether as depositors or reusers of surveys).

Recommendations of the Scientific Advisory Board at its May 2014 meeting

1. The SAB members suggest that, with regard to the acquisition of new surveys, the fields targeted should be those for which there is the greatest potential for secondary analysis and the greatest interest in this kind of research method (eg. sociology of health, sociology of education or criminology).

DIME Quali answer : We agree with the CTS that it is still too soon to target such specific fields of research. What is important above all is to increase the overall diversity of the surveys. This goal could be pursued in the longer term, when the instrument has been further consolidated and is better known.

2. The SAB members recommend that the DIME quali team should choose between two strategies. They should:

- Either carry on with the current intensive survey documentation. In which case it will not be possible to increase survey diversity given current resource levels. If this path were to be followed, it might be more efficient to focus on a subfield, on a niche (e.g. ethnographic surveys are very difficult to archive). Focusing on a subfield would gain that invisibility and would lower the pressure to be everywhere, at the expense of promoting secondary qualitative analysis;

- Or reduce the documentation of surveys to increase the number of holdings. This will also makes it possible to increase survey diversity.

DIME Quali answer : It has been decided, for the moment, to continue to focus on intensive survey documentation, while increasing survey diversity in terms of disciplines (towards sociology) and in terms of the methods employed (towards ethnography), though without moving into the sphere of anthropological surveys.

3. Secondary analyses are often conducted in closely related fields and match the initial orientation of the surveys. The SAB members suggest that this could have consequences for the contextualisation choices made when documenting surveys.

DIME Quali answer : The further development of the guide for the conduct of surveys on surveys, scheduled for 2018 or 2019, should tackle this issue.

4. The SAB members suggest using the repository project to detect popular surveys. It could be a signal that these surveys have strong potential for secondary analysis and should therefore also be documented and disseminated by DIME quali.

DIME Quali answer : The abandonment of the repository project makes this detection process obsolete.

5. Whatever strategy is employed, the SAB members recommend that the DIME quali team should:

- Expand the number of surveys held, in the interests of the scientific community: the more the better;

- Continue experimenting and promoting the secondary use of qualitative surveys, in particular by training PhD candidates and researchers and funding post-docs.

DIME Quali answer : These recommendations have been followed and will continue to be followed. We will add the expansion of training in research, which constitutes an important reservoir of potential reusers.

6. The SAB members suggest that the DIME quali team should clarify its position vis-a-vis other initiatives in France, for instance those underway in some *Maisons des sciences de l'hommes* (MSH – human science centres) and the national infrastructure specialising in the digital humanities, Huma-Num.

DIME Quali answer : No official position on the MSH has been established, with the exception of the PROGEDO policy, which integrated the MSH into the academic data platforms. The position on Huma-Num has been specified, in 2 respects. First, Huma-Num plays the role of intermediary (as the transferring office) between DIME quali and CINES for the long-term archiving of digital survey files. Second, CDSP's position vis-a-vis the archiPolis network was clarified until Huma-Num's funding for that network ceased (end of 2016). Since the beginning of 2017 and the decision not to renew archiPolis as a Huma-Num consortium, there has been no institutional link between DIME quali and Huma-Num via archiPolis, which is now pursuing its activities as an informal network.

7.The SAB members advise the DIME quali team to be very careful on the issue of confidentiality and to:

- rely on the primary researchers to anonymise data because they are very familiar with their survey;

- establish a set of guidelines for the different kinds of surveys, based on the team's experience;

- promote these guidelines, along with existing guidelines for quantitative surveys;

- conduct a project-by-project risk assessment that takes into account the extent to which interviewees are known to the public or the importance of the identity of the actors to the analysis.

DIME Quali answer : The recommendations of the CS have been followed on almost all the points (see activity report). The only point which requires further exploration concerns the establishment of a policy for the promotion of anonymisation protocols linked with the policy adopted for quantitative surveys at CDSP.

8. The SAB members recommend that the DIME quali team should use indicators to monitor the project that match the goals pursued. They suggest the following:

- Number of surveys

- Number of users

    * Students or trainees

    * Researchers conducting secondary analysis projects

    * International users

- Number and type of publications

- Types of funding

DIME Quali answer : We have not been able to make sufficient progress on this recommendation, but plan to make it a significant action point in the next two years.

DIME Quanti

Brief recap of the instrument and its purposes

The DIME-SHS quantitative instrument is a probability-based online panel dedicated to research in the human and social sciences (SHS). The purpose of ELIPSS (online longitudinal social science study) is to fill the gap in questionnaire-based survey resources available to French social science researchers. Funded by France's National Research Agency, the system is a free access, open source instrument available to the scientific community in France and abroad. It is based on a random sample of the population of metropolitan France, drawn by INSEE from the national census. Survey projects are submitted by research teams in response to calls for projects, and selected by a committee made up of researchers specialising in quantitative methods. Inspired by the Dutch LISS panel, ELIPSS – like similar German panels (GESIS panel and the German internet panel) – uses a probability-based sample and includes individuals without Internet connection. Its originality lies in the choice to provide all panel members – including those who already have an Internet connection – with a touchscreen tablet computer and a mobile Internet subscription, so that they can respond to questionnaires on a dedicated application. The advantages of this costly choice are twofold: 1) the panellists are not paid to answer surveys and can use the tablet for their personal needs, 2) the design of the questionnaires they respond to is entirely controlled.

The pilot study, conducted between 2012 and 2016 with approximately 1000 individuals, defined, tested and developed the different aspects of the system: the method of recruiting panellists, the methods of responding to questionnaires, the procedures for proposing surveys, the procedures for monitoring panellists. Since September 2016, the panel has included more than 3000 participants, equipped with a touchscreen tablet and a 4G subscription, who respond to monthly surveys devised by social science researchers on a variety of topics: political behaviours and opinions, leisure and cultural practices, environment, health and work…

Progress report (May 2014-May 2017)

Results of the pilot study

From April 2014 until the end of the pilot study in August 2016, 12 new transversal surveys designed by research teams were administered (marked * in the table below), 9 new waves of the Mobilisation Dynamics (Dynamob) survey, a new wave for two research surveys already administered previously (marked ** in the table below), two joint surveys of the European panels.

Table 1: List of surveys administered since March 2014

Dates	**Titles**	**Response rates (%)**
April 2014	Cross-National Replication of Question Design Experiments*	88.4
April 2014	Mobilisation Dynamics - wave 4	84.2
May 2014	Comparative Study of European panels & Digital Practices 2014	85.5
May 2014	Mobilisation Dynamics - wave 5	76.6
June 2014	Mobilisation Dynamics - wave 6	84.2
June 2014	Attitudes towards Alternative Partnership Arrangements*	86.4
July-August 2014	Ordinary categorisations and knowledge of society (set of jobs)*	88.6
September-October 2014	Tastes, morals, social groups*	89.5
November 2014	Ordinary categorisations and knowledge of society (set of portraits)*	84.0
December 2014 - January 2015	Extract from the European Social Survey & Mobilisation Dynamics - wave 7	90.4
February 2015	Lifestyle diversity and Energy Consumption Peaks*	86.1
March 2015	Mobilisation Dynamics - wave 8	75.5
April 2015	Annual Survey 2015	90.4
April 2015	Mobilisation Dynamics - wave 9	82.3

Dates	Titles	Response rates (%)
May 2015	Comparative Study of European Panels 2015 / Digital Practices 2015	86.9
June 2015	Mobilisation Dynamics - wave 10	78.5
July-September 2015	Cultural Practices, Media and Information Technologies – wave 2**	87.3
October 2015	Residential Strategies and Location Choices in France*	85.4
November 2015	Survey on Values, the Environment and Energy - wave 2**	83.5
November 2015	Mobilisation Dynamics - wave 11	76.5
December 2015 - January 2016	Mobilisation Dynamics - wave 12	80.7
February 2016	Practices and Representations regarding the State – wave 1*	89.1
March 2016	Mobilities and Relation to Space in the Life Cycle*	85
April-May 2016	Annual Survey 2016 & Digital Practices 2016	89
June 2016	Social Pricing for Water*	75.7
June 2016	Package Deliveries and E-Consumer Mobilities*	78.1
July-August 2016	Health Housing Diet Medicine Architecture*	81.7

Throughout this period, response rates remained above 80%. However, for a few surveys, the response rates were lower. This occured when the campaigns were shorter than the usual five weeks, which was generally the case for the pre-electoral waves of the Dynamob (Mobilisation Dynamics) survey, where the end date was restricted by the date of the

elections. The surveys in June 2016 also attracted a lower response. It should be noted that the Tarifeau (Social Pricing for Water) survey demanded significant cognitive efforts: the respondents were offered several scenarios for social prices on water, with variations in the beneficiary population, the method of funding this public policy, and the additional costs that the respondent would bear. Many of the panellists stopped answering when they reached these scenario-questions, which explains the lower response rate for the survey.

