



Workshop: Let's take a look at a DDI-documented paradata set 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 paradata set 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-02873594>

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

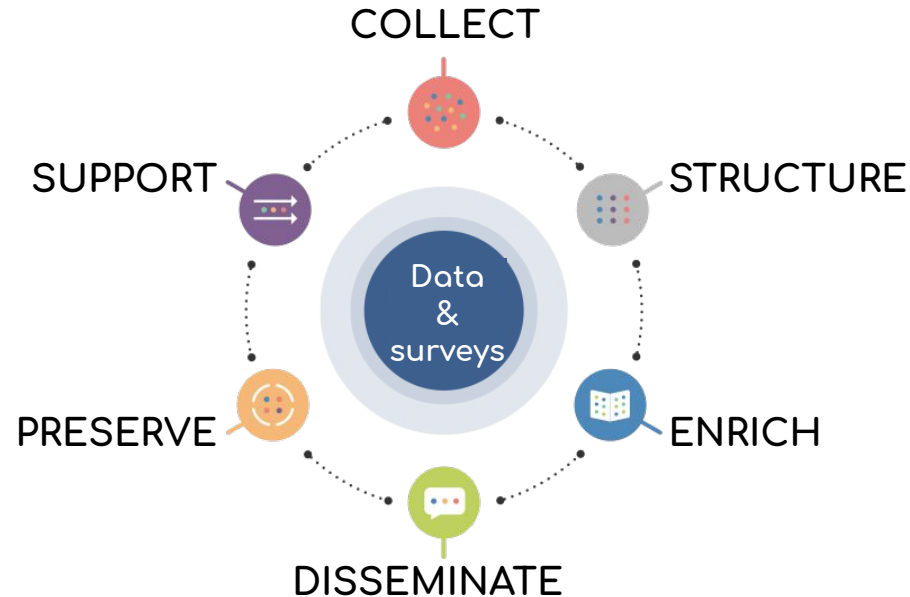
a DDI-documented paradata set
collected by an in-house panel
information system!

[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: [Amédé](#), 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:

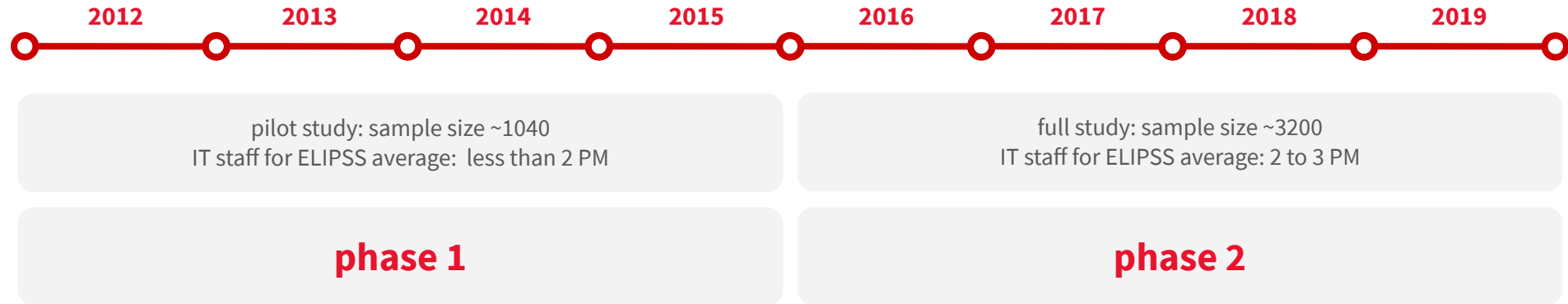
- web developers, web designer
- system administrator
- software project manager

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.

