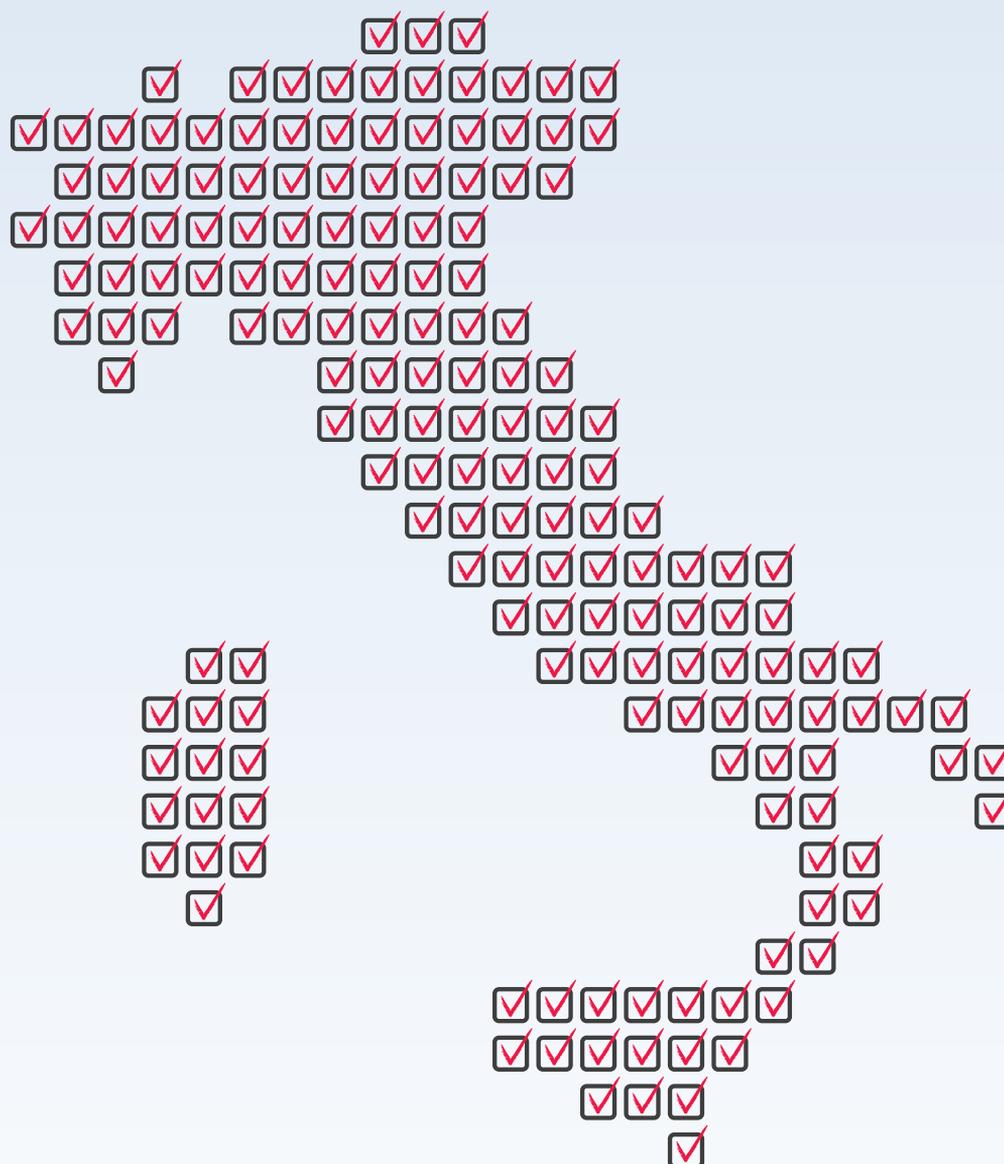


# Atti del 15° Censimento generale della popolazione e delle abitazioni



## 5 I sistemi informatici a supporto delle operazioni censuarie

*Atti a cura di:* Giuseppe Sindoni  
Fascicolo a cura di: Marina Venturi

Atti del 15° Censimento generale della popolazione e delle abitazioni  
5 - I SISTEMI INFORMATICI A SUPPORTO DELLE OPERAZIONI CENSUARIE

© 2016  
Istituto nazionale di statistica  
Via Cesare Balbo, 16 - Roma

978-88-458-1878-3 (elettronico)  
978-88-458-1877-6 (stampa)

Salvo diversa indicazione la riproduzione è libera,  
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),  
marchi registrati e altri contenuti di proprietà di terzi  
appartengono ai rispettivi proprietari e  
non possono essere riprodotti senza il loro consenso.



# Indice

	Pagina
<b>Presentazione dell'opera</b>	7
<b>Il 15° Censimento Generale della Popolazione e delle Abitazioni</b>	9
<b>Introduzione</b>	13
<b>Capitolo 1 - Il sistema di acquisizione delle liste anagrafiche comunali</b>	17
1.1 L'acquisizione delle Lac	17
1.2 Il sistema di controllo delle Lac (Traslac1)	20
1.3 Il sistema di controllo delle variazioni anagrafiche (Traslac2)	22
1.4 Il supporto ai comuni	25
<b>Capitolo 2 - Il sistema della rilevazione dei numeri civici</b>	27
2.1 L'architettura generale	27
2.1.1 Login	29
2.1.2 Pagina personale	29
2.1.3 Home page e funzionalità del menu	30
2.1.4 Gestione stradario	32
2.1.5 Gestione richieste toponimo	35
2.1.6 Upload files	36
2.1.7 Monitoraggio	37
2.1.8 Logout	38
<b>Capitolo 3 - L'architettura del sistema integrato di acquisizione dati e gestione della rilevazione</b>	39
3.1 L'architettura generale	39
3.2 Metodologia e strumenti di sviluppo software	41
3.3 L'architettura software del portale della rete	42
3.4 La base dati	43
3.4.1 La base di dati Sgr: struttura, popolamento e gestione	43
3.4.2 Amministrazione e tuning della base dati	45
3.5 Le viste materializzate	46
3.6 Test e prestazioni del sistema	48
<b>Capitolo 4 - Il sistema di acquisizione</b>	53
4.1 Caratteristiche generali	53
4.2 La modellazione del questionario	54
4.2.1 La modellazione della struttura del questionario	54
4.2.2 La modellazione dei percorsi di compilazione e il grafo del questionario	56
4.2.3 Come usare il grafo del questionario	59
4.3 Il Progetto del sistema	62
4.3.1 L'architettura	62
4.3.2 Le tipologie di domande	63
4.3.3 I percorsi di compilazione	66
4.3.4 I controlli	66
<b>Capitolo 5 - Il sistema di gestione della rilevazione</b>	69
5.1 Caratteristiche generali del sistema	69
5.2. Gestione della rete e del territorio	71
5.2.1 Le tipologie utente	71

	Pagina
5.2.2 L'organizzazione del lavoro	73
5.3 Gestione delle operazioni sul campo	76
5.4 I rapporti riassuntivi di monitoraggio	81
5.5 La raccolta dei microdati	83
5.5.1 I modelli riepilogativi	84
5.5.2 Il censimento degli edifici	85
5.5.3 Le minoranze linguistiche	86
5.6 Le operazioni di chiusura e invio materiale compilato	86
5.6.1 Le modalità operative	86
5.6.2 I controlli quantitativi sul ritorno dei questionari	87
5.6.3 La gestione delle eccezioni	88
5.7 Strumenti di comunicazione con la rete	89
5.7.1 Il portale documentale	89
5.7.2 Attività di helpdesk	90
5.7.3 Fornitura di dati di monitoraggio off line	91
<b>Capitolo 6 - Il sistema per il confronto tra censimento e anagrafe</b>	<b>93</b>
6.1 Gestione delle variazioni anagrafiche interattive	93
6.2 Le operazioni dei comuni	95
6.3 I prodotti del confronto	98
<b>Capitolo 7 - I sistemi per l'acquisizione, il monitoraggio e la verifica della qualità della registrazione e consultazione di dati e immagini</b>	<b>101</b>
7.1 La fornitura di dati e immagini	101
7.2 Le fasi del controllo di qualità della registrazione	102
7.3 Il sistema di acquisizione e monitoraggio della registrazione (Samor-Pop)	102
7.3.1 I controlli previsti dal sistema	103
7.3.2 L'architettura del sistema	105
7.3.3 Il modello dei dati	106
7.4 Il sistema di catalogazione delle immagini	107
7.4.1 Il sistema di controllo dell'injection sul Sistema Gestione Immagini	109
7.5 Le verifiche preliminari di completezza	110
7.6 I controlli di qualità a campione	110
7.6.1 La procedura software All-In-One	111
7.6.2 Le operazioni di controllo	113
7.6.3 Le linee guida della procedura di controllo	113
7.6.3.1 <i>La leggibilità dell'immagine</i>	114
7.6.3.2 <i>La verifica dei dati</i>	114
7.6.4 Il Sistema Samor-Cqc	116
7.7 Elementi di innovazione e punti di forza dell'intero processo	117
7.8 Corinto, un sistema software per la gestione delle immagini	118
7.8.1 L'architettura del sistema	119
7.8.2 Gestione delle immagini	119
7.8.2.1 <i>Recupero file da Documentum</i>	121
7.8.2.2 <i>Recupero file da File-System remoto.</i>	121
7.8.3 Utenti e funzionalità	121
7.8.4 Ricerca questionari	122
7.8.5 Ulteriori utilizzi	122
<b>Capitolo 8 - Il sistema di validazione dei dati</b>	<b>125</b>
8.1 I requisiti statistici	127
8.2 Il controllo del processo	130
8.3 L'interfaccia utente	131

	Pagina
8.4 L'utilizzo del sistema di gestione delle immagini	133
<b>Capitolo 9 - Il sistema di rendicontazione delle spese censuarie</b>	<b>135</b>
9.1 Modalità di accesso e di utilizzo del sistema	135
9.2 La struttura del modello elettronico	136
9.3 Le funzioni di controllo di invio e di supporto	138
9.4 Supporto ai comuni per il calcolo del contributo variabile	139
<b>Capitolo 10 - I sistemi di diffusione dei dati</b>	<b>141</b>
10.1 Caratteristiche generali	141
10.2 Architettura dei microdati	143
10.2.1 Etl di caricamento e riconciliazione	144
10.3 Architettura dei macrodati	145
10.3.1 Il modello Multidimensionale: il Data Warehouse Primario	145
10.3.2 I Fatti e le Dimensioni	147
10.3.3 Etl di caricamento delle tabelle dei fatti	150
10.4 I servizi offerti	150
10.4.1 I servizi offerti dal Data Warehouse Primario	150
10.4.2 La reportistica	151
10.4.3 Self-Service Business Intelligence	153
10.4.4 Procedure per l'estrazione dei dati	155
10.5 La progettazione dei cubi dati per la diffusione	157
10.5.1 Le fasi del processo, le logiche e la metodologia seguite	157
10.5.2 Caso 1: Famiglie per numero di componenti – livello comunale	158
10.5.3 Caso 2: Famiglie per numero di componenti e situazione abitativa della famiglia – livello provinciale	161
10.5.4 Caso 3: Stanze in abitazioni occupate da persone residenti per numero di stanze e numero di stanze ad uso professionale – livello provinciale	163
10.5.5 Considerazioni conclusive	165
10.6 Il data warehouse di diffusione	166
10.6.1 L'architettura tecnologica del data warehouse di diffusione	166
10.6.2 L'ambiente di navigazione	167
10.6.3 L'ambiente di amministrazione ed il caricamento dei dati	169
10.7 - Il Single Exit Point per la diffusione dati machine-to-machine	170
10.7.1 L'architettura tecnologica ed il workflow	170
10.7.2 La modellazione dei cubi e la mappatura Sdmx	172
10.8 La disponibilità di microdati per gli utenti esterni	172
10.8.1 La comunicazione dei microdati agli enti Sistan	173
10.8.2 L'accesso ai microdati attraverso il Laboratorio Adele	173
10.8.3 I file di microdati disponibili per il 15° Censimento della popolazione e delle abitazioni 2011	174
10.9 Pubblicazione dei dati in formato Linked Open Data	175
10.9.1 Introduzione	175
10.9.2 Ontologie di Dominio	176
10.9.2.1 <i>Ontologia delle Basi Territoriali</i>	176
10.9.2.2 <i>Ontologia delle Variabili Censuarie</i>	177
10.9.3 Schema Architetture e Implementazione	178
<b>Conclusioni</b>	<b>183</b>
<b>Riferimenti bibliografici</b>	<b>185</b>
<b>Per saperne di più</b>	<b>186</b>



## Presentazione dell'opera

Gli atti del 15° Censimento generale della popolazione e delle abitazioni, data la complessità e l'importanza dell'argomento, sono articolati in sei volumi che documentano l'operazione censuaria attraverso la normativa di riferimento e le principali soluzioni organizzative, metodologiche e tecniche. Ciascun volume è autonomo nella trattazione tematica e affronta in maniera dettagliata ed esaustiva il proprio argomento specifico, ma restano evidenti le interconnessioni delle fasi e dei processi di progettazione e produzione. La collana è corredata da un insieme di documenti di riferimento, disponibili sul sito web dell'Istat<sup>1</sup>.

I sei volumi che formano la collana sono i seguenti:

1. *Le norme, l'organizzazione e la progettazione finanziaria*
2. *I metodi campionari e le tecniche di rilevazione*
3. *I contenuti informativi della rilevazione, la validazione e diffusione dei dati*
4. *Gli strumenti di informazione territoriale*
5. *I sistemi informatici a supporto delle operazioni censuarie*
6. *La valutazione della qualità*

Il primo volume *Le norme, l'organizzazione e la progettazione finanziaria* contiene le norme, sia internazionali che nazionali, sulla base delle quali si è progettato e realizzato il Censimento. Vengono illustrati e descritti nei loro aspetti salienti il Piano Generale di Censimento, il quadro finanziario in termini di progettazione e costi, l'organizzazione della rete di rilevazione, gli interventi di formazione e gli strumenti di comunicazione operativa adottati, i processi di gestione dei servizi in appalto, ed infine le strategie di comunicazione con i cittadini.

Il secondo volume *I metodi campionari e le tecniche di rilevazione* partendo da una trattazione generale della strategia di rilevazione assistita da liste amministrative, descrive nel dettaglio il processo di preparazione delle liste anagrafiche di base e delle liste ausiliarie per il recupero della sotto-copertura anagrafica. Vengono illustrati il disegno di campionamento per la selezione delle famiglie che hanno ricevuto il questionario in forma ridotta, gli aspetti salienti delle tecniche di rilevazione adottate e i processi di calcolo dei pesi di riporto all'universo per le variabili campionarie e di stima degli errori.

Il terzo volume *I contenuti informativi della rilevazione, la validazione e diffusione dei dati* descrive i contenuti informativi acquisiti tramite i questionari e le unità che sono state oggetto della rilevazione, i processi di controllo, correzione e validazione che hanno portato alla produzione degli aggregati di diffusione, i processi messi in atto per la pubblicazione dei primi risultati, dei dati di popolazione legale e sua struttura demografica e dei dati socio-economici, secondo il piano di diffusione nazionale e quello del Regolamento europeo.

Il quarto volume *Gli strumenti di informazione territoriale* descrive l'insieme di informazioni territoriali prodotte prima, durante e dopo le operazioni censuarie e finalizzate sia all'organizzazione delle operazioni sul campo che alla produzione dell'informazione resa al paese. Vengono illustrati il processo di aggiornamento delle basi territoriali, i metodi di definizione delle aree di censimento per la strategia di campionamento, la preparazione degli archivi di numeri civici

<sup>1</sup> <http://www.istat.it/it/archivio/174654>

utilizzati a supporto delle operazioni censuarie, la Rilevazione dei Numeri Civici, che ha portato alla creazione di un archivio di indirizzi ed edifici nelle città maggiori e il censimento degli edifici e delle abitazioni.

Il quinto volume *I sistemi informatici a supporto delle operazioni censuarie* descrive l'infrastruttura tecnico-informatica a supporto della raccolta dei dati e del monitoraggio delle operazioni sul campo. Vengono descritte le applicazioni web attraverso le quali i comuni hanno potuto trasmettere all'Istat le liste anagrafiche e quelle per effettuare la Rilevazione dei Numeri Civici, il "sistema integrato" di acquisizione dei dati, gestione della rilevazione e repository della documentazione, i sistemi per l'acquisizione e la verifica dei dati provenienti dalla lettura ottica, il sistema per la gestione della validazione finale dei dati, il sistema di rendicontazione delle spese censuarie e i sistemi di diffusione dei risultati.

Il sesto volume *La valutazione della qualità* descrive la Post Enumeration Survey (PES) ossia la rilevazione campionaria areale volta a stimare gli indicatori di qualità del censimento richiesti dal regolamento europeo. La rilevazione post censuaria, infatti, ha avuto il principale obiettivo di stimare il numero di individui abitualmente dimoranti nel periodo di riferimento, il tasso di sotto-copertura e il tasso di sovra-copertura del Censimento. Il volume descrive gli aspetti organizzativi, metodologici e i risultati della rilevazione campionaria.

## Il 15° Censimento generale della popolazione e delle abitazioni

*A cura di Andrea Mancini*

Per la prima volta, nel 2011, tutti gli stati membri dell'Unione europea hanno condotto il censimento della popolazione seguendo le regole comuni stabilite da un nuovo specifico Regolamento del Consiglio e del Parlamento<sup>2</sup>. Questo faceva proprie le nuove Raccomandazioni della Conferenza degli statistici europei (Ces)<sup>3</sup>, ma introduceva anche vincoli cogenti nei tempi di esecuzione, di riferimento dei dati e di trasmissione dei risultati. Ulteriori norme comuni venivano stabilite dai regolamenti di attuazione, emanati dalla Commissione europea<sup>4</sup>, soprattutto in materia di maggior dettaglio delle classificazioni, definizioni e nomenclature che i paesi membri erano chiamati a rispettare.

Destinato a svolgersi in occasione del 150° anniversario dell'Unità d'Italia, il 15° Censimento generale della popolazione e delle abitazioni è stato progettato dall'Istat con il supporto di un Comitato consultivo<sup>5</sup> costituito da Luigi Biggeri, allora Presidente dell'Istat, e composto da rappresentanti dei Ministeri interessati, del Cisis-Regioni, dell'Upi, dell'Anci, dell'Usci e dell'Anpci, delle associazioni professionali degli ufficiali di anagrafe (Anusca e DeA), da esperti nominati dal Presidente dell'Istat, oltre che da molti colleghi delle varie direzioni dell'Istituto. Il censimento è stato progettato in modo da rispettare le regole stabilite dai Regolamenti dell'Unione e in particolare di riuscire a diffondere i risultati con maggiore tempestività delle precedenti tornate e comunque entro i prescritti ventisette mesi dalla fine del 2011, ma anche di ottenere elevati standard di qualità dei dati, di mantenere la confrontabilità dei dati rispetto alle precedenti edizioni censuarie, di contenere al massimo i costi di produzione e il fastidio statistico delle famiglie.

Per cogliere questi obiettivi, parzialmente contrastanti, l'Istat ha deciso di effettuare il 15° Censimento introducendo numerose e rilevanti innovazioni nei metodi, nelle tecniche e nell'organizzazione della rilevazione. Le novità sono descritte nei vari fascicoli che compongono gli "Atti del censimento", che, come tradizione dell'Istituto, intendono garantire la memoria storica della rilevazione insieme alle informazioni che ne rappresentano il prodotto.

Sotto il profilo metodologico, il censimento del 2011 ha conservato l'enumerazione completa e simultanea mediante rilevazione di campo, ma ha utilizzato, per la prima volta in Italia, una lista di partenza delle unità di rilevazione personalizzata in base ai dati nominativi delle Liste anagrafiche comunali (Lac) di tutti i comuni italiani, riferite al 31 dicembre 2010 e acquisite mediante apposita applicazione elettronica gestita sul web. Dunque si è trattato di una rilevazione attuata sul campo, ma largamente supportata fin dall'inizio dall'uso di dati amministrativi. Ciò ha consentito di introdurre due ulteriori innovazioni di rilievo, costituite dalla spedizione postale di circa 25 milioni di plichi contenenti i questionari di famiglia, indirizzati all'intestatario del foglio di famiglia sulla base

---

<sup>2</sup> Regolamento (Ce) n. 763/2008 del Parlamento europeo e del Consiglio del 9 luglio 2008 relativo ai censimenti della popolazione e delle abitazioni.

<sup>3</sup> Conference of European statisticians, recommendations for the 2010 Censuses of population and housing.

<sup>4</sup> Commission regulation (Ec) n. 1201/2009 of 30 November 2009 implementing Regulation (Ec) n. 763/2008 of the European Parliament and of the Council on population and housing censuses as regards the technical specifications of the topics and of their breakdowns; Commission regulation (Eu) n. 519/2010 of 16 June 2010 adopting the programme of the statistical data and of the metadata for population and housing censuses provided for by Regulation (Ec) n. 763/2008 of the European Parliament and of the Council; Commission regulation (Eu) n. 1151/2010 of 8 December 2010 implementing Regulation (Ec) No 763/2008 of the European Parliament and of the Council on population and housing censuses, as regards the modalities and structure of the quality reports and the technical format for data transmission.

<sup>5</sup> Comitato consultivo per la preparazione a livello comunale del 15° Censimento generale della popolazione e delle abitazioni di cui alla Deliberazione del Presidente dell'Istat n. 673/PER del 6 giugno 2007.

dei dati anagrafici, nonché dalla predisposizione di un complesso Sistema di gestione della rilevazione (Sgr) completamente elettronico, in grado di controllare in tempo reale l'andamento della rilevazione su tutto il territorio nazionale con riferimento a ciascuna singola unità di rilevazione e a ciascun singolo operatore di censimento. Questo sistema, completamente sviluppato in Istat, ha comportato ulteriori opportunità di innovazione sul piano della tecnica di rilevazione, che infatti è divenuta "multicanale", dando la possibilità a ciascuna famiglia di compilare il questionario in forma elettronica<sup>6</sup> o cartacea<sup>7</sup>. In questo modo, rispetto al censimento del 2001, il numero dei rilevatori è diminuito di circa il 40 per cento, essendo il loro compito limitato al recupero dei questionari delle famiglie non ancora rispondenti, alla ricerca di persone e famiglie non ancora iscritte nell'anagrafe del comune alla data di riferimento del censimento (8 ottobre 2011) e alla enumerazione delle abitazioni non occupate. I risultati sono stati positivi sotto i profili della tempestività, della qualità dei dati raccolti e del risparmio di spesa, in conseguenza di tassi di risposta del 22,6 per cento via uffici postali, del 31,7 per cento presso i centri comunali di raccolta, del 12,3 per cento attraverso i rilevatori e del 33,4 per cento via web. Circa 8,5 milioni di questionari sono stati raccolti attraverso l'applicazione on line. La cifra assoluta di questionari compilati on line in Italia è la seconda più alta a livello mondiale nella tornata censuaria 2010. Nei primi due mesi di operazioni il carico medio su Sgr è stato di circa 115 questionari completati al minuto, con un picco di 500 nei periodi di massimo carico.

Inoltre, Sgr ha consentito a ciascun Ufficio comunale di censimento di effettuare le operazioni di confronto tra censimento e anagrafe in modo del tutto simultaneo alla rilevazione in atto. Ciò ha prodotto tre rilevanti risultati innovativi. In primo luogo, ha dato modo all'Istat di conoscere in tempo reale i risultati del confronto a livello di singolo nominativo nel rispetto della privacy, grazie al fatto che i risultati nominativi del confronto individuale effettuato dall'Ufficio comunale di censimento venivano memorizzati all'interno di Sgr stesso. In secondo luogo, ha consentito all'Istat di usare i risultati del confronto per determinare con grande tempestività i dati definitivi di popolazione legale, distinti per sesso, anno di nascita e cittadinanza (italiana e straniera). Il 18 ottobre 2012, a circa un anno dalla data di riferimento del censimento e a cinque mesi dalla fine della rilevazione, Enrico Giovannini, allora Presidente dell'Istat, ha potuto inviare al governo questi dati per la pubblicazione in Gazzetta Ufficiale, avvenuta poi il 18 dicembre 2012, contemporaneamente alla diffusione sul sito dell'Istituto delle prime tavole di dati definitivi. In terzo luogo, ha permesso di fornire tempestivamente ai servizi demografici di ciascun comune le liste nominative degli irreperibili al censimento e dei censiti non iscritti in anagrafe per la verifica post censuaria dei registri locali di popolazione, così come prescritto dal regolamento anagrafico nazionale.

Di nuovo sotto il profilo del metodo statistico, il censimento del 2011 si è anche caratterizzato per l'uso di campioni di famiglie residenti in aree di censimento urbane alle quali somministrare un questionario in forma breve (35 quesiti in luogo degli 84 contenuti nel questionario in forma lunga). Dunque, una parte non marginale dei risultanti dati è costituita da stime statistiche che, tuttavia, hanno rispettato errori campionari molto contenuti<sup>8</sup>. Anche questa novità è stata resa possibile dall'uso iniziale delle Lac e da Sgr, contribuendo al contenimento dei costi e del fastidio statistico di circa otto milioni di famiglie, alle quali è pervenuto il questionario in forma breve.

Molte delle innovazioni brevemente ricordate sono state oggetto di sperimentazioni, anche ripetute, condotte tra il 2008 e il 2010 dall'Istat in comune con le amministrazioni locali che avrebbero poi agito come organi di censimento e segnatamente con gli uffici di statistica e i servizi

---

<sup>6</sup> Con restituzione via web e rilascio automatico della ricevuta.

<sup>7</sup> Con restituzione ad uno qualsiasi degli uffici postali sparsi sul territorio oppure ad uno dei centri comunali di raccolta predisposti da ciascun Ufficio Comunale di Censimento.

<sup>8</sup> Dell'entità degli errori si parla diffusamente nel volume 2.

demografici dei comuni. Di esse non si dà conto nei fascicoli degli atti del censimento, ma un'ampia documentazione è comunque pubblicata sul sito dell'Istituto<sup>9</sup>, in particolare con riferimento alla rilevazione pilota, che si è svolta alla fine del 2009 coinvolgendo un campione di 78 mila famiglie di 31 comuni, e all'acquisizione sperimentale delle liste anagrafiche, che si è tenuta nel 2010 in 1.855 comuni. Di fatto queste sperimentazioni, insieme con le attività del progetto di informazione e formazione ai comuni sulle innovazioni censuarie<sup>10</sup>, hanno svolto un ruolo assai rilevante nel preparare la rete di rilevazione degli organi territoriali di censimento ben prima dell'inizio delle operazioni e in un clima di diffusa cooperazione e condivisione con le strutture centrali del Sistema statistico nazionale che sicuramente ha poi favorito l'efficacia della rilevazione e la capacità di affrontare e gestire le emergenze in corso d'opera.

Peraltro le innovazioni introdotte con il 15° Censimento sono state rese possibili anche dalla tempestiva rideterminazione delle basi territoriali o, più propriamente, dal complesso degli adempimenti precensuari in materia ecografica, di onomastica stradale e numerazione civica previsti dalla legge anagrafica (L. 24 dicembre 1954, n. 1228) e dal suo regolamento di attuazione (D.P.R. 30 maggio 1989 n. 223). Nel caso del censimento del 2011 questi adempimenti hanno assunto un precipuo valore propedeutico alle progettate innovazioni di metodo e tecniche della rilevazione, poiché alla rideterminazione delle basi territoriali è stata per la prima volta affiancata nei centri abitati dei comuni con più di 20 mila abitanti una rilevazione esaustiva dei numeri civici con l'obiettivo di verificare territorialmente la presenza di numeri civici, edifici e relative unità abitative e di riconciliare su mappa le relative informazioni. Nel fascicolo 4 degli Atti sono documentate le attività di rideterminazione delle basi territoriali, di rilevazione dei numeri civici e di loro georiferimento alle sezioni di censimento, queste due ultime anche con riferimento alle operazioni avvenute durante il censimento in tutti i comuni con meno di 20 mila abitanti. Di fatto si è trattato di un investimento pubblico in una infrastruttura informativa nazionale che, anche grazie all'intensa collaborazione con l'Agenzia del territorio poi confluita nell'Agenzia delle entrate, ha consentito all'Istat di realizzare, come ricaduta del censimento e con effetti duraturi nel tempo, l'Archivio nazionale di numeri civici e strade urbane.

Condurre un censimento basato su tecniche di distribuzione postale, con scelta autonoma dei rispondenti in merito alla modalità di restituzione del questionario e con necessità di gestire servizi on line 24 ore su 24 per un periodo di tempo non breve, ha comportato la necessità di predisporre un'organizzazione più articolata rispetto a quella adottata nelle precedenti tornate censuarie. Anche nel 2011 l'organizzazione si è fondata sui due tradizionali pilastri della rete territoriale di rilevazione da un lato e dei servizi censuari in appalto dall'altro lato, ma le interconnessioni tra i due pilastri sono state di necessità assai più delicate e complesse. La rete territoriale ha continuato a essere incentrata sugli Uffici Comunali di Censimento, ma questi hanno dovuto rafforzare tutte le funzioni di back office, cioè quelle deputate a usare le numerose funzionalità di Sgr e a effettuare il confronto tra censimento e anagrafe in tempo reale, in cambio di un significativo contenimento delle funzioni tradizionali dei rilevatori. Per contro i servizi censuari in appalto hanno assunto dimensioni e complessità assai superiori a quelle delle volte precedenti. Ciò ha comportato uno sforzo assai rilevante da parte dell'amministrazione dell'Istat e segnatamente del Servizio per il coordinamento del supporto amministrativo ai censimenti (Sac), appositamente creato per specificare le caratteristiche dei servizi in appalto e per controllare la corretta esecuzione di tutti i contratti stipulati con società esterne a seguito di gare pubbliche. I principali hanno riguardato il servizio di web hosting per la restituzione via internet dei questionari compilati e per la gestione di Sgr, il servizio di distribuzione postale alle famiglie iscritte in Lac dei plichi contenenti i questionari

<sup>9</sup> <http://www.istat.it/it/censimento-popolazione/censimento-popolazione-2011>

<sup>10</sup> Progetto INFO/FORCENS [http://www.istat.it/it/files/2011/04/doc\\_1\\_2010.pdf](http://www.istat.it/it/files/2011/04/doc_1_2010.pdf)

censuari e di loro raccolta presso gli uffici postali una volta compilati, il servizio di stampa dei questionari e di registrazione in lettura ottica e tradizionale e di gestione delle immagini, il servizio di sportello telefonico per i rispondenti e per gli organi di censimento della rete territoriale. I fascicoli 2 e 5 degli Atti danno conto delle caratteristiche organizzative di questi servizi e delle numerose interconnessioni operative tra ciascuno di essi e con le altre funzioni svolte direttamente dall'Istat o affidate agli organi della rete territoriale nelle varie fasi di conduzione del censimento. Si vuole qui sottolineare come i risultati ottenuti dal censimento siano largamente dipesi dall'efficacia dell'azione svolta dalla direzione amministrativa dell'Istat e dal Servizio Sac.

Il fascicolo 6 degli Atti illustra in dettaglio i metodi, le tecniche e l'organizzazione della rilevazione campionaria areale di controllo della copertura del Censimento (Pes), condotta da un gruppo di lavoro del tutto autonomo dal Dipartimento dell'Istat responsabile del Censimento. I risultati testimoniano della qualità dei risultati ottenuti dal censimento, con un tasso di sottocopertura a livello nazionale pari all'1,45 per cento della popolazione censita e un tasso di sovracopertura pari allo 0,38 per cento. Per la popolazione di cittadinanza italiana il primo tasso si riduce allo 0,66 per cento, mentre sale all'11,23 per cento per la popolazione straniera a ulteriore conferma della difficoltà di enumerare questo segmento di popolazione residente anche a causa della sua maggiore mobilità sul territorio. Peraltro le innovazioni di metodo, tecnica e organizzazione hanno consentito di soddisfare gli obiettivi di maggiore tempestività richiesti dal Regolamento europeo e stabiliti dal Piano generale di censimento. Le operazioni di censimento sul campo si sono concluse a febbraio del 2012 in tutti comuni, ad eccezione dei nove ai quali l'Istat aveva concesso la possibilità di registrare in proprio i questionari compilati in forma cartacea. Entro aprile 2012 si sono concluse le operazioni di confronto tra dati censuari e dati delle anagrafi. Entro ottobre 2012 erano pronti i dati definitivi della struttura demografica della popolazione residente per singolo comune, sesso, anno di nascita e cittadinanza (italiana e straniera), poi pubblicati il 18 dicembre 2012 nel Dpcm relativo alla popolazione cosiddetta legale e diffusi in pari data sul sito dell'Istat. La trasmissione ad Eurostat di tutti gli ipercubi di dati per il Census Hub creato da Eurostat è stata completata entro la scadenza regolamentare della fine di marzo 2014, con aggiornamento di alcune celle a settembre dello stesso anno, come consentito dal Regolamento della Commissione. È probabile che il 15° Censimento sia stato di transizione tra quelli precedenti, condotti secondo il metodo tradizionale dell'enumerazione mediante rilevazione generale sul campo, e quelli futuri, conducibili sulla base di sistemi di microdati integrati desunti da una pluralità di fonti amministrative. Nel 2011 l'uso delle Liste anagrafiche comunali ha segnato una tappa decisiva per realizzare un registro statistico della popolazione residente, così come la sistematica acquisizione di dati sui numeri civici ha consentito di impiantare un archivio generalizzato tramite il quale georeferenziare qualsiasi informazione di fonte amministrativa o statistica dotata di indirizzo. Nel futuro sarà dunque possibile fondare il censimento italiano sull'uso dei dati amministrativi, così come accade già da tempo in altri paesi europei. La rilevazione sul campo potrà allora assumere un disegno campionario funzionale da un lato alla correzione degli errori di copertura delle fonti anagrafiche e dall'altro all'acquisizione delle variabili censuarie non disponibili nelle fonti amministrative. Grazie alla conseguente riduzione dei costi, sarà anche possibile aumentare la frequenza dell'offerta informativa statistica di dati riferiti a livelli territoriali locali e aggiornare annualmente le informazioni censuarie finora prodotte una volta ogni decennio. Questo è l'obiettivo di lungo periodo che ha influenzato le scelte dell'Istat in materia di metodi e tecniche per il censimento del 2011.

## Introduzione<sup>11</sup>

L'esecuzione del 15° Censimento della Popolazione e delle Abitazioni è stata caratterizzata da numerose innovazioni di metodo, di processo e tecnologiche, che già erano state in parte introdotte nel 6° Censimento dell'Agricoltura.

Ciò che ha differenziato in maniera sostanziale i censimenti dell'agricoltura del 2010 da quello della popolazione ed abitazioni del 2011 è stata la differente modalità di gestione dell'infrastruttura hardware e software di base, basata esclusivamente su risorse (umane e strumentali) interne all'Istat nel primo caso e sull'adozione di un'infrastruttura di web hosting presso fornitore esterno per il secondo, limitatamente alle fasi di acquisizione delle liste anagrafiche di supporto e di acquisizione dei questionari e gestione della rilevazione sul campo, fasi in cui il carico sui sistemi è stato particolarmente alto in quanto hanno coinvolto l'intera popolazione italiana (circa 60.000.000 di individui), poco meno di 90 mila operatori della rete di rilevazione e l'intero territorio (8.092 Comuni).

Una caratteristica rilevante della maggior parte dei sistemi informatici realizzati per supportare le attività censuarie è l'infrastruttura centralizzata e la fruibilità via web da parte dei differenti attori, che hanno potuto accedere ai vari sistemi semplicemente utilizzando un *personal computer* o *tablet* di ultima generazione, in presenza di una connessione a *internet* attraverso un *browser* per la navigazione a scelta tra i più diffusi sul mercato, senza la necessità di installare *software* aggiuntivi o effettuare particolari configurazioni sui dispositivi utilizzati.

Nel fascicolo vengono presentati i sistemi realizzati per supportare tutte le fasi del censimento, a partire dalle attività preliminari, fino alla fase di diffusione dei risultati. Non vengono trattati i sistemi di controllo e correzione, dei quali si tratta in un fascicolo a parte. Nel primo capitolo viene descritta l'applicazione web, denominata Starlac (Sistema di Trasmissione, Acquisizione e Reportistica delle Liste Anagrafiche comunali), attraverso la quale i Comuni hanno potuto trasmettere all'Istat le liste Anagrafiche individuali con riferimento al 1° gennaio 2011, mentre i soli Comuni di dimensione demografica maggiore o uguale a 20000 abitanti hanno utilizzato l'applicazione anche per trasmettere lo stock della popolazione residente alla mezzanotte dell'8 ottobre 2011. Vengono inoltre descritte tutte le procedure (Traslac1 e Traslac2) realizzate per controllare correggere i dati e produrre la lista anagrafica di base, necessaria sia per l'invio postale dei questionari cartacei alle famiglie sia per la gestione dell'intera rilevazione.

Il secondo capitolo è dedicato alla descrizione dei sistemi realizzati per effettuare la Rilevazione dei numeri civici (Rnc), attraverso la quale il Comune ha raccolto informazioni sulla numerazione civica e sugli edifici presenti sul proprio territorio. Tale rilevazione è stata svolta da ciascun Comune con almeno 20 mila abitanti o comune capoluogo di provincia ed è stata utilizzata nel processo censuario in quanto ha consentito il corretto riferimento territoriale di individui, famiglie, abitazioni ed edifici, agevolando gli operatori della rete nella reperibilità dei cittadini e dando strumenti aggiuntivi di controllo durante le operazioni di rilevazione.

I capitoli 3, 4, 5 e 6 sono dedicati al "sistema integrato" di acquisizione dei dati (Qpop), gestione della rilevazione (Sgr) e *repository* della documentazione (Rete), caratterizzato dalla scelta strategica di rendere il web elemento cardine del processo di rilevazione e di gestione. Infatti una tra le caratteristiche innovative nell'organizzazione e conduzione dell'ultimo censimento della popolazione e delle abitazioni è sicuramente rappresentata dall'introduzione di un sistema di acquisizione dati via web, che ha affiancato la tradizionale raccolta dati attraverso questionario cartaceo. Il sistema (Qpop) è una applicazione web che è stata utilizzata dai rispondenti per compilare il questionario censuario. L'applicazione riproduce tutte e tre le tipologie di questionario

---

<sup>11</sup> L'Introduzione è a cura di Marina Venturi.

cartaceo, ovvero le due form per le famiglie (long e short) e il questionario delle convivenze. Qpop è pienamente integrato con il sistema di gestione della rilevazione Sgr, un sistema rivolto agli operatori per la gestione e il monitoraggio della rilevazione in tutte le sue fasi. Sgr è un sistema di *workflow* distribuito nel quale ogni operatore ha potuto lavorare autonomamente rispettando allo stesso tempo determinate regole condivise. Questa modalità operativa ha prodotto benefici in termini di tempestività, costi e capacità di controllo centralizzato della rete censuaria, delle operazioni censuarie e dei dati rilevati. In Sgr sono state implementate oltre 80 funzioni.

Contestualmente alle operazioni di rilevazione, ciascun Ufficio comunale di censimento ha effettuato il confronto tra gli individui risultanti in Lac all'8 ottobre 2011 e gli individui trovati al censimento. Per svolgere questa attività particolarmente complessa, gli operatori hanno utilizzato un insieme di funzionalità di Sgr che consentiva loro di esaminare ogni singolo questionario compilato, web o cartaceo, per confrontare la lista A compilata dai cittadini (Lista delle persone abitualmente dimoranti nell'alloggio/convivenza) con le risultanze dell'anagrafe comunale alla data del censimento, presente nel sistema stesso, e produrre così le liste individuali L1 (Individui censiti e residenti al medesimo indirizzo), L2 (Individui residenti non censiti), L3 (Individui censiti non residenti), L4 (Individui censiti ma residenti presso un altro indirizzo).

Nel capitolo 7 si descrivono tutti i sistemi informatici realizzati per acquisire in formato elettronico i questionari cartacei e verificare l'aderenza tra i dati riportati sul modello cartaceo e quelli registrati nel database. Il processo di registrazione in outsourcing dei questionari cartacei è stato effettuato tramite lettura ottica e registrazione tradizionale. I dati registrati sono stati organizzati in più file e inviati all'Istat attraverso un sistema realizzato appositamente per l'acquisizione dei dati ed il monitoraggio del processo di registrazione (Samor-Pop). L'architettura informatica approntata nell'ultimo censimento ha permesso di eliminare i supporti fisici mediante l'acquisizione on line delle informazioni presenti sui modelli (dati) e delle immagini output del processo di scansione. Per la gestione della fase di controllo di qualità della registrazione è stata appositamente realizzata l'applicazione Samor-Cqc, mentre per la ricerca e il recupero delle immagini dei questionari è stata predisposta l'applicazione Corinto. In questo modo i modelli cartacei sono stati inviati al macero e Corinto, oltre ad essere utilizzato durante il processo di acquisizione e monitoraggio della registrazione, viene utilizzato ogni qual volta è richiesto un accesso agli atti da parte di enti autorizzati o da parte dei cittadini.

Nel capitolo 8 viene descritto il sistema per la gestione della validazione finale dei dati, operazione che interviene a valle del processo di controllo e correzione. Per questo è stato realizzato un applicativo Web di controllo del processo, in cui vengono riportati - per ogni fase - i vari stati di avanzamento, i passaggi di stato, le eventuali attività in parallelo, le funzionalità per tipo di utente.

Nel capitolo 9 si descrive Il Sistema di Rendicontazione delle Spese censuarie (Srs) realizzato con l'obiettivo di permettere agli Uffici comunali (Ucc) e Provinciali di Censimento (Upc) di adempiere all'obbligo della rendicontazione delle spese umane e logistiche per le attività censuarie. Per adempiere all'obbligo della rendicontazione, gli Uffici Comunali e Provinciali di Censimento sono stati tenuti a compilare il modello di rendicontazione delle spese censuarie esclusivamente con l'ausilio di Srs.

Il capitolo 10 è dedicato ai sistemi di diffusione dei risultati del Censimento. Il progetto di diffusione si è articolato su tre principali tecnologie: data warehousing, business intelligence e navigazione avanzata ed interattiva sul Web.

Le attività di data warehousing e business intelligence hanno supportato sia la fase di validazione dei dati sia la fase di diffusione. In particolare, la fase di diffusione ha riguardato due filoni distinti seppur integrati: la diffusione nazionale fino al livello di dettaglio comunale e la diffusione in ambito europeo. Essa ha previsto la realizzazione di un data warehouse primario (Dwp) a monte della diffusione Web e la creazione di una particolare istanza del data warehouse dell'Istituto -

I.Stat - la cui tecnologia è stata appositamente evoluta per rispondere alle esigenze specifiche del Censimento.

La diffusione per sezione di censimento è invece stata eseguita utilizzando la tecnologia Linked Open Data (Lod).

I principali prodotti di diffusione supportati da processi informatici sono stati i seguenti:

- ▶ tavole statistiche (ipercubi) da fornire ad Eurostat sulla base del Piano di diffusione richiesto;
- ▶ tavole statistiche (ipercubi) diffuse tramite il data warehouse sulla base del Piano di diffusione italiano;
- ▶ microdati validati comunicati agli enti Sistan tramite file piatti richiesti mediante il Contact Center dell'istituto e accessibili anche a ricercatori e studiosi di università, istituti o enti di ricerca attraverso il Laboratorio Adele;
- ▶ dati aggregati per sezione di censimento e per le altre aree sub-comunali in formato Lod;
- ▶ indicatori demografici forniti al Sistema Informativo Geografico dell'Istat.
- ▶ file per la ricerca con campione al 5 per cento.



## CAPITOLO 1

### Il sistema di acquisizione delle liste anagrafiche comunali<sup>12</sup>

L'acquisizione delle liste anagrafiche comunali (Lac) costituisce il punto di partenza per la costruzione delle liste necessarie all'invio postale dei questionari censuari e per la gestione della rilevazione nel sistema Sgr.

Per la rilevazione censuaria sono state effettuate due acquisizioni distinte delle liste anagrafiche. La prima trasmissione delle Lac, obbligatoria per tutti i Comuni e propedeutica all'invio dei questionari censuari tramite servizio postale, è riferita al 31.12.2010 e si è svolta tra gennaio e marzo 2011. La seconda trasmissione, obbligatoria per i comuni con più di 20 mila abitanti e facoltativa per gli altri, aveva lo scopo di acquisire le variazioni anagrafiche intercorse tra il 1° gennaio 2011 e l'8 ottobre 2011 in modo tale che la situazione anagrafica aggiornata fosse di maggiore supporto alla rilevazione censuaria e al successivo confronto tra i dati censuari e i dati riportati nelle Anagrafi comunali.

#### 1.1 L'acquisizione delle Lac

Per l'acquisizione dei dati delle Lac è stata creata un'applicazione web denominata Starlac (Sistema di Trasmissione, Acquisizione e Reportistica delle Liste Anagrafiche Comunali). Attraverso di essa, gli operatori comunali hanno potuto trasmettere all'Istat le liste anagrafiche in formato csv. Si riporta una breve descrizione del sistema.

Gli utenti sono classificati secondo le seguenti tipologie:

- ▶ l'amministratore del sistema;
- ▶ gli utenti dell'Istat centrale, con visibilità su tutto il territorio nazionale;
- ▶ gli utenti delle sedi territoriali Istat (a cui sono equiparati, in Starlac, gli utenti dell'Astat per la provincia autonoma di Bolzano e del Servizio Statistico della provincia autonoma di Trento);
- ▶ i responsabili comunali;
- ▶ i referenti tecnici comunali.

Mentre le tipologie sono fisse, il sistema offre ampia possibilità di gestire profili utente e le funzioni associate ad ogni profilo. In tal senso, per ogni tipologia, a partire da un profilo base possono essere creati altri profili e ad ogni profilo possono essere associate o disassociate le corrispondenti funzioni.

Nella gestione delle utenze per gli operatori comunali Starlac deve tenere conto della normativa vigente in materia di protezione dei dati personali (quelli degli operatori, ma soprattutto quelli contenuti all'interno delle Lac). In fase di primo accesso al sistema ogni utente, ricevute le credenziali, deve indicare una serie di dati: nuova password, nome, cognome, email, pec, email, domanda di sicurezza con relativa risposta, telefono.

In particolare, per quanto riguarda le utenze comunali avviene quanto segue:

- 1) in fase di inizializzazione del sistema viene creata, per ogni comune, un'utenza del tipo "responsabile comunale": le credenziali vengono inviate al Comune tramite posta elettronica ordinaria o pec;

---

<sup>12</sup> Il Capitolo è a cura di Simona Spirito. Il paragrafo 1.1 è stato redatto da Francesco Bosio; il paragrafo 1.2 è stato redatto da Manuela Marrone; il paragrafo 1.3 è stato redatto da Raffaella Rosati; il paragrafo 1.4 è stato redatto da Maria Pia Di Maio.

- 2) effettuato il primo accesso, il responsabile comunale deve specificare il responsabile per l'invio dei dati: può indicare sé stesso oppure fornire i dati di un altro individuo (incluso indirizzo di posta elettronica certificata);
- 3) qualora il responsabile comunale indichi sé stesso, viene abilitato all'invio delle Lac;
- 4) qualora venga indicata una persona diversa, il sistema crea una nuova utenza del tipo "referente tecnico comunale" (abilitato all'invio delle Lac) e invia, tramite pec, le credenziali all'indirizzo indicato precedentemente.

