



**HAL**  
open science

## Workshop: Let's take a look at a DDI-documented paradataset collected by an in-house panel information system!

Geneviève Michaud, Baptiste Rouxel

### ► To cite this version:

Geneviève Michaud, Baptiste Rouxel. Workshop: Let's take a look at a DDI-documented paradataset collected by an in-house panel information system!. EDDI 2019. 11th Annual European DDI Users Conference, Dec 2019, Tampere University, Finland. hal-02873594

**HAL Id: hal-02873594**

**<https://sciencespo.hal.science/hal-02873594v1>**

Submitted on 19 Dec 2022

**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.

---

Workshop:  
Let's take a look at

a DDI-documented paradata set  
collected by an in-house panel  
information system!

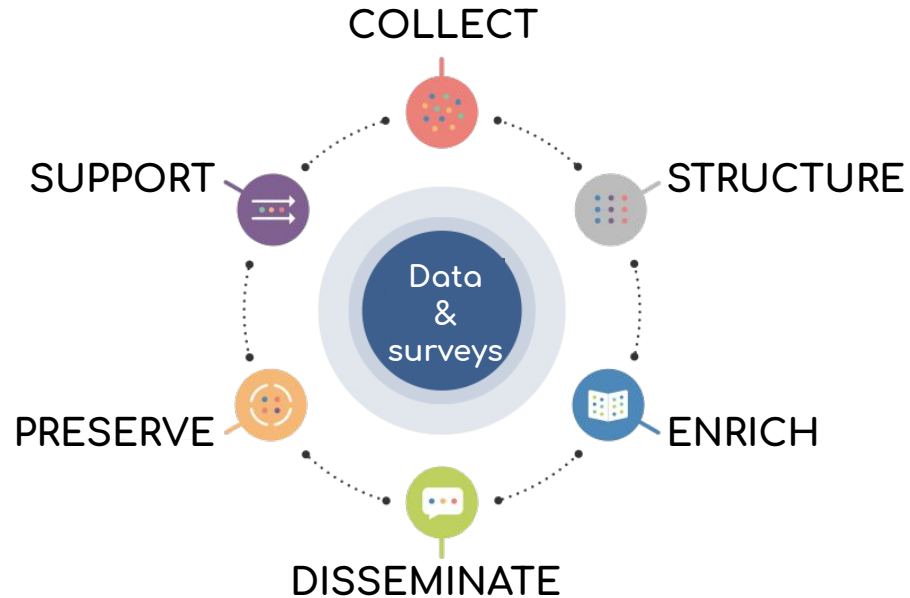
[Geneviève Michaud](#), [Baptiste Rouxel](#),  
EDDI 2019, 11th Annual European DDI Users Conference  
Tampere, Finland  
December 2, 2019

---

# The **E** **IPSS** project: context

---

# the CDSP, center for methods & data of Sciences Po



source: [Amédée](#), Carol Guinebert

---

## a grant, two concurrent projects



A 9 years project & 9 millions € grant from the french national agency, a compound project *EquipEx* **DIME-SHS** (ref: ANR-10-EQPX-19-01, 2011-2019).

The CDSP leads 2 sub-projects:

- [beQuali](#): a qualitative survey bank
- [ELIPSS](#): an internet longitudinal study

---

# ELIPSS project

ELIPSS is an **online panel that is representative of the population living in metropolitan France**. Panel members are randomly selected [...] and **provided with a touchscreen tablet and a mobile Internet subscription so that they can participate in monthly surveys**. [...]

The pilot study was composed of 1039 panel members recruited in 2012. 848 of them were still participating in the panel in September 2015. In 2016, 2500 new panel members have joined the ELIPSS panel.

The CDSP is in charge of the production process of the surveys, including the programming of the questionnaires, the panel management activities (incl. tier 1 mobile device technical support), the documentation and dissemination of data. For this purpose, the IT team is developing a tailor-made software suite.

---

# ELIPSS: internal resources and skills

A dedicated production, post production and preservation team:

- survey programmers and researchers support
- field, sample and fleet management support
- data and metadata scientists

A shared digital project team:

- web developers, web designer
- system administrator
- software project manager

---

# IT needs and resources

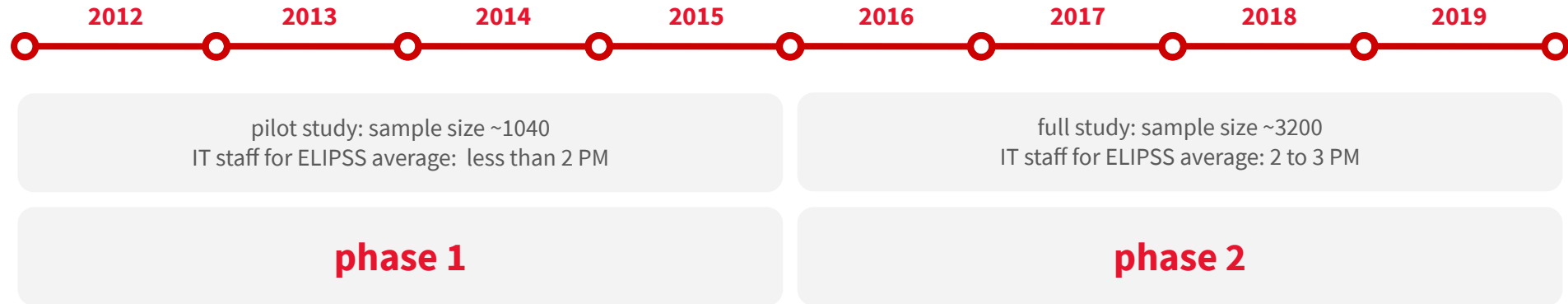
The CDSP's digital projects team is in charge of building and running an IT infrastructure:

