

Volume I

Gli argomenti di contesto

Parte I – Il Sistema informativo territoriale

(a cura di Giovanni Fazio; Francesca Monaco, Bruno Pizzolante e Giorgio Zamboni)

1. La costruzione del Sistema informativo comunale

(a cura di Giovanni Fazio)

1.1. Il Sistema informativo comunale: un datawarehouse aziendale

I sistemi informativi sono ormai ampiamente diffusi per il supporto alla gestione dei processi operativi delle organizzazioni che erogano servizi pubblici, come i Comuni.

Tali sistemi producono nel tempo enormi moli di dati, che costituiscono un patrimonio sia per le singole organizzazioni sia per la collettività.

I diversi archivi contengono la descrizione di differenti aspetti e concetti delle realtà che sono oggetto di servizi da parte delle istituzioni titolari dei dati.

Per analizzare approfonditamente i fenomeni associati ai concetti, in maniera da *creare informazione* (e quindi conoscenza) dai dati è opportuno integrare in un unico archivio i contenuti informativi sparsi nelle diverse banche dati.

A fronte del rapido cambiamento che coinvolge il sistema dei servizi moderno si rende dunque necessaria la creazione di apposite strutture che raccolgano, integrino e diffondano conoscenza (informazioni) al fine di coordinare meglio e più puntualmente le attività di supporto ai cittadini, alle famiglie ed all'impresa nell'ottica di migliorare la conoscenza puntuale del territorio da parte dei *decision maker* e quindi migliorare potenzialmente le azioni di politica attiva.

Il progressivo miglioramento delle metodologie di raccolta e conservazione dei dati nei sistemi informativi comunali consente di disporre di informazioni attendibili e aggiornate riguardanti gli individui e le famiglie sul territorio. Tale patrimonio informativo diventa abilitante per la realizzazione di alcuni strumenti che siano in grado di fornire una descrizione accurata e aggiornata della situazione del Territorio e informazioni puntuali sui target verso cui indirizzare politiche di intervento.

In ragione di ciò, il Comune di Como ritiene improcrastinabile “mettere a sistema” quell'ingente patrimonio informativo di cui si trova, in quanto Ente comunale, istituzionalmente a disporre, ed a tal fine ha stabilito di procedere alla costruzione del *Sistema informativo statistico comunale* (nel seguito indicato con l'acronimo S.I.C.), datawarehouse aziendale dell'Ente Comune di Como.

Questa volontà, politica e gestionale insieme, rappresenta una svolta storica per l'Ente, in quanto per la prima volta si pongono le basi per un approccio “scientifico” alla conoscenza del territorio comunale tramite la valorizzazione di tutta quell'informazione che si accumula, spesso in modo sordinato, all'interno delle varie banche dati comunali.

Da un punto di vista più prettamente tecnico, l'obiettivo primario della creazione di un Sistema informativo comunale consiste nella definizione di un modello di analisi dei dati, di un insieme di indicatori comuni e di un sistema statistico di supporto che possa essere facilmente alimentato ed utilizzato dalle banche dati locali competenti.

L'output di tale azione è rappresentato dal modello statistico e di monitoraggio dei dati amministrativi a livello locale prendendo a riferimento quale base fondamentale l'*archivio anagrafico comunale*.

Per poter disporre di informazioni complete e corrette è necessario studiare tutte le fonti dei dati disponibili e definire un *modello integrato* dei dati in esse contenuti in modo da correlare attributi conservati in archivi distinti. È inoltre necessario individuare le modalità più opportune per stabilizzare prima e rendere automatiche poi le forniture dei dati e gestire i processi di *qualificazione* dei dati disponibili.

Da ultimo si sottolinea l'importanza nella trattazione di dati amministrativi della costituzione di gruppi di lavoro che vedano continuamente la partecipazione di soggetti referenti delle amministrazioni che mettono a

disposizione i dati sia per meglio cogliere le dinamiche dei processi che li generano sia per divenire a una corretta interpretazione delle analisi che possono essere effettuate.

Nel seguito si riportano alcune definizioni dei principali componenti di un Sistema Informativo Statistico.

Datawarehouse

Un datawarehouse (termine inglese traducibile con magazzino di dati), è un archivio informatico contenente i dati di un'organizzazione. I datawarehouse sono progettati per consentire di produrre facilmente relazioni ed analisi.

Vengono considerati componenti essenziali di un sistema datawarehouse anche gli strumenti per localizzare i dati, per estrarli, trasformarli e caricarli, come pure gli strumenti per gestire un dizionario dei dati. Una definizione ampliata comprende inoltre gli strumenti per gestire e recuperare i metadati e gli strumenti di business intelligence.

William H. Inmon, colui che per primo ha parlato esplicitamente di datawarehouse, lo definisce come una raccolta di dati integrata, orientata al soggetto, variabile nel tempo e non volatile di supporto ai processi decisionali. L'integrazione dei dati costituisce la principale caratteristica distintiva del DW rispetto ad altri sistemi di supporto alle decisioni; secondo Inmon la raccolta di dati è:

- a) *integrata*: requisito fondamentale di un datawarehouse è l'integrazione dei dati raccolti. Nel datawarehouse confluiscono dati provenienti da più sistemi transazionali e da fonti esterne. L'obiettivo dell'integrazione può essere raggiunto percorrendo differenti strade: mediante l'utilizzo di metodi di codifica uniformi, mediante il perseguimento di una omogeneità semantica di tutte le variabili, mediante l'utilizzo delle stesse unità di misura;
- b) *orientata al soggetto*: il datawarehouse è orientato a temi aziendali specifici piuttosto che alle applicazioni o alle funzioni. In un datawarehouse i dati vengono archiviati in modo da essere facilmente letti o elaborati dagli utenti. L'obiettivo, quindi, non è più quello di minimizzare la ridondanza mediante la normalizzazione, ma quello di fornire dati organizzati in modo tale da favorire la produzione di informazioni. Si passa dalla progettazione per funzioni a una modellazione dei dati che consenta una visione multidimensionale degli stessi;
- c) *variabile nel tempo*: i dati archiviati all'interno di un DW coprono un orizzonte temporale molto più esteso rispetto a quelli archiviati in un sistema operativo. Nel DW sono contenute una serie di informazioni relative alle aree di interesse che colgono la situazione relativa a un determinato fenomeno in un determinato intervallo temporale piuttosto esteso. Ciò comporta che i dati contenuti in un DW siano aggiornati fino a una certa data che, nella maggior parte dei casi, è antecedente a quella in cui l'utente interroga il sistema. Ciò differisce da quanto si verifica in un sistema transazionale, nel quale i dati corrispondono sempre a una situazione aggiornata, solitamente incapace di fornire un quadro storico del fenomeno analizzato;
- d) *non volatile*: tale caratteristica indica la non modificabilità dei dati contenuti nel DW che consente accessi in sola lettura. Ciò comporta una semplicità di progettazione del database rispetto a quella di un'applicazione transazionale. In tale contesto non si considerano le possibili anomalie dovute agli aggiornamenti, né tanto meno si ricorre a strumenti complessi per gestire l'integrità referenziale o per bloccare record a cui possono accedere altri utenti in fase di aggiornamento.

Il datawarehouse, quindi, descrive il processo di acquisizione, trasformazione e distribuzione di informazioni presenti all'interno o all'esterno delle aziende come supporto ai decision maker. Esso si differenzia in modo sostanziale dai normali sistemi gestionali che, al contrario, hanno il compito di automatizzare le operazioni di routine. Si può notare che la definizione di Inmon precedentemente citata sia indifferente rispetto alle caratteristiche architettoniche dei sistemi transazionali e alla dislocazione fisica dei dati nei diversi database. Se il focus viene posto sulla capacità di supportare il processo decisionale, il datawarehouse può essere costruito secondo modalità differenti, che possono andare da una logica completamente accentrata a una logica completamente distribuita.

Gli elementi costitutivi dell'architettura sono:

- i) *i dati provenienti dai sistemi transazionali*: sono quell'insieme di dati elaborati dai sistemi transazionali dell'azienda. Essi possono essere contenuti all'interno dello stesso database o provenire da diversi database o anche esterni all'azienda. Spesso l'architettura di un datawarehouse prevede l'integrazione

dei dati interni con quelli esterni. L'utilizzo di questi ultimi consente di arricchire il patrimonio informativo;

- ii) *il data movement*: tale componente è responsabile dell'estrazione dei dati dai sistemi transazionali, della integrazione tra dati aziendali e dati esterni, del pre – processing dei dati, del controllo della consistenza dei dati, della conversione delle strutture dati, e dell'aggiornamento dei dizionari dei dati;
- iii) *il datawarehouse*: i dati estratti dagli archivi transazionali vengono memorizzati internamente al datawarehouse. Nel datawarehouse l'accesso ai dati è consentito in sola lettura. Tali dati hanno una dimensione storica e sono riferiti a soggetti di business. Essi possono essere memorizzati in un archivio centrale o in un datamart. Il termine datamart identifica un datawarehouse di dimensioni ridotte, specializzato per una particolare area di attività. All'interno della banca possono quindi esistere più datamart, aventi finalità diverse e orientati a coprire diverse aree. I dati contenuti nel datawarehouse possono essere aggregati e indicizzati per rispondere a specifiche necessità informative;
- iv) *i metadati*: costituiscono informazione aggiuntiva che arricchisce i dati contenuti nel datawarehouse. Spesso essi vengono chiamati in gergo “data about data” indicando la provenienza, l'utilizzo, il valore o la funzione del dato. A tale proposito vengono costituiti dei veri e propri *information catalog*. Questi ultimi sono i file che contengono i metadati. Il catalog consente di spiegare all'utente la natura dei dati nel datawarehouse, il loro significato semantico, da quali archivi essi provengono e la loro storicità;
- v) *l'utente finale*: i dati contenuti nel datawarehouse vengono presentati all'utente finale, il quale dispone di un insieme di strumenti per effettuare elaborazioni e produrre informazioni appropriate. I tool a disposizione dell'utente possono essere semplici generatori di *query* e *report*, interfacce grafiche che consentono la rappresentazione dei dati o sistemi di analisi dati più complessi.

Il datawarehouse è organizzato su quattro livelli architetturali:

1. *la trasformazione dei dati*: è il livello che si occupa di acquisire i dati e validarli;
2. *la preparazione e “stoccaggio” dei dati*: è il livello che fornisce i dati agli utenti e alle applicazioni analitiche;
3. *l'interpretazione e analisi dei dati*: è il livello, ad elevato valore aggiunto, che presiede alla trasformazione dei dati in informazioni aventi valore strategico;
4. *la presentazione dei dati*: è il livello, a basso valore aggiunto, che presiede alla presentazione finale agli utenti delle informazioni e quindi delle risposte cercate.

Nel suo complesso il datawarehouse è un sistema periferico, cioè non è necessario che risieda fisicamente sul sistema informativo centrale.

Olap

Olap, acronimo che sta per l'espressione *On – Line Analytical Processing*, designa un insieme di tecniche software per l'analisi interattiva e veloce di grandi quantità di dati, che è possibile esaminare in modalità piuttosto complesse.

È la componente tecnologica base del datawarehouse.

Gli strumenti *Olap* si differenziano dagli *Oltip* per il fatto che i primi hanno come obiettivo la performance nella ricerca e il raggiungimento di un'ampiezza di interrogazione quanto più grande possibile; i secondi, invece, hanno come obiettivo la garanzia di integrità e sicurezza delle transazioni.

La creazione di un database *Olap* consiste nell'effettuare una fotografia di informazioni (ad esempio quelle di un database relazionale) in un determinato momento e trasformare queste singole informazioni in dati multidimensionali. Eseguendo successivamente delle interrogazioni sui dati così strutturati è possibile ottenere risposte in tempi decisamente ridotti rispetto alle stesse operazioni effettuate su altre tipologie di database, anche perché il DB di un sistema *Oltip* non è stato studiato per consentire analisi articolate.

Una struttura *Olap* creata per questo scopo è chiamata “*cubo*” *multidimensionale*.

Ci sono diversi modi per creare un cubo, ma il più conosciuto è quello che utilizza uno schema “a stella”; al centro c'è la tabella dei “fatti” che elenca i principali elementi su cui sarà costruita l'interrogazione, e collegate a questa tabella ci sono varie tabelle delle “dimensioni” che specificano come saranno aggregati i dati.

Il calcolo delle possibili combinazioni di queste aggregazioni forma una struttura *Olap* che, potenzialmente, potrebbe contenere tutte le risposte per ogni singola combinazione. In realtà viene memorizzato solo un nu-

mero predeterminato di combinazioni, mentre le rimanenti vengono ricalcolate solo al momento in cui quella richiesta viene materialmente effettuata.

