

Il Sistema Integrato dei Registri @ Istat

Progettazione e realizzazione di una architettura
Ontology-Based

Mauro Bruno, Roberta Radini, Laura Tosco

Istituto Nazionale di Statistica - Istat

Sommario

- Il contesto
- Il Sistema Integrato dei Registri (SIR)
- Ontology Based Data Management System (OBDM)
- Realizzazione del SIR attraverso un OBDM
- Esempi di interrogazione del SIR

Il contesto

Programma di modernizzazione dell'ISTAT:

- Ufficialmente approvato nel 2016
- Profonda revisione della struttura organizzativa. Centralizzazione delle strutture trasversali (Informatica, Metodologia, Raccolta Dati, Diffusione)
- Profonda **revisione dei processi di produzione statistica**



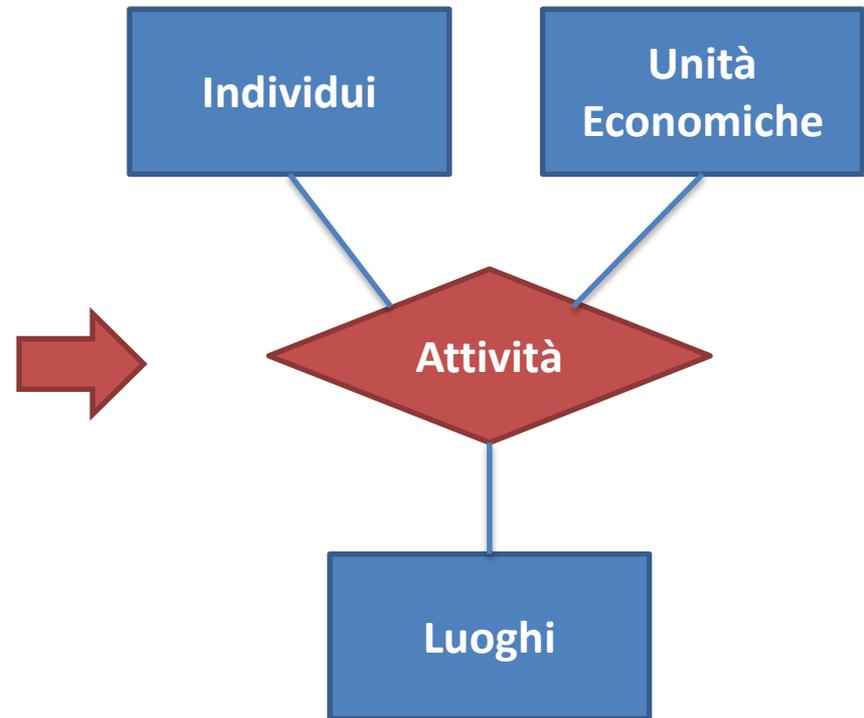
Sistema Integrato dei Registri (SIR) rappresenta uno dei pilastri della nuova architettura:

«*Single logical environment to support the **consistency** of statistical production processes, in particular **identification** and **estimation** for the whole integrated system of **units** and **variables***»

Sistema Integrato dei Registri (SIR)



Rappresentazione insiemistica



Rappresentazione Entità-Relazione

SIR: tipologie di registro

RSB (Registro Base)

contiene diverse popolazioni statistiche e il set minimo di variabili per caratterizzare le unità statistiche

RSE (Registro Esteso)

estende le informazioni di un RSB su una popolazione specifica

RST (Registro Tematico)

supporta più processi statistici trattando alcuni temi in modo consistente e condiviso

Units	Variables														
	Base			Extended						Them					
	X ₁	X _i	X _n	Y ₁	Y _i	Y _n	Y' ₁	Y' _i	Y' _n	Z ₁	Z _i	Z _n	Z' ₁	Z' _i	Z' _n
J ₁															
.....															
J _i															
.....															
J _m															
J _{m+1}															
.....															
J _{m+i}															
.....															
J _{m+p}															
.....															
U _{m+p+i}															