- from scratch
- short production cycles



---

# study production timeline



---

# methods and tools: phase 1

**Technical stack keystones:**  
Python, Django, Javascript,  
libraries (too many)

**Methods:** one IT meeting a week, one project meeting every 2 weeks. Redmine and "big chunks" tasks.



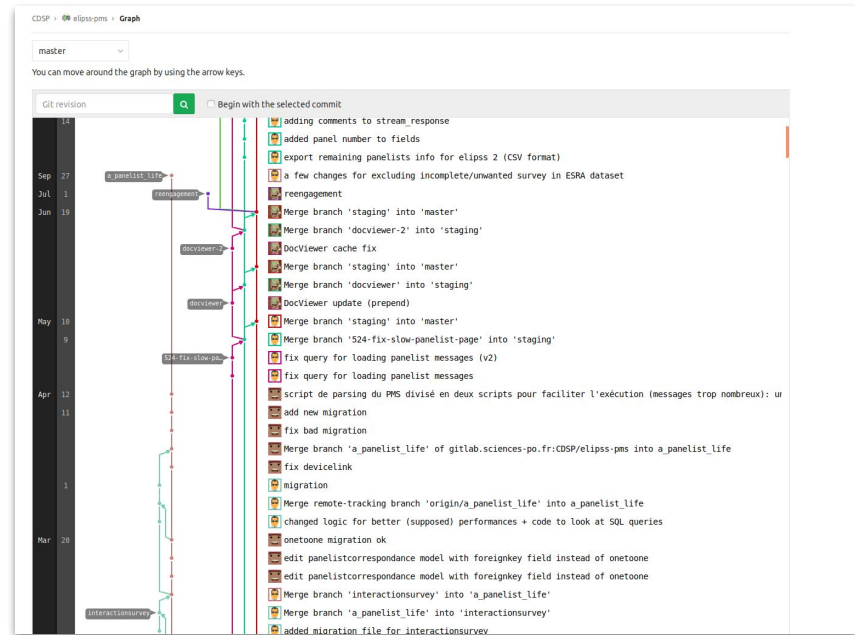
---

# methods and tools: phase 2

Code base rewriting (80%)

**Technical stack:** towards a more controlled technical stack

**Methods:** moving towards agile methods and sprint cycles. Specific sprints for code refactoring. Code review. Opting for gitlab and hiring a lead dev.



source: CDSP's gitlab instance, ELIPSS back office development branch graph

---

*“Any fool can write code that a computer can understand. Good programmers write code that humans can understand” — Martin Fowler*

# achievements

77 fielded surveys

of [which 61 are published](#)

and archived at the french long term  
archive certified repository ([CINES](#))

The screenshot displays the SciencesPo website's 'Online Resources' page. The header includes navigation menus for 'STUDENTS', 'FACULTY', 'BUSINESSES', 'ALUMNI & DONORS', and 'RESOURCES', along with social media icons and a 'FRANÇAIS' language option. The main navigation bar lists 'THE CDSP', 'EXPERTISE', 'PROJECTS', 'ONLINE RESOURCES', 'DATA DOWNLOAD', and 'DATA DEPOSIT'. The page content shows a filter sidebar on the left with sections for 'Filter by', 'Type', and 'Themes'. The 'Type' section has 'Qualitative survey /0' and 'Quantitative survey /53' (checked). The 'Themes' section lists various categories with counts, with 'ELIPSS /53' checked. The main content area displays a list of 53 online resources, including 'Digital practices - 6th wave (2018)', 'ELIPSS Annual survey - 6th wave (2018)', 'Attitudes and representations towards the State - 2nd wave (2017)', and 'Digital practices - 5th wave (2017)'. Each resource entry includes a title, a horizontal line, and the text 'Quantitative survey - [Year]'.

source: CDSP, online resources page, website (2019/11)