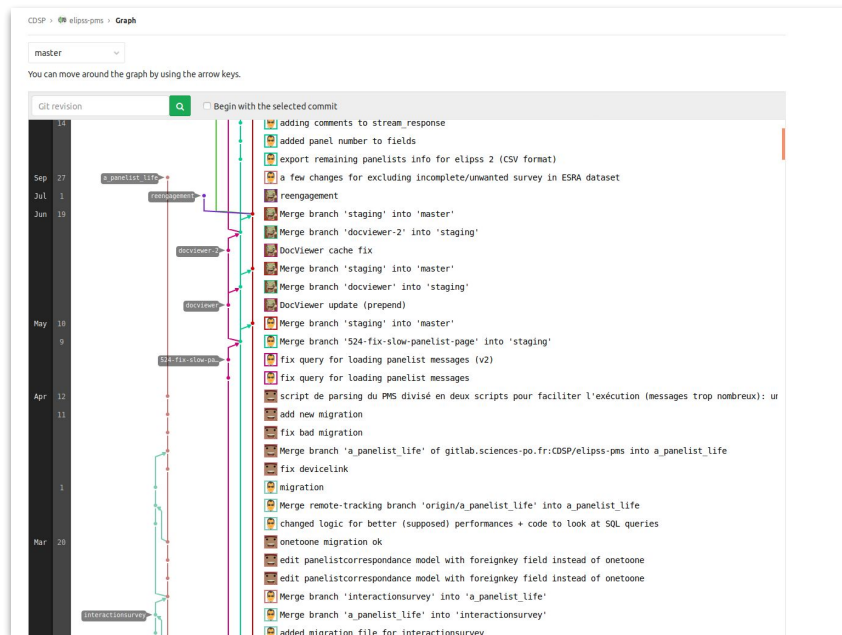


methods and tools: phase 2

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 website of the SciencesPo Center for Socio-Political Data (CDSP). The header includes navigation links for STUDENTS, FACULTY, BUSINESSES, ALUMNI & DONORS, and RESOURCES, along with social media icons and a language selector for FRANÇAIS. The main navigation bar lists THE CDSP, EXPERTISE, PROJECTS, ONLINE RESOURCES (highlighted), DATA DOWNLOAD, and DATA DEPOSIT. Below this, a breadcrumb trail shows 'Home > Resources'. The left sidebar contains a 'Filter by' section with options to reverse the order of items, filter by Type (Qualitative survey /0, Quantitative survey /53), and filter by Themes (Administration /3, Culture /14, Education /0, Elected representatives /6, Elections /1, ELIPSS /53, Environment /10, European Union /2, Everyday life /1). The main content area shows '53 online resources' and lists several surveys: '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 entry is followed by a link to the survey details.

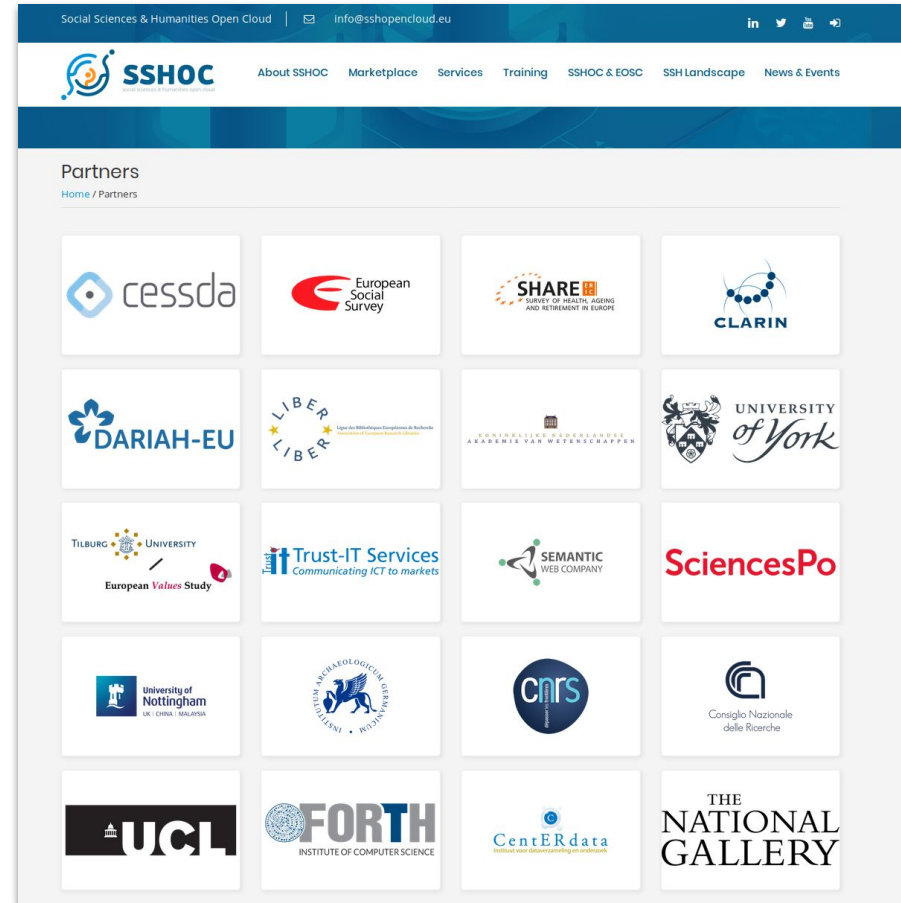
source: CDSP, online resources page, website (2019/11)

