

## PROGETTI IV CALL 2023/2024

<b>Titolo progetto</b>	Acquisizione e analisi di dati provenienti da nuove fonti per la costruzione di nuovi indicatori sulla capacità di innovazione delle unità economiche
<b>Descrizione</b>	<p>L'utilizzo dei dati brevettuali è ormai consolidato a livello internazionale per attività di analisi e di ricerca in tema di innovazione, performance e competitività delle imprese. I brevetti sono una proxy della propensione a innovare da parte delle imprese e della potenziale trasformazione degli investimenti in ricerca e sviluppo in prodotti innovativi. Le informazioni sui brevetti risultano fondamentali per individuare traiettorie tecnologiche di sviluppo, innovazioni emergenti e prodotti capaci di colmare nuovi bisogni nella società. Tale quadro conoscitivo amplifica la sua portata informativa se associato al tessuto produttivo che genera tali innovazioni, fornendo rilevanti indicazioni di policy.</p> <p>L'acquisizione dei dati sui brevetti e l'integrazione con quelli sulle unità economiche dei registri statistici valorizzano l'output statistico dell'Istat, ma è anche una risposta alla richiesta crescente da parte di utenti e di policy maker. Una dettagliata documentazione di nuove procedure automatizzate di acquisizione di dati da fonti alternative alle survey e procedure di record linkage per l'integrazione, rappresentano inoltre un bene pubblico di valore per la comunità scientifica, che sempre più spesso si trova ad affrontare l'esigenza di acquisire nuove informazioni complementari o aggiuntive a quelle della statistica ufficiale, e di raccordare informazioni prelevate da fonti diverse e disomogenee.</p>
<b>Obiettivi</b>	<p>L'obiettivo è produrre un quadro statistico di informazioni dettagliate sulla capacità innovativa delle imprese producendo nuove variabili statistiche sugli asset intangibili, brevetti in particolare, delle unità economiche. Con questo progetto si utilizzano i dati sui brevetti dello European Patent Office (EPO), da cui vengono estratte le informazioni anagrafiche del brevetto (applicant, inventor, tipo di brevetto, etc.) e informazioni specifiche sulla singola invenzione, per associare infine il brevetto alla corrispondente unità economica dei registri statistici Istat attraverso procedure di record linkage (non essendo disponibile nelle due fonti EPO e Asia una chiave di aggancio univoca presente in entrambi, come ad esempio il codice fiscale). Il prodotto finale del lavoro è rappresentato da indicatori sulle unità economiche titolari di domande e concessioni di brevetti, aggregate per caratteristiche strutturali (dimensione, attività economica, localizzazione) ed economiche.</p>
<b>Metodologia</b>	<p>Il processo di definizione e costruzione delle informazioni è suddivisa nelle fasi di:</p> <ol style="list-style-type: none"> <li>1) Acquisizione dei dati sulle domande e concessioni di brevetti di fonte Epo. La procedura di raccolta dati è basata su un processo data retrival automatizzato, tramite tecniche di text mining e machine learning per estrarre e strutturare le informazioni testuali, da integrare con i registri statistici. La procedura di acquisizione consente di acquisire i brevetti italiani e le relative informazioni anagrafiche. Dopo una prima fase di sperimentazione è stata quindi definita una nuova procedura che permette di raccogliere i dati sui brevetti EPO attraverso delle API progettate ad hoc.</li> <li>L'acquisizione dei dati sulle domande e concessioni di brevetti ha previsto le seguenti attività: a) analisi delle informazioni sui brevetti Epo, ed individuazione di una procedura automatizzata di data collection; b) raccolta delle informazioni XML; c) verifica e consolidamento delle procedure di raccolta dati. La procedura di data collection definita consente di replicare annualmente la raccolta dei brevetti EPO.</li> <li>2) Integrazione dei dati sui brevetti con quelli delle imprese dei registri statistici Istat tramite record linkage. Il match viene realizzata attraverso la combinazione di più fonti (big data, registri statistici).</li> <li>3) Produzione dell'output di informazioni integrate sui brevetti e relativi indicatori statistici.</li> </ol> <p>La metodologia di acquisizione è stata presentata alla conferenza UNECE Expert Meeting on Statistical Data Collection 2023", con il contributo "Data collection methods to produce new enterprise variables using new data sources"</p> <p><a href="https://unece.org/statistics/documents/2023/07/informal-documents/data-collection-methods-produce-new-enterprise">https://unece.org/statistics/documents/2023/07/informal-documents/data-collection-methods-produce-new-enterprise</a></p>

**Risultati ottenuti**

Il progetto permette di arricchire le informazioni e i processi per la produzione di statistiche ufficiali sui brevetti delle imprese, con un impatto sulla diminuzione del burden sulle imprese, sui tempi e costi di raccolta dati legati alle informazioni brevettuali, rendendo disponibili microdati sulle imprese e sulle invenzioni prodotte.

**Membri del team**

Sergio Salamone, Gianpiero Bianchi, Francesco Scalfati