Un sistema *Olap* permette di:

- i) studiare una grande quantità di dati;
- ii) vedere i dati da prospettive diverse;
- iii) supportare i processi decisionali.

Le funzioni di base di uno strumento *Olap* sono:

- a) *Slicing*: è l'operazione di rotazione delle dimensioni di analisi. È un'operazione fondamentale per analizzare totali ottenuti in base a dimensioni diverse o se si vogliono analizzare aggregazioni trasversali;
- b) *Dicing*: è l'operazione di estrazione di un subset di informazioni dall'aggregato che si sta analizzando. L'operazione di dicing viene eseguita quando l'analisi viene focalizzata su una 'fetta del cubo' avente particolare interesse per l'analista. In alcuni casi l'operazione di dicing può essere 'fisica' nel senso che non consiste solo nel filtrare le informazioni di interesse ma anche nell'estrarle dall'aggregato generale per distribuirne i contenuti;
- c) *Drill – down*: è l'operazione di 'esplosione' del dato nelle sue determinanti. L'operazione di drill – down può essere eseguita seguendo due diversi percorsi: la gerarchia costruita sulla dimensione di analisi oppure la relazione matematica che lega un dato calcolato alle sue determinanti;
- d) *Drill – across*: è l'operazione mediante la quale si naviga attraverso uno stesso livello nell'ambito di una gerarchia;
- e) *Drill – through*: concettualmente simile al drill – down, è l'operazione mediante la quale si passa da un livello aggregato al livello di dettaglio appartenente alla base dati normalizzata.

Ipercubo

Un cubo *Olap* è una struttura per la memorizzazione di dati che permette di eseguire analisi in tempi rapidi, superando un limite dei database relazionali. I database relazionali non sono adatti ad elaborazioni istantanee e alla visualizzazione di una grande mole di dati. Al contrario, sono adatti all'elaborazione di record derivanti dall'esecuzione di transazioni note come *Oltp*. Il cubo *Olap* può essere immaginato come un'estensione del foglio di lavoro a due dimensioni. Poiché in un cubo *Olap* possono esserci in generale più di tre dimensioni, il termine più appropriato è quello di ipercubo.

Il contesto

Si possono reperire in letteratura diverse definizioni di Sistema Informativo Statistico, tra le quali:

- i) definizione ufficiale (Unece, 2000): Sistema informativo orientato alla raccolta, memorizzazione, trasformazione e distribuzione d'informazione statistica;
- ii) definizione specifica (glossario Istat): Sistema informativo in cui l'utilizzo delle informazioni mira a finalità prevalentemente statistico – conoscitive, e che si caratterizza per il trattamento di dati aggregati ("macrodati") e per la particolare attenzione dedicata ai "metadati" e alle "meta – informazioni";
- iii) definizione operativa (Martini, 2001): Sistema informativo di dati statistici elementari che attribuiscono caratteri qualitativi a diverse unità statistiche connotate nel tempo e nello spazio.

Lo scopo di un Sistema Informativo Statistico, quale è il S.I.C., è quello di rendere fruibile l'informazione presente nelle banche dati amministrative, attraverso un processo complesso di trattamento dei dati elementari in esse presenti. Il sistema informativo così ottenuto, acquisisce anche la funzione di sistema di "retroazione e controllo" (Martelli, 2004) rispetto ai sistemi amministrativi dai quali vengono prelevati i dati.

Un Sistema Informativo Statistico viene costruito con l'obiettivo di monitorare con cadenza regolare i cambiamenti che intervengono sul fenomeno in osservazione, attraverso diversi punti di vista (studi economici – sociali) al fine di supportare i diversi livelli decisionali (politici e dirigenziali) nell'acquisizione di informazioni puntuali e creare quindi opportune conoscenze a supporto decisionale.

Nel corso degli ultimi anni sono stati affrontati diversi progetti, partecipati da ricercatori appartenenti a diverse discipline (sistemi di elaborazione dell'informazione, statistica, economia e management, giuridiche), che si sono concentrati su tre filoni di studio principali:

- i) il modello del sistema informativo;

- ii) la definizione delle metodologie di analisi;
- iii) i modelli di interpretazione e valutazione dei risultati.

Il modello del Sistema Informativo Statistico

L'elevata eterogeneità degli archivi coinvolti, vista da diversi punti di osservazione quali la loro struttura, il contenuto, la qualità dei dati ospitati ed il diverso livello di aggiornamento nel tempo, costituiscono il principale problema studiato ed affrontato per poter integrare i dati dei differenti archivi ed effettuare analisi di natura statistica a supporto decisionale sotto un profilo descrittivo, esplorativo e previsionale.

La metodologia utilizzata per integrare gli archivi prevede un'accurata analisi dello schema e del contenuto dei singoli database e la creazione di un unico archivio integrato nel quale far confluire i dati dei database delle singole istituzioni coinvolte. Contestualmente a queste operazioni, si sono studiati modelli e tecniche di analisi della qualità dei dati presenti negli archivi, allo scopo di identificare le operazioni di "messa in qualità" da eseguire affinché i dati di partenza raggiungano il livello qualitativo necessario allo svolgimento delle analisi fenomenologiche. Il percorso è proseguito con il disegno e la conseguente implementazione dell'architettura del sistema informativo statistico basato su datawarehouse.

Avendo introdotto i concetti di integrazione degli archivi e di qualità dei dati, si riportano dei brevi approfondimenti su tali temi.

L'integrazione dei dati

Il primo passo del processo di integrazione consiste nello svolgere un'analisi degli archivi amministrativi originali, con lo scopo di produrre una descrizione dello schema a livello concettuale sul quale l'archivio stesso è modellato. Completata la fase di analisi, spesso risulta necessario normalizzare i singoli archivi di origine, sia dal punto di vista della struttura che dal punto di vista dei contenuti.

- a) *la struttura*: la normalizzazione dello schema dell'archivio è un'attività che prevede una parziale riorganizzazione della struttura logica con cui sono organizzati i dati, allo scopo di semplificare lo schema dell'archivio di partenza e soprattutto ridurre le ridondanze presenti.
- b) *i contenuti*: la normalizzazione dei contenuti dell'archivio consiste nello standardizzare il contenuto degli archivi e nel ricondurre le possibili varianti di uno stesso dato a un unico valore.

Terminata la fase di normalizzazione, si passa alla progettazione di uno schema integrato; ovvero, la struttura di un unico archivio nel quale vanno a confluire le informazioni di tutti gli archivi di partenza. Tale schema viene progettato in modo che riesca a raggruppare le informazioni comuni, e ad integrare in modo sistematico le componenti eterogenee dei diversi archivi di partenza. Per esempio all'interno di un SIS lo schema integrato può prevedere una gestione unificata dell'anagrafica cittadini dei soggetti.

Una volta definito lo schema dell'archivio integrato, si procede alla definizione della "mappatura" sia tra gli schemi sia tra le istanze dei database, le quali guideranno le attività di estrazione dei dati dagli archivi normalizzati di partenza e il successivo caricamento nell'archivio integrato. La mappatura degli schemi è una attività di astrazione; definisce le corrispondenze tra archivi concentrandosi unicamente sullo schema e astraendo da dettagli (per esempio le codifiche utilizzate).

La mappatura delle istanze, partendo dal lavoro svolto a livello di schema, integra le informazioni considerando i dettagli di più basso livello come il formato dei dati (ad esempio una data può essere espressa nel formato 12/mar/1976 nell'archivio di origine mentre nell'archivio integrato viene espressa nel formato 12/03/1976), i sistemi classificatori usati per esprimere i valori di un dato e la granularità adottata all'interno dei sistemi classificatori. Il risultato derivante dalle attività sopra descritte, è la riconduzione a un formato normalizzato di diversi archivi trattati.

Gli archivi spesso hanno delle strutture molto complicate (un archivio amministrativo può essere composto anche da centinaia di tabelle differenti), e pertanto la suddivisione delle operazioni di mappatura in due livelli permette di potersi concentrare *in primis* sulle corrispondenze tra schemi per poi considerare in un secondo momento i dettagli di più basso livello.

La qualità dei dati

Gli archivi amministrativi, in genere, non si trovano in uno stato qualitativo sufficiente per svolgere correttamente delle analisi; infatti, si riscontrano errori e inconsistenze le quali, sebbene non influiscano sulle

“normali” attività amministrative, possono condizionare i risultati di analisi statistiche; prima di procedere a qualsiasi analisi è necessario quindi verificare lo “stato qualitativo” degli archivi.

In generale, la qualità dell’informazione si può così articolare:

- a) la qualità del modello concettuale con cui è organizzata l’informazione all’interno degli archivi;
- b) la qualità dei valori dei dati presenti negli archivi;
- c) la qualità con la quale l’informazione viene rappresentata esternamente.

Le principali analisi di qualità degli archivi amministrativi riguardano i primi due aspetti e, rispetto alla qualità del modello concettuale, si riportano alcune delle loro dimensioni qualitative di maggior interesse:

1. *la rilevanza*: (indica in che misura i dati sono rilevanti per gli scopi di analisi), ottenibilità (misura della facilità con cui le informazioni necessarie per l’analisi sono ottenibili dagli schemi a disposizione) e chiarezza della definizione del contenuto (indica se ogni termine nella definizione del modello è chiaramente definito o meno);
2. *lo scopo*: definito come il grado con cui il modello comprende i dati necessari per soddisfare le necessità di analisi;
3. *la granularità degli attributi* (indica il numero di possibili valori utilizzabili per rappresentare un concetto) e precisione dei domini rappresentati (livello di dettaglio delle misure o dello schema di classificazione che definiscono il dominio o i domini di riferimento).

In merito alla qualità del valore dei dati, si possono considerare le seguenti dimensioni di analisi:

- i. *l’accuratezza*: vicinanza del valore di un dato memorizzato negli archivi al valore effettivamente descritto dalla distrazione reale;
- ii. *l’aggiornamento*: misura l’adeguatezza della rapidità di aggiornamento dei dati memorizzati negli archivi, in realtà mutevoli nel tempo;
- iii. *la completezza*: indica in che misura un dato previsto a livello di schema è effettivamente presente all’interno dell’archivio;
- iv. *l’inconsistenza tra i valori*: si ha quando due attributi, tra i quali esiste una forma di dipendenza, presentano valori non compatibili.

Scopo dell’analisi di qualità degli archivi è stabilire se i dati contenuti presentano un livello qualitativo sufficiente allo svolgimento delle analisi preposte sulla base di alcuni metodi:

- a) la comparazione dei dati con le controparti reali, metodo molto efficace ma costoso, sia in termini economici sia in termini di tempo richiesto; non è quindi applicabile su larga scala;
- b) il Database bashing, prevede la comparazione di dati teoricamente simili presenti in due o più database diversi. Questo metodo è efficace per individuare dati discordanti, non è altrettanto efficace per la correzione;
- c) l’utilizzo di “data edits”, routines che verificano se i valori e le loro rappresentazioni soddisfano vincoli predeterminati. Questo metodo è molto utile in fase di analisi della qualità, ma non altrettanto in fase di correzione.

Occorre tuttavia sottolineare che le attività di “qualificazione” degli archivi hanno un effetto che non perdura nel tempo se non si interviene sulle cause sistematiche che generano nuovi errori negli archivi.

Le potenzialità di un datawarehouse comunale

A semplice titolo di esempio vengono riportati alcuni *snapshot* del S.I.C., con l’intento di fare intuire le potenzialità dello strumento, mostrando alcuni *snapshot* di Analisi *Olap* sulla Tarsu, alcuni *snapshot* di Analisi *Olap* sui residenti, nonché un’analisi territoriale con evidenziate le linee TPL (Trasporto Pubblico Locale) in associazione alla distribuzione della popolazione per quartieri: gli *snapshot* non possono ovviamente mostrare la gran dinamicità offerta dal prodotto di datawarehouse, per sua natura utilizzabile per essere realmente apprezzato nella sua grande potenzialità; le Analisi *Olap* (tabellari), veloci ed estremamente intuitive, sono fonte preziosissima di conoscenza del fenomeno trattato (popolazione, Tarsu, etc.), mentre l’analisi territoriale è intuitiva “per natura” nel mostrare la distribuzione di determinate grandezze sul Territorio.



