



HAL
open science

EQUIPEX DIME-SHS (2011-2016). Scientific Report - Convention ANR-10-EQPX-19-01

Anne Cornilleau, Anne-Sophie Cousteaux, Guillaume Garcia, Mathieu
Jacomy, Laurent Lesnard

► **To cite this version:**

Anne Cornilleau, Anne-Sophie Cousteaux, Guillaume Garcia, Mathieu Jacomy, Laurent Lesnard. EQUIPEX DIME-SHS (2011-2016). Scientific Report - Convention ANR-10-EQPX-19-01. [Rapport de recherche] CDSP. 2017, pp.16. hal-03429999

HAL Id: hal-03429999

<https://sciencespo.hal.science/hal-03429999>

Submitted on 14 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EQUIPEX

DIME-SHS

2011-2016



PROGRESS OF THE PROJECT

The purpose of the DIME-SHS “*équipement d'excellence*” (excellence facility) was to equip France with an innovative structure for the collection, processing and dissemination of three types of social science research data: quantitative, qualitative, and web. To establish this facility, we divided the project into three instruments: DIME quali, DIME quanti, and DIME web. In our view, the initial objective has globally been met: the content of this report will allow the project to be assessed instrument by instrument. However, we have not yet been able to apply multiple instruments to projects simultaneously and thereby to contribute to the development of connections between social science methodologies.

Establishment

DIME quali

The qualitative survey bank was built in two phases: first, preparation of the system fundamentals, then stabilization and development, which is still underway. Beginning in 2011, the different dimensions of this innovative project had first to be defined and developed. Starting with a first prototype developed between 2010 and 2011 as part of the National Research Agency's “Reanalysis” project, it took two and a half years to establish an initial version of the beQuali system. The IT application for accessing surveys was produced in mid-2013. It already comprised its current components (metadata presentation page; survey on survey module; document exploration and visualization tools; deposit registration, ordering and download interface; administrative functions), including the associated editorial website. A first version of the processes, methods and professional techniques needed was produced (classification and naming of individual documents; TEI-based interview transcriptions; documentation and contextualization through metadata input and the production of a “survey on the survey”; data anonymization; usage and deposit contracts and consent forms). This test was carried out on three surveys conducted by researchers at Sciences Po in 2011.

The subsequent years were spent improving the system, in different respects:

- An overhaul of the beQuali site was completed in 2015. The architecture and content of the editorial site pages were made simpler and more informative. For the access application, the tools used to consult surveys were improved (posting of a more accurate and manageable document inventory in Excel format; implementation of secure survey downloads 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; interface in English).
- Working procedures, methods and techniques were improved. The collection and usage contracts and consent form were made legally watertight with the help of a specialist lawyer. The anonymization procedure was formalized in consultation with CNRS' legal specialist on data protection and an ad hoc working group drawn from the Scientific and Technical Committee (CST). An in-house seminar was organized to decide the model and survey investigation method (written report and audio chaptering on the website). On the archiving and documentation side, the survey collection guidelines were set (definition of the documents necessary, or essential, to making a survey available) and the classification and naming plan improved, in order to adapt to the specificities of new surveys and to the need to raise awareness of survey deposit among researchers. 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 (sociopolitical data center), but returned to the originators or redirected to the appropriate archive departments. The issue of long-term archiving of digital files was anticipated, via a collaboration with the Huma-num Very Large Research Infrastructure, which acts as an interface between Equipex and CINES (national higher education computer center). The protocol for CINES deposit is currently being finalized. More generally, a process of case-by-case document review was undertaken, in order to maintain interoperability between beQuali and other systems or entities (Quetelet portal, CDSP website, CINES, archive software such as LIGEO, Dataverse archiPolis, etc.). A protocol for digitizing paper 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 digitized by this method.
- A formal survey deposit and selection procedure was established in 2015, in concert with the beQuali CST. Links were formed with the archiPolis network, in order to identify surveys that could potentially be included in the beQuali catalogue, and to raise awareness among researchers about keeping surveys for possible deposit.

The tool is now operational. The stumbling block preventing the deposit of new surveys, and therefore the readiness of researchers to share their surveys, seems to have been removed; in this respect, the primary cause of the difficulties encountered is the mismatch between the number of surveys awaiting processing and the size of the team, which is still small given the multiple and very time-consuming tasks involved in survey processing. The most salient issue today is the need to establish a proactive strategy to encourage use of the instrument, notably by developing the use of surveys for teaching, which seems to be the optimum method for attracting users.

DIME quanti

It has taken four years to establish ELIPSS (Internet-based longitudinal study for the social sciences). In fact, the test phase consisted in developing the full-scale panel. From the 4500 dwellings drawn randomly in an experimental INSEE process, 1039 people aged 18 to 75 agreed to take part and were administered 36 surveys between 2013 and 2016, to complete online using provided Samsung touchscreen tablets. The different dimensions of this innovative project had to be defined and developed:

- the method of recruiting the panel members (the pilot study tested three recruitment models: by letter, by telephone and face-to-face, and determined that the use of face-to-face investigators was the most effective method);
- automated procedures for tracking panel members (the protocols for reminding nonrespondents, by telephone and by letter, produced a survey response rate generally in excess of 80% and an attrition rate of less than 7% a year. The panel management software, developed in-house, centralizes the contact details of the panel members and the references of the tablets provided. It also stores contact history, technical incidents and completed surveys. Finally, it automatically generates reminder letters);
- the technical specifications, the development of the application and programs to conduct surveys on touchscreen tablets (drawing on the experiences of the LISS panel in the Netherlands and of INSEE, it was decided to use the Blaise software employed by national statistical institutes for their surveys, combined with an Internet-adapted stylesheet designed by a web designer, which allows respondents to complete the surveys on their tablets);
- the methods for system users to propose surveys (the first call for survey projects was issued in 2011 within the DIME-SHS consortium, and the DIME Quanti Scientific and Technical Council selected the first five surveys for production. Then, a further three calls for projects related to the pilot study, making a total of 13 possible additional research projects).

The pilot study was conducted by two providers (TNS SOFRES and Bouygues Télécom), selected through a call for tenders:

- with survey institutes for the face-to-face recruitment of the panelists and by telephone on the basis of a sample of households drawn from the census;
- with the telephone companies for the supply of the tablets and the associated Internet subscriptions.