In August 2016, there were still 795 panellists in the panel, representing an attrition rate of 23.4% after 44 surveys in 44 months.

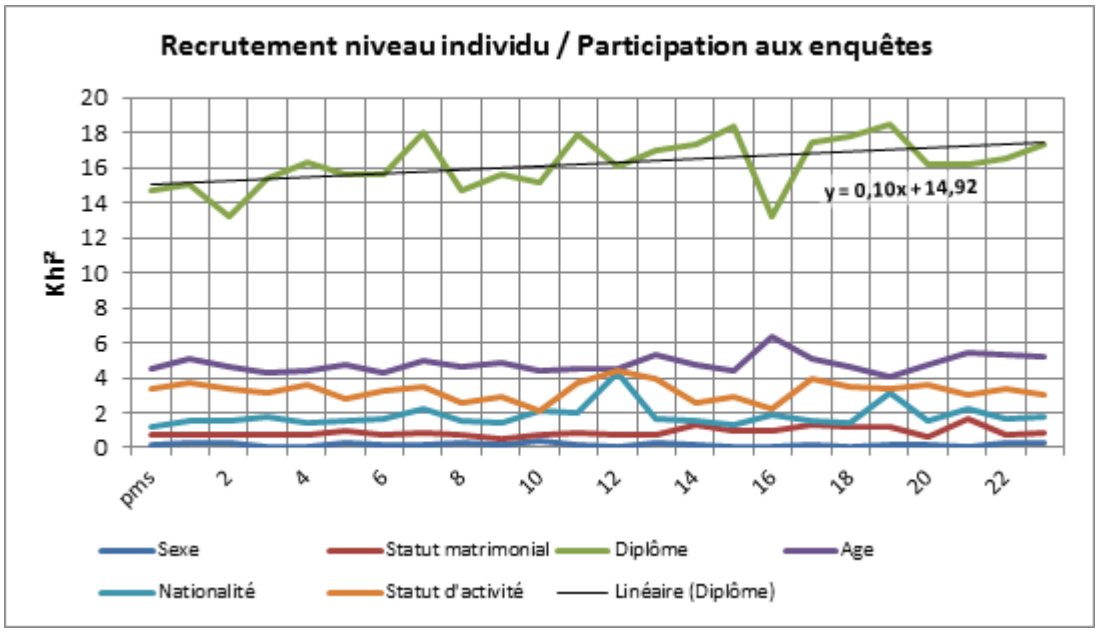
Initial analyses were carried out on the statistical quality of the pilot study by Stéphane Legleye and Nirinstoa Razakamanana, [2] based on data from the first 23 surveys administered between December 2012 and June 2015. Presented below, these analyses were prepared for the pilot study report submitted to the Conseil national de l'information statistique (CNIS – national statistical information council) in October 2015.

Changes in representativeness over the surveys

An important aspect of the statistical quality of the panel is the representativeness of the sample of respondents over the surveys and the attrition rate.

It would seem that the long-term representativeness is little affected by dropouts from the panel: the differences in distribution of certain sociodemographic variables relative to the census remain essentially stable. The only negative shift is in the distribution of educational qualifications across the panellists, as shown in Figure 1.

Figure 1: Representativeness of the panel in sociodemographic terms (sex, marital status, education, age, nationality, and employment status)

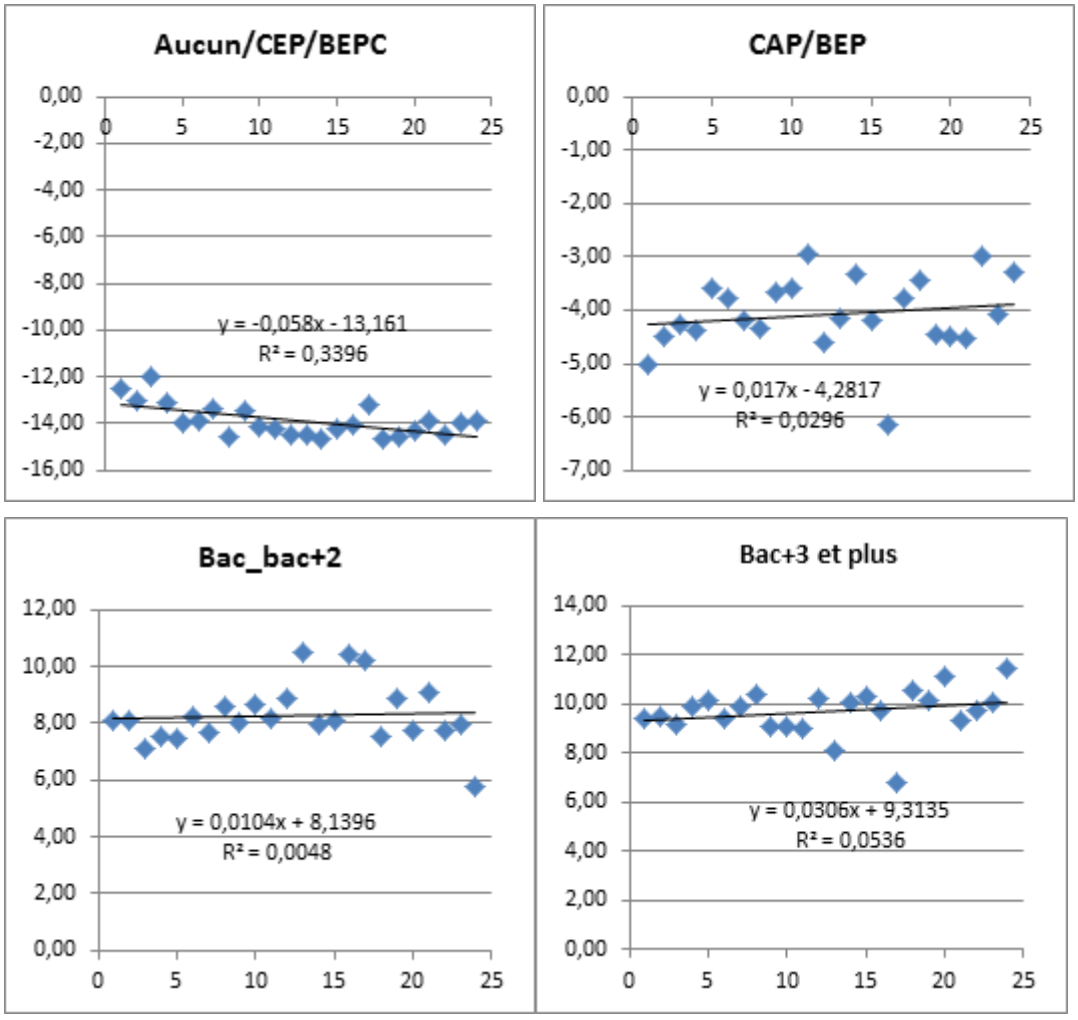


[2] Both in charge of these questions until February 2016, when the two of them left the project. The statistician position then remained vacant until 15 May 2017.

Key: PMS=Panel management system; final entry into the panel begins with registration in this tool. On the x-axis, the successive surveys; on the y-axis, the standardised chi-square distance relative to the census, for each variable, calculated for the participants in each survey.

A more fine-grained analysis of educational levels indicates that attrition and non-responses primarily lead to the loss of the least qualified participants, as can be seen from the graphs in Figure 2. These graphs represent the percentage difference in the educational levels of the panellists relative to the census in the course of the surveys.