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:

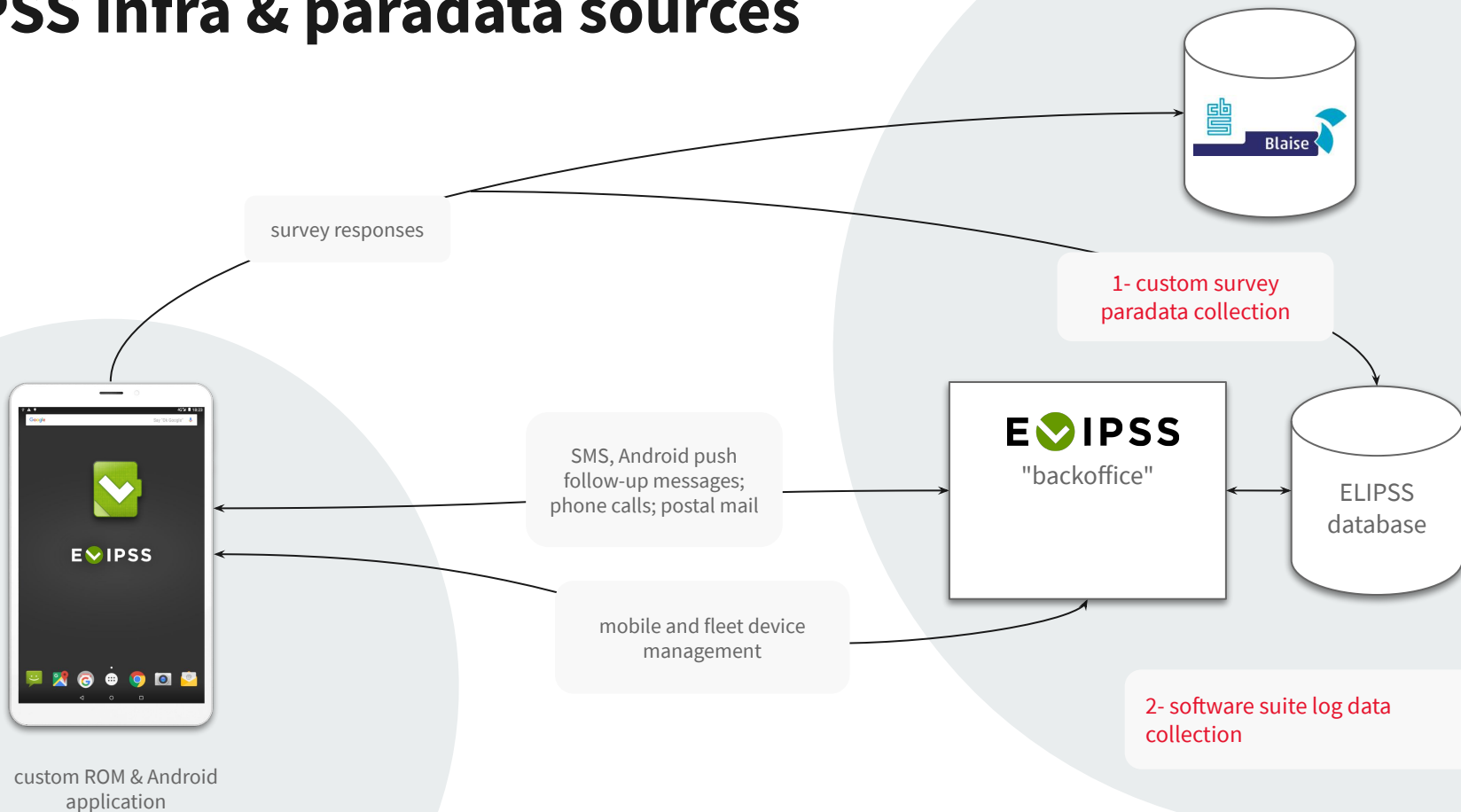
1. panel management,
2. survey and field management,
3. mobile device and fleet management,
4. Android Mobile application,
5. Blaise questions enhancement.



