

Oggetto: Relazione STM Gabrielli - Lussemburgo

Grazie al bando **STM 2016/2 call** ho avuto l'opportunità di effettuare una visita presso Eurostat, *Directorate B: Methodology; corporate statistical and IT services*, come riportato da documentazione allegata, fornita dall'ente ospitante.

Coerentemente con il bando del progetto, durante il mio soggiorno a Eurostat presso il Directorate B, ho approfondito, con il team metodologico di Eurostat, riguardo a come impiegare i dei Big Data nel contesto dello studio degli obiettivi per lo sviluppo sostenibile (SDGs).

Durante questa visita ho avuto l'opportunità di presentare il lavoro svolto durante il percorso di dottorato in tre occasioni pubbliche, a STATEC (Istituto Nazionale di statistica e studi economici di Lussemburgo) (21-11), un seminario interno (24-11) e un seminario presso l'Università del Lussemburgo (29-11). Inoltre ho interagito con diversi gruppo di lavoro di Eurostat che lavorano con i Big Data.

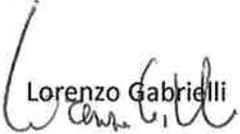
Da parte mia, inoltre, è stato dato un contributo alla preparazione di un abstract per la conferenza CES 2017 riguardante le linee guida per la formazione dei nuovi statistici nell'era della Data Science. L'elaborato tratta le competenze necessarie ai nuovi statistici per affrontare le tematiche tecniche ed etiche nell'uso delle nuove sorgenti dati per produrre statistica ufficiale. Il testo dà particolare importanza alla necessità di orientare la formazione verso un indirizzo multidisciplinare che bilanci accuratamente le hard skill (competenze principalmente tecniche) con le soft skill (capacità comunicative oltre che di curiosità investigativa).

Eurostat è molto interessata all'uso dei Big Data per la creazione di nuovi studi demografici e per prevedere lo sviluppo socio-economico di un territorio. Il periodo di visita è stato molto utile per confrontare i diversi modi di lavorare sui Big Data, da un lato l'approccio guidato dai dati che appartiene alla comunità informatica, dall'altro il modo di lavorare di EUROSTAT proprio della comunità statistica. Risulta chiaro come il disegno metodologico debba cambiare per misurare e valutare l'incertezza e l'errore che si ottengono nell'uso delle nuove sorgenti dati nei processi statistici tradizionali. Il disegno metodologico tradizionale tiene sotto controllo tutte le fasi, dalla raccolta alla elaborazione del dato. Se utilizziamo le nuove sorgenti dati abbiamo meno controllo sul processo di generazione del dato, oltre che delle informazioni demografiche aggiuntive. Nell'ambito di questa discussione sono emersi diversi punti da affrontare, ad esempio molti fenomeni osservati attraverso i Big Data non presentano una distribuzione gaussiana. Pertanto una sfida (o difficoltà) degli istituti di statistica sarà di adattare le competenze e modelli per misurare fenomeni che possono essere non gaussiani. Un altro aspetto importante è il bias di selezione. Gli istituti di statistica in grado di misurare il bias di selezione sui dati non strutturati? In aggiunta, come integrare fonti di dati raccolti con intervalli spazio-temporali diversi?

Concludo osservando che il periodo è stato di grande utilità perché ho avuto la possibilità di esaminare il processo di produzione statistico del dato a livello europeo, con particolare attenzione al processo di valutazione della qualità del dato. Grazie a quanto imparato, potrò valutare in modo analitico quanto sono rappresentativi i dati su cui sto lavorando.

Inoltre ho potuto verificare diversi punti di contatto con la task force Big Data, il gruppo che a Eurostat si occupa di utilizzare le nuove sorgenti dati per estrarre nuovi indicatori. È ad esempio molto interessante l'uso che viene fatto dei dati telefonici per classificare le zone, invece degli individui, in residenziali, lavorative o miste. La necessità, dal punto demografico, di annotare le zone invece degli individui è legata al fatto che per motivi di privacy in molti paesi non vengono più rilasciate informazioni sullo stesso dispositivo/utente per più di un giorno. Questo impedisce di caratterizzare gli individui ma non impedisce di osservare la vita di un territorio, elemento molto interessante per la statistica.

22 dic. 16


Lorenzo Gabrielli