Figure 2: Differences in educational levels (no diploma, vocational diploma, high school or bachelor diploma, and college graduate) between the population census and the responding panellists over the course of the surveys



Participation profiles

Following a survey, there are five possible states in response to the invitation to participate addressed to the panellists:

- the questionnaire was finished;

- the questionnaire was started but not finished (either because it was abandoned part way, or the questions were completed but the questionnaire was not confirmed after the last question);

- the panellist was not invited (either because s/he informed the ELIPSS team that s/he would be absent, or because the tablet was broken or lost);

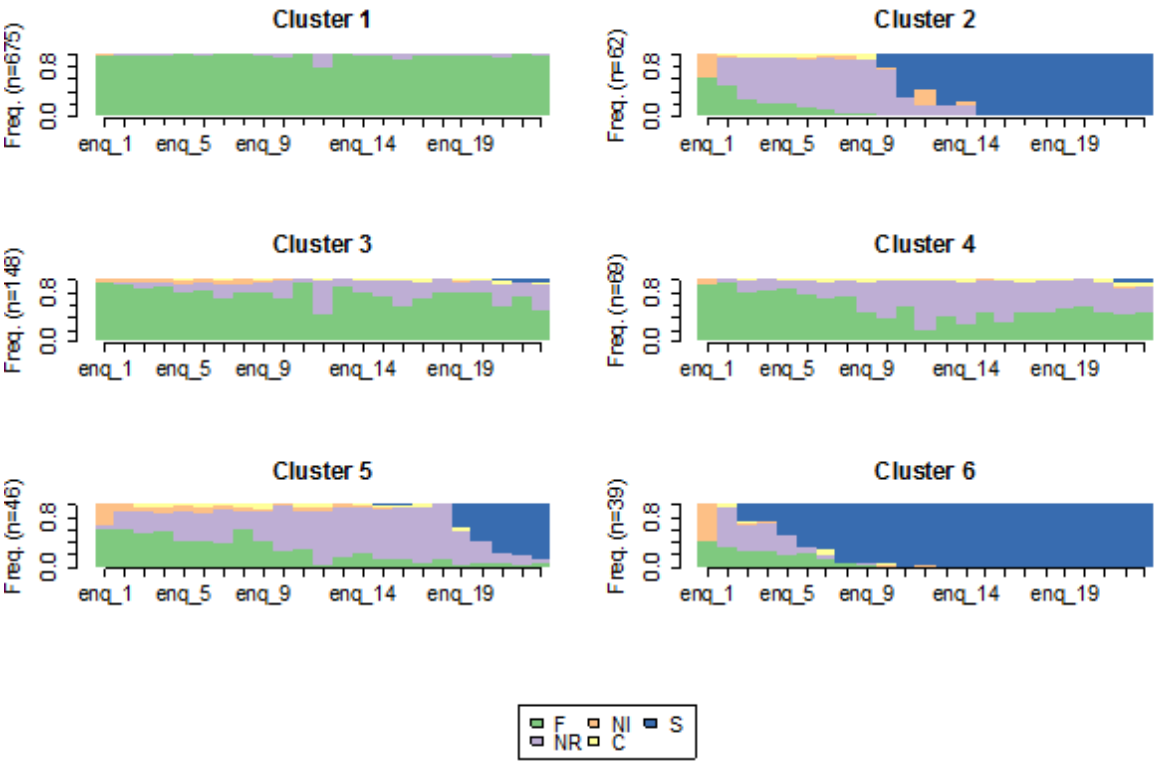
- the panellist did not reply;

- the panellist had definitively left the panel at the date of the survey (attrition).

By distinguishing between the 5 possible states of response to a survey (questionnaire finished F, questionnaire started but not validated C, not invited NI, not responding NR and left the panel S), we can study the participation trajectories of the panellists and try to group them into homogeneous categories, in order to obtain an effective description of their behaviours throughout the surveys. The optimal matching method was used for this purpose.[3]

If we look at the 23 surveys analysed, we can identify 6 clusters of participation trajectories (Figure 3).

Figure 3: Homogeneous clusters of participation trajectories in 23 ELIPSS surveys in 6 categories



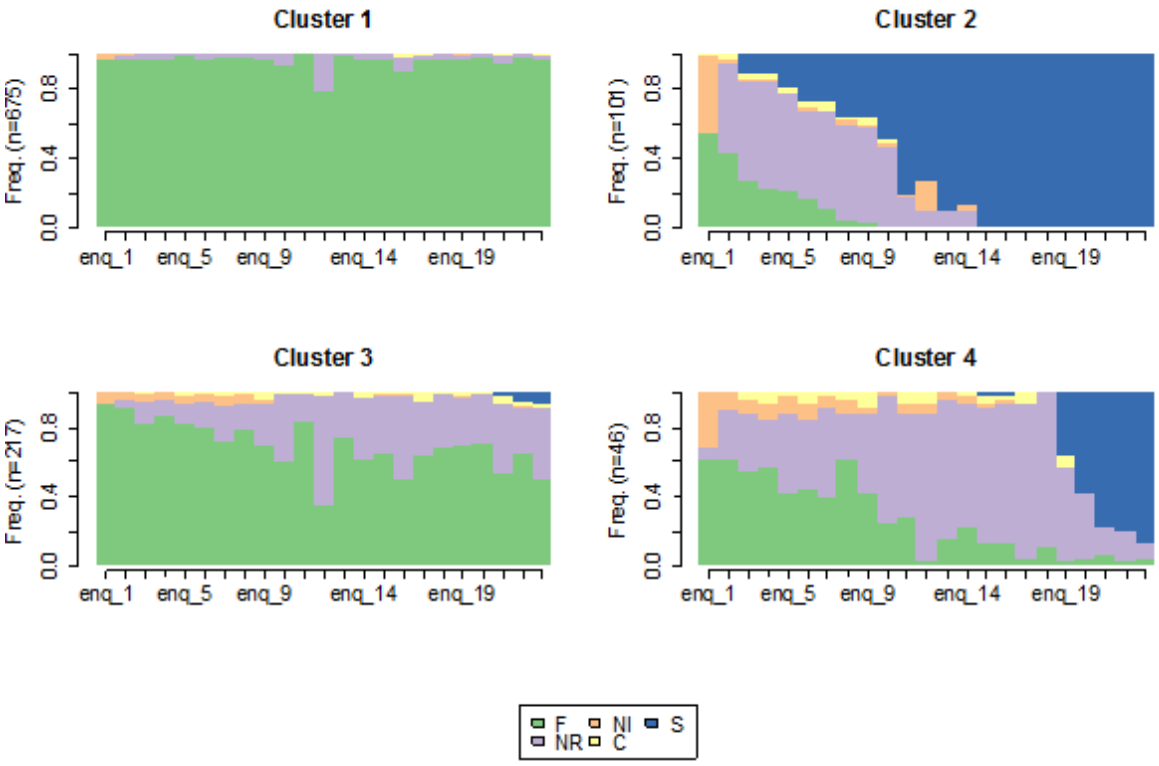
Cluster 1 consists of 65% of the initial 1039 panellists. These are the reliables: they responded to all or almost all the surveys. Clusters 3 and 4 respectively cover 14% and 7% of the original panellists: these are the panellists whose participation is not systematic but still high (cluster 3) or medium (cluster 4), but who are clearly on a downwards trend.

[3] This was done using the TraMineR software package.

Dropouts from the panel start to appear in these two clusters around the 20th survey. Cluster 2 (6% of the panellists), cluster 5 (4%) and cluster 6 (4%) identify three groups of panellists who dropped out of the panel, differentiated by departure date. Cluster 6 is the early dropouts, cluster 2 the intermediate dropouts (from around the 10th survey) and cluster 5 the late dropouts (from around survey 18).

This result can be summed up by grouping the closest clusters. This gives us 4 clusters (Figure 4): the cluster of reliables, which is intact, clusters 2 and 4 which cover dropouts according to departure stage, and cluster 3 which covers the panellists whose participation is medium and declining. The following breakdowns will be conducted on these 4 categories.

Figure 4: Homogeneous clusters of participation trajectories in 23 ELIPSS surveys in 4 categories



An analysis of the sociodemographic characteristics of the reliable panellists (cluster 1) shows that the panellists who were still present at the last survey (n=675) are more often people of French nationality than the rest of the panellists (92% vs 87%, p=0.007), are more often educated to beyond baccalaureate level (64% vs 57%, p=0.032), are more often married (46% vs 36%, p=0.001) and are also more often aged between 35 and 64 (69% vs 54%, p=0.001). In addition, these reliable panellists are less likely to live in social housing (13% vs 21%, p=0.001) or in an "urban renewal zone" (4% vs 8%, p=0.014), and more likely to live in homes with at least 4 rooms (68% vs 62%, p=0.028) or 5 rooms (46% vs 35%, p=0.001).

The socio-economic profile of the reliable panellists is therefore relatively privileged compared with the others.

Influence of individual characteristics and survey attributes on participation

The analyses above show both the influence of individual social characteristics (age, education, etc.) and of the characteristics of the surveys as described to the participants (questionnaire duration, subject matter, length of collection period) on the participation rate. Is it possible to separate the influence of these individual and survey-related factors?