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

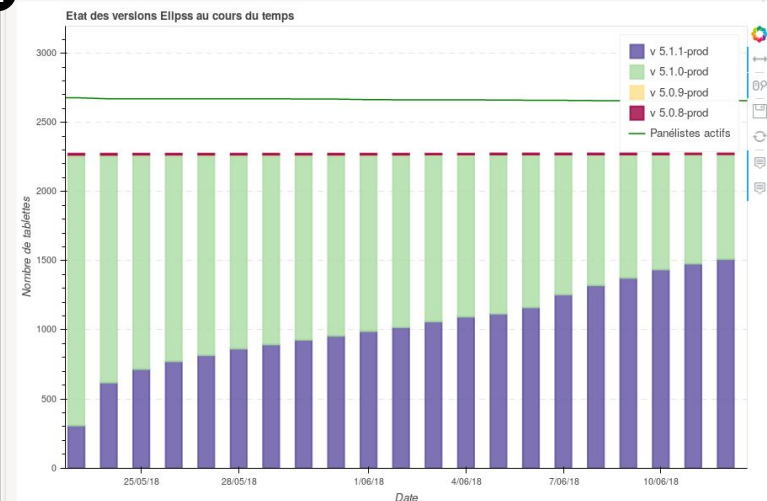
 version : 5.1.1-prod

3

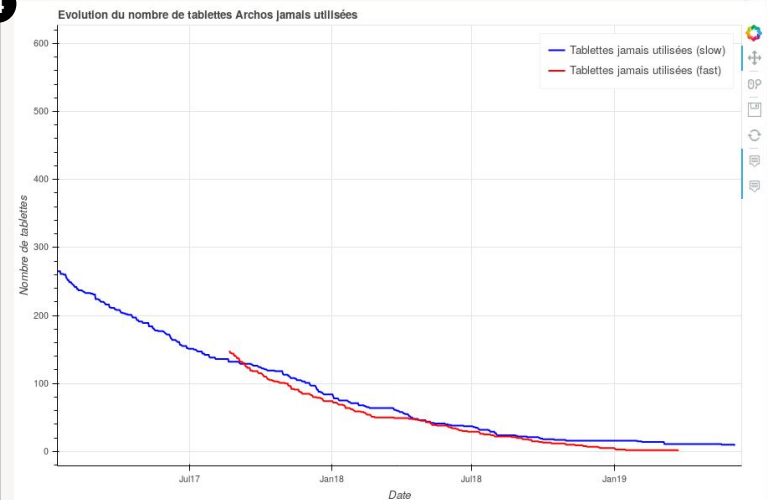
HISTORIQUE

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	

2



4



panel management

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

2

SURVEY: VOTRE PARTICIPATION À ELIPSS

NOUVEAU MESSAGE

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
survey (161)