The call for tenders with the phone companies was prepared in concert with a legal adviser specializing in the telecom markets. The specifications for the tablets and subscriptions, for their delivery to the panelists and for their management were drawn up at this stage. After the selection of Bouygues Télécom, several months of negotiations were needed to draw up the five-year contract.

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 instrument, while reducing the size of the panel from a minimum of 5000 to 3000 individuals for budget reasons.

In order to prepare for the expansion of the panel, the contract with Bouygues Télécom had first to be renegotiated in line with changes in the Internet service and tablet supply market, but this time without the (costly) assistance of a legal specialist, which prolonged the negotiations and delayed the second rollout phase by several months. Once the negotiations were completed, the rollout was put at risk when Archos supplied a tablet model that did not match the model agreed in the negotiations. This risk was handled by the operational team, in particular by the IT developers, who had to test this new model at short notice and adapt the tools previously developed. No legal solution could be found to relieve the project of the costs associated with this unexpected event at such a crucial moment in the rollout.

At the same time, to guarantee the best possible sample, an agreement was established with INSEE to draw 10,000 households from the census, after a favorable ruling was obtained from INSEE's board of directors and

the agreement of the Privacy Committee, and after the project was presented to the Commission for Demographics and Social Questions at CNIS. A public tender was issued for the face-to-face recruitment of 2700 new panelists. The service provided by the IPSOS Institute, selected in July 2015 at the end of this tendering process, was partly financed by the Université Sorbonne Paris Cité (USPC) community of universities and establishments (Comue).

Between January and July 2016, 2500 new panelists joined the system. They were provided with Archos tablets and trained in their use over the phone by the Bouygues Télécom service provider. With the 800 panelists from the pilot study (also equipped with the new tablet model), there were therefore 3300 individuals on the ELIPSS panel to complete the first research survey administered to the expanded panel in September 2016.

DIME Web

Since 2011, the Web instrument has led to the creation of several dedicated digital social science tools, and to the development of direct support for some 20 projects through assistance and training, and of indirect support for an indeterminate number of users of the tools around the world. The instrument was developed solely by two engineers recruited in 2011 and 2012. In the space of a few years, they created a range of free, open source tools, the most important of which is the Hyphe crawler, now used by researchers in several European countries and in North America. They also adapted the instrument to fit as closely as possible with researchers' needs. Implementation was therefore gradual and iterative, rather than following the model of a launch followed by operation. The development of Hyphe, together with the governance structure for the instrument, thus took place over several years, with a turning point in 2014 that we present in the next section.

Tracking

The project's Coordinating Executive is responsible for ensuring that the project proceeds smoothly. This body consists of the project coordinator, Laurent Lesnard; the head of the CDSP's digital projects team, Geneviève Michaud; the CDSP's secretary general, Marie d'Arcimoles; and the heads of the three instruments: Guillaume Garcia (quali), Anne Cornilleau and Anne-Sophie Cousteaux (quanti), and Mathieu Jacomy (web). The governance of DIME-SHS relies on several bodies with well-defined roles. They gather on a regular basis depending on the strategic or operational questions that arise. The Steering Committee, made up of the directors or presidents of the consortium's institutions, makes decisions on the overall direction of the project. The Scientific Advisory Board, composed of internationally recognized independent experts, provides general scientific guidance to the Coordinator and Steering Committee. At least three of the Scientific Advisory Board's members must be experts on ethics. In addition, for each instrument, we set up a Scientific and Technical Committee (CST). Composed of French-speaking experts, these committees are responsible for selecting the research projects to collect data (quanti and web – for the latter however, see below) or for prioritizing the archiving, documentation and dissemination of qualitative surveys. There is one last governance structure that needs to be created, a Users' Committee which will give us feedback on the services we provide. Across all these governance bodies combined, we have had 6 to 8 meetings per year since 2012. We have experienced many difficulties since the project started. The biggest is undoubtedly managing the human resources, in particular the hiring of new staff members. These difficulties take a heavy toll on the operational teams.

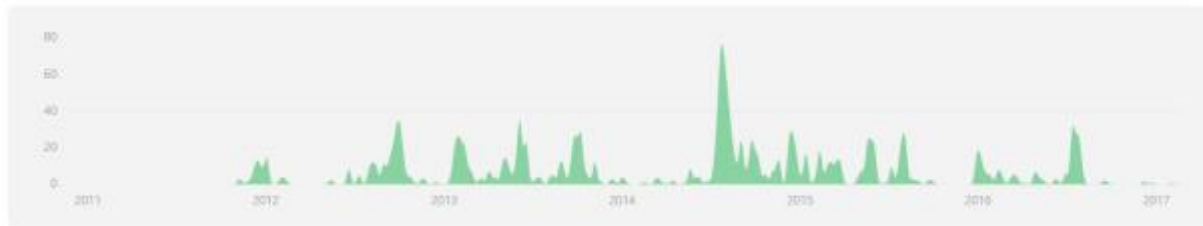
DIME Web

Constructing the offering and outlining the business plan entailed several successive stages of trial and error. The principle of a Scientific and Technical Committee to steer policy for the instrument was present from the start, and continues today, although the committee's role has not always been exactly the same. 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 tedious 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 between 2011 and 2017, combined with extensive day-to-day maintenance, achieving operational maturity in 2015 and wider dissemination in 2016. The

code contains more than 2000 commits and the website, set up in 2015 and including a demo, has attracted more than 7600 single users (according to Google Analytics).

Dec 19, 2010 – Mar 23, 2017

Contributions to master, excluding merge commits



“Commits” on Hyphe, visible on its GitHub repository

Implementation of the facility

- *Management structure*

DIME quali

The implementation of beQuali is currently reliant on a six person team, with a range of skills in social sciences, archiving and documentation. The operational team is responsible for coordinating the project, managing the entire survey processing chain (initial contact with producers, collection, classification and naming, selection, digitization, anonymization, editing, documentation and contextualization, online publication), as well as for liaising with the digital team on IT development and infrastructure. The team was constructed gradually. It often took time to hire the different profiles, the team was weakened by staff turnover, and several temporary staff were needed to perform certain tasks (digitization and digitization monitoring, notably OCR processing).

The operational team is supported by the DIME quali Scientific and Technical Committee (CST) in the selection of survey deposit proposals, and in issuing scientific, technical and strategic recommendations for the direction of the project. The CST currently consists of 11 people, with different profiles: research and education (political science and sociology) and IST (archiving and documentation). It meets 2 to 3 times a year.