								Measures				
anno	Descrizione	Tipo	Via	Quartiere Geo	Circoscrizione	Sezione censimento	Tipo Civico	nr righe	nr civici	nr contribuenti	superficie totale	Imposta
+anno	-Descrizione	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	329.470	8.860	53.251	34.622.641,00	53.130,9
	ABITAZIONI E LOCALI PERTINENZIALI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	289.332	8.231	49.648	27.134.786,00	35.723,0
	AREE SCOPERTE ADIBITE A PARCHEGGIO DISTR	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	734	38	56	430.507,00	154,2
	ASSOCIAZIONI FONDAZIONI ENTI MORALI PART	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	1.060	125	170	244.669,00	292,6
	AUTODEMOLIZIONE	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	5	1	2	2.446,00	6,2
	AUTOFFICINE RIPARAZ. PARTI MECCANICHE	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	336	108	148	57.116,00	134,9
	AUTOLAVAGGIO	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	35	9	14	5.248,00	15,3
	AUTORIMESSE-AUTONOLEGGI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	301	81	113	184.051,00	377,6
	BAR - PASTICCERIE	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	1.759	297	587	125.360,00	587,0
	CAMPEGGI E LIDI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	8	1	3	31.860,00	57,7
	CARCERE CASERME DISTRETTO	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	108	14	9	456.012,00	502,4
	CENTRI SPORTIVI PALESTRE E PISCINE	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	107	15	25	22.348,00	37,0
	CONVENTI E CONVITTI COMUNITA' RICOVERI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	191	30	29	386.732,00	924,2
	GRANDI MAGAZZINI SUPERMERCATI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	34	15	21	56.910,00	375,2
	IMPORTO CONCORDATO	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	189	86	126	54.787,00	52,3
	LABORATORI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	2.455	530	850	242.658,00	639,8
	LOC. E AREE DETASSATE INDUSTRI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	584	165	253	245.806,00	
	MAGAZZINI-TETTOIE PER DEPOSITI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	8.064	1.513	2.734	851.832,00	2.204,7
	MENSE - RISTORANTI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	768	149	261	100.770,00	1.256,6
	MINIGOLF	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	8		2	6.632,00	10,3
	NEGOZI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	7.757	945	2.077	481.952,00	2.230,4
	OSPEDALI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	18	3	6	136.664,00	512,9
	PENSIONI ALBERGHI E LOCANDE CON ALLG.OST	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	188	27	36	245.655,00	766,8
	SCUOLE CIRC.RICREATIVI ORATORI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	1.179	151	156	1.674.738,00	362,4
	SERRE	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	17	1	5	2.868,00	7,3
	TEATRI CINEMA E LOCALI DI DIVERTIMENTO	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	106	20	29	78.077,00	84,1
	TEATRI CINEMA E LOCALI DI DIVIRTIMENTO	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	16	12	14	11.230,00	10,6
	UFFICI E STUDI	+Tipo	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	14.111	1.654	3.785	1.350.927,00	5.803,9

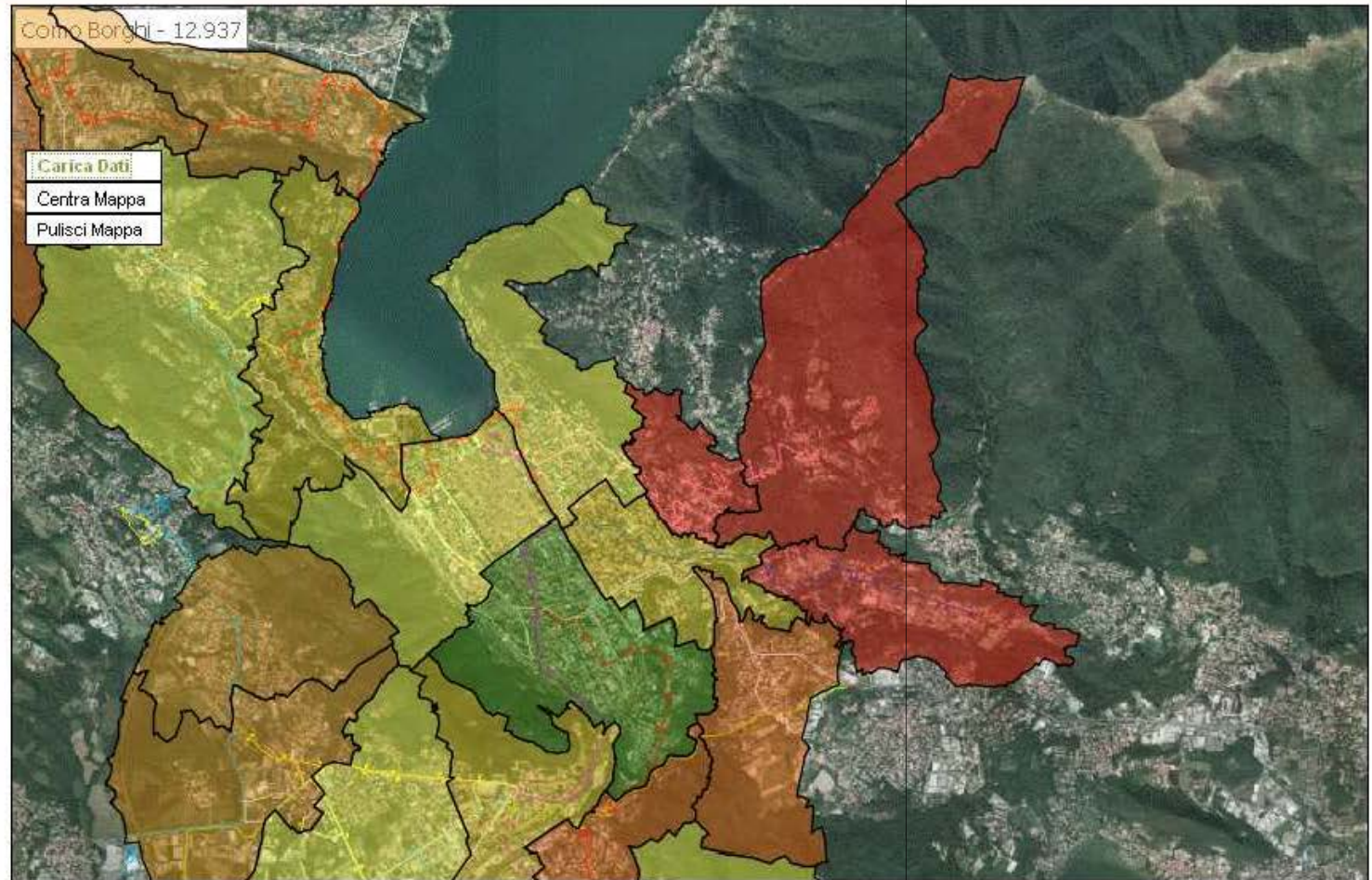
2003	-Descrizione	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+45.880	+8.522	+42.108	+4.847.567,00	+6.121.444,55
	ABITAZIONI E LOCALI PERTINENZIALI	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+40.434	+7.636	+38.380	+3.780.983,00	+4.118.077,39
	AREE SCOPERTE ADIBITE A PARCHEGGIO DISTR	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+102	+19	+32	+65.638,00	+14.984,90
	ASSOCIAZIONI FONDAZIONI ENTI MORALI PART	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+147	+101	+133	+42.768,00	+42.810,17
	AUTOFFICINE RIPARAZ. PARTI MECCANICHE	-Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+44	+34	+44	+7.381,00	+14.553,00
		#null	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+2	+1	+2	+352,00	+811,31
	AL BASSONE	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+170,00	+345,97	
	ALFONSO OLDELLI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+92,00	+221,49	
	ASIAGO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+108,00	+146,53	
	BATTISTA TETTAMANTI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+70,00	+142,46	
	BELLINZONA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+70,00	+142,46	
	BRIANTEA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+74,00	+150,60	
	CANTURINA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+2	+2	+2	+207,00	+421,27	
	CARLO BOSELLI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+54,00	+109,90	
	DANTE ALIGHIERI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+759,00	+1.544,64	
	DEI MULINI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+68,00	+138,39	
	DEL GELSO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+300,00	+610,53	
	DEL LAVORO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+2	+2	+2	+269,00	+547,44	
	DONATORI DI SANGUE	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+102,00	+207,58	
	ENRICO PESSINA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+61,00	+124,14	
	FILIPPO TURATI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+240,00	+488,42	
	FORNACE	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+113,00	+272,05	
	FRANCESCO ANZANI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+81,00	+164,84	
	FRANCESCO VIGANO*	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+87,00	+177,05	
	FRATELLI ROSSELLI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+14,00	+28,49	
	GABRIELE D'ANNUNZIO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+44,00	+89,54	
	GIAN BATTISTA SCALABRINI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+2	+1	+2	+275,00	+559,66	
	INNOCENZO XI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+435,00	+885,27	
	LUIGI DOTTESIO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+290,00	+590,18	
	MASSENZIO MASIA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+758,00	+1.963,27	
	MENTANA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+112,00	+269,64	
	NAPOLEONA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+2	+2	+2	+379,00	+771,30	
	OLTRECOLLE	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+25,00	+50,88	
	PANNILANI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+48,00	+97,68	
	PASQUALE PAOLI	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+3	+2	+3	+798,00	+1.624,01	
	ROMA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+503,00	+1.023,66	
	RONCHETTO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+20,00	+40,70	
	SAN FRANCESCO D'ASSISI	-Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+47,00	+95,65	
		Camerlata	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+47,00	+95,65	
	VARESINA	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+3	+3	+3	+328,00	+667,51	
	ZEZIO	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+1	+1	+1	+28,00	+28,49	
	AUTOLAVAGGIO	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+3	+1	+3	+534,00	+1.285,60
	AUTORIMESSE-AUTONOLEGGI	+Via	+Quartiere Geo	+Circoscrizione	+Sezione censimento	+Tipo Civico	+51	+42	+47	+15.378,00	+24.600,87

Layer Disponibili: Autolinee, Istruzione, Parrocchie, Stradario, Ortofoto_TerraItaly_NR

Città di Como

Selezionare i valori di soglia:

0	≤		<	1428.57
1428.57	≤		<	2857.14
2857.14	≤		<	4285.71
4285.71	≤		<	5714.29
5714.29	≤		<	7142.86
7142.86	≤		<	8571.43
8571.43	≤		<	10000



1.2. Dai Data base di settore al Datawarehouse aziendale

1.2.1. *Il patrimonio informativo comunale “ieri”*

Storicamente ciascun settore comunale ha sempre definito in totale autonomia la tipologia di dati da trattare e conservare nei propri archivi di settore. L'unico *driver* nella scelta di tali tipologie era, ed ancor oggi in molti casi è, il particolare servizio che il settore in oggetto eroga all'utente.

La mancanza di una regia unica ha prodotto negli anni una duplicazione incontrollata di dati personali (e sensibili) all'interno di svariate banche dati comunali.

Questo consolidato e diffuso *modus operandi* ha causato due problemi di non poco conto:

- a) il medesimo individuo/impresa può comparire in modo differente su banche dati differenti (mancanza di controllo incrociato)
- b) l'immotivata e non necessaria memorizzazione di dati sensibili.

Mentre al secondo problema ha messo mano il garante della Privacy, obbligando gli Enti a una profonda rivalutazione delle necessità di memorizzazione di dati sensibili, la differente rappresentazione dello stesso soggetto a seconda della banca dati presa in considerazione è rimasto un problema grave e diffuso in molti Enti pubblici, locali e non.

Il *codice fiscale* ha rappresentato la chiave di volta per dotare tutte le banche dati di un codice identificativo unico ed inequivocabile di riconoscimento personale, anche grazie alla centralità dell'Agenzia delle Entrate che rappresenta l'Ente certificatore unico di tale codice primario.

Il codice fiscale (dove applicato con dovizia) ha risolto solo una parte del problema succitato, ma ha lasciato irrisolta la questione dei dati accessori della persona non contemplati dal codice fiscale stesso, che possono ovviamente permanere *diversi* su banche dati *diverse*.

L'informatizzazione dell'Ente ha rappresentato la *conditio sine qua non* per mettere mano finalmente all'annoso problema.

Infatti non è realisticamente ipotizzabile l'aggancio tra banche dati differenti senza una preventiva informatizzazione delle stesse; solo la trasformazione da banca dati a “*database*” permette l'automatizzazione della correlazione fra campi.

Un esempio classico è rappresentato dalla banca dati Tarsu, che nasceva non correlata alla banca dati dell'Anagrafe. La mancanza di correlazione ha portato nel recente passato ad accumulare notevoli e diffusi problemi relativi al numero civico dell'individuo pagante il tributo, che poteva dunque risultare diverso tra l'archivio Tarsu e Anagrafe.