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
panel (353)

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 Maileva ☒

▼ R3 (19) 21 mai 2019 13:50 Maileva ☒

▼ R4 (14) 21 mai 2019 13:48 Maileva ☒

▼ 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 Maileva ☒

▲ R3 (18) 8 avril 2019 10:03 Maileva ☒
Relance_3_08-04-2019 avec AR
Date réception Maileva : 8 avril 2019 10:05
Statut Maileva: OK
RECUPERER PDF RECUPERER ORIGINAL

▼ Liste utilisateurs (18)

▼ R3 (18) 8 avril 2019 10:03 Maileva ☒

▼ R2 (26) 8 avril 2019 10:01 Maileva ☒

▼ R1 (47) 8 avril 2019 10:00 Maileva ☒

3

(163) SOMNOLENTS À APPELER (0) SOMNOLENTS APPELÉS

Genevieve Assistant

Mobile:
Fixe:
Liens Profil Tickets (0 / 1)

Issue de l'appel:
Contact téléphonique ▼

Commentaire :
Contact téléphonique : |

ENREGISTRER

HISTORIQUE DES RELANCES

Date	Type	Issue	Commentaire	Appelant
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

1

Rechercher un paneliste

Name

Prénom

Identifiant

Email

Chercher par n° de téléphone

Chercher une tablette

19

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 (dépúblier) 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_psycosoc - Groupe : 14
Durée indicative : 20 minutes
2460 sélectionnés (changer) Publiée (dépúblier) 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 (dépúblier) Le : 21 Mai 2019 Terrain : 23 Mai 2019 - 30 Juin 2019

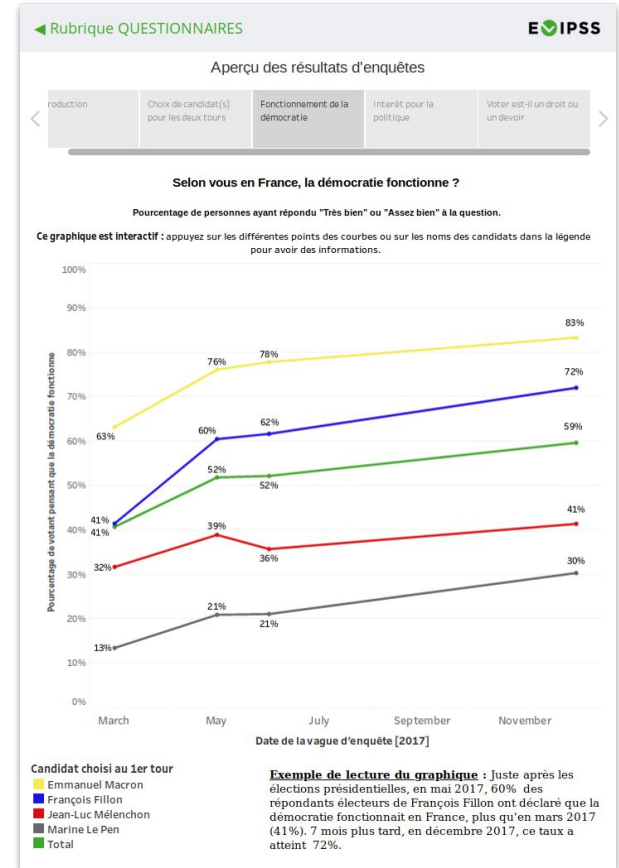
2

SURVEY: QUESTIONS D'ACTUALITÉ APRÈS LES ÉLECTIONS EUROPÉENNES



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

4G 17:25

EIPSS

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

Avec E, âgée de 3 ans, quand la télévision parle d'un attentat qui vient de se produire, vous auriez plutôt tendance à...

☐ Laisser E regarder la TV autant qu'elle veut

☐ Laisser E voir un peu ce qui passe à la TV

☐ Exposer le moins possible E à ce qui passe à la TV

☐ Ne pas exposer du tout E à ce qui se passe à la TV

Je ne sais pas Je refuse de répondre

PAGE PRÉCÉDENTE PAGE SUIVANTE

4G 17:25

EIPSS

Quel est votre département de résidence actuelle ?