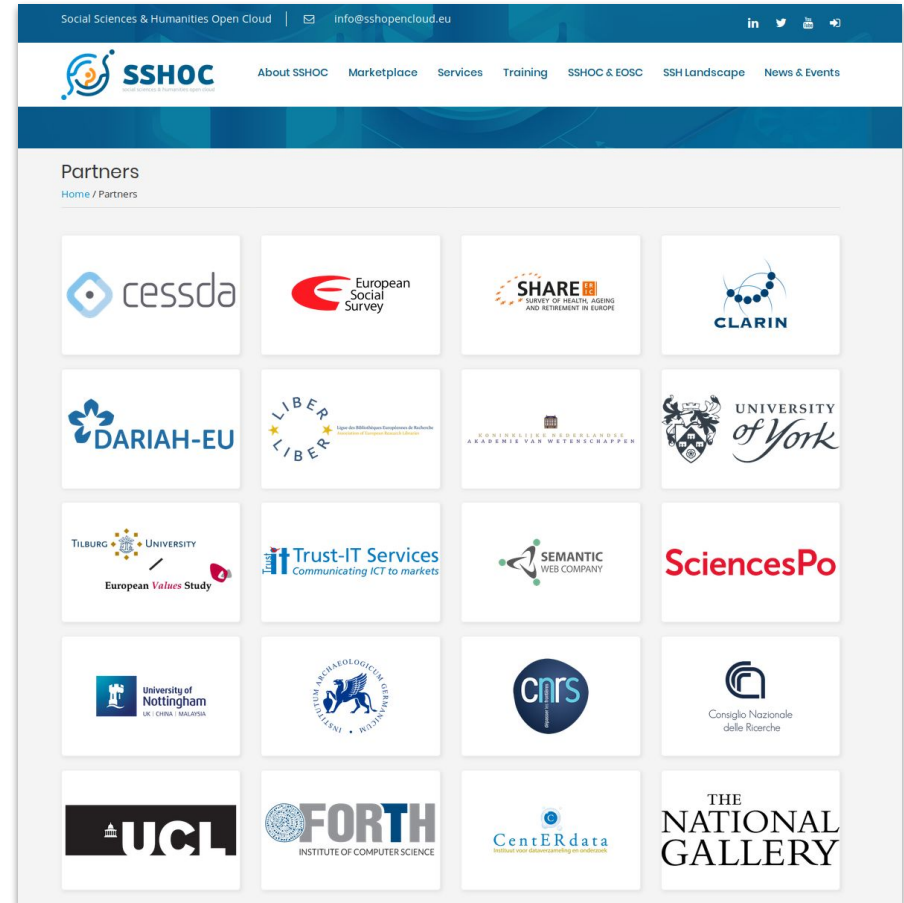
---

# follow up

H2020 **SSHOC** & **ESS-SUSTAIN-2** involvement:

[SSHOC](#) Task 4.1: Conception and development of a sample management system for cross-national web surveys (status: first deliverable submitted).

ESS-SUSTAIN-2 Task 6.2: Build infrastructure for sample management and questionnaire administration for the ESS panel (jan. 2020).



---

# **the software development project**

---

# ELIPSS software suite

Software suite components are all in-house built, except blaise 4 survey platform.

Software suite components:

1. panel management,
2. survey and field management,
3. mobile device and fleet management,
4. Android Mobile application,
5. Blaise questions enhancement.



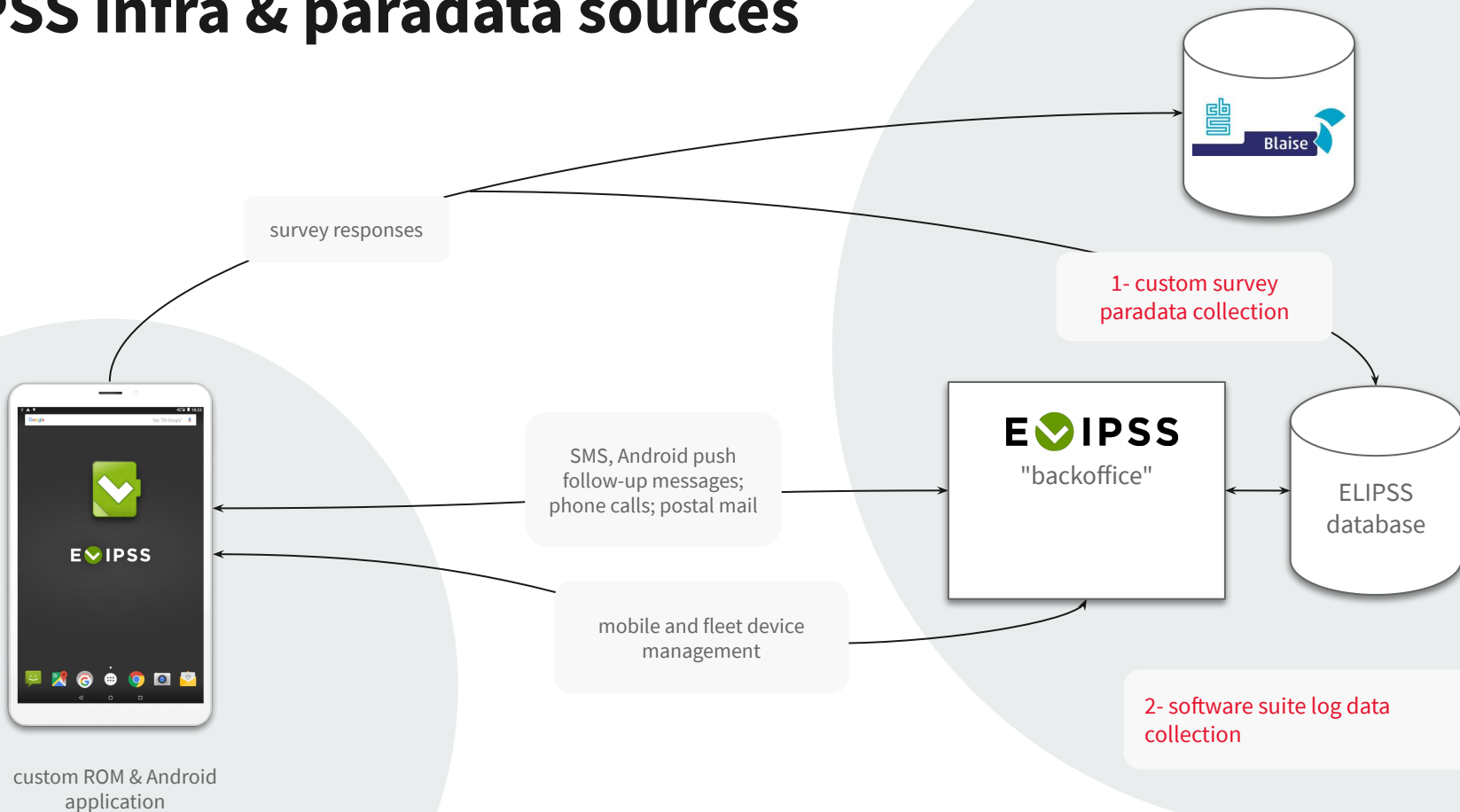


---

# ELIPSS software suite specifications

Let's draft ELIPSS software suite **feature list**, and go for a short hands-on [agile specifications session](#)