To do this, we need to conduct a process of multilevel modelling. Indeed, since ELIPSS is a cohort, the panellists are asked to take part in all the surveys and a single individual's successive participations are correlated with each other and therefore not independent. As well as providing predictive clues to the panellists' participation throughout the panel, this kind of analysis could make it possible to establish more effective panel management, for example by helping us to define reminder protocols for the panellists most likely to drop out.

Given the differences in behaviour in participation trajectories observed between homogeneous clusters represented in Figure 4, it would seem helpful to distinguish the clusters before undertaking the analysis. Indeed, for the reliable panellists, the effect of time is almost zero, since their participation is virtually systematic. For the panellists whose participation is less systematic, it diminishes linearly with time. So measuring the effects of time on participation as calculated for all the panellists would be very deceptive, and insufficient to describe very contrasting participation behaviours. So we will identify 3 clusters: the cluster of reliables (n=675), cluster 3 of medium and declining participants (n=217) and clusters 2 and 4 of dropout panellists, differentiated by the date of their exit from the panel (n=101 and n=46)

We can reasonably formulate the following hypotheses regarding the results:

1. The effects of individual sociodemographic variables and of survey characteristics are smaller for the reliables cluster.
2. The effects of survey characteristics are more marked for the other clusters.

Hypothesis 1 is based on the fact that the average trajectory of the reliable panellists is almost exclusively spread, with no significant variation perceptible (see cluster 1 in Figure 4). In consequence, all the panellists systematically participate and neither their individual characteristics nor the characteristics of the surveys have an impact. Hypothesis 2 is based on the same approach applied to the other clusters: since their participation is not systematic, it is very likely that the individual characteristics and the survey characteristics have a greater impact on participation.

We will take age and education as the individual sociodemographic characteristics;[4] and membership of the reliables cluster at the fifth survey, questionnaire duration, duration of collection and the sensitive nature of the surveys, as the survey-related attributes.

In these models, the statistical unit is the survey, and the variable modelled is participation in the survey. Now a single individual's participations in successive surveys are not independent of each other, since they are linked with the individual's profile and motivations. In other words, there is a correlation between the individual's successive participations. The multilevel model takes into account this non-independence of the

[4] Other variables were tested: sex, nationality, recruitment method, residence in social housing. Their effect is very rarely significant, and their presence reduces the quality of the model and provides very little information.

observations over time. The so-called fixed effect variables are the characteristics of the surveys (duration, period, sensitive topic) and the individual characteristics of the panellists (age, education). Each individual is considered to have an average level of participation specific to them over the 23 surveys, but participation varies with time. Time is incorporated through the introduction of a survey counter variable (from 1 to 23) which has a linear effect on participation (the participation rate declines continuously with the number of surveys) but varies from one individual to another. Average participation and time are random variables.[5]

The results for the three clusters of individuals (reliables, declining participation, gradual departure) are presented in Table 2 below.

Table 2: Multilevel models of participation in the 23 surveys according to membership of a trajectory cluster (OR or significance)

	Reliables n=675		**Declining participation n=217**		**Gradual dropout n=147**	
Effect	OR	Pr > |t|	**OR**	**Pr > |t|**	**OR**	**Pr > |t|**
Time (+1 survey)	1.00	0.8076	**0.90**	<.0001	**0.52**	<.0001
Age (18-24)	0.64	0.1315	**0.46**	0.0036	2.41	0.1665
Age (25-34)	0.75	0.244	**0.54**	0.0119	**4.90**	0.0086
Age (35-44)	1.01	0.9503	**0.58**	0.0234	**3.32**	0.0442
Age (45-54)	1.01	0.9524	0.69	0.1289	1.83	0.3208
Age (55-64)	1.28	0.3007	0.63	0.1056	1.14	0.8413
Age 65-75 (ref.)	1.00	.	1.00	.	1.00	.
Educ (no education)	1.16	0.4557	0.85	0.3255	0.54	0.1995
Educ (vocational)	1.36	0.0822	1.15	0.3076	**0.37**	0.0552
Educ (high school)	1.27	0.1064	**1.27**	0.0579	0.63	0.3422
Educ (higher ed.) (ref.)	1.00	.	1.00	.	1.00	.

[5] The modelling was carried out using the SAS V9.4. GLIMMIX procedure.

	Reliables n=675		**Declining participation n=217**		**Gradual dropout n=147**	
Sensitive topic	1.06	0.5858	0.91	0.2607	1.08	0.696
Not reliable at time 5	**0.09**	<.0001	**0.63**	0.0144	**0.01**	<.0001
Questionnaire duration						
3rd third =30 min	**2.16**	<.0001	0.89	0.3183	1.60	0.0757
2nd third =20-25 min	**1.89**	<.0001	**1.41**	<.0001	1.05	0.8456
1st third =5-15 min (ref.)	1.00	.	1.00	.	1.00	.
Collection period						
3nd third =32-100 days	**3.93**	<.0001	**2.53**	<.0001	0.94	0.7898
2nd third =24-31 days	**4.23**	<.0001	**2.19**	<.0001	1.06	0.8167
1st third =12-23 days (ref.)	1.00		1.00	.	1.00	.

Note : OR=odds ratio. For the three models, the difference tests on average participation, on changes in participation over time are significant (p<0.0001); the clusters can be identified in Figure 7: cluster 1, cluster 3 and clusters 2&4.

The results show overall that hypothesis 1 is verified: age is significant for the clusters of non-reliables, but not for the reliables. For the panellists with declining participation, relative to people aged 65 to 75, the younger people on average participate less (OR<1). For the dropout panellists, by contrast, panellists aged 25-44 participate a little more than those aged 65-75 (OR>0). On the other hand, the result for education is less convincing, although the trend is similar: the effect of education is closer to the significance threshold in the groups which are not reliable (the tendency being towards lower participation amongst those with less education), whereas there is no significant education effect in the reliables cluster.

And finally, hypothesis 2 is refuted: questionnaire duration and the length of the collection period have a more frequent and stronger impact in the reliables cluster. Indeed, the effect is even zero in the dropout cluster, which is probably explained by their general indifference to the panel and to its surveys.

The length of the collection period has a bigger impact than the duration of the questionnaire, which is not surprising insofar as the lengthening of the collection period

is more important and offers more opportunities to participate. In detail, we find that a high questionnaire duration increases the chances of participation among the reliables (OR>2 for a duration of 30 minutes vs. 5-15 minutes); among the panellists with declining participation, the effect is noticeable for a response time between 20 and 25 minutes compared with a duration of less than 15 minutes (OR=1.4). Among the dropouts, an extreme duration tends to increase the relative likelihood of participating in a short duration (OR=1.6), but the effect is not significant.

Finally, the sensitivity of the survey topic clearly has no impact: the OR never significantly vary from 1.

In parallel, the effects of time (the succession of surveys) has a pronounced effect on the cluster of medium participants, whose participation declines (OR=0.90), and a more clearly marked effect on the gradual dropout cluster (OR=0.52), whereas it is zero in the reliables cluster (OR=1.00), which was expected given the participation trajectories of these clusters of panellists.

Preparing the enlargement of the panel

On the basis of the operational team's assessment of the recruitment of the pilot panel (report produced in June 2014) and the recommendations of the Scientific Advisory Board (issued in May 2014), the DIME-SHS Steering Committee decided in May 2015 to move to the development phase of the DIME-Quanti tool, while reducing the size of the panel from a minimum of 5000 to 3000 individuals for financial reasons.

Despite the initial schedule, recruitment of the new panellists could not be undertaken in 2014. The main reason for this delay was the renegotiation of the contract with the telephone operator which was providing the tablets and the subscriptions. Conducted by the operational team, this renegotiation to set new tariff conditions for the subscriptions and to find a new tablet model, lasted more than a year. The amendment was only signed on 15 September 2015. Unfortunately, the Archos tablets delivered at the end of November 2015 did not match the model ordered in September and tested by the IT team for two months. With recruitment of the new panellists due to start in a few weeks, these non-compliant tablets had to be kept. The timetable was maintained through an immense effort (in particular by the IT team) to check the compatibility of the new tablets in Android 5 and at the price of a delay on the other aspects of the recruitment process (final field preparation with the survey institute, development and scheduling of the telephone training script for panellists). The commercial negotiation to obtain financial compensation, undertaken with the help of Sciences Po's legal department, lasted 4 months and yielded no results.