Il file in formato csv con cui vengono inviati i dati è strutturato in base alle specifiche fornite dall'Istat:

- ▶ ad ogni riga del file corrisponde un individuo residente nel comune;
- ▶ i campi, corrispondenti ai dati dell'individuo, sono di lunghezza variabile e separati dal carattere “;”;
- ▶ ciascun individuo viene identificato tramite codice univoco;
- ▶ ciascuna residenza, famiglia o convivenza, è individuata da un codice famiglia o convivenza.

Per l'acquisizione del file l'applicazione Starlac guida l'utente nell'ambito di 5 fasi:

- 1) controlli preliminari: viene effettuato un primo set di controlli sulle prime 10 righe del file csv, con l'obiettivo di evitare che vengano inviati file nel formato errato (ad es. in formato xls/xlsx al posto di csv) o con struttura errata. Da un punto di vista tecnico questi controlli vengono effettuati in modo diverso a seconda che si tratti di comuni al di sopra o al di sotto di 100 mila abitanti. Se i controlli preliminari falliscono l'acquisizione del file csv non va a buon fine e viene inviato, tramite pec, un corrispondente resoconto all'utente comunale di riferimento;
- 2) acquisizione dei dati: il file viene trasmesso all'Istat e i dati vengono inseriti nelle tabelle del Database. Entrambe le operazioni avvengono in modo diverso a seconda che si tratti di comuni al di sopra o al di sotto dei 100 mila abitanti;
- 3) riclassificazione dei dati: ai comuni viene data la possibilità di utilizzare nel file csv una classificazione diversa da quella standard per i campi "codice sesso", "codice stato civile" e "codice relazione di parentela". In tal caso, nella fase di riclassificazione l'utente deve fornire gli elementi per poter interpretare la classificazione utilizzata;
- 4) controlli sui dati trasmessi, che sono di vario tipo:
  - controlli per singolo individuo, ossia sulla valorizzazione di determinate variabili (ad es. verifica della presenza di cognome e nome), sulla correttezza del formato di singole variabili (ad esempio data di nascita) o della loro classificazione (ad es. codice di cittadinanza), sulla coerenza tra variabili (ad es. tra codice fiscale e gli altri dati anagrafici);
  - controlli effettuati su più individui (ad es. famiglia con più di un intestatario, famiglia senza intestatario, codice fiscale ripetuto per più intestatari).

Per ogni tipo di controllo il sistema conta i record per i quali l'esito è negativo: nella maggior parte dei casi sono stabilite soglie di accettazione, anche in base all'ampiezza demografica, al di sopra delle quali il controllo viene considerato non superato. L'accettazione del file inviato dai comuni è subordinata al superamento di tutti i controlli: in questo caso l'acquisizione della Lac viene considerata "terminata con successo" e il sistema invia, tramite pec, una corrispondente ricevuta all'utente comunale. Se almeno uno dei controlli dà esito negativo, il sistema invia un avviso di mancata accettazione tramite pec al Comune.

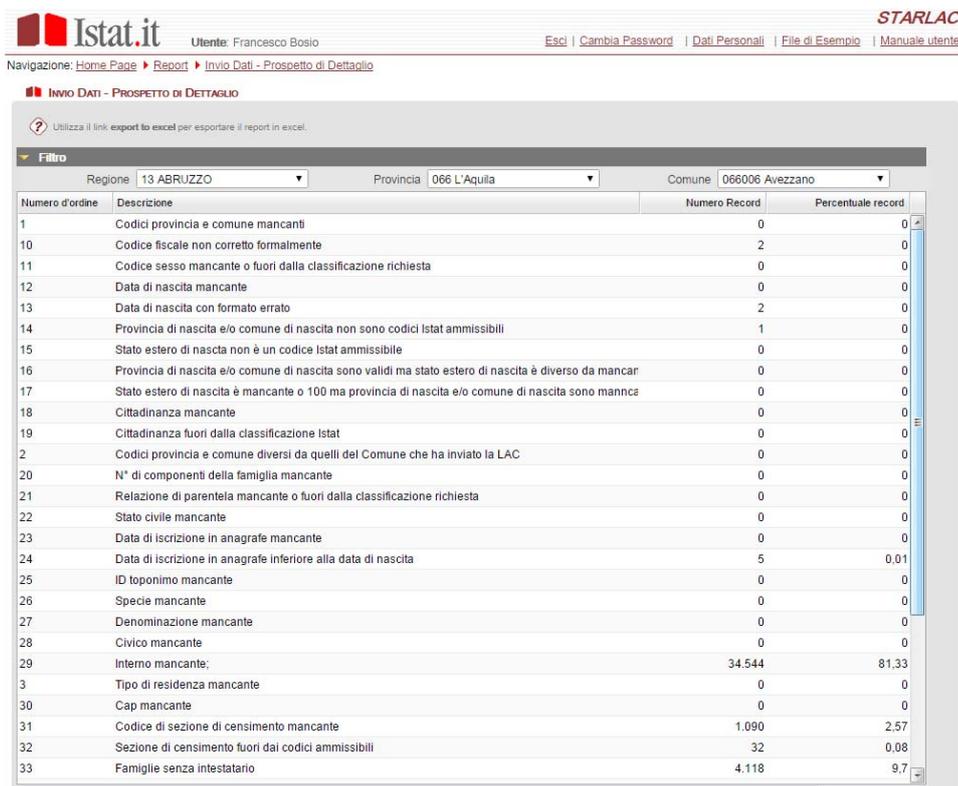
Starlac dispone anche di funzioni di reportistica e monitoraggio, mediante le quali gli operatori possono avere una panoramica sullo stato dell'acquisizione delle Lac nel territorio di propria competenza. I vari prospetti e report comprendono:

- ▶ il monitoraggio dei comuni, con l'indicazione dello stato in cui si trova la relativa acquisizione della Lac, con la possibilità di scaricare il corrispondente avviso inviato al Comune;
- ▶ un "Prospetto generale", con dati riepilogativi dell'acquisizione delle Lac;
- ▶ un "Prospetto di dettaglio" contenente i risultati ottenuti nella fase di controllo secondo diversi livelli di aggregazione (per l'intero territorio, per regione, provincia o comune); a livello comunale è anche possibile visualizzare liste di individui per i quali un determinato controllo ha dato esito negativo.

Nella figura 1.1 si riporta un esempio di Prospetto di dettaglio:

Il sistema è implementato utilizzando un Rdbms Oracle, facendo uso di oggetti Pl/Sql quali funzioni e procedure. L'applicazione web è realizzata in un ambiente java/jwt con l'utilizzo di Javascript e della piattaforma Struts. Affinché l'applicazione web con codice Javascript sia eseguibile su diversi browser viene impiegato Google Web Toolkit (Gwt). Per l'interfacciamento con l'Rdbms viene impiegato il framework Hibernate.

Figura 1.1 - Esempio di prospetto di dettaglio



Navigation: Home Page | Report | Invio Dati - Prospetto di Dettaglio

Utente: Francesco Bosio | Esci | Cambia Password | Dati Personali | File di Esempio | Manuale utente

**INVIATO DATI - PROSPETTO DI DETTAGLIO**

Utilizza il link [export lo excel](#) per esportare il report in excel.

Filtro: Regione: 13 ABRUZZO | Provincia: 066 L'Aquila | Comune: 066006 Avezzano

Numero d'ordine	Descrizione	Numero Record	Percentuale record
1	Codici provincia e comune mancanti	0	0
10	Codice fiscale non corretto formalmente	2	0
11	Codice sesso mancante o fuori dalla classificazione richiesta	0	0
12	Data di nascita mancante	0	0
13	Data di nascita con formato errato	2	0
14	Provincia di nascita e/o comune di nascita non sono codici Istat ammissibili	1	0
15	Stato estero di nascita non è un codice Istat ammissibile	0	0
16	Provincia di nascita e/o comune di nascita sono validi ma stato estero di nascita è diverso da mancar	0	0
17	Stato estero di nascita è mancante o 100 ma provincia di nascita e/o comune di nascita sono manna	0	0
18	Cittadinanza mancante	0	0
19	Cittadinanza fuori dalla classificazione Istat	0	0
2	Codici provincia e comune diversi da quelli del Comune che ha inviato la LAC	0	0
20	N° di componenti della famiglia mancante	0	0
21	Relazione di parentela mancante o fuori dalla classificazione richiesta	0	0
22	Stato civile mancante	0	0
23	Data di iscrizione in anagrafe mancante	0	0
24	Data di iscrizione in anagrafe inferiore alla data di nascita	5	0,01
25	ID toponimo mancante	0	0
26	Specie mancante	0	0
27	Denominazione mancante	0	0
28	Civico mancante	0	0
29	Interno mancante:	34.544	81,33
3	Tipo di residenza mancante	0	0
30	Cap mancante	0	0
31	Codice di sezione di censimento mancante	1.090	2,57
32	Sezione di censimento fuori dai codici ammissibili	32	0,08
33	Famiglie senza intestatario	4.118	9,7

In una prima versione di Starlac tutti i controlli preliminari venivano effettuati localmente, tramite Javascript e invocando funzioni dell'Api di Html5. Sempre facendo uso di Html5 l'invio del file veniva realizzato dividendo quest'ultimo in "pacchetti" di lunghezza fissa, che venivano inviati e memorizzati in un campo Clob di un apposita tabella sul Database (trasmissione del file) e successivamente inseriti in un'altra tabella (tabella "INDIVIDUO") avente una riga per ogni individuo (elaborazione del file).

Tale impostazione aveva lo svantaggio di non poter essere utilizzata nelle versioni dei browser in cui non erano supportate le Api Html5 utilizzate, creando difficoltà a molti operatori comunali e generando un eccessivo carico di assistenza nei loro confronti.

Per minimizzare tali svantaggi l'impostazione sopra descritta è stata mantenuta unicamente per i Comuni con una popolazione maggiore o uguale a 100 mila abitanti, mentre per quelli più piccoli è stata utilizzata una strategia alternativa, che non prevede l'utilizzo delle Api Html5:

- ▶ la trasmissione del file avviene tramite una semplice operazione di upload;
- ▶ i controlli preliminari avvengono a valle dell'upload, e non in locale;
- ▶ l'inserimento nella tabella INDIVIDUO avviene leggendo direttamente il file inviato.

L'impostazione descritta è totalmente trasparente all'utente che invia il file.

## 1.2 Il sistema di controllo delle Lac (Traslac1)

Dopo l'acquisizione delle Liste Anagrafiche Comunali (Lac) riferite al 31.12.2010 è stato necessario effettuare una serie di operazioni, finalizzate a:

il controllo, la correzione e la standardizzazione dei dati;

- ▶ determinare il tipo di questionario per ciascuna unità di rilevazione (famiglia o convivenza);
- ▶ preparare i dati per la stampa e la spedizione dei questionari nonché per la compilazione via web;
- ▶ caricare il sistema informatico per la gestione della rilevazione (Sgr).

Si elencano i passi che sono stati eseguiti:

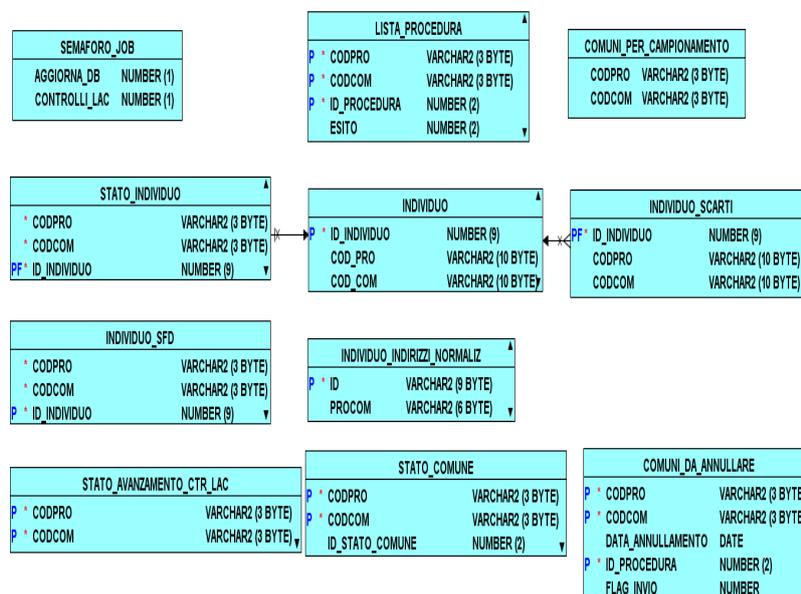
- 1) confronto tra popolazione Lac e Posas (Popolazione comunale residente al 31 dicembre per sesso età e stato civile): ha lo scopo di individuare Lac con anomalie legate a errori commessi in fase di predisposizione o di trasmissione del file da parte dei comuni. Il numero di residenti da Lac viene confrontato con l'analogo ammontare derivante dalla rilevazione Posas; vengono considerate anomale le Lac con scarti superiori ad una certa soglia, determinata in base all'ampiezza demografica;
- 2) controllo indirizzi palesemente errati: ha l'obiettivo di selezionare Lac anomale per via di un'elevata percentuale di indirizzi palesemente errati che non consentirebbero la spedizione del questionario. Qualora la frequenza degli indirizzi palesemente errati rispetto alla numerosità della popolazione superi una determinata soglia, si chiede un nuovo invio della Lac;
- 3) normalizzazione e geocodifica degli indirizzi: la normalizzazione (standardizzazione degli indirizzi), che permette il confronto con altre fonti amministrative, non è finalizzata alla spedizione dei questionari poiché si è deciso di affidarsi alla normalizzazione della ditta incaricata della spedizione. La geocodifica consente di recuperare la sezione di censimento qualora sia mancante o inaffidabile in Lac,
- 4) individuazione dei residenti in famiglia: si identificano correttamente i residenti in famiglia differenziandoli dai residenti in convivenza e si individuano eventuali anomalie relative alla numerosità delle famiglie.
- 5) Selezione degli intestatari di famiglia: si attribuisce l'intestatario alla famiglia in cui l'informazione da Lac sia errata o mancante;
- 6) secondo controllo indirizzi: si differenzia dal primo controllo sugli indirizzi perché viene effettuato in corrispondenza degli intestatari;
- 7) terzo controllo indirizzi (normalizzati): viene analizzato l'esito della normalizzazione. I comuni con percentuale di indirizzi non normalizzati superiore ad una certa soglia (5 per cento o 10 per cento) vengono sottoposti a correzione interattiva degli indirizzi per poi essere nuovamente sottoposti a normalizzazione;
- 8) individuazione dei "senza fissa dimora", iscritti all'Aire (Anagrafe degli Italiani Residenti all'Estero), nomadi: ha lo scopo di individuare in anagrafe i soggetti per i quali non verranno stampati i questionari;

- 9) preparazione dei dati per il campionamento: vengono predisposti i dati ed effettuati gli opportuni conteggi per l'estrazione dei campioni in ciascuna area di censimento e la corretta localizzazione delle famiglie campione;
- 10) predisposizione dei dati per la ditta di spedizione: vengono preparati i dati da inviare alla ditta che si occupa della spedizione dei questionari. Una volta eseguita la standardizzazione degli indirizzi da parte della ditta, viene eseguito il monitoraggio di qualità della normalizzazione.
- 11) predisposizione dei dati per la stampa dei questionari personalizzati: viene generata un'apposita tabella di dati contenente tutte le informazioni propedeutiche alla stampa dei questionari personalizzati, da inviare alle famiglie o all'ufficio comunale di censimento;
- 12) stima delle scorte: ha lo scopo di quantificare i questionari destinati alla parte di popolazione che non può essere censita tramite questionario personalizzato (famiglie senza fissa dimora, famiglie nomadi del comune di Roma e potenziali irreperibili di Roma, famiglie il cui intestatario è presente in Lac senza nominativo o senza indirizzo, individui che costituiscono sotto-copertura delle Lac, famiglie presenti in Lac senza codice, i terremotati del comune L'Aquila che risultano alla data del censimento residenti in altro comune, potenziali famiglie per le quali la consegna postale non vada a buon fine, potenziali rispondenti che dichiarano di aver smarrito o deteriorato il questionario;
- 13) predisposizione dei dati per la gestione delle convivenze: preparazione dei dati per l'individuazione delle convivenze e del relativo questionario, distinguendo tra convivenze anagrafiche e in convivenza (infatti i dati riferiti ad alcune tipologie di convivenze sono gestite dai ministeri);
- 14) preparazione dati per il caricamento in Sgr (sistema di gestione della rilevazione): predisposizione dei dati per selezionare i questionari da caricare sul diario di sezione;

I passi sopra elencati sono gestiti con procedure PL/Sql che interrogano un database Oracle creato ad hoc per l'elaborazione delle Lac.

Nella figura 1.2 si mostra uno schema delle principali tabelle Oracle coinvolte nelle varie elaborazioni:

Figura 1.2 - Schema della base dati



Di seguito si elencano le tabelle e le loro funzioni all'interno del flusso di esecuzione dei controlli sulle Lac:

- 1) individuo: contiene le Lac inviate dai comuni ed è l'input di tutti i passi di controllo delle Lac;
- 2) semaforo Job: segnala il corretto aggiornamento sul database delle Lac inviate da ciascun comune. Tabella di input ed output nel passo relativo al confronto tra popolazione Lac e Posas;
- 3) lista procedura: gestisce l'inizio e la fine (con o senza errori) delle varie fasi di controllo della Lac inviata. È una tabella di input ed output di tutti i passi di controllo delle Lac;
- 4) stato individuo: per ogni individuo contiene indicatori di qualità del record in relazione a ciascun passo eseguito, campi ricalcolati corrispondenti a variabili già presenti in Lac che sono state corrette, nuove variabili necessarie alla realizzazione dei vari passi. È l'output di tutti i passi di controllo delle Lac;
- 5) stato comune: indica per ogni comune l'ultima fase di controllo lanciata (con o senza errori) ed è l'output di tutti i passi di controllo delle Lac;
- 6) stato avanzamento: contiene per ogni comune variabili di conteggio e indicatori di qualità relativi all'esecuzione di ciascun passo. È l'output di tutti i passi di controllo delle Lac;
- 7) individuo indirizzi normaliz: contiene per ciascun individuo gli indirizzi normalizzati e indicatori di qualità della normalizzazione. È la tabella di output del passo di normalizzazione e geo-codifica degli indirizzi ed input per le fasi di controllo degli indirizzi;
- 8) comuni per campionamento: contiene la lista dei comuni con almeno 20000 abitanti da sottoporre a campionamento. È la tabella di input nel passo relativo alla preparazione dati per il campionamento;
- 9) individuo sfd: contiene gli individui senza fissa dimora ed è l'input del passo di individuazione dei senza fissa dimora;
- 10) individuo scarti: contiene gli individui scartati per superamento di soglia di errore ed è l'output di tutti i passi di controllo delle Lac;
- 11) comuni da annullare: contiene i comuni per i quali le Lac non hanno superato determinati controlli e presentano anomalie tali che necessitano di un nuovo invio. È l'output di tutti i passi di controllo delle Lac.

### **1.3 Il sistema di controllo delle variazioni anagrafiche (Traslac2)**

L'acquisizione della seconda versione delle Lac ha avuto lo scopo di individuare le variazioni anagrafiche e caricarle nel sistema Sgr.

I comuni che avevano l'obbligo di inviare all'Istat la lista anagrafica aggiornata alla data del censimento sono tutti quelli con popolazione residente al 1° gennaio 2010 superiore o uguale ai 20 mila abitanti (dato derivante dalla rilevazione Posas) e quelli della Provincia di Trento indipendentemente dalla dimensione demografica. Per i comuni con numero di abitanti compreso tra 5 mila e 19.999 l'invio era facoltativo. I comuni della Provincia di Bolzano hanno inviato la Lac aggiornata nel mese di settembre e per essi sono state adottate delle procedure ad hoc.

Le principali fasi del processo di controllo e correzione del secondo scarico sono descritte qui di seguito.

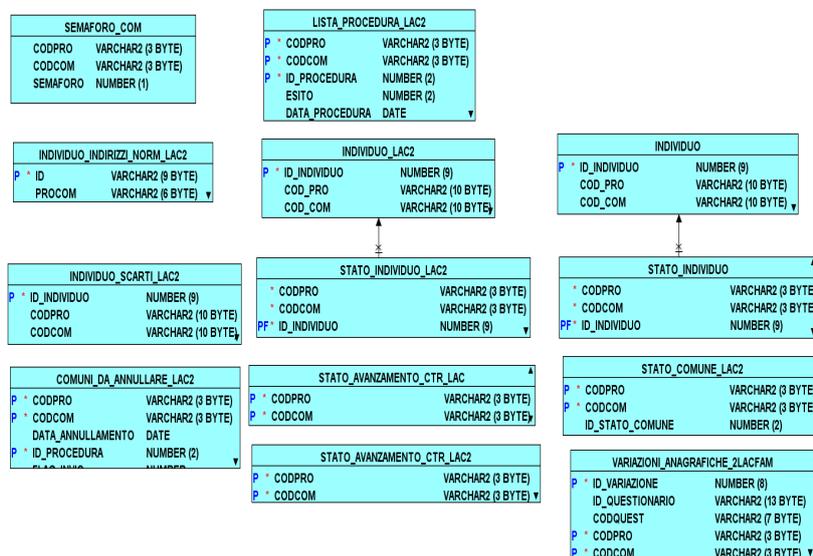
- 1) confronto tra popolazione Lac prima versione e popolazione Lac seconda versione: consiste nell'individuazione di comuni anomali, per i quali la variazione di popolazione tra primo e secondo scarico sia nulla oppure superiore ad una determinata soglia, variabile in base alla dimensione del comune;
- 2) analisi della data di iscrizione: si verifica che ciascun comune abbia inviato la Lac aggiornata, pertanto si controlla la presenza di individui con data di iscrizione compresa tra 31.12.2010 e 08.10.2011;

- 3) individuazione dei residenti in famiglia: analogamente a Traslac1, si distinguono correttamente i residenti in famiglia rispetto ai residenti in convivenza;
- 4) individuazione degli intestatari di famiglia: analogamente a Traslac1, si attribuisce l'intestatario alle famiglie per le quali l'informazione da Lac sia errata o mancante;
- 5) linkage tra prima e seconda versione e verifica della validità delle chiavi di aggancio: si effettua il linkage per individuo tra la Lac della prima versione e quella della seconda, allo scopo di individuare le variazioni anagrafiche che devono essere caricate e gestite in Sgr. Non essendo presente un'unica chiave di aggancio, il linkage viene eseguito in più passi utilizzando diverse variabili di aggancio e verificandone la validità;
- 6) individuazione delle variazioni anagrafiche: effettuato il linkage, è possibile determinare le variazioni anagrafiche a livello di individuo e di famiglia/convivenza. Vengono individuate le seguenti variazioni anagrafiche: nuova famiglia/convivenza, nuovo componente in famiglia/convivenza, componente cancellato da famiglia/convivenza, cambio domicilio famiglia/convivenza, famiglia/convivenza cancellata, cambio intestatario di famiglia, cambio responsabile di convivenza;
- 7) normalizzazione degli indirizzi per il caricamento in Sgr: vengono sottoposti a normalizzazione (standardizzazione degli indirizzi) soltanto gli indirizzi relativi a famiglie/convivenze esistenti che hanno cambiato indirizzo o intestatario oppure a famiglie/convivenze nuove;
- 8) creazione tabella per caricamento in Sgr: la tabella delle variazioni anagrafiche, che devono essere caricate su Sgr, contiene i record riferiti a:
  - gli individui appartenenti a famiglia/convivenza nuova;
  - i nuovi componenti di famiglia/convivenza già esistente;
  - i componenti cancellati da famiglia/convivenza già esistente;
  - gli intestatari di famiglia/convivenza già esistente che ha cambiato indirizzo;
  - gli intestatari appartenenti a famiglia/convivenza cancellata;

i nuovi intestatari appartenenti a famiglia/convivenza già esistente. I passi sopra elencati sono gestiti mediante procedure PL/Sql che interrogano un database Oracle, creato ad hoc per l'elaborazione sulla seconda versione delle Liste Anagrafiche Comunali (Lac).

Le principali tabelle Oracle coinvolte nelle varie elaborazioni sono riportate nella figura 1.3:

Figura 1.3 - Schema delle tabelle Oracle utilizzate nelle elaborazioni della seconda versione delle Lac



I passi sopra elencati sono gestiti mediante procedure PL/Sql che interrogano un database Oracle, creato ad hoc per l'elaborazione sulla seconda versione delle Liste Anagrafiche Comunali (Lac).

Le principali tabelle Oracle coinvolte nelle varie elaborazioni sono riportate nella figura 1.3:

Di seguito viene mostrata una breve descrizione delle principali tabelle e delle loro funzioni.

- 1) semaforo\_com: segnala il corretto aggiornamento sul database del secondo invio delle Lac da ciascun comune;
- 2) lista\_procedura\_lac2: gestisce l'inizio e la fine (con o senza errori) delle varie fasi di controllo della seconda Lac inviata. È la tabella di input ed output di tutti i passi di controllo della seconda versione delle Lac;
- 3) individuo\_indirizzi\_norm\_lac2: contiene gli individui del secondo invio delle Lac, con indirizzi sottoposti a normalizzazione. È la tabella di output del passo di normalizzazione degli indirizzi per il caricamento in Sgr;
- 4) individuo: contiene le Lac del primo invio dei comuni, costituisce l'input dei seguenti passi: individuazione dei residenti in famiglia e linkage tra primo e secondo scarico;
- 5) individuo\_lac2: contiene le Lac del secondo invio dei comuni ed è l'input di tutti i passi di controllo del secondo scarico delle Lac;
- 6) stato\_individuo: per ogni individuo del primo scarico delle Lac, contiene le informazioni riepilogative risultanti dalle varie fasi di controllo, è l'input di tutti i passi di controllo della seconda versione delle Lac;
- 7) stato\_individuo\_lac2: analogamente alla corrispondente della prima versione, contiene indicatori di qualità del record in relazione a ciascun passo eseguito, campi ricalcolati corrispondenti a variabili già presenti in Lac che sono state corrette, nuove variabili necessarie alla realizzazione dei vari passi. È l'output di tutti i passi di controllo della seconda versione delle Lac;
- 8) stato\_comune\_lac2: indica per ogni comune l'ultima fase di controllo lanciata (con o senza errori) ed è l'output di tutti i passi di controllo della seconda versione delle Lac;

- 9) stato\_avanzamento\_ctr\_lac: contiene, per ogni comune, variabili di conteggio e indicatori di qualità relativi all'esecuzione di ciascun passo della prima versione. È l'input del passo di confronto tra popolazione in Lac della prima e seconda versione;
- 10) stato\_avanzamento\_ctr\_lac\_2: per ogni comune contiene variabili di conteggio e indicatori di qualità relativi all'esecuzione di ciascun passo della seconda versione. È l'output di tutti i passi di controllo della seconda versione delle Lac;
- 11) individuo\_scarti\_lac2: contiene gli individui scartati per superamento di soglia di errore ed è l'output dei passi di controllo della seconda versione delle Lac;
- 12) comuni\_da annullare\_lac2: contiene i comuni per i quali le Lac della seconda versione non hanno superato i controlli e necessitano di un nuovo invio ed è l'output dei passi di controllo della seconda versione delle Lac;
- 13) variazioni\_anagrafiche\_2lacfam: contiene le variazioni anagrafiche che devono essere caricate e gestite in Sgr. È l'input del passo di creazione tabella per il caricamento in Sgr.

#### 1.4 Il supporto ai comuni

Al fine di garantire il corretto svolgimento dell'operazione di invio delle Liste Anagrafiche Comunali, l'Istat ha messo a disposizione dei comuni vari canali di comunicazione. Sono state inviate diverse circolari attuative e informative, sia durante la fase preliminare di costruzione del file contenente le informazioni anagrafiche da inviare, sia durante le varie fasi di acquisizione.

Le circolari potevano essere visualizzate e scaricate direttamente dall'interfaccia di Login dell'applicazione Starlac cliccando sul link di interesse (figura 1.4).

Figura 1.4 - Home dell'applicazione Starlac



**LOGIN**

**Istat.it**  
 Benvenuti nel sistema ISTAT STARLAC 3.2. Per accedere usa le tue credenziali di accesso.  
 Se hai dimenticato la Password, puoi utilizzare il link [Recupera Password] per resettarla.

**Per l'invio della LAC i comuni con oltre 100.000 abitanti devono utilizzare:  
 Mozilla Firefox (versione 15 o succ) o Chrome (versione 25 o succ.).**

Nome Utente

Password

[Recupera Password]

[Circolare per Resp. Serv. Demografici con meno di 50.000 ab](#)  
[Circolare per Resp. Serv. Demografici con piu' di 50.000 ab](#)  
[Circolare per Resp. Uff. Statistica con meno di 50.000 ab](#)  
[Circolare per Resp. Uff. Statistica con piu' di 50.000 ab](#)

Un ulteriore ausilio per l'utente comunale che accedeva all'applicativo per l'invio della Lac era la possibilità di consultare il manuale utente e di scaricare un esempio di lista anagrafica cliccando sui rispettivi hyperlink (figura 1.5) descritti di seguito:

- ▶ “manuale Utente”, mediante il quale è possibile visualizzare il manuale in Acrobat Reader oppure scaricarlo in formato *pdf*;
- ▶ “file di Esempio”, attraverso il quale si può scaricare un file di esempio comprendente pochi records.

Figura 1.5 - Hiperlink per download



In aggiunta al materiale informativo, è stato predisposto un servizio di helpdesk telefonico e telematico al quale i comuni potevano rivolgersi in caso di difficoltà.

Per quanto concerne il supporto telematico è stato attivato l'indirizzo mail [Starlac@istat.it](mailto:Starlac@istat.it), mediante il quale era possibile dettagliare la problematica riscontrata attraverso schermate o altro materiale illustrativo (figura 1.6).

Al fine di massimizzare l'efficienza del servizio, i recapiti telefonici sono stati differenziati in modo da supportare al meglio le operazioni delle diverse figure coinvolte: una linea, infatti, è stata dedicata ai Responsabili e ai Referenti tecnici dei comuni mentre l'altra è stata destinata a supportare le Software House (le società incaricate dai comuni per la gestione delle Lac).

Figura 1.6 - E-mail di riferimento

<b>E-mail: <a href="mailto:starlac@istat.it">starlac@istat.it</a></b>	
<b>☎ per i Responsabili e Referenti tecnici:</b>	06-46734520
<b>☎ per le Software House:</b>	06-46734439
<b>☎ Istat Sedi Regionali:</b>	<small>[Elenco telefonico ed email dei referenti Istat regionali]</small>

La linea dedicata ai Responsabili e Referenti tecnici è stata a sua volta divisa in due macro unità distinte per tipologia di problematiche. Una volta composto il numero, una voce guidata segnalava:

- ▶ il tasto 1 per le delucidazioni di carattere tecnico-statistico, quali le modalità di invio, scadenze, rinvio di circolari;
- ▶ il tasto 2 per le difficoltà di carattere tecnico-informatico, come ad esempio reimpostare la password o l'incompatibilità del browser, oppure la necessità di essere guidato durante la fase di riclassificazione delle variabili.

I telefoni sono stati attivi dal Lunedì al Venerdì dalle 9:00 alle 13:00 e il Martedì e Giovedì fino alle 16:00.

Grazie a questa fitta rete di canali di ausilio e grazie al costante monitoraggio da parte dell'Istat è stato possibile mantenere sempre alta l'attenzione sull'acquisizione delle Liste Anagrafiche. Conseguentemente l'intera operazione si è conclusa con successo entro i tempi stabiliti.

## CAPITOLO 2

### il sistema della rilevazione dei numeri civici <sup>13</sup>

Tra le attività preparatorie del 15° Censimento generale della popolazione e delle abitazioni che hanno riguardato i comuni, gli adempimenti topografici ed ecografici sono risultati di rilevante importanza per la nuova strategia censuaria. I primi afferiscono alla creazione delle basi territoriali e i secondi alla revisione dello stradario e della numerazione civica.

Il corretto riferimento territoriale di individui, famiglie, abitazioni ed edifici è stato fondamentale sia per la reperibilità dei cittadini e sia per aver offerto la possibilità di effettuare controlli nelle operazioni di rilevazione.

Attraverso la Rilevazione dei numeri civici (Rnc) il Comune ha raccolto informazioni sulla numerazione civica e sugli edifici presenti sul proprio territorio. Le unità di rilevazione della Rnc sono due:

- ▶ il numero civico, costituito dal numero che contraddistingue l'*accesso esterno* che dall'area di circolazione immette direttamente o indirettamente alle unità immobiliari (abitazioni, esercizi, uffici, ecc.);
- ▶ l'edificio, che costituisce il contenitore delle altre unità rilevate ai censimenti (abitazioni, famiglie, persone, uffici, stabilimenti, ecc.);

Per la rilevazione sul campo, il Rilevatore ha avuto a disposizione:

- ▶ la mappa della sezione di censimento in cui sono riportati i limiti della sezione stessa, le foto aeree aggiornate del territorio, i grafi stradali e la cartografia catastale con i poligoni che indicano gli edifici con i relativi codici di edificio preassegnati. I codici di edificio sono stati generati e assegnati automaticamente a tutti i poligoni con superficie maggiore di 20 metri quadri presenti nella cartografia di fonte catastale;
- ▶ il modello Istat Rnc.1 precompilato per sezione di censimento con i tratti di arco di strada e numeri civici secondo le informazioni risultanti dagli archivi (stradari e indirizzari) comunali;
- ▶ il modello Istat EDI.1 per la rilevazione degli edifici.

La Rnc è stata svolta da ciascun Comune con almeno 20 mila abitanti o comune capoluogo di provincia.

#### 2.1 L'architettura generale

Il Sistema informatico Rnc completo per la gestione dei numeri civici è composto dai seguenti moduli:

- ▶ accesso al sistema e recupero password;
- ▶ informazioni;
- ▶ parte dell'applicazione disponibile a tutti senza necessità di login:
  - diffusione documentazione;
  - diffusione documentazione sull'applicazione;
  - Faq;
  - visualizzazione delle domande più frequenti con relative risposte
  - contatti;
  - visualizzazione dei contatti dei responsabili delle unità regionali e del centro Istat;

<sup>4</sup> Il capitolo è cura di Francesco Di Pedè. Il paragrafo 2.1 e i sottoparagrafi 2.1.1, 2.1.2, 2.1.3, 2.1.4, 2.1.5, 2.1.6, 2.1.7, 2.1.8 sono stati redatti da Edoardo Patruno.

- ▶ gestione utenze: I responsabili comunali e gli utenti di livello superiore creano, modificano e cancellano utenze dei rilevatori del proprio Comune oltre a visualizzare i propri dati. Il responsabile del progetto gestisce le utenze dei responsabili comunali e visualizza i dati delle unità regionali mediante le seguenti funzioni:
  - gestione rilevatori;
  - assegnazione schede di rilevazione ai rilevatori comunali;
  - gestione responsabili comunali;
  - gestione unità regionali.

Il responsabile del progetto visualizza e modifica i dati delle utenze delle unità regionali:

- ▶ monitoraggio: i responsabili comunali e gli utenti di livello superiore monitorano l'andamento della rilevazione per il territorio di loro competenza
- ▶ invio e scarico file: i responsabili comunali e gli utenti di livello superiore possono inviare file con informazioni su indirizzi ed edifici e scaricare i file stessi o il file degli scarti dei dati che il sistema non è riuscito a processare.
- ▶ gestione rilevazione: parte principale del sistema Rnc, gestisce l'associazione degli edifici agli indirizzi e il completamento delle informazioni attraverso le seguenti funzioni:
  - selezione schede di rilevazione: i rilevatori e gli utenti di livello superiore selezionano le schede di rilevazione tra quelle di loro competenza;
  - scarico delle schede di rilevazione: le schede selezionate possono essere salvate sul pc dell'utente;
  - gestione scheda di rilevazione: i rilevatori e gli utenti di livello superiore completano le informazioni sugli indirizzi della scheda di rilevazione e associano gli indirizzi a uno o più edifici;
  - gestione edifici: i rilevatori e gli utenti di livello superiore completano le informazioni sugli edifici delle schede di rilevazione;
  - gestione richieste nuovi toponimi: i rilevatori e gli utenti di livello superiore richiedono che possa essere inserito nel sistema un nuovo toponimo; che possa essere cambiato il nome di un toponimo; che possa essere aggiunto un indirizzo;
  - validazione schede di rilevazione: i rilevatori e gli utenti di livello superiore validano le informazioni sugli indirizzi e gli edifici della scheda di rilevazione;
  - gestione stradario: I responsabili comunali e gli utenti di livello superiore visualizzano lo stradario del loro comune e modificano i dati dei toponimi e degli indirizzi;
  - gestione richieste toponimi: i rilevatori non possono aggiungere nuovi toponimi, ma solo richiedere l'aggiunta o la modifica, spetta ai responsabili comunali e gli utenti di livello superiore la validazione delle richieste.

L'architettura della soluzione software proposta è una tipica applicazione client/server costituita da un client che implementa l'interfaccia utente con alcune funzionalità di elaborazione e di comunicazione e da un server che fornisce una serie di servizi come la gestione e l'accesso ai dati di un database. L'applicazione web si basa su elementi software standard indipendenti dalle caratteristiche della particolare applicazione e dalla piattaforma software e hardware su cui viene eseguita, inoltre si sviluppa su tre livelli logico-funzionali:

- ▶ livello di presentazione - rappresenta l'interfaccia utente dell'applicazione e si occupa di acquisire dati e visualizzare risultati (Html, Ajax e Javascript);
- ▶ livello intermedio - si occupa delle elaborazioni dei dati in base alla cosiddetta business logic, cioè all'insieme delle regole per cui i dati sono considerati significativi e le loro relazioni consistenti; le elaborazioni del livello intermedio generano i risultati richiesti dall'utente (Php);
- ▶ livello dati – il gestore di database (Oracle).

### 2.1.1 Login

La homepage del sistema Rnc presenta una form che si compone di due campi per l'inserimento delle credenziali di accesso (username e password) ed un link tramite il quale rigenerare la parola chiave nel caso in cui l'utente dimentichi quella scelta al primo accesso. Per questioni di sicurezza tutte le volte che un utente effettua il primo login al sistema è obbligato a definire una nuova parola chiave per i successivi accessi. Una volta compilati i campi, è sufficiente fare click sul pulsante 'Login' per visualizzare la propria pagina personale (figura 2.1).

Al di sopra della form di accesso è presente un menù le cui voci sono: 'Login' e 'INFO'. Per visualizzare la form di login è necessario cliccare sulla voce 'Login'; attraverso la voce 'INFO' è possibile ottenere informazioni sul progetto ('PROGETTO'), salvare localmente la documentazione relativa all'attività di rilevazione ('DOCUMENTAZIONE'), ottenere i contatti degli uffici territoriali Istat ('CONTATTI') e poter consultare le Frequently Asked Questions ovvero le domande più frequenti (Faq). Per rigenerare la propria password nel caso di smarrimento, l'utente deve cliccare sul link posto sotto il pulsante 'Login' ed inserire la propria username ed email con le quali è registrato. Il sistema invia alla casella di posta elettronica dell'utente un messaggio contenente una password generata dal sistema, che deve successivamente essere modificata. Ogni volta che un utente effettua un tentativo di accesso al sistema compare sullo schermo un 'alert' che lo avvisa circa l'esito dell'operazione comunicandogli anche la necessità di modificare la password nel caso in cui si accede al sistema per la prima volta.

Figura 2.1 - Homepage Rnc



I responsabili comunali devono, dopo aver definito una nuova parola chiave, fornire le proprie generalità e i dati del proprio Comune.

### 2.1.2 Pagina personale

Effettuato con successo il login e terminate le formalità legate al primo accesso si raggiunge la propria pagina personale, illustrata in figura 2.2

Figura 2.2 - Homepage dell'utente



La struttura della pagina è identica per tutti i profili (Amministratore, Unità regionale, Responsabile comunale, Rilevatore, Compilatore) ma sono differenti le operazioni disponibili per ciascuno dei 5 profili. È dunque differente il menù a tendina presente sul banner rosso che presenta alla sua sinistra un logo che indica la posizione dell'utente all'interno del sistema.

Al di sotto del menù a tendina viene indicato lo username dell'utente, il suo profilo ed eventualmente le richieste legate ai toponimi ricevute o effettuate (ciò dipende dal ruolo dell'utente, infatti il rilevatore ed il compilatore possono inviare richieste ma non riceverle e quindi processarle).

Oltre a queste indicazioni è possibile utilizzare una serie di 'tab' posti al di sotto delle informazioni utente, attraverso i quali visualizzare lo stato della rilevazione a livello aggregato e di singolo utente. Queste informazioni di monitoraggio sono chiaramente legate al profilo dell'utente e alla regione e comune d'appartenenza dello stesso. I valori che vengono mostrati in tabella sono presentati in valore assoluto ed in valore percentuale. Una volta visualizzata una tavola è possibile ottenere informazioni più dettagliate attraverso il pulsante "DETTAGLI" presente in corrispondenza di ogni singola voce mostrata. Nel caso in cui l'utente è un rilevatore o compilatore visualizza le informazioni relative alle sole schede di sua competenza e le richieste da lui effettuate (figura 2.3).

Figura 2.3 - Tavole di monitoraggio



### 2.1.3 Home page e funzionalità del menu

La barra dei pulsanti è lo strumento attraverso il quale ogni utente può muoversi all'interno dell'applicazione. Non tutti i pulsanti sono visibili per tutte le tipologie di utente. In base al ruolo ricoperto dall'utente che ha effettuato l'accesso al sistema alcuni pulsanti sono visibili mentre altri non compaiono sulla barra.

I pulsanti condivisi da tutti i profili sono: 'HOME', 'GEST\_PROFILLO', 'GEST\_RILEVAZIONE', 'MONITORAGGIO', 'INFO', 'LOGOUT'. Il pulsante 'GEST\_UTENZE' non compare sulla barra per i rilevatori e per i compilatori (figura 2.4)

Figura 2.4 - Barra dei pulsanti



Il pulsante 'GEST.PROFILO' (gestione profilo) permette all'utente di modificare la propria password attraverso la compilazione di una form predefinita. L'operazione di modifica della password prevede

l'inserimento della password da cambiare ed il duplice inserimento della nuova password per ragioni di sicurezza. La nuova password non può coincidere con la vecchia password.

Il passaggio del puntatore del mouse su 'GEST.UTENZE' (non visibile per i rilevatori e per i compilatori) mostra 3 differenti pulsanti:

- ▶ 'GESTIONE RILEVATORI';
- ▶ 'GESTIONE COMPILATORI';
- ▶ 'GESTIONE RESPONSABILE COMUNALE';

Cliccando su 'GESTIONE RILEVATORI' viene mostrato l'elenco dei rilevatori registrati per un determinato comune che coincide con quello del responsabile comunale che ha effettuato l'accesso al sistema oppure con quello scelto attraverso un apposito filtro dall'utente regionale o dall'amministratore.

Mentre l'amministratore può scegliere sia la regione, che la provincia, che il Comune, l'utente regionale può selezionare solo provincia e comune.

Una volta ottenuto l'elenco dei rilevatori, ad ogni nominativo sono associati ulteriori pulsanti, che permettono l'assegnazione delle schede di rilevazione ('ASSEGNA'), la modifica delle generalità del rilevatore ('EDITA') e la cancellazione del rilevatore ('CANCELLA'). È possibile inserire un nuovo rilevatore attraverso il pulsante 'NUOVO RILEVATORE' posto al di sopra dell'elenco rilevatori. Questo pulsante è visibile solo quando è stato selezionato un comune. Per il responsabile comunale il pulsante di aggiunta nuovo rilevatore è sempre visibile. L'icona di MsExcel a sinistra del pulsante 'NUOVO RILEVATORE' genera un file in formato x/s che contiene l'elenco dei rilevatori visualizzati.

Cliccando su 'GESTIONE COMPILATORI' viene mostrato l'elenco dei compilatori registrati per un determinato comune, che coincide con quello del responsabile comunale che ha effettuato l'accesso al sistema, oppure con quello scelto attraverso un apposito filtro dall'utente regionale o dall'amministratore.

Per questa funzione valgono considerazioni analoghe a quelle descritte per 'GESTIONE RILEVATORI'.

Per procedere con la cancellazione di un compilatore è necessario cliccare sul pulsante 'CANCELLA' e confermare l'operazione. Si può cancellare un compilatore se e solo se non ha schede assegnate. La modifica dei dati di un compilatore si effettua compilando la form che appare dopo aver cliccato il pulsante 'EDITA' e dopo aver confermato l'operazione. L'assegnazione delle schede di rilevazione ad un determinato compilatore avviene cliccando il pulsante 'ASSEGNA'. Questa operazione apre una form all'interno del quale sono mostrati gli elenchi delle schede assegnate al compilatore prescelto e le schede ancora assegnabili.

Il pulsante 'GESTIONE RESPONSABILE COMUNALE' permette la modifica dei dati relativi al responsabile comunale e al comune. Nel caso in cui l'utente che ha effettuato l'accesso sia l'amministratore oppure sia un utente regionale, prima di poter effettuare modifiche deve selezionare il comune mediante un apposito filtro.

Il pulsante 'GESTIONE UURR', visibile solo dall'amministratore del sistema, permette la modifica dei dati relativi all'unità regionale. Il pulsante 'EDITA', posto su ogni riga della tabella che elenca le unità regionali, apre una pagina di modifica delle informazioni.

Cliccando il pulsante 'INDIRIZZI' relativo ad un toponimo si ottiene una nuova tabella dove vengono elencati gli indirizzi legati al toponimo precedentemente scelto. Per ogni indirizzo sono visualizzati il civico, l'esponente, gli edifici, la provenienza e la destinazione d'uso e alcuni pulsanti che permettono la modifica e l'assegnazione di uno o più edifici. In fondo alla tabella invece sono disponibili pulsanti per la modifica, l'aggiunta, l'eliminazione e la ricerca di un indirizzo, ed il ricaricamento della tabella nella base dati.

Il pulsante contrassegnato dalla lettera 'E' permette l'editing dell'indirizzo direttamente sulla tabella; con il bottone 'S' si salvano le modifiche effettuate, mentre con il pulsante 'A' si annullano le modifiche in corso. Queste stesse operazioni possono essere effettuate anche con i pulsanti situati in fondo alla tabella. Il pulsante 'ASS.E' è invece legato alla possibilità di associare al corrispondente indirizzo uno o più edifici. Non è possibile inserire indirizzi con lo stesso civico ed esponente.

#### 2.1.4 Gestione stradario

La funzionalità 'GESTIONE STRADARIO' non è disponibile per i rilevatori e per i compilatori, in quanto la possibilità di modificare, cancellare oppure aggiungere nuovi toponimi è data ai soli utenti comunali, regionali e all'amministratore del sistema. La visualizzazione dei toponimi è legata al filtro geografico posto al di sopra della tabella e che è strettamente legato agli attributi geografici dell'utente che ha effettuato l'accesso. I toponimi elencati nella tabella possono essere filtrati digitando nel campo dug e/o nel campo toponimo del testo che viene utilizzato come chiave di ricerca aggiornando di conseguenza la visualizzazione dell'elenco delle voci dello stradario (figura 2.5).