# ELIPSS infra & paradata sources



# device and fleet management

1. hardware and software real-time inventory
2. fleet management board: ELIPSS Android application version
3. panelist equipment history
4. fleet management board: hardware stock

**1**

IMEI : 354438070000000 - IMEI2 : 354438070000000

Status: Utilisable

Line number: 077777777

Last network: WIFI (11/03/2019 15:07)

Android version: 5.1

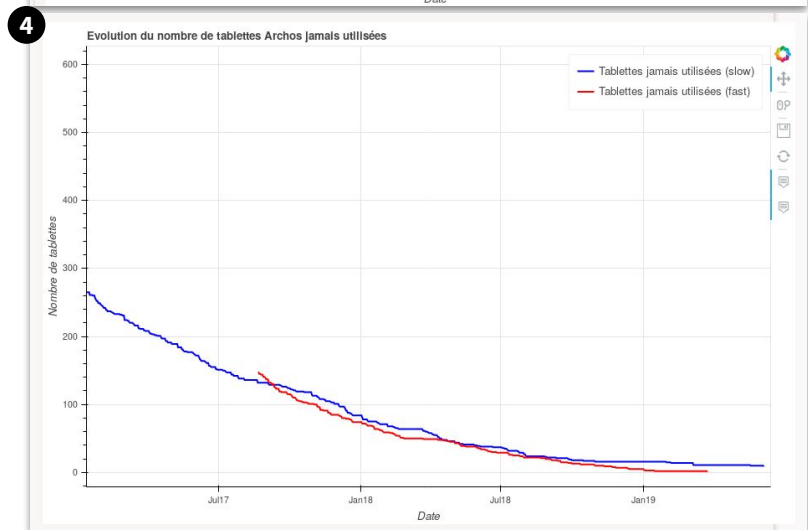
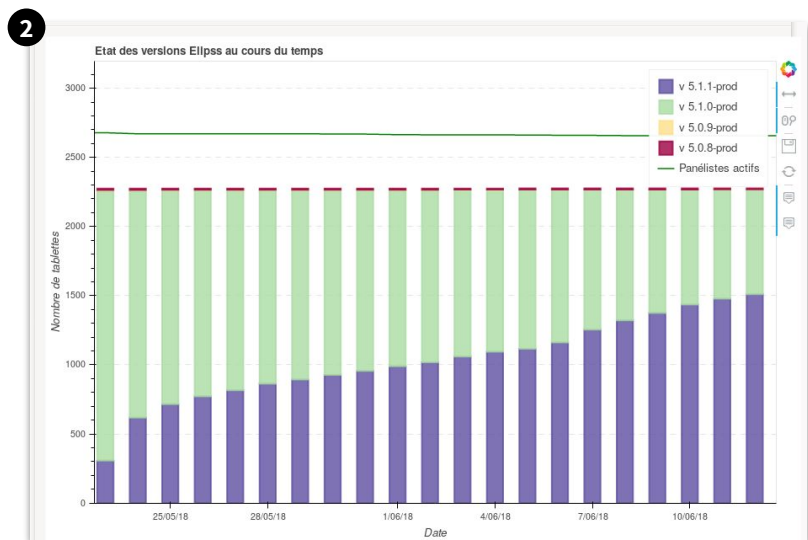
ELIPSS APP

version: 5.1.1-prod

**3**

HISTORIQUE

30/08/2019	en cours...	Utilisable	354438070024470	p1332	Panéliste	v.5.1.1-prod / WIFI (22/10/2019 23:49)
14/11/2018	30/08/2019	Cassée	354438070041219	p4117	Panéliste	v.5.1.1-prod / WIFI (27/07/2019 20:50)
03/01/2018	14/11/2018	Cassée	354438070015890	p1018	Recyclage	v.5.1.1-prod / LTE (06/11/2018 07:25)
03/04/2017	03/01/2018	Cassée	354438070108075	p2612	Recyclage	v.5.0.9-prod / LTE (18/12/2017 16:59)
29/02/2016	03/04/2017	Cassée	354438070035138	p0450	Recyclage	



# panel management

1. custom search engine
2. multi-mode messages
3. multi-agent follow up call
4. integrated follow up postal letter module

**1**

### Rechercher un panéliste

Name

Prénom

Identifiant

Email

Chercher par n° de téléphone

Chercher une tablette

**2** SURVEY: VOTRE PARTICIPATION À ELIPSS

Tous les messages Messages SMS Messages Application Messages Push

2 LOTS

ven 13 Sep 2019, 16:04  
Bonjour, Vous avez jusqu'à dimanche pour compléter le formulaire disponible sur l'application pour continuer à faire partie du panel ELIPSS. Par avance merci !  
(NR + C)\*panel 1 + ED + MO + EP (161) **survey**

ven 28 Jul 2019, 17:14  
Bonjour, le formulaire contenant les nouvelles conditions de participation est disponible jusqu'au 30 juin dans l'application.Par avance merci L'équipe ELIPSS  
NR + C:survey 117 & panel 1+ ED+ JBP + EP + CM + MO (353) **panel**

