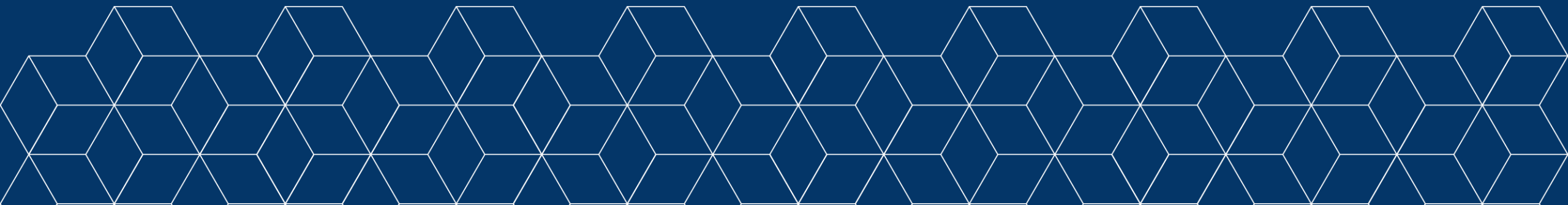


M. BIAGETTI*

Interoperability of Administrative and Survey Data for Policy Evaluation: INAPP's Experience in Economic and Social Statistics

ASA Conference 2026 UniTE, Teramo Italy September 17th, 2026

***INAPP, Rome Italy**



SUMMARY

MOTIVATION

LITERATURE

DEFINITION

METHODOLOGY

DATA SOURCES

RESULTS

SEMANTIC INTEROPERABILITY

LIMITS

CONCLUSIONS AND FURTHER RESEARCH

Identification Needs Infrastructure

40%

of former unemployed get a job six months
after a training program with a course to
follow.

But was it thanks to that training program?

- The identification problem — known, and solvable in principle

We never observe the same unit with and without treatment. A credible counterfactual must be constructed. The toolkit exists: DiD, PSM, RDD, event studies.

- Randomization is largely unavailable: labor policy is designed at distinct LoGs: Central Government, Regions and INPS: there is little room for pilots.
- But the data problem results in a binding constraint.
- Constructing a counterfactual requires linking beneficiaries → firms → outcomes.
- Administrative data are fragmented across archives, formats and institutions. Passive policies lie in central archives; active policies lie in fragmented regional ones.
- Even within one archive, fields are not self-interpreting: “*azienda di riferimento*” means “last employer” for one benefit and “current employer” for another.
- **The method is not the bottleneck. The bottleneck is the information base on which the method must be applied.**



From Fragmentation to Interoperability

THE GAP

- A broad but fragmented information base: census and survey sources produced by different institutions
- Linking beneficiaries, firms, and outcomes remains difficult and costly
- Structural, semantic, and governance obstacles

THE THESIS

The interoperability between administrative and sample data – governed by shared standards and supported by AI – constitutes a critical enabling condition for a rigorous, causal, forward-looking evaluation.



Four Key Strands of Relevant Literature



Causal Inference & Quasi-Experimental Designs

- Roth, Sant'Anna, Bilinski & Poe (2023, J. Econometrics) — DiD synthesis
- de Chaisemartin & D'Haultfœuille (2020, AER; 2023, survey)
- Callaway & Sant'Anna (2021); Sun & Abraham (2021)
- PSM / RDD: Cattaneo & Titiunik (2022); Imbens & Wooldridge (2009)



Administrative data & linked employer–employee

- Abowd, Kramarz & Margolis (1999, Econometrica) — AKM
- Abowd & Kramarz (1999, Handbook of Labor Economics, vol. 3) “The Analysis of Labor Markets Using Matched Employer-Employee Data”
- Card, Heining & Kline (2013, QJE)
- Haltiwanger, Lane & Spletzer (1999, AER) – linked employer-employee data & productivity
- Herzog, Scheuren & Winkler (2007, Wiley) — Data Quality and Record Linkage Techniques



Interoperability & open data

- Hodapp & Hanelt (2022, J. of Information Technology)
- Tangi et al. (2023, JRC); European Interoperability Framework
- Reg. (UE) 2024/903 — Interoperable Europe Act



Labor Market Intelligence & NLP/ML

- Mezzanzanica & Mercorio (2019); Boselli et al. (2017)
- Li et al. (2020, NER) → survey skill extraction (2024)
- Zhang et al. (2022/2023, ESCO); Clavié & Soulié (2023)



What They Give and What They Do Not



Causal Inference & Quasi-Experimental Designs

- Valid estimators for staggered treatments, heterogeneous effects, weak parallel trends
- Nothing about where the variables come from. GIGO



Administrative data & linked employer-employee

- The right unit of analysis: firm $\times$ worker $\times$ time. Firm and worker fixed effects; productivity
- Assumes the linkage already exists. Says little about how to build it across institutions, formats, identifiers