Tapez au choix le code du département (ex. 75) ou bien ses trois premières lettres (ex. PAR) puis validez la proposition (ex. 75 - Paris)

Mar

06 - Alpes-Maritimes

17 - Charente-Maritime

51 - Marne

52 - Haute-Marne

76 - Seine-Maritime

SUIVANTE

Mar March Mary

q w e r t y u i o p

a s d f g h j k l

z x c v b n m ! ?

123 , 123

4G 17:22

EIPSS

Habituellement, pour vous informer sur l'actualité, quel(s) moyen(s) utilisez-vous de préférence ? (Trois réponses possibles)

Classez les réponses par ordre d'importance, en commençant par le moyen le plus important (1 pour le plus important, 3 pour le moins important).

3 ☒ La télévision

☐ La radio

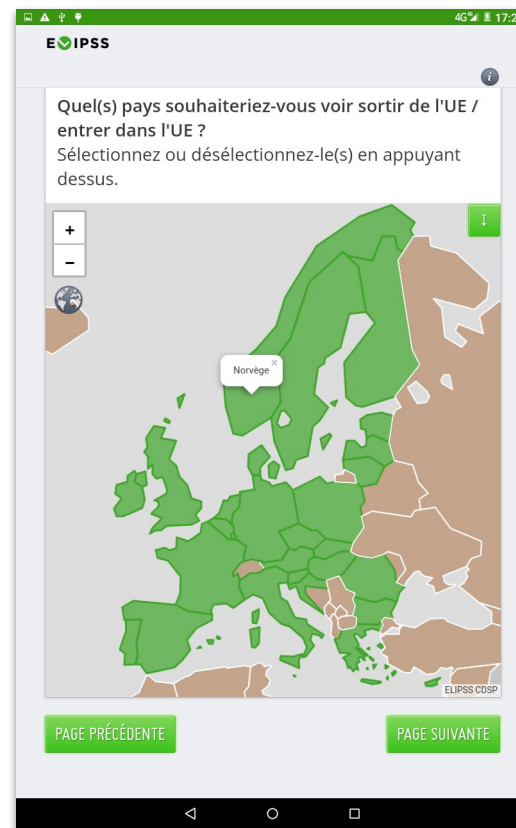
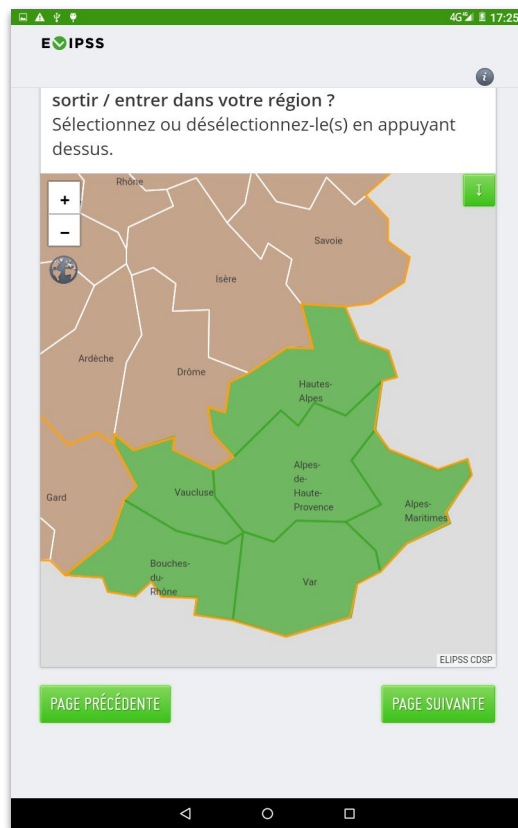
2 ☒ La presse écrite imprimée

☐ Les sites de médias sur Internet ou leur compte Twitter

1 ☒ D'autres sites Internet

☐ Les réseaux sociaux (Facebook, Twitter, etc.)

