



## Record Linkage

Percorso formativo: "Trattamento di archivi"

<b>OBIETTIVI FORMATIVI</b>	Arricchimento degli strumenti a disposizione degli statistici per le attività di produzione delle diverse direzioni, con l'obiettivo di un miglioramento complessivo degli <i>output</i> produttivi.
<b>DESTINATARI</b>	Statistici
<b>PREREQUISITI</b>	Conoscenza di: variabili aleatorie (normale multivariata), teoria dell'inferenza (stimatori di massima verosimiglianza, teoria dei test)
<b>DOCENTI</b>	Il corso sarà svolto con docenza interna.
<b>METODI</b>	Le lezioni saranno di natura sia teorica, sia applicativa
<b>LETTURE CONSIGLIATE</b>	<p>Scanu M. (2003) <i>Metodi Statistici per il record linkage</i>, Metodi e Norme - n.16, Istat</p> <p>Batini C., Scannapieco M. (2006) <i>Data Quality: Concepts, Methods, and Techniques</i>, Springer.</p> <p>Scheuren .F.J., Winkler E.W. (2007); <i>Data Quality and Record Linkage Techniques</i>, Springer</p>
<b>IMPATTO ORGANIZZATIVO ATTESO</b>	<p>Il corso si propone di introdurre i metodi di <i>record linkage</i> all'interno dell'Istituto. Il <i>record linkage</i> è un processo importante per l'integrazione di dati provenienti da fonti diverse. La sua complessità dipende da diversi aspetti, principalmente legati all'assenza di identificatori univoci o alla presenza di errori negli stessi identificatori. Per queste ragioni, la disponibilità di procedure standardizzate e assistite all'interno dell'Istituto per condurre studi e progetti di <i>record linkage</i> rappresenta un elemento importante per valutare la qualità, ridurre i costi e incrementare la comparabilità delle informazioni provenienti da diverse fonti..</p>

---

## PROGRAMMA

### 1-a giornata - Introduzione al *record linkage*

9.00-13.30

- ✚ 9.00-10.00 Presentazione del corso
- ✚ 10.00-11.00 Formalizzazione di un problema di *record linkage*
- ✚ 11.15-12.00 Fasi del *record linkage*
- ✚ 12.00-13.30 Approccio deterministico e probabilistico. Riepilogo e chiusura

### 2-a giornata - *Record linkage* probabilistico

9.00-13.30

- ✚ 9.00-10.30 Breve riepilogo. Modelli statistici per il *record linkage*
- ✚ 10.30-12.00 Regole di decisione: l'approccio di Fellegi-Sunter ed estensioni
- ✚ 12.00-13.30 Il modello a classi latenti per la stima dei parametri delle regole di decisione

14:30-17:30

- ✚ 14.30-15.30 Pre-processamento e standardizzazione dei dati
- ✚ 15.30-16.30 Merge, bloccaggio e altri metodi per la riduzione della complessità
- ✚ 16.30-17.30 Sessioni esercitative (casi di studio). Riepilogo e chiusura.

### 3-a giornata - *RELAIS* un software per il *record linkage*

9.00-13.30

- ✚ 9.00-10.00 Breve riepilogo. *Matching* 1:1
- ✚ 10.00-12.00 *RELAIS*: uno strumento per il *record linkage*
- ✚ 12.00-13.30 Sessioni esercitative (*RELAIS*).

14:30-17:30

- ✚ 14.30-17.30 Sessioni esercitative (casi di studio e *RELAIS*). Riepilogo e chiusura.

### 4-a giornata - Qualità del *record linkage*

9.00-13.30

- ✚ 9.00-10.00 Breve riepilogo. Qualità nel *record linkage* e valutazione dell'errore
- ✚ 10.00-11.00 Analisi statistica di dati abbinati
- ✚ 11.15-13.30 Sessioni esercitative (casi di studio). Riepilogo e chiusura.