Figura 2.5 - Gestione stradario

The screenshot shows the 'RNC - Rilevazione di controllo dei Numeri Civici' web application. The main content area includes a search section with the following elements:

- Buttons: AGGIUNGI NUOVA STRADA, AGGIUNGI NUOVA DUG
- Input fields: DUG: \_\_\_\_\_, TOPONIMO: \_\_\_\_\_
- Section: Scarica stradario in formato xls

DUG	TOPONIMO	
VICOLO	ABBONDANZA	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ACHILLE BARILATTI	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ACHILLE CAMPANILE	MODIFICA CANCELLA AGGIUNGI SEZIONI
CONTRADA	ACQUESALATE	MODIFICA CANCELLA AGGIUNGI SEZIONI
CONTRADA	ACQUEVIVE	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ADIGE	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ADRIANO ARIANI	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ALBERICO GENTILI	MODIFICA CANCELLA AGGIUNGI SEZIONI
CONTRADA	ALBEROTONDO	MODIFICA CANCELLA AGGIUNGI SEZIONI
VIA	ALCIDE DE GASPERI	MODIFICA CANCELLA AGGIUNGI SEZIONI
GALLERIA	ALESSANDRO LUZIO	MODIFICA CANCELLA AGGIUNGI SEZIONI

Completato

Come mostrato in figura 2.6, per aggiungere una nuova strada è necessario compilare la form che appare a video dopo aver cliccato il pulsante 'AGGIUNGI NUOVA STRADA'.

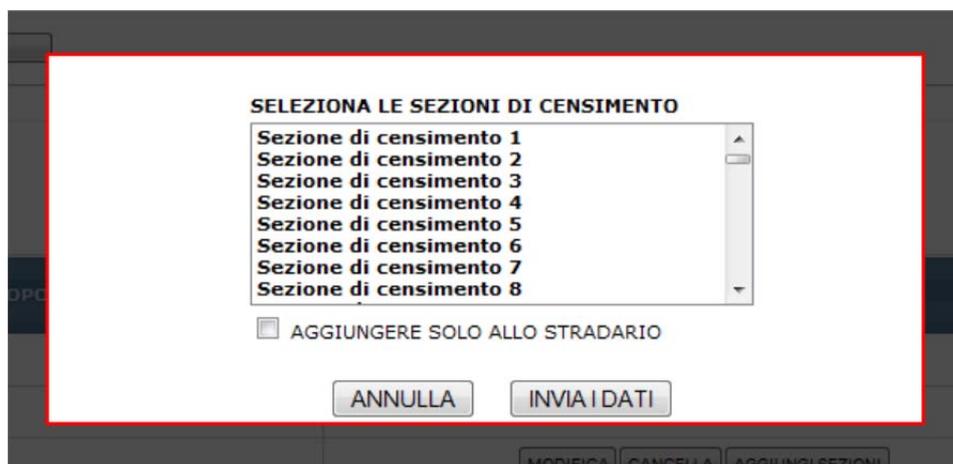
Figura 2.6 - Gestione stradario - Aggiungi nuova strada



È sufficiente, infatti, scegliere la dug, digitare il toponimo ed effettuare l'invio dei dati dopo aver scelto le sezioni di censimento (figura 2.7) cui fa riferimento la strada da inserire nel database.

Se non si conoscono le sezioni di censimento per la strada che si vuole inserire è necessario cliccare sulla casella in corrispondenza della dicitura "AGGIUNGERE SOLO ALLO STRADARIO"; in questo modo la strada viene aggiunta allo stradario ma non compare su nessuna scheda. In questo caso le possibilità per poter inserire la strada su una o più schede sono due: a) un rilevatore effettua una richiesta di inserimento toponimo esistente; b) il responsabile comunale associa la strada ad una o più sezioni.

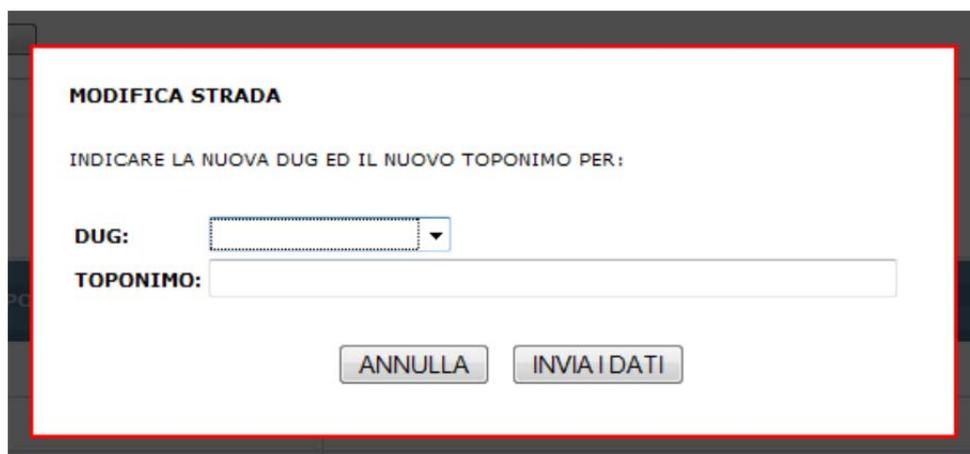
Figura 2.7 - Gestione stradario - Scelta sezioni di censimento



Attraverso il pulsante "MODIFICA", si è in grado di poter modificare un toponimo e la relativa dug. Subito dopo aver cliccato sul bottone in questione e confermata la volontà di procedere con

l'operazione di modifica di un determinata strada, compare una form dove inserire il nuovo dug ed il nuovo toponimo (figura 2.8). L'operazione di rinomina della strada viene confermata cliccando sul pulsante "INVIA I DATI".

Figura 2.8 - Gestione stradario - Modifica strada



**MODIFICA STRADA**

INDICARE LA NUOVA DUG ED IL NUOVO TOPONIMO PER:

**DUG:**

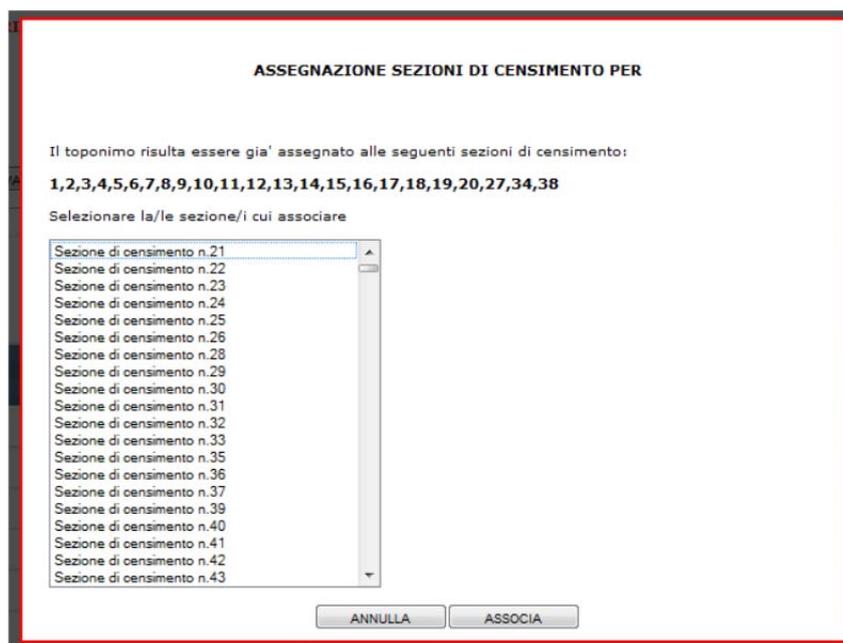
**TOPONIMO:**

Con il pulsante "CANCELLA" si ha la possibilità di eliminare dallo stradario di un comune una determinata strada e di conseguenza tutti i civici/edifici ad essa associati.

Prima di procedere con l'eliminazione di una strada il sistema chiede conferma dell'operazione all'utente.

Cliccando sul bottone "AGGIUNGI SEZIONI" l'operatore può associare ad una strada una o più sezioni (escluse ovviamente tutte quelle che fanno riferimento a schede chiuse e/o convalidate). Nella figura 2.9 è mostrato il pannello tramite il quale è possibile effettuare l'operazione in esame.

Figura 2.9 - Gestione stradario - Aggiungi sezioni



**ASSEGNAZIONE SEZIONI DI CENSIMENTO PER**

Il toponimo risulta essere già assegnato alle seguenti sezioni di censimento:  
**1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,27,34,38**

Selezionare la/le sezione/i cui associare

Sezione di censimento n.21  
Sezione di censimento n.22  
Sezione di censimento n.23  
Sezione di censimento n.24  
Sezione di censimento n.25  
Sezione di censimento n.26  
Sezione di censimento n.28  
Sezione di censimento n.29  
Sezione di censimento n.30  
Sezione di censimento n.31  
Sezione di censimento n.32  
Sezione di censimento n.33  
Sezione di censimento n.35  
Sezione di censimento n.36  
Sezione di censimento n.37  
Sezione di censimento n.39  
Sezione di censimento n.40  
Sezione di censimento n.41  
Sezione di censimento n.42  
Sezione di censimento n.43

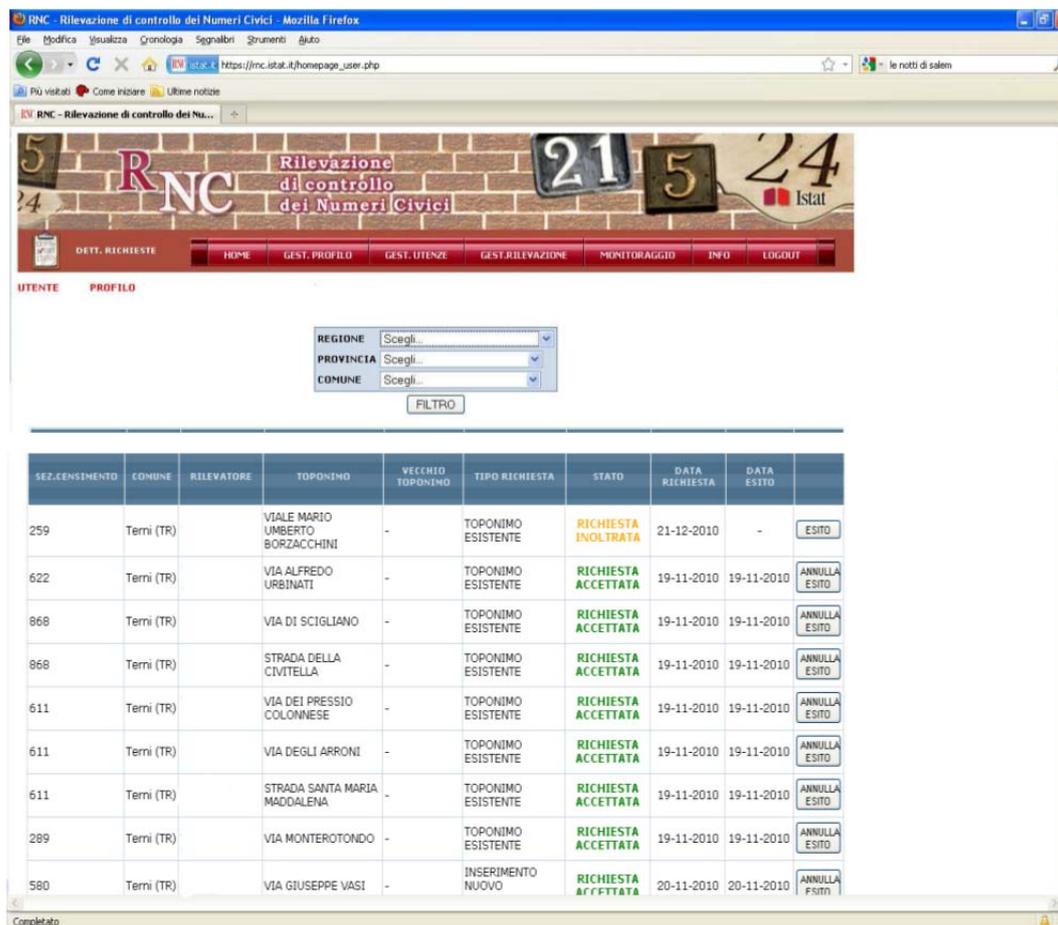
Il sistema mostra per la strada scelta tutte le sezioni di censimento associate ed una listbox che contiene le sezioni associabili. Dalle sezioni associabili sono escluse quelle che appartengono a schede la cui compilazione è chiusa.

### 2.1.5 Gestione richieste toponimo

È possibile gestire le richieste legate ai toponimi solo ed esclusivamente se il profilo dell'utente che ha effettuato l'accesso al sistema è quello di amministratore oppure di responsabile comunale. Il rilevatore infatti può solo verificare lo stato delle proprie richieste ed eventualmente eliminarle attraverso l'apposito pulsante 'ELIMINA'. Mentre il responsabile comunale vede solo ed esclusivamente le richieste per il proprio comune, l'amministratore del sistema può selezionare la regione, la provincia ed il comune di interesse. La tabella dove sono elencate le richieste mostra per ognuna di esse la sezione di censimento, il comune, il rilevatore, il tipo di richiesta ed il toponimo interessato, oltre allo stato della richiesta e alla data nella quale la richiesta è stata processata. Nel caso di utenti con profilo di amministratore oppure di responsabile comunale sono visibili per ogni richiesta due pulsanti ('ESITO' e 'SCHEDA') che permettono di processare la richiesta e di visualizzare la scheda di rilevazione.

Una volta che la richiesta viene accettata oppure rifiutata o estesa, il pulsante 'ESITO' viene sostituito da un nuovo pulsante con la dicitura 'ANNULLA ESITO' finalizzato all'annullamento dell'accettazione, del rifiuto o dell'estensione della richiesta (figura 2.10).

Figura 2.10 - Gestione richieste Toponimo



The screenshot shows the 'RNC - Rilevazione di controllo dei Numeri Civici' web application. The main content area displays a table of requests for toponym management. The table has the following columns: SEZ. CENSIMENTO, COMUNE, RILEVATORE, TOPONIMO, VECCHIO TOPONIMO, TIPO RICHIESTA, STATO, DATA RICHIESTA, DATA ESITO, and a button for 'ESITO'.

SEZ. CENSIMENTO	COMUNE	RILEVATORE	TOPONIMO	VECCHIO TOPONIMO	TIPO RICHIESTA	STATO	DATA RICHIESTA	DATA ESITO	ESITO
259	Terni (TR)		VIALE MARIO UMBERTO BORZACCHINI	-	TOPONIMO ESISTENTE	RICHIESTA INOLTATA	21-12-2010	-	ESITO
622	Terni (TR)		VIA ALFREDO URBINATI	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
868	Terni (TR)		VIA DI SCIGLIANO	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
868	Terni (TR)		STRADA DELLA CIVITELLA	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
611	Terni (TR)		VIA DEI PRESSIO COLONNESE	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
611	Terni (TR)		VIA DEGLI ARRONI	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
611	Terni (TR)		STRADA SANTA MARIA MADDALENA	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
289	Terni (TR)		VIA MONTEROTONDO	-	TOPONIMO ESISTENTE	RICHIESTA ACCETTATA	19-11-2010	19-11-2010	ANNULLA ESITO
580	Terni (TR)		VIA GIUSEPPE VASI	-	INSERIMENTO NUOVO	RICHIESTA ACCETTATA	20-11-2010	20-11-2010	ANNULLA ESITO

Figura 2.11 - Gestione richieste Toponimo - Esito



Come mostrato in figura 2.11 una volta cliccato sul pulsante 'ESITO' il responsabile comunale o l'amministratore possono decidere di accettare la richiesta oppure di rifiutarla.

È prevista tuttavia una terza possibilità che consiste nell'accettare la richiesta ed estenderla, e quindi renderla disponibile, anche alle altre schede che possono essere selezionate da un elenco che viene mostrato nel momento cui viene selezionata questa terza opzione. Una volta effettuata la scelta è sufficiente cliccare sul pulsante 'INVIA I DATI' per rendere effettivo l'esito della risposta alla richiesta.

### 2.1.6 Upload files

L'opzione di upload files è disponibile sia per gli utenti con profilo 'comunale' sia per gli utenti con profilo 'regionale', oltre che per l'amministratore del sistema, e consente loro di caricare file di due tipi: a) file che contengono le informazioni legate agli indirizzi, agli edifici ed alle scale; b) shapefile degli edifici. È possibile indicare solo file con formato zip e con una grandezza inferiore od uguale a 2 Megabyte. In figura 2.12 è mostrata la form di upload con il filtro geografico che consente di selezionare provincia e comune per l'utente 'regionale' o 'amministratore'. Prima di poter effettuare l'upload dei files è necessario effettuare una richiesta all'amministratore dell'applicazione indicando la volontà di poter procedere con la convalida massiva delle schede anziché convalidare una scheda alla volta. Inviata la richiesta, questa viene presa in consegna dall'amministratore del sistema. I file .zip relativi agli edifici, indirizzi, scale ed edifici-indirizzi, possono essere caricati in più fasi ed è possibile effettuare l'upload singolarmente (figura 2.13). Una volta effettuato l'upload di uno o più file di questo tipo, viene avviata una procedura di controllo che mostra a video gli eventuali errori riscontrati. Questi errori potranno essere scaricati in formato x/s.

Figura 2.12 - Upload files

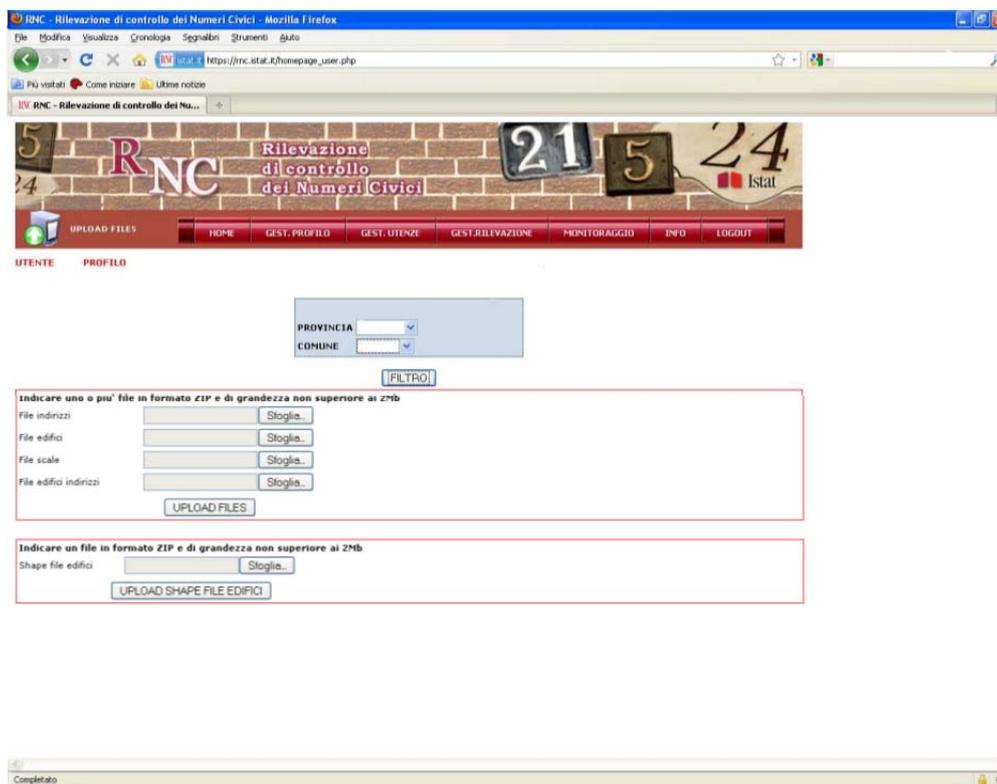


Figura 2.13 - Maschera di Upload completato



### 2.1.7 Monitoraggio

Le informazioni di monitoraggio visualizzabili grazie all'apposito pulsante mostrato in figura 2.14, come già detto in precedenza, sono legate al profilo dell'utente e alla regione e provincia d'appartenenza dello stesso. L'opzione 'MONITORAGGIO GENERALE' mostra dati sia in valore assoluto sia in valore percentuale e grazie al pulsante "DETTAGLI" presente in corrispondenza di ogni riga è possibile avere una visione particolareggiata della situazione in esame. Un utente un rilevatore o compilatore visualizza le informazioni relative alle sole schede di sua competenza e le richieste da lui effettuate. I dati mostrati in tabella riguardano essenzialmente i rilevatori/compilatori

e le schede loro assegnate, le richieste (ed il loro stato) effettuate da ogni singolo rilevatore ed il numero di utenti che fa riferimento ad un determinato comune (figura 2.15).

La voce 'MONITORAGGIO DI DETTAGLIO' permette di esaminare l'operato di ogni singolo rilevatore mostrando lo stato della sua attività di indagine in riferimento agli indirizzi, agli edifici ed ai toponimi di sua competenza. Le informazioni mostrate a video possono essere visualizzate offline grazie alla possibilità di generare file x/s cliccando sulle icone poste al di sopra delle tabelle. La voce 'AVANZAMENTO DI CONVALIDA' è una funzionalità disponibile solo per l'amministratore del sistema che può controllare lo stato di avanzamento della rilevazione per ogni singolo Comune.

Figura 2.14 - Barra dei pulsanti – Monitoraggio



Figura 2.15 - Monitoraggio generale



### 2.1.8 Logout

La barra dei pulsanti si chiude con il bottone 'LOGOUT', attraverso cui l'utente esce dal sistema tornando alla pagina principale dell'applicazione Rnc e mostrata in figura 2.16.

Figura 2.16 - Barra dei pulsanti – Logout



## CAPITOLO 3

### L'architettura del sistema integrato di acquisizione dati e gestione della rilevazione<sup>14</sup>

La realizzazione del 15° Censimento generale della popolazione e delle Abitazioni è stata caratterizzata dalla scelta strategica di rendere il web elemento cardine del processo di rilevazione e di gestione.

In particolare sono state sviluppate applicazioni web fortemente integrate che hanno permesso di gestire le diverse fasi della rilevazione:

il questionario on line (Qpop), il sistema utilizzato dai cittadini per fornire i dati censuari all' Istat, che ha permesso di raccogliere circa 8.500.000 questionari, pari al 33 per cento della totalità;

un portale di tipo prettamente documentale (Rete) per la gestione delle comunicazioni alla rete di rilevazione da parte dell'Istituto;

un sistema rivolto agli operatori della rete, per la gestione e il monitoraggio della rilevazione in tutte le sue fasi (Sgr).

L'integrazione tra il questionario on line e il sistema di gestione della rilevazione ha consentito un costante monitoraggio in tempo reale dell'andamento della rilevazione in tutte le sue fasi.

Nei prossimi paragrafi si descrive l'architettura software in generale, presentando una panoramica delle tecnologie e delle metodologie implementate.

#### 3.1 L'architettura generale

Uno degli aspetti di maggiore complessità nell'implementazione di un sistema business-critical di tali dimensioni e importanza è la difficoltà di integrare varie fonti di dati (liste anagrafiche comunali, archivio degli indirizzi e numeri civici, liste ausiliarie e dati provenienti dal Vettore Postale) e la capacità di garantire all'utenza la disponibilità dell'intero sistema senza interruzioni del servizio (24h per 7 giorni) in particolare durante le ore di picco.

Questi risultati sono stati possibili solo attraverso la progettazione e l'implementazione di un'architettura flessibile, basata su sistemi replicati e con un meccanismo di bilanciamento del carico, che si appoggi a un middleware consolidato in cui le componenti possano essere integrate facilmente e senza massicce rielaborazioni del codice sorgente.

Poiché la natura degli applicativi censuari si discosta notevolmente, per dinamicità e complessità delle soluzioni da implementare, da quella delle soluzioni Ict normalmente utilizzate per le attività standard delle pubbliche amministrazioni centrali, si è ricorso al coinvolgimento di un partner tecnologico in grado di soddisfare dei requisiti non comuni. La scelta è caduta su un operatore nazionale, che ha ospitato sui suoi sistemi di hosting gli applicativi e i database realizzati da Istat.

Quella che inizialmente era nata come un'architettura di hosting è stata in seguito potenziata e trasformata in una vera e propria *cloud* privata, con pile applicative dedicate, che si sono dimostrate adeguate a sopperire all'insorgenza di requisiti sempre più 'demanding' dovuti all'elevato afflusso di utenti dei sistemi on line nel momento della messa in esercizio dei sistemi.

Come mostrato in figura 3.1, Sgr è stato il nucleo dell'architettura complessiva, integrando i suoi moduli con l'applicazione del questionario (Qpop), usata sia dai cittadini rispondenti che dagli

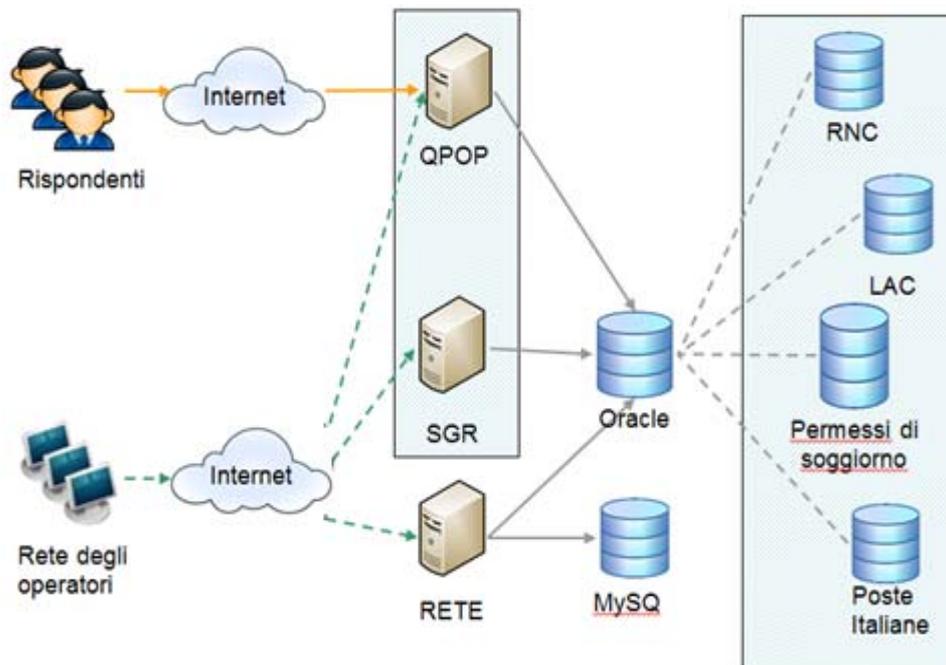
---

<sup>14</sup> Il capitolo è a cura di Maura Giacommo. Il paragrafo 3.1 è stato redatto da Marco Silipo; il paragrafo 3.2 è stato redatto da Mauro Bruno; il paragrafo 3.3 è stato redatto da Daniele Ingretolli e Rossana Vignola; il sottoparagrafo 3.4.1 è stato redatto da Linda Corradi; il sottoparagrafo 3.4.2 è stato redatto da Roberta Benedetti; il paragrafo 3.5 è stato redatto da Francesco Bosio; il paragrafo 3.6 è stato redatto da Bernardo Palazzi, Manlio Paterniti e Ferdinando Ruggiero.

operatori autorizzati della Rete di rilevazione, impegnati nel data entry dei questionari cartacei di loro competenza.

La gestione dei dati per Sgr e Qpop si è basata su una infrastruttura Oracle molto complessa, che è stata accuratamente ottimizzata per permettere la grande mole di accessi prevista da parte di operatori e rispondenti (potenziali 25 milioni di famiglie). Quello che virtualmente era visto dagli applicativi come un unico database, a livello fisico era ospitato da un set di macchine che andavano costantemente tenute allineate, gestendo aspetti critici come la *fault tolerance* e la *high availability*.

Figura 3.1 - Architettura generale del sistema di acquisizione (Qpop) e gestione (Sgr)



Le tre componenti dell'architettura generale del sistema sono Qpop, Sgr e Rete.

L'integrazione dei sistemi ha riguardato prettamente il sistema di autenticazione e la gestione degli stati del questionario.

Per quanto riguarda l'autenticazione, è stato predisposto un meccanismo di single-sign-on che permetteva ad un operatore della rete, autenticato su Sgr e con l'autorizzazione al data entry, di entrare in maniera trasparente su Qpop senza necessità di autenticarsi nuovamente per lavorare sui questionari di propria competenza.

Lo stesso meccanismo, basato su tabelle Db, permetteva agli operatori di Sgr di autenticarsi sul Portale della Documentazione, Rete, senza bisogno di un'ulteriore immissione di username e password.

Per quanto concerne gli stati di lavorazione del questionario e la gestione dei questionari cartacei ritornati all'Istat (anche attraverso il vettore postale) l'integrazione è avvenuta grazie alla possibilità per gli applicativi di condividere i dati sullo stesso database. Nel momento in cui un cittadino sottometteva il proprio questionario attraverso Qpop, le informazioni erano già disponibili per Sgr. Viceversa quando un questionario veniva immesso dall'operatore attraverso Sgr, ciò disabilitava automaticamente la compilazione on line da parte del cittadino.

Per quanto riguarda i questionari ritornati dai cittadini negli Uffici Postali, è stato predisposto un invio giornaliero da parte delle Poste Italiane attraverso file Excel elaborati automaticamente

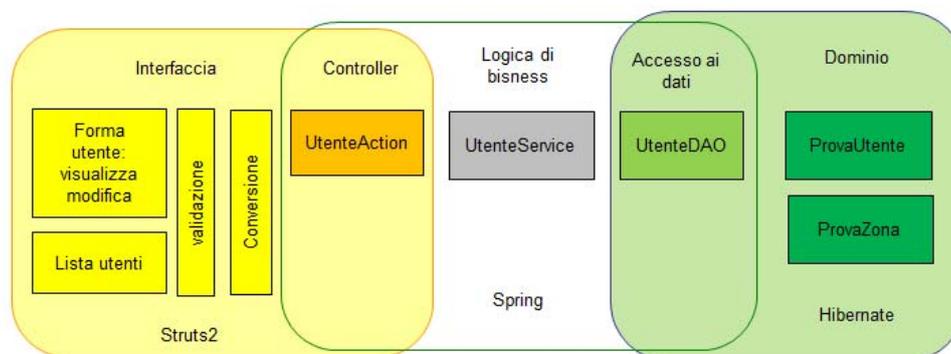
attraverso un Etl che andava ad aggiornare il database di Sgr, cambiando opportunamente lo stato dei questionari per l'intero sistema e permettendo un costante monitoraggio del numero dei questionari ritornati.

L'applicazione on line ha di per sé permesso una facile e veloce compilazione del questionario da parte dei cittadini, in quanto permetteva all'utente di seguire un percorso guidato, condizionato dalle risposte date man mano ai quesiti, verificando la correttezza formale e semantica dei valori immessi attraverso un set di regole predefinite a livello di database e lette dinamicamente dall'applicativo.

### 3.2 Metodologia e strumenti di sviluppo software

L'infrastruttura software di Sgr e Qpop si è basata sul design pattern Model-View-Controller e sulla tecnologia Java Enterprise Edition (Jee) (figura 3.2). La scelta di tale linguaggio è stata dettata sia da motivazioni tecnologiche che di valorizzazione delle competenze già presenti in Istituto, maturate durante il 6° Censimento Generale dell'Agricoltura. Il team di sviluppo responsabile della realizzazione del sistema di gestione della rilevazione ha infatti prodotto un software migliorato in termini di sicurezza e qualità, riutilizzando le componenti principali dell'infrastruttura e al tempo stesso adottando framework open source ampiamente diffusi (Struts2, Spring and Hibernate). L'utilizzo di tali framework ha determinato un impatto estremamente positivo in termini di pulizia del codice, semplicità e velocità di sviluppo e test, robustezza dell'applicazione, manutenibilità.

Figura 3.2 - Infrastruttura software di Sgr e Qpop



La figura mostra l'integrazione tra i layer applicativi e i framework utilizzati. Il layer di accesso alla base dati ha lo scopo di astrarre la complessità del Dbms facilitando le operazioni di interrogazione attraverso un semplice modello a oggetti Java, chiamati "beans": gli sviluppatori delegano al framework la generazione delle query Sql, utilizzando la sintassi Java.

In particolare il framework per l'Object-relational mapping (Hibernate) permette di semplificare tutte le funzioni di accesso alla base dati, creando un livello che associa gli oggetti utilizzati nel programma a tabelle della base dati, gestendo in maniera trasparente il caricamento degli oggetti e il loro salvataggio su base dati. L'utilizzo di questo tipo di framework consente di limitare e semplificare notevolmente il codice per l'accesso alla base dati, che risulta di conseguenza anche più facilmente manutenibile.

Il framework Spring permette la gestione dichiarativa delle dipendenze tra le classi: gli oggetti utilizzati nell'ambito di una classe vengono automaticamente creati dal framework al momento dell'uso (autowiring). Attraverso file di configurazione xml, Spring consente la gestione della connessione alla base dati, rendendola trasparente al programmatore; inoltre il framework

permette di gestire in modo dichiarativo le transazioni. Pertanto Spring, richiedendo al programmatore la sola definizione di file di configurazione per realizzare funzioni complesse e ripetitive, riduce e semplifica il codice da scrivere, rendendolo più robusto, manutenibile e meno soggetto ad errori.

Struts2 permette di organizzare una applicazione web in livelli, separando le funzionalità dei componenti di presentazione e di logica applicativa secondo un'organizzazione chiara e strutturata. Fornisce funzionalità di base per la gestione semplificata dell'acquisizione dei dati delle form e la loro validazione. Inoltre offre numerose librerie che implementano componenti di uso comune nelle pagine web, riducendo il lavoro dello sviluppatore alla composizione e configurazione di tali componenti. L'utilizzo di questo tipo di framework è essenziale per la realizzazione di una applicazione web complessa.

Le pagine dell'applicazione sono state rese dinamiche attraverso l'utilizzo della tecnologia Ajax, che permette una stretta interazione tra gli elementi dell'interfaccia utente, ad esempio pulsanti, liste, menù a tendina, etc., e le conseguenti operazioni effettuate lato server. Ciò ha consentito di guidare puntualmente l'utente nell'uso del sistema poiché in ogni momento venivano visualizzate tutte e sole le informazioni di interesse.

Tecnologie differenti sono state utilizzate per il portale Rete che è stato realizzato attraverso Typo3, un Content Management System open source ampiamente diffuso, scritto in Php e basato su una base dati MySql. L'integrazione tra Sgr e Rete, necessaria affinché un utente potesse accedere ad entrambi i sistemi utilizzando le medesime credenziali, è stata realizzata attraverso un apposito script di sincronizzazione delle basi dati.

L'implementazione di Sgr e Qpop è stata realizzata applicando consolidate regole di sicurezza (Owasp 2013, SafeCode 2013) quali ad esempio: meccanismi di autenticazione, uso di algoritmi di crittografia, livelli di autorizzazioni basati sulla profilazione utente, protezione delle Url per impedire accessi non autorizzati alle varie funzionalità. Relativamente all'autenticazione, poiché le applicazioni sono state installate su diversi server applicativi Tomcat in cluster, è stato implementato un meccanismo di Single Sign On, realizzato attraverso i realm di Tomcat e il `j_security_check`. Le password sono state cifrate (come hash Sha-256) evitando in tal modo che eventuali attacchi potessero avere accesso alle credenziali in chiaro degli utenti. L'autorizzazione è stata realizzata in modo dichiarativo attraverso un file di configurazione, il deployment descriptor, messo a disposizione dalla Java Enterprise Edition: i profili applicativi venivano associati alle rispettive funzionalità tramite la semplice definizione di tag Xml. Ciascun utente di Sgr o di Qpop, autenticato sul sistema, poteva in tal modo accedere solo alle funzionalità di sua competenza; inoltre opportune logiche di business sono state implementate per consentire di visualizzare le sole informazioni relative al proprio territorio o ai questionari assegnati.

### **3.3 L'architettura software del portale della rete**

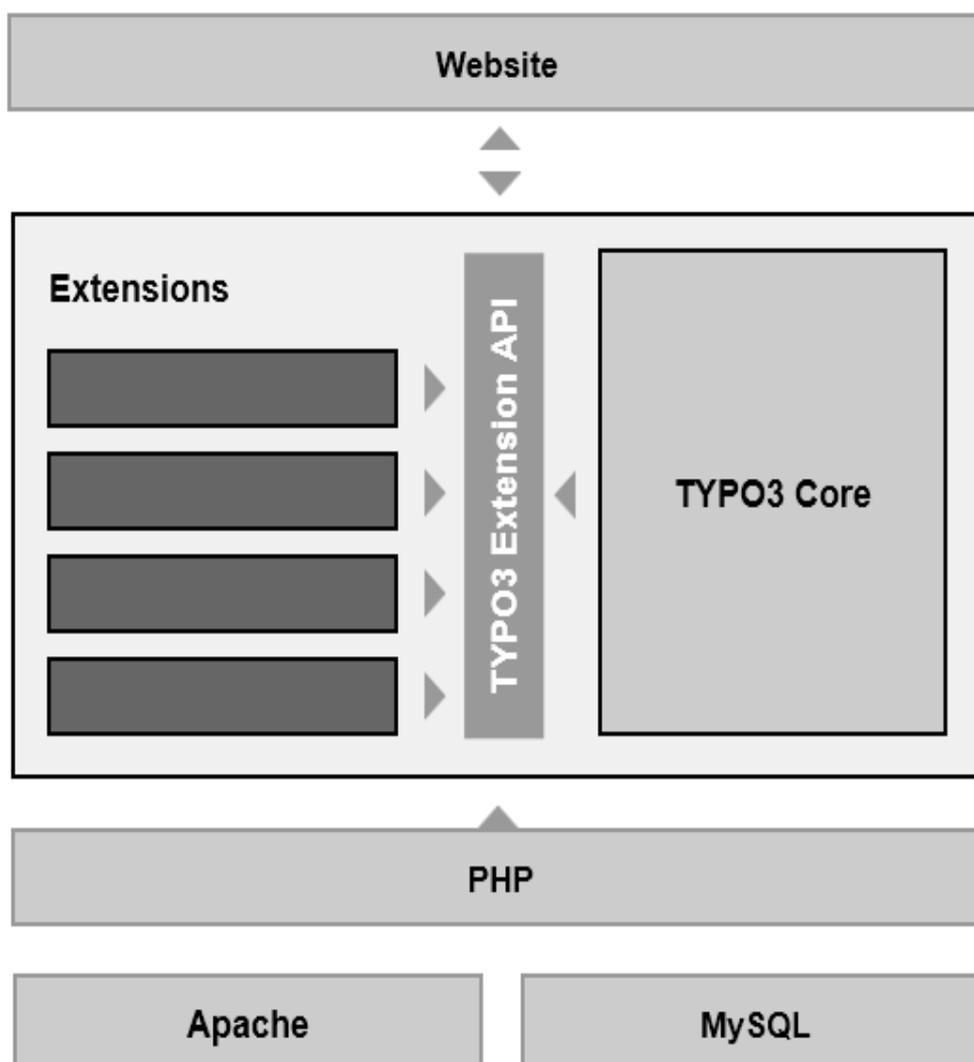
Typo3 è un Cms open source, ovvero un content management framework gratuito, progettato per semplificare la creazione di siti web e consentirne l'aggiornamento a personale non tecnico. È scritto completamente in php mentre tutti i contenuti e le configurazioni di un sito sono gestite in un database mysql. Typo3 è stato rilasciato sotto licenza Gpl ed è un progetto che viene sviluppato continuamente (figura 3.3).

La gestione del sito si basa sul Typo3 Extension Manager incluso nel Typo3-core per installazione, attivazione e disattivazione di estensioni in una determinata installazione base. L'installazione standard permette di gestire le funzionalità di base per un sito, come ad esempio l'inserimento di testi, di immagini o di link a dei documenti, ma volendo aggiungerne di più complesse si può ricorrere a estensioni che permettono di integrare molte altre funzioni. Molte estensioni sono disponibili nel repository di typo3 e possono essere velocemente importate e installate nel backend

di Typo3. L'Extension Kickstarter riduce notevolmente il tempo necessario per creare una nuova estensione Typo3 e automatizza la creazione di interfacce per i record personalizzati di database nel backend. Con questa estensione è stato possibile creare estensioni personalizzate per la gestione di tabelle aggiuntive rispetto a quelle standard di Typo3.

La gestione delle utenze è avvenuta in modo parallelo al sistema Sgr; infatti, poiché Typo3 usa una propria tabella per la gestione degli accessi, si è dovuto creare una procedura per mantenere le utenze del portale allineate a quelle di Sgr. Così ogni volta che veniva creata o eliminata un'utenza Sgr questa veniva replicata o eliminata dalla tabella degli utenti di Typo3. Per il resto invece il portale della rete è stato gestito in modo indipendente.

**Figura 3.3 - Architettura typo3**



### 3.4 La base dati

#### 3.4.1 La base di dati Sgr: struttura, popolamento e gestione

Per la gestione del Sistema di monitoraggio della rete di rilevazione è stata realizzata una base di dati Oracle versione 11g ospitata fisicamente in hosting, composta da oltre un centinaio di tabelle, circa cinquecento indici e vincoli di unicità e congruenza, oltre trecento Oracle store procedure e

function, decine di trigger per la generazione dei dati storicizzati e altrettante viste materializzate di ausilio per l'estrazione rapida dei dati.

Da un punto di vista concettuale è possibile suddividere il database di Sgr in categorie tematiche di dati gestiti e mantenuti nel corso dell'attività censuaria:

- ▶ i dati delle sezioni di censimento e degli itinerari di sezione;
- ▶ i dati provenienti da Lac (Liste Anagrafiche Comunali) di famiglie e convivenze;
- ▶ le sotto-coperture anagrafiche provenienti da Lifa (Liste ausiliarie famiglie anagrafiche) e da Rnc (Rilevazione dei numeri civici);
- ▶ i dati riguardanti gli indirizzi;
- ▶ i dati inerenti gli edifici;
- ▶ i dati di gestione della rete di rilevazione;
- ▶ i dati di log e storicizzazione generati ed utilizzati dal sistema.

In ogni tabella è stato creato un vincolo chiave di unicità attraverso cui le strutture sono state messe in relazione tra loro e grazie al quale è stata garantita la congruenza e la consistenza dei dati.

Dovendo contenere grandi quantità di dati, sono stati inseriti nei punti di maggiore accesso alla base di dati una serie di indici per l'ottimizzazione, in termini di performance temporali, delle procedure di estrazione ed elaborazione dei dati utilizzate dalle applicazioni di acquisizione on line e dal Sistema di Gestione e monitoraggio della rete di Rilevazione.

Il popolamento della base di dati è stata un'attività complessa e lunga; i processi Oracle di caricamento si sono diversificati in base alla fase di lavorazione ed al tipo di dato trattato. Il primo elemento di difficoltà è stato dato dalla pluralità delle fonti a cui attingere per il popolamento delle tabelle: basi di dati Oracle, file tipo csv, file formato excel, ecc. Primaria è stata quindi anche l'attività di integrazione delle informazioni che ha coinvolto in sinergia le unità di competenza statistica ed informatica.

Con l'obiettivo di snellire la fase di controllo e minimizzare al massimo gli eventuali ricicli dei dati nella base di dati di produzione in hosting (eliminazione e successivo reinserimento), è stata realizzata una struttura di base dati interna Istat, posta come passo intermedio tra le varie fonti di acquisizione ed il sistema di produzione finale.

Data l'elevata numerosità dei dati in termini di occorrenze, le procedure previste per l'inserimento nelle strutture tabellari hanno avuto tempi di esecuzione non indifferenti, in particolar modo per i grandi comuni. Per agevolare la migrazione e trasformazione dei dati senza gravare troppo sulla rete informatica dell'Istituto, sono stati studiati ed implementati job Oracle con esecuzioni notturne asincrone direttamente sui server virtuali dedicati. In questa attività di preparazione della base di dati è stata importante la collaborazione con i sistemisti dello hosting che hanno supportato con continuità il gruppo Istat dedicato alla gestione ed al monitoraggio dei job Oracle nelle fasi di caricamento massivo, avvenuto principalmente nelle ore notturne della giornata.

Fondamentale e capillare è stata la doppia fase di controlli di congruenza e consistenza, sia a livello intermedio che conclusivo, sulla totalità della base di dati realizzata attraverso routine Oracle previste a fine caricamento (figura 3.4)

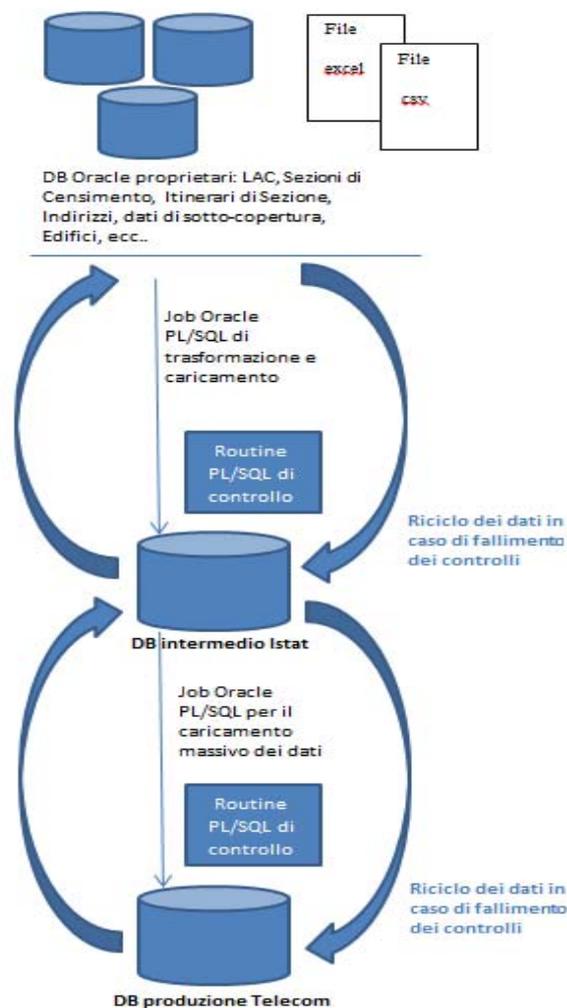
Le informazioni inerenti le sezioni di censimento, gli indirizzi, le Lac di famiglie e convivenze, gli edifici e tutto l'insieme di informazioni utili al sistema Sgr per il suo funzionamento, sono state tra le prime ad essere caricate all'interno delle base di dati con una operazione continua nel tempo.

Un secondo ciclo di popolamento dati è stato effettuato a Censimento avviato, quindi parallelamente all'utilizzo già possibile, a cittadini e rilevatori, del sistema di acquisizione on line Qpop e di Sgr. Questa fase successiva di inserimento, trasformazione e controllo dei dati è stata molto delicata e più lenta della precedente attività informatica.

Per non interferire con i sistemi censuari già aperti agli utenti, le operazioni si sono concentrate nelle sole ore notturne della giornata e le verifiche di coerenza e consistenza dei dati non sono state effettuate solo a fine ciclo di caricamento, ma anche per ogni singolo dato migrato, mantenendo quindi la base dati sempre in condizioni di funzionamento ottimale e sicuro.

Il Sistema Sgr durante tutto il periodo di esercizio dell'applicazione web è stato continuamente monitorato, mantenuto e aggiornato come è accaduto per la fase di revisione dei dati anagrafici, di indirizzo e sezioni di censimento, a fronte della seconda acquisizione della Lac, in cui si è proceduto a modificare ed integrare tutte le informazioni anagrafiche variare nel tempo rispetto al primo caricamento.

Figura 3.4 - Flusso dei controlli tramite PI/Sql



### 3.4.2 Amministrazione e tuning della base dati

Le attività di amministrazione e *tuning* della base dati Oracle si sono concentrate principalmente durante le fasi di caricamento, controllo e correzione dei dati censuari.

