Abstract – Master di I livello: Analisi Dati per la Business Intelligence e Data Science
Anno accademico 2024/2025

Candidato: Norina Cristina Nazare

Titolo tesi: **Open research data in ambito biomedico: il possibile ruolo dei dati sintetici come fattore di bilanciamento tra Open Science e protezione dei dati sensibili.**

La crescente centralità dei dati nella ricerca scientifica contemporanea ha trasformato profondamente il modo in cui la conoscenza viene prodotta, condivisa e riutilizzata. Nel contesto dell'*Open Science*, l'apertura dei dati rappresenta un valore imprescindibile per garantire trasparenza, replicabilità e progresso collettivo. Tuttavia, quando tali dati riguardano persone, corpi, malattie o condizioni cliniche, l'apertura incontra limiti etici e giuridici che non possono essere ignorati. La protezione dei dati personali, in particolare di quelli sensibili, impone infatti un equilibrio delicato tra accessibilità e tutela e tra innovazione e responsabilità. Questa tesi esplora tale equilibrio, analizzando il ruolo emergente dei dati sintetici come possibile strumento di mediazione tra le esigenze dell'*Open Science* ed i vincoli della *data protection*. Dopo aver ricostruito il quadro teorico dell'OS ed il contesto normativo europeo (GDPR, Codice *Privacy*, AI *Act*, EHDS), il lavoro approfondisce le tecniche di generazione sintetica dei dati, i loro vantaggi ed i loro limiti, con particolare attenzione al settore biomedico. I dati sintetici vengono presentati come potenziale ibrido: artificiali nella forma, ma statisticamente coerenti con il reale, capaci di preservare la *privacy* pur mantenendo utilità scientifica. La parte applicativa della tesi è dedicata al progetto europeo MAIBAI, che sviluppa un *framework* metrologico completo per il *benchmarking* di sistemi AI, basati su immagini diagnostiche. Attraverso l'analisi dei metadati DICOM, della generazione sintetica e delle metriche applicate di *privacy*, il caso di studio mostra come i dati sintetici possano contribuire a superare la scarsità di dati sanitari condivisibili, supportando la ricerca senza compromettere la riservatezza dei pazienti.

Tale elaborato si pone l'obiettivo di evidenziare come i dati sintetici non siano una soluzione definitiva, ma un tassello fondamentale per costruire un ecosistema scientifico responsabile, in cui apertura e protezione non siano poli opposti, bensì dimensioni complementari di una stessa visione etica e sostenibile della ricerca.