RSB

RSE

RST

SIR: Data Architecture

Registro Base: Attività

ID_Activity	Ref_IDIndividual	Ref_IDBusiness	Ref_IDPlace	TimeInterval	ActivityType

Registro Tematico: Relazioni di lavoro

ID_LEED	ID_Employer (Ref_IDIndividual/RefID_Business)	ID_Employee (Ref_IDIndividual)	LEED_Type	TimeInterval

Registro Base: Unità Economiche

ID_Business	Location (Ref_IDPlace)

Registro Base: Individui e Famiglie

ID_Individual	Residence (Ref_IDPlace)

Registro Base: Luoghi

ID_Place

SIR: what's missing?

Sistema integrato dei registri:

- Ambiente di supporto dei processi di produzione
- Permette di identificare e stimare le unità statistiche
- Garantisce l'integrità dei dati a livello di record

Come accedere ai dati in modo **integrato** e **indipendente** dall'eterogeneità (semantica & tecnologica) delle fonti dati?



**Ontology Based Data Management
System**

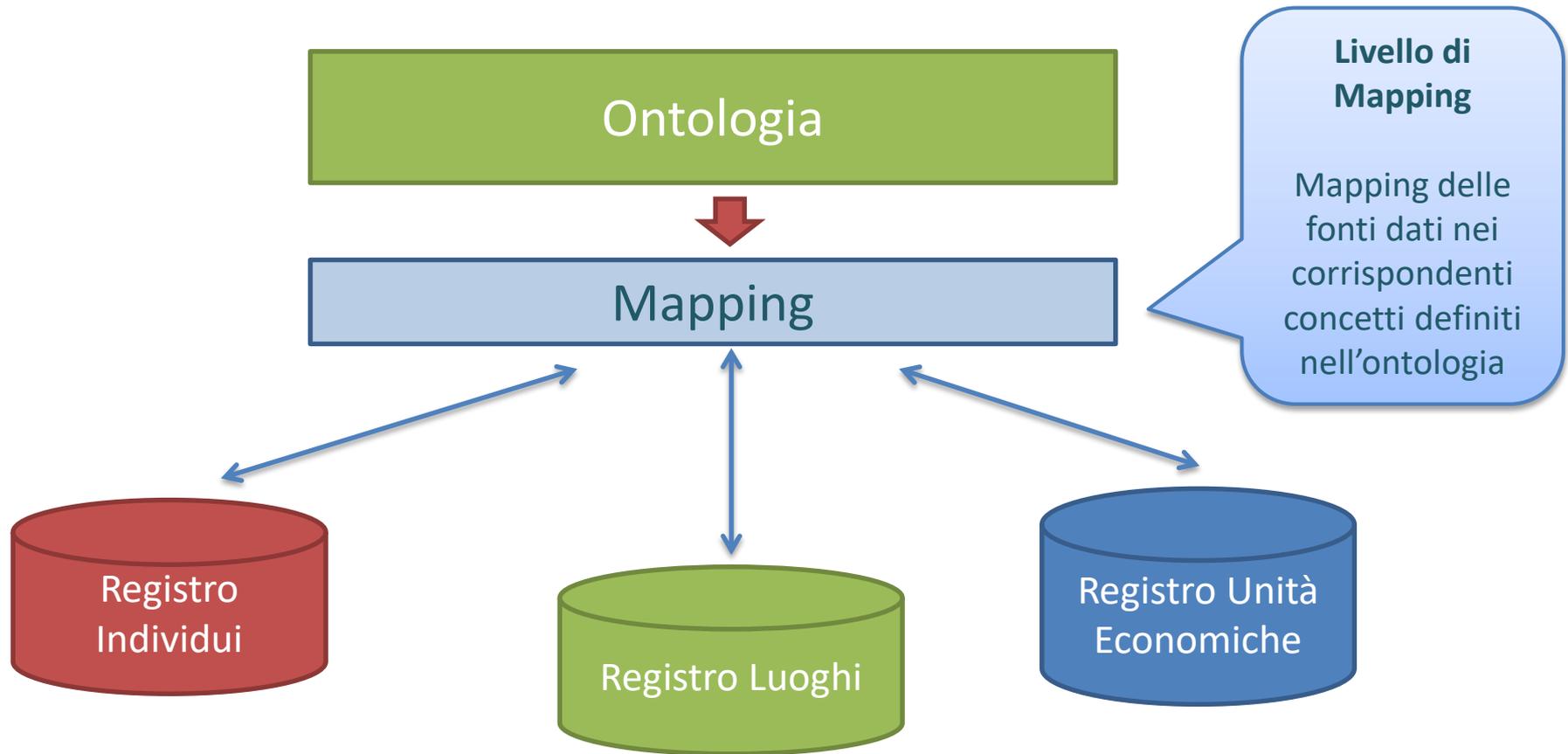
Ontology Based Data Management System (OBDM)