L'attività degli amministratori della base dati ha comportato, in particolare, la creazione di 10 utenze Oracle, sia di sviluppo che di esercizio, afferenti a più database localizzati nei vari server interni all'Istituto. Tale creazione è stata subordinata a preliminari operazioni di individuazione dei

server Linux più adatti per le esigenze di dimensionamento e delle istanze Oracle più appropriate per le condizioni di utilizzo delle utenze da creare.

La fase di acquisizione dei dati censuari, provenienti sia dai questionari compilati on line, sia da quelli cartacei preventivamente scansionati tramite processi di lettura ottica, ha comportato una costante attività di monitoraggio dell'effettiva funzionalità ed efficacia delle operazioni messe in atto per il caricamento dei dati stessi sulle varie tabelle; tale attività si è tradotta principalmente nello svolgere operazioni di controllo dei processi Oracle connessi all'esecuzione di job, funzioni e stored procedure.

Contestualmente alla fase di acquisizione e alla successiva fase di gestione dei dati, sono state effettuate continue operazioni di monitoraggio ed ottimizzazione degli spazi su disco, sia tramite controllo visivo attraverso i software di amministrazione delle basi dati a disposizione dell'Istituto, sia tramite la creazione di procedure e job Oracle progettati e sviluppati ad hoc. Frequenti sono stati i contatti con il gruppo dei sistemisti Linux per eventuali richieste di allargamento dei filesystem che ospitavano le varie utenze Oracle; contatti a cui spesso hanno fatto seguito operazioni di estensione e adattamento di tablespace Oracle laddove necessario, in funzione delle mutevoli esigenze di spazio derivanti dalle attività di correzione e controllo dei dati censuari.

Periodicamente sono state inoltre effettuate operazioni di import ed export delle utenze Oracle o di porzioni di esse, sia per la creazione di backup specifici non inquadrabili nell'ambito delle politiche standard di backup dell'Istituto, sia per la migrazione su istanze ritenute di volta in volta più efficienti e più adatte ad ospitare i dati censuari, a seguito di sopravvenute e modificate necessità di gestione, aggiornamento e futura diffusione dei dati stessi.

Occorre infine ricordare come per alcune delle suddette attività di amministrazione necessarie al corretto ed efficiente funzionamento dei database Oracle, con particolare riferimento a quelle legate al controllo dei processi Oracle di caricamento e correzione dei dati censuari, è stato fatto tesoro dell'esperienza acquisita durante lo svolgimento delle stesse attività effettuate nel corso della Rilevazione censuaria pilota del 2009.

### 3.5 Le viste materializzate

Per la realizzazione dei rapporti riassuntivi è stato sviluppato un meccanismo ad hoc, che ha previsto, tra l'altro, l'utilizzo di insiemi di viste materializzate.

Data la dimensione dei dati gestiti (nell'ordine di decine di milioni di righe per tabella), infatti non era pensabile visualizzare i rapporti riassuntivi mediante query ad hoc direttamente sui microdati. Ciò ha obbligato a creare tabelle che contenevano, già calcolati, i dati da visualizzare nei rapporti. Il ricalcolo veniva effettuato periodicamente; in tal modo i rapporti riassuntivi non contenevano mai informazioni aggiornate in tempo reale.

È stato scelto un approccio in cui, a partire dai dati della rilevazione (**microdati**) e tramite vari passaggi che prevedono l'utilizzo di viste materializzate e tabelle temporanee, venivano calcolate le tabelle finali, contenenti i dati da visualizzare nei rapporti riassuntivi (figura 3.5).

Questo approccio ha consentito di ottenere che:

- 1) i rapporti riassuntivi contenessero dati coerenti tra loro. Per far ciò era necessario che essi si riferissero tutti allo stesso istante di tempo;
- 2) il divario di tempo tra l'istante a cui era riferito il rapporto riassuntivo e quello in cui era visualizzato fosse minimo.

Per ottenere quanto specificato al punto 1), in generale, occorre predisporre periodicamente di una copia dei microdati (o di un sottoinsieme di essi) riferita ad un determinato istante di tempo. Da questa "fotografia" dei dati sono poi calcolati quelli visualizzati nei rapporti riassuntivi. Tale "fotografia" è rappresentata dalle **viste materializzate di primo livello**, ottenute direttamente dai

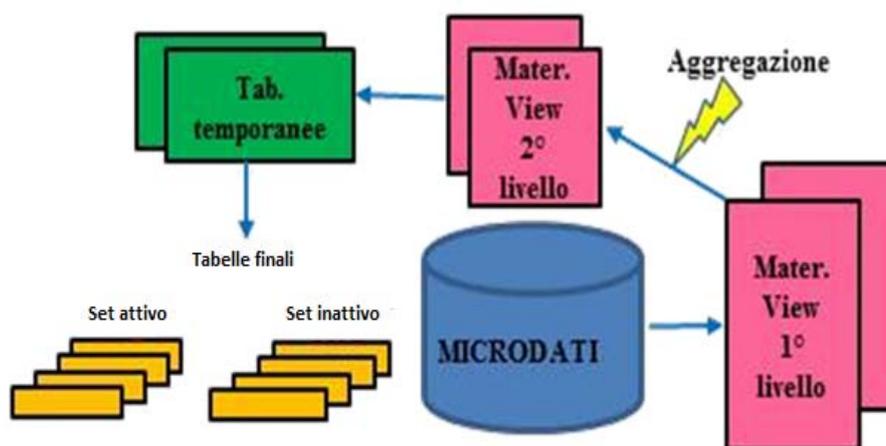
microdati. In essa i dati non sono aggregati e lo spazio occupato da queste viste è paragonabile a quello occupato dalle tabelle dei microdati. La loro sincronizzazione si ottiene utilizzando l'apposito oggetto Oracle Refresh Group.

A partire dai dati delle viste materializzate di secondo livello vengono create quelle di **secondo livello**, tramite aggregazione dei dati.

Per minimizzare i tempi di refresh di tutte le viste materializzate (sia quelle di primo livello che quelle di livello successivo) viene utilizzata l'opzione **Refresh Fast**. Essa avviene attraverso l'utilizzo degli Materialized View Log, che memorizzano le modifiche ai dati a partire dall'ultimo Refresh effettuato.

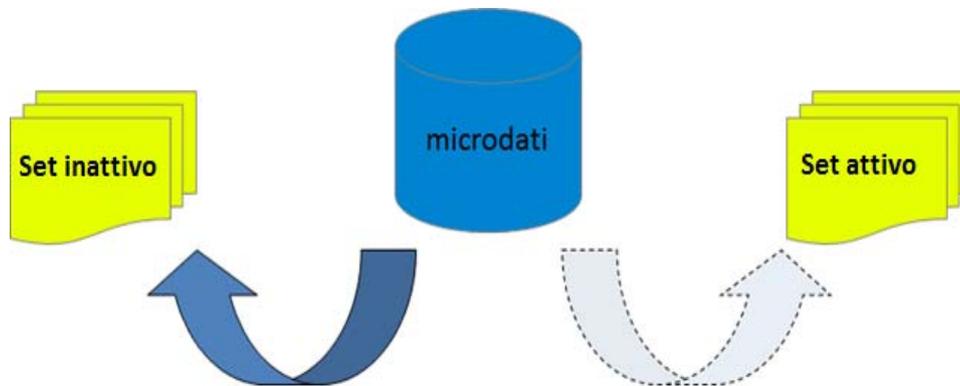
A partire dai dati delle viste di primo livello si ottengono, per **aggregazione**, quelle di **secondo livello**.

Figura 3.5 - Architettura delle viste materializzate dei rapporti riassuntivi



Come è noto, le viste materializzate sono ottenute tramite una semplice query Sql a partire dai dati sottostanti – per tali query, oltretutto, vi sono molte restrizioni, soprattutto quando viene utilizzato il Refresh Fast. Data la complessità a volte insita nei rapporti riassuntivi, era impossibile calcolare i dati attraverso un solo comando Sql. Per tale motivo è stato necessario spesso passare attraverso **più passaggi - fino a 4 livelli di viste materializzate**, laddove ogni vista veniva ottenuta a partire da quella di livello superiore. In alcuni casi, quando la complessità del rapporto riassuntivo richiesto lo richiedeva, è stato anche necessario far uso di **codice procedurale PL/Sql** direttamente su dati già aggregati, in modo da minimizzare il tempo di elaborazione richiesto. L'output di tali procedure è stato immagazzinato in apposite **tabelle temporanee**. **Le tabelle di dati finali sono state suddivise in due insiemi**. In ciascun istante uno era accessibile per la consultazione (l'insieme attivo) mentre l'altro veniva aggiornato (l'insieme inattivo) (figura 3.6). Terminato l'aggiornamento di un insieme di tabelle, avveniva lo **switch**: al successivo refresh dei dati, veniva aggiornato l'insieme precedentemente consultabile e reso accessibile l'altro.

Figura 3.6 - Le tabelle finali dei rapporti riassuntivi



### 3.6 Test e prestazioni del sistema

Lo scopo primario delle attività di test delle applicazioni e dei sistemi è stato di dotare l'Istituto di un supporto specializzato nella revisione qualitativa dell'architettura software per la gestione informatica del Censimento Generale della Popolazione, in particolare prima che esso venisse messo in esercizio e nel supporto evolutivo durante lo svolgimento della rilevazione.

Le attività di test funzionali e la loro automatizzazione hanno avuto come obiettivo il miglioramento della qualità del software sviluppato, tramite un processo continuo di verifica e segnalazione agli sviluppatori di problemi individuati nelle fasi di sviluppo e pre-release del software.

La code review si è occupata della ricerca di possibili criticità e vulnerabilità presenti nel codice e nella verifica del rispetto delle "best practice" internazionali (ad es. Owasp, Nist) in materia di sviluppo ed è stata svolta con il supporto di una società di consulenza esterna all'Istituto.

Il servizio di test di carico per il dimensionamento dell'infrastruttura è servito per verificare che l'infrastruttura, in hosting presso un Isp (Internet Service Provider) all'interno del contratto quadro Spc, fosse adeguatamente dimensionata per gli ingenti carichi di accesso al software censuario. Queste attività di test sono state svolte da una società esterna direttamente presso l'Isp.

La strategia adottata per i test funzionali ha permesso la verifica della correttezza funzionale rispetto ai requisiti del committente e ha contribuito alla realizzazione di controlli mirati alla valutazione di altri fattori di qualità come ad esempio: affidabilità, usabilità (si vuole valutare la facilità d'uso dell'applicazione da parte dell'utente finale), efficienza, sicurezza, interoperabilità, prestazioni, robustezza, in modo da garantire un prefissato livello di qualità del prodotto.

La progettazione dei test per il software a supporto della rilevazione censuaria si riassume nei seguenti punti:

- 1) analisi dei documenti delle specifiche funzionali, dei requisiti e dei documenti di progetto,
- 2) individuazione delle criticità per ogni funzionalità prevista;
- 3) definizione dei test case mirati alla valutazione delle singole funzionalità dell'applicazione. Per ogni test case, in base alle specifiche di progetto, occorre identificare le pre-condizioni che devono essere soddisfatte prima dell'esecuzione del test, l'insieme di azioni da intraprendere per effettuare il test, e il comportamento atteso dall'applicazione in seguito a tali azioni (risultato atteso);
- 4) nella definizione dei test case bisogna tener presenti gli opportuni "criteri di copertura":
  - ogni funzionalità deve essere eseguita almeno una volta;
  - per ogni funzionalità si devono creare azioni che simulino i differenti casi d'uso;

- definito il dominio dei dati per i campi compilabili dagli utenti si effettuano test case ottenuti selezionando:
  - valori in posizione centrale;
  - valori di frontiera (tra sottodomini);
  - valori “speciali”;
  - Precondizioni e postcondizioni per test case positivi e negativi (invalid input data, invalid output data);
- 5) aggregazione di più test case in modo da realizzare un test plan specifico per ciascuna funzionalità completa;
- 6) avvio della fase iterativa di test e tracciamento dei test effettuati (test run): occorre tracciare l'identificativo del test case in esame, e il risultato (positivo o negativo) corredato da alcune informazioni aggiuntive (data, versione del prodotto);
- 7) implementazione di procedure automatiche per la ripetibilità dei test e la validazione di casi limite;
- 8) in caso di riscontro di un'anomalia, segnalazione del bug con tutte le informazioni utili allo sviluppatore per riprodurre l'anomalia, insieme alla descrizione del comportamento errato (Risultato effettivo). I test, in presenza di eventuali variazioni delle specifiche, vanno rivisti sin dalla definizione dei test case e vanno ripetuti una volta corretta l'anomalia e in generale ad ogni nuovo rilascio parziale o definitivo, in modo da verificare la correzione di errori già segnalati o l'introduzione di nuovi errori (test di regressione);
- 9) rilascio della documentazione relativa ai risultati delle campagne di test.

A supporto dell'attività di test sono stati utilizzati i seguenti strumenti:

- ▶ Selenium Software: Framework per il test del software per applicazioni web. Selenium fornisce uno strumento per registrare e riprodurre test, attraverso un add-on di Firefox, senza la necessità di apprendere uno specifico linguaggio di scripting. La riproduzione dei test, al fine di testarne la compatibilità è possibile nella maggior parte dei moderni browser web, in modo semi automatico, tramite l'utilizzo dell'estensione server di Selenium;
- ▶ Bugzilla è un bugtracker, cioè un applicativo software usato generalmente dai programmatori per tenere traccia delle segnalazioni di bug all'interno del codice, in modo che tali errori siano facilmente tracciabili e controllabili, con una descrizione della riproducibilità e dei dettagli ad essi correlati, e dunque più facilmente risolvibili. Bugzilla inoltre permette agli utenti di riportare direttamente report di errori contribuendo, di fatto, al perfezionamento del prodotto in oggetto. Tale strumento dovrebbe essere auspicabilmente utilizzato anche nella fase di beta testing del progetto.

Vista l'importanza della rilevazione censuaria e data la presenza di società di consulenza esterna, l'attività di test è stata documentata con l'utilizzo di report formali. A tal proposito sono stati presentati tutti i test condotti, sia superati (pass) che falliti (fail), mentre non sono trattati i pretest; questa tipologia di test è stata omessa poiché troppo ridondante e poco interessante ai fini della reportistica. Il vantaggio principale della redazione di un documento di reportistica è stata la descrizione di tutti i test condotti, in contrapposizione al tool Bugzilla in cui sono riportati solamente i test che hanno comportato un fallimento del sistema.

Un test run è stato indicato dal nome del test case di riferimento seguito dalla data di esecuzione; in ciascuna tabella è stata inserita una voce per distinguere agevolmente tra test run e test case. I campi presenti nelle schede sono:

- ▶ test name – Il nome del test case;
- ▶ data esecuzione – Data esecuzione del test e orario;

- ▶ descrizione – Descrizione sintetica dell’obiettivo del test;
- ▶ tester – Funzione in test e riferimento al caso d’uso (nome+versione) sul sistema di gestione dei documenti di progetto (Rational Composer);
- ▶ risultato atteso – Specifica da parte del tester del risultato atteso;
- ▶ risultato effettivo – Se coincide con il risultato atteso si può omettere (Pass) se no si descrive l’errore (Fail + descrizione errore);
- ▶ criticità – Livello di criticità come indicato su Bugzilla;
- ▶ precondizioni – Lista di condizioni che devono essere soddisfatte prima dell’esecuzione del test, potrebbero essere dettagliate nella test suite (poiché comuni a differenti test case);
- ▶ testing steps – Dettaglio dei passi per eseguire il test mediante un file del tool scelto per il test (Selenium) in allegato;
- ▶ test Case / Test Run – Indica se è un test case o un test run;
- ▶ test Case di Riferimento – Se è un test run, il nome del test case a cui si riferisce;
- ▶ ruolo – Il ruolo dell’utenza con cui è stato eseguito il test.

Un esempio di reportistica è indicato nella figura 3.7.

**Figura 3.7 - Esempio di documentazione per la segnalazione di un test non superato**

UseCase: **Operatori - Sgr\_ASS\_Assegnazione Rilevatori ai Coc**

Test Name	Sgr_ASS_AssegnazioneRilevatoriAiCoc_TastoAnnulla20110523_0521
Data Esecuzione	23-mag-11
Descrizione	Funzionamento del tasto Annulla nella ricerca Coc
Tester	Msp
Risultato Atteso	Quando si preme il tasto Annulla ci si aspetta che i campi del form vengano cancellati
Risultato Effettivo (Pass – Fail)	Fail: premendo il tasto -Annulla il form non viene cancellato-
Criticità	Normal
Precondizioni	Si fa l-autenticazione-->GestioneOperatori-->Assegnazioni-->Rilevatori ai Coc--> Inserisco - abc- nel campo Id operatore--> inserisco -123- nel campo cognome--> Faccio click sul pulsante -Annulla-.
Testing Steps	
TestCase/TestRun	TestRun
TestCase di Riferimento	Sgr_ASS_AssegnazioneRilevatoriAiCoc_TastoAnnulla
Ruolo	ducc respucc
Note	<a href="http://censagr-appl/bugzilla/show_bug.cgi?id=31">http://censagr-appl/bugzilla/show_bug.cgi?id=31</a>

Nel report di documentazione delle attività di test sono stati indicati in modo aggregato le attività di test run con il relativo esito (figure 3.8, 3.9, 3.10).

Figura 3.8 - Esiti delle attività di test run

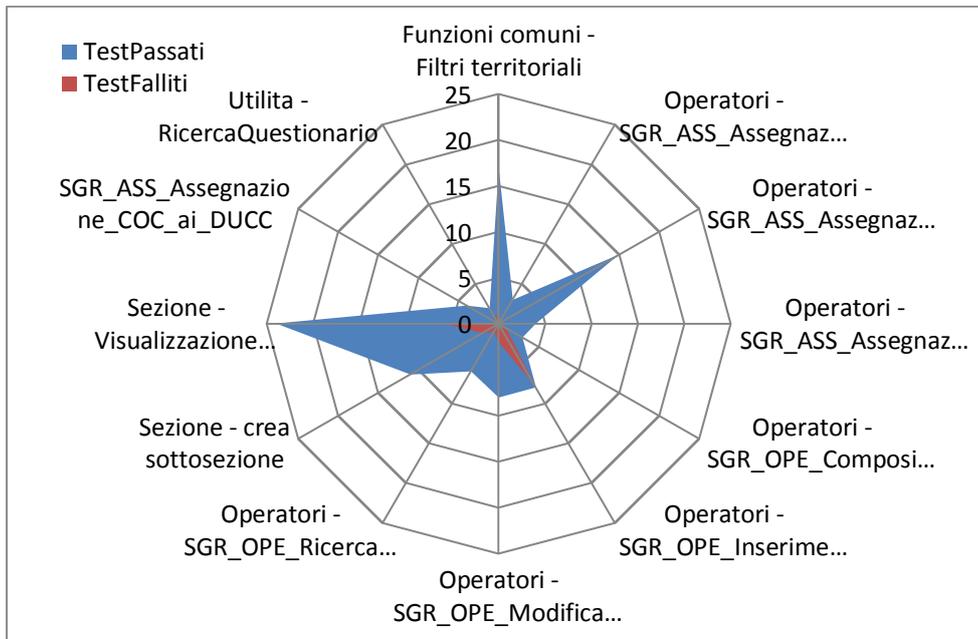


Figura 3.9 - Test passati

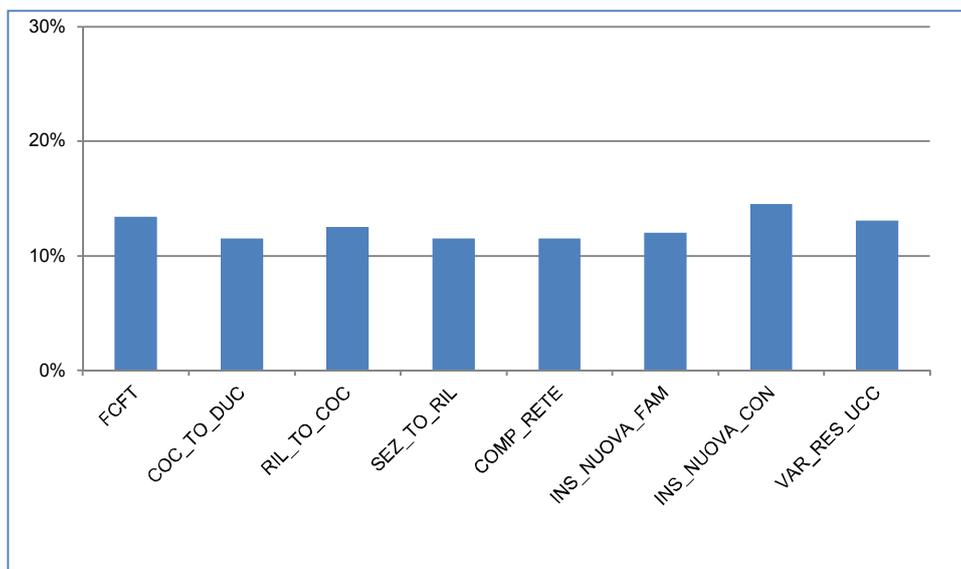
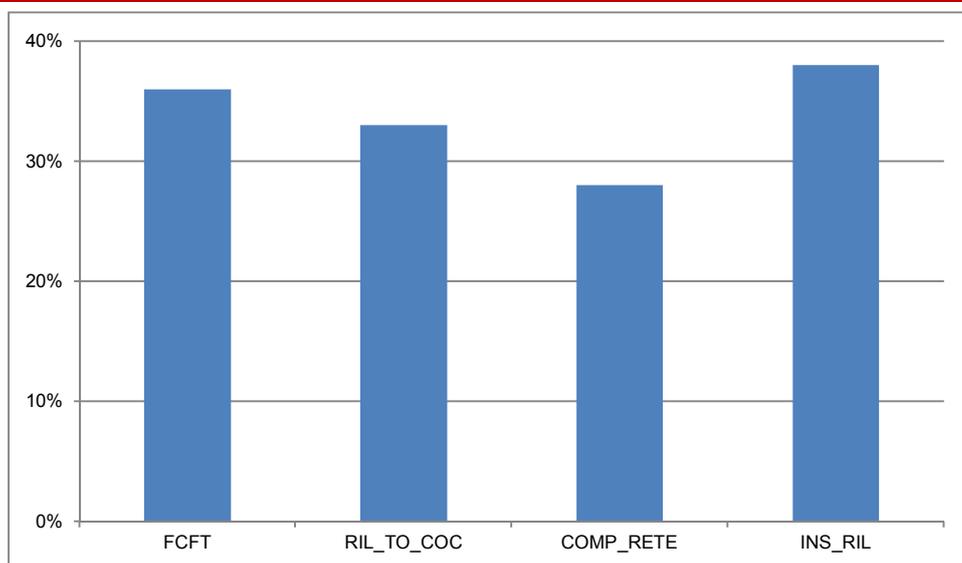


Figura 3.10 - Test falliti



## CAPITOLO 4

### il sistema di acquisizione<sup>15</sup>

#### 4.1 Caratteristiche generali

Una tra le caratteristiche innovative nell'organizzazione e conduzione dell'ultimo censimento della popolazione e delle abitazioni è sicuramente rappresentata dall'introduzione di un sistema di acquisizione dati via web, che ha affiancato la tradizionale raccolta dati attraverso questionario cartaceo.

Il sistema (Qpop) è una applicazione web che è stata utilizzata dai rispondenti per compilare il questionario censuario in modo concettualmente equivalente al questionario cartaceo. Un codice di autorizzazione è stato generato automaticamente e inviato a ogni rispondente, che lo ha utilizzato per creare la propria utenza per l'accesso al sistema. L'applicazione riproduce tutte e tre le tipologie di questionario cartaceo, ovvero le due form per le famiglie (long e short) e il questionario delle convivenze.

Qpop è pienamente integrato con il sistema di gestione della rilevazione Sgr; non appena un rispondente completava la compilazione di un questionario attraverso Qpop, lo stato di lavorazione del questionario veniva immediatamente aggiornato nel server centrale e visibile agli operatori attraverso Sgr.

Data l'evidente criticità dell'applicazione, la massima accuratezza è stata destinata al progetto tecnico del sistema. In primo luogo tutta l'applicazione è basata su un livello di metadati, appositamente concepito per consentire la memorizzazione su base dati sia di tutta la struttura del questionario (testi e tipologie delle domande, percorsi di risposta) che dei controlli di coerenza. Questa scelta ha consentito di poter efficacemente scalare sui tre differenti questionari e, nell'ambito del progetto tecnico complessivo, di implementare rapidamente le varie tipologie di domande presenti nel questionario e l'architettura dei controlli. Inoltre, questa soluzione ha rappresentato una solida base per un riutilizzo futuro dell'intera architettura applicativa.

Anche dal punto di vista dell'architettura interna del sistema sono state introdotte innovazioni significative per rispondere agli stringenti requisiti lato utente. Da un lato era di primaria importanza consentire all'applicazione di scalare a un grande numero di utenti contemporanei (la platea potenziale era costituita da circa 25 milioni di famiglie rispondenti), dall'altro di garantire la massima robustezza (nessun errore applicativo visualizzato dagli utenti) e la sicurezza dei dati, un aspetto critico in una applicazione esposta a un vasto pubblico e dall'alta visibilità mediatica. Sono state adottate tecniche sofisticate nel progetto dell'applicazione per a) ridurre il carico verso la base dati anche in presenza di un alto numero di utenti connessi all'applicazione e b) eseguire la validazione dei dati in diversi punti dell'applicazione per garantire la consistenza dei dati immessi.

Circa 8 milioni e 500 mila questionari sono stati raccolti attraverso l'applicazione on line, corrispondenti a più del 33 per cento del totale dei questionari attesi. La cifra assoluta di questionari compilati on line in Italia è la seconda più alta a livello mondiale nel round censuario 2010<sup>16</sup>. Il carico medio nei primi due mesi di operazioni è stato di circa 115 questionari completati al minuto, con un picco di 500 nei periodi di massimo carico.

<sup>15</sup> Il capitolo è a cura di Marina Venturi. Il paragrafo 4.1 è stato redatto da Antonino Virgillito; il paragrafo 4.2 è stato redatto da Leonardo Tininini; il sottoparagrafo 4.3.1 è stato redatto da Loredana Giunta; i sottoparagrafi 4.3.2, 4.3.3, 4.3.4 sono stati redatti da Francesco Cuccia.

<sup>16</sup> Fonte: "Measuring Population and Housing - Practices of UNECE countries in the 2010 round of censuses (2013)" - <http://www.unece.org/stats/census.html>

Il capitolo è articolato in due parti: nella prima (paragrafo 4.2 e sottosezioni) viene presentata la modellazione del questionario, la seconda (paragrafo 4.3 e sottosezioni) è dedicata al progetto del sistema.

## 4.2 La modellazione del questionario

In questa sezione viene illustrata la tecnica di modellazione utilizzata per i questionari elettronici del Censimento della popolazione e abitazioni. Pur nascendo con esigenze molto specifiche e per certi aspetti uniche, in particolare garantire la massima efficienza di gestione di dati e metadati in presenza di decine di migliaia di accessi contemporanei al sistema, la tecnica è stata sviluppata in modo da risultare, per quanto possibile, generalizzabile a contesti molto diversi da quello del Censimento stesso. I requisiti fondamentali che hanno portato allo sviluppo di tale tecnica per la descrizione formale della struttura del questionario attraverso tabelle di metadati sono elencate nel seguito.

- ▶ supporto al multilinguismo attraverso il caricamento delle varie versioni localizzate dei testi direttamente nella base di metadati;
- ▶ modellazione e memorizzazione nella base di metadati dei vari percorsi di compilazione del questionario, corrispondenti alle risposte fornite dall'intervistato. In sostanza si è voluto rappresentare, attraverso un apposito formalismo, informazioni del tipo: "Se alla domanda X si è risposto che l'individuo è nato all'estero, passare a compilare la domanda Y";
- ▶ sfruttamento delle informazioni di cui al punto precedente attraverso un meccanismo automatizzato che consentisse di attivare/disattivare le domande nell'interfaccia utente;
- ▶ realizzazione di un meccanismo di corrispondenze automatiche, basate su standard di nomenclatura e le proprietà di *reflection* di Java, tra domanda e componenti (*bean*) Java fino alla specifica colonna corrispondente nella tabella dei microdati;
- ▶ massimizzazione del riuso interno del codice programmatico attraverso un processo di astrazione basato su un numero limitato di *tipi di domanda* e relativi moduli applicativi. In sostanza una volta implementato il modulo che realizza il comportamento (visualizzazione nell'interfaccia utente e logica applicativa) di un certo tipo di domanda (ad esempio una "domanda con scelta multipla semplice"), esso può essere riutilizzato (ovviamente con parametri diversi) per implementare tutte le domande di quello stesso tipo.

### 4.2.1 La modellazione della struttura del questionario

La prima famiglia di metadati oggetto della modellazione è quella che potremmo chiamare la componente statica del questionario. Essa corrisponde di fatto alla sequenza ordinata dei singoli quesiti unitamente a tutti i testi ad essi associati. Gli obiettivi principali di questa tipologia di metadati sono:

- ▶ memorizzazione della struttura di base del questionario, costituita essenzialmente dall'elenco dei quesiti che lo compongono, ed eventualmente delle corrispondenti modalità di risposta predefinite, con relativi testi nelle varie localizzazioni (lingue) previste;
- ▶ realizzazione di una corrispondenza automatica tra i vari componenti dell'interfaccia utente visualizzati, i componenti costitutivi dell'applicazione (Java Beans) e i metadati scritti/letti nella base dati;
- ▶ realizzazione di alcuni controlli di base sull'input dell'utente.

Si elencano di seguito le componenti fondamentali dedicate alla modellazione della componente statica del questionario e le loro funzioni.

*Il tipo di questionario.* Con questo termine si indica di fatto la sequenza di quesiti che costituiscono un intero questionario o una sua sezione (ad esempio la sezione I del questionario "long form").

Ogni tipo di questionario è identificato nella base dati da un codice univoco di identificazione e da una descrizione testuale che ne chiarisce brevemente il contenuto.

*Le domande.* Questa componente fondamentale di modellazione rappresenta non solo ogni singolo quesito presente nel questionario, ma anche altre componenti complementari, come titoli di sezione, istruzioni di compilazione e perfino componenti *virtuali*, invisibili all'utente e finalizzate, in casi particolari, alla corretta realizzazione del flusso di compilazione. Nella tabella delle domande sono memorizzate varie informazioni relative al quesito, in particolare la tipologia (quesito con modalità di risposta singola ed elenco di possibili risposte predefinito, quesito con risposta numerica oppure testuale, quesito con risposta di tipo data, etc.), i testi ad esso associati nelle varie versioni localizzate (italiano, tedesco e sloveno), se si tratta o meno di una domanda con risposta opzionale da parte del compilatore. Si è parlato di testi (e non di singolo testo) associati ad un certo quesito, perché spesso il quesito viene formulato attraverso più testi diversi con finalità e stili di visualizzazione diversi, ad esempio un testo principale in neretto attraverso il quale viene formulata la domanda vera e propria, un altro scritto in piccolo e/o colorato che fornisce informazioni ausiliare (tipicamente degli esempi al compilatore), ecc.

*La composizione del questionario.* La composizione vera e propria di un tipo di questionario è memorizzata in una tabella apposita che mette in relazione la tabella dei tipi di questionario con quella delle domande. Ne consegue che la stessa domanda può essere condivisa da più tipi di questionario (tipicamente questo accade quando una stessa domanda è presente sia nella long form che nella short form) semplicemente inserendo più record per la medesima domanda (e tipi di questionario diversi) in questa tabella. Tale tabella descrive (attraverso un apposito indice numerico) anche l'ordine secondo il quale una certa domanda è presente in un determinato questionario: in generale la stessa domanda può essere presente in posizioni diverse in tipi di questionari diversi.

*I tipi di domanda.* Come detto sopra, ogni domanda è associata ad uno specifico tipo, che ne determina la visualizzazione nell'interfaccia utente, il comportamento e la memorizzazione delle relative risposte nella base dati. Se ad esempio una domanda è del tipo "domanda con scelta singola da una lista di risposte", il frammento di codice corrispondente prevede la visualizzazione di una serie di *radio button* con associate le corrispondenti risposte possibili (anch'esse memorizzate nella base dati), e la scrittura/lettura di un codice numerico nella tabella dei microdati, corrispondente alla scelta del rispondente. Tutte le domande di tipo "domanda con scelta singola da una lista di risposte" utilizzano i medesimi moduli software, massimizzando il riuso e la robustezza del codice prodotto. Di fatto quindi il catalogo dei tipi di domanda supportati (memorizzati in una tabella specifica) rappresenta la potenza di modellazione di una certa versione del software. All'aumentare del numero (ma anche e soprattutto della parametricità e flessibilità) dei tipi di domanda supportati, cresce la capacità da parte del software di modellare e gestire questionari via via più complessi e con quesiti anche molto particolari. Un esempio è il tipo di domanda relativo al titolo di studio, che provoca l'attivazione di un modulo software in grado di supportare la codifica assistita, utilizzando un motore di ricerca basato su un dizionario e un algoritmo di similarità tra stringhe.

*Le modalità di risposta.* Se il tipo di una domanda prevede una serie di risposte prefissate (in forma di semplice elenco o matriciale, con la possibilità di esprimere una sola scelta o con l'opzione di risposta multipla), tali risposte, unitamente ai corrispondenti testi nelle varie localizzazioni possibili, sono modellate dal concetto di modalità di risposta e memorizzate in una coppia apposita di tabelle (di cui una utilizzata solo nel caso di quesiti a risposta multipla).

#### ***4.2.2 La modellazione dei percorsi di compilazione e il grafo del questionario***

Tra le informazioni memorizzate nella componente statica del questionario, per ogni domanda è presente anche un indice di sequenza che specifica l'ordine secondo il quale i vari quesiti vengono visualizzati all'utente nella pagina web. La sequenza di quesiti presenti in un questionario non rappresenta quasi mai il percorso effettivo di compilazione di un rispondente (lo rappresenta soltanto se tutti i rispondenti devono rispondere a tutti i quesiti presenti, indipendentemente dalle risposte date). Nella pratica la compilazione prevede spesso "salti" con sezioni del questionario che devono essere compilate esclusivamente (o, viceversa, del tutto tralasciate) da alcune tipologie di rispondenti (con determinate caratteristiche e/o che hanno risposto in un certo modo a uno o più dei quesiti proposti).

Si esamina ora la tecnica utilizzata per rappresentare i percorsi di compilazione, ovvero la componente dinamica del questionario. Tale tecnica si basa sull'osservazione che, dal punto di vista della sequenza di compilazione, il questionario può essere rappresentato da un grafo, più precisamente da un Dag (grafo diretto aciclico), in cui i nodi sono costituiti dalle singole domande e gli archi (orientati) permettono di rappresentare i possibili percorsi di compilazione. Più precisamente gli archi sono etichettati in modo che l'etichetta sull'arco tra il nodo-domanda A e il nodo-domanda B specifichi la condizione che deve essere verificata affinché la compilazione (qualora proceda in modo corretto) passi dalla domanda A alla domanda B. Se non è presente alcuna etichetta, la transizione dal nodo-domanda A al nodo-domanda B non è soggetta ad alcuna condizione, è cioè obbligata e fa parte del normale percorso sequenziale di compilazione del questionario.

Un esempio reale, corrispondente al frammento di questionario long form, è riportato in figura 4.1. Il sotto-grafo corrispondente a tale frammento è mostrato in figura 4.2. Per facilitare la lettura del diagramma si è aggiunta una breve descrizione ai blocchetti che rappresentano le domande (si noti che per modellare il quesito 1.4 si sono resi necessari 3 blocchetti-domanda di tipo diverso). Le etichette condizione sono state espresse in forma discorsiva, anche se ovviamente nella realtà è stato utilizzato un formalismo specifico, tale da poter essere direttamente interpretato dal programma che implementa il questionario on line.



rappresentare tutte le condizioni richieste per la corretta modellazione dei flussi di compilazione. D'altra parte, anche nel caso delle tipologie di condizioni, il sistema prevede un approccio di tipo modulare (analogo a quello visto per le tipologie di domande), per cui è possibile estendere le tipologie delle condizioni di transizione supportate, prevedendo un modulo di interpretazione delle stesse più complesso e sofisticato (ad esempio quello realizzato più recentemente per i questionari del Censimento industria e servizi).

Ad ogni modo il formalismo utilizzato ha consentito di realizzare un efficace compromesso tra potere espressivo (era comunque necessario rappresentare tutte le condizioni di diramazione del questionario) e efficienza elaborativa nella valutazione delle condizioni logiche. Questo secondo aspetto è molto importante perché la valutazione delle condizioni sugli archi è alla base del meccanismo di abilitazione e disabilitazione delle domande nell'interfaccia web delle macchine client, per le quali la semplicità e l'elevata efficienza è un requisito irrinunciabile.

Così come tutti i dati che riguardano le domande, anche i dati riguardanti i percorsi di compilazione (che di fatto, visto che i nodi del grafo coincidono con le domande, si riducono agli archi e alle relative etichette-condizioni) sono memorizzate nella base dati in una tabella specifica. In pratica ogni record presente in questa tabella corrisponde biunivocamente ad un arco del grafo che rappresenta i vari percorsi di compilazione del questionario.

Da un punto di vista progettuale vale la pena sottolineare che, analogamente a quanto si fa, ad esempio, quando si effettua la progettazione concettuale di una base dati con un diagramma Entity-Relationship, anche per un questionario elettronico si può iniziare la progettazione costruendo un diagramma concettuale che descrive i percorsi di compilazione e utilizzare il linguaggio di alto livello (anche non formale) che più si ritiene opportuno per esprimere le condizioni sugli archi. A partire da questa descrizione di alto livello si possono già applicare una serie di verifiche più o meno complesse che consentono di effettuare una prima validazione sulla correttezza o meno della concettualizzazione stessa, ad esempio le seguenti:

- ▶ per ogni nodo-domanda deve essere presente almeno un arco in entrata e almeno un arco in uscita (fanno eccezione soltanto i due nodi che rappresentano l'inizio e la fine del questionario);
- ▶ il grafo nel suo complesso deve essere un Dag (grafo diretto aciclico): non devono quindi essere presenti percorsi che generano cicli (in un questionario cartaceo questo corrisponderebbe ad un'istruzione di compilazione che fa tornare il rispondente ad una domanda precedente);
- ▶ se è presente un solo arco uscente da un certo nodo, l'arco deve essere privo di etichetta (la transizione è l'unica possibile e deve quindi essere incondizionata);
- ▶ se sono presenti più archi uscenti dal medesimo nodo, essi devono essere tutti etichettati, ed esprimere condizioni mutuamente esclusive (se una certa configurazione di risposte rendesse vere più condizioni in uscita dal nodo, si sarebbe in presenza di una indeterminatezza di comportamento, perché non si saprebbe quale arco/percorso seguire);
- ▶ in uscita dal medesimo nodo deve essere presente al più un arco con una condizione di tipo "else"/"altrimenti";
- ▶ se non è presente una esplicita condizione di tipo "else" le varie condizioni in uscita dal medesimo nodo devono essere esaustive: qualunque sia la combinazione di risposte fornite dall'utente deve quindi esistere una e una sola condizione tra quelle degli archi uscenti che viene resa vera da tale combinazione di risposte.

Proseguendo l'analogia con la progettazione di una base dati, a partire dalla rappresentazione concettuale di un questionario si possono effettuare dei passi di trasformazione che portano alla sua definizione logico-fisica (il vero e proprio grafo del questionario), dipendente in maniera più o

meno marcata dallo specifico formalismo utilizzato per esprimere le condizioni sugli archi. Se il formalismo prevede condizioni complesse, le trasformazioni sono minime, ma si rende necessario un parser sofisticato e complesso per consentire la valutazione delle condizioni durante il normale funzionamento del questionario elettronico.

Se, come nel nostro caso, il formalismo prevede un formalismo relativamente elementare, si possono rendere necessarie delle trasformazioni più significative, ad esempio l'introduzione di nodi virtuali, che non alterano tuttavia il comportamento del questionario così prodotto in termini di percorsi di compilazione. D'altra parte questo sforzo aggiuntivo è ripagato dal fatto che risulta enormemente semplificata la valutazione delle condizioni logiche. Vale la pena notare inoltre che nella pratica le trasformazioni richieste e le introduzioni di nodi virtuali sono state minime, per cui il modello concettuale del questionario è risultato quasi identico a quello logico-fisico.

#### 4.2.3 Come usare il grafo del questionario

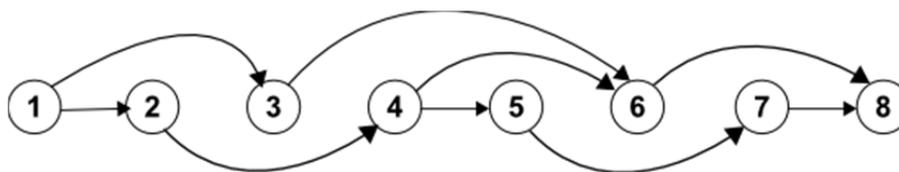
La concettualizzazione dei percorsi di compilazione attraverso la definizione del grafo del questionario è sicuramente interessante di per sé. Tuttavia ciò che giustifica realmente lo sforzo necessario per la sua costruzione e verifica è il fatto che tale modellazione può essere utilizzata sia per attivare/disattivare le varie sezioni e domande del questionario, sia per determinare se le informazioni memorizzate ad un certo momento rappresentano o meno un percorso di compilazione corretto, eventualmente effettuando i "reset" di alcune risposte a seguito di modifiche. Come si è visto in precedenza il grafo del questionario è un Dag (grafo diretto aciclico), che gode di molte importanti proprietà, tra le quali quella di maggior rilevanza nel contesto specifico è la seguente:

**Proprietà (ordinamento topologico)** – Dato un Dag è sempre possibile definire una relazione di ordinamento topologico  $<$  sui nodi del grafo stesso, tale che se  $(u,v)$  è un arco (orientato) del Dag si ha sempre  $u < v$ .

Equivalentemente si può supporre di definire una funzione  $s$  che associ ad ognuno degli  $N$  nodi del grafo un valore intero distinto dell'intervallo  $[1, N]$ . Tale funzione definisce un ordinamento topologico sui nodi del Dag, se dato un generico arco orientato  $(u,v)$ , si ha sempre  $s(u) < s(v)$  (in questo caso il simbolo  $<$  rappresenta il normale operatore di confronto tra interi).

Intuitivamente si può descrivere questa proprietà immaginando di disporre i nodi del grafo lungo una retta secondo l'ordine fissato dalla funzione  $s$  (figura 4.3). In base alla proprietà descritta, una volta disposti i nodi secondo l'ordinamento topologico, gli archi del Dag sono tutti orientati "da sinistra a destra".

Figura 4.3 - Ordinamento topologico dei nodi di un Dag



In effetti, nel caso del grafo di questionario, ottenere il suo ordinamento topologico è molto semplice, perché uno (tra i molti esistenti) corrisponde esattamente all'ordine secondo il quale le domande compaiono, pagina dopo pagina, colonna dopo colonna, dall'alto verso il basso, nel questionario cartaceo. Questo corrisponde peraltro alla considerazione che in un questionario cartaceo si fa sempre in modo che, seguendo il normale flusso di compilazione, non si debba mai "tornare indietro". L'informazione sull'indice topologico può quindi essere ricavata una volta per

tutte in fase di progettazione, memorizzata stabilmente nella base dati e utilizzata estensivamente per le finalità che si vedranno.

Nel seguito si può dunque supporre che ai nodi del grafo del questionario sia associato un indice intero secondo l'ordinamento topologico del Dag corrispondente e che gli archi del Dag siano descritti da una lista di coppie di interi ordinata secondo l'indice del nodo sorgente e del nodo destinazione. Ad esempio la lista (ordinata) corrispondente al grafo di figura 4.3 è quella riportata in figura 4.4.

Vista la corrispondenza stabilita, nel seguito si parla indifferentemente di quesito o domanda del questionario, di nodo ad essa associato, o anche di indice associato al nodo stesso dalla relazione di ordinamento topologico. Si può dire che un nodo D del grafo del questionario è compilato se l'utente ha inserito una risposta per la corrispondente domanda del questionario (e non compilato/non specificato altrimenti).

La compilazione di un certo numero di domande fa sì che alcuni archi (che usano tali nodi nelle proprie etichette) non risultino più percorribili per la compilazione del questionario. In altre parole è possibile dire che la compilazione "taglia" progressivamente una serie di cammini che inizialmente facevano parte del grafo di questionario, rendendo un certo numero di nodi (domande) non più raggiungibili da un percorso di compilazione corretto, e quindi, in definitiva, non più compilabili. La progressiva compilazione del questionario determina quindi una modifica dello stato di ciascun nodo-domanda, che dipende sia dalla struttura del grafo che dai nodi compilati (cioè dalle risposte specifiche già fornite dall'utente).

**Figura 4.4 - La lista ordinata di coppie di indici che rappresenta il grafo di Figura 4.3**

Sorgente	Destinazione
1	2
1	3
2	4
3	6
4	5
4	6
5	7
6	8
7	8

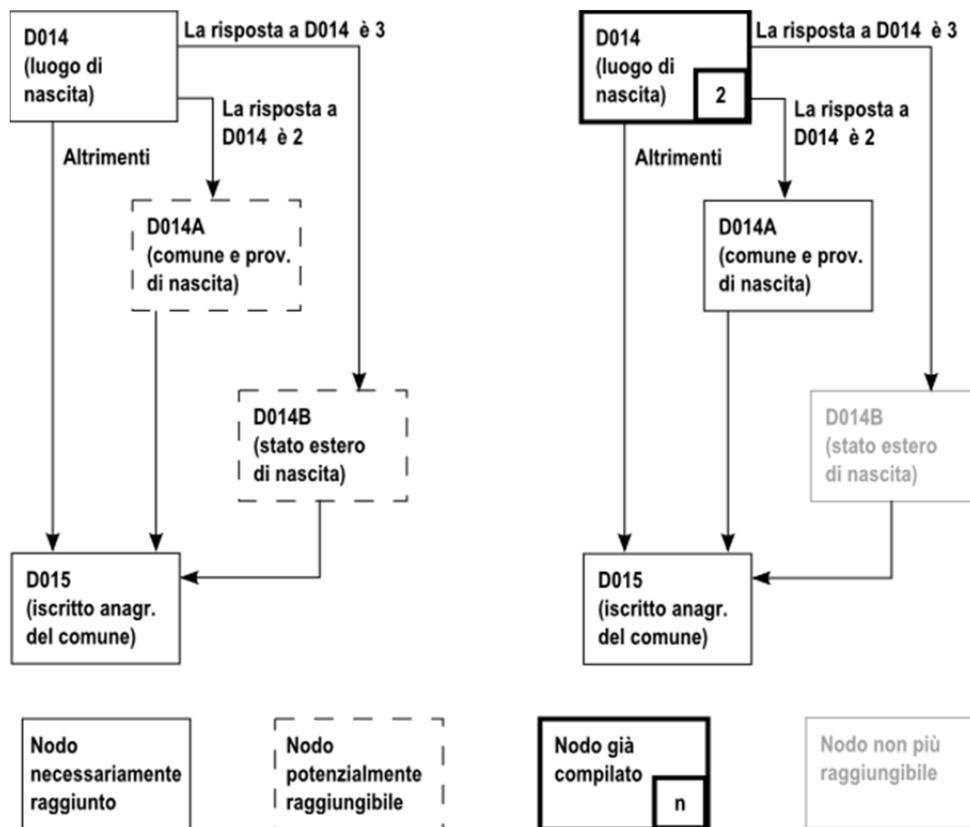
La situazione è illustrata graficamente in figura 4.5, dove si considera nuovamente il sotto-grafo di figura 4.2 (corrispondente al frammento di questionario di figura 4.1). Nella parte sinistra della figura si suppone che la compilazione non sia ancora arrivata al frammento di questionario rappresentato e che in particolare non sia stata ancora fornita una risposta alla domanda-nodo D014 (luogo di nascita). I nodi D014 e D015 sono mostrati con tratto continuo, perché saranno necessariamente raggiunti dai successivi passi della compilazione. Non esistono infatti sequenze di risposte (a cui corrispondono naturalmente percorsi di compilazione) che consentono di escludere D014 e D015 dalla compilazione. Al contrario D014A e D014B sono mostrati tratteggiati perché solo potenzialmente raggiungibili: esistono infatti alcune sequenze di risposte che ne determinano la compilazione e altre che fanno sì che non debbano essere compilato.