Dal 2008 la procedura informatizzata che gestisce la Tarsu “pesca” direttamente nel database anagrafico il corretto numero civico. Se non trova il numero civico solleva un allarme, significativo che tale numero civico non è ancora stato inserito in banca dati (p.es: nuova abitazione).

Per risolvere dunque l'annoso problema non vi è altra via che la correlazione delle banche dati comunali nell'ottica di scovare ed eliminare le ridondanze potenzialmente pericolose.

1.2.2. *Il patrimonio informativo comunale “da oggi”*

Siamo alla registrazione di un vero e proprio passaggio culturale: da banche dati di “settore” a “banca dati comunale”.

L'Utente cittadino/impresa si rivolge al proprio Comune per un numero di servizi sempre più ampio e disparato, in stagioni e in età diverse della propria vita civile e lavorativa, lasciando le proprie tracce in un numero sempre più alto di banche dati comunali; è dunque l'Utente cittadino/impresa la “*primary key*” all'interno del patrimonio informativo comunale.

Si provi ad immaginare in quante banche dati comunali possiamo essere registrati come individui nell'arco della nostra vita di cittadino: 1) anagrafe registro nascite; 2) anagrafe residenza; 3) elettorale; 5) politiche educative *Nidi* (se frequentante l'asilo nido comunale); 6) politiche educative *Mense* (se utilizzatore); 7) biblioteca (se utilizzatore); 7) anagrafe stato civile (in caso di matrimonio); 9) Ici (se proprietario di seconda casa); 10) Tarsu; 11) patrimonio (se utilizzatore di beni immobili comunali); 12) servizi sociali

(assistenza domiciliare minori, disabili, anziani, assistenza scolastica, etc.); 13) anagrafe registro morti; 14) banca dati dei cimiteri.

Si consideri, ancora in quante banche dati comunali possiamo essere registrati come lavoratori, artigiani o professionisti: 1. edilizia privata (Dia, permesso di costruire, etc.); 2. Cosap; 3. servizi finanziari (se fornitore); 4. attività produttive.

Ad oggi tutte queste banche dati, ad eccezione dell'Attività Produttive, sono informatizzate e quindi sono *database*, condizione base per poterne pensare una correlazione.

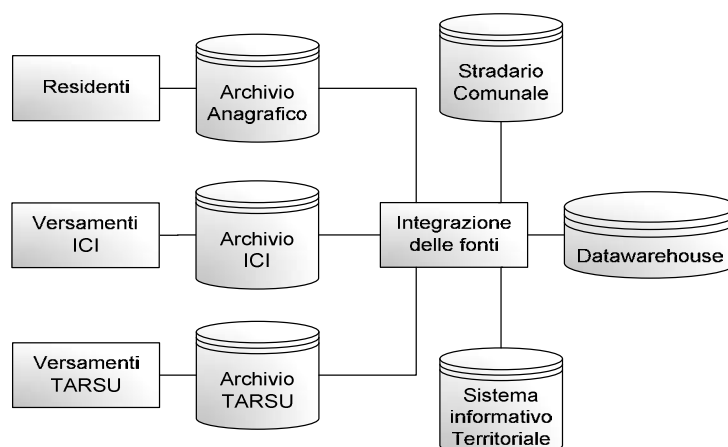
I database attualmente correlati in tempo reale sono: 1) anagrafe registro nascite; 2) anagrafe residenza; 3) elettorale; 4) anagrafe stato civile (in caso di matrimonio); 5) Tarsu; 6) anagrafe registro morti; 7) banca dati cimiteri.

I restanti database possono quindi contenere teoricamente ridondanze incongruenti e pericolose; in attesa dunque di costruire correlazioni automatiche tra i diversi applicativi che gestiscono i diversi database, una via possibile per individuare le discrepanze tra database è quella di provare ad importarli all'interno di un datawarehouse, operazione che infatti impone il confronto tra le diverse rappresentazioni dello stesso dato

1.3. Le banche dati comunali coinvolte

Un Sistema Informativo Statistico si fonda innanzitutto sui dati contenuti al suo interno: la conoscenza delle loro caratteristiche, la loro affidabilità e la loro accuratezza sono elementi fondamentali per la costituzione di un sistema in grado di monitorare in modo efficiente le evoluzioni di fenomeni territoriali locali e di rispondere in modo efficace alle domande che emergono.

All'interno della figura seguente vengono illustrate le principali fonti informative che attualmente contribuiscono alla costruzione del Sic per il Comune di Como.



I dati contenuti negli archivi amministrativi sono principalmente utilizzati per produrre analisi descrittive con particolare attenzione ad analisi longitudinali, che si riescono a produrre in quanto gli archivi sono aggiornati periodicamente e contengono informazioni riferite a cambiamenti nelle unità/entità di analisi (come nascite, morti, migrazioni, ecc.) o nei loro attributi (cambiamento dello stato civile, ecc.); inoltre, poiché i dati sono rilevati in modo costante è possibile seguire le unità nel tempo e quindi ottenere analisi di flusso, anche relativamente a molti anni (analisi di durata dei processi, analisi delle transizioni).

Particolare menzione occorre fare relativamente al collegamento tra il Sic e il S.I.T. (Sistema Informativo Territoriale) di questo Ente, che rende possibile:

1. la visualizzazione su *cartografia ufficiale dell'Ente* di tutte le misure legate al territorio (dirette o calcolate) derivanti dalle banche dati trattate;
2. l'utilizzo di tutti quei "layer" già sviluppati (o di futuro sviluppo) come: *i*) reticolo idrico maggiore e minore; *ii*) rete fognaria; *iii*) percorsi del trasporto pubblico locale; *iv*) quartieri; *v*) zone censuarie; *vi*) circoscrizioni; *vii*) parrocchie.

Per maggiori dettagli sul SIT, consultare il capitolo successivo.

Archivio anagrafico

La scelta dell'*Archivio anagrafico* può essere considerata una scelta obbligata, almeno per quanto concerne l'individuo – cittadino che si relaziona con l'Ente Comune.

Infatti, al di là degli aspetti sanitari (*ma non quelli legati al mondo dell'assistenza sociale*), degli aspetti legati alla sfera lavorativa (*ma non quelli legati all'esercizio commerciale*), e di quelli legati all'istruzione secondaria (*tutta*) e primaria (*tranne gli aspetti relativi a trasporto scolastico e mense*), ogni momento della vita di un individuo – cittadino viene fotografato da una banca dati comunale, e quindi, l'identità della persona (*con attributi classici annessi*) nonché il suo posizionamento sul territorio e le sue relazioni familiari diventano la base insostituibile per qualunque tipo di indagine territoriale.

Nell'ambito di tale Archivio Anagrafico, importanza strategica assume il particolare database della *Toponomastica*, sulla cui integrità e qualità si fonda la maggior parte di tutte le analisi territoriali.

Archivio Ici

La scelta dell'Ici fu fatta per avere una foto della distribuzione sul territorio (e della sua evoluzione negli anni) di uno dei proventi più significativi per l'Ente.

Inoltre, l'Ici rappresenta(va) uno dei possibili riscontri per eventuali indagini tese all'individuazione di falsi casi di "*difficoltà economica*", condizione che frequentemente permette al cittadino di usufruire di agevolazioni tariffarie nell'accesso ad alcuni servizi comunali (asili nido, mense, etc.)

Ma l'analisi di tale banca dati ha rilevato subito un problema di non facile soluzione, rappresentato dal legame che esiste tra il proprietario pagante il tributo in questione e l'ubicazione dell'immobile per il quale viene versato il tributo. Infatti, l'ubicazione dell'immobile non viene effettuata tramite indirizzo (via + numero civico) ma secondo le "coordinate catastali", e cioè sezione censuaria – mappale per i terreni e sezione censuaria – mappale – subalterno per i fabbricati.

Questa peculiarità, associata alla mancanza di *freschezza* e affidabilità del database catastale, impedisce di utilizzare in modo proficuo questo database.

Allo stato attuale questa banca dati ha perso gran parte della sua valenza in seguito all'abolizione dell'Ici "prima casa"; permane ovviamente un interesse "storico" per l'archivio così come composto.

Archivio Tarsu

Contrariamente a quanto si pensasse all'inizio delle attività per la costituzione del Sic, la Tarsu si sta rilevando una banca dati estremamente utile per effettuare una serie di analisi economiche del territorio, in attesa di poter "mettere a sistema" la pleora di dati legati al Settore Attività Produttive che ancora oggi vengono memorizzati in modo non coordinato su semplici fogli Excel.

Difatti, il buon livello di dettaglio delle categorie economiche, quali:

1. abitazioni e locali pertinenziali,
2. aree scoperte adibite a parcheggio,
3. associazioni fondazioni enti morali,
4. autofficine riparaz. parti meccaniche,
5. autolavaggio,
6. autorimesse – autonoleggi,
7. bar – pasticcerie,
8. campeggi e lidi,
9. carcere caserme distretto,
10. centri sportivi palestre e piscine,
11. conventi e convitti comunità ricoveri,
12. grandi magazzini supermercati,
13. importo concordato,
14. laboratori,

15. locali e aree detassate industriali,
16. magazzini – tettoie per depositi,
17. mense – ristoranti,
18. minigolf,
19. negozi,
20. ospedali,
21. pensioni alberghi e locande con alloggio e ostello,
22. scuole circoli ricreativi oratori,
23. serre,
24. teatri cinema e locali di divertimento,
25. uffici e studi,

permette già di visualizzare sulla cartografia la distribuzione di alcune macrocategorie di servizi, quali appunto i negozi, i supermercati, i parcheggi, i ristoranti e gli alberghi.

La banca dati della Tarsu permette inoltre di visualizzare sulla cartografia la distribuzione di alcuni indici importanti per la stesura di un Pgt, quale per esempio il rapporto residenziale/servizi per quartiere e/o zona censuaria e/o circoscrizione.

Gli sviluppi “a brevissimo”

Nel breve periodo si vuole integrare all'interno del Sic anche il database della “Denuncia dei Redditi” proveniente dall'Agenzia delle Entrate, nonché quello delle Politiche Educative (servizio Asili nido e mense); con tale integrazione si potrà dunque procedere ad alcune analisi territoriali relative alla distribuzione della ricchezza nonché ad alcuni controlli relativi alle agevolazioni richieste per i servizi erogati delle Politiche Educative.

1.4. Il Sistema informativo comunale: un lavoro a “tendere”

Il Sistema informativo comunale nasce a prescindere dal Piano di governo del territorio, ma quest'ultimo rappresenta il primo grande obiettivo che il datawarehouse aziendale, opportunamente calibrato, può concorrere a raggiungere.

L'evoluzione di un Sistema informativo comunale non è predicibile nelle sue macroattività di medio periodo, in quanto seguirà e cercherà di soddisfare le necessità informative dell'Ente mano a mano che queste si paleseranno; solo l'obiettivo finale del Sic è prevedibile, anzi, auspicabile: la totale integrazione di tutte le banche dati comunali.

1.5. Le politiche di aggiornamento dell'Ipercubo

L'aggiornamento di un Ipercubo in genere dipende dal grado di necessità di “real time” legata alle analisi che vi si effettuano.

Le politiche di aggiornamento del Sistema informativo comunale sono legate ad esigenze di tipo informativo, al grado di aggiornamento desiderato e all'effettiva disponibilità dei dati e degli archivi.

In particolare, per quanto riguarda le fonti informative attualmente gestite dal Sic del Comune di Como:

- a) l'anagrafe comunale viene aggiornata con cadenza mensile in base alla disponibilità delle informazioni (tale tempistica non è comunque vincolante ma può essere adattata agli adempimenti amministrativi in corso);
- b) la banca dati Ici viene caricata su richiesta ogni qual volta si presenti un aggiornamento;
- c) la banca dati Tarsu viene caricata su richiesta ogni qual volta si presenti un aggiornamento;
- d) la Toponomastica viene caricata su richiesta ogni qual volta si presenti un aggiornamento.

Considerati gli attuali volumi informativi interessati e i tempi di caricamento e trattamento ad essi relativi si è optato per un caricamento totale a fronte di ogni nuova fornitura in modo da assicurare il completo aggiornamento dell'informazione; tecniche di caricamento incrementale avrebbero infatti allungato i tempi di caricamento e complicato il processo di aggiornamento.

1.6. La bonifica delle banche dati

Il medesimo dato (codice fiscale, data di nascita, etc.) spesso è contenuto in più banche dati aziendali e, non di rado, non coincide; la bonifica delle banche dati che trattano in parte lo stesso dato diventa di conseguenza “conditio sine qua non” per la loro integrazione.