OBDM: Permette di accedere, integrare e gestire differenti fonti di dati

Basato su una architettura a tre livelli:

- **Ontologia:** rappresentazione formale e concettuale del dominio di interesse
- **Sorgenti dati:** sorgenti dei dati input generalmente eterogenee sia dal punto di vista semantico che tecnologico
- **Mapping:** permette di specificare le corrispondenze tra le sorgenti dati e l'ontologia

Ontology Based Data Management System (OBDM)



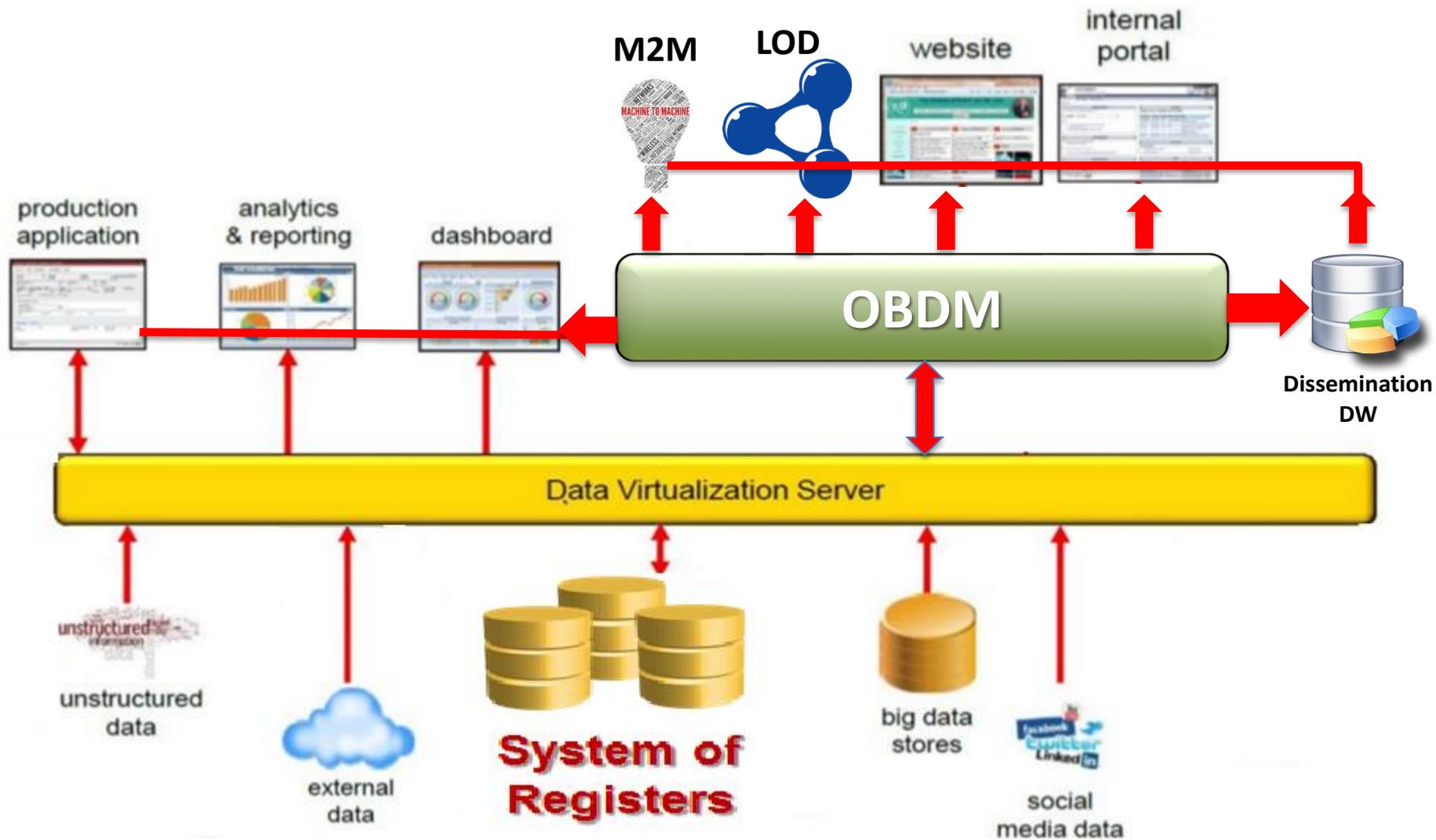
OBDM è progettato al fine di garantire l'integrità all'interno di ciascun registro (**intra**) e tra registri (**inter**)