DIME Quanti

A team of a dozen people is responsible for implementing the ELIPSS instrument, with varying skills in social sciences, statistics and IT. The operational team is in charge of coordinating the project, managing the panelists and the stock of tablets, producing and disseminating the surveys, analyzing the statistical quality of the panel, and developing the IT and infrastructure. The team was formed gradually. In 2014, the decision was made to create three new positions (two engineers and one developer). For greater efficiency, the team – initially distributed between two partners in the consortium (Sciences Po and INED) – was gradually shifted to Sciences Po (the panel managers joined Sciences Po in 2015, the statistician will arrive in 2017). It often took time to hire the different technical profiles needed for the project. This was particularly difficult in the case of survey statisticians and developers, where the recruitment process could take several months, during which the positions remained vacant. The operational team is supported by the DIME quanti Scientific and Technical Committee (CST) in selecting survey projects, in issuing scientific, technical and strategic recommendations, and in testing the surveys on tablets. The CST currently consists of 15 researchers from different social science disciplines (sociology, political science, economics, social psychology), specializing in questionnaire-based surveys. It meets 2 to 3 times a year.

DIME Web

A dedicated Dime Web CST monitors the project selection process and the direction of the instrument, through one annual meeting and regular email interchanges. The instrument is also represented on the Equipex Scientific Advisory Board. The operational team consists of two engineers who pool their working time with the Sciences Po medialab, an approach that notably provides access to more varied skills (designers, expert JavaScript developers, researchers...).

- *Facility management methods (specific training, human resources, etc.)*

DIME quali

Since 2015, survey deposit proposals have been sought through national calls for proposals (one or two a year), combined with corresponding laboratory visits. A method and support documents have been developed for this purpose: proposal guide, deposit form (completed by survey providers, with the help of the BeQuali team if necessary), selection form (completed by the members of the CST), 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 being prepared for summer 2017). Once a survey is selected, BeQuali carries out all the necessary processing steps, with the help of the producers of the survey. At the end of this process, the surveys are made available free of charge to the whole scientific community, for download through the Quetelet Network portal, and in parallel on the beQuali website for online consultation.

DIME quanti

Survey proposals are made in response to calls for projects (generally once a year). The surveys are assessed and selected by the DIME quanti CST. The ELIPSS team then assists the selected research teams in designing the questionnaire and adapting it to the instrument, with particular attention to the longitudinal dimension and the specificities of the tablet mode. The ELIPSS team also manages the different stages of survey production: scheduling the questionnaire, testing, collection of data via the ELIPSS application, field monitoring and reminders, providing an exploitable data file. At the end of a 12 month exclusivity period for the research team that jointly designed the survey, the data are made available free of charge to the whole scientific community via the Quetelet Network portal.

- *Conditions of access to the facility for education*

DIME quali

Surveys in the beQuali catalogue can be used free of charge for university education, provided that the teachers and students concerned sign a contract of reuse. Courses in qualitative methodology using 3 surveys in the catalogue – When French, English and (French-speaking) Belgians talk about Europe; The French and Politics; the French Parliament and the European Union (1993-2005): Europe understood through Parliamentary roles – have been run by several members of the team since 2013, as part of the Sciences Po Masters Degree in Sociology, and more recently the Masters Degree in “Social Uses of Law and Legal Communication” at Assas. Similar methodological courses have been run elsewhere, notably in Grenoble. An open working group on the teaching of methods from qualitative survey archives has been running since 2016. The development of uses of the instrument, in particular for teaching purposes, has become a priority as a means of expanding the user network.

DIME quanti

The ELIPSS panel can be used for university education. The survey on “Cultural practices, media and information technologies”, conducted in July 2015, was designed with the students on the Sociology Masters programme at Sciences Po. Data made available from ELIPSS surveys are also used in the teaching of statistical methods applied to the social sciences (Sciences Po, École Normale Supérieure de Lyon), where the teacher and students sign a license for the use of the data. In addition, 5 PhD candidates have entered requests to reuse data for their theses.

DIME Web

Education encompasses different realities that the instrument covers at several levels. One-off requests for help, which are also often requests for guidance, are dealt with at Médialab’s monthly open-doors workshop. This is often the occasion for initial contact, which then develops into a full service role. Dime Web gives courses to groups of researchers (and PhD candidates), either as part of a project support mission, or independently. Since 2014, a fee has been charged for these courses. Training was also provided for a company in 2015 under different conditions and by way of experiment. Finally, the Dime Web and in particular the Hyphe tools are taught in Masters classes by research academics and in French secondary schools (*lycées*) through Sciences Po’s FORCCAST programme. There is no charge for these classes and they are not paid for by the facility except through the support given to the teachers and, sometimes, the hosting of Hyphe structures. It should be recalled that Hyphe is a free, open source tool, which can be freely downloaded and installed without us being informed. We often learn about it subsequently when users come to us for support and training.

- *Conditions of access to the facility for research by a user inside or outside the consortium (charges, time quotas for different types of user, etc.)*

DIME quali

Before the formal procedure for the deposit and selection of surveys could be established, the surveys were collected directly by means of interpersonal or professional networks, at Sciences Po or beyond (INED, archiPolis). Since the end of 2015, calls for survey 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 raise excessively excessive difficulties of data confidentiality, sensitivity, ownership or legibility; the languages used in the survey must be within the capacity of the team; the source researchers must undertake to take part in certain operations (validation of the limits of the material to be retained and of the classification plan, definition of an anonymization plan, scientific contextualization). A deposit agreement must be signed for each survey deposited. The entry is processed by the beQuali team free of charge. Use of the surveys is also free. Use is subject to status and project conditions, and to the signature of an agreement covering the conditions of reuse (undertaking not to break data confidentiality, to adopt a respectful attitude to the survey originators, to cite the source when reusing data, etc.).

DIME quanti

The first three calls for survey projects, during the establishment of the system, were confined to research teams that were members of a DIME-SHS Equipex partner institution. Since 2015, calls for projects have been open to the whole scientific community in France and abroad. In order not to restrict the number of proposals received under calls for projects, at a time when the system is yet to become familiar to the scientific community, surveys are conducted free of charge. Indeed, according to the business plan produced in 2014 for the DIME-SHS excellence facility, very few research teams have sufficient funding to cover the costs of an ELIPSS survey (estimated at around €100,000 for 30 minutes of questioning with 3000 panelists, not including payroll). The risk of this is that it would no longer be possible to select survey projects on scientific grounds, a situation experienced by the LISS panel in the Netherlands since payment for use was introduced. By contrast, public institutions chosen by CST DIME quanti, which have a budget for statistical studies, are asked to fund the survey jointly.

