Accessing Privately-Held Data: The French Experience

Several Business Cases

Insee

Measuring, understanding

OUTLINE OF THE TALK

USE OF SCANNER DATA AT INSEE

TRANSACTION CARD DATA

MOBILE NETWORK OPERATOR DATA

BANK ACCOUNT DATA

NEW BUSINESS CASES AND IDEAS

01 USE OF SCANNER DATA AT INSEE

The introduction of Scanner Data was a major project

- A long process (2010-2020), with methodological, legal and technical aspects
- That came after successful implementations elsewhere
  → *NL (2002), NO (2005), CH (2008), and others*
- In a context of wide international interest
  → *Estat guide (2017); expert group at UNECE (2018)*

Important issues to be addressed:

- Sustainable data access;
- Usability of data streams.

THE LONG ROAD TO ACCESS SCANNER DATA

Problem for accessing the data: scanner data are intangible assets of the companies.

→ *No obligation to give access*

- **2012: contacts with four retailers** (40% of the mass distribution market)

  → *Access on an experimental basis, on the basis of agreements*

- **2016: Amendment of the statistical law**

  → *possibility to make certain PHD transmission mandatory (when replacing an existing survey)*

- **2017: Transmission of scanner data made mandatory**

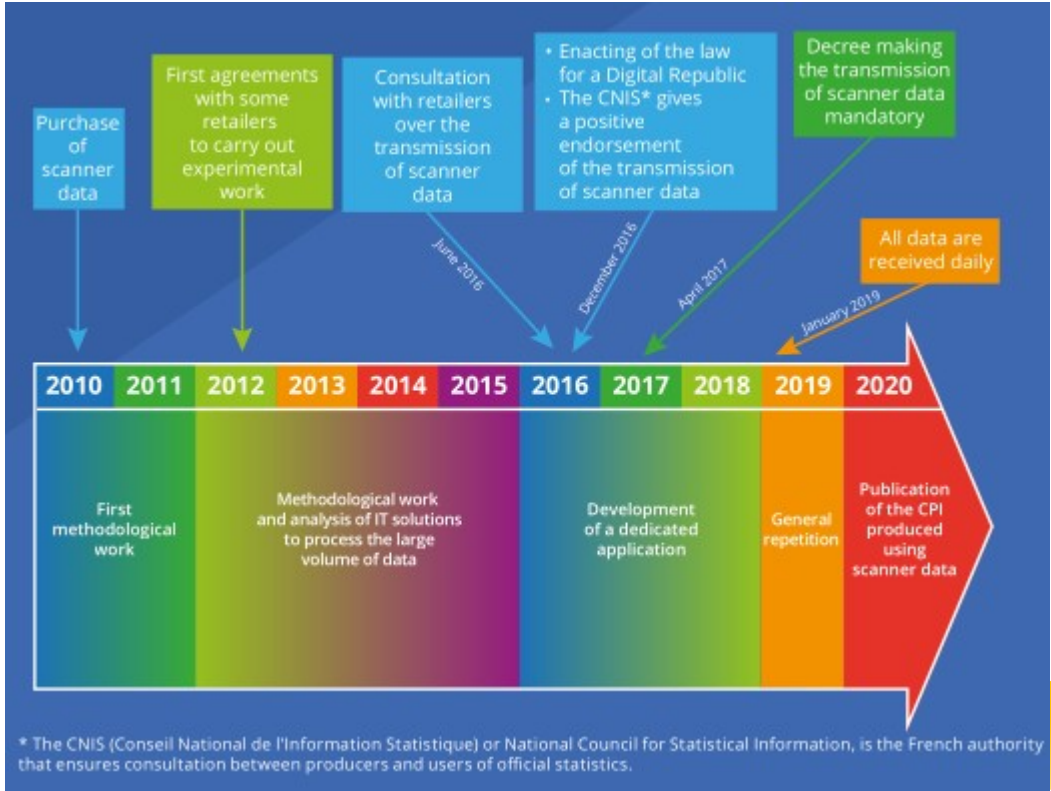
  → *by decree, after positive opinion of the NUC*

