



HAL
open science

Preservation workflow of the French probabilistic ELIPSS panel at the Center for Socio-Political Data

Alina Danciu, Emmanuelle Duwez, Mathieu Olivier

► **To cite this version:**

Alina Danciu, Emmanuelle Duwez, Mathieu Olivier. Preservation workflow of the French probabilistic ELIPSS panel at the Center for Socio-Political Data. Workflow Webinar Series 2024, Digital Preservation Coalition, Feb 2024, Online, France. <hal-04514632>

HAL Id: hal-04514632

<https://sciencespo.hal.science/hal-04514632v1>

Submitted on 21 Mar 2024

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.



Distributed under a Creative Commons CC BY-SA 4.0 - Attribution - ShareAlike - International License

Preservation workflow of the French probabilistic ELIPSS panel at the Center for Socio-Political Data

Digital Preservation Workflow Webinars 2024 - Episode 1

Alina Danciu, Emmanuelle Duwez, Mathieu Olivier
Sciences Po, Center for Socio-Political Data, CNRS (CDSP)

CENTER FOR SOCIO-POLITICAL DATA



- CDSP, one of the 11 research centers of Sciences Po is jointly supported with financial and human means by both Sciences Po and the CNRS.
- CDSP shares around 400 datasets on its Banque de données, 100 being collected in the ELIPSS panel CDSP is responsible for
- CDSP has received the CoreTrustSeal in 2023

CDSP'S PRESERVATION MISSION

CDSP's legal statuses : "To insure functions of data dissemination and data preservation"

(in French "Elle assure les fonctions de diffusion et d'archivage de données").

<https://cdsp.sciences-po.fr/fr/le-cdsp/presentation/>

CDSP'S DESIGNATED COMMUNITY

- CDSP's designated community mainly includes **French researchers and students from the social sciences academic community** (Sociology, Political Science, Economics, etc.).
- Around a quarter of our data users are non-French speaking
- Translation of our study-level metadata

DATA: THE ELIPSS PANEL (1)

A high-quality probability online panel for the scientific community since 2012

- 2012: initial recruitment round (1,039 participants)
- 2016: expansion recruitment round (817 first round + 2,514 additional participants)
- 2020: expansion recruitment round (1,391 first rounds + 890 additional participants)
- 2023: expansion recruitment round (1,736 first rounds + 905 additional participants)

INSEE recruited a random sample of individuals aged over 18 with a fixed address in metropolitan France who can read French

DATA: THE ELIPSS PANEL (2)

A large production...

- Cross-sectional or longitudinal projects with no thematic restrictions
- Monthly survey, 10 waves per year
- 9 waves designed by research applicants; 1 wave for the ELIPSS core questionnaire
- 20 minutes duration per wave

... to data dissemination for research

DATA: THE ELIPSS PANEL (3)

The life cycle of an ELIPSS survey



IN A FEW WORDS...

- More than 100 surveys, available on The *Banque de données of the CDSP* (data.sciencespo.fr)
- With few exceptions, anonymised data is shared to the scientific community
- The two standards we use for data sharing and management are OAIS (Open Archival Information System) and DDI (Data Documentation Initiative)

PRESERVATION ACTIONS

- Data and metadata preserved in non proprietary formats (CSV, XML) and data is managed using R, Nesstar and Dataverse
- The very complete documentation goes up to the level of the variable:

- Question text
- Universe
- Interviewer instructions
- Response domains
- Notes



DATA DOCUMENTATION INITIATIVE

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks.

<https://ddialliance.org/>

THE IMPORTANCE OF METADATA

- The ELIPSS surveys are documented using DDI-Codebook
- The entire data lifecycle is managed by the CDSP, so we reuse as much metadata as possible (example : when documenting data, we reuse responses from the programming stage).
- Using open standards like DDI allows us to preserve data and metadata

QUESTIONS ?

alina.danciu@sciencespo.fr

emmanuelle.duwez@sciencespo.fr

mathieu.olivier@sciencespo.fr

SOURCES

Matthieu Olivier, Marta Veljkovic, Andrew Zola, Ettore Recchi. Before, During and After the Deluge: Monitoring the French Resident Population's Response to the Covid-19 Pandemic with a Probability-Based Longitudinal Survey. 9th conference of the European Survey Research Association (ESRA), Jul 2023, Milan, Italy. <hal-04266079>

Emmanuelle Duwez. ELIPSS, a French Internet probability-based panel. 10 years of challenges. 9th conference of the European Survey Research Association (ESRA), Jul 2023, Milan, Italy. fhal04228822f