In parallel, we obtained INSEE's agreement for the national statistics Office to supply us with a sample of 10,000 dwellings drawn randomly from the population census. This agreement was signed in October 2015 following a favourable ruling by INSEE's Board of Directors and the agreement of the Privacy Committee, and after the project had been presented to CNIS's Demographics and Social Affairs Committee:

- It introduced a clause on ethics and the respect for individual privacy, reflecting the recommendations of the Data Protection Commission and the Constitutional Council on Diversity Measurement. These essentially concerned the non-use of ethno-racial criteria, and the obligation, where necessary, to employ open questions and to ask for the express consent of the panellists.

- It emphasised the requirement for the presence of an INSEE representative on the DIME quanti Scientific and Technical Committee.

- It set an information procedure: the programme of completed surveys is to be sent to the CNIS every year and a report submitted to INSEE two years after the delivery of the sample.

The whole of 2015 was therefore marked by preparations for the enlargement of the Elipss online panel. A contract was established in April 2015 for the selection of a survey institute to carry out the face-to-face recruitment of 2700 new panellists starting in January 2016.

This new phase of the project involved numerous IT developments to manage the fleet of tablet computers, to incorporate the new panellists into our information system and to adapt the questionnaire response app to the new tablet.

Recruitment of the new panellists

The target set for the Ipsos survey institute was to recruit 2700 individuals from a sample of 10,000 dwellings drawn by INSEE from the 2014 census. These 10,000 dwellings were broken down into 1 main sample and 7 reserve samples of different sizes and had been drawn from 300 zones (corresponding to 475 municipalities) across the whole of metropolitan France excluding Corsica. At each dwelling, a face-to-face interviewer had to randomly draw an inhabitant aged between 18 and 79,[6] and propose participation in the ELIPSS panel.

149 interviewers received one day of training on how to conduct the interview, in which their task was to persuade the individual to take part in the ELIPSS panel, to demonstrate the use of the tablet and to have the contract for participation in ELIPSS signed.

Before the interviewers' visit, a notice letter was sent to each household at the dwelling in the sample.

The rules for attempting to contact households were as follows:

- at least 4 home visits at different times of the week and of the day before classifying the dwelling as a non-contact, including at least one attempt on a weekday evening and one attempt on a Saturday,

- at least 15 days of attempts before classifying the dwelling as a non-contact.

The collection of data on the composition of the household, and the selection and recruitment of the individual, had to be done through a face-to-face interview. Under no circumstances could these activities be carried out by telephone.

Once an individual had signed the agreement, he or she was given the tablet in a second phase after the recruitment had been approved by the ELIPSS team.

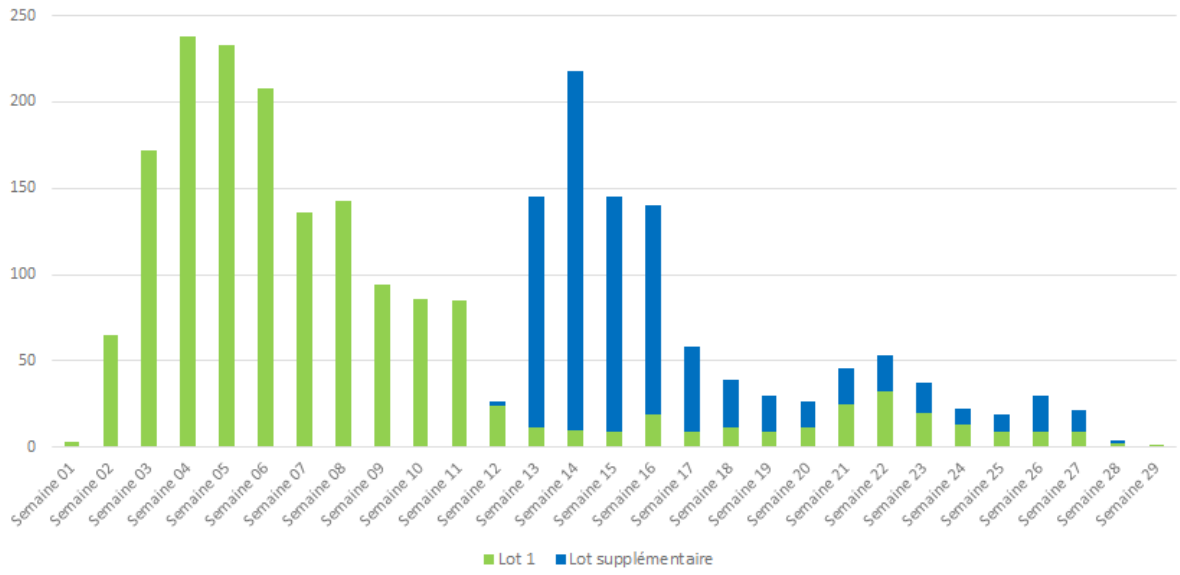
In the event of refusal by the household or by an individual, interviewers had to ask a number of questions in order to gather information on the main sociodemographic factors, interest in politics and Internet equipment, and themselves fill in further information.

[6] The upper age limit for the eligibility of the panellists was raised to 79 to take into account the ageing of the pilot panel (it had previously been 75).

On average, each interviewer had 60 addresses to process.

The face-to-face fieldwork lasted from 8 January to 18 July. After a good start, the processing of the main sample of 6000 addresses underwent a rapid decline. The ELIPSS team decided to open only 3000 reserve addresses in order not to exceed the 2700 panellists who would subsequently have to be supplied with a tablet and a 4G subscription. Recruitment resumed with this additional batch but was not as high as forecast.

Figure 5: Number of panellists recruited per week of fieldwork (main sample, in green, and reserve sample, in blue).



The main difficulty lay in the high rate of un-contactables, defined as "short-term absence" (at 19% well above the 10-15% obtained in equivalent studies). The acceptance rate at household level was 43.5%. 2523 individuals aged 18 to 79 signed the ELIPSS participation agreement, giving a recruitment rate of 32% once ineligible households and individuals had been excluded.

Table 3: Result of last contact at household level

	% of the 9000 addresses processed
Non-surveyable household	**11.9%**
Wrong address	2.7%
Address corresponding to a company, a government office	0.5%

Dwelling vacant, unoccupied, address no longer exists, under construction, second home	3.6%
Outside scope (age, language, state of health, moving in 3 months, away more than 6 months)	4.1%
Other ineligible	1%
Household eligible but not surveyed	**24.5%**
No answer	19.0%
Home inaccessible	4.8%
Other reason (appointment, health problem)	0.7%
Household processed	**63.6%**
Refusal to participate (household)	20.1%
Agreement to complete household composition table and selection of an individual	43.5%
including: Non-eligible individual	*1.9%*
Individual eligible but not surveyed	*1.4%*
Refusal to participate (individual)	*12.2%*
Recruitment completed	*28%*
TOTAL	**100%**

The male/female distribution corresponds to the 2012 Population Census. On the other hand, inhabitants of Île-de-France, Nord-Pas-de-Calais and Provence-Alpes-Côte d'Azur are underrepresented, whereas those from the Western regions are overrepresented among the panellists. Younger people aged 18 to 34 are underrepresented in the sample (21.5% instead of 28.3%), while people aged 45 to 54 are overrepresented. However, the representation of older people is largely in line with the census. There are also more working people among the new panellists (69.2%) than in the general population aged 18-79 (64.7%). Finally, single person households are underrepresented, being more difficult to contact. The main underrepresentations are the same here as in the other face-to-face surveys, suggesting that there is nothing in the system that balances out these distortions.

At first sight, the new panellists seem particularly connected: 84% use the Internet every day or almost every day (this rate is 62% in the 2015 TIC survey with people aged 15 and above living in mainland France and France's overseas dominions and territories).