Nella parte destra della figura si suppone invece che sia stata fornita una risposta al quesito D014 (il nodo D014 risulta compilato ed è stata fornita la modalità di risposta 2), mentre non sono ancora state fornite risposte per le domande successive. I nodi D014A e D015 sono mostrati con tratto continuo, perché saranno necessariamente raggiunti dai successivi passi della compilazione. A partire dal nodo attualmente compilato D014 non esistono infatti sequenze di risposte (a cui corrispondono percorsi di compilazione) che consentono di escludere D014A e D015 dalla compilazione. Al contrario D014B è mostrato in grigio (similmente a quanto accade nell'interfaccia utente del questionario) perché in seguito alla compilazione della domanda-nodo D014, tale nodo diventa irraggiungibile: sulla base delle risposte fornite e della struttura del questionario l'utente non dovrà rispondere a tale domanda, a meno che ovviamente ritorni sui suoi passi e fornisca qualche risposta diversa alle domande precedenti, in particolare alla D014.

Lo stato di ogni nodo determina quindi se una certa domanda deve essere abilitata alla compilazione e/o lo stile di visualizzazione. D'altra parte, come si è visto, lo stato dipende dalle risposte via via fornite dall'utente e deve quindi essere continuamente aggiornato, in corrispondenza di qualunque modifica sul questionario operata dall'utente.

Ne consegue che tale procedura di aggiornamento dello stato deve essere molto rapida per garantire un ricalcolo praticamente istantaneo e non appesantire l'interfaccia utente con fastidiosi ritardi. Il fatto che un grafo del questionario sia un Dag è fondamentale da questo punto di vista, perché si può dimostrare che per aggiornare lo stato di tutti i nodi è sufficiente una singola scansione di tutti gli archi della lista secondo l'ordine determinato dall'ordinamento topologico. In termini più formali l'aggiornamento si può quindi ottenere in tempo lineare nel numero di archi presenti nel grafo, vale a dire in maniera molto efficiente.

Figura 4.5 - Lo stato dei nodi a seconda delle risposte fornite

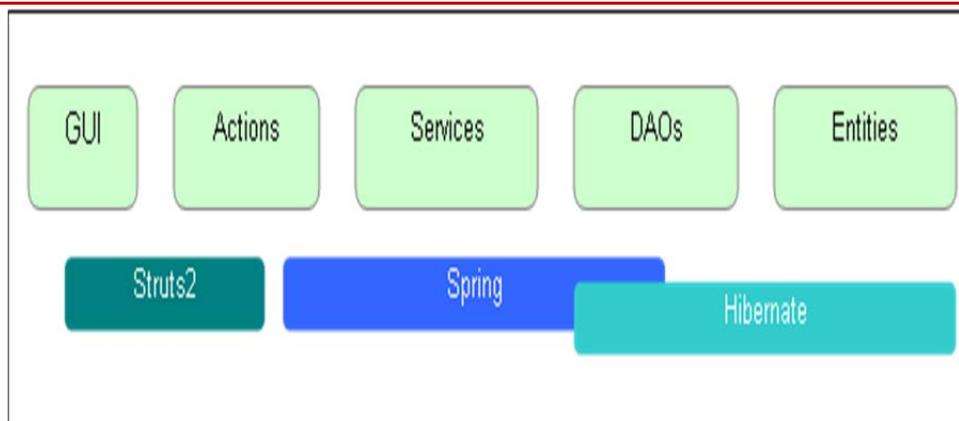


## 4.3 Il Progetto del sistema

### 4.3.1 L'architettura

Qpop è una applicazione web implementata attraverso la piattaforma Java2 Enterprise Edition, organizzata in un'architettura a 5 strati basati sul design pattern Model-View-Controller. L'applicazione sfrutta tre framework open source: Struts2, Spring e Hibernate. La progettazione di applicazioni a strati è illustrato nella figura 4.6.

Figura 4.6 - l'architettura a 5 strati di Qpop



Le funzionalità di ogni strato sono di seguito dettagliate:

- ▶ **Gui:** pagine JSP che implementano l'interfaccia grafica utente. Esse possono essere pagine per l'invio di dati al server, che devono essere processati da una action, o il relativo risultato;
- ▶ **Actions:** classi Java la cui esecuzione viene attivata da una chiamata Http. Esse ricevono i dati dalla richiesta Http ed eseguono alcune operazioni di elaborazione lato server chiamando i services;
- ▶ **Services:** classi Java che gestiscono le transazioni sul database, realizzati attraverso sequenze di chiamate a Dao;
- ▶ **Data Access Objects (Dao):** classi Java che implementano il cosiddetto Crud (Create-Read-Update-Delete), cioè le operazioni di database collegati a uno o più oggetti di dominio;
- ▶ **Entities:** classi Java che rappresentano il dominio, cioè la struttura dati su cui si basa l'applicazione web.

La realizzazione di ciascuno strato è supportata da uno o più framework Java, secondo quanto già descritto nel capitolo 3.

All'avvio dell'applicazione tutti i metadati per i tre questionari (testi delle domande e delle risposte, dizionari riguardanti province e comuni o paesi stranieri, navigazione e regole di validità, ecc...) vengono caricati nella memoria principale. I metadati memorizzati nella cache vengono direttamente utilizzati in sede di visualizzazione delle pagine del questionario, limitando, in tal modo, gli accessi al database al solo salvataggio dei dati inseriti dagli utenti, un principio fondamentale di progettazione per ottenere elevata scalabilità.

In pratica quando il sistema mostra una pagina del questionario, esso innanzitutto carica tutti i metadati relativi alle domande di quella pagina. Tali metadati includono tutti i testi (tenendo conto anche della localizzazione) e la diversa tipologia (check, multi response, testi liberi).

A ogni tipo di domanda è associato un template. Un template specifica come un tipo di domanda deve essere costruita in termini di Html ed è associato ad uno specifico Java bean per implementare il corrispondente trattamento dei dati.

Sono stati definiti oltre 20 tipi domanda. L'utilizzo massiccio dei template insieme con i metadati ha permesso di raggiungere la massima ottimizzazione del codice, ed una risposta più rapida ai cambiamenti nelle specifiche che si sono verificati nel corso del progetto.

I dati inseriti dagli utenti prima di essere salvati devono superare i controlli di qualità. Tali controlli vengono effettuati innanzitutto lato client attraverso dei controlli javascript che consentono una correzione immediata del dato errato. Gli stessi controlli vengono poi ripetuti lato server per una verifica finale prima del salvataggio. Questo passo è stato fondamentale per motivi di sicurezza, un utente malintenzionato avrebbe potuto aggirare i controlli sulla pagina modificando i dati inviati attraverso la richiesta Http, creando inconsistenza dei dati o addirittura a minacce alla sicurezza (ad esempio iniezione di codice Sql malizioso per rilevare dati sensibili dal database).

L'operazione di salvataggio sfrutta ampiamente le peculiarità dei framework utilizzati, con l'ausilio della Reflection di Java. I dati inviati tramite Http vengono salvati automaticamente da Struts2 in un oggetto Java (l'entità). A quel punto, Hibernate è responsabile del mapping fra il bean e la relativa tabella del database, e della scrittura automatica delle query Sql per l'inserimento / aggiornamento dei dati. La gestione delle transazioni nel contempo è gestita dal framework Spring. L'uso massiccio di tali framework ha consentito di produrre codice più pulito, quindi più facile da scrivere, testare e mantenere, consentendo così la realizzazione di una applicazione robusta.

#### 4.3.2 Le tipologie di domande

Nel sistema di acquisizione sono previsti otto tipi di domande.

- ▶ inserimento testo libero (figura 4.7);

Data la domanda, l'utente può scrivere liberamente in un campo di testo la propria risposta, che può essere di tipo numerico oppure testuale.

Figura 4.7 - Inserimento testo libero

**6.3. Specificare l'indirizzo per esteso senza abbreviazioni**  
 (ad esempio PIAZZA GIUSEPPE VERDI 1)  
 VIALE DELL'OCEANO PACIFICO 171

- inserimento testo con auto-completamento (figura 4.8);

Data la domanda, l'utente può scrivere in un campo di testo la propria risposta. Non appena si digitano i primi tre caratteri viene proposta una lista di possibili opzioni da cui selezionare quella opportuna.

Figura 4.8 - Inserimento testo con auto-completamento

**1.4.a. specificare la provincia e il comune**  
 Provincia: BOL Comune:   
 Bologna  
 Bolzano/Bozen

- selezione risposta singola da lista (radio button) (figura 4.9);

Data la domanda, l'utente può selezionare una sola risposta tra quelle proposte. La visualizzazione delle opzioni è ad elenco con evidenziazione delle varie voci mediante un circoletto. In alcuni casi è prevista un'opzione "altro" per permettere all'utente di "personalizzare" la propria risposta rispetto alle opzioni proposte.

Figura 4.9 - Selezione risposta singola da lista

**1.1. Indicare il tipo di alloggio**

- Abitazione
- Altro tipo di alloggio (baracca, roulotte, camper, ecc.)
- Alloggio presso sede diplomatica o consolare
- Struttura residenziale collettiva (hotel, casa di riposo, ecc.)

- selezione risposta singola da elenco (menù a tendina) (figura 4.10);

Data la domanda, l'utente può selezionare una sola risposta tra quelle proposte. La visualizzazione delle opzioni è con menù a tendina.

Figura 4.10 - Selezione risposta singola da elenco

**1.4.b. specificare lo stato estero**  
 Selezionare Stato Estero  
 Algherianistan  
 Ajman (territorio)  
 Aland (territorio)  
 Alaska (territorio)  
 Albania  
 Algeria  
 Altro territorio  
 Altura di Golan (territorio)  
 American Oceania (territorio)  
 Americane del Pacifico, isole (territorio)  
 Andamane e Nicobare (territorio)  
 Andora  
 Angola  
 Anguilla (Regno Unito)  
 Anipuan (territorio)  
 Antigua e Barbuda  
 Antigua (territorio)  
 Antille Olandesi, isole (Paesi Bassi)  
 Apollide

- selezione risposte multiple da lista (check box) (figura 4.11);

Data la domanda, l'utente può selezionare più di una risposta tra quelle proposte. La visualizzazione delle opzioni è ad elenco con evidenziazione delle varie voci mediante un quadratino.

In alcuni casi è prevista anche un'opzione "altro" per permettere all'utente di "personalizzare" la propria risposta rispetto alle opzioni proposte.

Figura 4.11 - Selezione risposte multiple da lista

**2.5. Indicare se l'abitazione dispone di**  
 (sono possibili più risposte)

- 1  Cucina (con caratteristiche di stanza)  
[locale progettato e attrezzato per la preparazione dei pasti che ha le caratteristiche di stanza]
- 2  Cucinino  
[locale progettato e attrezzato per la preparazione dei pasti che non ha le caratteristiche di stanza]
- 3  Angolo cottura in stanza destinata a più attività  
[spazio progettato e attrezzato per la preparazione dei pasti inserito in un locale con caratteristiche di stanza con varie funzioni (soggiorno, salotto, ecc.)]
- 4  Non dispone di cucina, cucinino o angolo cottura

► inserimento dati in lista (figura 4.12);

Data la domanda, la risposta prevede l'inserimento di una riga di dati all'interno di una lista di dimensione non fissa. È prevista la correzione dei dati inseriti a livello sia di singolo campo sia di intera riga (eliminazione della riga).

Figura 4.12 - Inserimento dati in lista

COGNOME	NOME	SESSO		DATA DI NASCITA			LUOGO DI NASCITA	CITTADINANZA		
		M	F	Giorno	Mese	Anno		Italiana	Straniera o apolide	
VERDI	FILIPPO	<input checked="" type="radio"/>	<input type="radio"/>	24	12	1955	BRESSANONE	<input checked="" type="radio"/>	<input type="radio"/>	Intestatario
BIANCHI	RACHELE	<input type="radio"/>	<input checked="" type="radio"/>	23	12	1953	ANDRIA	<input checked="" type="radio"/>	<input type="radio"/>	<input type="button" value="Rimuovi"/>
		<input type="radio"/>	<input type="radio"/>					<input type="radio"/>	<input type="radio"/>	

► selezione risposte multiple da tabella di check box (figura 4.13);

Data la domanda, l'utente può selezionare più di una risposta tra quelle proposte all'interno di una struttura tabellare in cui le opzioni si trovano agli incroci di riga e colonna.

Figura 4.13 - Selezione risposte multiple da tabella di check box

**4.2. Indicare quale impianto (o impianti) di riscaldamento ha l'abitazione e - per ciascun impianto - il combustibile o l'energia che lo alimenta**  
 (sono possibili più risposte)

	Metano, gas naturale	Gasolio	GPL (Gas Petrolio Liquefatto)	Combustibile solido (legna, carbone, ecc.)	Olio combustibile	Energia elettrica	Altro combustibile o energia
Impianto centralizzato ad uso di più abitazioni	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Impianto autonomo ad uso esclusivo dell'abitazione	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Apparecchi singoli fissi (camino, stufa, radiatore, pompe di calore, ecc.) che riscaldano l'intera abitazione, o la maggior parte di essa	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Apparecchi singoli fissi (camino, stufa, radiatore, pompe di calore, ecc.) che riscaldano alcune parti dell'abitazione	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

► testo descrittivo con selezione da lista di ricerca (figura 4.14);

Data la domanda, l'utente può inserire in un campo di testo una descrizione della risposta che intende fornire. A questo punto viene mostrata una lista di possibili risposte da cui scegliere: la lista

è ottenuta attraverso una ricerca per similitudine con la descrizione fornita. È questo ad esempio il caso della domanda relativa al titolo di studio conseguito in cui è richiesta la corretta dicitura.

Figura 4.14 - Testo descrittivo con selezione da lista di ricerca

Selezionare la descrizione esatta del titolo di studio dall'elenco sottostante:

- Laurea magistrale Ingegneria elettronica
- Laurea magistrale Ingegneria elettronica e fotonica
- Laurea magistrale Ingegneria elettronica e delle telecomunicazioni
- Laurea magistrale Ingegneria elettronica per l'automazione
- Laurea Ingegneria elettronica (vecchio ordinamento)
- Laurea magistrale Ingegneria elettronica per l'automazione e le telecomunicazioni
- Ingegneria elettronica (Laurea specialistica nuovo ordinamento)
- Laurea magistrale Ingegneria elettronica per l'industria e l'innovazione
- Laurea magistrale Ingegneria e tecnologie per le telecomunicazioni e l'elettronica
- Laurea magistrale Ingegneria elettronica e telecomunicazioni per lo sviluppo sostenibile
- Ingegneria elettronica per l'automazione (Laurea specialistica nuovo ordinamento)
- Ingegneria elettronica e fotonica (Laurea specialistica nuovo ordinamento)
- Ingegneria elettronica e delle telecomunicazioni (Laurea specialistica nuovo ordinamento)
- Laurea magistrale Ingegneria elettrica

chiudi

**5.5 . Con riferimento alla risposta fornita alla domanda 5.3 specificare per esteso il titolo di studio conseguito**

Se al quesito 5.3 è stata indicata una delle modalità 5, 6, 7, 8, 9, 10, 11, 12, 14 o 16, selezionare la descrizione tramite il pulsante "cerca". Se al quesito 5.3 è stata indicata una delle modalità 13, 15 o 17, scrivere una descrizione estesa del titolo conseguito (ad esempio, diploma di qualifica professionale agraria, diploma di maturità tecnica industriale, diploma di accademia di belle arti, laurea in matematica vecchio ordinamento, laurea magistrale in giurisprudenza, ecc.) e selezionare la descrizione esatta tramite il pulsante "cerca".

Scrivere una descrizione estesa del titolo di studio conseguito:

Titolo di studio selezionato:

### 4.3.3 I percorsi di compilazione

Al fine di semplificare la compilazione del questionario da parte degli utenti, nella realizzazione del sistema di acquisizione si sono perseguiti due obiettivi principali:

- ▶ indirizzare i rispondenti verso domande pertinenti: ad esempio, sottoporre gruppi differenti di domande a chi si dichiara occupato e a chi si dichiara non occupato;
- ▶ evitare di scendere in domande dettagliate se non necessario: è inutile, ad esempio, richiedere informazioni sul titolo di studio se il rispondente ha dichiarato di non possedere alcun titolo di studio.

A tale scopo sono state individuate varie domande filtro che permettono di saltare uno o più quesiti successivi se sono verificate alcune condizioni, definendo così dei percorsi di compilazione.

Il sistema gestisce i filtri automaticamente e la creazione dei percorsi di compilazione è dinamica: ogni risposta data o ogni correzione di una risposta data in precedenza contribuisce alla definizione immediata di un ben determinato percorso di compilazione che il sistema attua attraverso l'attivazione e/o la disattivazione automatica di gruppi di domande.

In questo modo, ad esempio, per una persona con meno di 15 anni automaticamente si disabilita la compilazione dei quesiti sulla condizione professionale o non professionale.

### 4.3.4 I controlli

Il sistema di acquisizione prevede un insieme di controlli eseguiti direttamente in fase di compilazione. In questo modo si è garantiti della correttezza dei dati già alla fonte.

I controlli riguardano innanzitutto l'inserimento dei dati ritenuti essenziali.

Ci sono domande alle quali è obbligatorio rispondere perché sono ritenute irrinunciabili per la qualità della rilevazione o rilevanti perché definiscono i percorsi di compilazione. In caso di

mancata risposta il sistema impedisce di completare la compilazione: si tratta di domande definite per questo anche come “bloccanti”.

Ad esempio è bloccante la domanda riguardante lo stato civile, perché dalla risposta dipende l’attivazione del percorso relativo alle domande specifiche per chi ha contratto matrimonio.

Altri controlli riguardano invece la correttezza dei dati inseriti nei campi a testo libero. Se ad esempio un campo prevede l’inserimento di un valore numerico o di una data, il sistema verifica che quanto inserito corrisponda al tipo previsto e in caso contrario visualizza un opportuno messaggio di errore che impedisce il completamento della compilazione.

Accanto alle domande “bloccanti” ci sono domande definite “non bloccanti”, quelle che sono ritenute non di importanza primaria (ad esempio la domanda relativa allo stato civile prima dell’ultimo matrimonio). Per questo tipo di domande c’è completa libertà di scelta di risposta, fermo restando il controllo di correttezza in caso di risposta.

Ci sono poi controlli di coerenza in corrispondenza di alcune informazioni inserite in sezioni diverse del questionario. Ad esempio c’è un controllo di coerenza tra le informazioni su sesso, data di nascita e cittadinanza fornite nella lista delle persone abitualmente dimoranti nell’alloggio e le stesse informazioni riportate nei singoli fogli individuali. Anche in questo caso un opportuno messaggio di errore impedisce il completamento della compilazione.

Accanto ai controlli di tipo bloccante di cui si è parlato, il sistema di acquisizione prevede anche una serie di controlli “non bloccanti” definiti “accertamenti”. Questi non segnalano errori ma producono dei messaggi di “attenzione” (warning) perché viene rilevata una possibile incongruenza di dati e si chiede una conferma da parte del rispondente.

Un esempio è dato dal caso in cui una persona abbia dichiarato di avere meno di 15 anni e che dichiara anche di essere coniugata. Sebbene non comune, potrebbe essere un caso possibile per cui il sistema chiede conferma di quanto dichiarato attraverso un opportuno messaggio. A questo punto il rispondente o corregge l’incongruenza, se sussiste, oppure conferma la correttezza della risposta e prosegue con la compilazione.



## CAPITOLO 5

### IL SISTEMA DI GESTIONE DELLA RILEVAZIONE<sup>17</sup>

#### 5.1 Caratteristiche generali del sistema

Il Sistema di Gestione della Rilevazione, Sgr, è stato concepito come un unico ambiente con molteplici funzioni, organizzate in menù e sottomenù e utilizzato da varie tipologie di utenze, con differente visibilità territoriale e operatività funzionale. Il sistema si presentava in modo personalizzato all'utente che effettuava il login: ogni operatore visualizzava infatti solo le funzioni di proprio interesse e le informazioni che era autorizzato a trattare in base all'ambito territoriale di propria competenza e al proprio ruolo all'interno dell'indagine.

Ad ogni utente Sgr inoltre metteva a disposizione da una parte informazioni aggiornate a diversi livelli di aggregazione e dall'altra uno strumento per il lavoro collaborativo, basato su un workflow distribuito nel quale ogni operatore ha lavorato autonomamente seguendo un percorso ben delineato e standardizzato. Sgr ha contribuito alla riduzione dei possibili errori statistici attraverso la verifica puntuale dei potenziali errori delle liste pre-censuarie, un monitoraggio "in itinere" della rilevazione e il supporto al lavoro sul campo dei rilevatori.

La fase preliminare all'avvio della rilevazione ha comportato il popolamento del sistema con informazioni relative alle famiglie da censire e successivamente la distribuzione del lavoro sul territorio attraverso la definizione della rete di rilevazione e l'assegnazione agli operatori dei questionari da trattare. Il primo grande cambiamento rispetto al passato è legato all'utilizzo delle Liste Anagrafiche Comunali (Lac): tutti i questionari, prima di essere spediti sono stati prestampati con le informazioni provenienti da tali archivi e consegnati agli intestatari dei fogli di famiglia tramite vettore postale. La scelta del recupero mirato della sotto-copertura o della individuazione della sovra-copertura, inoltre, ha fatto emergere la necessità di caricare all'interno del Sistema tutti i segnali di sotto-copertura/sovra-copertura disponibili, per indirizzare in modo puntuale i rilevatori. Pertanto sono stati caricati nel sistema sia i dati delle liste anagrafiche comunali che dati provenienti da fonti amministrative e dalla rilevazione dei numeri civici, i cosiddetti segnali che dovevano essere controllati e chiusi dal rilevatore. Relativamente alla distribuzione del lavoro sul territorio, il sistema, per gestire la capillarità territoriale italiana, caratterizzata da realtà geografiche, socio-demografiche e abitative molto differenti, ha previsto varie tipologie di utenze, con differente visibilità territoriale (nazionale, regionale, provinciale, comunale) e operatività funzionale. Sgr ha consentito ai responsabili territoriali di creare e gestire, in completa autonomia, la propria rete di rilevazione secondo queste caratteristiche.

Sgr ha guidato e supportato l'intera rilevazione. È stato previsto, infatti, uno specifico flusso degli "stati" del questionario, per cui ogni stato rappresentava una fase di lavorazione del questionario: era possibile passare da uno stato ad un altro attraverso funzioni di Sgr o azioni esterne al sistema (ad esempio il completamento della compilazione del questionario web). Ogni funzione aveva degli stati che ne erano preconditione, ovvero per l'utilizzo della funzione su un dato questionario esso doveva trovarsi in un determinato stato, e uno stato che ne era post-condizione, ovvero lo stato in

---

<sup>17</sup> Il capitolo è a cura di Maura Giacommo. I paragrafo 5.1 e il sottoparagrafo 5.6.1 sono stati redatti da Giulia Vaste; il paragrafo 5.2 è stato redatto da Eleonora Sibilio; il paragrafo 5.3 è stato redatto da Lucia Toti; il paragrafo 5.4 è stato redatto da Enrico Orsini; il paragrafo 5.5 è stato redatto da Maura Giacommo; il sottoparagrafo 5.5.1 è stato redatto da Marco Amato; il sottoparagrafo 5.5.2 è stato redatto da Annalisa Cesaro; il sottoparagrafo 5.5.3 è stato redatto da Roberta Radini; il sottoparagrafo 5.6.2 è stato redatto da Marco Polizzi; il sottoparagrafo 5.6.3 è stato redatto da Claudio Giovannini; il sottoparagrafo 5.7.1 è stato redatto da Rossana Vignola; il sottoparagrafo 5.7.2 è stato redatto da Cristiano Francone; il sottoparagrafo 5.7.3 è stato redatto da Paolo Astorri e Marco Ciccacci.

cui la funzione poneva il questionario. Di fatto quindi Sgr era un sistema di workflow distribuito nel quale ogni operatore lavorava autonomamente rispettando tuttavia, in modo a lui trasparente, determinate regole condivise. Questa modalità operativa ha prodotto benefici in termini di tempestività, costi e capacità di controllo centralizzato della rete censuaria. Nel contempo ha reso gli operatori delle rete indipendenti nella gestione delle operazioni loro affidate.

Il cuore del sistema era rappresentato dal “Diario di sezione”, un pannello di controllo che recepiva in un unico ambiente tutti i segnali delle varie fonti di dati e che conteneva per ogni singolo questionario le informazioni riguardanti l’anagrafica delle famiglie, lo stato di lavorazione del questionario e le date nelle quali venivano effettuate le operazioni che comportavano un cambiamento nello stato.

Il “Diario di sezione” è risultato uno strumento completo: sia per il rilevatore per avere un supporto al lavoro sul campo, continuamente aggiornato anche in base alle compilazioni web portate a termine dai rispondenti e alle restituzioni dei cartacei, sia per gli operatori di più alto livello, come ad esempio i coordinatori comunali, per monitorare l’andamento della rilevazione in ogni sezione di censimento.

Tramite Sgr era inoltre possibile inserire nel sistema i dati relativi alle variabili primarie (maschi, femmine, stranieri, tipologia di abitazione, ecc.); ciò ha permesso la pubblicazione dei primi risultati provvisori in tempi brevi, senza dover aspettare la registrazione e l’acquisizione in formato elettronico di tutti i questionari cartacei.

La tracciabilità di qualsiasi operazione effettuata sul sistema ha permesso all’Istituto Nazionale di Statistica e agli uffici territoriali di censimento (dell’Istat e provinciali) di avere costantemente un quadro completo relativo all’andamento delle operazioni e allo stato di avanzamento delle varie attività presso ciascun ufficio comunale di censimento. Ciò ha consentito anche di verificare il lavoro di ogni operatore e di ogni comune, permettendo azioni puntuali e tempestive ove necessario. D’altra parte il sistema ha reso disponibili dei report per avere informazioni anche di minor dettaglio, dati aggregati a livello territoriale o per operatore che hanno permesso all’Istituto Nazionale di Statistica di supervisionare il rapporto col vettore postale, l’operatività dei singoli comuni, province, Uffici Regionali di Censimento, avendo così uno strumento che forniva informazioni in tempo reale, anche di supporto a decisioni strategiche.

Sgr è stato arricchito con funzionalità non strettamente legate al censimento della popolazione: un menù è stato progettato e realizzato esclusivamente per la gestione di una ulteriore indagine, la rilevazione degli edifici, svolta parallelamente al censimento, e per la gestione della rilevazione sulle minoranze linguistiche per le sole province di Trento e Bolzano; altre funzioni hanno supportato fasi successive al censimento, necessarie all’aggiornamento delle anagrafi. Infine Sgr è stato reso disponibile sia in italiano che in tedesco.

Sgr è organizzato in aree all’interno delle quali sono presenti specifiche funzioni. Nel complesso sono state definite 9 aree e sviluppate 78 funzioni.

Le aree definite sono:

- ▶ operatori, per la definizione e la gestione della rete di rilevazione e l’assegnazione delle sezioni di censimento ai rilevatori;
- ▶ variazioni anagrafiche, per l’inserimento, per i soli comuni di limitata dimensione demografica, delle variazioni anagrafiche intercorse tra il 1 gennaio 2011, data di riferimento della prima Lac, e la mezzanotte tra l’8 e il 9 ottobre 2011, data di riferimento del Censimento;
- ▶ rapporti, per la produzione dinamica, a intervalli regolari e predefiniti di tempo, dei rapporti di monitoraggio sull’andamento della rilevazione sia per operatore che per territorio;
- ▶ sezione, contenente il diario di sezione del rilevatore ed alcune funzioni per la gestione di eventuali modifiche territoriali a livello di sezione;

- ▶ edifici, per l'inserimento di alcune informazioni chiave relative alla rilevazione sugli edifici, parallela a quella censuaria sulla popolazione;
- ▶ confronto censimento anagrafe, per la gestione delle verifiche individuali tra le persone censite in un comune e quelle residenti registrate nell'Anagrafe del comune stesso e la produzione delle 4 liste individuali: L1 censiti residenti; L2 censiti non residenti; L3 residenti non censiti; L4 censiti residenti nel comune ma ad altro indirizzo rispetto a quello della Lac;
- ▶ questionari, per la registrazione degli arrivi cartacei, il data-entry dei modelli cartacei per i soli comuni abilitati, mediante l'accesso automatico al medesimo questionario elettronico messo a disposizione delle famiglie, e la registrazione dei modelli riepilogativi contenenti alcune variabili anagrafiche;
- ▶ utilità, contenente una serie di funzioni di utilità generale, trasversali alle varie aree, come ad esempio le funzioni di ricerca famiglia o di stampa questionari o la funzione di chiusura del Censimento;
- ▶ ritiro scatole, per la gestione delle confezioni e spedizioni dei questionari cartacei alla ditta di lettura ottica/registrazione.

## 5.2. Gestione della rete e del territorio

### 5.2.1 Le tipologie utente

L'area Operatori del Sistema di Gestione della Rilevazione comprende le funzioni per la definizione e gestione della rete di rilevazione sul territorio e le funzioni per effettuare le assegnazioni del lavoro che gli operatori dovevano svolgere in base alla loro abilitazione sul sistema.

Sono state definite 10 tipologie di operatore:

- 1) **Ucc**: responsabile dell'Ufficio comunale di censimento. Era obbligatorio e unico e aveva visibilità sull'intero comune. Quando erano presenti più comuni in forma associata e un unico Ucc, questi riceveva tante user-id e password per ciascun comune di propria competenza;
- 2) **Ducc**: Delegato dell'ufficio comunale di censimento. Poteva essere definito solamente per i Comuni con più di 150 mila abitanti e per quelli che appartenevano agli Ucc in forma associata. Relativamente alle funzioni di visualizzazione e aggiornamento del diario, aveva visibilità sulla porzione di territorio coperta dai coordinatori a lui assegnati (o dai rilevatori a lui assegnati qualora nel comune non ci fossero coordinatori). In questo secondo caso, era comunque necessario creare all'interno di Sgr un Coc fittizio (poteva essere lo stesso Ducc), in quanto non si potevano assegnare direttamente Ril ai Ducc;
- 3) **Coc**: COordinatore Comunale. Nell'ambito delle funzioni di visualizzazione e aggiornamento diario aveva visibilità sulla porzione di territorio coperta dai rilevatori a lui assegnati;
- 4) **Ril**: RILevatore. Contestualmente alle funzioni di visualizzazione e aggiornamento diario, aveva visibilità su una o più sezioni o sottosezioni a lui assegnate;
- 5) **Upc**: Ufficio Provinciale di Censimento. Aveva visibilità sulla provincia e su tutti i comuni di competenza della propria provincia;
- 6) **Pre**: PREfettura. Aveva visibilità per la propria prefettura per i soli rapporti riassuntivi;
- 7) **Rit**: Responsabile Istat Territoriale. Aveva visibilità sulla regione di competenza e su tutti i comuni di questa;
- 8) **Urc**: Ufficio Regionale di Censimento. Aveva visibilità sulla regione di competenza e su tutti i comuni di questa;
- 9) **Istat**: Istat nazionale. Aveva visibilità sull'intero territorio nazionale;
- 10) **Verde**: Operatore Numero Verde. Aveva visibilità sull'intero territorio nazionale.

Alcune tipologie di utente sono state ulteriormente classificate. In particolare agli operatori della tipologia Urc, Coc e Ril è stato assegnato un profilo applicativo specifico associato ad un insieme di funzioni di Sgr che potevano essere utilizzate secondo tale profilo. Nella tabella 5.1 si riportano tutti i profili abbinati alle tipologie operatore:

**Tabella 5.1 - Associazione tipologia profilo degli operatori**

<b>Tipologia operatore</b>	<b>Profilo applicativo</b>	<b>Descrizione profilo applicativo</b>
Ucc	Ucc	Responsabile Ufficio Comunale di Censimento
Ducc	Ducc	Delegato Ufficio Comunale di Censimento
Urc	Rurc	Responsabile Ufficio Regionale di Censimento
	Vurc	Vice responsabile dell'Ufficio Regionale di Censimento
	Curc	Componente di back office dell'Ufficio Regionale di Censimento
Rit	Rit	Responsabile Istat Territoriale
Upc	Prefettura	Responsabile di prefettura con visibilità provinciale
Pre	Prefettura	Responsabile di prefettura che accede ai soli rapporti riassuntivi
Istat	Istat centrale	Utente Istat
Coc	Coc	Coordinatore Comunale
	Cor	Coordinatore con funzioni anche di configurazione della rete
Ril	Lavoro sul campo	Accedeva alle funzioni per gestire il lavoro sul campo e alle funzioni di lista edifici
	Media operatività	Accedeva anche alle funzioni di gestione dei modelli riepilogativi, registrazione arrivo cartaceo e gestione lista edifici
	Alta operatività	Accedeva anche alle funzioni di inserimento della lista A, registrazione arrivo cartaceo, gestione lista edifici, rapporti riassuntivi e tutte le funzioni del CoR
	Ccr	Figura di back office comunale per le funzioni di gestione lista edifici
	Ccr alta	Figura di back office comunale per le funzioni di gestione lista edifici con inserimento della lista A
	Edifici	Accedeva alle sole funzioni di gestione lista edifici

### 5.2.2 L'organizzazione del lavoro

Attraverso le funzioni interne all'area Operatori (figura 5.1), per ogni operatore è stato possibile creare in modo dinamico ed indipendente la propria rete di rilevazione assegnando, gestendo o modificando le figure operative facenti parte della rete territoriale.

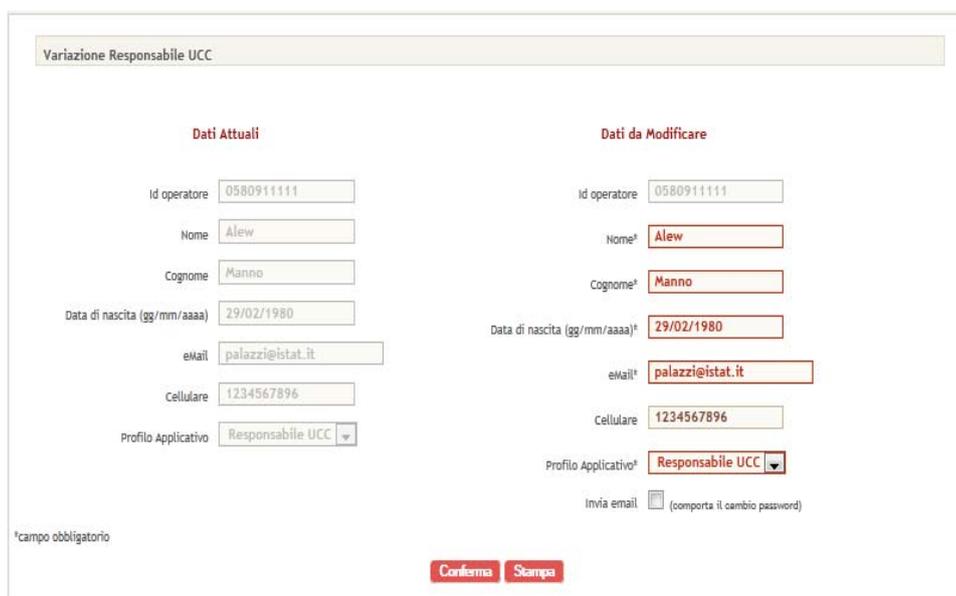
Figura 5.1 - Menu dell'area Operatori



Di seguito si fornisce una descrizione delle funzioni sviluppate:

- 1) **variazione del responsabile Ucc:** consente di modificare il responsabile Ucc inserendone uno nuovo in caso di cessazione di quello attuale (figura 5.2);

Figura 5.2 - Maschera per la Variazione del responsabile Ucc



- 2) **gestione delegati Ucc:** attraverso questa funzione, è possibile inserire delegati Ucc compilando i campi della maschera riportata nella figura 5.3. È anche possibile cancellare delegati Ucc o modificare le variabili associate a ciascuno di essi;

Figura 5.3 - Maschera per inserimento Delegati Ucc

Inserimento Delegati UCC

Nome\*

Cognome\*

Data di nascita (gg/mm/aaaa)\*

eMail\*

Cellulare

Profilo Applicativo\*

\*campo obbligatorio

- 3) **gestione Coc:** attraverso questa funzione, è possibile inserire coordinatori Coc compilando i campi presenti nella maschera riportata nella figura 5.4. Inoltre gli operatori abilitati possono eseguire la cessazione di un coordinatore Coc dalla rete oppure modificarne i dati anagrafici precedentemente inseriti;

Figura 5.4 - Inserimento Coc

Inserimento COC

Nome\*

Cognome\*

Data di nascita (gg/mm/aaaa)\*

eMail\*

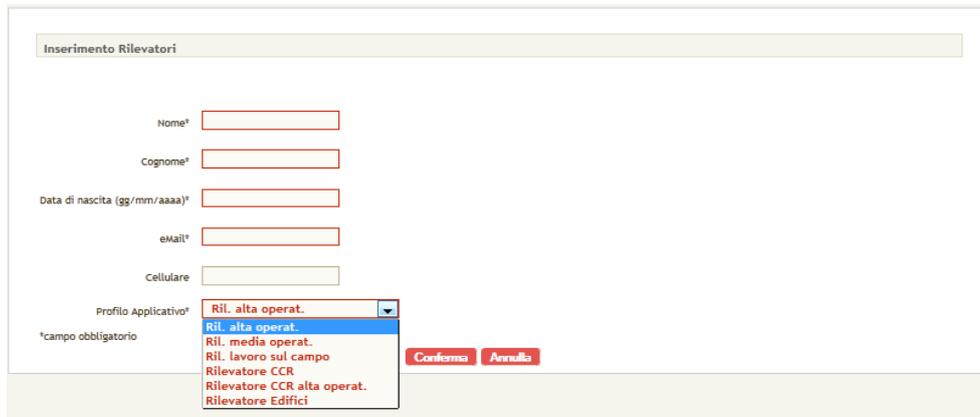
Cellulare

Profilo Applicativo\*

\*campo obbligatorio

- 4) **gestione rilevatori:** consente di inserire un rilevatore (figura 5.5), scegliendo tra i vari profili, ed effettuare la modifica/cessazione in un qualsiasi momento, procedendo come di consueto. In generale, quando si effettua una cancellazione, vengono contestualmente eliminati anche tutti i legami tra quell'operatore ed eventuali altri operatori e/o sezioni assegnate;

Figura 5.5 - Inserimento rilevatori



Inserimento Rilevatori

Nome\*

Cognome\*

Data di nascita (gg/mm/aaaa)\*

eMail\*

Cellulare

Profilo Applicativo\* 

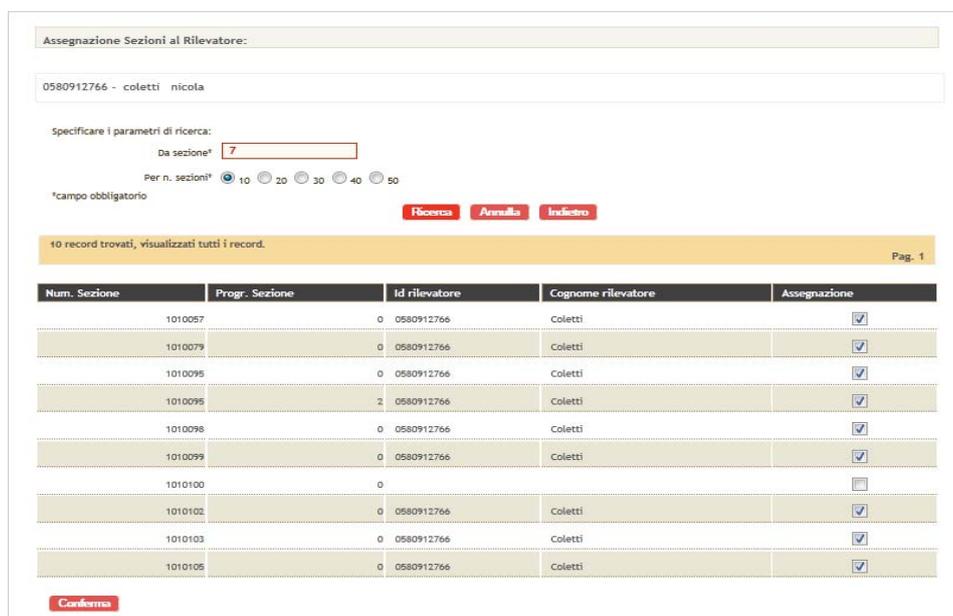
- Ril. alta operat.
- Ril. media operat.
- Ril. lavoro sul campo
- Rilevatore CCR
- Rilevatore CCR alta operat.
- Rilevatore Edifici

\*campo obbligatorio

5) **assegnazioni:** le operazioni di assegnazione, indispensabili per permettere un corretto andamento della rilevazione e un attento monitoraggio, consentono di assegnare o de-assegnare le sezioni ai rilevatori, i Coc ai Ducc e i rilevatori ai Coc. Tali operazioni devono essere necessariamente effettuate prima dell'inizio della rilevazione sul campo e dopo aver inserito i dati anagrafici degli utenti. Si suddividono in:

- **assegnazioni Sezioni ai Rilevatori** (figura 5.6): ha lo scopo di assegnare o de-assegnare a ciascun rilevatore una o più sezioni del comune di competenza. Le sezioni possono essere assegnate esclusivamente ai profili applicativi di Rilevatore alta operatività, Rilevatore media operatività e Rilevatore lavoro sul campo. Qualora fosse stato necessario assegnare una sezione a più rilevatori, si sarebbero dovute definire prima le sottosezioni, attraverso opportune funzioni, all'interno della sezione che si desiderava suddividere;

Figura 5.6 - Maschera per l'assegnazione delle sezioni al rilevatore



Assegnazione Sezioni al Rilevatore:

0580912766 - coletti nicola

Specificare i parametri di ricerca:

Da sezione\*

Per n. sezioni\*  10  20  30  40  50

\*campo obbligatorio

10 record trovati, visualizzati tutti i record. Pag. 1

Num. Sezione	Progr. Sezione	Id rilevatore	Cognome rilevatore	Assegnazione
1010057	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010079	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010095	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010095	2	0580912766	Coletti	<input checked="" type="checkbox"/>
1010098	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010099	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010100	0			<input type="checkbox"/>
1010102	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010103	0	0580912766	Coletti	<input checked="" type="checkbox"/>
1010105	0	0580912766	Coletti	<input checked="" type="checkbox"/>

- **Assegnazioni Coc ai Ducc:** la funzione ha lo scopo di assegnare al Delegato Ucc i Coordinatori comunali all'interno del comune di competenza. In caso di coordinatori già assegnati, questi si possono deselezionare e rendere effettiva tale operazione. In modo analogo, si può procedere nel caso di coordinatori da assegnare;
  - **assegnazioni rilevatori ai Coc:** tramite questa funzione si possono assegnare al Coordinatore comunale i Rilevatori per il comune di competenza. Le operazioni da effettuare sono simili a quelle illustrate nelle funzioni precedenti;
- 6) **composizione rete di rilevazione:** permette di consultare la composizione della rete comunale di rilevazione ovvero tutti gli operatori definiti in un determinato comune, il loro profilo utente e i loro dati anagrafici;
- 7) **cerca operatore:** rende possibile una ricerca degli operatori della rete di rilevazione del Comune di competenza e la visualizzazione dei dettagli di un singolo operatore.

### 5.3 Gestione delle operazioni sul campo

Il cuore di Sgr è rappresentato dal diario di sezione e dalle funzioni di alimentazione e di aggiornamento di questo. Il diario di sezione, funzione centrale del sistema di monitoraggio, ha permesso infatti agli operatori la gestione delle attività sul campo. Inizialmente popolato utilizzando i dati provenienti dalle Lac ed aggiornato sulla base dei dati pervenuti tramite il sistema di acquisizione, ha consentito di tenere una traccia puntuale dei cambiamenti di stato di ogni questionario di rilevazione. Inoltre, all'interno del diario di sezione sono state riportate anche una serie di informazioni aggiuntive, sotto forma di righe del diario, provenienti da altre fonti amministrative (Lifa e Rnc) relative a tutti gli individui che presentavano segnali di presenza sul territorio coperto dalla sezione rappresentata nel diario e assegnata ad un determinato rilevatore. Tutte le righe del diario dovevano essere verificate dal rilevatore e risolte. Alla base del diario di sezione è il concetto di "stato" del questionario o più in generale di "stato" della riga del diario; nell'ambito di un flusso di lavorazione sequenziale di ogni singolo modello, lo stato rappresenta il punto in cui esso si trova e ne definisce anche le caratteristiche. Inoltre sono state implementate nel sistema le regole che consentivano la transizione da uno stato all'altro o il ripristino, in caso di errore, di uno stato precedente.

Il passaggio di stato poteva essere fatto o in modo automatico oppure dall'operatore stesso, sempre tramite il diario di sezione. Nel primo caso le informazioni potevano provenire da altre funzioni del sistema, da informazioni pervenute da attori esterni, quali il vettore postale, oppure dal sistema di acquisizione. Nel secondo caso la scelta era gestita manualmente e vincolata ad un numero limitato di stati congruenti con quello della riga. Lo stato del diario inoltre poteva essere finale, nel caso in cui l'operatore aveva concluso la sua lavorazione della famiglia, o intermedio.

Sono stati modellati nel sistema quattro diversi workflow, corrispondenti ad oltre 40 stati, suddivisi in base alla fase di lavorazione e alla provenienza del dato precaricato:

- ▶ per famiglie precaricate;
- ▶ per dati precaricati con situazioni particolari (senz'altro, convivenze e famiglie molto numerose);
- ▶ per dati provenienti da archivi amministrativi;
- ▶ per dati provenienti dalla rilevazione dei numeri civici.

I quattro workflow, la gestione delle eccezioni e del ritorno allo stato precedente sono state realizzate tramite un sistema di metadati mappato all'interno della base dati.

Le diverse caratteristiche dello stato, anche esse mappate all'interno della base dati, governavano in modo automatico le funzioni dell'applicazione rendendo trasparente allo sviluppatore la logica di processo.

Il flusso principale è quello costituito dalle famiglie precaricate di seguito quindi verrà mostrato tale flusso suddiviso per fase di lavorazione.

Fase di consegna: tutti i questionari, giudicati eleggibili alla spedizione, sono stati passati al vettore postale. Sono escluse da questo invio le convivenze, le famiglie numerose ed i senza tetto. Questi ultimi hanno avuto la consegna fatta direttamente dal rilevatore ed un loro specifico flusso. Il vettore postale ha inoltre inviato dei file di aggiornamento circa l'avvenuta consegna dei questionari due volte al giorno, indicando i soli casi di fallita consegna. Questi questionari sono stati poi riportati agli uffici comunali e il rilevatore ha tentato di effettuare una seconda consegna. Anche questi passaggi sono stati modellati nel flusso di lavorazione (figura 5.7).

