The Eurostat simplification initiative and its implications for research

Mathilde Godard

CNRS, Dauphine-PSL

ESAC-NUC Conference, October 20, 2026

Short Bio

- **Mathilde Godard**, CNRS, LEDa, Université Paris Dauphine, PSL.
- I'm an applied microeconomist with interests in health, labor and public economics.
- Researchers' representative, CNIS (French National User Council).

The Eurostat simplification initiative and its implications for research

1. **Reduce the burden of data collection**
2. **Modernise statistical production**, incl. make greater use of existing administrative, privately-held or web data.
3. **Reassess statistical reporting requirements**, incl. the possible discontinuation of some surveys.

With important implications for research...

Key issues for research

- **Information content:** administrative/web data may not capture the same information as surveys. Alternative data sources should be adequate.
- **Data access:** researchers need access to the administrative data if they replace discontinued surveys.
- **Researcher consultation:** researchers should continue to be involved in shaping data collection.
- **Cost and complexity**

Plan of the talk

1. Taking a step back : from design to organic data.
2. Implications of the "simplification" process for research.
3. Focus on a case study : discontinuation of the "Formation et Qualification Professionnelle" survey in France.

From design to organic data

- A shift from design data (collected for a specific statistical or research purpose) to organic data (generated as a by-product of systems that automatically record transactions and other activities in society – the web, an administration, a business etc.).
- Current rise in data due to increase in organic data.
- Whether the latter may be converted into relevant information is not guaranteed (Grove, 2011).
- Some examples in organic data in research :
  - Data from the web : JOCAS (Job Offers Collection and Analysis System), by DARES.
  - The rise of administrative data.

Web data & modern web-scraping tools

How to collect data on job vacancies?

- **Previously**
  - Reliance on traditional administrative data from the Public Employment Service (PES) – STMT.
  - Covered job vacancies submitted to the PES by employers, rather than the broader online job market.
  - Access to detailed data was more difficult for researchers and generally required a formal agreement with the PES.
- **Now**
  - **JOCAS collects online job postings** daily from a selection of major job websites.
  - Uses **web scraping**, automated text classification and deduplication.
  - **Open access and online computing**
    - Freely accessible to students, researchers and official statisticians via the SSPCloud platform.
    - Data hosted remotely; users can run their code directly on the platform, with substantial computing and storage resources.

The rise of admin. data in research

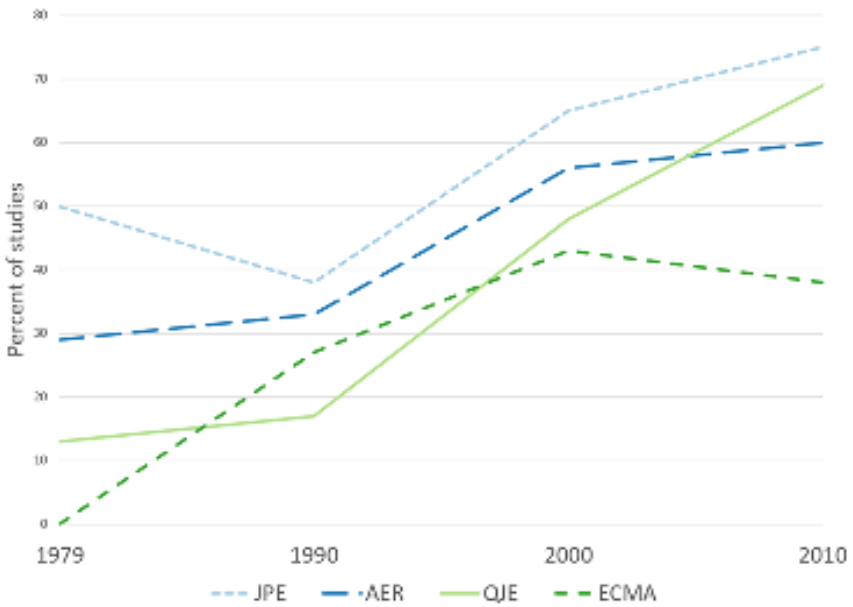

Figure 1: Share of studies conducted in high-income countries that use administrative data, among studies published in the four top US journals in Economics (Journal of Political Economy, American Economic Review, Quarterly Journal of Economics, Econometrica). Source: Chetty (2012).

Very useful properties of admin data

- Measure certain features objectively (contact with a system or provider, e.g. GP consultation).
- Often more reliable than self-reported information (e.g. nb of GP consultations over the past 12 months).
- Avoid typical biases in surveys : social desirability, recall bias, attrition...
- Broad coverage, routine collection $\Longrightarrow$ huge size, granular in time and space.
- Enables researchers to detect effects that are small or heterogeneous between groups, without loss of statistical power.

Potential challenges from the researcher's point of view

So, what are the potential challenges of statistical producers' increased reliance on administrative data?

1. **Measurement:** Administrative data do not always capture the same concepts as surveys.
2. **Access:** Data users (incl. researchers) may not retain access to administrative data.
3. **Cost:** If they do, they typically incurr a much higher cost (at least in France)
4. **Complexity:** Administrative data are often difficult to navigate and exploit.
5. **Consultation:** Data users should continue to be involved in consultations about statistical data production : from adding survey variables to developing new administrative data linkages.

A case study : the FQP survey 1/3

Discontinuating the FQP survey : the rationale

- **A historical survey:** conducted six times since 1964, most recently in 2014–2015.
- **Declining comparative advantage:**
  - Social mobility and educational attainment can largely be studied using other sources.
  - Training-related questions are increasingly covered by dedicated surveys.
- **Limited use of its distinctive features:**
  - The four main topics were rarely studied jointly, despite being covered by a single survey.
  - Its rich retrospective calendar of professional trajectories remained valuable, but potentially subject to recall bias.

Source: Insee, "Bilan des utilisations passées de l'enquête FQP", May 2024.

A case study : the FQP survey 2/3

Replacing FQP: new data, new challenges

- **A new, hybrid statistical system:**
  - The Labour Force Survey, with selected FQP questions added periodically.
  - Administrative data linked to the survey to reconstruct employment trajectories:
    - employee and self-employed records;
    - benefit records, including unemployment-related spells.
  - Other surveys and longitudinal data sources for education and training.

Sources: Insee (2024); CNIS, Employment, Qualifications and Labour Income Committee, 28 May 2024.

A case study : the FQP survey 3/3

Replacing FQP: consulting data users

- **Researchers were consulted through the CNIS Employment committee.**
  - The Insee proposal already acknowledged that no single alternative source could fully replace FQP.
  - No opposition, with only one suggestion by researchers – which INSEE had already anticipated and addressed.
- **But access and data fragmentation remain concerns:**
  - Multiple access procedures and requests across data providers.
  - Information scattered across databases, rather than available in one linked dataset covering the same individuals.

Sources: Insee (2024); CNIS, Employment, Qualifications and Labour Income Committee, 28 May 2024.

Thanks for your attention!