NOUVEAU MESSAGE

**3** (163) SOMNOLENTS À APPELER (0) SOMNOLENTS APPELÉS

Name: [redacted]  
Prénom: [redacted] **Genevieve Assistant**  
Mobile: [redacted]  
Fixe: [redacted]  
Liens Profil Tickets (0 / 1)

Issue de l'appel: Contact téléphonique

Commentaire: Contact téléphonique

ENREGISTRER

HISTORIQUE DES RELANCES

Date	Type	Issue	Commentaire	Appelant
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

**4** PANELISTES ÉLIGIBLES POUR UNE RELANCE :

ENVOYER LES RELANCES

Relance 1 (219)

Relance 2 (22)

Relance 3 (0)

Relance 4 (13)

VAGUES DE COURRIERS DE RELANCE ENVOYÉS IL Y A MOINS D'UN MOIS : REFRESH INFO

▼	R3 (19)	21 mai 2019 13:50	Malleva	■
▼	R3 (19)	21 mai 2019 13:50	Malleva	■
▼	R4 (14)	21 mai 2019 13:48	Malleva	■
▼	R4 (14)	21 mai 2019 13:48	Manuel	INCONNU ▼
▼	R4 (16)	8 avril 2019 10:05	Manuel	ENVOYE ▼
▼	R4 (16)	8 avril 2019 10:05	Malleva	■
▲	R3 (18)	8 avril 2019 10:03	Malleva	■
Relance_3_08-04-2019 avec AR Date réception Malleva : 8 avril 2019 10:05 Statut Malleva: OK				
REGENERER PDF RECUPERER ORIGINAL				
▼ Liste utilisateurs (18)				
▼	R3 (18)	8 avril 2019 10:03	Malleva	■
▼	R2 (26)	8 avril 2019 10:01	Malleva	■
▼	R1 (47)	8 avril 2019 10:00	Malleva	■

# survey management

1. questionnaire publication module
2. survey stats dashboard

1

QUESTIONNAIRES (85) CREATE A SURVEY

**117. Votre participation à ELIPSS** - 201906\_formulaire\_2 - Groupe : 15  
Nouvelles conditions et formulaire de participation

1831 sélectionnés (changer) [Publiée \(dépublier\)](#) Le: 25 Juin 2019 Terrain: 25 Juin 2019 - 15 Sept 2019

**116. Votre participation à ELIPSS** - 201906\_formulaire - Groupe : 15  
Nouvelles conditions et formulaire de participation

2453 sélectionnés (changer) [Brouillon \(publier\)](#) Le: 19 Juin 2019 Terrain: 20 Juin 2019 - 30 Juin 2019

**115. Questions d'actualité après les élections européennes** - 201905\_cee - Groupe : 14  
Durée indicative : 20 minutes

2445 sélectionnés (changer) [Brouillon \(publier\)](#) Le: 27 Mai 2019 Terrain: 27 Mai 2019 - 30 Juin 2019

**114. Culture et personnalité** - 201906\_psysoc - Groupe : 14  
Durée indicative : 20 minutes

2460 sélectionnés (changer) [Publiée \(dépublier\)](#) Le: 21 Mai 2019 Terrain: 23 Mai 2019 - 30 Juin 2019

**113. Naturbanisation** - 201905\_NatInterest - Groupe : 14  
Durée indicative : 10 minutes

2460 sélectionnés (changer) [Publiée \(dépublier\)](#) Le: 21 Mai 2019 Terrain: 23 Mai 2019 - 30 Juin 2019

2

SURVEY: QUESTIONS D'ACTUALITÉ APRÈS LES ÉLECTIONS EUROPÉENNES

Participation Exports

69% 30%

Terminés 1653 / 2365

Commencés 40 / 2365

Non-répondants 672 / 2365

Sélectionnés 2405

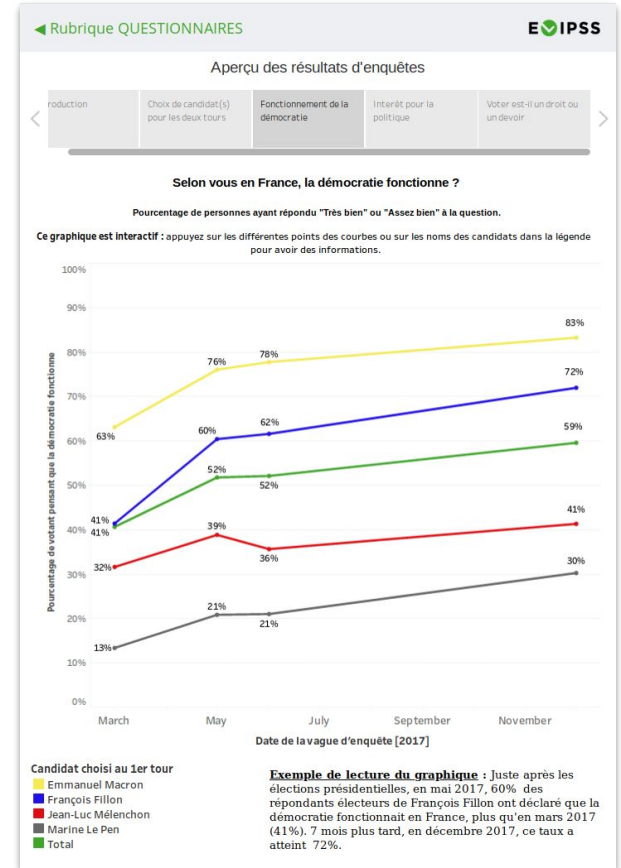
STATISTIQUES DATES  
(27 MAI - 30 JU)