I questionari restituiti secondo i diversi canali previsti ( al centro di ritiro postale, al centro comunale di raccolta, via web) sono stati registrati una volta giunti al centro comunale, dando il primo segnale di restituzione, necessario sia per il monitoraggio, sia per evitare al rilevatore di recarsi dalla famiglia (figura 5.8).

Tutti i questionari non arrivati al comune dopo una certa data, sono stati lavorati inviando il rilevatore sul campo con l'obiettivo di ottenere la restituzione del questionario. Tutti i questionari sono stati revisionati dopo l'arrivo al comune e ne sono state registrate le variabili primarie tramite i modelli riepilogativi. I questionari sono poi stati passati alla fase successiva del confronto con le anagrafi (figura 5.9).

Figura 5.7 - Workflow della fase di consegna del questionario

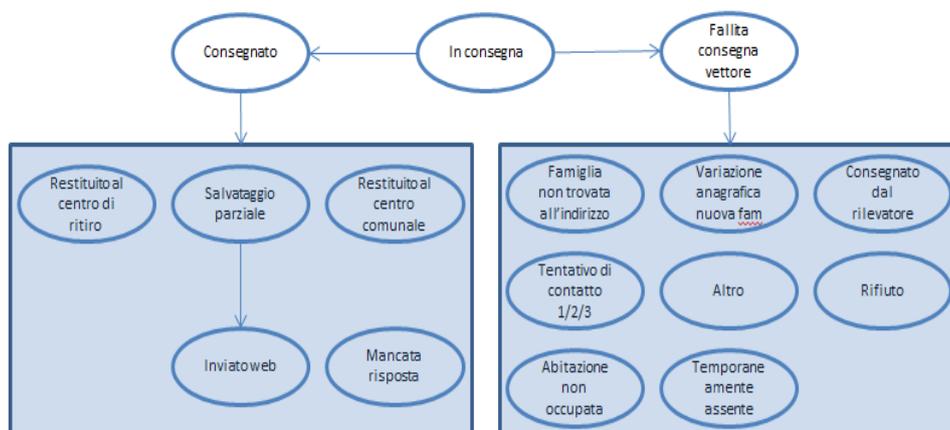


Figura 5.8 - Workflow della fase di restituzione del questionario

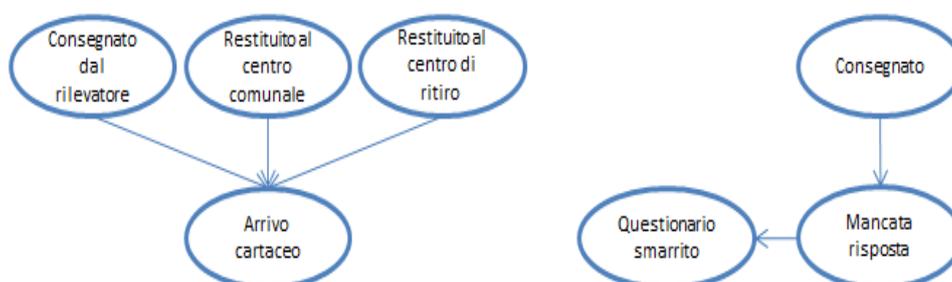
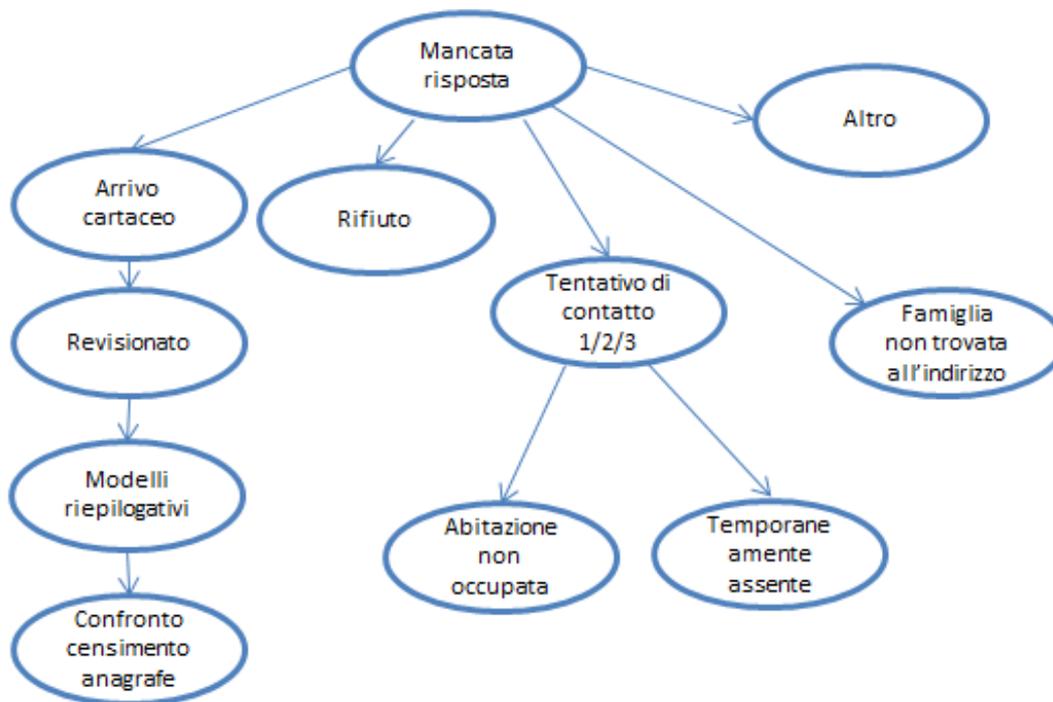


Figura 5.9 - Workflow della fase di lavoro sul campo



Il diario di sezione ha permesso quindi di monitorare e di modificare, dove possibile, le fasi di lavorazione tramite una serie di funzioni che ne hanno consentito la visualizzazione, l'aggiornamento e la correzione in casi eccezionali (figura 5.10):

- ▶ **visualizzazione.** La funzione permette all'operatore di effettuare una ricerca per visualizzare il diario della sezione di competenza;
- ▶ **visualizzazione dello storico del questionario** (figura 5.11): all'interno di ogni singolo questionario è sempre possibile visualizzare tutti i cambiamenti di stato, con l'identificativo dell'operatore che li ha effettuati e la data associata;
- ▶ **aggiornamento.** Questa funzione permette all'operatore di aggiornare il diario della sezione di competenza. L'informazione principale che può essere aggiornata è quella relativa allo stato del questionario; a seconda del tipo di stato, dal diario di sezione è possibile accedere ad altre funzioni, come ad esempio alla funzione di compilazione dei modelli riepilogativi, se lo stato del questionario si trova nella condizione "revisionato".

È inoltre possibile annullare le operazioni di cambiamento di stato effettuate.

Figura 5.10 - Esempio di diario di sezione

Diario di Sezione: Visualizzazione

Parametri della ricerca  
 Sezione: 1010001 , Sottosezione: 0  
 Rilevatore: rilroma1 rilroma1ta - 0580911521

Filtri per la ricerca  
 Stato questionario: Tutti  
 Indirizzo: Tutti

85 record trovati, visualizzati 1 - 10. [Prima/Prec.] 1, 2, 3, 4, 5, 6, 7, 8 [Succ./Ultima]

Codice questionario	Tipo modello	Fam/Conv o denom. convivenza	Cognome intestatario	Nome intestatario	Indirizzo	Civico	Esponente	Interno	Stato questionario	Data questionario	Storico del questionario
0580918619407	L	Fam	CONTINI	IOLANDA	VIA DEL QUIRINALE	00028			In consegna	31-08-2011	
0580918619406	L	Fam	CONTINI	IOLANDA	PIAZZA DEL QUIRINALE	00028			In consegna	08-06-2011	
0580917331930	L	Fam	OCCHIALINI	MAURIZIO	PIAZZA DEL QUIRINALE	00028	SC. B Pl. T	IN. 1	In consegna	08-06-2011	
0580917331931	L	Fam	OCCHIALINI	MAURIZIO	VIA DEL QUIRINALE	00028	SC. B Pl. T	IN. 1	In consegna	08-06-2011	
0580917378483	L	Fam	MONTESI	SANDRO	PIAZZA DEL QUIRINALE	00028	SC. C Pl. 4	IN. 20	In consegna	08-06-2011	
0580918631103	L	Fam	PUPPO	PAOLA	VIA DEL QUIRINALE	00028	SC. C Pl. 5	IN. 21	In consegna	08-06-2011	
0580918631102	L	Fam	PUPPO	PAOLA	PIAZZA DEL QUIRINALE	00028	SC. C Pl. 5	IN. 21	In consegna	08-06-2011	
0580918604699	L	Fam	GUIDI	ROBERTA	VIA DEL QUIRINALE	00028	SC. C Pl. 3	IN. 17	In consegna	08-06-2011	
0580918604698	L	Fam	GUIDI	ROBERTA	PIAZZA DEL QUIRINALE	00028	SC. C Pl. 3	IN. 17	In consegna	08-06-2011	
0580918600884	L	Fam	IACOPONI	SERGIO	VIA DEL QUIRINALE	00028			In consegna	08-06-2011	

Esporta dati:  

Figura 5.11 - Esempio di visualizzazione dello storico del questionario

Storico questionario Chiudi

Codice questionario	Stato questionario	Data questionario
0580918619407	Revisionato	09-08-2011
0580918619407	Restituito cartaceo	09-08-2011
0580918619407	In consegna	08-06-2011

Esporta dati:  

Nel normale flusso di lavorazione della rilevazione, per le varie unità si sono rese indispensabili le seguenti funzionalità, atte a gestire quelle situazioni (definite appunto eccezioni) che altrimenti non sarebbe stato possibile gestire nel sistema:

- ▶ **inserimento nuova riga.** Questa funzione permette all'operatore di inserire all'interno di una sezione/sottosezione nuove righe riguardanti:
  - abitazioni non occupate;
  - nuove famiglie/convivenze;
  - persone temporaneamente residenti.

Nella figura 5.12 è riportato un esempio di maschera per l'inserimento di una nuova famiglia.

Figura 5.12 - Maschera per inserimento nuova famiglia/convivenza

The screenshot shows a web form titled "Inserimento nuova riga". The form contains the following fields and controls:

- Sezione/Sottosezione:** Text input field containing "2070092".
- 0:** Text input field containing "0".
- Tipo Riga:** Text input field containing "Nuova Famiglia/Convivenza".
- Famiglia/Convivenza:** Dropdown menu.
- Codice Questionario:** Three text input fields, each containing "000".
- Cognome capofamiglia/Denominazione convivenza:** Text input field.
- Nome:** Text input field.
- Codice Fiscale:** Text input field.
- Indirizzo:** A sub-form containing:
  - Specie(DUG):** Dropdown menu.
  - Denominazione:** Text input field.
  - Civico:** Text input field.
  - Esponente:** Text input field.
  - Interno:** Text input field.
- Buttons:** "Inserimento nuova riga" and "Nuovo".
- Footnote:** "\*Campo Obbligatorio".

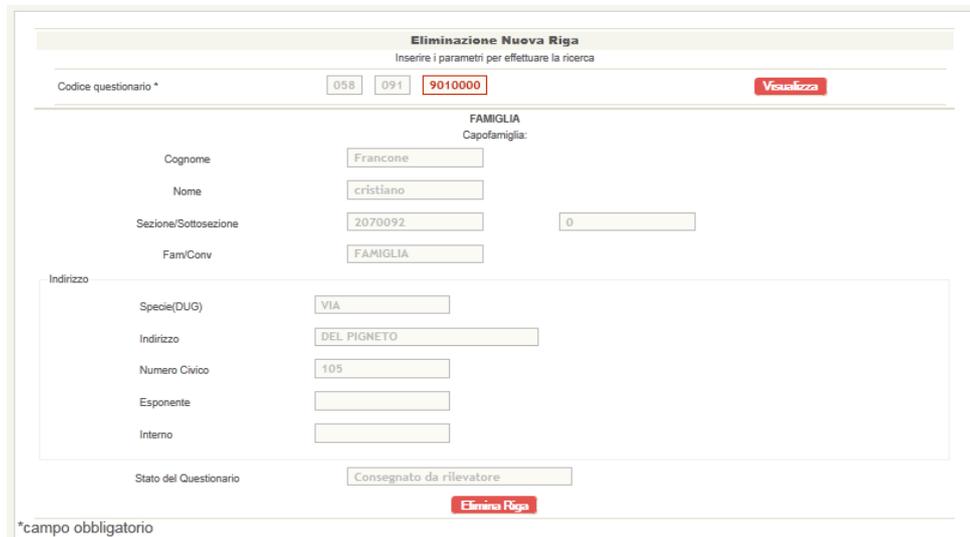
- ▶ **eliminazione nuova riga.** La funzione consente di eliminare una riga precedentemente inserita nel diario di una sezione.

Le nuove righe inserite che si riferiscono ad abitazioni non occupate non possono essere eliminate attraverso l'utilizzo di questa funzione. Per eliminarle è necessario utilizzare la funzionalità di modifica stato, presente all'interno della funzione "Aggiornamento diario".

Nella figura 5.13 è riportato un esempio di maschera per l'eliminazione di una riga precedentemente inserita.

- ▶ **spostamento dei questionari tra sezioni.** Questa funzione permette di effettuare lo spostamento di gruppi o di un singolo questionario da una sezione ad un'altra;
- ▶ **sostituzione questionario per smarrimento.** Questa funzione permette di creare una copia del questionario in caso di smarrimento dello stesso. Poiché ogni questionario associato ad una famiglia ha un codice identificativo univoco, la sostituzione in caso di smarrimento comporta anche la generazione di un nuovo identificativo e l'attribuzione del nuovo codice al questionario duplicato;
- ▶ **ripristina password per questionario on line.** Questa funzione permette all'operatore di generare una nuova password con le credenziali necessarie alla compilazione del questionario on line.

Figura 5.13 - Maschera per l'eliminazione di una riga



#### 5.4 I rapporti riassuntivi di monitoraggio

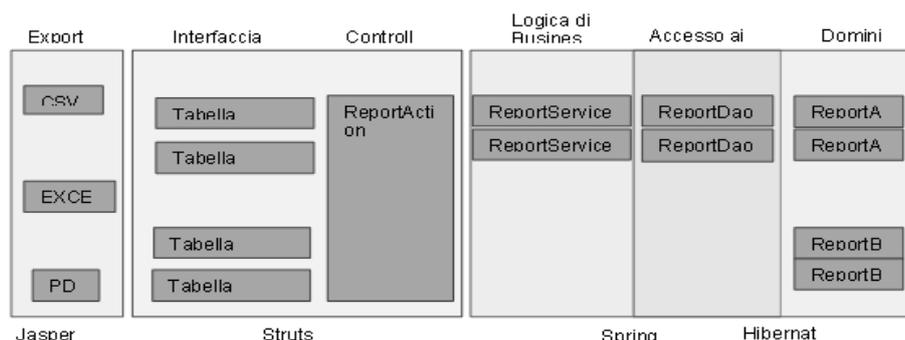
I rapporti riassuntivi hanno costituito la parte di vero e proprio monitoraggio della rilevazione, al fine di consentire una valutazione in tempo quasi reale dell'evoluzione di tutte le diverse fasi del processo censuario. Il controllo della rilevazione è stato effettuato mediante indicatori classificati territorialmente e raggruppati in forma tabellare, per riassumere sia in termini di quantità (assoluta o percentuale) sia per area tematica, il processo di rilevazione.

L'accessibilità alle funzioni di monitoraggio è stata definita per ruolo, livello e ambito territoriale dell'operatore della rete di rilevazione. Inoltre, il sistema gestiva un'elevata mole di dati quasi in real-time ed è stato fruibile per un elevato numero di utenti.

I diversi report di monitoraggio realizzati, ognuno associato ad una determinata fase del processo di rilevazione, erano navigabili tra i livelli territoriali ammessi e scaricabili nei formati *pdf*, *xls* e *csv*.

L'infrastruttura tecnologica è stata ideata come un modulo software integrato all'architettura complessiva del sistema di gestione della rilevazione. Quindi strutturato mediante i design pattern informatici quali Mvc (Model - View - Controller), Dao (Data Access Object) e IoC (Inversion of Control), come mostrato nella figura 5.14

Figura 5.14 - Architettura dei rapporti di monitoraggio



I dati visualizzati, erano letti da tabelle di macrodati popolate mediante un complesso sistema di viste materializzate. Nel dettaglio, per ogni rapporto riassuntivo erano presenti due tabelle di macrodati, di cui alternativamente, una era aggiornata e l'altra era in fase di aggiornamento. Per ottimizzare tale meccanismo, il modulo è stato progettato in modo da selezionare automaticamente la tabella dei macrodati opportuna.

L'interfaccia web, costituita a partire dagli strumenti propri dell'architettura generale, come Struts2, JQuery e le librerie DisplayTag, è stata personalizzata per permettere di rappresentare tabelle con intestazioni di colonna multiple, navigabili tra i livelli territoriali mediante la tecnica del drill-down e del roll-back, user-friendly.

Lo scarico dei dati è stato implementato aggiungendo la libreria Java JasperReports e il relativo plugin di Struts2, che hanno permesso di snellire tutta la gestione di esportazione dei formati dei file di dati, strutturandoli secondo template costruiti a partire dal tool designer IReport.

È stato inoltre previsto un meccanismo di salvataggio in memoria delle liste di dati, tra la fase di visualizzazione in pagina e la fase di esportazione del file. Tale salvataggio permetteva la creazione del file, evitando la re-interrogazione delle tabelle del database, migliorando così le prestazioni del download. Contemporaneamente, tale meccanismo evitava possibili problemi di disallineamento tra i dati visualizzati in pagina e quelli esportati nel file. Il derivante appesantimento nell'uso della memoria, è stato gestito mediante un processo di svuotamento automatico all'uscita dalla pagina Html. Infine è stata automatizzata l'intercettazione delle eccezioni con relativo invio e-mail di segnalazione errori.

La figura 5.15 rappresenta un esempio di rapporto riassuntivo realizzato attraverso il modulo. La struttura è analoga per tutte le funzioni di monitoraggio. Nello specifico, la navigazione tra i possibili livelli territoriali si effettua cliccando sull'icona a destra, per la fase di drill-down, e sul tasto indietro per il roll-back, mentre le funzioni di export nei formati *pdf*, *xls* e *csv* sono effettuate cliccando sulle apposite icone.

Figura 5.15 - Esempio di rapporto riassuntivo

Indicatore Andamento Rilevazione

122 record trovati, visualizzati 1 - 22. [Prima/Prec.] 1, 2, 3, 4, 5, 6 [Succ./Ultima]

Provincia: Roma - Aggiornato al: 27/07/2012 09:00

Comuni	Percentuale sul numero di famiglie e convivenze da LAC									Totale Questionari Ricevuti					
	Questionari Web			Consegnati a centro raccolta			Consegnati a rilevatore			Consegnati al punto di ritiro			LAC	Sottoc	Totale
	LAC	Sottoc	Totale	LAC	Sottoc	Totale	LAC	Sottoc	Totale	LAC	Sottoc	Totale			
AFFILE	26,73	44,71	3,45	68,17	0,15	0,15	0,30	2,25	0,00	2,25	93,84	3,60	97,45		
AGOSTA	19,28	38,67	0,00	38,67	16,02	3,25	19,28	15,78	0,00	15,78	89,76	3,25	93,01		
ALBAHO LAZIALE	25,62	25,19	2,41	27,59	5,38	2,40	7,78	27,83	0,32	28,15	84,03	5,12	89,15		
ALLUMIERE	64,86	20,47	4,49	24,96	0,05	0,05	0,11	6,46	0,00	6,46	91,84	4,54	96,39		
ANGUILLARA SABAZIA	32,25	35,72	3,52	39,24	1,19	0,54	1,73	16,48	0,01	16,49	85,65	4,07	89,72		
ANTICOLI CORRADO	44,01	30,28	1,21	31,59	10,89	4,36	15,25	3,92	0,00	3,92	89,11	5,66	94,77		
ANZIO	49,52	4,72	2,94	7,65	0,52	0,62	1,14	27,19	0,11	27,30	81,95	3,66	85,61		
ARCINAZZO ROMANO	57,48	30,37	2,22	32,59	0,00	0,00	0,00	4,74	0,00	4,74	92,59	2,22	94,81		
ARDEA	35,08	9,89	1,49	11,38	5,66	5,65	11,31	23,84	0,15	24,00	74,47	7,29	81,76		
ARICCIA	23,18	29,37	1,43	30,80	5,69	1,35	7,05	23,66	0,04	23,69	81,90	2,82	84,72		
ARSOLI	82,40	3,87	2,13	6,00	0,13	2,13	2,27	3,87	0,00	3,87	90,27	4,27	94,53		
ARTENA	17,01	0,28	0,00	0,28	59,88	3,46	63,34	9,86	0,00	9,86	87,04	3,46	90,49		
BELLEGRA	29,22	43,51	1,62	45,13	2,84	2,11	4,95	18,02	0,00	18,02	93,59	3,73	97,32		
BRACCIANO	65,57	8,21	3,78	11,99	1,30	1,89	3,19	13,24	0,00	13,24	88,32	5,67	93,99		
CAMERATA NUOVA	62,64	0,00	0,00	0,00	18,32	1,47	19,78	6,96	0,00	6,96	87,91	1,47	89,38		
CAMPAGNANO DI ROMA	41,93	29,85	7,27	37,12	0,28	0,06	0,34	12,23	0,04	12,27	84,28	7,38	91,66		
CAIALE MONTERANO	31,73	41,37	2,30	43,67	1,81	3,23	5,04	14,08	0,05	14,14	88,99	5,59	94,58		
CANTERANO	30,59	19,41	5,29	24,71	24,71	1,76	26,47	4,12	0,00	4,12	78,82	7,06	85,88		
CAPENA	18,63	46,62	6,37	52,99	3,77	1,01	4,78	11,98	0,16	12,14	81,00	7,54	88,55		
CAPRANICA PREHESTINA	41,56	20,99	0,41	21,40	14,81	2,06	16,87	4,94	0,00	4,94	82,30	2,47	84,77		
CARPINETO ROMANO	48,75	33,62	0,46	34,08	11,21	2,60	13,81	1,68	0,00	1,68	95,26	3,06	98,32		
CASAPE	73,23	1,05	2,62	3,67	9,97	3,67	13,65	1,31	0,00	1,31	85,56	6,30	91,86		

Indietro

Tutte le sezioni della Provincia

Se la lista dei dati visualizzata è costituita da un numero di record tale da non permetterne la visualizzazione agevole nella pagina web, la lista stessa viene paginata mediante la barra di navigazione evidenziata in alto, nella figura 5.15.

I rapporti riassuntivi realizzati per il monitoraggio della rilevazione sono 11:

- 1) stato spedizioni e consegne;
- 2) mancate consegne;
- 3) restituzione questionari per canale;
- 4) indicatore andamento rilevazione;
- 5) stato lavorazione questionari;
- 6) riepilogo attività rilevatori;
- 7) riepilogo attività operatori.
- 8) variazioni anagrafiche
- 9) lista edifici
- 10) confronto censimento anagrafe
- 11) modelli riepilogativi

I livelli territoriali di navigazione consentiti sono 5:

- 1) intero territorio nazionale: l'operatore visualizza tutte le informazioni suddivise per regione, provincia, comune e sezione;
- 2) regionale: l'operatore visualizza le informazioni suddivise per province, comuni e sezioni della regione di competenza;
- 3) prefettura: l'operatore visualizza le informazioni suddivise per comuni e sezioni della prefettura di competenza;
- 4) provinciale: l'operatore visualizza le informazioni suddivise per comuni e sezioni della provincia di competenza;
- 5) comunale: l'operatore visualizza le informazioni suddivise per le sezioni del comune di competenza.

Per ogni rapporto è possibile salvare i dati in tre formati, cliccando sulle apposite icone :

- 1) file *pdf*: che riporta i dati visualizzati nella pagina corrente (salvataggio per pagina);
- 2) file *xls*: che riporta i dati visualizzati nella pagina corrente (salvataggio per pagina);
- 3) file *csv* (comma separator value): che riporta i dati dell'intero rapporto;
- 4) per i rapporti "stato spedizioni e consegne", "mancate consegne", "restituzione questionari per canale", "indicatore andamento rilevazione", "stato lavorazione questionari" "riepilogo attività lavoratori" e "riepilogo attività operatori" è possibile anche scaricare un unico file *csv* così organizzato:
  - a) tutti i comuni d'italia (utenza nazionale);
  - b) tutti i comuni di una regione (utenza regionale);
  - c) tutte le sezioni di una provincia (utenza provinciale);
  - d) tutte le sezioni di una prefettura (utenza di prefettura).

In definitiva, le funzioni di monitoraggio hanno gestito un'elevata mole di dati quasi in tempo reale garantendo accessibilità e fruibilità ad un elevato numero di utenti.

### 5.5 La raccolta dei microdati

Durante il censimento della popolazione, oltre alla rilevazione delle famiglie sono state condotte altre rilevazioni, alcune delle quali hanno riguardato tutti i comuni d'Italia altre solo alcuni di essi. Il Sistema di gestione della Rilevazione ha permesso di inserire tutte le informazioni necessarie per coadiuvare tutte le indagini sia nella fase di raccolta dati sia in quella di monitoraggio.

In particolare per il censimento degli edifici sono state raccolte le principali informazioni derivanti dal modello cartaceo e contestualmente sono state predisposte liste, simili al diario di sezione, che consentissero il monitoraggio puntuale delle informazioni. Inoltre le province di Trento e Bolzano hanno potuto effettuare il monitoraggio della rilevazione delle minoranze linguistiche utilizzando un pannello di controllo analogo al diario di sezione, che aveva un suo work flow semplificato. In questo paragrafo infine, viene descritta anche il modulo che ha permesso, per il censimento della popolazione, di inserire le sole variabili primarie, ossia quelle anagrafiche necessarie alla prima diffusione dei dati.

### 5.5.1 I modelli riepilogativi

Non appena terminate le operazioni di revisione dei modelli acquisiti in modalità cartacea, i rilevatori dovevano inserire nella base dati alcune variabili, le cosiddette variabili primarie, in modo da poter anticipare la diffusione dei primi risultati del Censimento senza essere costretti ad aspettare il completamento delle attività di raccolta, acquisizione elettronica, controllo, correzione e validazione di tutti i dati del censimento.

I dati acquisiti tramite la compilazione web popolavano direttamente le informazioni dei modelli riepilogativi ed il rilevatore aveva la sola possibilità di monitorare le informazioni inserite dalla famiglia. Per consentire l'inserimento delle variabili primarie è stata realizzata la funzione 'Modelli riepilogativi' che è composta da tre sotto-funzioni:

- ▶ inserimento modelli riepilogativi;
- ▶ modifica modelli riepilogativi;
- ▶ rapporto riassuntivo del modello riepilogativo.

La funzione 'inserimento modelli riepilogativi' (figura 5.16) permette all'operatore di inserire il numero delle persone abitualmente dimoranti e non, il tipo di alloggio ed il numero di famiglie relative al questionario e i dati sul numero di Pad e Nad per sesso e cittadinanza italiana o straniera.

Figura 5.16 - Maschera di inserimento modello riepilogativo

Modelli Riepilogativi

Codice questionario: 058 091 8581253

Nome: BELLINI STEFANIA

Indirizzo: VIA TARQUINIO COLLATINO 00016

Sezione/Sottosezione: 2240065/0

Tipo Alloggio: nessun alloggio Numero Famiglie: Nessuna

Pers. abitualmente dimoranti (PAD) Pers. non abitualmente dimoranti (NAD)

Numero di maschi Numero di maschi

Numero di femmine Numero di femmine

Numero di stranieri

Totale PAD: 0 Totale NAD: 0

Salva

La funzione 'Modifica modelli riepilogativi' ha lo scopo di modificare le informazioni inserite in precedenza dall'operatore durante la funzione 'Inserimento modelli riepilogativi'. L'operatore ha la possibilità di modificare i seguenti dati: il numero delle persone abitualmente dimoranti e non, il tipo di alloggio ed il numero delle famiglie relative al questionario.

La funzione “Rapporto riassuntivo del modello riepilogativo” fornisce informazioni a livello aggregato territoriale dei dati contenuti nel modello riepilogativo, in questa funzione sono visibili anche i dati provenienti dalla compilazione on line.

### **5.5.2 Il censimento degli edifici**

In questa sezione vengono descritti gli aspetti principali del sistema informatico a supporto del Censimento degli Edifici 2011.

A tal fine si specifica che per edifici si intendono costruzioni dotate di una propria struttura indipendente, contenente spazi stabilmente utilizzati da persone per abitazione e/o produzione di beni e servizi, con le eventuali pertinenze.

Si ricorda altresì, quale attività propedeutica al Censimento, la Rilevazione dei Numeri Civici (Rnc), che è stata effettuata in tutti i Comuni e centri abitati al di sopra dei 20 mila abitanti ed ha permesso di definire un indirizzario arricchito delle informazioni indispensabili per la localizzazione delle unità statistiche sul territorio e il monitoraggio delle operazioni censuarie.

Le operazioni sul campo hanno previsto la raccolta delle informazioni da parte dei rilevatori specializzati sulla rilevazione edifici. Ciascun rilevatore si è avvalso di un modello di rilevazione cartaceo (Istat CP.ED), nonché di apposite funzioni a supporto nell’ambito di Sgr, in cui era possibile effettuare il monitoraggio delle operazioni, visionando la lista degli edifici censiti, il numero del questionario cartaceo di ciascun edificio, i toponimi di riferimento. Attraverso Sgr è stato inoltre possibile scaricare l’itinerario di sezione derivante dall’Archivio Nazionale Stradari e Civici, ed il Diario di Sezione. Nello svolgimento delle operazioni censuarie, il rilevatore nel percorrere le strade (prima su un lato, poi sull’altro, nel caso in cui entrambi i lati di una strada siano compresi all’interno della stessa sezione) ha censito gli edifici incontrati, accertandosi della presenza di entrate su altri lati dell’isolato; registrando contestualmente le strade completate grazie all’ausilio dell’Itinerario di sezione.

Nello specifico, in Sgr sono state messe a disposizione le seguenti funzioni:

- 1) inserimento nuovo edificio;
- 2) modifica/eliminazione edificio;
- 3) lista edifici;
- 4) rapporto riassuntivo edificio.

Attraverso la prima funzionalità è possibile inserire il codice questionario di un nuovo edificio censito, nonché alcune informazioni relative, quali il fatto di essere un edificio residenziale o meno, nonché la lista dei toponimi associata. Le informazioni relative al toponimo sono codificate nel sistema informativo sottostante Sgr, in particolare esse sono suddivise nell’indicazione della specie, della denominazione, del civico e dell’esponente. Tra i vantaggi legati ai toponimi così strutturati si ricorda l’efficienza introdotta nelle ricerche degli indirizzi nell’ambito del sistema web; infatti si è ridotto lo spazio di ricerca al solo indirizzo, fissando i valori degli altri campi costituenti il toponimo. A tal proposito, si evidenzia che sono state introdotte, ad ausilio del rilevatore sia in fase di inserimento che di ricerca, tecniche di autocompletamento automatico, limitando la quantità di possibili indirizzi consigliati all’utente lato client nei casi in cui lo spazio dei possibili risultati fosse eccessivamente oneroso. In caso di non presenza nell’indirizzario codificato di cui sopra di uno specifico toponimo, attraverso Sgr si è data la possibilità di inserire/eliminare nuovi toponimi attraverso specifiche maschere, così da arricchire eventualmente la base informativa relativa allo stradario. Si evidenzia che, in fase di raccolta di tali informazioni tramite il sistema Sgr, ciascun edificio inserito/modificato è associato con l’identificativo del particolare operatore che ha trattato il questionario di interesse.

Le informazioni inserite relative ad un edificio sono successivamente modificabili/eliminabili attraverso la seconda funzione, dando indicazione del numero del questionario associato.

La lista degli edifici associati ad un dato toponimo è altresì visualizzabile/scaricabile attraverso la funzione omonima.

La quarta funzione infine consente il monitoraggio delle operazioni di rilevazione attraverso un apposito rapporto riassuntivo, in cui vengono conteggiati gli edifici censiti, suddivisi tra edifici residenziali e non residenziali.

### **5.5.3 Le minoranze linguistiche**

La funzione Diario Minoranze Linguistiche è la seconda voce nel menu “sezione” ed è riservata ai soli operatori delle province autonome di Trento e Bolzano. Questa funzione permette di gestire i questionari relativi alle minoranze linguistiche.

La prima schermata, “Ricerca Questionario Minoranze Linguistiche”, consente di specificare i filtri per ottenere una selezione dei questionari. L’utente ha la possibilità di specificare o un’intera sezione o il codice questionario specifico. Il bottone “Cerca” rimanda l’utente ad una nuova schermata, “Aggiornamento Diario Minoranze Linguistiche”, che presenta la lista dei questionari coerenti con i criteri di ricerca prima definiti. Per ogni questionario vengono mostrate alcune proprietà (codice questionario, numero di sezione, proprietà famiglia o convivenza, codice del rilevatore, nome e cognome dell’intestatario, numero di fogli consegnati/aggiunti/ritirati, data di consegna e data di ritiro, stato del questionario, indirizzo dell’intestatario). Da questa schermata è inoltre possibile effettuare una nuova selezione sui questionari basata sull’indirizzo e sullo stato del questionario.

Per ogni questionario mostrato è inoltre possibile accedere ad una pagina di dettaglio tramite cui l’utente ha la possibilità di modificare le proprietà del questionario. Questa schermata permette di modificare lo stato del questionario, la data di consegna e il numero di fogli aggiunti e ritirati.

## **5.6 Le operazioni di chiusura e invio materiale compilato**

### **5.6.1 Le modalità operative**

Sgr ha avuto un ruolo chiave per le operazioni di chiusura delle attività censuarie, in particolare per il ritiro del materiale cartaceo presso le sedi comunali. Tale processo ha coinvolto diversi attori: i responsabili comunali per la preparazione del materiale da inviare, il vettore postale per il ritiro del materiale e il trasporto al punto di raccolta, il gestore del punto di raccolta per la verifica e il tracciamento dell’arrivo del materiale, da sottoporre poi a lettura ottica. In particolare il processo di spedizione prevedeva i seguenti passi:

- 1) predisposizione da parte del comune del materiale da inviare in apposite scatole e secondo le disposizioni dell’Istat;
- 2) stampa della documentazione necessaria per la spedizione;
- 3) prenotazione del ritiro;
- 4) scarico da parte del vettore postale delle prenotazioni di ritiro;
- 5) ritiro da parte del vettore postale.

Gli uffici comunali hanno predisposto una o più spedizioni, ognuna composta da una o più scatole, confezionate secondo le indicazioni fornite da Istat. Per ogni ritiro da parte del vettore postale i comuni hanno compilato e stampato tramite Sgr:

- ▶ un documento di trasporto (ddt) per ogni spedizione;
- ▶ una distinta per ogni scatola della spedizione.

Il Ddt era quindi associato a più distinte e conteneva le informazioni legate alla spedizione, ovvero il centro di ritiro con denominazione e indirizzo (es. Ufficio Comunale di Censimento di Roma 3 – Via Ravà, 121), la data di ritiro, il numero di scatole per tipologia da ritirare, eventuali note utili agli incaricati del vettore postale. Era presente anche un flag per segnalare se la spedizione corrente fosse l’ultima prevista per il comune. La distinta conteneva invece le informazioni sul contenuto della scatola, in particolare le quantità dei modelli spediti e un flag per segnalare il termine della spedizione di un modello.

Per preparare la documentazione di trasporto l'operatore comunale ha utilizzato le specifiche funzioni di Sgr per:

- ▶ creare un documento di trasporto;
- ▶ creare le distinte del documento di trasporto;
- ▶ ove necessario modificare o cancellare una distinta;
- ▶ chiudere il documento di trasporto (con tale operazione il documento di trasporto e le relative distinte non erano più modificabili e veniva prenotato il ritiro da parte del vettore postale);
- ▶ stampare il documento di trasporto e le relative distinte.

Sgr ha supportato l'operatore nella predisposizione della documentazione, attraverso form parzialmente precompilati e dinamici, che consentivano: la creazione dell'identificativo del Ddt, associato al codice provincia e codice comune dell'operatore connesso; la selezione del centro di ritiro tra quelli disponibili; la creazione dell'identificativo univoco Istat della distinta e la sua associazione al relativo Ddt; la selezione della tipologia di scatola da preparare, con la visualizzazione dei relativi modelli inseribili nella scatola; la scelta della data in cui prenotare il ritiro; la generazione del codice a barre, secondo le specifiche richieste dal vettore postale, per l'individuazione univoca della scatola; la stampa di tutta la documentazione di trasporto, ovvero Ddt e relative distinte.

Il vettore postale ha avuto accesso a Sgr con un profilo utente specifico che consentiva di scaricare l'elenco dei Ddt da ritirare, in un certo giorno o in un'area territoriale prescelta, e di stampare la documentazione di trasporto. In tal modo è stato possibile gestire in modo automatico e snello il processo di interazione tra i comuni e il vettore postale, consentendo al contempo all'Istat di avere una visione completa e dettagliata del processo di spedizione delle scatole e di chiusura delle operazioni censuarie sul campo.

### ***5.6.2 I controlli quantitativi sul ritorno dei questionari***

In fase di restituzione dei questionari all' Istat da parte degli Uffici Comunali di Censimento sono stati eseguiti molteplici controlli quantitativi per assicurarsi che non si verificassero errori o disguidi. Da un lato era necessario evitare ritardi e inconvenienti col vettore, dall'altro si presentava la necessità di eseguire una verifica di coerenza finale relativamente alle attività degli uffici in tempo utile perché potessero ancora essere presi dei provvedimenti.

I controlli quantitativi venivano applicati in più momenti:

- ▶ in fase di chiusura di ciascun Ddt in Sgr;
- ▶ in fase di chiusura delle operazioni di ritiro, sempre in Sgr;
- ▶ internamente all' Istat per confrontare le informazioni provenienti dagli Ucc con le analoghe pervenute dal vettore.

Alla chiusura di ciascun Ddt il sistema verificava, per ciascuna tipologia, che il totale dei modelli per i quali era stato richiesto il ritiro sino a quel momento non fosse superiore al totale dei modelli lavorati dal comune all'interno del diario di sezione o della lista degli edifici.

Alla chiusura delle operazioni di ritiro, cioè alla chiusura del Ddt sul quale il comune dichiarava che si trattava dell'ultima richiesta di ritiro, venivano eseguiti i controlli più approfonditi di seguito elencati:

- ▶ confronto fra i totali di questionari per ciascuna tipologia lavorati in fase di rilevazione ed i corrispondenti aggregati per i quali era stato richiesto il ritiro;
- ▶ verifica di dettaglio tra unità di rilevazione pervenute ed attese.

Le discordanze venivano segnalate all'utente in modo che il comune potesse effettuare le necessarie azioni correttive. I dati aggregati sono stati valutati tramite un criterio basato su soglie di accettabilità in quanto erano ammissibili piccoli scostamenti tra le quantità finali.

Il sistema Sgr, inoltre, visualizzava dei rapporti riassuntivi che consentivano, ai vari livelli territoriali, di monitorare l'andamento delle operazioni nel massimo dettaglio.

Successivamente al ritiro, venivano eseguiti fuori linea controlli di dettaglio fra le quantità dichiarate dai comuni e le corrispettive pervenute dal vettore incaricato del ritiro. I relativi risultati dovevano essere ottenuti in tempo reale per poter riscontrare, rimanendo nei tempi contrattuali, eventuali problemi al vettore così che si potesse intervenire in tempo. Il rispetto delle scadenze era, chiaramente, fondamentale perché vincolato dal contratto. Venivano, innanzitutto, caricate nella base dati le informazioni inviate dal vettore tramite file dal tracciato prestabilito. I controlli erano eseguiti mediante procedure batch che producevano in output report resi disponibili con cadenza costante al personale Istat coinvolto nei rapporti col vettore.

La fase di ritiro è stata controllata capillarmente e ciò ha permesso di limitare al massimo i ritardi e le contestazioni sia con i comuni sia col vettore esterno.

### **5.6.3 La gestione delle eccezioni**

La procedura di confezionamento dei plichi da parte dell'Ucc e la contestuale predisposizione dei relativi documenti per l'invio del materiale ai Centri di acquisizione ottica, rispondeva ad un'Informativa Istat di cui il personale comunale incaricato era tenuto a rispettare le linee guida, elaborate al fine di ridurre le eventuali anomalie (eccezioni) indicate dagli esempi descritti:

- ▶ discordanze fra la quantità effettiva del materiale da ritirare ed il dettaglio indicato sulle distinte e/o il documento di trasporto (Ddt);
- ▶ errori di inserimento plichi all'interno di scatole di tipologia non prevista per gli stessi;
- ▶ necessità di modifica distinte e/o Ddt;
- ▶ mancato o parziale ritiro da parte del corriere.

La sola riapertura di un Ddt chiuso era, invece, un'operazione permessa esclusivamente al Rit (presente con un Responsabile per Regione) e al personale Istat addetto all'assistenza ed al monitoraggio di questa specifica fase operativa. Tale operazione veniva svolta ugualmente tramite Sgr. La riapertura, però, poteva essere eseguita in automatico solo se il Ddt non era stato ancora scaricato dal vettore incaricato del ritiro, altrimenti le modifiche nei dati dovevano essere richieste tramite il supporto telefonico. Le eventuali eccezioni rispetto alla procedura standard avrebbero infatti potuto comportare non solo disguidi nel processo di ritiro, ma soprattutto difficoltà a vari livelli con il fornitore, in quanto si rischiava di andare in deroga al contratto in essere. Era stato quindi allestito un supporto Istat telefonico e telematico chiamato a coordinare – in caso di richiesta di assistenza – le operazioni relative alla produzione della documentazione da parte dell'Ucc tramite il sistema Sgr e a risolvere, insieme al personale Ucc, ogni altra problematica, come ad esempio la predisposizione di richieste di ritiro multiplo da eseguirsi, da parte del corriere, nell'arco di due o più giornate. Gli Uffici di censimento inoltre dovevano segnalare telefonicamente eventuali errori nella compilazione delle distinte o dei Ddt commessi successivamente allo scarico da parte del vettore. In tal modo le informazioni venivano passate al servizio di supporto del vettore per intervenire tempestivamente sul processo di ritiro. Successivamente venivano anche aggiornate le informazioni nel sistema Sgr in modo da mantenere coerenti i dati in tutti gli archivi.

## 5.7 Strumenti di comunicazione con la rete

### 5.7.1 Il portale documentale

Il Portale della Rete Censuaria è stato realizzato con l'obiettivo di fornire un supporto a tutti gli operatori della rete censuaria (responsabili comunali, rilevatori, personale Istat, ecc.). Gli utenti abilitati potevano accedere e trovare le informazioni necessarie per il corretto svolgimento della rilevazione.

La struttura delle pagine del portale (figura 5.17) è composta da due menu orizzontali posizionati nella parte alta e da un'area dei contenuti posta in primo piano.

Figura 5.17 - Homepage del portale



Istat.it

**RETE** Il Sito per la Rete Censuaria

Home **Materiale di rilevazione** Strumenti Documenti Domande&Risposte Video **SGR**

Contatti Formazione Questionari in giacenza Questionari non consegnati Mappa Cerca Logout

Sei in: Home

**NEWS**

06.04.2012  
**Riattivazione del sistema SGR**  
 Il sistema SGR è stato riattivato dopo un fermo di circa 20 ore reso indispensabile per effettuare operazioni di manutenzione straordinaria. L'attività è stata necessaria a causa di un inconveniente tecnico che purtroppo ha... **[more]**  
 Category: **Comunicazioni di servizio**

23.02.2012  
**Questionari in giacenza**  
 E' stato inserito il file relativo all'ottavo invio dei questionari in giacenza all'Istat. **[more]**  
 Category: **Aggiornamenti**

30.01.2012  
**Aggiornamento Area Documenti**  
 Sono state inserite nel sito due nuove informative. La prima riguarda la chiusura dei canali di restituzione mentre la seconda contiene le istruzioni per la generazione delle distinte di spedizione e del documento di trasporto... **[more]**  
 Category: **Aggiornamenti**

29.12.2011  
**Informativa del 28 Dicembre 2011**  
 Nella'area Documenti/Informative e comunicazioni è stata inserita l'informativa del 28 Dicembre 2011 riguardante i questionari in giacenza in Istat dal 16 al 27 dicembre 2011. **[more]**  
 Category: **Aggiornamenti**

**Archivio**

**Istat - Istituto nazionale di statistica**  
 Via Cesare Balbo 16 00184 - Roma tel. +39 06 46731

Attraverso il primo menu (quello posizionato nella riga in alto) si può navigare tra le varie aree del portale ("Home", "Materiale di rilevazione", "Strumenti", "Documenti", "Domande&Risposte", "Video", "Sgr"). La descrizione dettagliata dei contenuti informativi nelle varie aree è riportata nel fascicolo 1.

Inoltre l'area questionari non consegnati (figura 5.18) mette a disposizione files contenenti le informazioni sui questionari non consegnati selezionabili per comune.

Figura 5.18 - Area Questionari non consegnati

The screenshot shows the Istat website interface for the 15th Census. At the top right is the Istat.it logo. Below it is a navigation bar with the following items: Home, Materiale di rilevazione, Strumenti, Documenti, Domande&Risposte, Video, SGR, Contatti, Formazione, Questionari in giacenza, Questionari non consegnati, Mappa, Cerca, and Logout. The 'Questionari non consegnati' item is highlighted. Below the navigation bar, there is a section titled 'Sei in: Questionari non consegnati'. This section contains three dropdown menus: 'Regione' (seleziona una regione), 'Provincia' (seleziona una provincia), and 'Comune' (seleziona un comune). At the bottom of the page, there is contact information for Istat - Istituto nazionale di statistica: Via Cesare Balbo 16 00184 - Roma tel. +39 06 46731.

### 5.7.2 Attività di helpdesk

Per garantire agli operatori l'assistenza necessaria nella gestione dei sistemi web (Qpop, Sgr, Rete) sono stati attivati diversi canali di comunicazione e messi a disposizione strumenti aggiuntivi. Il primo di questi strumenti è rappresentato dal manuale per l'utilizzo del Sistema di Gestione della Rilevazione, reso disponibile on line e in formato pdf, dalla homepage di Sgr.

Il manuale è stato uno strumento fondamentale per la formazione di tutti gli operatori ed è stato predisposto per accompagnare l'utente in ogni fase della rilevazione, partendo dalle istruzioni necessarie per effettuare l'autenticazione, sino ad arrivare alle operazioni di chiusura del censimento. Il manuale è stato appositamente strutturato per offrire l'assistenza necessaria ad ogni tipologia di operatore, dal responsabile dell'ufficio provinciale di censimento, fino al singolo rilevatore. È stato quindi necessario evidenziare, per ogni funzione chi poteva accedere alla stessa e le operazioni permesse. Per cercare di rendere, il più chiaramente possibile, tutti i passaggi da eseguire sul sistema di gestione e rilevazione (Sgr), nel manuale si è fatto un largo uso di immagini esplicative.