In addition, the projects selected covered by an agreement on joint data production, which in particular defines the roles of each party during the production phase and the conditions for the use of the data (citation, exclusivity for 12 months followed by availability for research, etc.).

DIME Web

Free, open source tools such as Hype can not only be freely downloaded and installed by third parties, but in some cases are directly functional online, as is the case with the Table2Net utility, which received 8700 hits in 2016 (<http://tools.medialab.sciences-po.fr/table2net/>). Since 2014, a fee has been charged for access to the facility for support, training or joint research development, applicable to both members and nonmembers of the consortium. The instrument is not yet working at full capacity, so there is currently no limit on user numbers. When this situation is reached, the CST will be responsible for choosing projects.

- *Integration of the Equipex into a regional, national, European or global infrastructure, more broadly positioning of the Equipex within international competition.*

At the French level, the activities of DIME-SHS are coordinated with and assisted by the two Very large research infrastructure in social sciences and humanities: Progedo and Huma-num. At European level, DIME quanti is cooperating with other academic probability based Internet panels (Netherlands, Germany, Norway, and Iceland). A proposal was recently submitted in response to the call for Design Studies (Infradev-01-2017 H2020).

Scientific results achieved since the start of the project

- *through use of the facility by members of the consortium*

DIME quali

Between 2011 and 2015, before the establishment of the formal deposit and selection procedure, 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). In the case of one of these surveys, processing had to be stopped after the producing researcher was dismissed. Currently, 6 of the 9 surveys are already available, and the processing of the last 3 is in its final stage. They tackle a variety of topics, such as: electoral behaviors, opinion and political institutions; educational and residential choices; the family, the couple and sexuality; gender; the professions.

In terms of reuse, members of the consortium have conducted experiments in the teaching of the methods. A research team from ANR Réanalyse has reanalyzed the survey on "The French and politics" (publication pending). One political science thesis has used some of the materials from the survey "The French Parliament and the European Union (1993-2005): Europe understood through parliamentary roles."

DIME quanti

During the first three calls for projects issued within the DIME-SHS consortium (2011-2014), 21 survey projects were deposited. 13 projects, representing 92 researchers, were accepted by the DIME Quanti CST, and of these, 4 projects took advantage of the longitudinal dimension of the system by drawing on several waves of surveys. In the two subsequent calls for projects, open to the whole scientific community, 7 projects were backed by a consortium team, i.e. 1/5 of the projects deposited. In other words, 27 surveys were produced for consortium member teams, on a variety of topics such as cultural practices, electoral behaviors and political opinions, fertility, family relations, health, social stratification, the environment and energy, residential choices.

DIME Web

Since 2011, 30 projects have been submitted and 21 chosen, mainly in France but also elsewhere in Europe. 14 of the 21 chosen projects were from consortium members, including 11 for different Sciences Po laboratories (CEE, CEVIPOF, CERI...). One article is currently being revised with CEVIPOF and a note has been published on the CEVIPOF site. In addition, members of the operational team have signed or co-signed 13 academic publications since 2011, including one article on the Gephi network analysis software on PlosOne in 2015 (lead author), and one article presenting Hyphe to the ICWSM 2016 conference (lead author).

- *through use of the facility by non-consortium members*

DIME quali

The formal procedure for the deposit and selection of surveys was established in 2015. Some 15 proposals were submitted for two calls for deposits, and 8 surveys were selected by the CST. Two of them are surveys by a researcher who is a member of a consortium laboratory (Paris V Descartes), the other 6 from non-consortium laboratories. Since 2015, a significant effort has been made to diversify the catalogue in disciplinary terms (sociology surveys now outnumber political science surveys) and in institutional terms (surveys from Sciences Po laboratories are now outnumbered by those from other university laboratories). The new surveys tackle new themes such as immigration, heritage, hospitality, social movements, parenthood, territorial governance, extreme right-wing activism, poverty and exclusion. With regard to levels of reuse, the proliferation of users combined with the inevitable interval before results are published, means that the impact cannot yet be measured.

DIME quanti

In 2015, 32 projects were deposited by producing teams from 27 research institutions under two calls for projects issued beyond the consortium partners (one for surveys conducted on the pilot group and the other for surveys conducted with the expanded panel between 2016 and 2017). Following these two calls, DIME Quanti chose 16 projects, representing 106 researchers. Of these selected projects, 11 surveys have already been produced, tackling new themes such as mobilities, attitudes to the state, the relations between housing and diet, waste, etc. Finally, a European collaboration that has been continuing since the start of the project has led to the production of two comparative surveys conducted in 2014 and in 2015 in panels similar to ELIPSS in Germany and the Netherlands.

DIME Web

Since 2011, 7 of the 21 projects selected have come from non-members of the consortium. One article on these projects has been published in the journal Netcom. In addition, a paper was delivered at the APSA 2015 conference by a German researcher who did not contact us, but which was identified during our monitoring process

- *specify the Equipex's contribution to the achievement of these results (could those results have been achieved without the facility?)*

DIME quali

At present, there is no other facility equivalent to beQuali in France, at least in the world of public research and education. The sharing of "raw" survey material between researchers is still a rare practice, in particular because until now there has been no mediation system that could do this, other than through personal links between researchers or research groups that agree to pool their data. In addition, beQuali offers a response to a number of problematic issues in the reuse of qualitative surveys, which are not or cannot be handled individually by local research or IST actors: restoring the potential for raw materials to be reused by working on the documentation and contextualization of surveys, applied to an expanded body of archives; covering digitization costs in the case of non-digital archives; handling the problems of data confidentiality (retrospective anonymization or obtaining consent from survey subjects; management and contractualization of access requests; management of file access security).

DIME quanti

The surveys and data produced through the ELIPSS panel are unique. There is no equivalent system in France, i.e. an online panel based on a random sample drawn from the general population in such a way as to include individuals originally without access to the Internet. While probabilistic Internet panels have been developed in Europe and the USA, ELIPSS is the only one to supply the same tablet to all participants and thereby to achieve uniform online data collection. Some teams that have used the ELIPSS system might perhaps have had the financial resources to administer their questionnaire by employing a survey institute. However, most of the teams that have used the ELIPSS panel have grasped a unique opportunity to conduct a survey on a random sample representative of the population living in mainland France. In addition, certain innovative projects required specific IT developments (e.g. the use of microphone and camera, navigation on maps, the creation of a survey in the form of a game), involving much more than a simple exchange of questions and answers.