Table 4: Sociodemographic profile of the new panellists

	New panellists age 18-79	General population
Sex		(1)
Male	47.7	48.4
Female	52.3	51.6
Age		(1)
18-24	7.5	11.3
25-34	14.0	17.0
35-44	20.3	18.7
45-54	24.0	18.9
55-64	17.8	17.7
65-74	13.1	11.6
75-79	3.2	4.8
Employment status		(1)
Active (job, apprenticeship, unemployed)	69.2	67.2
Retired	20.6	22.9
Student, other inactive	10.2	12.4

	New panellists age 18-79	General population
Education		*(2)*
No diploma	25	*32.2*
Vocational diploma	14	*24.0*
High school diploma	20	*16.7*
Higher education	41	*27.1*
Number of people in household		(1)
1	25.1	31.2
2	35.6	34.0
3	14.9	15.4
4	16.1	12.9
5 or more	8.3	6.6
Internet use		*(3)*
In the last 3 months	94	*78.0*
Every day or almost every day	84	*62.1*
Region		(1)
Nord-Pas-de-Calais	3.9	6.4
Picardie	3.4	3.0
Haute-Normandie	3.1	2.9
Basse-Normandie	2.7	2.3
Ile-de-France	13.0	18.9

	New panellists age 18-79	General population
Champagne-Ardenne	1.9	2.1
Lorraine	4.4	3.8
Alsace	2.9	3.0
Bretagne	7.0	5.1
Pays de la Loire	6.8	5.7
Centre	4.5	4.0
Bourgogne	2.7	2.6
Franche-Comté	1.9	1.9
Poitou-Charentes	4.4	2.8
Limousin	1.4	1.2
Auvergne	2.8	2.2
Rhône-Alpes	10.0	10.0
Aquitaine	6.5	5.3
Midi-Pyrénées	5.0	4.7
Languedoc-Roussillon	5.0	4.3
Provence-Alpes-Côte d'Azur	6.7	7.9

Sources:

(1) INSEE, Population Census, 2012 , scope: Mainland France, population aged 18 to 79

(2) INSEE, Population Census, 2013 , scope: Mainland France, non-school population aged 15 and above

(3) INSEE, Household Survey on Information and Communication Technologies, 2015, mainland France and overseas dominions, population aged 15 and above

In conclusion, to explain a lower than hoped for recruitment rate, we may note the difficulties in contacting households and managing the survey network (in particular in motivating the interviewers to finish processing their addresses). These problems are not specific to the Ipsos institute. There is little competition in the market for organisations capable of conducting face-to-face surveys in mainland France. Only two or three organisations can conduct this type of survey. It is also possible that the increase in the number of households already equipped with a tablet has reduced the attractiveness of the offer since the pilot study in 2012-2013.

The surveys

The DIME Quanti Scientific and Technical Committee (CST) on which the operational team relies for the selection of the surveys, has been renewed twice since May 2014, the terms being for 2 years, renewable.

In September 2014, six new members joined the CST, while two members terminated their membership. At the end of the CST's second term in autumn 2016, four new members joined the 11 members who wished to renew their terms (two left the committee).

The CST is currently made up of the following experts:

- Rémy Caveng, lecturer in sociology at the University of Picardie Jules Verne (University Centre for Research on Public Action and Politics, European Centre for Sociology and Political Science and Laboratory of Quantitative Sociology)
- Joanie Cayouette-Remblière, researcher in sociology at INED
- Jérôme Cubillé, researcher at EDF R&D
- François Denord, researcher in sociology at CNRS (European Centre for Sociology and Political Science)
- Céline Goffette, lecturer and researcher at the Laboratory of Quantitative Sociology at CREST (Centre for Economic and Statistical Research)
- Anne Jadot (chair of the CST), lecturer in political science at the University of Lorraine
- Dominique Joye, professor of sociology at the University of Lausanne
- Cécile Lefèvre, professor of sociology at University Paris Descartes (Centre for Research on Social Bonds)
- Muriel Letrait, research officer at University Paris Descartes (Centre for Research on Social Bonds)
- Pierre Mercklé, lecturer in sociology at ENS Lyon (Max Weber Centre)
- Gaël de Peretti, INSEE administrator, head of the Information Collection and Processing Division at INSEE's Department of Statistical Methods
- Élise Tenret, lecturer in sociology at University Paris Dauphine
- Karine van der Straeten, director of research in economics and political science at CNRS (Toulouse School of Economics)
- Loup Wolff, INSEE administrator, head of the Department of Surveys, Forecasting and Statistics at the Ministry of Culture and Communication

- Sonja Zmerli, professor of political science at Sciences Po Grenoble (PACTE)

Since the first report to the DIME-SHS Scientific Advisory Board, three calls for projects have been organized, in 2015 and 2016, in which 40 survey projects were submitted. The DIME Quanti CST selected 19 survey projects and 4 are currently being assessed.

The last two calls for projects concerned the surveys with the enlarged panel. In the first call for projects for the enlarged panel, 19 projects were submitted, the largest number achieved since ELIPSS began, whereas 9 projects were submitted under the 2016 procedure. It is likely that the enlargement of the panel automatically increased the number of submissions in 2015, whereas the 2016 procedure was less anticipated by the academic community.

Table 5: Survey selection

	Calls for the pilot study				**Calls for the enlarged panel**	
	2011	**2012**	**2013**	**2015-1**	**2015-2**	**2016**
Number of proposals	**5**	**8**	**7**	**12**	**19**	**9**
Number of projects accepted	3	4	2	6	10	3
Number of projects accepted after revision and resubmission	1	1	2	0	0	0
Number of projects rejected	1	3	3	6	9	2
Number of projects undergoing assessment	0	0	0	0	0	4

The subjects tackled since June 2014 have been similar to those of the surveys already conducted between 2012 and 2014 – e.g. cultural practices, electoral behaviours and political opinions, health, the environment – but have also included new topics such as family relations, residential choices, mobilities, attitudes to the state, the connections between housing and diet, wastage, leisure, etc.

Among these surveys, several drew on new methods for which new IT developments were made. A study on geographical mobilities led to the development of a geographical map on which the panellists were asked to select the territorial areas (country, *département*,

region) where they had travelled. This function is going to be adapted for a forthcoming survey on panellists' perceptions of territorial levels.

The taking of photos as part of the questionnaire has also been developed to explore the connection between dietary habits and home layout, in particular the kitchen. This function is being used again for a survey on the watching of audiovisual series, with panellists being asked to take a photograph of the place from which they watch the series.

Several measures have been taken to ensure the protection of personal data in ELIPSS within the framework of France's "Informatique et libertés" data protection act:

- during calls for projects, each survey project submitted is examined by the CNRS data protection expert;
- the non-response options available to respondents have been revised in this respect, and since July 2015 the express consent of respondents is obtained at the beginning of each questionnaire;
- the members of the ELIPSS system followed a seminar dedicated to the "informatique et libertés" data protection act;
- the procedures for systematic matching with the annual survey have been reviewed and additional matching requests are examined on a case-by-case basis (see below);
- with the enlargement of the panel, the declaration on the processing of personal data has been updated with the CNRS Data Protection Department.

Data dissemination

In general, the survey data, enriched with the variable and modality headings, with the sociodemographic variables from the annual survey, with the geographical information drawn from the INSEE census and with the weighting variables (both transversal and, where applicable, longitudinal) are sent to the research team within two months of the end of the fieldwork. However, since September 2016 (date of the first survey on the enlarged panel), in the absence of a statistician for more than a year, first versions of the data files have had to be sent to the research teams without weighting.

In agreement with the DIME quanti Scientific and Technical Committee, it was decided that certain modules of the annual survey (civil status, work and education, description of the household, dwelling and neighbourhood, income and assets, social life, leisure) would be systematically matched to the surveys. For the modules relating to subjects that are more sensitive in terms of data protection law (politics, religion, health), only certain variables (having a religion, importance of religion in life, religious education and childhood, interest in politics, self-positioning on the left/right scale, perceived state of health, practice of sport) are automatically sent with the survey data. In order to have access to the other variables in the annual survey, teams must request and justify matching when they submit the survey project under the call for projects procedure. These matching requests are examined by the DIME quanti Scientific and Technical Committee and by the legal expert at the CNRS Data Protection Department.

For the pilot study, the adjustment procedure defined by Stéphane Legleye and Nirintsoa Razakamanana at the National Institute of Demographic Studies (INED) comprises two

steps: first adjustment for total non-response (ATNR)[7] then calibration weighting.[8] Three weighting calculations were proposed: adjustment without ATNR, with maximum ATNR, and with ATNR at household level only. The latter weighting was the one chosen. Despite increasing the variance, it improves the representativeness of most of the adjustment variables, with adjustment being used to correct the remaining differences. If we compare the difference in distribution of the sample consisting of the 1039 panellists in the pilot study, with the population census, depending whether the initial weighting (survey weight) is used, or the final weighting adjusted for total non-response, before adjustment, we observe an improvement for age, sex, French nationality, education, but a slight deterioration for geographical distribution. Nonetheless, the sum of these differences is greatly improved. It turns out that in terms of distribution of the weights, this procedure of total non-response adjustment before calibration weighting is superior to direct adjustment.