Stabilire dunque quale sia la banca dati dove il dato replicato si possa considerare “vero” rappresenta l’attività più complessa e organizzativamente impegnativa nella costruzione dell’Ipercubo.

Per poter dunque garantire i requisiti di qualità necessari, gli archivi amministrativi vengono caricati, trattati ed integrati in modo da disporre di informazioni coerenti ed aggiornate a tutti i diversi livelli di analisi. In particolare, ad oggi, nel corso delle analisi preliminari sono stati individuati *problemi di qualità* riguardanti:

- a) la ricostruzione degli stock annuali dei cittadini del Comune;
- b) la ricostruzione dei nuclei familiari;
- c) la corrispondenza tra lo stradario anagrafico e quello cartografico;
- d) le aggregazioni territoriali utilizzate nello stradario;
- e) la corrispondenza tra lo stradario e gli indirizzi riportati nei versamenti Ici;
- f) la corrispondenza tra lo stradario e gli indirizzi riportati nei versamenti Tarsu.

Per poter migliorare la qualità complessiva delle informazioni vengono quindi svolte, al momento del trattamento, le seguenti attività:

- i) riclassificazione, normalizzazione, deduplica e selezione dei record individuali dell’anagrafe,
- ii) integrazione dei record anagrafici con lo stradario,
- iii) ricostruzione degli stock annuali a partire dall’anno 1990,
- iv) ricostruzione dei nuclei familiari per l’anno corrente,
- v) integrazione dei record familiari con lo stradario,
- vi) riclassificazione, normalizzazione, deduplica e selezione dei record Ici,
- vii) riclassificazione, normalizzazione, deduplica e selezione dei record Tarsu,
- viii) integrazione dei record Tarsu con lo stradario,
- ix) integrazione dei record individuali con i record Tarsu e Ici,
- x) integrazione dei record familiari con i record Tarsu e Ici.

Di seguito si riporta un esempio dei principali trattamenti attualmente operati sulle banche dati fornite, dettagliato con le numerosità relative all’ultima iterazione eseguita.

Trattamento dei record individuali dell’anagrafe

- a) caricamento della fornitura: caricamento selettivo dei dati forniti dall’anagrafe e riconduzione al tracciato di caricamento standard; generazione dei metadati e normalizzazione del tracciato (*210.180 record interessati*);
- b) eliminazione dei soggetti non residenti: eliminazione dei soggetti catalogati come non residenti dal dataset (*87.120 soggetti eliminati, 123.060 soggetti rimanenti*);
- c) aggiornamento della condizione di cittadino: verifica della valorizzazione del campo e correzione dei casi errati o mancanti;
- d) aggiornamento degli identificatori del coniuge, di paternità e di maternità: verifica della valorizzazione dei campi e correzione dei casi errati o mancanti;
- e) deduplica dei record individuali: deduplica logica in base all’identificatore individuale dopo quella fisica in fase di caricamento (*122.783 soggetti rimanenti*);
- f) ricostruzione degli stock annuali dal 1990 al 2009: ricostruzione degli stock annuali in base alle date di ingresso e uscita dalla banca dati (*1.447.766 record complessivi*);
- g) aggiornamento degli indirizzi e correlazione alla banca dati territoriale: l’integrazione del dato territoriale è condizionata dalla corrispondenza tra l’indirizzo anagrafico e il corrispondente identificativo geografico, avvenuto secondo le seguenti modalità:
 - i. indirizzi georeferenziati: 11.132,
 - ii. indirizzi stradario anagrafico: 12.437,
 - iii. record corrispondenti: 9.332,

- iv. *record presenti nello stradario anagrafico ma non georeferenziati: 3.094,*
- v. *record georeferenziati ma non presenti nello stradario: 1.816.*

Aggregazione dei record familiari (l'aggregazione è possibile per il solo anno 2008)

- a) *aggregazione delle famiglie: aggregazione selettiva dei record anagrafici in base al numero di famiglia (37.711 famiglie, relative al solo anno 2008 poiché il dato non è storicizzato);*
- b) *calcolo del capo famiglia: definizione delle composizioni familiari e identificazione dei capo famiglia;*
- c) *calcolo delle variabili familiari: calcolo di indicatori aggiuntivi riferiti a ciascun nucleo familiare (numero componenti, figli, ecc.).*

Aggiornamento dei record individuali dell'anagrafe

- a) *aggiornamento del tipo di famiglia calcolato al passaggio precedente,*
- b) *aggiornamento della nazionalità della famiglia calcolata al passaggio precedente.*

Trattamento e integrazione record Tarsu

- a) *caricamento della fornitura: caricamento selettivo dei dati forniti e riconduzione al tracciato di caricamento standard; generazione dei metadati e normalizzazione del tracciato (393.491 record interessati);*
- b) *aggiornamento degli identificatori individuali: verifica della valorizzazione del campo e correzione dei casi errati o mancanti;*
- c) *deduplica dei record: deduplica logica del dataset (329.470 record rimanenti);*
- d) *aggregazione degli individui – Tarsu: distribuzione del dataset sugli individui e integrazione con i record anagrafici (303.916 record individuali interessati nei diversi stock annuali);*
- e) *aggregazione delle famiglie – Tarsu: distribuzione del dataset sulle famiglie e integrazione con i record familiari (31.919 record interessati nell'anno 2008).*

Trattamento e integrazione record Ici

- a) *caricamento della fornitura: caricamento selettivo dei dati forniti e riconduzione al tracciato di caricamento standard; generazione dei metadati e normalizzazione del tracciato (254.840 record interessati);*
- b) *deduplica record: deduplica logica del dataset (138.974 record rimanenti);*
- c) *aggregazione individui – Ici: distribuzione del dataset sugli individui e integrazione con i record anagrafici (60.334 record individuali interessati nei diversi stock annuali);*
- d) *aggregazione famiglie – Ici: distribuzione del dataset sulle famiglie e integrazione con i record familiari (nessun record interessato per l'anno 2008 in quanto non presente in fornitura).*

Generazione delle strutture di analisi

- a) *individui (1.447.766 record distribuiti dal 1990 al 2009),*
- b) *famiglie (37.711 record nell'anno 2008),*
- c) *Tarsu (329.470 record distribuiti dal 2003 al 2009),*
- d) *Ici (138.974 record distribuiti dal 1993 al 2005).*

I trattamenti sopra descritti hanno permesso di individuare alcuni problemi di non poca rilevanza sulle banche dati, che di seguito vengono riportati per evidenziare l'estrema importanza dell'integrazione fra banche dati nell'ottica del consolidamento e del miglioramento della qualità del "dato comunale".

Lo stradario comunale

Ad oggi esistono tre diverse banche dati che trattano la toponomastica del Comune di Como:

- a) *la "Gestione Territorio" dell'Anagrafe,*
- b) *il DB Access "temporaneo" dell'Anagrafe,*
- c) *il DB Cartografico Istat 2001,*

d) Il DBT (Appalto R.A.T.I. ancora in essere) in fase di validazione.

La semplice elencazione delle banche dati sopraesposta evidenzia di per sè una situazione “di transizione” per questo tipo di dato, che risulta assolutamente fondamentale per legare gli individui al territorio.

Un primo esempio di problema individuato è quella relativa alla georeferenziazione dei numeri civici; a valle di un certosino lavoro di confronto e deduplicazione questo è il risultato ottenuto: *i*) indirizzi georeferenziati: 14.572 (dopo la deduplica), *ii*) indirizzi “Gestione Territorio”: 12.455, *iii*) record corrispondenti: 11.459, *iv*) record presenti nella “Gestione Territorio” ma non georeferenziati: 996.

Ergo, rimangono solo 996 numeri civici per i quali manca la corrispondenza con una coordinata geografica, e l’Ufficio competente è colmerà tale lacuna nel brevissimo.

La Tarsu non pagata

Da una prima analisi effettuata incrociando Tarsu e Anagrafe, per l’anno 2008 risulta che al 15,4% delle famiglie comasche (31.919 su 37.711) non risulta associato nessun pagamento Tarsu.

Attivato dunque l’ufficio competente, ad oggi, dopo alcune prove a campione, si sono individuate alcune casistiche di *falsi positivi* come: *i*) a ruolo il coniuge deceduto; *ii*) a ruolo a nome del padre xxx yyy, emigrato; *iii*) regolare a ruolo, ancorché il contribuente abbia presentato denuncia di cancellazione per trasferimento all’estero e, anagraficamente, al momento risulti ancora residente;

Rimane, ovviamente, una nutrita serie di casi al vaglio degli uffici competenti.

1.7. La struttura del Sistema informativo comunale

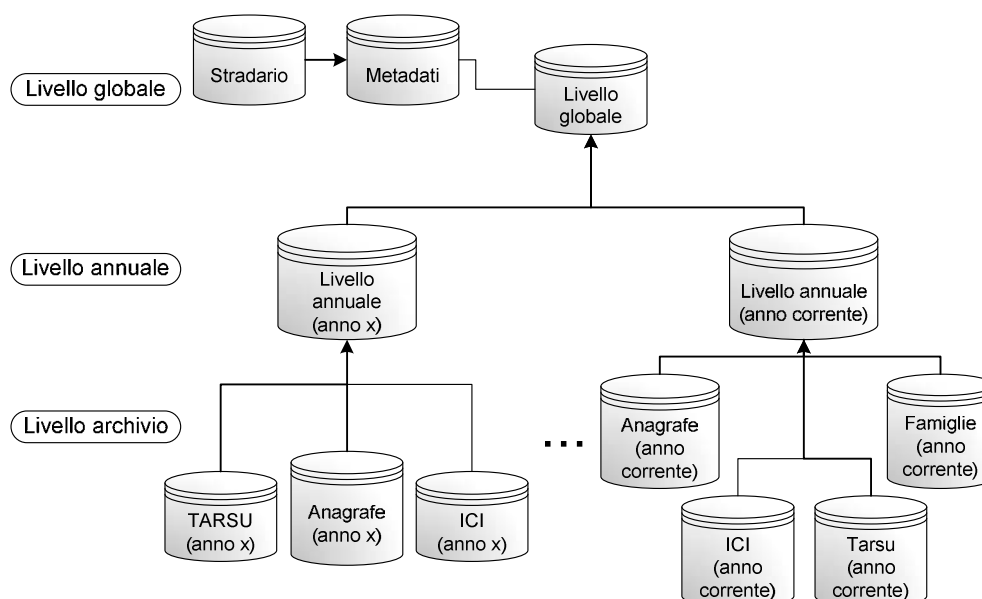
Perché un Sistema informativo comunale risponda appieno alle esigenze espresse è necessario individuare soluzioni tecnologiche in grado di offrire elevata configurabilità, affidabilità e flessibilità nell’introduzione di nuove funzionalità personalizzate nel tempo.

Questo paragrafo descrive la soluzione architeturale studiata e realizzata, già utilizzata anche in altri contesti, di studio e di ricerca, a seguito di sviluppi e adattamenti effettuati. La soluzione, posizionata nell’ambito dei sistemi di “Business Intelligence”, è costituita da un datawarehouse composto da moduli software realizzati in ambiente open source .

Architettura del datawarehouse

Il datawarehouse è la banca dati su cui poggia l’intero Sistema informativo comunale: la sua corretta progettazione e realizzazione è fondamentale per poter ottenere i risultati programmati.

Il datawarehouse è in grado di coniugare le caratteristiche eterogenee delle forniture che concorrono a popolarlo, garantendo nel contempo prestazioni ottimali e disponibilità delle informazioni.



Per raggiungere tali obiettivi il datawarehouse è progettato su tre livelli distinti, come rappresentato in figura precedente

- un livello archivio contenente le singole forniture (fonti amministrative) caricate nel datawarehouse che possono essere gestite, trattate ed analizzate singolarmente;
- un livello annuale in cui viene realizzato il primo livello di integrazione suddividendo le informazioni in base al periodo di osservazione; tale livello mantiene sia informazioni di flusso sia di stock riguardanti l'anno di interesse;
- un livello globale che integra tutte le informazioni disponibili e offre una visione completa dell'intera base di dati.