DIME Web

A minority of projects could also have been run by similar teams in Europe, such as the Digital Methods Initiative in Amsterdam, but at a higher cost because of the distance. However, most of the projects supported, based on the use of Hyphe, would have been difficult to conduct elsewhere without modifying the research design. Since the specificity of Hyphe is its application to a quali-quantitative approach in social sciences, it would have been difficult for the researchers to have their needs met by industrial tools, and no academic alternative exists, no doubt because of the development costs.

Contributions of the Equipex and its position in site policy

DIME-SHS was supported from the beginning by Université Sorbonne Paris Cité (USPC, a group of 14 higher education institutions). DIME-SHS is part of the project to create a Data Institute at USPC. Finally, DIME-SHS received a grant of 496,800 euros from USPC. At the Sciences Po level, DIME-SHS is completely integrated in the site policy.

HUMAN RESOURCES:

From the start of the project, the people recruited under the Dime-SHS framework were technical personnel, mostly at a level equivalent to design engineer. Apart from a transversal project manager and a statistician, they can be divided into three distinct categories:

- Engineers or research officers specializing in research data (qualitative, quantitative, web)
- "Digital project" engineers (web developers and designers, system and network engineers collaborating transversally with the teams of the three Equipex instruments)
- Technical staff responsible for managing the panel and the stock of digital tablets for the ELIPSS instrument

Most of the staff were hired on fixed term contracts (largely 18 or 36 months), two on apprenticeship contracts. Over the period, 5 were switched to permanent contracts, funded by the Equipex (including one who left the project at the end of 2014). Initially financed out of Equipex funds, the salaries of three members of staff are now funded by their employer, although they continue to contribute to the project: 1 contractual engineer employed by MENESR (ministry of education, higher education and research); 2 "IT project" engineers now with

permanent contracts with FNSP (national foundation for political sciences).

In all, 27 different people have been hired since 2011 on Equipex funds or under a joint funding arrangement, i.e. the equivalent of 10.3 full-time employees on average per year since 2011.

The list of employees set out below does not take account of the working time of permanent staff (FNSP, CNRS and MENESR) of the CDSP who, since the beginning of the project, have been closely involved in the establishment and operation of the instruments. Nonetheless, they are represented in the graph below.

In addition, while the list specifies the new hirings each year, it does not reflect the significant staff turnover and the – often lengthy – vacancies that have marked the project. That is why the number of full-time working equivalents has been specified for each year. Likewise, the graph below, based on staff numbers as at June 30, shows the number of staff actually present on a given date and the changes over the period 2011-2016. Finally, it should be noted that while the number of staff assigned to the project from Equipex funds during the pilot phase was satisfactory, it was not sufficient over the period 2015-2016, which notably corresponded to the recruitment of 2500 additional panelists, placing significant extra workload on the teams. As a result, the project will only be fully staffed from 2017, with four additional hirings planned, two of which have already taken place as of March 2017.

Staff hired with Equipex funds in 2011: 3

- 1 transversal project manager
- 1 quantitative data engineer
- 1 web data engineer

Full-time equivalent worked: 0.7

2012 – Staff hired with Equipex funds: 11

- 1 transversal project manager (transition to permanent contract)
- 2 quantitative data engineers (including 1 hired during the year)
- 2 panel managers (hired during the year)
- 2 qualitative data engineers (including 1 hired during the year)
- 2 web data engineers (including 1 hired during the year)
- 2 “digital project” engineers (hired during the year)

Full-time equivalent worked: 6.7

2013 – Staff hired with Equipex funds: 15

- 1 transversal project manager
- 3 quantitative data engineers (1 hired during the year with joint funding by Gis Quetelet/Progédo and 1 transition to permanent contract)
- 2 panel managers
- 2 qualitative data engineers
- 2 web data engineers
- 4 “digital project” engineers (including 2 hired during the year, 1 with joint funding by FNSP)
- 1 statistician (hired during the year, position jointly funded by INED)

Full-time equivalent worked: 13.4

2014 – Staff hired with Equipex funds: 19

- 1 transversal project manager
- 5 quantitative data engineers (3 hired during the year with joint funding by Gis Quetelet/Progédo and 1 transition to permanent contract)
- 3 panel managers (including 2 hired during the year)
- 2 qualitative data engineers (including 1 hired during the year)
- 2 web data engineers (including 1 transition to permanent contract)
- 5 “digital project” engineers (including 2 hired during the year, 1 with joint funding by FNSP)
- 1 statistician (joint funding by INED)

Full-time equivalent worked: 13.5

2015 – Staff hired with Equipex funds: 18

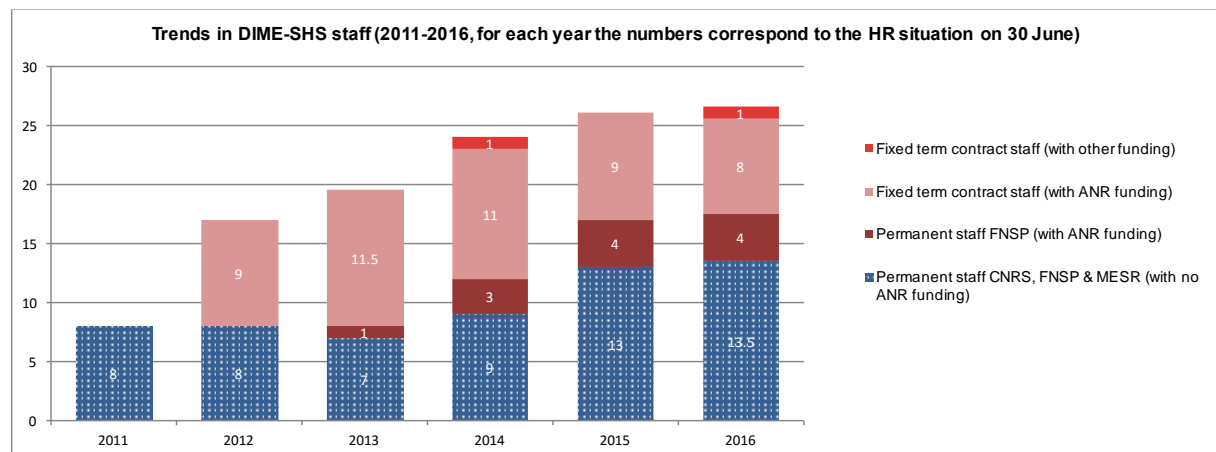
- 5 quantitative data engineers (including 1 jointly funded by Gis Quetelet/Progédo)
- 2 panel managers
- 2 qualitative data engineers (including 1 hired during the year)
- 2 web data engineers (including 1 transition to permanent contract)
- 6 “digital project” engineers (including 2 hired during the year, 1 with joint funding by FNSP and 1 with joint funding by USPC)
- 1 statistician (joint funding by INED)