Interoperability & open data

- Legal, organizational and technical architecture for data exchange. Four levels: protocols, formats, services, platforms
- Too generic. Designed for service delivery, not for identification. Silent on semantics and on evaluation use



Labor Market Intelligence & NLP/ML

- Extracts skills, occupations, tasks from job postings, CVs, training offer. Scales coding beyond manual capacity
- Accuracy and transparency are unresolved. Extracts concepts, but does not reconcile them across archives

STRONG ON THEIR OWN TURFS BUT SILENT ON THE INTEROPERABILITY-FOR-EVALUATION



Interoperability and Coding Systems

FOUR TECHNICAL LEVELS WHICH ARE NOT ENOUGH

- **Protocols** — same transport/communication standards (they can physically exchange bytes)
- **Data formats** — shared structure so the receiver can parse the fields
- **Services** — access to functions of other systems (callable functions through APIs)
- **Platforms** — software/hardware environments can collaborate

These four things must match before data can pass between systems, from mechanical to meaningful

SHARED SEMANTIC LAYER

- **CP2021 · ATECO · NACE** (activities and professions)
- **ESCO · ISCO** (skills and jobs)
- **EQF · ISCED** (qualifications and educational level)
- **Labor Atlas (Atlante del Lavoro)** as backbone

Matching format is not matching meaning. Words must mean the same thing across datasets so that they must map to a common reference vocabulary



Pillar I - Empirical Backbone: Counterfactual + DiD

Counterfactual from survey

RIL survey asks firms what they would have done without the incentive and therefore splits them between treated and would-be-untreated firms.

Must Hold
Stated preference consistent with revealed behavior

Check
Pre-treatment behavior of “would-be-untreated” firms must match non-requesters



Comparison group

Similar firms which didn't request any incentive. The key characteristics are usually, the type of management, sector, workforce, degree of innovation, that of export (this is a non-exhaustive list).

Must Hold
Common support, covariate balance

Check
Overlap regions of propensity score
 $|SMD| < 0.1$ after matching



DiD + FE

DiD with usually multiple fixed effects on the integrated longitudinal data (Firms above all)

Must Hold
Parallel trends

Check
No anticipation, placebo treatment date, Event-study leads jointly zero

Cases: Incentives for hiring (2022) and the Youth Guarantee Program (PGG, 2023).



INAPP's Data Assets Overview

ADMINISTRATIVE

- INPS Archives (Beneficiaries database)
- Compulsory Communications (COB)/ SISCO
- ORBIS (*Latu sensu*) — commercial database providing standardized financial statements for corporations
- Same as AIDA for Italy
- ASIA — Register of active firms (ISTAT)

SURVEYS

- RIL — Survey on Firms and Labor
- PLUS — Labor Supply
- ISTAT — Labor Force Survey (ISTAT)
- ICP — Sample Survey on Professions (INAPP+ISTAT)
- + Official Stats

THIS DATA WEALTH IS MADE / CAN BE MADE INTEROPERABLE



The Integrated Employer–Employee Database (Brunetti-Ricci, 2023)

RIL 2018

COB/SISCO

ORBIS

ASIA

*linkage through
tax code→*

Around 13,000

Firms in the final integrated sample.
The unit of analysis is the firm while the outcome is the share of people under 30 among new hired people and labor productivity



Hiring Incentives: Coarse DiD Means

Outcome: share of new hirings on total employment · treated (use incentives) vs control (didn't use incentives) · RIL panel 2010/2015/2018, firms with 9+employees; 2,500 firms

	Pre — Treated		Pre — Control		Post — Treated	Post — Control
	2010	2015	2010	2015	2018	2018
Mean	0.106	0.108	0.112	0.086	0.182	0.114
Std. Dev.	0.177	0.160	0.192	0.174	0.136	0.193
N of obs	297	302	1,699	1,645	309	1,625

The share of new hirings increased from 10.7 to 18.2 with the incentives. Controls remained constant. Consistently, the causal estimated effect is positive and statistically significant and led by the manufacturing sector. (TWFEs)

Source: Brunetti, Martino & Ricci (2022), International Journal of Manpower 43(7), 1646–1669.



Youth Guarantee: Outcome Variables

	GG = 0		GG = 1		Would have hired anyway		Would have hired fewer		Would not have hired	
	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
<30 yrs net flows	-0.014	0.28	0.018	0.14	0.035	0.15	-0.058	0.119	0.021	0.059
Lab prod	10.68	0.76	10.73	0.48	10.80	0.481	10.81	0.468	10.79	0.53
Nobs	12,599		530		394		93		43	

Under 30-year-old net flows slightly positive for treated units (0.018), negative for controls (-0.014). But the causal effect – estimated through OLS and PSM - is negative. YG is not successful in increasing the under-30 share of total new hirings, nor labor productivity.

Source: Brunetti & Ricci (2023), Sinappsi XIII, n.1, pp. 88–100 — data RIL-ORBIS-COB 2018.