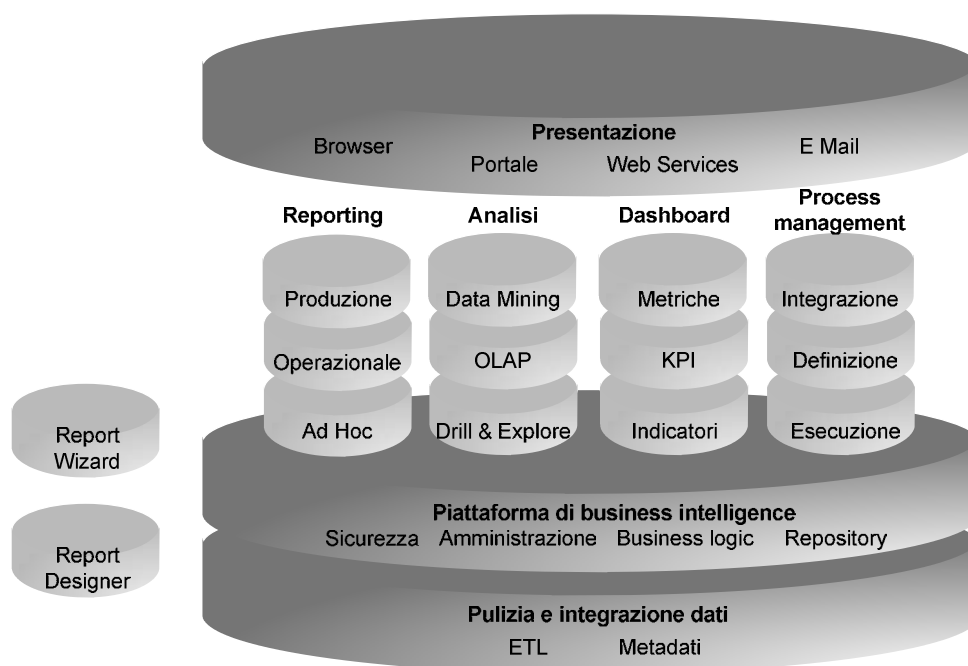
Oltre ai livelli descritti sono stati realizzati, per rispondere a specifiche esigenze di analisi, alcuni *datamart* contenenti sottoinsiemi informativi di particolare interesse, che consentono di selezionare l'informazione in base a un periodo di osservazione o a particolari caratteristiche richieste o ancora operando al suo interno particolari trattamenti.

Il datawarehouse contiene inoltre le strutture finalizzate alla gestione dei metadati e dei sistemi classificatori utilizzati. Essi consentono di mettere in relazione archivi che riportano contenuti analoghi (ad esempio la nazionalità dei cittadini) ma classificate in modo differente, per mezzo di appositi sistemi di transcodifica che riportano per ciascuna voce il termine corrispondente adottato nelle altre tassonomie, analogamente a quanto avviene per i sistemi di transcodifica e traduzione linguistica.

Componenti dell'architettura complessiva

L'architettura del sistema è articolata in tre livelli:

- livello dati*: si occupa di qualità, pulizia e integrazione dei dati e contiene anche gli strumenti finalizzati all'estrazione, al trattamento e al caricamento dei dati e dei metadati;
- livello applicativo*: contiene le applicazioni di gestione dei contenuti (sicurezza, amministrazione, repository), i componenti di gestione dei processi (motore di workflow) e i moduli finalizzati alla realizzazione delle analisi (motore *Olap* (On Line Analytical Processing), navigazione multidimensionale, reporting statico e dinamico, query on line, cruscotti, indicatori (KPI – Key Performance Indicator, ...));
- livello presentazione*: gestisce le modalità di visualizzazione e trasmissione dell'informazione per mezzo dell'interfaccia web.



Nel seguito vengono descritte, in estrema sintesi, le principali funzionalità che una soluzione come quella identificata rende disponibili.

Analisi dati

Le funzionalità di analisi consentono di:

- a) semplificare l'esplorazione dell'informazione attraverso operazioni di selezione di variabili di interesse e l'applicazione sulle stesse di "filtri" attivati con l'ausilio di funzioni parametrizzate; confrontare (incrociare) i dati e approfondire il livello di analisi in base ai fatti e alle dimensioni definite;
- b) fornire rapide risposte a interrogazioni complesse permettendone la definizione dinamica senza la necessità di conoscere il codice specifico;
- c) offrire funzionalità avanzate di elaborazione di rapporti grazie all'integrazione realizzata dalla piattaforma di datawarehouse, con specifici strumenti di reporting;
- d) consentire l'esportazione dei risultati delle analisi nei formati più diffusi, il loro salvataggio per un futuro utilizzo e la loro condivisione.

Tutto ciò è possibile grazie alla visualizzazione dei dati in cubi multidimensionali e alla possibilità di navigare il loro contenuto in modo dinamico selezionando di volta in volta le informazioni di interesse e le modalità di visualizzazione. Ciò consente di approfondire il livello di analisi entrando nel dettaglio senza dover avere a che fare con l'intera popolazione e selezionando solo la porzione di interesse in modo dinamico a seconda delle esigenze del momento.

Cruscotti direzionali

I cruscotti direzionali visualizzano in modo sintetico ed efficace informazioni di particolare interesse per l'utente di alto livello che non ha la possibilità di lavorare direttamente sui dati ma deve poter ricavare in poco tempo le informazioni di proprio interesse.

Georeferenziazione

La georeferenziazione è una modalità di visualizzazione delle informazioni particolarmente efficace che consente di collocare sul territorio le informazioni di interesse, accedendovi per mezzo della mappa interattiva: una volta individuato un punto di interesse è possibile richiedere ulteriori informazioni, fornite per mezzo di grafici e report contestualizzati.

Reporting

I report consentono di sintetizzare e visualizzare i risultati delle analisi fornendone una visione completa e omogenea. I report possono essere supportati da grafici al loro interno e possono essere di tipo statico (definiti a priori in fase di progettazione) o dinamico (riguardanti cioè informazioni selezionate direttamente dall'utente secondo criteri specifici); il sistema prevede anche una funzionalità di query on line che permette l'accesso all'intera banca dati, facilitando la selezione degli attributi di interesse, l'applicazione di filtri e la progettazione delle modalità di visualizzazione.

Estrazione trasformazione e caricamento

Le procedure di estrazione, trasformazione e caricamento sono spesso particolarmente complesse ed è necessario poterle progettare in modo attento, facilitando il riutilizzo di componenti e procedure precedentemente implementate ma anche l'integrazione rapida di nuovi trattamenti e consentendo di poter eseguire processi articolati e di monitorarne i risultati. L'utilizzo di appositi tool visuali consente di:

- a) accrescere la produttività grazie a un'architettura estensibile e basata su standard tecnologici che consente di evitare l'adozione di metodologie proprietarie nell'applicazione dei processi di ETL;
- b) creare complesse trasformazioni in un ambiente grafico senza dover generare alcun tipo di codice;
- c) utilizzare un approccio repository based che consenta il riutilizzo delle interrogazioni, trasformazioni e componenti e contemporaneamente la gestione dei modelli, delle connessioni e dei log;
- d) utilizzare sorgenti dati eterogenee, incluse le più comuni applicazioni, piattaforme di database, sia open source sia proprietari, flat file e documenti Excel

Modelli di analisi multidimensionale

I modelli descritti nel seguito sono stati progettati e realizzati per il funzionamento all'interno della piattaforma.

Strutture fisiche

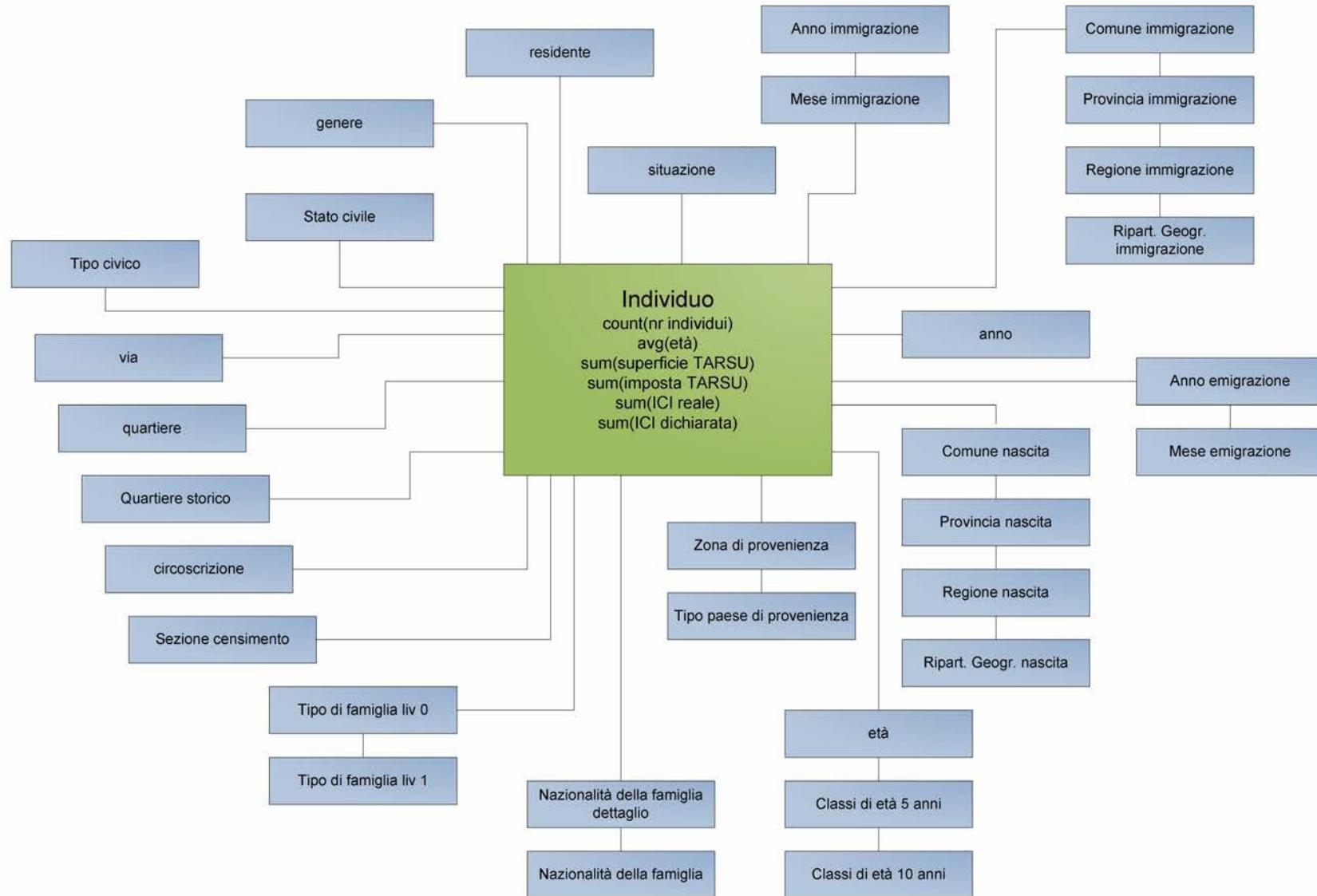
Attualmente i modelli multidimensionali sono basati su 4 tabelle fisiche e 3 viste:

- a) Tabelle: Dw_famiglia, Dw_ici, Dw_individuo, Dw_tarsu
- b) Viste:
 - Dw_ici_corrprec (selezione di dw_ici limitata agli ultimi due anni disponibili)
 - Dw_individui_stranieri (selezione di dw_individuo)
 - Dw_tarsu_corrprec (selezione di dw_tarsu limitata agli ultimi due anni disponibili)