OBDM: Principali caratteristiche

Caratteristiche principali di OBDM:

- **Trasparenza:** rispetto alle sorgenti dati. Non è necessario conoscere **come** & **dove** sono memorizzati i dati
- **Vista globale:** i concetti sono definiti globalmente in **modo univoco**. Concetti definiti in diversi domini devono essere modellati sull'ontologia
- **Consistenza:** l'utilizzo di una rappresentazione formale permette di evidenziare eventuali inconsistenze semantiche e nei dati (**funzionalità di reasoning**)

ODBM: Architettura



Vantaggi di un SIR basato su OBDM

Livello ontologico:

- Gestisce l'accesso al sistema offrendo **servizi** all'utente finale
- Permette di **governare la fase di data-integration** garantendo la qualità dei dati integrati (l'ontologia non è utilizzata solo a livello documentale)
- Offre **funzionalità di reasoning** permettendo di inferire nuova conoscenza
 - gli utenti statistici possono “scoprire” pattern impliciti che permettono di comprendere più a fondo i dati da analizzare

Vantaggi di un SIR basato su ontologie

Integration layer:

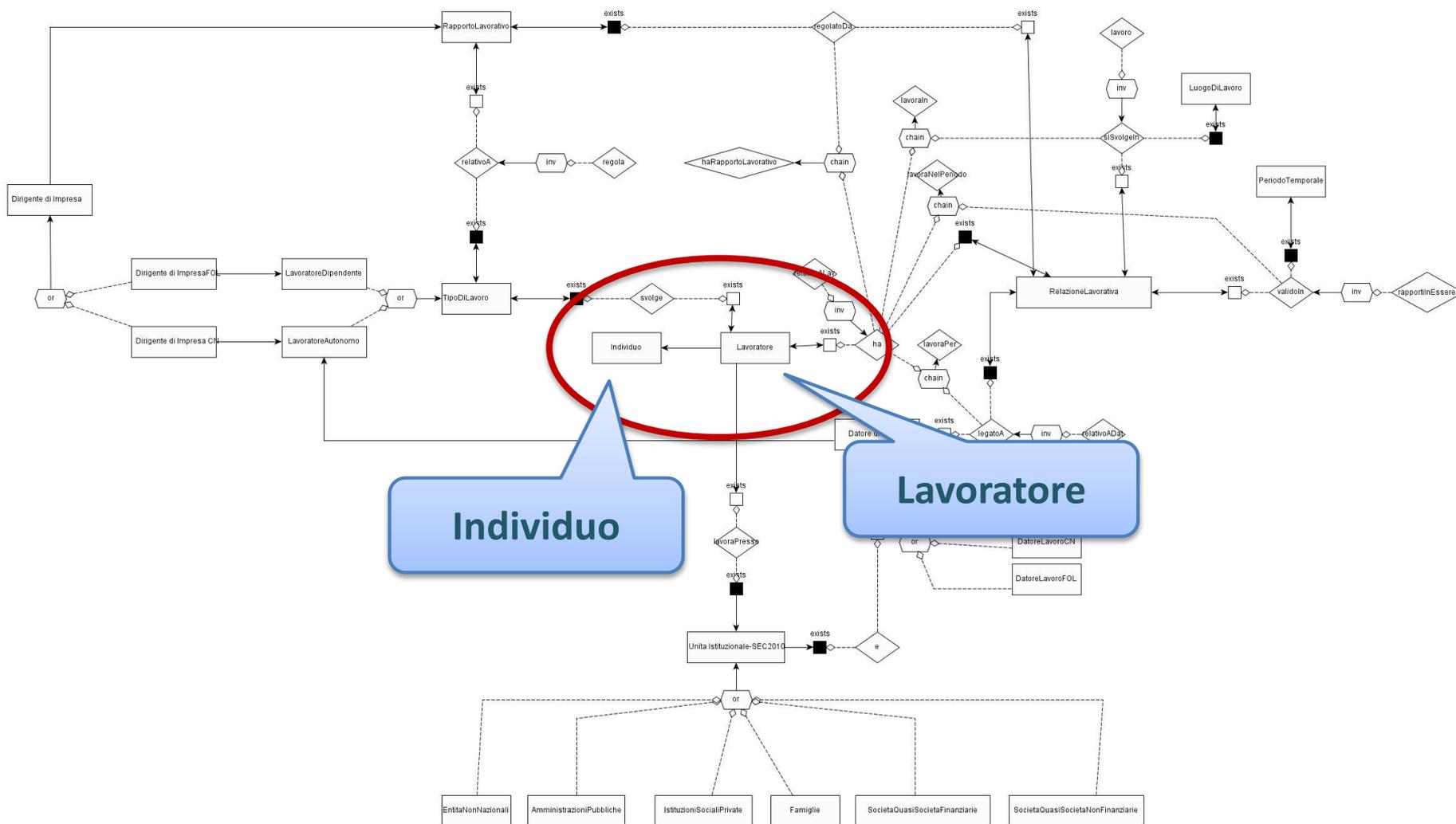
- Permette di “virtualizzare” le sorgenti dati
- Esegue interrogazioni on-the-fly

Metadati:

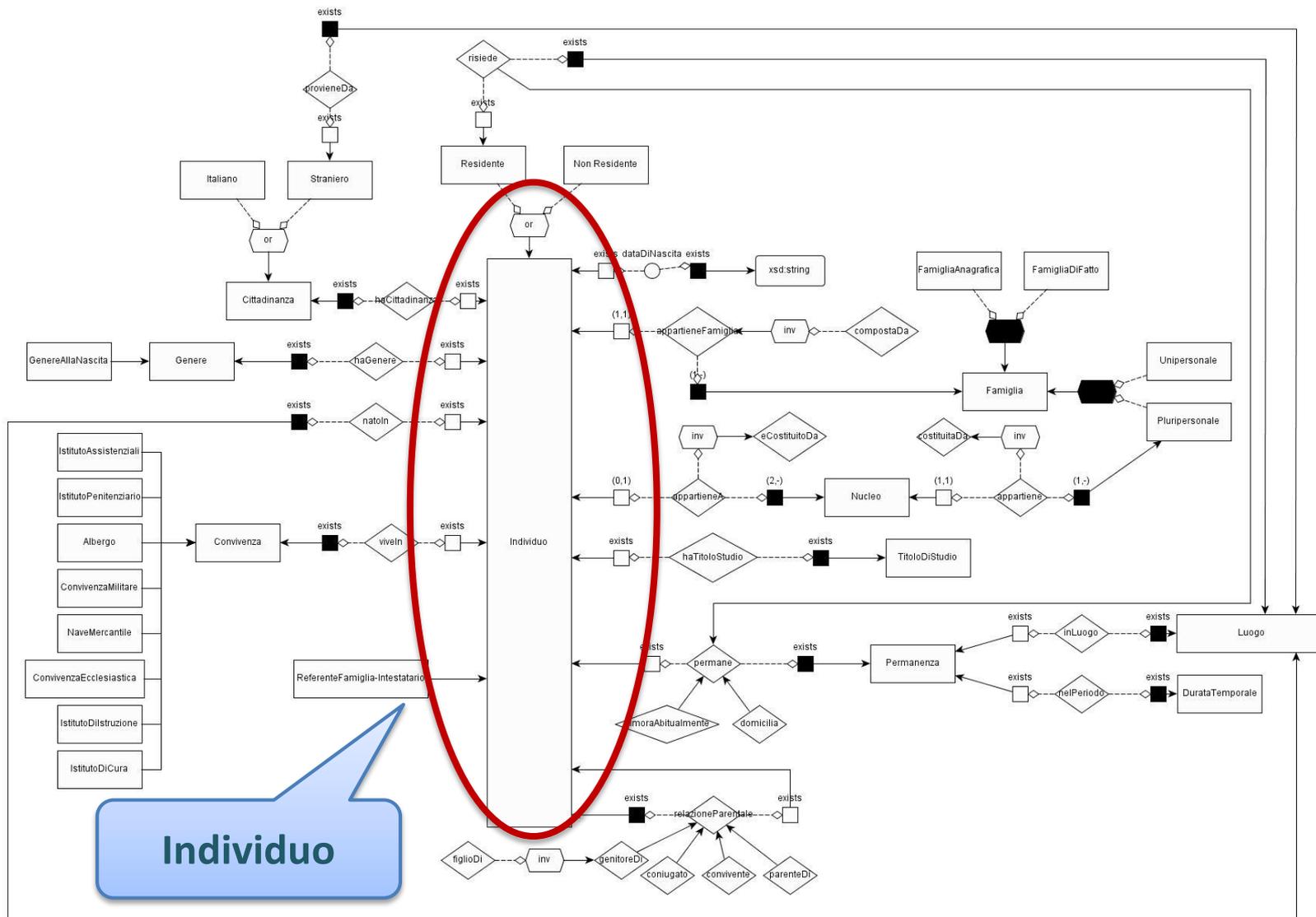
- Formalizzazione dei metadati consente l'accoppiamento di dati & metadati
- Modellazione formale consente di accedere ai metadati tramite sistemi IT