Full-time equivalent worked: 14.2

2016 – Staff hired with Equipex funds: 17

- 5 quantitative data engineers
- 3 panel managers (including 1 hired during the year)
- 1 qualitative data engineer
- 2 web data engineers
- 5 “digital project” engineers (including 1 hired during the year and 1 with joint funding by FNSP)
- 1 statistician (joint funding by INED)

Full-time equivalent worked: 13



SOCIO-ECONOMIC IMPACT

Links with the regional research and innovation environment

DIME Web

The link with the local research and innovation environment is based exclusively on technology exchange. The free and open source technologies used and/or developed by Dime Web and more broadly by the medialab are also being developed by an ecosystem of private actors united by the network topic and also often by the practice of data visualization. The spin-offs often come through the (two-way) transfer of skills at conferences or informal meetings, or less frequently through contracts, such as the ByteClub startup’s production of the Hyphe Browser module thanks to funding from FORCCAST, with Dime Web project management.

Users (outside Equipex)

DIME quali

The only categories that have had access to the beQuali surveys – entirely free of charge – are members of the scientific and academic community in the broad sense, in other words researchers, teachers, engineers and students, for projects entailing the reuse of data for purposes of research or training in research. Access is obtained after surveys are identified on the beQuali website and an access request is submitted to the beQuali team, which examines the validity of the requests (status of the applicant and nature of the request). Access rights are granted for one or more specific surveys, after a contract of reuse has been signed. This allows the

user to explore the documents in the requested surveys, on the beQuali website, and to download the files from the Quetelet portal.

List of institutions to which the applicants belonged: Sciences Po, IUT Rodez, University of Paris 5 Descartes, EPHE, INIST, University of Alger, University of Amsterdam, AgroParisTech, University of Lille 2, University of Strasbourg, MIT, University of Evry, University of Paris 2, University of Geneva, University of Aix Marseille, MMSH, University of Paris 8, INRA, ENS, University of Grenoble, EHESS, University of Paris 10, Catholic University of Louvain, Saint Louis University of Brussels, University of Bremen.

DIME quanti

Since 2015, non-consortium users have had the same access as consortium users to the ELIPSS panels data production service (previously, access was restricted to consortium teams): access is based on calls for survey projects and selection by the DIME quanti Scientific and Technical Committee. Many universities in France are represented by the teams that produced the surveys (Universities of Montpellier, Cergy, Avignon, Rouen, Le Havre, Lyon 2, Lille 2, Nantes, Bretagne Sud, Versailles, Paris 7, Paris 8, Paris 10, Paris Dauphine; Political Studies Institutes of Bordeaux and Grenoble), as well as universities around the world (Stanford University, UCLA, University of Texas, University of Lausanne, Swiss Federal Institute of Technology, Brussels Free University, University of Mannheim, University of Bergen, University of Utrecht, University of Goettingen). Other research institutes are represented, such as Ined, Inra, Inserm, Irstea, Iffstar, Insee, the Employment Studies Centre, the Ministry of Culture, the Paris-Malaquais Higher National School of Architecture, the Lille Higher National School of Architecture and Landscape, Lille Regional Teaching Hospital, or Lyon's Saint-Joseph Saint-Luc Hospital. Teams are often multidisciplinary, with projects that therefore cover a wide variety of disciplines and topics.

In addition, the data produced via ELIPSS and made available for reuse since 2015 have attracted 50 requests for access by more than 50 individual users. These individual users come primarily from non-consortium institutions (16 non-consortium institutions out of the 19 represented).

DIME Web

It is difficult to retrace who has used the free, open source tools we have developed, since we do not require registration. On the other hand, we have clues that enable us to broadly assess the spread and reputation of those tools. First, we know of instances of Hyphe installed in various laboratories or universities, generally because they have contacted us to help them install the software. These include UCLA (USA), King's College (UK), Aalborg University and ITU (Denmark), UNIL (Switzerland), FMSH, EHESS, Universities of Lille 3, Rennes 2, Paris Nanterre, Paris Descartes and Paris-Est Marne La Vallée (France). In addition to this, there are cases of Hyphe being used by actors otherwise unknown to us and/or from outside the academic community. For example, the activists at Utopies Concrètes produced a map of community organizations.¹ An analysis of *fake news* conducted by an American journalist using Hyphe was identified on Medium.² A German researcher published an analysis for a conference on the far right in his country, using Hyphe.³ Apart from these identified cases, we do not know how many there are that we know nothing about. On the other hand, we have an idea of the influence of our tools through the statistics on online hits. The Hyphe presentation website, which is in English and contains a demo, has attracted 7600 single users since it went live in 2015, for 13,000 sessions. More than 5000 of these sessions originated in France, 2000 in Russia, 1000 in the UK and Denmark, the other countries in the top 10 being the US, Germany, Italy, Canada, Brazil and Spain. Our other small online tools also attract visitors. Table2Net (for extracting networks from tables) has had 17,500 single visitors, and ScienceScope (scientometric utilities) 7500, since 2011.