Longitudinal weights are also calculated. For the repeated Dynamob (Mobilisation Dynamics) surveys, two weights are systematically provided: a weighting for respondents to all the waves and a weighting for which the benchmark is the respondents to the first survey conducted in September 2013.

Since 2015, 9 surveys conducted in 2013 and 2014 have been made available to the scientific community via the Quetelet Network portal:

- Annual ELIPSS survey - wave 1 (2013)
- Mobilisation Dynamics - wave 1 (2013)
- Cultural practices, media and information technologies - wave 1 (2013)
- Fertility, contraception, sexual dysfunctions (2013)
- Survey on values, the environment and energy - wave 1 (2013)
- Health, work and environment Survey on exposure to inorganic dusts - wave 1 (2013)
- Mobilisation Dynamics - wave 2 (2013)
- Relations between generations through the prism of standards of solidarity and social justice (2014)
- Annual ELIPSS survey - wave 2 (2014)

These surveys gave rise to some fifty requests from different users. It should be noted that the ELIPSS (2013) annual survey and the survey on Cultural Practices, Media and Information Technologies (2013) were both among the five surveys most resquested at the Sociopolitical Data Centre (CDSP) in 2016.

Specially tailored files created from several ELIPSS surveys corresponding to different research projects were requested for secondary analyses: one in order to study the

[7] Two techniques were used: decision tree modelling and homogeneous response groups (HRG). 39 variables were used in the modelling (geographical level, dwelling level, household level).

[8] The variables used are sex, age, nationality, education and geographical zone (ZEAT).

inconsistencies in responses between surveys, the other to look at the digital practices of older people and their impact on family relations.

Finally, a major effort to harmonise data documentation was conducted at the beginning of 2017, linked with discussions currently underway at CDSP on the new version of the new DDI-L international documentation standards.

Promoting the instrument

The ELIPSS system has been presented many times since 2013, at major French institutions (INSEE, INED, CEREQ), in seminars and (French Society of Statistics, French Association of Political Science, the universities of Lausanne and Mannheim), at conferences in France and abroad (International Conference of Blaise Users, European Survey Research Association - ESRA, European Sociological Association - ESA, French Sociology Association - AFS, Webdatanet, INSEE's Statistical Methodology Day, World Association for Public Opinion Research - WAPOR, European Congress of Methodology). In addition, researchers who have conducted a survey through the ELIPSS panel or have reused data from an ELIPSS survey have presented their results in some 30 papers at seminars and conferences, in particular at the two workshops focusing on the DIME quanti instrument that marked the 10th anniversary of CDSP celebrated in Paris in December 2016.

Mélanie Revilla from Pompeu Fabra University in Barcelona spent several weeks at CDSP in 2014 to study the population of panellists who had no access to the Internet before joining the ELIPSS panel. Her aim, in particular, was to analyse their sociodemographic profile and response behaviour. The central question was the impact of their inclusion in a probability-based Internet panel, especially in terms of representativeness. The following article was the outcome of this collaboration: Mélanie Revilla, Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Pablo de Pedreza, "What is the gain in a probability based online panel of providing internet access to sampling units who previously had no access", *Social Science Computer Review*, vol.34, n°4, p.479-496.

Collaborations with similar systems in Europe have been pursued. Two surveys conducted in a collaboration between the German Internet Panel (GIP), the GESIS Panel, the LISS Panel and ELIPSS were completed in 2014 and 2015, and a third is scheduled for summer 2017. In 2014, questions were extracted from the comparative surveys of the European Social Survey (ESS), Survey of Health, Ageing and Retirement in Europe (SHARE), Programme for the International Assessment of Adult Competencies (PIAAC) and the European Election Study 2014 (EES). The 2015 survey is based on a questionnaire provided by the "Reforms Monitor" public interest group. Finally, the 2017 survey is extended to other panels: the Norwegian Citizen Panel (NCP) and Iceland's SSRI Online Panel. This survey is part of the call for contributions to the "Governance for the future" project in the 2018-2020 Work Programme and consists of *questions that are thematically relevant to orientations under this call*.

In parallel with these surveys, collaboration has included the deposit of two projects (one in 2016 which was not selected and one in 2017 currently being assessed) within the framework of calls for "European Research Infrastructures" projects under the European Commission's 2018-2020 Work Programme.

Finally, a jointly authored article comparing the four panels – ELIPSS, the GESIS Panel, GIP and the LISS Panel – was written and published in 2016 (Annelies Blom, Michael Bosnjak, Anne Cornilleau, Anne-Sophie Cousteaux, Marcel Das, Salima Douhou, Ulrich Krieger, "A

Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe", *Social Science Computer Review*, vol.34, n°1, p.8-25).

Prospects for 2017-2019

As things stand, the Internet subscriptions provided for panellists will end in April 2019, as will the contracts of the two panel managers, which will also mark the end of the surveys administered in ELIPSS.

Since the call for projects issued in 2016 led to the submission of 9 projects, once the DIME Quanti Scientific and Technical Committee has completed its selection, there will still be several months left to administer a few surveys. The DIME Quanti Scientific and Technical Committee has therefore decided that a final call for projects will be issued from July 2017 to November 2017 for surveys to be administered at end 2018/beginning 2019.

The primary focus of the next two years will be on making surveys available and on communication around the ELIPSS data. Data documentation will be the priority, together with the translation of that documentation. The plan is also to produce survey reports for each field. The model for these reports will be developed at the end of 2017.

The long-term archiving of surveys, currently in an exploratory phase, will also constitute one of the lines of development for the team. In addition, the team will focus on scientific development activities, firstly by feeding back the results to the panellists and secondly through the production of exploratory results for publication on the DIME Quanti website.[9]

These prospects are only viable if the ELIPSS team remains at its current level (see table below). However, if vacancies recur (as was the case for the role of statistician for more than a year), the team's priorities will have to be reviewed for production of the ELIPSS surveys to continue. In particular, the team responsible for IT developments, shared across all CDSP's activities, is chronically understaffed and specific developments may be reviewed if necessary.

[9] http://quanti.dime-shs.sciences-po.fr/en/

Table 6: The operational team

		2016				2017				2018				2019			
		T 1	T 2	T 3	T 4	T 1	T 2	T 3	T 4	T 1	T 2	T 3	T 4	T 1	T 2	T 3	T 4
ELIPSS Team	Coordination	Anne Cornilleau															
		Anne-Sophie Cousteaux															
	Panel management					Charlotte Montcharmont											
		Elodie Pétorin															
	Survey production and documentation	Alina Danciu															
		Emmanuelle Duwez															
		Alexandre Mairot															
		Mathieu Olivier															
						Jean-Baptiste Portelli											
	Statistics					Thomas Pilorin											
Combined CDSP team	IT development and infrastructure	Quentin Agren															
		Alexandre Chevallier															
		Geneviève Michaud															
		Romain Mougin															
		Jérémy Richard (half-time at CDSP)															
						Baptiste Rouxel (apprentice, half-time at CDSP)											

Vacant position or no recruitment budget

Recommendations of the Scientific Advisory Board at its May 2014 meeting

To finish, here are the responses we can make to the recommendations issued for the DIME quanti instrument by the Scientific Advisory Board in May 2014:

- *decide in favour of a longitudinal or transversal direction for the project and therefore for the surveys proposed*

This question, posed to DIME quanti's Scientific and Technical Committee, following the recommendation of the DIME-SHS Scientific Advisory Board, has not been decided. The calls for projects have remained open to transversal and longitudinal surveys (up to a ceiling of 60 minutes of surveys per year) and selected projects of both types.

- *conduct an in-depth analysis of non-response, in particular during recruitment, for example through a survey of non-responders*

During the recruitment of the new panellists in 2016, a questionnaire was developed and offered to people who refused to participate in the ELIPSS process. A detailed analysis of the recruitment results is planned with the statistician who has just joined the project.

- *propose English language documentation for the data files you disseminate, starting with the annual survey*

The documentation of the 9 surveys currently disseminated is being translated. This only entails translation of the study description and the titles of the variables. There are no plans to translate the questions and response options, because this would involve making sure that the questions asked in French and English are comparable, which is more than just a translation process.

- *study the possibility of disseminating the detailed and collected files through CASD (secure data access centre)*

Knowing the complex procedures for accessing confidential data through CASD, we prefer the procedures that we have established, in concert with CNRS's legal expert, for the dissemination of standard ELIPSS survey files and other files. In fact, these procedures correspond to our users' demand and we have so far received no requests for access to highly detailed files.

- *consider matching with administrative data*

This recommendation is not applicable to the French case, by contrast with the Netherlands, a fact that was pointed out to the Scientific Advisory Board at the meeting in May 2014.