Date	Nombre de réponses
27/05	2600
28/05	1800
29/05	1500
30/05	1200
31/05	1000
1/06	900
2/06	800
3/06	1000
4/06	900
5/06	800
6/06	700
7/06	600
8/06	500
9/06	400
10/06	300
11/06	200
12/06	100
13/06	50
14/06	20
15/06	10
16/06	5
17/06	10
18/06	15
19/06	15
20/06	15
21/06	10
22/06	10
23/06	10
24/06	10
25/06	10
26/06	10
27/06	15
28/06	15
29/06	10
30/06	15

STATISTIQUES HEURES

# other modules

- online recruitment module
  - 2012, user: household
- sample import module
  - 2015, user: polling institute
- online training module
  - 2016, user: call center
- survey statistics
  - Tableau public® integration, 2018

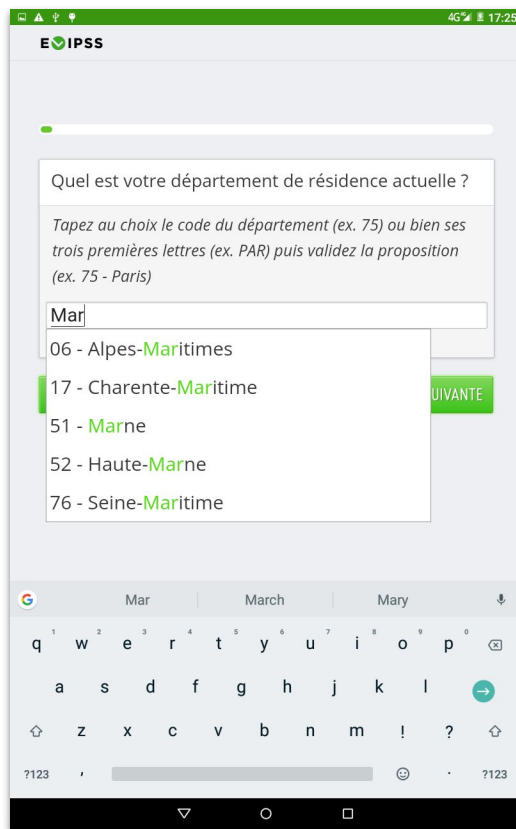
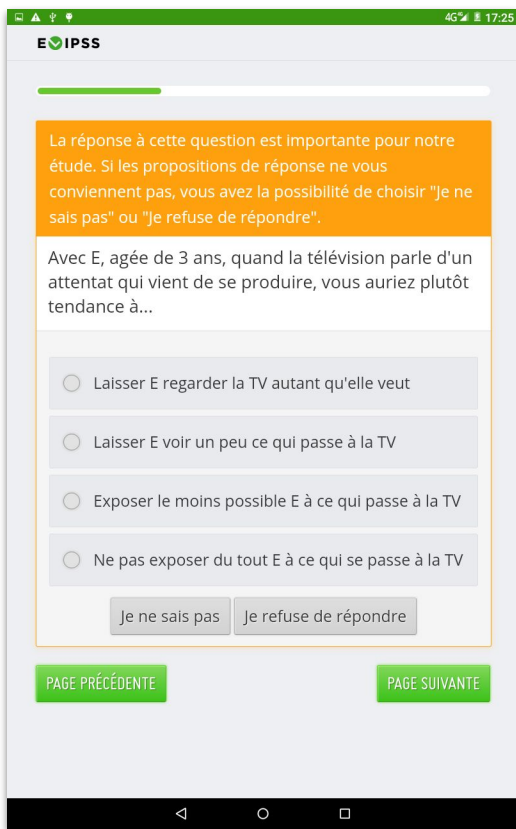


source: integrating Tableau public visualisations into Android ELIPSS application