- **2019: Full-scale test exercise**

  → *exhaustive data received daily*

- **2020: CPI publication including scanner data**

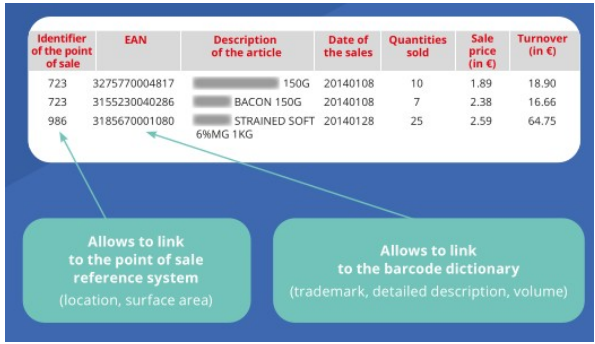
  → **Ready just before the Covid 19 crisis!**



The volume of data calls for adaptation of treatments.

1,7bn records monthly

10% of total CPI weighting (processed food products, cleaning products and hygiene and beauty products)



Data point: barcode, date, quantity, price

Need specific IT system, adjusted to big data

- In the case of France, INSEE receives *exhaustive* data from retailers
- Use of on premise big data infrastructure (Hadoop)

Automation of statistical treatments is required:

- Corrections of marginal modification of prices (eg. change of packaging)
  → source of bias if not treated (“shrinkflation”)
- Replacement of discontinued products in the CPI basket

SCANNER DATA: MAIN TAKEAWAYS

- Number of price records at physical outlets monthly divided by 2 (100k vs. 200k)
- Important increase in quality at detailed level, thanks to use of exhaustive data
- In the amended law: no transmission of microdata to researchers or to authorities
- The law also provides that the data is available free of charge
  → but compensation for the processing
- The dialogue with stakeholders is complex (different professional associations, different processing services to make data available, etc.)

→ Many new fields can be explored: spatial comparisons, new retailers, hard discounters, prices of organic food ...

… but need resources to make all these analyses!

02 TRANSACTION CARD DATA

GENESIS OF THE PARTNERSHIP (2018-2019)

GIE CB

- Main payment card platform for French banks
- 70-80% of card transactions made by residents within the national territory

Initial GIE CB reaction (summer 2019)

- Interest of the technical team (market share tracking at the activity level)
- Blocked by the Board of Directors (confidentiality, image, data usage, cost)

CATALYTIC EFFECT OF THE COVID-19 CRISIS (2020)

Decisive acceleration

- March 2020: Pandemic shock, lockdowns → Urgent need to measure economic impact in real-time
- Traditional sources ineffective in the short term
- GIE CB exceptionally agrees to provide aggregated data (weekly/daily, by MCC)

Immediate use for short-term monitoring of consumption

- Economic situation March 26, 2020: Household consumption estimated 35% below its "normal" level
- Demonstrate the operational value of CB data (responsiveness + sufficient quality … at least to measure a huge shock)
- Establish a relationship of trust between the two institutions

Partnership made durable … but with limitations

- Transmission of aggregated data
- Indirect access (via researchers) to anonymized transactional data (finest level)
- No data on consumption by non-residents
- Payment of data provision costs

Operational result

- Timeliness and precision of short-term indicators (retail turnover)
- Useful during major economic shocks, less reliable under normal conditions
- Interest for research: consumption-related mobility, impact of remote work

03 MOBILE NETWORK OPERATOR DATA

GENESIS OF THE PARTNERSHIP (2016-2019)

Context

- Exploring new sources to measure population presence and mobility
- Need for high-resolution data (time/space)

First Steps (2016-2019)

- Key Partnership: INSEE – Orange Lab (research) – Eurostat (2016-2022)
- Facilitator: methodologists in the 3 institutions
- Detailed (pseudonymized) & antenna-aggregated data; secure access
- Starting with methodological research