Impact of Evaluated Policies

Here the linkage is «easy»
Structured sources linked by a shared key:
the tax code



Incentives to Hirings

- Significant Increase in the share of new hirings in the short run.
- Positive effect led by the manufacturing sector

Brunetti, Martino & Ricci (2022)



Youth Guarantee

- Decreases net youth employment in the short run.
- Also reduces labor productivity
- Unsuccessful.

Brunetti & Ricci (2023)

Opposite outcomes → analysis design and target of measure are fundamental.



**BUT MOST OF THE INFORMATION BASE HAS NO SHARED KEY
AND ISN'T STRUCTURED AT ALL**



Within One Archive

Semantic Treatment of INPS Administrative Archives (De Vincenzi & De Blasio 2017)

17,36 mln

Coarse records (extracted in 2017)

↓ -6,4% *incomplete / duplicates*

16,25 mln

Normalized records (SIP_DATA)

record = treatments, not individuals (around 11,19 mln beneficiaries)

- “Polycentricity” of infos: passive policies → Central Gov’t/INPS (archivio nazionale unico); active policies → fragmented regional archives. Hurdle number 1.
- Normalization: deleting incomplete/invalid/inconsistent/duplicate records.
- Semantic interpretation: even within one archive, the meaning of the data isn't self-evident — one can't interpret a field without knowing the administrative procedure that generated it. For instance, "*azienda di riferimento*" means "last employer" for an unemployment benefit but "current employer" for a CIG spell: the label is identical, the meaning isn't. Hurdle 2.

Eg.: thousand of fixed-term contract teachers with the same employer (CED of the Italian Ministry of Economy and Finance, based in Latina as the NASpl was filled in by taking into account the MEF data center). This is a systematic error that only a qualitative analysis could explain back in 2017.

ALL THIS WAS BASICALLY HANDMADE



SEMANTIC INTEROPERABILITY

Second Pillar: AI-Assisted Semantic Linkage – Across Heterogeneous Sources

SQL

Join between harmonized databases (eg. Accreditalento ↔ Atlante).

API

Machine-readable endpoints to enable tools to communicate (e.g., Atlas ↔ Professions Portal)

NLP / ML

It links concepts extracted from unstructured sources (job postings, CVs, training offer) — Sophia project.

Growing Complexity but AI reduces the use of transcoding tables



Semantic Interoperability: What comes out from Sophia

Extraction of entities from unstructured sources (job postings, CVs, training offer) and cross-domain linking:



Skill mismatch can be at least partially compensated by linking concepts beyond classifications.



- DiD rests on parallel trends, firm FEs, staggered-treatment corrections. If they fail so does identification. We got to be careful of self-selection (see YG examples, where the slight positive effect is driven by firms which would have hired anyway)
- Survey counterfactual: likely perception bias; self-reported responses.
- Administrative linkage: cost, incomplete coverage, unobservable factors.
- Administrative data designed for management purposes → they can't capture all relevant dimensions (*partial dimension-blindness*).



CONCLUSIONS AND FURTHER RESEARCH

- Interoperability is an enabling condition for rigorous, causal, and forward-looking evaluation.
- When adopting an employer–employee approach, outcomes might be opposite: general hiring incentives work has been shown to work, while the YG has not. Consequently, what matters is the design of the measure.
- AI connects heterogeneous semantic domains — but the limits of reliability and transparency must be acknowledged even though it is useful when no shared key exists to link concepts across sources to investigate and solve skill mismatch
- INAPP's profile: an institution that operates with continuity and rigor at the intersection of statistics and AI.



Further Research in Interoperability: An Agenda

FOR INAPP - INTERNALLY

- Interoperability by design & by default (i.e., interoperability from the onset, build new tools and datasets so that linkage is the default rather than an afterthought)
- Shared semantic tier (Atlas as overarching tool, i.e. adopt one common vocabulary)
- AI-assisted record linkage/entity resolution (i.e., ML to match records and resolve entities where no clean shared key exists)
- Privacy by design (i.e., bake data protection in, since linking administrative archives raises confidentiality stakes)

FOR CITIZENS AND OTHER INSTITUTIONS (WHAT INTEROPERABILITY SHOULD DELIVER OUTWARD)

- Interoperability as a Service: open APIs, PDND (i.e. expose the data through machine-readable endpoints on Italy's national digital data platform, so others can plug in.)
- Public semantic hub (eg., a shared reference point connecting professions, skills, qualifications, demand)
- Access to stats in natural language (i.e., let non-experts query the data in plain language which is the true AI/NLP payoff, for the general public).
- Shared standards to onboard external data (i.e., common formats so outside organizations can attach their data to the ecosystem)

From measurement to prediction:

Middle term prediction (2024-2030) by sector, region, profession and educational level can foresee skill mismatches.





THANKS FOR YOUR ATTENTION



www.inapp.org