PAGE PRÉCÉDENTE PAGE SUIVANTE



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:

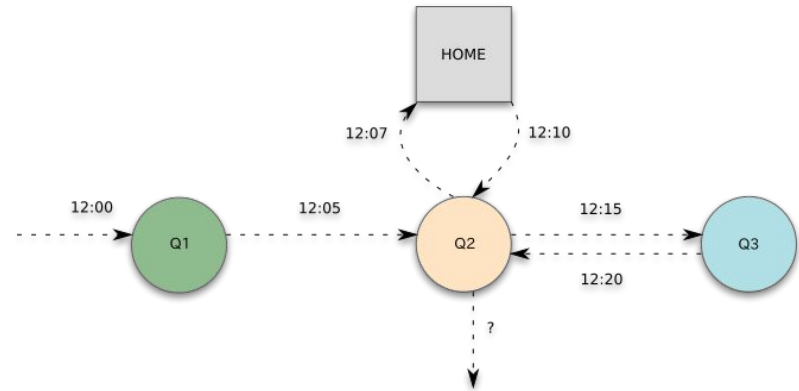
- incoming and outgoing phone calls
- multi-mode batch or individual contacts
 - postal mail
 - SMS
 - Android push messages
 - ELIPSS in-app messages and emails cc
- unavailability time (hardware replacement, illness ...)



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.

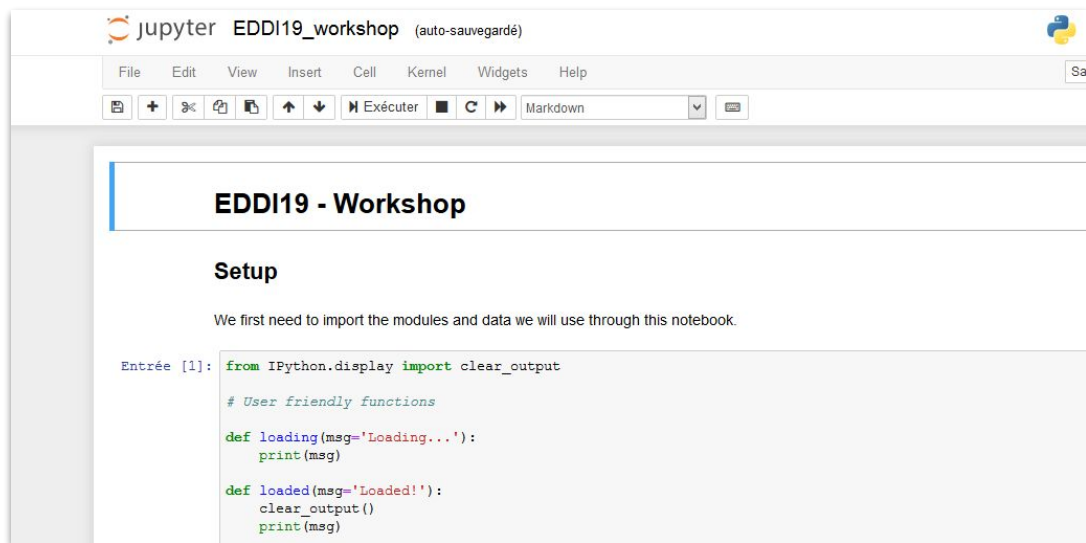
- timezone issues between survey platform Blaise and ELIPSS web application
 - timestamps value adjustment due to time zone settings
- missing values:
 - missing events due to 4G connection stability issues