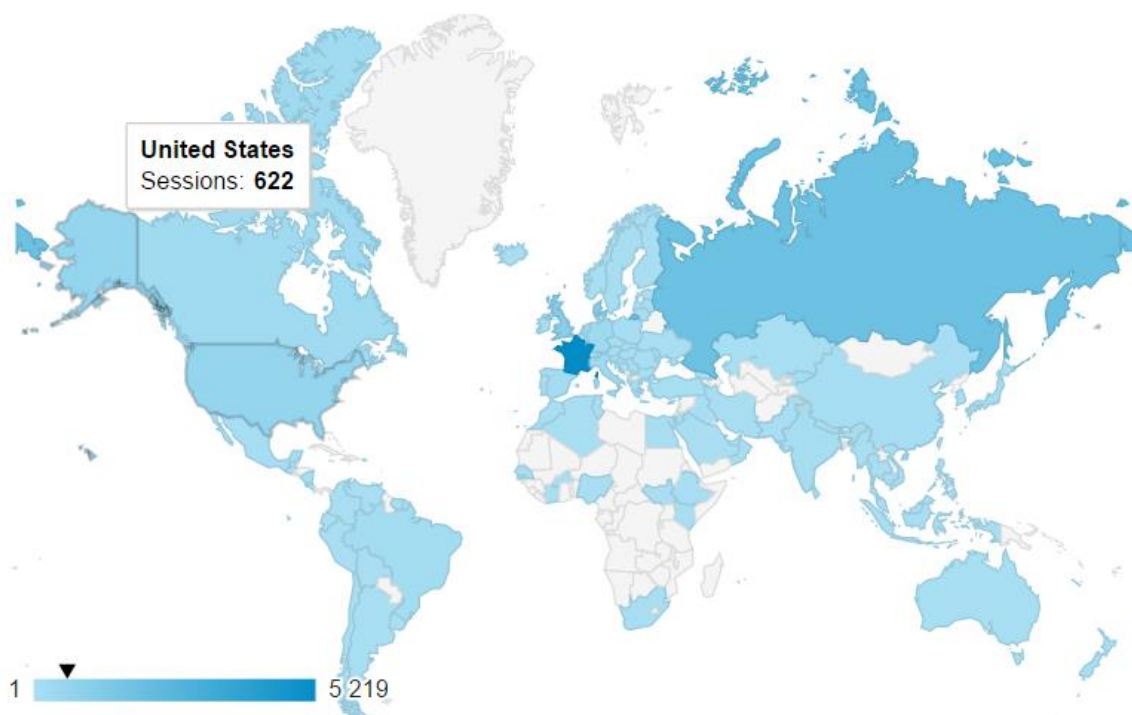
¹ <http://utopies-concretes.org/>

² <https://medium.com/@d1gi/the-election2016-micro-propaganda-machine-383449cc1fba#.ge10i1y8p>

³ <http://www.kai-arzheimer.com/my-apsa-2015-paper-the-afds-facebook-wall-as-a-hub-for-right-wing-mobilisation-in-germany/>



Trend in hits on the Hyphe showcase site (with demo) since its launch in 2015



Geographic data from the Hyphe presentation website (since 2015)

Development (creation of start-ups, companies, patents and licenses)

None

Reputation, actions to promote the Equipex

The primary target is the academic community. Promotional measures therefore take the typical forms specific to that environment, in other words contributions to conferences or laboratory seminars, and publications. Other actions are less common, such as laboratory visits for DIME quali, monthly discovery workshops and training workshops for DIME web, or a long article presenting the CDSP and the Equipex in the national newsletter of the Institute for humanities and social sciences (InSHS), part of the National Centre for Scientific Research (CNRS), which was published in November 2016.

Essentially, the resources allocated to promoting the reputation of the Equipex – which can be valued at a total of €97,000 for the period in question – cover the following activities:

- Travel by the teams, in France and abroad, for visits to laboratories, contributions to scientific conferences or professional networks (approximately €54,000, including the cost of missions carried out by the CST)
- Development of Internet portals for Dime-Shs and for each instrument (around €18,000 in web setup and design services since the start of the project)
- Organization of scientific events (approximately €17,000 for 2011-2016). In particular, two public information meetings have been held: 1) in 2012, for the inauguration of Dime-Shs; 2) in 2016, to mark the 10th anniversary of CDSP.
- On a smaller scale, publication of communication media (posters, brochures, etc.), for a total amount of around €8000

For the upcoming period (2017-2019), the budget also includes two significant operations:

- Translation into English of the surveys conducted on the Elipss panel in order to increase their use (€30,000)
- Closing conference for Dime-Shs (€30,000)

DIME quali

As of today, 6 surveys are available and 11 are in process, at various stages (from initial data collection to prepublication phase), making a total of 17 surveys. Apart from the processing of surveys, and hence the expansion of the catalogue, a significant proportion of beQuali's activity consisted in promoting the instrument and integrating it into its environment. A large number of public presentations were made to the archive and documentation communities, in France and abroad (EDDI, IASSIST, CINES, MATE-SHS, AAF, ADBS, URFIST, etc.), and in the social science research world (AFS, AFSP, visits to a dozen laboratories), etc. Five publications written by members of the team were produced or accepted. An open seminar on the topic of "The reuse of qualitative surveys in the social sciences. Practices and methodological issues" was set up in 2015; five sessions have been organized with guest researchers, on different identified forms of re-use of qualitative survey archives (secondary analysis, reanalysis, history of the social sciences, education).

It is still too soon to determine the profile of the researchers who agreed to deposit their surveys, or the profile of users. With regard to the deposit of surveys, it would nevertheless seem that there is greater receptiveness amongst researchers approaching retirement or already retired, who perhaps find it easier to dispose of their archives for legacy purposes. Generally speaking, although it is too early to judge, calls for deposit proposals seem more effective when combined with awareness raising activities by the beQuali team in the form of laboratory visits, especially when they are subsequently backed up by local documentation departments. With regard to re-uses, again it is too soon to say, although clusters of use are already emerging (education, methodological research, secondary analysis) to the detriment of other uses (reanalysis for verification purposes, re-analyses). Two cases of use for doctoral theses show that established researchers are not the only possible users with regard to research projects, and that the reuse of surveys at PhD level is possible and should perhaps be encouraged.

DIME quanti

Promotional activities relating to the DIME quanti instrument consist in publicizing the system and the surveys produced at the time of calls for projects, when data are made available, and at seminars and conferences. They are primarily conducted by the team, but researchers who present their results also contribute to the reputation of the DIME quanti instrument. These promotional activities are primarily addressed to social science researchers. Between December 2012 and December 2016, 48 surveys were carried out with the panelists. Today, 7 surveys are available to the scientific community via the Quetelet Network portal, and have generated 51 requests for data from 50 users. The ELIPSS system has been presented numerous times since 2013, at major French institutions (INSEE, INED, CEREP), at seminars (French Statistical Society, French Association of Political Science, University of Lausanne, University of Mannheim), at conferences in France and abroad (International conference of Blaise users, French language polling conference, European Survey Research Association - ESRA, European Sociological Association - ESA, French Association of Sociology - AFS, Webdatanet, INSEE 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. In order to communicate with our (future) users, we have developed a website aimed at researchers and students (<http://quanti.dime-shs.sciences-po.fr/>) and we regularly send out newsletters on the ELIPSS project.