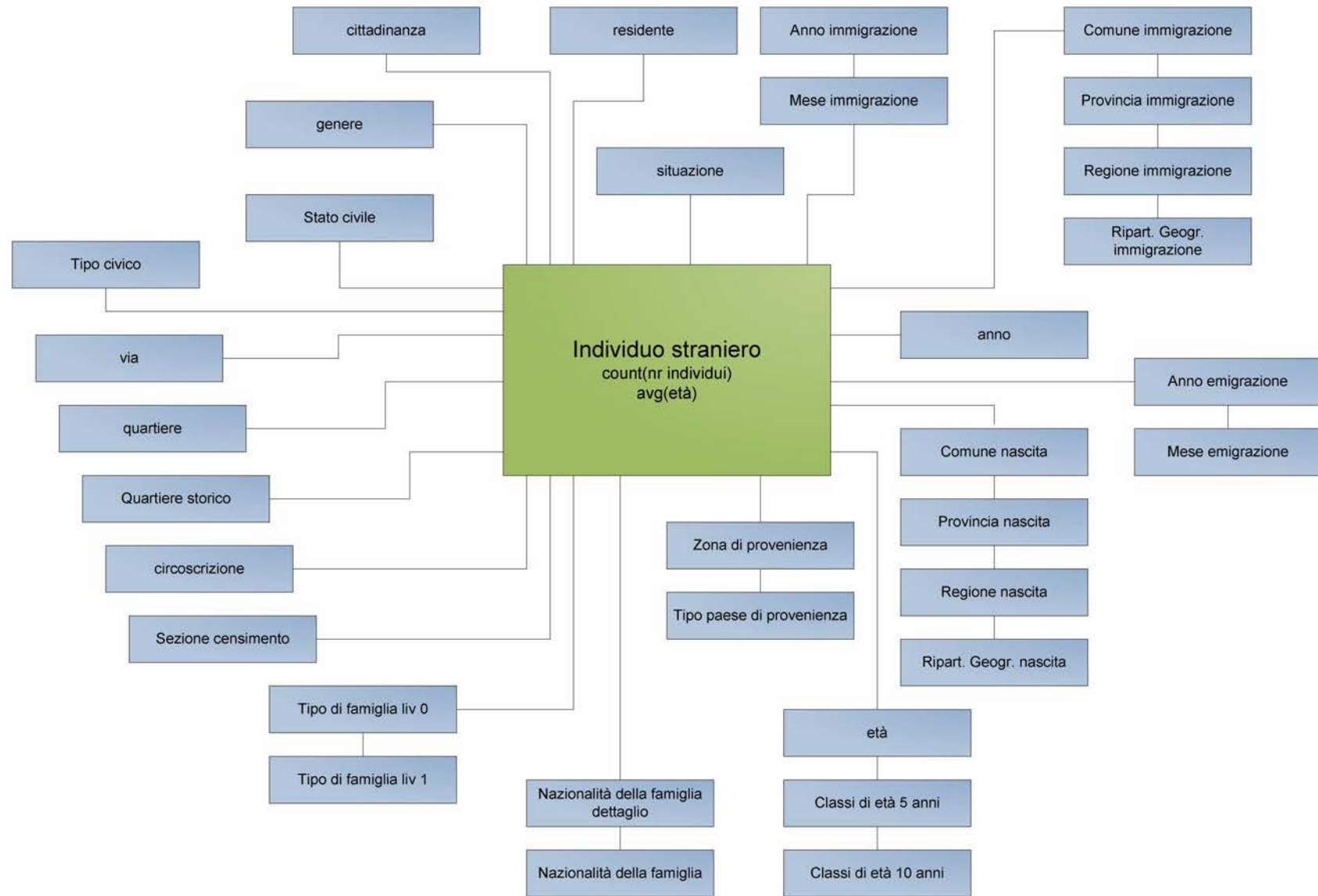
Modelli multidimensionali

Sulle tabelle descritte in precedenza sono stati progettati attualmente sette modelli multidimensionali descritti nel seguito.

