



LabInn IV Call

---

**Acquisizione e analisi di dati  
provenienti da nuove fonti  
per la costruzione di nuovi  
indicatori sulla capacità di  
innovazione delle unità  
economiche**

Sergio Salamone  
Gianpiero Bianchi  
Francesco Scalfati

---

# Il progetto

---

**L'obiettivo** è produrre un quadro statistico di informazioni dettagliate sulla capacità innovativa delle imprese producendo nuove variabili statistiche basate sulle informazioni dei brevetti delle unità economiche.

**Viene realizzato** attraverso la combinazione di più fonti (big data, registri statistici). La fonte di dati principale utilizzata sono i brevetti depositati presso l'Ufficio Europeo dei Brevetti (EPO).

**Si basa** su tecniche di text mining e machine learning per estrarre e strutturare le informazioni testuali, ed integrarle con i registri statistici.

**L'output** è relativo alla produzione di statistiche su aggregati di unità economiche che brevettano invenzioni, a livello dimensionale, territoriale, settoriale, e l'evoluzione nel tempo del fenomeno.

---

LABINN IV CALL

ROMA, GIUGNO 2023

ACQUISIZIONE E ANALISI  
DI DATI PROVENIENTI DA  
NUOVE FONTI PER LA  
COSTRUZIONE DI NUOVI  
INDICATORI SULLA  
CAPACITÀ DI  
INNOVAZIONE DELLE  
UNITÀ ECONOMICHE

SERGIO SALAMONE

GIANPIERO BIANCHI

FRANCESCO SCALFATI

# Step del processo statistico di produzione

---

**Raccolta dei dati** attraverso l'utilizzo di API per l'estrazione e conversione delle informazioni da file xml in formato testo, la pulizia dei dati estratti e la memorizzazione in file di testo

**Integrazione dei dati** attraverso una procedura di record linkage per abbinare i micro-dati sui brevetti con i dati disponibili nel registro statistico delle imprese (ASIA).

Il processo di abbinamento dei dati prevede le fasi di pre-processing, standardizzazione e cleaning dei dati, l'indicizzazione e individuazione delle possibili coppie di abbinamenti, la classificazione in match, no-match e possibili match, e la valutazione della classificazione.

**Produzione di indicatori di output** che quantificano il valore degli asset intangibili delle imprese relativamente alla proprietà intellettuale e alla creazione di valore economico

---

LABINN IV CALL

ROMA, GIUGNO 2023

ACQUISIZIONE E ANALISI  
DI DATI PROVENIENTI DA  
NUOVE FONTI PER LA  
COSTRUZIONE DI NUOVI  
INDICATORI SULLA  
CAPACITÀ DI  
INNOVAZIONE DELLE  
UNITÀ ECONOMICHE

SERGIO SALAMONE

GIANPIERO BIANCHI

FRANCESCO SCALFATI