DIME Web

The researchers who are most receptive to digital methods are young social science researchers, who are more comfortable with digital technology. They are the ones best able to position their projects and to find funding. Historically, the most receptive areas are social network analysis and controversy mapping. We are naturally close to these communities in the Paris region, where our credibility is increasing from one project to the next. Our objective is to expand our user base to the other European countries, where the medialab is a benchmark in these domains. We have therefore committed working time to reach these potential users, with the hope of results within a few months or years, depending on the outcomes of quests for funding. Since 2011, but especially since 2014, courses and papers on the subject of Hyphe have been delivered in different countries in Europe and beyond: 14 in Paris and 5 elsewhere in France, 5 in the UK and the Netherlands, 2 in the USA, in Sweden and in

Belgium. In addition to this, a stay of several months in Denmark by one of the team's engineers, jointly financed by Denmark's DigHumLab, gave rise to 10 seminars or workshops in 4 different universities in Denmark and Sweden.

SECURING THE FUTURE OF THE EQUIPEX

DIME Quali

At the current stage of development of beQuali, the intended business model (institutional subscription to access surveys) cannot be introduced. This will be possible when the catalogue is sufficiently large.

DIME Quanti

As stated above, the cost of a survey (€100,000 for 30 minutes of questioning with 3000 panelists, excluding payroll) means that full self-funding is not a serious option. Indeed, a market study showed that it would take the total operational budgets of all of France's social science laboratories to fully finance the panel. It should be specified that the issue here is not the cost of a survey: given the high quality of the sample (a random sample drawn from the census, low attrition, high participation rate), this cost is very competitive by comparison with a similar method (by way of example, the French field of the European Social Survey costs more than €600,000 for a 60 minute face-to-face interview with around 2000 people). The problem is more to do with the limited resources of social science researchers. As was pointed out earlier, the decision was made not to ask researchers for a funding contribution, but to ask for this in the case of requests originating in institutions.

DIME Web

The business plan for the Web instrument does not provide revenues that match the costs of the two instruments, but the production of free and open source tools generates value that is not only financial, whether in the form of reputation, image, publication opportunities or educational programs. The largest costs come from long-term support for projects. On the one hand, however, they also cost the most in working time, and they work when the researchers plan the funding in advance, at the project design stage. On the other hand, one-off support or training actions, often financed from budget residues and financially more advantageous, are not so common and tend to be less interesting scientifically. The most promising market relates to the use of Hyphe. The potential user base seems bigger than we expected, because it goes beyond the framework of academia. For the moment, the spread of Hyphe is restricted by the difficulty of installation, which deters users and costs us much too much in day-to-day working time. This is the main bottleneck. We therefore plan to adapt Hyphe in 2017 and 2018 to a *Software as a Service* model, which will allow a scale transition based on automation. This change should permit a different economic model, based on wider dissemination and more numerous opportunities for paid training.

LIST OF THE PROJECT'S 10 MAJOR PUBLICATIONS

- Publications linked with the design and development of the facility

- Both Anne, Cadorel Sarah (2015), "Pour en finir avec l'original ? Des effets du numérique sur les archives scientifiques : le cas de beQuali", in Jean-François Bert et Marc J. Ratcliff (eds.), *Frontières d'archives : recherches, mémoires, savoirs*, Éditions des Archives Contemporaines, pp. 157-164
- Both Anne, Garcia Guillaume (2014), "Le chercheur, l'archiviste et le webmaster : la polyphonie patrimoniale ? Le cas de beQuali, banque d'enquêtes qualitatives en sciences sociales", in DUFRENE Bernadette (eds.), *Patrimoines et humanités numériques : quelles formations ?*, Berlin, Lit Verlag.
- Duchesne Sophie, Garcia Guillaume, (2014) "beQuali : une archive qualitative au service des sciences sociales", in Cornu Marie, Formageau Jérôme (eds.), *Les archives de la recherche, pratique des acteurs et enjeux juridiques*, Paris, L'Harmattan.
- Annelies Blom, Michael Bosnjak, Anne Cornilleau, Anne-Sophie Cousteaux, Marcel Das, Salima Douhou, Ulrich Krieger (2016) "A Comparison of Four Probability-Based Online and Mixed-Mode Panels in Europe", *Social Science Computer Review*, vol.34, n°1, p.8-25.
- Mélanie Revilla, Anne Cornilleau, Anne-Sophie Cousteaux, Stéphane Legleye, Pablo de Pedreza (2016), "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.
- Jacomy, M., Girard, P., Ooghe-Tabanou, B., & Venturini, T. (2016). Hyphe, a Curation-Oriented Approach to Web Crawling for the Social Sciences, *ICWSM 16*.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6), 1–18.
<http://doi.org/10.1371/journal.pone.0098679>

- Publications linked with the use of the facility

C. Cavalin, P-A. Rosental, M. Vincent, M. Olivier, V. Bonnetterre (2014) "L'enquête ELIPSSilice : passation en population générale d'un questionnaire élaboré pour mesurer l'exposome minéral (projet SILICOSIS)", *Revue des Maladies Respiratoires* 32 (supplément), p. A155.

Nicolas Douay, Aurélien Reys et Sabrina Robin (2015) "L'usage de Twitter par les maires d'Île-de-France"
Netcom 29-3/4 | 2015, 275-296 2015

Anja Thomas (2016), *The 'European Integration Paradox': Comparing EU Practice and Discourse on the Role of Parliaments in the EU in the Assemblée nationale and the Bundestag Across Time*, PhD thesis in Political Science, Sciences Po.

M.C Le Pape, M. Portela, E. Tenret (2016) "Quand on aime, on ne compte pas". Difficultés méthodologiques et stratégies de questionnement dans les enquêtes sur les aides matérielles et financières apportées aux jeunes adultes par leur famille Dossier de la DREES n°4, 37 p.

Tournay, V., Blasimme, A., & Jacomy, M. (2016) The Formation of Digital Networks in Controversial Areas of Biotechnology: Discussing the notion of public opinion through web-based big-data analysis Ethox Seminar.

FREE COMMENTS

DIME-Quali

The coordinator of the DIME quali instrument considers that the indicators used are insufficient to obtain a full understanding of all the work that has been done, not only on the surveys (time and back-and-forth with the researcher needed to process all dimensions of the surveys), but also on raising awareness in the target communities, given the impact of resource shortages on team stability and the continuity of the work. Beyond this, there is the fact that we are developing something that had to be invented from scratch, and that new problems emerged in the course of events, which makes it impossible to meet all the individual objectives announced at the start, in conditions that are themselves shifting.