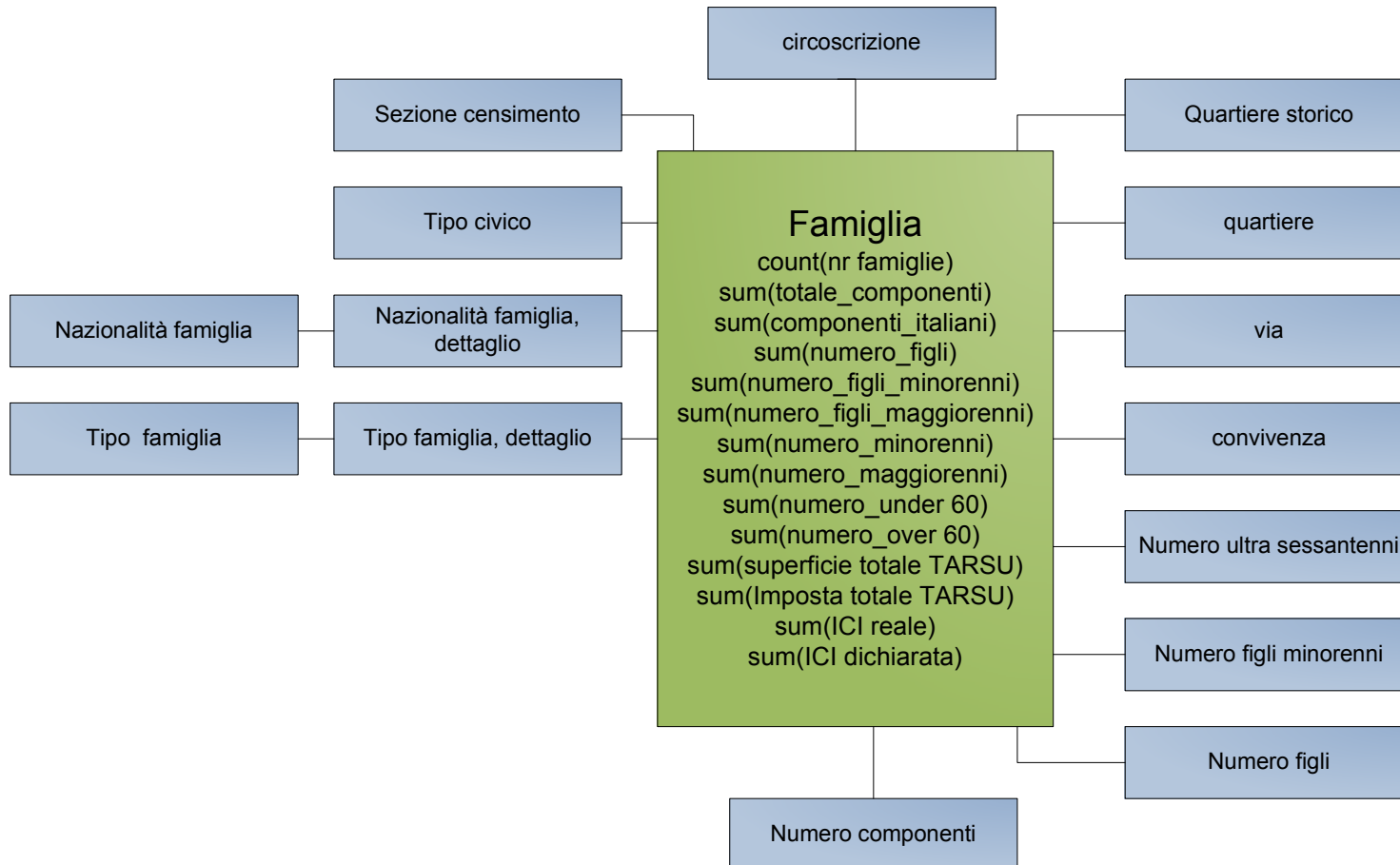
Individui

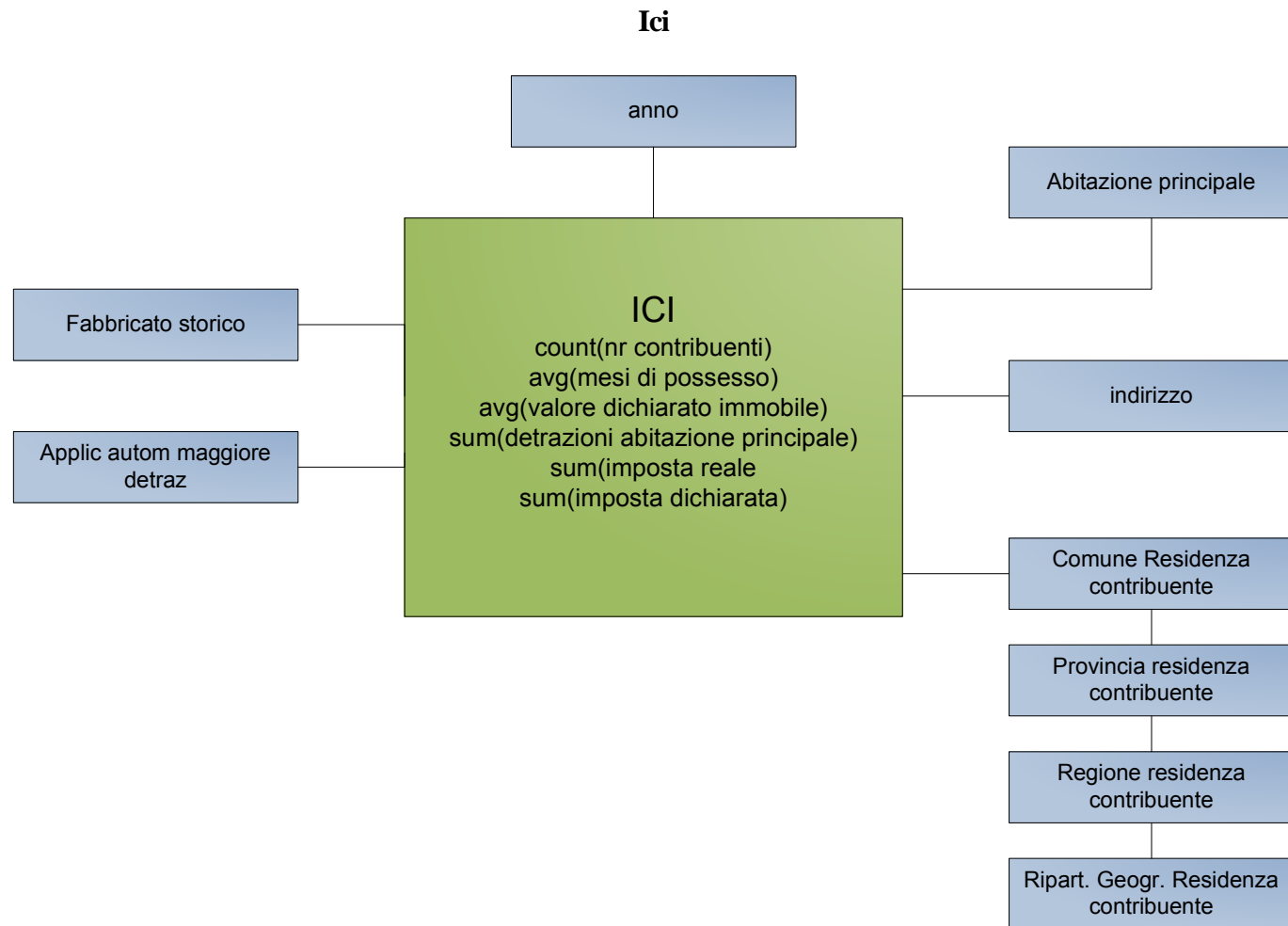


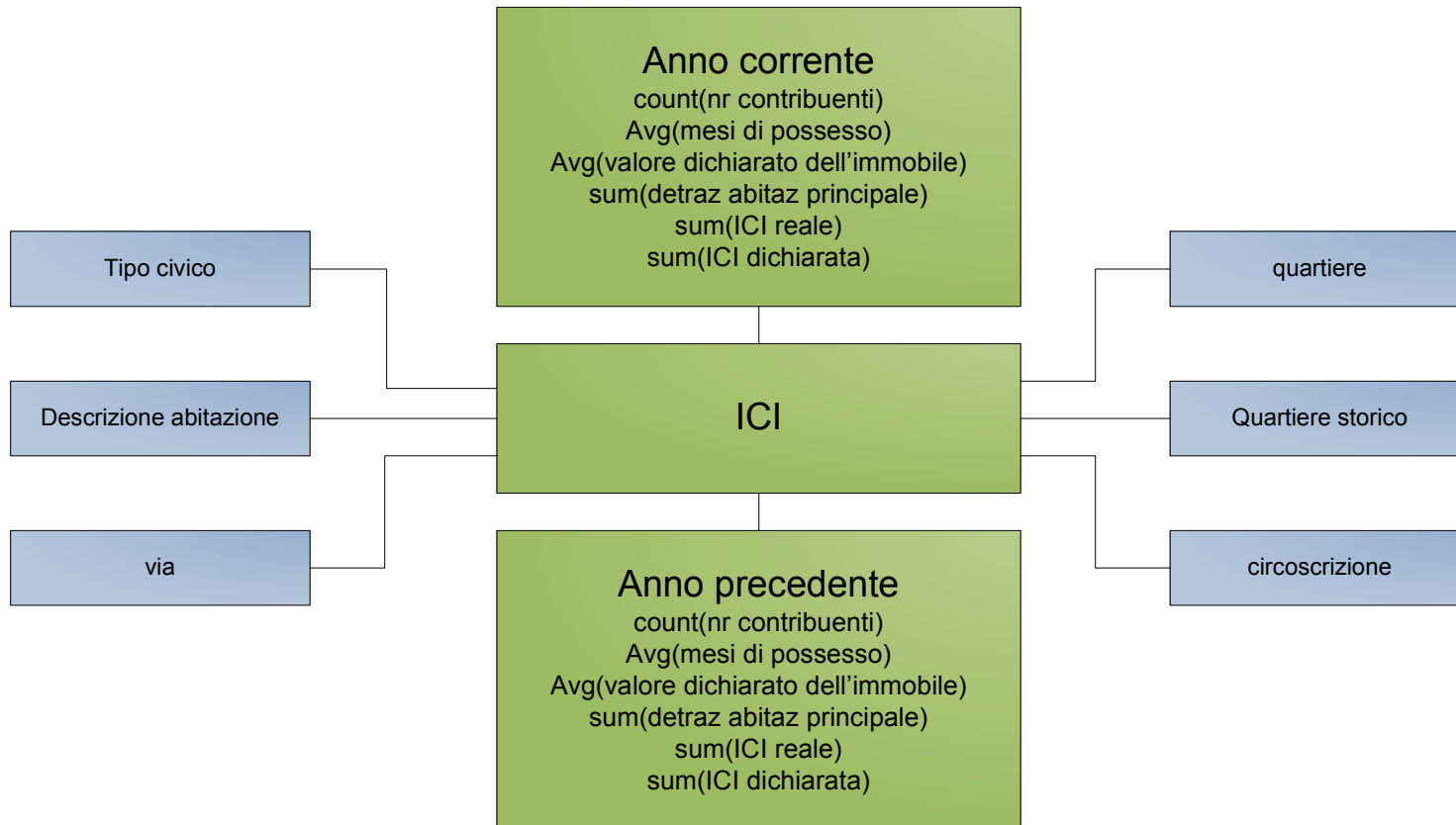
Individui stranieri

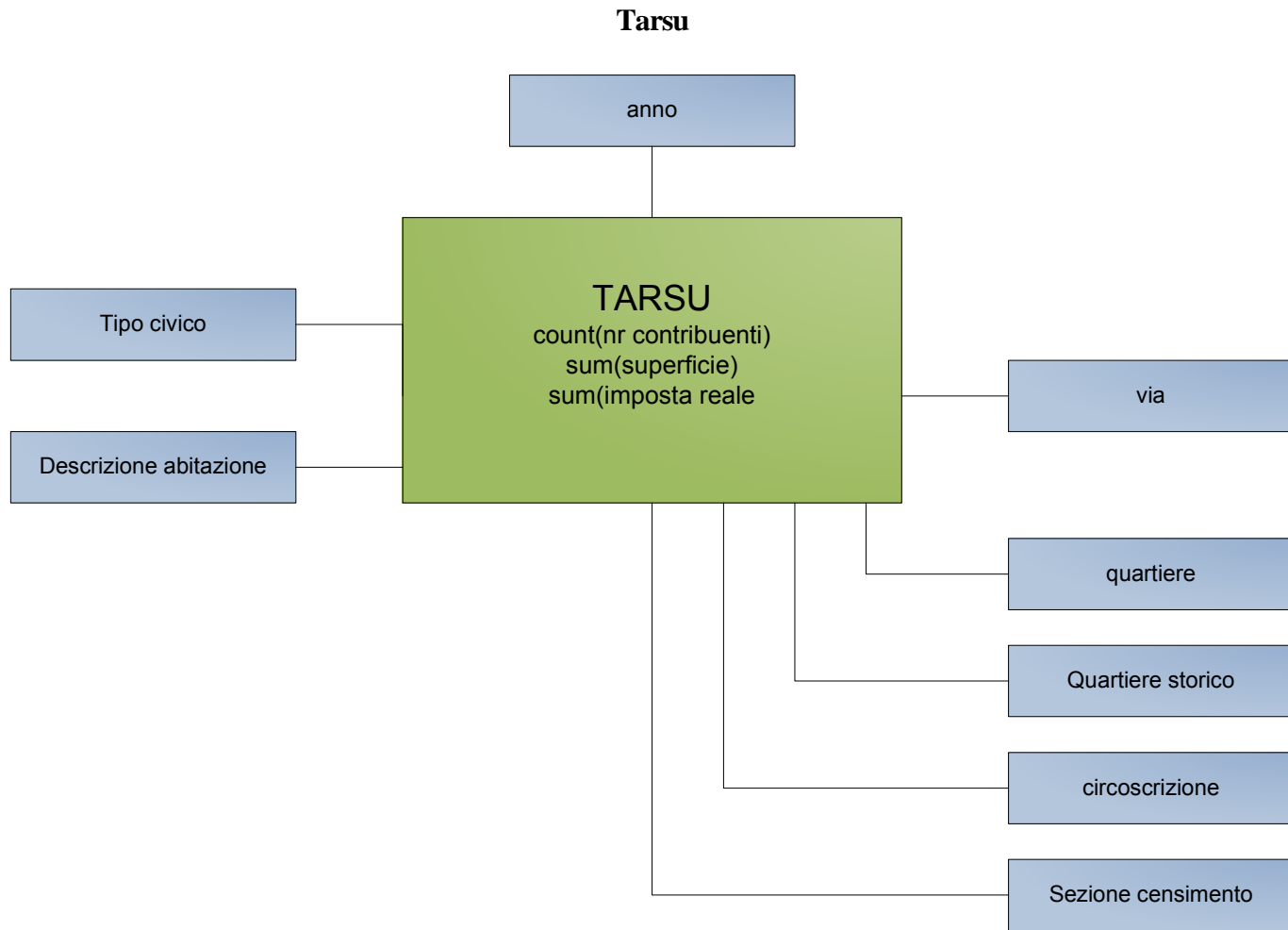


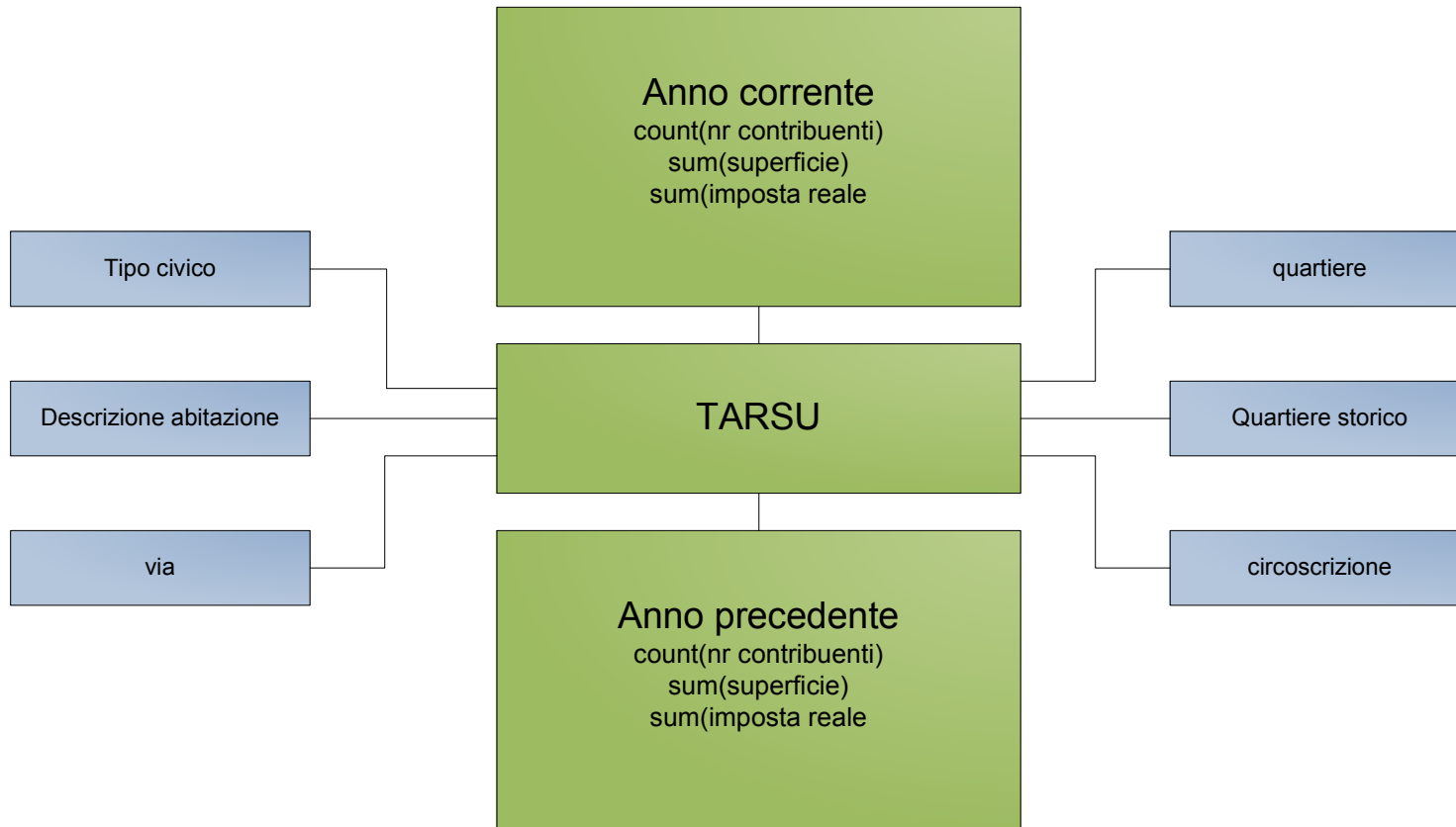
Famiglie





Ici (ultimi due anni)



Tarsu (ultimi due anni)

Infrastruttura hardware

Il Sic dal punto di vista architetturale è attualmente distribuito su due server virtuali contenuti:

- a) Il datawarehouse e più in generale il data layer (processi di ETL e integrazione dei dati);
- b) La piattaforma applicativa di Business Intelligence.

Infrastruttura software

Il Sic è interamente costruito sfruttando componenti open source di larga diffusione in ambito applicativo. In particolare i diversi layer sono costituiti dalle seguenti componenti:

- a) Data layer
 - MySQL (DBMS e datawarehouse) www.mysql.com
 - Talend (ETL) www.talend.com
- b) Application layer
 - Jboss (application server) www.jboss.org
 - Pentaho (BI platform) www.pentaho.org
 - Pentaho Reporting (reporting) reporting.pentaho.org
 - Mondrian (motore multidimensionale) mondrian.pentaho.org
 - JFreeChart (grafici e dashboard) www.jfree.org/jfreechart
 - ArcGIS Javascript API (georeferenziazione) www.esri.com/software/arcgis/arcgisonline/web_mapping_apis.html

1.8. Riferimenti bibliografici

- Barrow R., 2004, *OECD short term economic statistics export group – Using administrative data in short – term statistic: sub annual industry surveys at statistics New Zealand*, New Zealand.
- Brackfield D., 2005, *OECD short – term economic statistics expert group. Task force on administrative data: progress report*, OECD OCDE, Francia.
- Buzzigoli L., Innocenti R. e Martelli C., 2002, “La statistica a supporto della governance a livello locale: il caso del comune di Firenze”.
- Calzaroni M., Giovannini E. e Sorce A., 2001, *Il Sistema Informativo Statistico sulle imprese dell’Istat: problematiche e potenzialità*, Roma.
- Eurostat, Gennaio 1997, *Proceedings of the seminar on the use of Administrative Sources for Statistical Purposes*, Lussemburgo.
- Eurostat, 2002, “The use of administrative sources”, in *Business register recommendation manual*.
- Filippucci C., 2002, *Strategie e modelli per il controllo e la qualità dei dati*, FrancoAngeli, Milano.
- Graziani C., 2008, *Sistemi informativi statistici: metodi e modelli di progettazione*, Università di Milano Bicocca.
- Martelli C., 2004a, *Sistemi informativi statistici a supporto dello studio e l’analisi dei distretti rurali*, Firenze.
- Martelli C., 2004b, “Uso statistico dei dati amministrativi a livello locale: problemi e prospettive”, in Aa. Vv., *Strategie di informazione dei dati statistici e analisi economiche per il governo locale*, FrancoAngeli, Milano.
- Martini M., 1999, “Censimenti, dati amministrativi e Sistema Statistico delle imprese basato sui registri”, in *Atti del convegno della Società Italiana di Statistica*, Udine.
- Martini M., 2000, “L’integrazione statistica di registri e indagini”, in Aa. Vv., *Tecnologie informatiche e fonti amministrative nella produzione di dati*, FrancoAngeli, Milano.
- Mezzanzanica M. e Lovaglio P., 2008, *Numeri al lavoro. Il sistema statistico del mercato del lavoro: metodologie e modelli di analisi*, FrancoAngeli, Milano.
- Myrskylä P., 1999, “New statistics made possible by the use of registers”, *Statistical Journal of the United Nations*.
- Pemper R.J., 1997, *Compilation and presentation of labour statistics based on administrative records*, Bangkok.