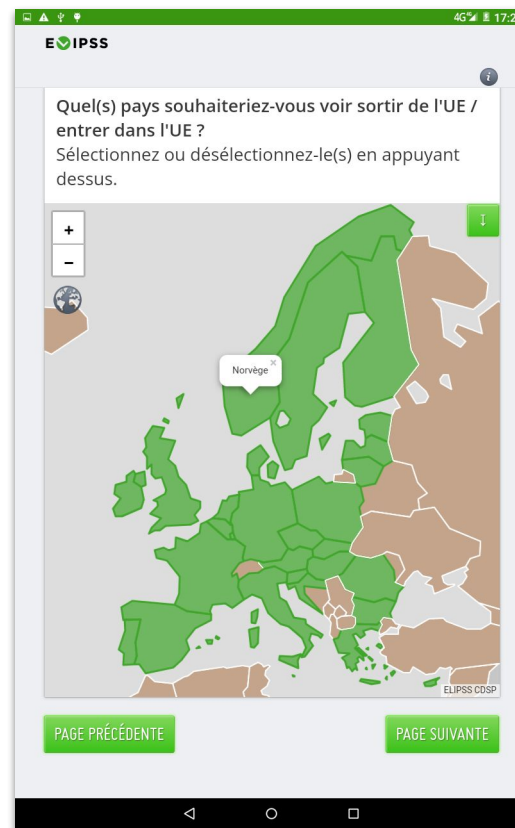
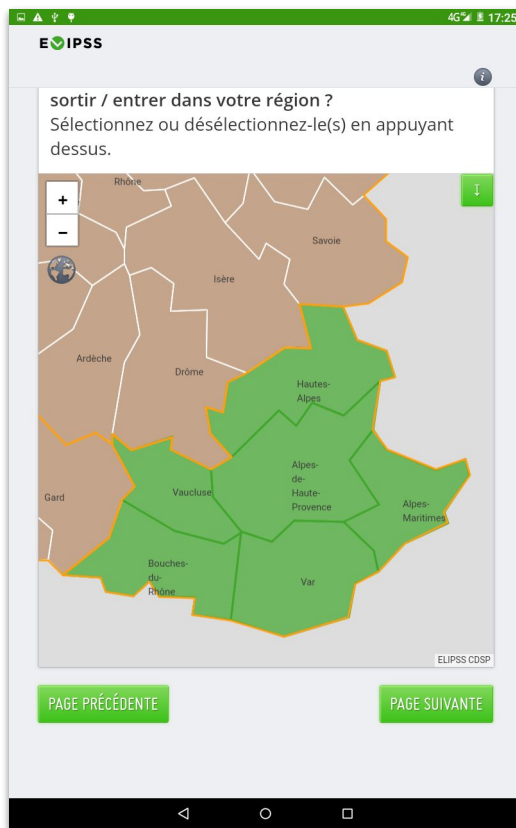
## Android application & questions enhancement



ELIPSS Android app homepage







---

# preserving paradata

---

# dataset design: defining two timelines

We designed a (para)dataset as a double timeline:

- a **panelist timeline:**

Collecting ELIPSS software use **logs** to track panel management events, focusing on **interactions**.

- a **survey timeline:**

Capturing **survey events** through ELIPSS mobile app.

---

# dataset design: defining scope

Dataset granularity:

- various sources and variables: split in 2 distinct datasets

Defining a study time period:

- stabilized sample size
- agreed panel management procedures
- agreed field management procedures

---

# paradata sources: 1. interactions logs

Software suite logs keep records of:

- incoming and outgoing phone calls
- multi-mode batch or individual contacts
  - postal mail
  - SMS
  - Android push messages
  - ELIPSS in-app messages and emails cc
- unavailability time (hardware replacement, illness ...)

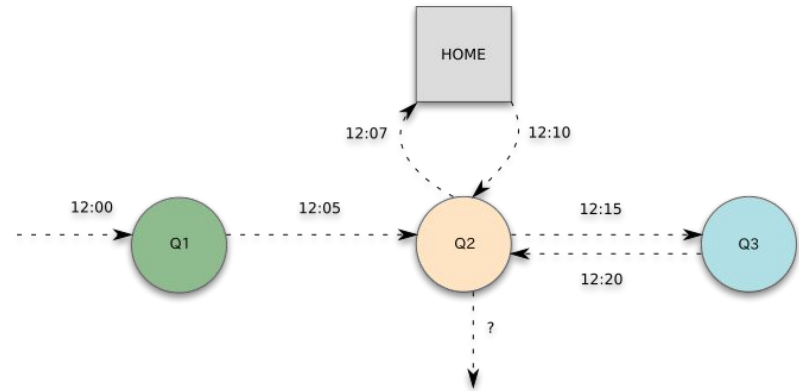


---

## paradata sources: 2. survey events

Ad-hoc survey audit trail files keep record of:

- questions time-stamps
- survey specific customised question events, for example:
  - cumulated video time view
  - number of map area selection tries
- △ "home" selection, device sleep / timeout sequences



*drafting response scenarii documentation, CDSP, 2018*

---

# dataset production expected issues

Raw interactions traces capture an ongoing project. **Processes and methods change over time**, an harmonisation process has been performed to address:

- **catch-all note field**: free text without labels or categories
- **labels**

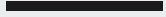
---

# dataset production unexpected issues

ELIPSS infrastructure is made of several components. This may introduce discrepancies.

- timezone issues between survey platform Blaise and ELIPSS web application
  - timestamps value adjustment due to time zone settings
- missing values:
  - missing events due to 4G connection stability issues