DIME Web

A brief reminder of the functioning and purpose of Dime Web

The *Dime Web* instrument helps social scientists to harvest and process digital traces. At the confluence of *Digital Methods*, *Computational Social Sciences* and *Digital Humanities*, it sees digital technology as a new opportunity to study the social domain. With the help of the Sciences Po medialab, its two research engineers develop tools and provide advice and methodological assistance to social scientists. Their contribution comprises both digital devices and human support.

The *Dime Web* team has developed many instruments over the years, all free and open source. The most prominent is *Hyphe*, a curation-oriented web crawler specifically developed for quali-quantitative sociology. A number of smaller tools were also developed for various needs, such as the Twitter harvester *Gazouilloire*. It also supports preexisting tools like *Gephi* or other teams' products like *DMI TCAT*.

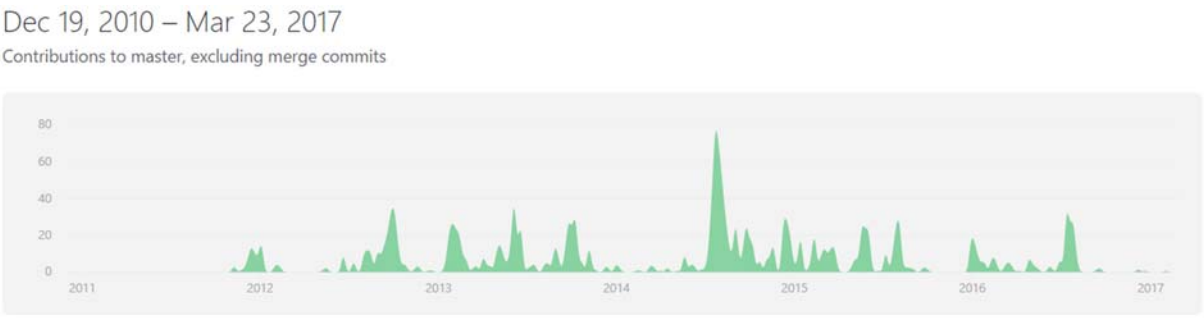
The team also welcomes and supports a wide range of queries, from the most uninformed to the most advanced. Based on the observation that digital literacy is unequally distributed among scholars, it sees an important part of its mission as raising awareness about the specificities and issues of working with digital traces. The help it provides can be to support the formalisation of a project, to steer scholars towards the best tools or data sources, to teach them how to use digital instruments and answer their methodological questions, and sometimes to partner in a research project. Most of the advanced projects started as small-scale experiments, which emphasises the importance of fostering discussion with scholars.

Dime Web's action can be seen as *upstream resource pooling*: users contribute to the development of large software projects like *Hyphe* in exchange for using them as a consortium would, except that these tools are preemptively developed by the team. The instrument has been asking its users to contribute financially since 2013. This money is allocated to time-consuming tasks like training, technical maintenance and support, but also funds software development. This permits its adoption of free and open source policy, which ensures a wider impact on the research community by reaching researchers beyond the official Equipex user pool.

Progress report (May 2014 – May 2017)

The last scientific advisory board (May 2014) took place during an important turn in Dime Web's life. Until 2014, projects were selected on the basis of a dossier whose content was coordinated with that of the *Elipss* instrument. This approach proved unsatisfactory for three reasons: the impression of a one-stop shop was not backed up by reality, since the CSTs (scientific and technical committees) of the two instruments generally gave contradictory results (acceptance by one, rejection by the other), the dossier was too unwieldy for micro-experiments, which discouraged applicants, and applicants were sometimes incapable of formulating their requirement without support from us. The approach was therefore revised in the period 2013-2014. In 2013, a first experiment in

paid support was conducted, and subsequently applied to the whole instrument in 2014. The conclusion was that big projects can only be implemented after small experiments with projects costing a few thousand euros, often drawing on budget residues and requiring quick decisions (in order not to lose the funding). For this reason, the task of supporting small projects was assigned to the operational team, and the CST adopted an annual oversight role. *Hyphe* was developed in the course of a dozen intensive sessions up to 2017, combined with extensive day-to-day maintenance, achieving operational maturity in 2015 and wider dissemination in 2016. The code is the result of more than 2000 commits. Set up in 2015, the website, which includes a fully functional demo of *Hyphe*, has attracted more than 8100 single users (according to Google Analytics).



Hyphe commits (code contributions) over the years

Since 2014, 13 projects have been supported, mainly in France but also elsewhere in Europe. Eight of the 13 chosen projects originated from consortium members, including 4 from different Sciences Po laboratories (CEE, CEVIPOF, CERI...). One article is currently being revised with CEVIPOF and a note has been published on the CEVIPOF website. In addition, members of the operational team have signed or co-signed 11 academic publications since 2014, including one article on the *Gephi* network analysis software on Plos One in 2015 (lead author), and one article presenting *Hyphe* to the ICWSM 2016 conference (lead author).

During its last session, the scientific advisory board gave 3 recommendations for Dime Web:

1. to position itself vis-a-vis Big Data initiatives
2. to follow the niche strategy that has been adopted
3. to offer training: it creates demand and attracts resources

A first move was to strengthen our bonds with Science Po's *Executive Education* team, in order to give more training and reach a more varied public. In the end, this move did not bring so many opportunities but the few that were consolidated helped us to position ourselves in relation to private sector actors. For these actors, "Big Data" is a very generic term that covers many different realities. We learned to use a different vocabulary if necessary and not be too dogmatic about our academic culture, since in the private sector too, digital literacy is very unequally distributed. Training the private sector is an opportunity, but our main problem lies elsewhere, as we will explain.

A second move was to promote the *Hyphe* crawler as an innovative tool for digital sociology. We wrote and published a paper, we developed numerous features, we put a website online and we held a number of workshops (including one at the Digital Humanities 2016 conference) and seminars in European research teams. This enhance

our identity and our visibility. Our main problem now is that *Hyphe* is not scalable enough to keep up with demand. We develop this issue in the next section.

It is worth noting that as the project matured we had more and more interactions with other research teams, and our academic purpose naturally settled as a *complement* to "Big Data" initiatives. The main difference lies in the qualitative dimension of our tools, where the goal is not to extract patterns but to engage with the data. Just as *Gephi* helps user to discover their networks, *Hyphe* helps users to discover the content and structure of the web. Our tools are recognised as exploratory devices in the tradition of Tukey's *Exploratory Data Analysis*.

In its last session, the scientific advisory board also began a discussion about the legal and ethical issues around *scraping*, a matter on which European legislation has also recently changed. We decided to acquire more information and to look closely at any scraping tasks requested. However, no scraping has been requested since 2014, probably because *Hyphe* has become more visible and because *Twitter* data seemed a more fruitful hunting ground for many researchers. We also had productive discussions with the Commision Nationale de l'Information et des Libertés (CNIL – data protection commission) correspondents from both CNRS and Sciences Po.

Prospects for 2017-2019

We need greater visibility to reach a wider public and we have learned that this comes with satisfied users and relevant academic publications. *Hyphe* has gathered momentum over the years and now seems our best opportunity to consolidate our identity and scale up. We plan to conduct our own research using *Hyphe*, to publish a paper highlighting its features in the context of social science research. External demand now exceeds our capacity because installing and sustaining *Hyphe* instances costs time and money. We have encouraged users to install it on their own servers (with the help of their university IT services) but this has turned out to be difficult and ultimately to require a lot of our time as well. We lost users because of this, but we have learned that many people are very keen to install it.

We plan to develop a "*Software as a Service"* (Saas) version of *Hyphe*. This will look like an online platform where users could pay for spawning their own instance of *Hyphe* for a limited time. The fee would at least pay for the server and possibly generate an additional margin to contribute to Dime Web's operating budget. This system would allow us to scale up much more efficiently, with the expected benefit of at least giving more training sessions. We have not yet established a business plan.

In order to transform *Hyphe* in this way, we need to replace a number of incompatible components and invest our energy in packaging technologies to simplify software installation and maintenance. We also need to find a platform for spawning servers (inside or outside Sciences Po) and integrate the code with it. Finally we need to develop a web interface for the service and adjust our public communication. This will require both time and specialist skills that are not present in our team or the medialab. We plan to hire a developer to help us and possibly some consultancy. An initial prototype of Hyphe Saas is scheduled for the first trimester of 2018 with closed testing in late 2018 and public release in spring 2019.

Appendix

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.