L'Istat, inoltre, ha messo a disposizione degli operatori diversi canali di comunicazione, per trasmettere informazioni e disposizioni cui attenersi, al fine di un corretto svolgimento delle attività da realizzare durante il censimento: un servizio di help-desk telefonico e un servizio telematico, attraverso i quali è stato possibile richiedere informazioni e assistenza. Entrambi i servizi sono stati operativi per tutto il periodo del censimento, ed hanno garantito agli operatori tutte le spiegazioni e le chiarificazioni sull'utilizzo dei sistemi. Questi strumenti sono stati indispensabili perché hanno consentito la segnalazione tempestiva delle eccezioni che si potevano verificare sul campo e la loro veloce risoluzione. Per il Sistema di helpdesk telefonico gli operatori avevano accesso diretto al sistema e funzioni dedicate, che permettevano in moltissimi casi di risolvere l'eccezione verificatasi. Per il servizio telematico invece esisteva un helpdesk di primo livello che smistava al personale più tecnico le problematiche di carattere informatico. Per effettuare tale servizio è stato necessario formare personale interno alla conoscenza sia del sistema informatico che della gestione dell'indagine.

Gli stessi canali hanno avuto anche l'obiettivo di permettere una comunicazione costante con gli Upc durante tutta la durata delle operazioni censuarie.

L'ultimo, ma non meno importante mezzo di assistenza e supporto messo a disposizione dell'Istat sono state le Frequently Asked Questions, meglio conosciute con la sigla Faq, che sono una serie

di risposte stilate rispetto alle domande più frequenti. Con questo canale si è cercato di dare una risposta veloce alle problematiche più comuni. Le Faq sono state pubblicate sul portale della rete, cioè sul sito web adibito alla gestione di tutto il materiale e del flusso documentale, e aggiornate costantemente durante la rilevazione.

### **5.7.3 Fornitura di dati di monitoraggio off line**

Nonostante in Sgr siano stati resi disponibili numerosi rapporti di monitoraggio e numerose liste disponibili on line e memorizzabili localmente, per le esigenze specifiche dei Comuni di informazioni di dettaglio che, a causa della elevata numerosità, non potevano essere scaricate direttamente da internet, sono stati resi disponibili, ai Comuni che ne facevano richiesta, file in formato csv. In particolare sono stati predisposti i seguenti file:

Diario di sezione, Diario di sezione esteso, Diario storico, Lista edifici, Lista censiti altro indirizzo, Lista censiti non Lac, Lista censiti, Lista non censiti Lac, Report data entry, Microdati (alloggio, famiglie coabitanti, fogli individuali\_conv, fogli individuali, listaA\_conv, listaA, listaB, NonAbitDim\_conv\_).

Per i Comuni che ne facevano richiesta i suddetti file venivano aggiornati e inviati con cadenza settimanale o bisettimanale.

I file csv sono stati inviati come archivio zip/rar mediante il sito ufficiale di data exchange dell'Istat, Indata. Per far questo sono state definite 4 utenze temporanee. Caricati gli archivi su una delle quattro utenze Indata a disposizione, si è impostata una password di accesso, che veniva comunicata separatamente al referente del Comune, per eseguire il download.

Terminato il download da parte del referente del Comune, si provvedeva a modificare la password dell'utenza, rendendola disponibile per un invio successivo allo stesso Comune o a un altro Comune.

Durante la rilevazione sono stati predisposti e inviati ai Comuni oltre 2 mila 500 file su richiesta di 320 Comuni.



## CAPITOLO 6

### Il sistema per il confronto tra censimento e anagrafe<sup>18</sup>

Nell'ambito del 15° Censimento generale della popolazione e delle abitazioni ciascun Ufficio Comunale di Censimento ha effettuato il confronto tra gli individui risultanti in Lac all'8 ottobre 2011 e gli individui trovati al censimento, contestualmente alle operazioni di rilevazione. A tal fine ci si è potuto avvalere di una specifica funzionalità del Sistema di Gestione della Rilevazione (Sgr) che consentiva di esaminare ogni singolo questionario compilato, Web o cartaceo, per confrontare la lista A compilata dai cittadini (Lista delle persone abitualmente dimoranti nell'alloggio/convivenza) con le risultanze dell'anagrafe comunale alla data del censimento presente nel sistema stesso.

Man mano che gli Uffici comunali di censimento eseguivano le operazioni di confronto, avveniva la compilazione automatica in Sgr di bilanci quantitativi per sintetizzarne gli esiti. Questo ha reso immediatamente disponibili e valutabili i saldi quantitativi relativi alle persone censite e iscritte in anagrafe, alle persone censite e non ancora iscritte in anagrafe, alle persone non trovate al censimento ma iscritte in anagrafe.

Oltre ai bilanci quantitativi, al termine delle operazioni di confronto, gli uffici comunali di censimento hanno potuto prelevare dal sistema le liste individuali necessarie per l'aggiornamento dell'anagrafe, in modo da usufruire del lavoro svolto in fase di rilevazione. In particolare, hanno potuto analizzare tempestivamente le liste nominative riguardanti i due aggregati di popolazione su cui effettuare la revisione post-censuaria:

- 1) gli individui censiti come dimoranti abitualmente nel Comune ma non presenti in Lac e "candidati" ad essere iscritti in anagrafe;
- 2) gli individui dichiarati non trovati al censimento ed "eleggibili" ad essere cancellati dall'anagrafe.

Le funzionalità di confronto tra censimento e anagrafe messe a disposizione in Sgr venivano abilitate soltanto per i comuni per i quali risultavano terminate le attività di inserimento delle variazioni anagrafiche manuali oppure per i quali l'Istat aveva completato il processo di elaborazione delle variazioni anagrafiche tramite processi batch. In questo secondo caso i dati anagrafici venivano aggiornati nel sistema a seguito dell'analisi delle due liste anagrafiche comunali pervenute all'Istat: quella relativa alla data del 31 dicembre 2010 e quella riferita alla data del censimento. In questo modo i dati anagrafici e le risultanze censuarie avevano lo stesso riferimento temporale.

La modalità operativa di confronto via Web, nuova rispetto ai precedenti censimenti, ha consentito all'Istat di coordinare e monitorare il processo per far sì che fosse effettuato in modo omogeneo sul territorio, rapido e corretto.

#### 6.1 Gestione delle variazioni anagrafiche interattive

Selezionando la funzione "Variazioni anagrafiche" sul pannello di Sgr si consentiva ai Comuni con popolazione inferiore ai 20 mila abitanti di registrare le variazioni anagrafiche intercorse tra il 1° gennaio 2011 e la mezzanotte dell'8 ottobre 2011 ai fini di un corretto utilizzo della funzione Confronto Censimento Anagrafe. I comuni con popolazione superiore, invece, inviavano una seconda lista anagrafica comunale riferita alla data del censimento e l'Istat si occupava di ricavare le variazioni anagrafiche tramite un processo batch.

<sup>18</sup> Il capitolo è a cura di Daniela Casale. Il paragrafo 6.1 è stato redatto da Roberta Radini; il paragrafo 6.2 è stato redatto da Cristiano Maione; il paragrafo 6.3 è stato redatto da Maria Pia Di Maio.

La funzione interattiva prevedeva alcune predeterminate operazioni, come si vede nella figura 6.1.

Figura 6.1 - Pannello delle variazioni anagrafiche in Sgr



“Inserimento nuova famiglia” e “Inserimento nuova convivenza” permettevano di inserire un nuovo nucleo familiare o una nuova convivenza all’interno di una sezione di censimento.

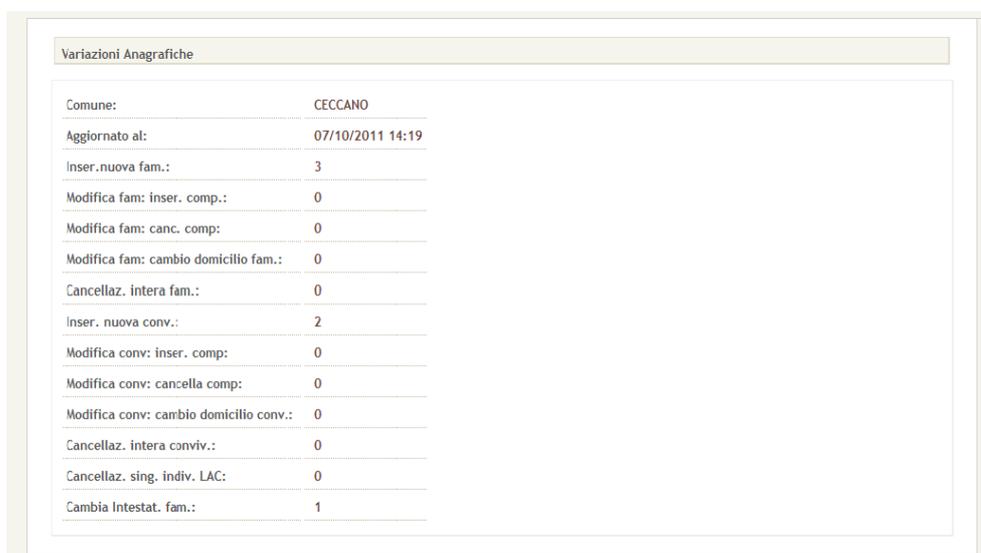
“Modifica famiglia” e “Modifica convivenza” consentivano all’operatore di modificare i componenti di un nucleo familiare/convivenza, eliminare o aggiungere un componente oppure modificare l’intestatario del nucleo stesso. Tramite la funzione di modifica, l’operatore poteva anche eliminare una famiglia o una convivenza, dopo aver specificato quale fosse il motivo della cancellazione tra una lista predefinita di possibili causali. Era, infine, prevista la registrazione del cambio di domicilio della famiglia/convivenza. Le variazioni anagrafiche di inserimento, modifica ed eliminazione di famiglie e convivenze avevano un effetto immediato nel diario di sezione, dove venivano opportunamente inserite o modificate delle righe in corrispondenza delle unità lavorate.

La funzione “Rapporto riassuntivo su variazioni anagrafiche” permetteva la visualizzazione del complesso delle variazioni effettuate sul territorio del comune di pertinenza come si può vedere nella figura 6.2.

La funzione “Chiudi Variazioni”, infine, permetteva al responsabile comunale di dichiarare la chiusura della fase di inserimento delle variazioni anagrafiche. In tal modo il comune poteva passare alla fase di confronto censimento anagrafe.

Il sistema offriva la possibilità di esportare in formato Excel o Pdf l’elenco degli individui coinvolti nelle variazioni.

Figura 6.2 - Rapporto riassuntivo delle variazioni anagrafiche



Variazioni Anagrafiche	
Comune:	CECCANO
Aggiornato al:	07/10/2011 14:19
Inser.nuova fam.:	3
Modifica fam: inser. comp.:	0
Modifica fam: canc. comp:	0
Modifica fam: cambio domicilio fam.:	0
Cancellaz. intera fam.:	0
Inser. nuova conv.:	2
Modifica conv: inser. comp:	0
Modifica conv: cancella comp:	0
Modifica conv: cambio domicilio conv.:	0
Cancellaz. intera conviv.:	0
Cancellaz. sing. indiv. LAC:	0
Cambia Intestat. fam.:	1

## 6.2 Le operazioni dei comuni

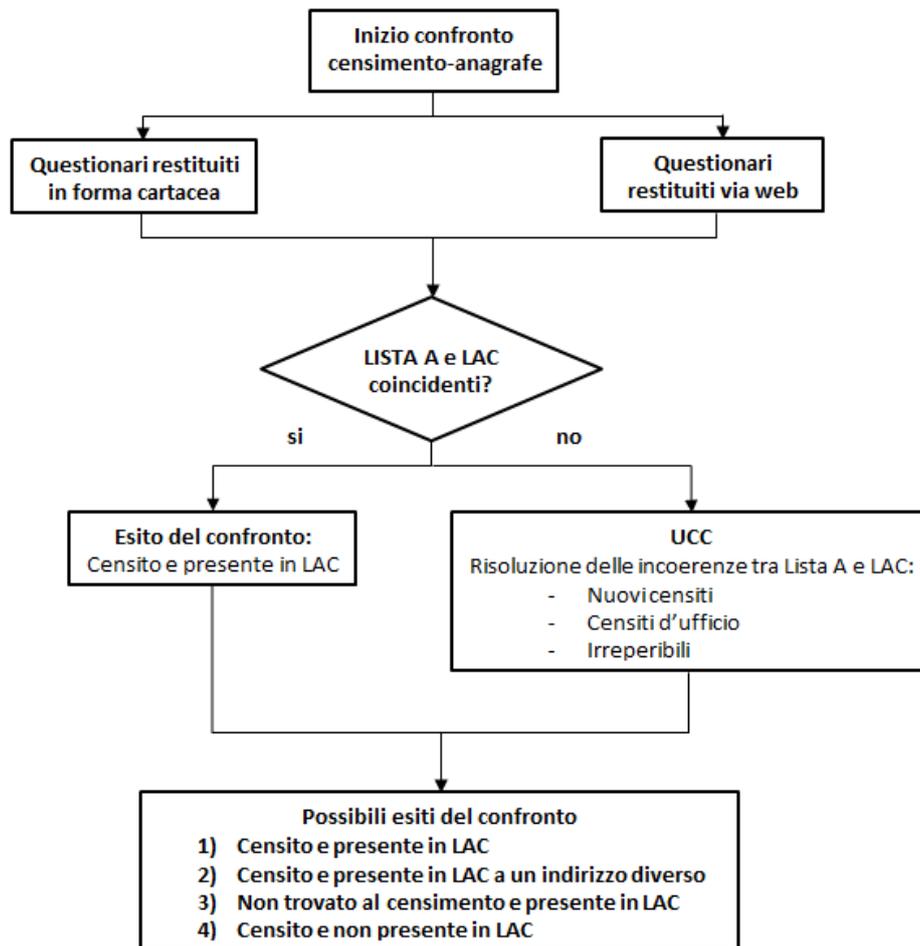
All'interno del Sistema di Gestione della Rilevazione gli uffici comunali di anagrafe potevano adempiere alla registrazione delle attività di confronto censimento anagrafe. Lo schema in figura 6.3 sintetizza il flusso di lavoro. Queste operazioni venivano effettuate dal back office dell'Ufficio Comunale di Censimento, dopo la revisione qualitativa dei questionari.

Il sistema distingueva i questionari compilati in forma cartacea e quelli compilati via Web dai rispondenti, trattandoli in maniera differente.

Per i questionari cartacei gli operatori comunali hanno avuto la possibilità di confrontare gli individui presenti nella Lista Anagrafica Comunale (Lac) presente nel sistema Sgr con i dati identificativi delle persone dimoranti abitualmente nel Comune, riportati nei singoli questionari di rilevazione nella lista A. Per ciascun individuo presente in Lac è stato possibile confermare la presenza sul questionario e dichiararlo censito, oppure riscontrarne l'assenza e dichiararlo irreperibile al censimento. Un'apposita funzione ha anche consentito di riconoscere e conteggiare gli individui censiti come dimoranti abitualmente ma non iscritti in anagrafe, identificandoli come nuovi individui. L'operatore ha avuto a disposizione funzioni che consentivano di riconoscere e trattare eventuali individui censiti più volte nel Comune, oppure di identificare gli individui censiti ad un indirizzo dello stesso Comune ma diverso da quello risultante in anagrafe.

La figura 6.4 mostra un esempio di confronto su un questionario cartaceo.

Figura 6.3 - Flusso di lavoro del confronto censimento anagrafe



Per i questionari compilati via Web, invece, Sgr ha eseguito automaticamente il confronto fra i dati presenti in anagrafe e quelli compilati on line dai rispondenti, consentendo all'operatore di effettuare opportune verifiche per evitare casi di doppio conteggio e assicurare il riconoscimento della coerenza o meno con i dati anagrafici. La figura 6.5 mostra un esempio di confronto su un questionario Web.

Ogni volta che, in fase di confronto, veniva inserito in un questionario un individuo "nuovo" rispetto alla Lac, il sistema automaticamente eseguiva una ricerca all'interno del comune per accertarsi che non si trattasse di un doppione. Tale ricerca, basata su algoritmi di vicinanza statistica, proponeva all'operatore alcuni nominativi quali possibili duplicazioni lasciando in tal modo all'utente la possibilità di confermare e poi eliminare il doppione oppure di dichiararlo non veritiero. In tal modo il sistema offriva agli operatori gli strumenti necessari per assicurare nelle fasi successive il corretto aggiornamento dell'anagrafe comunale.

Figura 6.4 - Esempio di confronto su un questionario cartaceo

<p>██████ EVELINA Femmina nato a RIVAROLO DEL RE ED UNITI (Cremona) il 1/4/1914</p> <p>██████</p>	<p>██████ EVELINA Femmina nato a RIVAROLO DEL RE ED UNITI il 1/4/1914 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>
<p>██████ ROLANDA Femmina nato a BUDUÏA (Pordenone) il 23/7/1927</p> <p>██████</p>	<p><b>Non Censito</b></p> <p></p>
<p>██████ IRGIA Femmina nato a VIADANA (Mantova) il 13/3/1920</p> <p>██████</p>	<p>██████ IRGIA Femmina nato a VIADANA il 13/3/1920 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>
<p>██████ ANGELO Maschio nato a CASALMAGGIORE (Cremona) il 25/12/1914</p> <p>██████</p>	<p>██████ ANGELO Maschio nato a CASALMAGGIORE il 25/12/1914 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>
<p>██████ FRANCESCO Maschio nato a MARIANA MANTOVANA (Mantova) il 13/2/1923</p> <p>██████</p>	<p>██████ FRANCESCO Maschio nato a MARIANA MANTOVANA il 13/2/1923 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>
<p>██████ IRMINIA Femmina nato a RIVAROLO DEL RE ED UNITI (Cremona) il 25/8/1923</p> <p>██████</p>	<p>██████ IRMINIA Femmina nato a RIVAROLO DEL RE ED UNITI il 25/8/1923 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>
<p>██████ ANTONIO Maschio nato a SABBIONETA (Mantova) il 12/7/1909</p> <p>██████</p>	<p>██████ ANTONIO Maschio nato a SABBIONETA il 12/7/1909 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p> <p></p>

Figura 6.5 - Esempio di confronto su un questionario web

Lista Anagrafica Comunale	Lista Corrispondenti Confermati	LISTA A
<p>Operatori ▾   Variazioni anagrafiche ▾   Rapporti ▾   Sezione ▾   Edifici ▾   Questionari ▾   Confronto Censimento-Anagrafe ▾   UI/URIA ▾</p>		
<p>Confronto Questionari Web</p> <p>Codice questionario: ██████ Tipo: Famiglia</p> <p>Nominativo intestatario: ██████ Indirizzo: VIA ROANA n°2</p> <p>Il questionario risulta essere chiuso</p> <p><a href="#">Torna alla lista</a> <a href="#">Apri questionario</a></p>		
<p>██████ AZZUZ Maschio nato a MAROCCO il 1/1/1962 Cittadino/a straniero/apolide</p> <p><b>Eliminato</b></p>	<p>██████ AZZUZ Maschio nato a MAROCCO il 1/1/1962 Cittadino/a straniero/apolide</p> <p><b>Eliminato</b></p>	
<p>██████ LEYBIRA Femmina nato a MAROCCO il 01/01/1940 Cittadino/a straniero/apolide (MAROCCO)</p> <p>██████</p>	<p>██████ LEYBIRA Femmina nato a MAROCCO il 1/1/1940 Cittadino/a straniero/apolide</p> <p><b>Censito</b></p>	
<p>██████ IKRAM Femmina nato a BASSANO DEL GRAPPA il 09/10/2006 Cittadinanza Italiana</p> <p>██████</p>	<p>██████ IKRAM Femmina nato a BASSANO DEL GRAPPA il 9/10/2006 Cittadinanza Italiana</p> <p><b>Censito</b></p>	

Data l'importanza di questa funzione, all'interno del sistema Sgr era presente un'apposita voce del menu principale con diverse scelte selezionabili, come mostrato nella figura 6.6.

Figura 6.6 - Menu del confronto censimento anagrafe in Sgr



Dal Pannello di controllo si accedeva alle funzionalità di confronto vero e proprio per i questionari cartacei e web. Gli individui ai quali andava prestata maggiore attenzione, quali per esempio i possibili doppioni, erano raccolti anche in una lista unificata di nominativi da controllare in modo che gli addetti potessero più facilmente accedere a tale elenco invece di dover riaprire in lavorazione ogni singolo questionario. Altre voci del menu consentivano agli operatori di visualizzare la lista anagrafica comunale e di monitorare lo stato di avanzamento del confronto tramite un apposito rapporto riassuntivo.

### 6.3 I prodotti del confronto

Il confronto produceva in tempo reale alcuni prodotti di grande utilità per gli uffici di censimento: i bilanci quantitativi con gli esiti del confronto a livello di individui e di famiglie/convivenze e le liste individuali.

I bilanci contavano gli esiti del confronto sui questionari per i quali risultava terminata la lavorazione. Una procedura automatica aggiornava gli aggregati una volta al giorno, rendendo disponibile gli esiti del confronto di tutti i questionari lavorati e chiusi ad una certa data. Dopo la chiusura del confronto su tutti i questionari del comune, questi bilanci hanno rappresentato la certificazione del confronto svolto dal comune.

Il bilancio poteva essere visualizzato anche per singola sezione di censimento all'interno del comune oppure, se l'operatore collegato ne aveva l'autorizzazione e la visibilità territoriale, anche a livelli superiori al comune, quali la provincia, la regione o l'intero territorio italiano. In questo modo è stato possibile tenere sotto controllo in tempo reale la popolazione censita rapportata a quella presente nelle anagrafi comunali con particolare riferimento ai saldi quantitativi relativi alle persone censite e iscritte in anagrafe, alle persone censite e non ancora iscritte in anagrafe, alle persone non trovate al censimento ma iscritte in anagrafe.

La figura 6.7 illustra un esempio di bilancio per gli individui residenti.

Figura 6.7 - Esempio di bilancio degli individui residenti

Bilancio Confronto Censimento-Anagrafe / Prospetto Individui Residenti in Famiglia

Comune di IVREA, sezione 1 - Aggiornato al: 22/11/2011 19:02

	Totale	Donne	Uomini	Cittadini stranieri	Bambini in età 0-5 anni	Persone in età 20-29 anni
A - Persone censite come residenti all'8/10/2011	0	0	0	0	0	0
A1 - di cui presenti in LAC all'8.10.2011	0	0	0	0	0	0
A11 - di cui censite nel Comune ma ad un indirizzo diverso da Lac	0	0	0	0	0	0
A12 - di cui censite d'ufficio perché temporaneamente assenti	0	0	0	0	0	0
A13 - di cui censite d'ufficio per violazione dell'obbligo di risposta	0	0	0	0	0	0
B - SALDO n.1 Persone censite nel comune ma non presenti in LAC (A - A1)	0	0	0	0	0	0
C - Persone presenti in Lac all'8/10/2011	0	0	0	0	0	0
C1 - di cui censite come residenti	0	0	0	0	0	0
D - SALDO n.2 Persone iscritte in Lac e irreperibili al censimento (C-C1)	0	0	0	0	0	0
Numero totale di persone da sottoporre a revisione post-censuaria dell'anagrafe (B+D)	0	0	0	0	0	0

Oltre ai dati aggregati dei bilanci, il confronto produceva liste individuali che potevano essere visualizzate o memorizzate sul proprio Pc. Si trattava dei seguenti elenchi:

- 1) individui presenti in Lac e censiti allo stesso indirizzo dell'anagrafe;
- 2) individui presenti in Lac ma censiti ad un indirizzo diverso da quello dell'anagrafe;
- 3) individui presenti in Lac ma non censiti, quindi irreperibili;
- 4) individui non presenti in Lac ma censiti, quindi nuovi individui.

Dal confronto di queste liste con l'anagrafe locale del comune gli addetti potevano ricavare un quadro dettagliato della situazione e potevano anche accorgersi di errori ed incongruenze in tempo per effettuare correzioni su Sgr. Anche tramite l'analisi incrociata delle liste, gli operatori potevano riscontrare incoerenze sulle quali intervenire.

Nella fattispecie potevano verificarsi le seguenti circostanze:

- ▶ medesimo individuo presente sia nella lista "non censiti" sia nella lista dei "nuovi individui": tale circostanza poteva verificarsi se l'individuo si era censito in una famiglia/convivenza diversa da quella riscontrata nella Lac e l'operatore, in fase di confronto, non aveva riconosciuto il duplicato;
- ▶ medesimo individuo presente nelle liste "censiti altro indirizzo" e "nuovo individuo": tale circostanza poteva verificarsi se l'individuo si era censito contemporaneamente in due famiglie/convivenze entrambe diverse da quella riscontrata nella Lac.
- ▶ medesimo individuo presente nelle liste "censiti" e "nuovo individuo": tale circostanza poteva verificarsi se l'individuo si era censito contemporaneamente sia sul questionario previsto da Lac che uno diverso.

In tutti i casi sopra descritti non era stato correttamente riconosciuto il duplicato, ma l'esame delle liste consentiva di accorgersene. Queste incoerenze, infatti, erano da attribuire ad una erronea lavorazione dei questionari durante l'elaborazione del confronto censimento-anagrafe. In particolare durante la fase di inserimento dei nuovi individui in lista A, il sistema attivava la verifica riscontrando se l'individuo fosse già presente in Lac su un altro questionario; in questi casi l'operatore poteva confermare l'associazione o ignorare il suggerimento. Nel caso avesse confermato l'associazione e il nominativo fosse già stato censito in un altro questionario, avrebbe generato un doppione, altrimenti un nominativo da verificare. Quindi, qualora l'utente avesse mancato di effettuare tale riscontro, l'individuo sarebbe stato censito in modo erroneo e, di conseguenza, non veniva collocato nelle liste corrette.

Anche l'analisi dei bilanci permetteva di evidenziare errori ed incoerenze a livello di totali comunali costituendo un campanello di allarme per gli Uffici di censimento.

Al termine delle operazioni di confronto, infine, le liste individuali diventavano definitive e potevano essere utilizzate dagli Ucc per l'aggiornamento dell'anagrafe sulla base delle risultanze del censimento.

## CAPITOLO 7

### I sistemi per l'acquisizione, il monitoraggio e la verifica della qualità della registrazione e consultazione di dati e immagini<sup>19</sup>

#### 7.1 La fornitura di dati e immagini

Il processo di registrazione in outsourcing dei questionari cartacei del 15° Censimento Generale della Popolazione e delle Abitazioni è stato effettuato tramite lettura ottica e registrazione tradizionale. I dati registrati sono stati organizzati in più file (denominati file-tracciato) secondo tracciati record la cui descrizione è dettagliata nel manuale per la registrazione dei modelli di rilevazione<sup>20</sup> (denominato manuale di registrazione). I file-tracciato sono stati a loro volta raggruppati e compressi in file con estensione .zip, indicati come blocchi di fornitura (blocchi logici), e trasferiti dall'azienda di registrazione all'Istat tramite Vpn (Virtual Private Network), in una specifica area posta su un server presso il Ced Istat, raggiungibile dal sistema realizzato per l'acquisizione dei dati ed il monitoraggio del processo di registrazione (Samor-Pop). Di seguito si riporta la descrizione dei blocchi.

Un blocco di fornitura è definito in modo da permettere l'esecuzione di controlli di qualità su classi di caratteri registrati nelle due modalità, tradizionale e lettura ottica. Un singolo blocco logico è quindi suddiviso in due blocchi fisici, il primo relativo ai modelli sottoposti a lettura ottica e il secondo a quelli sottoposti a registrazione tradizionale.

La dimensione di un blocco logico è determinata dal numero dei modelli di famiglia (CP.1) in esso contenuti. Questi devono riguardare possibilmente interi Comuni, fino al conseguimento di una dimensione complessiva di circa 200 mila modelli. Per i grandi Comuni l'insieme dei modelli è suddiviso nel minor numero di blocchi logici, così che ciascuno di essi possa rispettare la dimensione massima prestabilita.

La fornitura delle immagini è avvenuta in modo asincrono rispetto a quella dei dati. Comunque, la consegna dei blocchi logici di dati è successiva al completamento dell'invio dei file delle immagini dei questionari di riferimento al fine di consentirne il controllo, permettendo il riscontro sulle immagini.

Le immagini (file multitiff) risultanti dal processo di scansione dei questionari cartacei sono state consegnate tramite file compressi con estensione .zip (denominati pacchetti), trasferiti tramite Vpn nel repository, predisposto dall'Istat. In particolare, ogni pacchetto si compone di:

- ▶ un file xml "batch.xml", descrittore del file zip, in cui è indicato il nome dei file multitiff contenuti nel pacchetto;
- ▶ file multitiff contenenti ciascuno le immagini relative alle pagine di un singolo questionario;
- ▶ file xml, associati a ciascun multitiff, contenenti ciascuno i metadati di un singolo questionario.

Le immagini dei questionari, infine, sono state trasferite e archiviate nel repository del sistema di gestione documentale fornito dallo stesso fornitore dei servizi di lettura ottica (denominato Documentum). Il trasferimento dei questionari nel repository di Documentum è avvenuto tramite il

---

<sup>19</sup> Il capitolo è a cura di Giancarlo Carbonetti. I paragrafi 7.1 e 7.3 sono stati redatti da Gianpiero Bianchi; i paragrafi 7.2, 7.5, 7.6 e 7.7 sono stati redatti da Giancarlo Carbonetti; il paragrafo 7.4 e i sottoparagrafi 7.6.1 e 7.6.4 sono stati redatti da Marco Broccoli; i sottoparagrafi 7.6.2 e 7.6.3 sono stati redatti da Novella Cecconi; il paragrafo 7.8 è stato redatto da Francesco Amato.

<sup>20</sup> Bianchi et al. 2014. "Il processo di acquisizione dei dati del censimento generale della popolazione e delle abitazioni del 2011: guida alla registrazione dei modelli di rilevazione". In Istat working papers, N. 2

processo di *injection*. Lo *injector* è un'applicazione *standalone*, prevista da Documentum, per processare da una directory sorgente i file zip contenenti i multitiff dei questionari. Per l'esecuzione dell'*injector* è stato predisposto un file di configurazione che descrive non solo il data-mapping degli oggetti, ma anche come questi vengano processati, la directory sorgente e di output, il numero di thread, le credenziali dell'utente Documentum, la generazione di report ed un eventuale parametro di scheduling.

## 7.2 Le fasi del controllo di qualità della registrazione

Il controllo di qualità della registrazione di dati e immagini è stato articolato in tre distinte fasi:

- ▶ fase 1: acquisizione e verifica di conformità dei dati al manuale di registrazione;
- ▶ fase 2: verifiche preliminari sulla completezza della fornitura di dati e immagini;
- ▶ fase 3: verifica dell'accuratezza dei dati e della leggibilità delle immagini tramite i controlli di qualità a campione (Cqc).

Le tre fasi sono sequenziali; il passaggio alla successiva fase di verifica richiede il superamento dei controlli previsti nella fase precedente.

Per le azioni della Fase 1 è stata sviluppata un'applicazione software generalizzata (Samor-Pop) per l'acquisizione e il controllo dei dati presenti nei "blocchi" di dati trasmessi dall'azienda di registrazione. Le principali caratteristiche di tale sistema sono le seguenti:

- ▶ è parametrizzabile per differenti tracciati record e controlli formali;
- ▶ gestisce i file di dati inviati dal fornitore (archiviazione e versioning);
- ▶ effettua i controlli previsti dal manuale di registrazione e, per i file che hanno superato tali controlli, memorizza i dati nelle tabelle della base di dati;
- ▶ fornisce le informazioni di monitoraggio di processo (informazioni sullo stato di lavorazione dei blocchi), di sintesi (statistiche di produzione) e di esito (con un dettaglio sugli eventuali errori).

Il caricamento e il controllo formale dei dati sono precedenti alla fase del controllo di qualità e alla relativa accettazione. Contestualmente al controllo dei dati, vengono eseguite la catalogazione dei file delle immagini e le relative verifiche di conformità agli standard definiti in sede contrattuale, tramite l'applicazione "Catalogo delle immagini".

## 7.3 Il sistema di acquisizione e monitoraggio della registrazione (Samor-Pop)

L'applicazione generalizzata Samor-Pop consente l'acquisizione e il controllo dei dati presenti nei blocchi di dati trasmessi dall'azienda di registrazione. Più precisamente, verifica che i file, trasferiti dall'azienda nell'apposita area di appoggio, siano conformi ai tracciati record descritti nel manuale di registrazione, fornisce i risultati dei controlli formali nella stessa area su appositi file di log in formato testo, e consente l'immissione in un database relazionale dei dati rilevati.

L'accesso a Samor-Pop è regolato tramite utenza, password e ruolo. In base ai ruoli attribuiti ad un utente, l'applicazione consente di usufruire delle funzioni previste per lo stesso. Nel caso in esame i ruoli di interesse sono:

- ▶ amministratore: che ha il compito di configurare il sistema;
- ▶ superutente: che è l'utenza per l'azienda di registrazione;
- ▶ gestore controllo di qualità: che si occupa del controllo di qualità dei dati;
- ▶ gestore del monitoraggio: che ha il compito di monitorare lo stato di lavorazione dei modelli di rilevazione.

Samor-Pop consente la definizione delle utenze del sistema e notifica le credenziali agli utenti creati. La password è generata automaticamente e deve essere obbligatoriamente cambiata al

primo accesso. L'applicazione dispone di controlli di sicurezza per gli accessi al fine di evitare intrusioni. Dopo tre tentativi falliti viene proposta una stringa di caratteri da immettere, al decimo tentativo viene disabilitato temporaneamente l'indirizzo IP del computer che tenta la connessione. L'applicazione mette a disposizione le funzioni per il caricamento delle definizioni dei tracciati record a partire da file xls, e gestisce dinamicamente le regole che consentono i controlli sui dati, pertanto può essere utilizzata per acquisire e controllare i dati provenienti da qualsiasi rilevazione. Nel caso del 15° Censimento Generale della Popolazione e delle Abitazioni, i file risultanti dal processo di registrazione dei questionari cartacei sono in formato testo a lunghezza fissa e seguono precise regole di nomenclatura. Il processo di controllo di tali file può essere attivato sia in modalità batch, ad orari prestabiliti, sia attraverso l'interfaccia web dell'applicazione. In ogni caso il processo di controllo avviene in background, consentendo l'utilizzo continuo dell'applicazione. Gli esiti del controllo possono essere interrogati consultando direttamente i file di log predisposti dal sistema oppure accedendo alle pagine dell'interfaccia web dedicate ai risultati delle elaborazioni. Le indicazioni sugli errori formali riscontrati consentono di individuare il file, i tracciati, il questionario, la riga, il campo e il tipo di errore riscontrato.

In caso di esito negativo il sistema memorizza i dati in tabelle di appoggio e segna il file zip come "controllo della registrazione negativo". Il blocco, una volta corretto, dovrà essere nuovamente sottoposto dall'azienda di registrazione a file transfer e controlli formali. In caso di esito positivo il sistema registra la data di consegna del blocco, utilizzata per monitorare i livelli di servizio previsti per l'azienda di registrazione; memorizza i dati dei questionari nelle tabelle definitive, strutturate secondo quanto definito nei tracciati record previsti dal manuale di registrazione; etichetta il file zip come "controllo di registrazione positivo" rendendolo disponibile per il successivo controllo di qualità. Le pagine di reportistica dell'interfaccia web dell'applicazione mettono a disposizione degli utenti con ruolo di amministratore o gestore del monitoraggio alcuni report contenenti le seguenti informazioni:

- ▶ numero di questionari elaborati per Comune e tipo di modello di rilevazione;
- ▶ Comuni completati per Provincia.

### **7.3.1 I controlli previsti dal sistema**

I tipi di controllo previsti dal sistema Samor-Pop sono:

- ▶ controlli di nomenclatura: il sistema verifica che siano rispettate le regole di nomenclatura definite per i file oggetto della fornitura;
- ▶ controlli formali sui dati: per ciascun campo il sistema verifica che siano rispettate le regole relative a tipo (numerico/alfanumerico), lunghezza, dominio, ecc.;
- ▶ controlli di congruenza: il sistema verifica la congruenza tra i dati e il relativo modello per ogni tipologia di modello di censimento.

Il processo di controllo dei dati, attivato dall'applicazione, dopo aver verificato la presenza nell'area di trasferimento dei file zip corrispondenti ai blocchi di fornitura, esegue per ciascuno di essi i seguenti passi:

- 1) Verifica che la denominazione del file zip rispetti la seguente regola di nomenclatura:  
`{AAAAMMGG}_{IDBLOCCO}_{TIPO}_{PROGR}_{ULTIMO}.zip`  
dove:

- AAAAMMGG rappresenta la data di creazione indicata dall'azienda di registrazione;
- IDBLOCCO rappresenta l'identificativo univoco assegnato al blocco logico;
- TIPO è riferito al tipo di registrazione: 1 se tradizionale; 2 se Ocr;

- PROGR è il numero progressivo di frazionamento del blocco fisico, assegnato nel caso in cui il blocco fisico debba essere partizionato perché supera la dimensione massima di byte consentita;
  - ULTIMO è un flag che identifica l'ultimo frammento del blocco fisico.
- 2) Verifica che il file zip non sia mai stato trattato o che non abbia già superato i controlli. Infatti se la sua versione precedente ha superato i controlli formali o addirittura ha superato il controllo di qualità non deve essere riprocessato.
  - 3) Decomprime il file zip e verifica che la denominazione dei file-tracciato estratti rispetti la seguente regola di nomenclatura:  
`{AAAAMMGG}_{RRPPPCCC}_{PROGR}_{TRK}_{COMPLETO}.txt`  
 dove:
    - AAAAMMGG rappresenta la data di creazione indicata dall'azienda di registrazione e deve essere la stessa del blocco fisico corrispondente;
    - RRPPPCCC rappresenta il codice territoriale Istat (RR = codice regione, PPP = codice provincia, CCC = codice comune);
    - PROGR è il numero progressivo di partizionamento dell'insieme dei questionari di un comune;
    - TRK è il codice del tracciato record;
    - COMPLETO è un flag di completezza che indica se tutti i questionari del comune sono stati consegnati dall'azienda di registrazione.

Al termine dei controlli di nomenclatura, il processo verifica che il tracciato sia tra quelli previsti per il corrispondente blocco e che siano presenti i file-tracciato obbligatori:

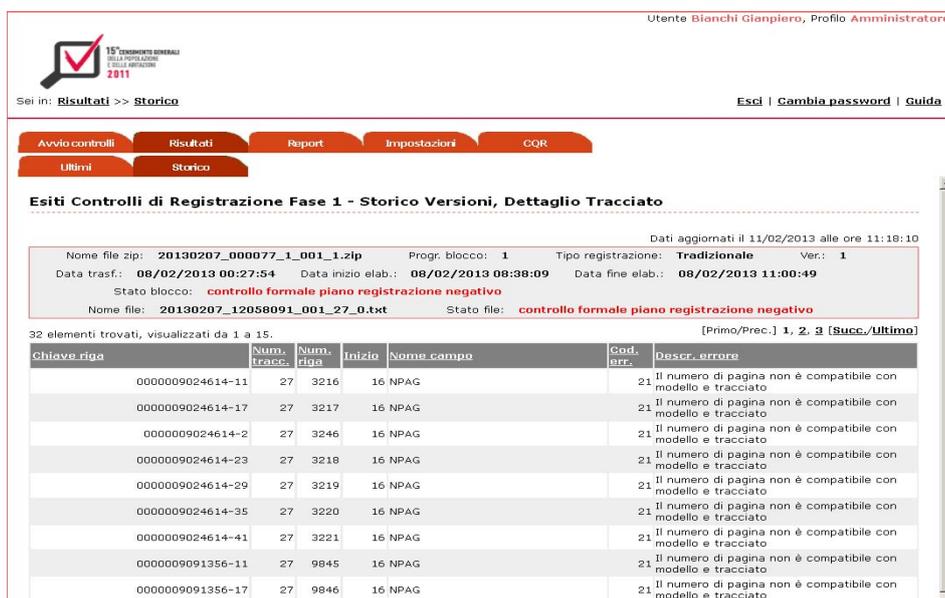
- 1) Memorizza nel database le informazioni relative al blocco dati e ne calcola la versione.
- 2) Passa il controllo alla stored procedure che estrae i dati dai file-tracciato, li appoggia nelle tabelle del database ed esegue i controlli formali e i controlli di congruenza previsti dal manuale di registrazione su tutti i record dei file-tracciato contenuti nel blocco. In caso di esito negativo di uno dei controlli, memorizza nel log le indicazioni sul tipo di errore riscontrato e sulla posizione (blocco, tracciato, riga) e passa al record successivo.
- 3) Al termine dell'elaborazione della stored procedure, se almeno uno dei controlli è fallito memorizza lo stato del blocco e dei file-tracciato come "errato"; altrimenti se l'esito dei controlli è positivo:
  - memorizza i dati contenuti nei file-tracciato nelle corrispondenti tabelle definitive, strutturate secondo quanto definito nei tracciati record previsti dal manuale di registrazione;
  - memorizza lo stato del blocco e dei file-tracciato come "controllo formale superato";
  - per ogni singolo questionario, memorizza nel registro invii il tracciato inviato ed il riferimento al blocco di appartenenza;
  - memorizza l'esito del processo nel log;
  - registra la data di consegna del blocco.

Nei file di log prodotti dall'applicazione durante la fase di controllo dei file di dati sono presenti le informazioni sugli errori riscontrati. Tali informazioni consentono di individuare il blocco, il file, il/i tracciato/i, il questionario, la riga, la colonna di inizio del campo, il nome del campo, il codice e il tipo di errore riscontrato. In particolare, il contenuto del file è organizzato come segue:

- 1) la prima riga del file contiene dati generali del blocco fisico (nome, versione, data inizio elaborazione, data di fine elaborazione, stato elaborazione);

- 2) le successive righe riportano i dati per individuare gli errori riscontrati (nome file-tracciato, chiave identificativa riga, progressivo riga, colonna inizio campo, nome campo, codice errore, descrizione errore);
- 3) Il sistema mette a disposizione due diverse modalità di consultazione dei risultati dei controlli:
- 4) consultazione con accesso diretto ai file di log prodotti da Samor-Pop nell'area di trasferimento dei blocchi dati;
- 5) consultazione con accesso tramite interfaccia web previa autenticazione (figura 7.1).

Figura 7.1 - Pagina web di Samor-Pop per la consultazione dei risultati dei controlli



Utente Bianchi Gianpiero, Profilo Amministratore

Sei in: Risultati >> Storico Esci | Cambia password | Guida

Avvio controlli | Risultati | Report | Impostazioni | CQR

Ultimi | Storico

**Esiti Controlli di Registrazione Fase 1 - Storico Versioni, Dettaglio Tracciato**

Dati aggiornati il 11/02/2013 alle ore 11:18:10

Nome file zip: 20130207\_000077\_1\_001\_1.zip Progr. blocco: 1 Tipo registrazione: Tradizionale Ver.: 1  
 Data trasf.: 08/02/2013 00:27:54 Data inizio elab.: 08/02/2013 08:38:09 Data fine elab.: 08/02/2013 11:00:49  
 Stato blocco: controllo formale piano registrazione negativo  
 Nome file: 20130207\_12058091\_001\_27\_0.txt Stato file: controllo formale piano registrazione negativo

32 elementi trovati, visualizzati da 1 a 15. [Primo/Prec.] 1, 2, 3 [Succ./Ultimo]

Chiave riga	Num. tracce	Num. riga	Inizio	Nome campo	Cod. err.	Descr. errore
0000009024614-11	27	3216	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-17	27	3217	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-2	27	3246	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-23	27	3218	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-29	27	3219	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-35	27	3220	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009024614-41	27	3221	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009091356-11	27	9845	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato
0000009091356-17	27	9846	16	NPAG	21	Il numero di pagina non è compatibile con modello e tracciato

### 7.3.2 L'architettura del sistema

L'applicazione in oggetto è una web application a tre livelli, realizzata in Java 2 Enterprise Edition, con Jakarta Tomcat 6.0 come Web Container e Oracle 11g come Rdbms. L'elenco delle principali librerie utilizzate e le relative versioni è il seguente:

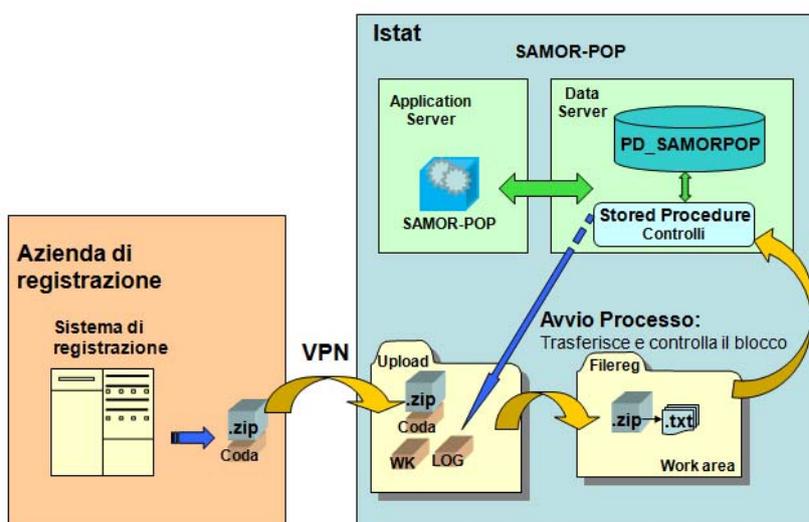
- ▶ Struts 2.1.8 (Mvc framework);
- ▶ MyBatis 3.0.6 (Java persistence framework);
- ▶ Spring 3.0.6 (per gestire l'interazione con il database);
- ▶ Jasypt 1.6 (Java simplified encryption);
- ▶ Simplecaptcha 1.1.1. (per la generazione di Captcha nella login).

Il processo di acquisizione dei blocchi di dati è avviabile sia in modalità interattiva attraverso l'interfaccia web di Samor-Pop, sia attivando un modulo software che periodicamente controlla la presenza dei file nell'area di trasferimento. Il numero di processi contemporaneamente attivi è regolato in base ad un parametro di sistema. Dopo l'avvio del processo di controllo, il blocco viene trasferito nell'area di lavoro dell'applicazione e sottoposto ai controlli formali e di coerenza della stored procedure (figura 7.2). Inoltre, per ogni blocco in lavorazione viene creato nell'area di consegna dei dati un file di presa in carico del blocco fisico, con la seguente denominazione: suffisso "WK\_" seguito dal nome del file del blocco fisico corrispondente.

Il sistema Samor-Pop accede alle seguenti directory:

- ▶ Upload: che raccoglie i file zip trasferiti dall'azienda di registrazione; quelli prodotti dall'applicazione per la presa in carico del blocco; quelli di log prodotti dall'applicazione con i risultati dei controlli;
- ▶ Filereg: che è una work area utilizzata dai processi di controllo di Samor-Pop.