Health Emergency

- Critical need to track massive movements (lockdowns)
- Traditional sources inadequate

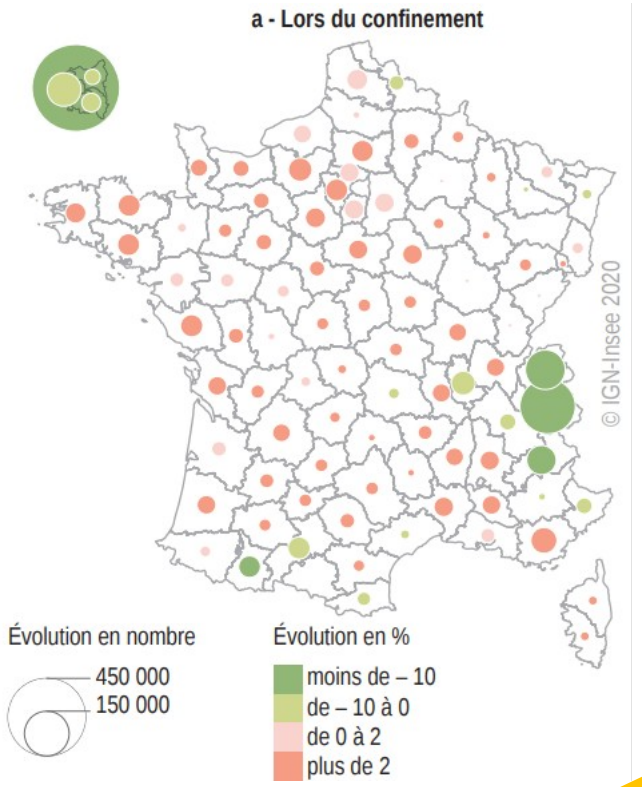
Exceptional Mobilization

- Active solicitation of operators by Insee
- Time-limited partnerships with 3 operators (Orange, SFR, Bouygues)
- Access: aggregated data (counts), adjusted by operators then by Insee
- 3 months free access

Decisive Outcomes

- Near real-time indicators on population distribution

Change in the number of residents



Partnership Reconfiguration

- End of exceptional Covid collaborations. Distanced relations with some operators
- Continuation with Orange, but access to aggregated data for research only

Research Partnership (MOBITIC Project)

- Multi-stakeholder collaboration: INSEE – Gustave Eiffel University – Orange
- Data: Aggregated and adjusted by the operator

Key Issues

- Ensuring sustainable data access
- Continuing data exploration for official statistics purposes

04 BANK ACCOUNT DATA

Partnership with two banks

- Partnerships during the covid-19 crisis with two banks
- Access to large anonymized client panels (300.000 each)
- Data: daily individual account data on income, expenditure and savings

Key findings

- Our traditional measure of the poverty rate: stable during the pandemic
- Users were skeptical
- Thanks to the bank account data: month-end overdrafts have not increased during the pandemic (consumption drop, savings surge, relative income stability)
- Other analyses: fuel price shock impact and distributional effects of subsidies, cross-border shopping

05 NEW BUSINESS CASES AND IDEAS

Another business case:

Smart meters (at the frontier between administrative data and privately-held data)

Some new ideas

- Data from Doctolib, the application for appointments with doctors: delay before the visit
- Data from the applications used by schools: grades, absences,...
- Data from the softwares used by farmers: inputs in the agricultural sector

	Scanner data	Transition card data	Mobile network operator data	Bank account data	Smart meters data
Regular production	Yes	Yes	No	Yes	Yes (work in progress)
Free-of-charge	Yes (by law) But compensation	No	Yes (Covid crisis) then No !	Yes	Yes (administrative data)
Cooperation with researchers	No	Yes	Yes	Yes	
Data sharing	No (by law)	No	No	No	
IT processing of raw data	Yes, on premise	No (except for researchers)	No	No	Yes

Join us on

insee.fr

Jean-Luc Tavernier
Director General
INSEE
europe-international@insee.fr

Insee
Measuring, understanding