—



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

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

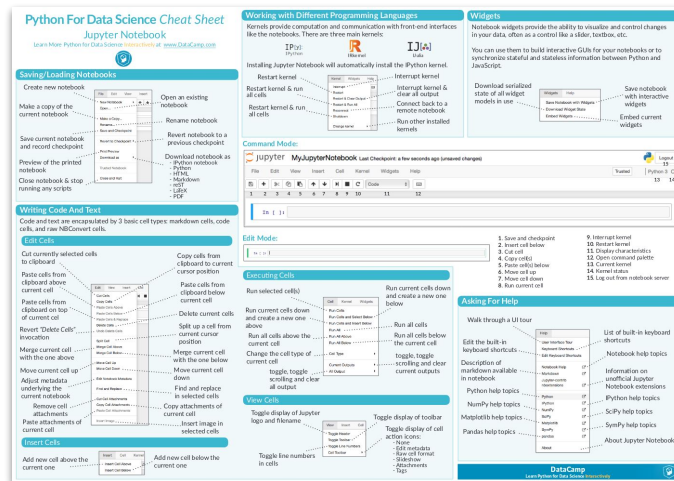
# User friendly functions

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

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

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

Jupyter cheat sheet



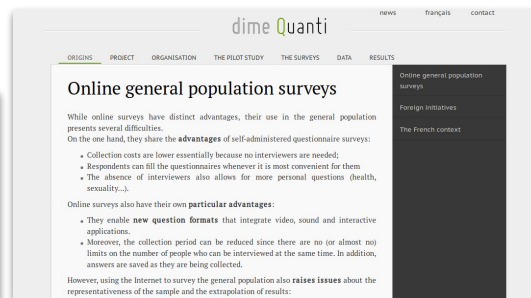
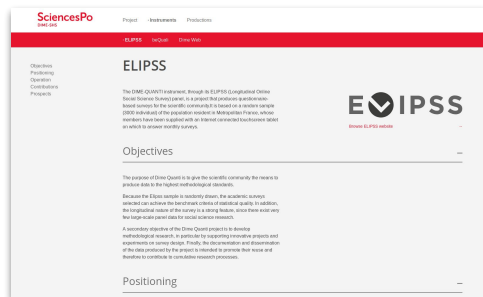
source: <https://www.datacamp.com/community/tutorials/tutorial-jupyter-notebook>

hands-on documentation session

source material for documenting

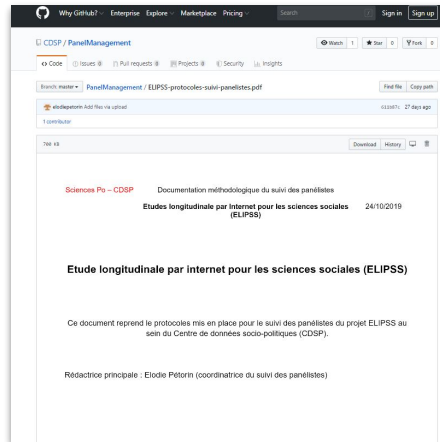
This very document can help.

As well as project-related web sites:



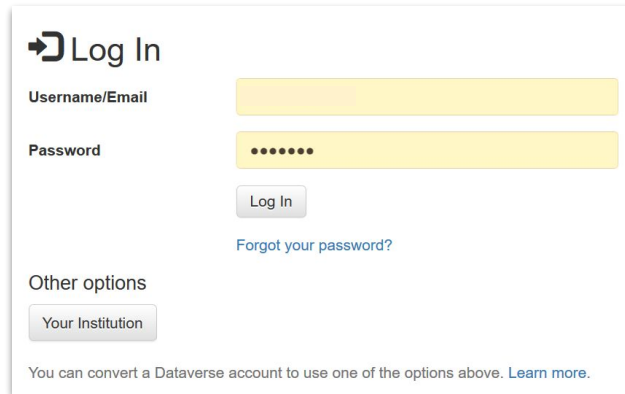
source material for documenting

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 image shows a screenshot of the Dataverse 'Log In' web form. The form is titled 'Log In' with a key icon. It contains two input fields: 'Username/Email' and 'Password'. The 'Password' field is masked with dots. Below the password field is a 'Log In' button. To the right of the 'Log In' button is a link that says 'Forgot your password?'. Below these fields is a section titled 'Other options' with a button labeled 'Your Institution'. At the bottom of the form, there is a note: '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!