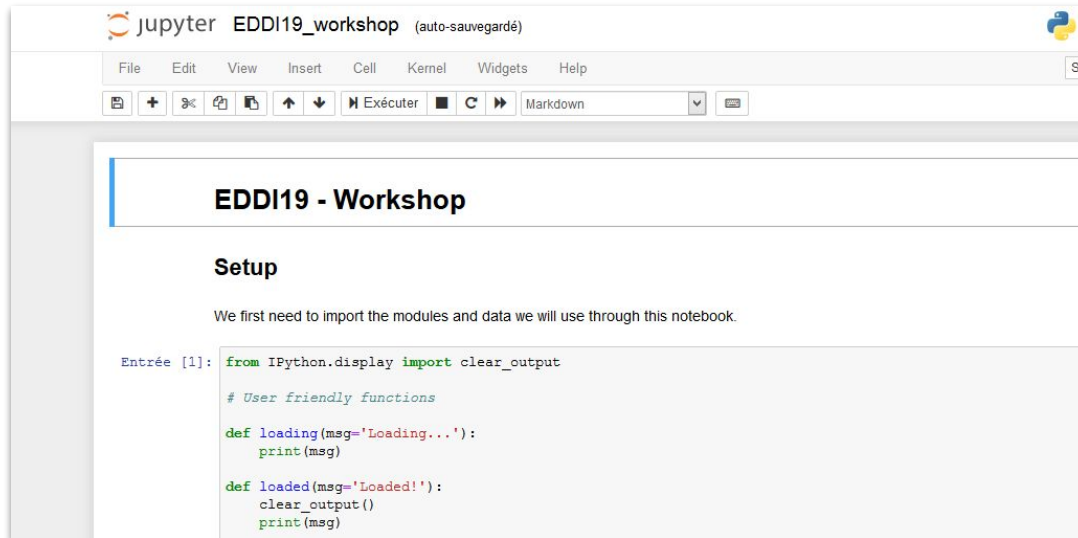


---

# **hands-on data visualisation session**

---

# the (para)dataset: hands-on session



```
jupyter EDDI19_workshop (auto-sauvegardé)
File Edit View Insert Cell Kernel Widgets Help
+ %< %> %< %> Exécuter C Markdown

EDDI19 - Workshop

Setup

We first need to import the modules and data we will use through this notebook.

Entrée [1]: from IPython.display import clear_output

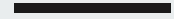
# User friendly functions

def loading(msg='Loading...'):
    print(msg)

def loaded(msg='Loaded!'):
    clear_output()
    print(msg)
```

<https://github.com/CDSP-SCPO/eddi19-workshop>



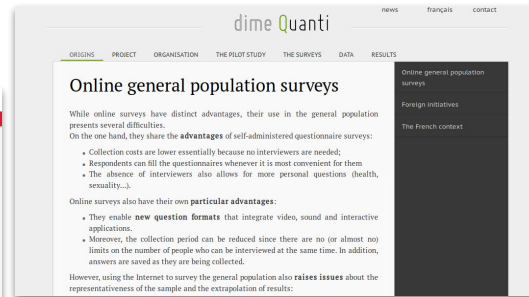
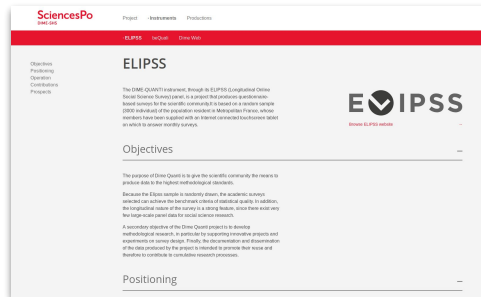


# **hands-on documentation session**

# source material for documenting

This very document can help.

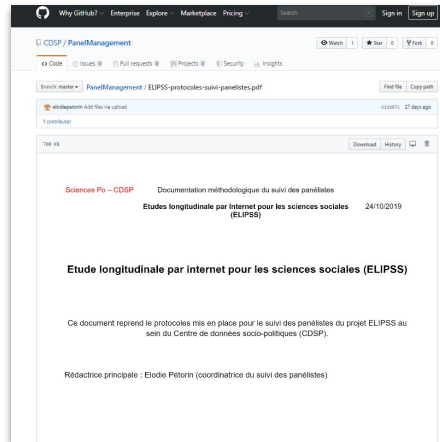
As well as project-related web sites:



---

# source material for documenting

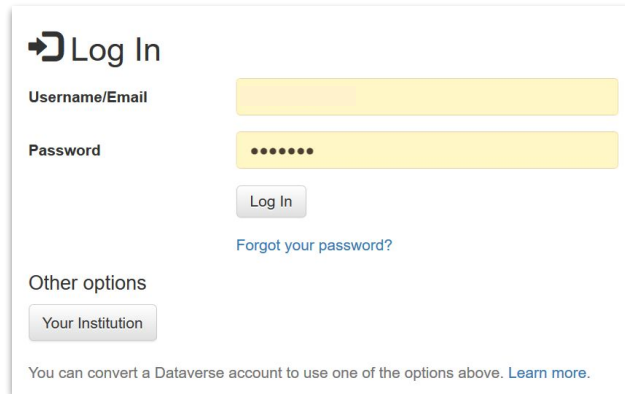
Project related public sources:  
panel and field management  
procedures, on Gitlab (in french)



---

# is [DDI-C] suitable for documenting a paradata set?

Let's try with our tool of choice [Dataverse](#) online web form.



The screenshot shows the Dataverse login interface. It features a 'Log In' header with a right-pointing arrow icon. Below the header are two input fields: 'Username/Email' and 'Password'. The 'Password' field is masked with dots. A 'Log In' button is positioned below the password field. A blue link for 'Forgot your password?' is located below the 'Log In' button. Under the heading 'Other options', there is a 'Your Institution' button. At the bottom, a note states: 'You can convert a Dataverse account to use one of the options above. [Learn more.](#)'

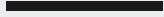


---

# specific issues while documenting a paradata set?

Documenting such a para-dataset raises specific issues.

- documenting funding,
- documenting project,
- documenting instrument
  - how to document an infrastructure
  - Software Management Plan



**thank you!**