Esempio di interrogazione del SIR

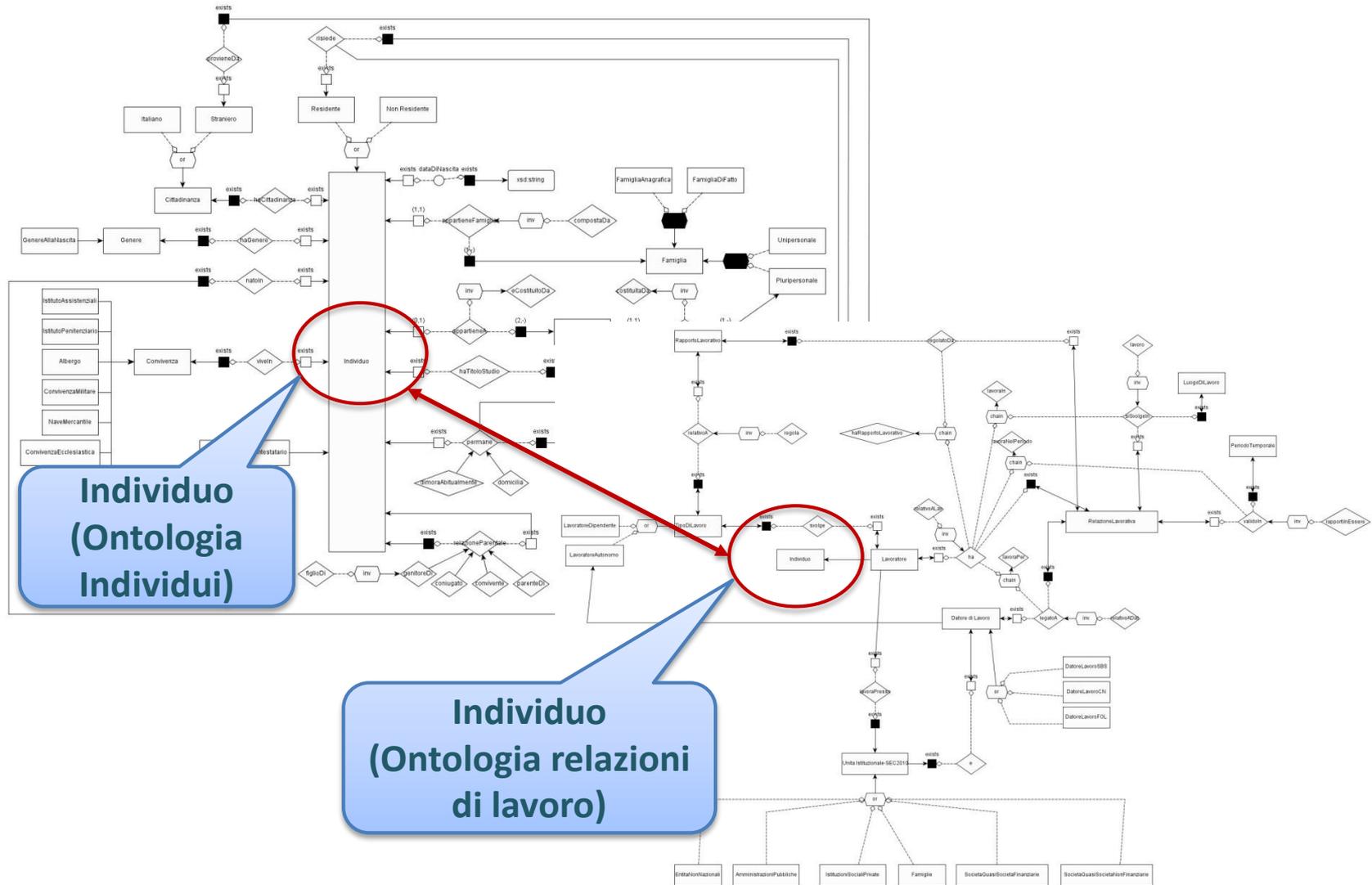
Ontologia delle Relazioni di lavoro



Ontologia degli Individui



Integrazione dati: stesso concetto



Interrogazione attraverso l'ontologia

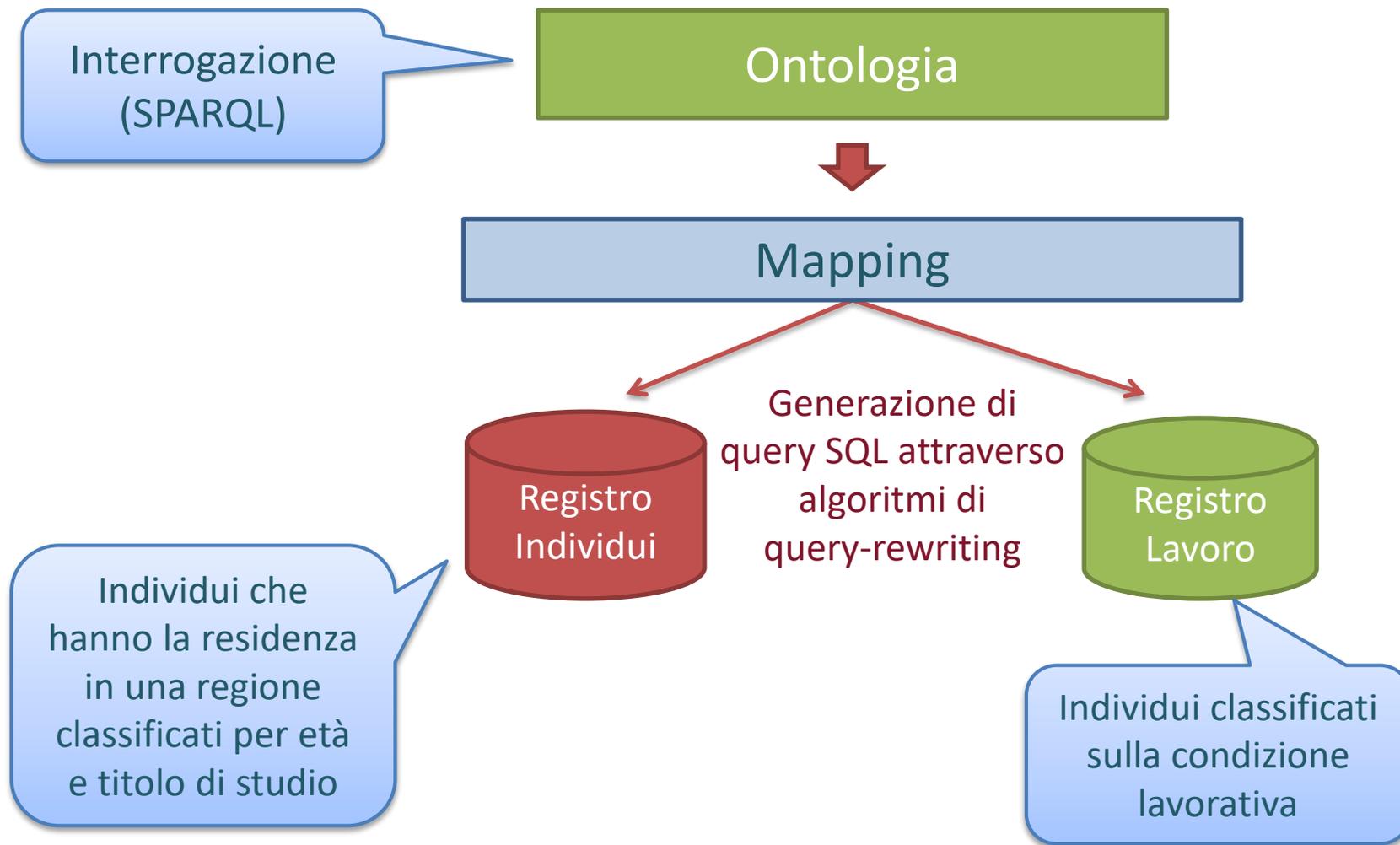
Vorremmo ottenere una risposta alla seguente domanda:

“le persone che hanno la residenza in una certa regione e classificarle per età, titolo di studio e condizione occupazionale”



Non è necessario conoscere **come & dove** sono memorizzati i dati

Interrogazione attraverso l'ontologia



Grazie per la vostra attenzione!