Figura 7.2 - Schema del processo di acquisizione e controllo dei blocchi di dati



### 7.3.3 Il modello dei dati

Il modello della base di dati necessaria alle funzionalità dell'applicazione Samor-Pop è rappresentabile in 4 aree concettuali:

- 1) Metadati Tracciati: contenente tutte le definizioni dei tracciati record indicate nel manuale di registrazione. In particolare, le entità coinvolte sono:
  - MD\_TRACCIATI per la descrizione dei tracciati record;
  - MD\_CAMPI\_TRACCIATI per la descrizione dei campi dei tracciati record;
  - T\_STATO\_TRACCIATI per lo stato di inserimento della definizione del tracciato record;
  - T\_PARTIZIONI per l'elenco dei blocchi logici;
  - MD\_TIPI\_CONTROLLI per l'elenco dei tipi di controllo da effettuare sui record dei file di registrazione.
- 2) Gestione Acquisizione e Controllo: contenente le informazioni associate ai file-tracciato ed al loro stato di lavorazione. In particolare, le entità coinvolte sono:
  - FILEZIP\_PERVENUTI per il registro dei file zip pervenuti;
  - FILE\_PERVENUTI per il registro dei file-tracciato pervenuti e delle relative elaborazioni;
  - PACCHI per le informazioni presenti nei modelli di supporto alla rilevazione;
  - LOG\_ELABORAZIONE\_FILE per il registro delle elaborazioni di controllo dei file pervenuti;
  - SA\_TABELLA per i dati contenuti nei file-tracciato;
  - T\_JOB RUN per lo stato di elaborazione del processo di controllo;
  - COMUNI per l'elenco dei Comuni italiani;
  - PROVINCE per l'elenco delle Province italiane;

- REGIONI per l'elenco delle Regioni italiane;
  - MD\_PARAMETRI per i parametri del sistema;
  - MD\_PARAMETRI\_TIPI per la classificazione dei parametri del sistema.
- 3) Tabelle Blocco Cq: contenente le informazioni identificative del blocco logico e lo stato di lavorazione. In particolare, le entità coinvolte sono:
- BLOCCO\_CQ per le informazioni sul blocco logico;
  - VERSIONE\_BLOCCO\_CQ per la versione del blocco logico;
  - PACCHI per le informazioni presenti nei modelli di supporto alla rilevazione;
  - SA\_RECORD\_LOTTO\_CQ per i record presenti nel blocco logico;
  - STATO\_LOTTO\_CQ per lo stato di elaborazione del blocco logico;
  - T\_INDAGINE per la descrizione del tipo di indagine.
- 4) Utenti: contenente la profilatura utenti e relativi permessi. In particolare, le entità coinvolte sono:
- UTENTI per gli utenti del sistema;
  - T\_PROFILI per le classi di utenza;
  - T\_FUNZIONI per le funzioni del sistema;
  - COMUNI per l'elenco dei comuni italiani;
  - ENTI per l'ente di appartenenza degli utenti (Istat oppure azienda di registrazione con le sue eventuali sedi).

#### 7.4 Il sistema di catalogazione delle immagini

Nel precedente censimento ciascun blocco di verifica dei dati e immagini veniva consegnato dal fornitore del servizio in maniera sincrona tramite la spedizione di supporti magnetici. Il processo permetteva di circoscrivere alla singola fornitura la fase di controllo della qualità e imponeva altresì all'Istituto una più complessa gestione di conservazione off-line dei supporti fisici ed un aggravio di tempi nelle fasi di caricamento sui server locali dei Dvd consegnati.

L'architettura informatica approntata nell'ultimo censimento ha permesso di eliminare i supporti fisici mediante l'acquisizione on line delle informazioni presenti sui modelli (dati) e delle immagini output del processo di scansione. Il trasferimento dei pacchetti zip (immagini in formato multitiff ed i metadati in formato xml) è avvenuto tramite Vpn su una Nas (Network Attached Storage) dedicata di 40 terabyte.

Tale modello di trasferimento asincrono tra dati e immagini dello stesso blocco, ha così imposto di progettare un sistema di catalogazione delle immagini ad hoc. Il database del catalogo delle immagini ha quindi fornito informazioni e servizi indispensabili per tutto il processo di controllo di qualità.

Gli utilizzi principali di tale sistema di catalogazione sono:

- ▶ la verifica in tempo reale della qualità delle immagini fornite (ad esempio pagine mancanti all'interno di questionari);
- ▶ il controllo dei metadati a corredo delle immagini fornite;
- ▶ il processo di "versioning" dei file forniti nelle successive fasi di acquisizione;
- ▶ la verifica preliminare del numero di immagini ancora mancanti al momento della consegna dei blocchi di dati;
- ▶ la selezione delle immagini oggetto di verifica campionaria dei blocchi mediante il sistema "All-In-One";
- ▶ l'individuazione delle immagini per la successiva fase di controllo interattivo effettuata con l'applicazione Corinto nell'ambito della successiva fase di controllo e correzione dei dati;

- ▶ la memorizzazione delle informazioni utili per la contabilizzazione delle immagini scansionate ai fini del pagamento del servizio al fornitore;
- ▶ il punto di partenza al termine della fornitura di tutti i questionari per il processo di *injection* delle immagini nel *repository* del sistema di gestione documentale (Documentum).

Il veicolo di trasmissione è composto da file di formato zip. Al loro interno la struttura prevedeva un file di metadati in formato xml, contenente la provenienza territoriale dei questionari allegati. Per ciascun file immagine in formato multitiff, rappresentante una lista, un questionario o parte di esso, veniva sempre allegato un file di metadati in formato xml contenente le informazioni utili a corredo della fornitura (ad esempio ID\_Pacco, ID\_Distinta, Codice\_Comune, dati anagrafici del compilatore, ecc).

Il sistema prevedeva due distinte fasi di lavorazione definite come catalogazione e verifica delle immagini. La prima fase, la catalogazione dei pacchetti zip depositati in un'area di scambio, memorizzava nel database Oracle di Samor-Cqc tutte le informazioni necessarie al controllo del processo di acquisizione dell'ultima fornitura delle immagini (versioning). I principali controlli eseguiti in questa fase riguardavano:

- ▶ la presenza dell'identificativo provincia tra i metadati del file zip;
- ▶ l'effettiva presenza del file immagine indicato nella lista;
- ▶ la presenza tra i metadati dei campi chiave tra cui ID\_Questionario, Tipo\_Questionario, Tipo\_Questionario\_Short, ID\_Pacco, ID\_Quest\_Comune e ID\_Distinta;
- ▶ il controllo che la dimensione dell'immagine sia diversa da zero byte;
- ▶ la verifica formale degli identificativi dei questionari in termini di lunghezza e presenza di caratteri non numerici;
- ▶ la verifica della coerenza del numero pagine del file multitiff con il tipo di modello di rilevazione corrispondente.

La verifica dei file immagine, processo molto più oneroso in termini elaborativi, analizza la presenza di anomalie all'interno dei singoli multitiff. I principali controlli eseguiti per ogni pagina dei file sono:

- ▶ la verifica della presenza di pagine bianche;
- ▶ la verifica della risoluzione minima di 300 Dpi dell'immagine;
- ▶ la scansione dei barcode della pagina al fine di verificare la corrispondenza biunivoca tra codice questionario e singola pagina inserita nel corrispondente multitiff.

Entrambi i processi memorizzavano nella corrispondente tabella del catalogo delle immagini le informazioni di sintesi per l'individuazione dei problemi. Mediante due tabelle analitiche di log venivano memorizzate le anomalie e le segnalazioni riscontrate. Tali informazioni erano disponibili in qualsiasi momento alla verifica puntuale da parte di operatori preposti al controllo.

La schedulazione dei processi, per tutto il periodo della fornitura, ha coinvolto contemporaneamente tre server. Il primo attivava il processo di catalogazione alle ore 8 e alle ore 20. Il secondo forniva supporto al primo, eseguendo un processo suppletivo di catalogazione alle ore 3 qualora il primo non avesse terminato la lavorazione dei nuovi arrivi iniziati alle ore 20; inoltre eseguiva il processo di verifica alle ore 21 ovvero una ora dopo l'avvio del processo di catalogazione avviato nello stesso giorno dal primo server. Il terzo server, impiegato solo nei periodi di picco delle forniture, supportava i primi due, eseguendo il processo di catalogazione alle ore 12 e 24 e la verifica alle ore 3. Inoltre è stato utilizzato un cluster di dodici personal computer in rete nei periodi in cui la fornitura ha toccato i picchi massimi di trasferimento immagini.

La scelta degli orari è stata dettata da una filosofia di ottimizzazione della componente hardware nelle ore di chiusura degli uffici. Gli stessi server durante il giorno venivano usati massicciamente per le attività di controllo di qualità campionaria con la procedura “All-In-One”.

Al fine di migliorare la qualità della fornitura, quotidianamente (alle ore 7) veniva eseguita una export del contenuto delle tabelle Oracle del catalogo, che veniva depositato in una specifica area, accessibile all’azienda fornitrice del servizio, come strumento di controllo alla produzione.

#### ***7.4.1 Il sistema di controllo dell’injection sul Sistema Gestione Immagini***

La fase di caricamento, definita injection, sul Sistema Gestione Immagini è stata eseguita dall’azienda fornitrice attingendo ai file memorizzati nel repository di 40 terabyte e mediante la consultazione delle informazioni presenti nel sistema di catalogazione delle immagini.

Tale prima fase di caricamento ha impiegato circa due mesi e mezzo per il completo espletamento ed è avvenuta dopo complesse fasi di consultazione tra Istat e azienda fornitrice del servizio per le verifiche tra i contenuti presenti nel catalogo delle immagini ed i report provinciali del fornitore relativi al caricamento delle 109 province (Bolzano escluso), sul un totale di quasi 43 milioni di multitiff, recuperando gli eventuali file ancora mancanti e aggiornando per alcuni di essi i metadati mancanti.

Inizialmente si prevedeva una verifica campionaria comparativa dei multitiff a video come avvenuto con la progettazione della procedura definita All-In-One, come descritta per il controllo della qualità della registrazione dei dati. Date le difficoltà nell’operabilità di questa modalità, si sono cercate soluzioni migliori in termini di praticità ed efficienza.

Nell’informatica forense, gli algoritmi di hash, in particolare Sha1 e Md5, sono largamente utilizzati per validare e in qualche modo firmare digitalmente i dati acquisiti, tipicamente le copie forensi. La recente legislazione impone infatti una catena di custodia che permetta di preservare i reperti informatici da eventuali modifiche successive all’acquisizione: tramite i codici hash è possibile in ogni momento verificare che quanto repertato sia rimasto immutato nel tempo. Se i codici hash corrispondono, entrambe le parti in un procedimento giudiziario hanno la certezza di poter lavorare sulla stessa versione dei reperti, garantendo quindi un’uniformità di analisi e in genere di risultati. I risultati dei codici hash sono ormai calcolati di default dalla maggioranza dei software per acquisizione forense e allegati alle copie forensi salvate.

Con lo stesso principio, le versioni originali dei multitiff presenti all’interno dei file zip nel file system dei 40 terabyte associati al sistema di catalogazione delle immagini, sono stati la fonte della copia certificata presente all’interno del Sistema Gestione Immagini.

Sono state sviluppate due distinte applicazioni software. La prima che calcolava il valore del Md5 di tutti i multitiff presenti all’interno degli zip nei 40 terabyte, memorizzando il codice hash all’interno di una tabella del database di Samor-Cqc. La seconda che come parametro di input aveva il path di una cartella presente sul secondo file system a nostra disposizione di complessivi 5 terabyte, che è stato utilizzato per l’estrazione massiva dei multitiff dal Sistema Gestione Immagini tramite l’applicativo Documentum. Il software calcolava il valore del Md5 dei file presenti nella cartella specificata e lo memorizzava in corrispondenza dell’istanza relativa al nome del file trattato su un’altratabella dello stesso database.

L’aspetto innovativo del processo è l’assoluta certezza della verifica, l’impossibilità dell’errore umano, l’estrema velocità di esecuzione rispetto al controllo visivo e la possibilità di giungere alla verifica esaustiva dei quasi 43 milioni di file.

### 7.5 Le verifiche preliminari di completezza

In questa seconda fase si distinguono tre sotto-fasi parallele:

- ▶ fase 2.1: conteggi;

Il software conta il numero di modelli consegnati e caricati, per comune e per tipo modello. Il confronto con i dati provenienti da fonte esterna (Sgr) ha permesso di fare verifiche più puntuali sulla completezza della fornitura.

- ▶ fase 2.2: codifiche;

Con riferimento alle stringhe alfa-numeriche del Comune, dello Stato Estero e del Titolo di Studio, il software verifica il rispetto del livello minimo di restituzione delle relative codifiche: rispettivamente il 95 per cento, il 90 per cento e l'85 per cento delle stringhe codificabili, secondo quanto indicato nell'Allegato C al capitolato tecnico della gara d'appalto.

- ▶ fase 2.3: corrispondenza tra dati e immagini;

L'applicazione sw verifica, a fronte dei modelli consegnati e caricati, la presenza del file immagine corrispondente, segnalando quelli mancanti (dati senza immagini). Inoltre, dal sistema di catalogazione delle immagini si ha l'evidenza delle immagini non conformi e delle immagini con la presenza di almeno una pagina bianca (assenza della singola pagina per difetto di scansione o per assenza del supporto cartaceo).

### 7.6 I controlli di qualità a campione

In questa fase, da ciascuna componente fisica (lettura ottica; Registrazione Tradizionale) si è proceduto all'estrazione di un campione rappresentativo di modelli (dati e immagini) per verificare la leggibilità delle immagini e misurare i livelli di accuratezza dei dati tramite il controllo di qualità a campione (Cqc), per diverse tipologie di variabili e di modelli.

Tabella 7.1 - Schema riassuntivo dei 2 tipi di registrazioni praticati per i differenti tipi di modello di rilevazione del 15° Censimento della popolazione e delle abitazioni, e tipologie di variabili sottoposte al controllo di qualità a campione

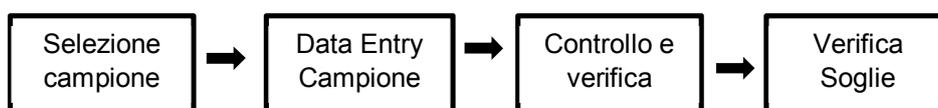
Tipo di Registrazione	Tipo Modello	Tipologia di Variabile
Lettura Ottica	CP.1	Precodificati
		Quantitativi
		Data di nascita
		Comune (codice)
		Stato Estero (codice)
	Titolo di Studio (codice)	
	CP.ED	Precodificati
		Quantitativi
Data Entry	CP.2	Battute
		Comune (codice)
		Stato Estero (codice)
		Titolo di Studio (codice)
	Liste A/B di CP.1 e CP.2	Battute

### 7.6.1 La procedura software All-In-One

Il controllo di qualità della registrazione dei modelli, nel tradizionale processo operativo avviene in quattro fasi distinte (figura 7.3):

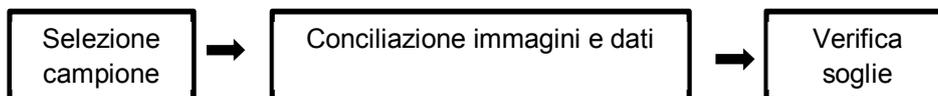
- ▶ la selezione del campione oggetto di verifica;
- ▶ la digitazione dei questionari del campione scelto;
- ▶ il controllo delle discordanze tra i due file dati dello stesso questionario;
- ▶ il conteggio degli errori riscontrati e la verifica di accettazione della soglia massima d'errore.

Figura 7.3 - Schema riassuntivo tradizionale delle fasi di controllo di qualità



La soluzione innovativa definita “All-In-One” è stata dettata dalla fornitura esaustiva della scansione digitalizzata di tutti i questionari (figura 7.4). Il workflow integra le fasi di data entry del campione e della relativa fase di controllo delle discordanze tra due set di dati in una unica fase definita come “conciliazione immagini e dati”.

Figura 7.4 - Nuovo schema riassuntivo delle fasi di controllo di qualità



La scelta è maturata dalla disponibilità del set completo delle pagine di tutti i questionari trattati in formato multitiff. Tale ristrutturazione del processo ha come principale vantaggio la totale assenza della fase di re-digitazione dei questionari oggetto di verifica, che rappresenta la fase più onerosa in termini di tempo e risorse.

I tempi di verifica ed i relativi costi sono stati ridotti in maniera sostanziale. Il personale interno Istat ha più agevolmente effettuato la fase di conciliazione mediante un confronto a video dei dati registrati dal fornitore e la digitazione dei soli errori rilevati con la corrispondente immagine scansionata.

È evidente che questa soluzione ha come presupposto l'affidabilità degli operatori addetti alla verifica; tuttavia, trattandosi di personale interno e facilmente monitorabile, si ritiene che questa soluzione abbia garantito un livello di controllo del processo nettamente superiore rispetto a delegare le stesse funzioni alla ditta fornitrice, cioè il soggetto da controllare.

Le figure professionali preposte a tale verifica sono state due: gli operatori ed i revisori. I primi hanno preso visione, per ciascun campione, di tutte le pagine afferenti ai questionari selezionati. Per le pagine ove sono state riscontrate discrepanze tra l'immagine e il dato registrato, l'operatore ha corretto nella sezione di sinistra il valore presente sull'immagine originaria (figura 7.5). I secondi avevano il compito di verificare i falsi positivi, in quanto la seconda fase è stata svolta visualizzando esclusivamente le pagine dei questionari oggetto di modifica da parte degli operatori (figura 7.6).

Il numero di utenti connessi contemporaneamente al sistema ha raggiunto le 40 unità. Il sistema estrae dal database i riferimenti della prima pagina da visualizzare per la verifica, estrae dal repository campionario l'immagine destinata all'operatore connesso e la carica nella zona sinistra

della pagina di visualizzazione. In base al tipo di modello dell'immagine, al tipo di registrazione e al numero pagina, l'applicazione carica il template corrispondente nella sezione destra. Utilizzando le tabelle di metadati del sistema e dei dati del campione, dalle tabelle presenti nel database vengono mappati gli oggetti della form.

Come nell'esempio della figura 7.6, la pagina è così composta:

- ▶ testata: Nome\_Blocco, Nome\_File\_Immagine, Nome\_Utente, Profilo\_Utente.
- ▶ zona sinistra: immagine pagina compilata;
- ▶ zona destra: form di modifica dati;
- ▶ piè di pagina: pulsanti di azione.

Figura 7.5 - Esempio di schermata a video della procedura All-In-One

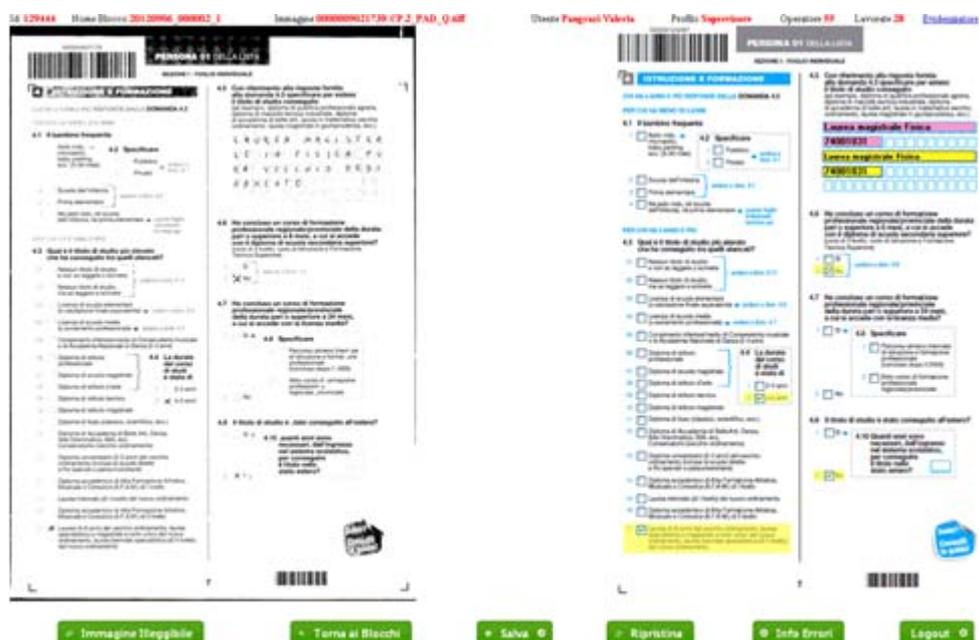
The screenshot displays two side-by-side forms for data verification and modification. The left form is a preview of a completed form, and the right form is the active editing form. Both forms contain sections for 'STATO CIVILE E MATRIMONIO' and 'CITTADINANZA'. The right form has yellow highlights on certain fields (e.g., 'FRANCIA', '999') and a red highlight on the 'FRANCIA' field, indicating data verification. A red 'Dubbio? Consulta la guida!' button is visible in the bottom right of the right form. At the bottom of the interface, there are five green buttons: 'Immagine illegibile', 'Torna ai Blocchi', 'Salva', 'Ripristina', and 'Logout'.

Per agevolare il compito di verifica agli utenti, la presenza di dati nella form è evidenziata dallo sfondo di colore giallo.

Nel caso di verifica da parte dei revisori, le modifiche eseguite in prima fase dagli operatori vengono messe in evidenza dal colore rosa. Per ciascuna modifica effettuata dal personale di verifica, il sistema memorizza un record di log all'interno del database con le seguenti informazioni: Progressivo\_Persona (solo nel caso delle liste), Codice\_Quesito, Nome\_Variabile, Valore\_Postel, Valore\_Operatore, Battute\_Errate (numero di battute errate calcolate).

Tali informazioni, mediante un complesso algoritmo di conteggio, permettono al sistema Samor-Cqc di determinare le soglie descritte più avanti.

Figura 7.6 - Esempio di schermata a video della procedura All-In-One per i revisori



### 7.6.2 Le operazioni di controllo

Come detto, il Cqr è stato articolato secondo tre fasi sequenziali; se i blocchi di consegna (dati e immagini) superavano le prime due fasi di verifica si procedeva con la Fase 3 relativa ai Controlli di Qualità a Campione (Cqc) tramite la procedura All-In-One che permetteva di verificare, su campioni rappresentativi di modelli (sia per la parte dati che di quella immagini), la leggibilità delle immagini e l'accuratezza dei dati.

Le verifiche hanno riguardato entrambe le modalità di registrazione (lettura ottica; registrazione manuale), i diversi tipi di modello e le principali tipologie di variabili.

La procedura All-In-One ha permesso ai coordinatori dell'apposito gruppo di lavoro di organizzare, gestire e monitorare la lavorazione dei campioni, estratti dai blocchi di dati, da sottoporre al controllo di qualità, la lavorazione degli stessi da parte degli operatori e la successiva fase di verifica e validazione degli errori da parte di un sotto-gruppo di revisori.

### 7.6.3 Le linee guida della procedura di controllo

Prima di procedere alla lavorazione a regime dei blocchi consegnati dall'azienda di registrazione, un primo gruppo di operatori ha effettuato un insieme di test sia per collaudare la componente software della procedura All-In-One, evidenziando eventuali anomalie e/o suggerimenti di miglioramento, sia per raccogliere e classificare le casistiche di errori più ricorrenti, utili alla stesura di un "Manuale di Linee Guida procedura ALL IN ONE"<sup>21</sup>, documento di riferimento per supportare il lavoro degli operatori preposti ai controlli tramite la procedura.

Nel "Manuale di Linee Guida procedura All-In-One" sono descritte, relativamente ai casi più ricorrenti, le regole di comportamento per rilevare e correggere gli errori, da seguire durante lo

<sup>21</sup> A cura di Novella Cecconi. Hanno collaborato alla definizione delle regole di correzione dell'errore (tramite lo studio e l'analisi di casi concreti relativi a blocchi di test consegnati da Postel) per la stesura del manuale: Claudia Bianchi, Francesco Borrelli, Fabiana Cecconi, Luana De Felici, Epifania Fiorello, Sara Giavante, Valeria Pangrazi e Marcella Pietrantonio.

svolgimento dell'operazione nei controlli sia dei dati ottenuti con la registrazione tradizionale che di quelli risultanti dalla lettura ottica.

Nel manuale si distinguono regole generali, che fanno riferimento in parte alle indicazioni per la compilazione del questionario<sup>22</sup> e in parte alle regole riportate nel manuale per la registrazione dei modelli, e regole particolari, emerse dopo vari confronti tecnici avvenuti con il fornitore al fine di condividere, per i casi più critici, quelli da giudicare come veri errori e quelli da non segnalare come tali. Infatti alcune tipologie di errore erano dovute a comportamenti sbagliati da parte dei cittadini durante la compilazione e quindi non dovevano essere imputate alla ditta come errori della fase di registrazione dei dati.

Di seguito si riportano in modo sintetico i casi e le modalità con cui si è provveduto alla correzione dell'errore.

### **7.6.3.1 La leggibilità dell'immagine**

L'immagine di una singola pagina di un generico modello di censimento si definisce leggibile, per definizione, se è chiaramente evidente la volontà espressa dal cittadino e se esiste un'associazione univoca delle informazioni espresse dal cittadino agli specifici campi/sezioni del modello, a prescindere dalla carenza di visibilità dei riquadri per le biffature. Laddove non esistono le condizioni di leggibilità, la singola pagina viene dichiarata illeggibile.

A riguardo l'operatore All-In-One poteva classificare l'immagine illeggibile tramite il software, secondo una delle seguenti tipologie:

- 1) immagine illeggibile e dati assenti;
- 2) immagine illeggibile e dati presenti;
- 3) immagine non conforme (strappata, macchiata, piegata o resa in modo tale da non permettere di leggere i dati in modo completo) e dati assenti;
- 4) immagine non conforme e dati presenti;
- 5) pagine discordanti (per esempio: immagine riportante la pagina 1 e schermata riportante i dati della pagina 2) e dati assenti;
- 6) pagine discordanti e dati presenti.

### **7.6.3.2 La verifica dei dati**

I casi e le modalità per effettuare la correzione dell'errore hanno riguardato:

#### **A. Casi generali**

- i differenti tipi di biffature;
- i quesiti dicotomici a risposta singola pre-codificata (sì/no; maschio/femmina);
- i quesiti a risposta numerica;
- le liste A/B dei CP.1 e le liste dei CP.2;
- le informazioni inserite al di fuori dei campi preposti in Ocr;
- i codici nei modelli di edificio e l'indirizzo nei CP.1 e nei CP.2.

#### **B. Casi particolari**

- le caselle annerite non totalmente nel modello lette come biffature;
- le caselle annerite nel modello lette come caratteri;
- la presenza di segni di annullamento del foglio letti come biffature;
- l'assenza di biffature nel modello riportate come dati;
- la presenza di lettere/numeri nei dati e assenti nel modello;

---

<sup>22</sup> Tali indicazioni sono presenti nella pagina 4 dei modelli di rilevazione del censimento.

- gli errori di stampa nel modello riportati come biffature;
- le biffature troppo grandi nel modello che producono altre biffature;
- i caratteri leggibili ma non comprensibili.

Come accennato, alcuni casi particolari sono stati evidenziati per ricorrenti errori di compilazione da parte di cittadini. Gli errori più frequenti commessi dai cittadini e riscontrati durante l'All-In-One sono stati:

- le caselle biffate per errore e poi annerite non totalmente;
- i caratteri ripassati su caratteri errati;
- le biffature troppo grandi.

### C. Le codifiche

In base al capitolato tecnico la ditta fornitrice era tenuta a restituire, oltre alle stringhe relative alle variabili Comune, Stato Estero e Titolo di Studio, anche i relativi codici con una percentuale minima di restituzione. Le regole di comportamento seguite per le verifiche sulla codifica delle suddette variabili sono state differenti a seconda del tipo di registrazione dei dati (Ocr; tradizionale). Nel caso di lavorazione di blocchi di dati di tipo Ocr, l'operatore All-In-One aveva il compito di controllare l'esattezza del codice restituito dalla ditta di registrazione, indipendentemente dalla correttezza o meno della stringa così come riportata nel tracciato record. L'errore è stato conteggiato solo nel caso di codice errato e non di stringa errata o assente. Nel caso della registrazione tradizionale invece, poiché i dati forniti sono stati il risultato del lavoro degli operatori addetti al Data Entry, l'errore è stato conteggiato nelle battute relative ad ogni stringa riportata o meno, sia nel caso di errore o meno del codice, sia in assenza del codice. Pertanto, nel caso di una codifica errata o assente, sono stati imputati all'azienda di registrazione sia l'errore di codice che di battuta.

Per la verifica della correttezza delle codifiche la procedura ha permesso all'operatore All-In-One di visualizzare a video, in un opportuno menù a tendina, i dizionari ufficiali Istat relativi alle classificazioni delle tre variabili fornite all'azienda di registrazione per la codifica.

Una volta conclusa la fase di individuazione e correzione degli errori dei dati di un blocco, lo stesso veniva sottoposto ad una fase di revisione per individuare ed eliminare i falsi positivi (falsi segnali di errore). Questa operazione aveva lo scopo di validare in modo definitivo il risultato del lavoro svolto dagli operatori All-In-One. Sono state sottoposte alla fase di revisione per un ulteriore controllo a video, soltanto le pagine contenenti almeno una evidenza errata; tale attività è stata assegnata ad un ristretto e più esperto gruppo di revisori. Anche per tale operazione il software è stato predisposto in modo da rendere semplice e immediata la visualizzazione di tutti i campi oggetto di revisione. Le correzioni effettuate dall'operatore All-In-One apparivano al revisore in un box di colore rosa. Inoltre, in aiuto al revisore, compariva un pulsante di "info errori" che permetteva l'apertura di una finestra nella quale era riportata la lista delle variabili rilevate errate, il valore restituito dalla ditta di registrazione, quello suggerito dall'operatore come correzione, il conteggio delle battute errate e l'eleggibilità al conteggio errori di ciascuna variabile.

In pratica, per ogni evidenza di errore, il revisore si trovava di fronte tre possibili situazioni:

- 1) dato Postel errato e dato operatore All-In-One corretto;
- 2) dato Postel errato e dato operatore All-In-One errato;
- 3) dato Postel corretto e dato operatore All-In-One errato.

Nei primi due casi si trattava di un reale errore di registrazione: il revisore non era tenuto a correggere il dato dell'operatore ma doveva solamente confermare la segnalazione di errore. Nel terzo caso si era in presenza di un falso positivo: il revisore doveva correggere l'erronea segnalazione di errore fatta dall'operatore. Di fatto, il revisore provvedeva a "bonificare" il falso segnale di errore scontandolo dal conteggio finale della percentuale di errori.

La fase di revisione, oltre a segnalare i falsi positivi per un più esatto calcolo degli errori del fornitore, ha permesso di valutare anche la qualità del lavoro svolto dagli operatori All-In-One e di intervenire tempestivamente in itinere nel caso di errori sistematici da parte degli stessi. Terminata la fase di revisione, il sistema procedeva al conteggio finale degli errori, fornendo così tutti gli elementi oggettivi per decidere in favore dell'accettazione o del rifiuto del blocco di dati sottoposto al Cqr.

#### **7.6.4 Il Sistema Samor-Cqc**

L'acronimo di Samor-Cqc deriva dalla fusione di due componenti. La prima ereditata dal sistema di acquisizione e monitoraggio della registrazione denominato Samor-Pop. La seconda che indica il controllo di qualità campionaria appunto Cqc.

La filosofia del riuso dei sistemi già in produzione ha consentito di ereditare lo schema dati del sistema di acquisizione (Samor-Pop). Essendo il campione estratto per ciascun blocco dati da verificare un sottoinsieme della popolazione universo, la scelta più razionale è stata quella di duplicare lo schema evitando così la fase di progettazione concettuale e rendendo speculari i due ambienti.

Questa scelta ci ha permesso di creare una vista delle tabelle del database di Samor-Pop all'interno di quello di Samor-Cqc al fine di popolare le tabelle del campione direttamente dal primo, riducendo la ridondanza dei dati e agevolando lo sviluppo delle procedure di upload del campione nella base dati di lavoro (Samor-Cqc).

Il sistema gestionale informativo Samor-Cqc ha permesso di coordinare tutte le fasi di lavorazione del controllo di qualità della registrazione. Dalla ricezione dei blocchi dati, solitamente provinciali, il sistema Samor-Pop ha gestito la prima fase di lavorazione relativa al caricamento e ai controlli formali di coerenza e di consistenza del blocco dati. I blocchi con esito positivo sono stati messi a disposizione delle successive fasi di lavorazione e verifica svolte da Samor-Cqc.

L'unità elementare del processo è sempre il blocco dati. Dapprima si è controllato il livello minimo di assegnazione del codice (codici e descrizioni). Dal database di Samor-Pop sono state conteggiate le valorizzazioni delle stringhe presenti nei relativi attributi del titolo di studio, provincia/comune e stato estero (denominatore del rapporto). Al tempo stesso si è conteggiata la relativa valorizzazione di quelle che sono state codificate con il corrispondente codice Istat (numeratore del rapporto).

Tali informazioni quantitative memorizzate nel sistema dovevano rispettare i livelli di accettazione definiti nel contratto. Se il rapporto era al di sotto in almeno uno dei livelli di accettazione, si rifiutava l'intero blocco senza effettuare nessuna selezione del campione. Restava a carico del supervisore del processo la possibilità di forzatura di tale controllo con la conseguente continuazione dell'intero processo di lavorazione.

In caso di esito positivo, il sistema prevedeva la funzione di selezione del campione di verifica. In base ai criteri descritti nel documento metodologico allegato al capitolato tecnico del contratto, la dimensione del campione di controllo da estrarre per ciascun blocco era determinata tenendo conto del Deff<sup>23</sup> per ciascuna delle tipologie di variabili da controllare. La numerosità campionaria veniva calcolata in base a più fattori quali: la numerosità dei modelli concernenti i sotto-blocchi delle famiglie di modelli (CP.1, CP.ED, CP.2, e CP1/2 lista A e B), la relativa soglia di accettazione, l'impostazione del Deff attribuito per ciascuna delle tipologie di controllo da effettuare, il conteggio di battute medie per tipologia di variabile e modello.

---

<sup>23</sup> Il Deff (*design effect*) è un parametro di aggiustamento della varianza campionaria che si introduce quando si adotta un disegno di campionamento differente da quello causale semplice. In questo contesto, poiché sono stati selezionati interi questionari (cluster di variabili da sottoporre a controllo) il Deff viene impiegato per ampliare la numerosità del campione finale utile al controllo statistico di qualità.

La fonte informativa in input era contenuta nel set di dati memorizzati nel database di Samor-Pop. Il processo produceva in output la duplicazione nelle corrispondenti tabelle su Samor-Cqc del set di dati relativi ai soli questionari selezionati, partendo dalle omonime tabelle definite nel database Samor-Pop per effetto dell'applicazione dell'algoritmo di selezione stocastica del campione. Per ciascuno dei questionari selezionati nel campione, mediante la lettura del sistema di catalogazione delle immagini, il sistema prevedeva la copia del file immagine multitiff su un repository di lavoro ai fini della successiva fase di conciliazione delle immagini e dei dati tramite la procedura software All-In-One già descritta nel paragrafo 7.6.1.

Al termine della fase di controllo degli operatori e della successiva fase di verifica dei revisori di tutte le immagini relative al blocco dati, il sistema prevedeva la chiusura della fase di conciliazione del blocco e il conseguente conteggio delle soglie di accettazione. Il blocco poteva essere composto da modelli relativi ai quattro gruppi di verifica (CP.1, CP.ED, CP.2 e CP1/2 lista A e B) a seconda che si trattasse di registrazione Ocr o tradizionale.

I controlli indipendenti verificati in base alla categoria delle variabili erano: CP.1-Precodificati, CP.1-Quantitativi, CP.1-DataNascita, CP.1-CodiceIdentificativo, CP.1-Comune, CP.1-TitoloStudio, CP.1-StatoEstero, CP.ED-Precodificati, CP.ED-Quantitativi, CP1/2listaA/B-Caratteri, CP.2-Caratteri, CP.2-Comune, CP.2-TitoloStudio e CP.2-StatoEstero.

Il superamento di una o più soglie d'errore per ciascun sotto-blocco portava al rifiuto dell'intero blocco dati. Nel caso di esito positivo di tutti i controlli, la procedura passava alla notifica dell'esito di accettazione del blocco, nel caso contrario veniva attivata la procedura di notifica di rifiuto. Inoltre si è ritenuto necessario, in situazioni d'emergenza, lasciare a coloro che erano preposti al controllo di qualità di poter accettare un blocco anche in condizione di violazione delle soglie di accettazione.

L'intero sistema Samor-Cqc è stato progettato dal punto di vista funzionale come "Data Driven". Tutti i processi sono stati gestiti mediante parametri e attributi definiti nel database. La gestione delle soglie d'accettazione, come la definizione del Deff per la determinazione della numerosità campionaria, la definizione delle utenze e l'associazione dati-template-immagini come metadati sono definite all'interno della base di dati, e consentono di rendere generalizzabile l'utilizzo dell'applicazione.

Il modello dei dati, come già rappresentato graficamente nella Figura 7.3, evidenzia al suo interno i metadati per il funzionamento della procedura All-In-One, la gestione delle utenze e delle fasi di lavorazione dei blocchi.

### 7.7 Elementi di innovazione e punti di forza dell'intero processo

L'intero processo di controllo della qualità della registrazione di dati e immagini del 15° Censimento Generale della popolazione e delle abitazioni è stato contraddistinto da molti elementi di innovazione, rispetto alle procedure in vigore in Istat e all'analoga esperienza percorsa in occasione del precedente censimento.

Di seguito, se ne illustrano alcuni:

- ▶ tempestività dell'acquisizione e dei controlli formali dei dati rispetto alla consegna;
- ▶ riutilizzo dello schema di dati del Db di Samor-Pop per Samor-Cqc;
- ▶ il Sistema di Catalogazione delle Immagini;
- ▶ conciliazione a video di dati e immagini, nella fase del Cqc, con la digitazione solo delle discordanze;
- ▶ memorizzazione delle informazioni utili al conteggio delle battute;
- ▶ interazione immediata del sistema Samor-Pop con il sistema di controllo e correzione dei dati (Scdd);

- ▶ possibilità di effettuare in modo tempestivo, sui blocchi di test, specifiche analisi per l'individuazione di errori sistematici nei dati.

Il buon esito finale delle operazioni di verifica è stato inoltre raggiunto grazie anche ai seguenti punti di forza del processo:

- ▶ progettazione del piano di registrazione utile ad un efficiente riutilizzo dello schema dei dati;
- ▶ generalizzazione del sistema di acquisizione dei dati;
- ▶ parallelismo dei diversi processi preposti al Cqr al fine di una efficiente lavorazione;
- ▶ schedulazione dei processi relativi alla catalogazione e alla verifica delle immagini;
- ▶ invio automatico di mail informative sulla conclusione dei processi intermedi e finali del Cqr a fini di monitoraggio;
- ▶ istruzione e formazione delle squadre di operatori proposti al Cqc.

Tutti gli elementi messi in evidenza come peculiari per il processo adottato per l'effettuazione dei controlli di qualità rappresentano un bagaglio di conoscenza ed esperienza che potrà certamente tornare utile se si dovranno effettuare altre attività per la verifica di forniture di dati e immagini provenienti da ditte esterne.

### 7.8 Corinto, un sistema software per la gestione delle immagini

Il software Corinto è una web application, accessibile da browser, per la ricerca e il recupero delle immagini dei questionari del 15° Censimento Generale della Popolazione e delle Abitazioni. Essa è stata sviluppata nell'ambito dell'attività del Controllo e Correzione del Censimento (figura 7.7).

Figura 7.7 - Visualizzazione questionario

**SCCD CORINTO**  
Gestione Immagini Popolazione 2011

Questionario: 0520220000187 [Nuova Ricerca](#)

Modello Questionario: 0520220000187-CP.1B\_6P\_Q [Stampa](#) Pagina: 1 di 32 Scala:

25%

0520220000187 SEZIONE I - NOTIZIE SU FAMIGLIA E ALLOGGIO

**I.1 Indicare il tipo di alloggio**

Abitazione

Altro tipo di alloggio (baracca, roulotte, camper, ecc.)

Alloggio presso sede diplomatica o consolare

Struttura residenziale collettiva (hotel, casa di riposo, ecc.)

**1.2 Quali è la superficie dell'abitazione?** 134

Considerare la superficie interna dell'abitazione ovvero la superficie del pavimento di tutte le stanze dell'abitazione, compresi i bagni, le cucine, i vani accessori ed esclusi i terrazzi, i balconi e le pertinenze (pomme, ad esempio, cantine, soffitte, garage e box auto)

**1.3 L'alloggio è occupato da**

Una sola famiglia

Due o più famiglie che coabitano

**1.4 Riportare le informazioni della/e altre/e famiglia/e coabitante/i**  
(codice questionario, cognome e nome dell'intestatario del Foglio di famiglia)  
(Se nell'alloggio coabitano più di quattro famiglie chiamare il numero verde gratuito 800.066.701)

Codice questionario* (famiglia/e coabitante/i)	Intestatario
	Cognome
	Nome
	Cognome

Le funzionalità specifiche di Corinto sono:

- ▶ Ricerca questionari tramite form web;
- ▶ Visualizzazione on line delle immagini del questionario;
- ▶ Download di un questionario immagine;
- ▶ Stampa della pagina in visualizzazione;
- ▶ Ricerca massiva in batch tramite form;
- ▶ Ricerca massiva in batch tramite file;
- ▶ Export dei questionari;
- ▶ Ricerca profilata per tipologia di utente;
- ▶ Monitoraggio delle ricerche svolte;
- ▶ Logging attività;
- ▶ Gestione utenti.

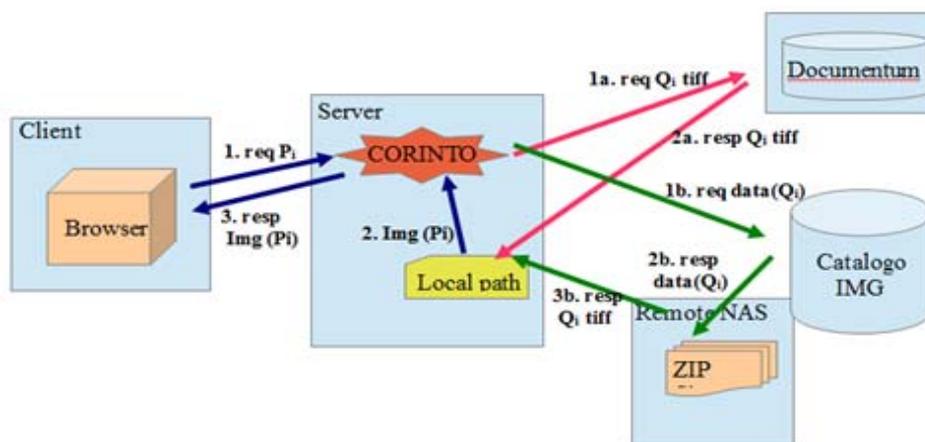
### 7.8.1 L'architettura del sistema

L'applicazione in oggetto è una web application a tre livelli, realizzata con tecnologia Java 2 Enterprise Edition, che gira su Application Server Tomcat 5 e accede a un Rdbms Oracle 11g (figura 7.8).

L'elenco delle principali librerie utilizzate e le relative versioni è il seguente:

- ▶ Struts 2.1.8 (Mvc framework);
- ▶ MyBatis 3.0.6 (Java persistence framework);
- ▶ Java Advanced Imaging (Jai) Api;
- ▶ JQuery, JQueryUI.

Figura 7.8 - Schema logico



### 7.8.2 Gestione delle immagini

Un questionario immagine è un file multitiff, il cui nome ha il formato del tipo

{id questionario}-{tipo modello}.tiff e

{id questionario}-{tipo modello AGG}-{id questionario padre}.tiff per i modelli aggiuntivi.

Nello specifico ID Questionario e Tipo Modello sono:

- ▶ ID Questionario, o Codice Questionario: è un codice numerico che può variare da 8 a 13 cifre, a seconda del modello del questionario, e identifica univocamente il singolo questionario completo;

- Tipo Modello: rappresenta le componenti del questionario per modello come riportato in tabella 7.2

Tabella 7.2 - Tabella dei tipo modello

<b>Tipo Modello</b>	<b>Descrizione</b>
CP.1_3P_Q	Sezione I e Sezione II del modello CP.1_3P
CP.1_6P_Q	Sezione I e Sezione II del modello CP.1_6P
CP.1_AGG	Modello CP.1_AGG
CP.1B_3P_Q	Sezione I e Sezione II del modello CP.1B_3P
CP.1B_6P_Q	Sezione I e Sezione II del modello CP.1B_6P
CP.1B_AGG	Modello CP.1B_AGG
CP.1_6P_SC	Modello CP.1_6P_SC
CP.2_PAD_Q	Sezione I del modello CP.1_PAD
CP.2_PAD_AGG	Modello CP.2_PAD_AGG
CP.2_NAD	Modello CP.2_NAD
CP.ED	Modello CP.ED
CP.1_3P_L	Lista del modello CP.1_3P
CP.1_6P_L	Lista del modello CP.1_6P
CP.1B_3P_L	Lista del modello CP.1B_3P
CP.1B_6P_L	Lista del modello CP.1_6P
CP.2_PAD_L	Lista del modello CP.2_PAD
CP.2_PAD_P	Prima pagina del modello CP.2_PAD

I questionari cartacei che hanno la sezione lista sono stati riprodotti in due file multitiff distinti, ad esempio per un questionario CP.1\_3P si ha il file idquestionario\_CP.1\_3P\_L.tiff, che contiene la sola sezione lista e il file idquestionario\_CP.1\_3P\_Q.tiff contenente la Sezione I e la Sezione II del questionario. Quindi un questionario, o meglio una componente del cartaceo, è univocamente identificabile dall' ID questionario (IDQuestionario) e dal tipo modello (tipoModello) riportato in tabella 7.2.

Quando dalla applicazione viene richiesto di visualizzare una pagina di un questionario o viene richiesto di effettuare il download di un intero questionario, l'applicazione deve essere in possesso del file immagine del questionario.

Il sistema gestisce il recupero delle immagini dei questionari da due repository diversi, interfacciandosi con il Cms Emc Documentum e accendendo a un filesystem remoto (figura 7.8). Attraverso due parametri di configurazione (QI\_DOCUMENTUM e QI\_FILESYSTEM), l'amministratore decide quale priorità dare al repository, infatti il sistema accede prima alla fonte con priorità maggiore e, in caso di questionario non trovato, effettua la ricerca sul secondo. In caso di non disponibilità di uno o di entrambi i repository, è possibile escluderlo nella ricerca, settando a zero il relativo parametro di configurazione. La scelta del repository è del tutto trasparente all'utente finale.

Il file recuperato viene salvato su una directory temporanea del server, da qui, a seconda della richiesta, Corinto o fa il download dell'intero file tiff sul client via Http oppure, tramite la libreria Java Jai, vi accede, cerca la pagina voluta e la visualizza a video.

Il file viene conservato in locale per velocizzarne il recupero delle pagine, evitando così ripetuti accessi in remoto. Dopo un intervallo di tempo di inutilizzo, configurabile come parametro, il file viene rimosso, ciò evita l'accumulo di grandi file inutilizzati sul server.

Per individuare la pagina di un questionario da mostrare, Corinto utilizza un algoritmo di mapping tra le pagine fisiche del questionario, il tipo modello, e la posizione della persona nel questionario.

### 7.8.2.1 Recupero file da Documentum

Documentum è il Cms (Content Management System) fornito dalla ditta appaltatrice del servizio di lettura ottica come sistema gestionale documentale per le immagini dei questionari del Censimento. I questionari sono archiviati in Documentum mediante un processo batch di injection, durante il quale viene caricato, per ogni questionario, sia il file multitiff contenente l'immagine del questionario e sia un file xml che contiene una serie di metadati che andranno a comporre il data-model rappresentativo del questionario all'interno del Cms.

Per recuperare un questionario immagine da Documentum, Corinto utilizza le Api della libreria Dfc (Documentum Foundation Classes ) che, dopo aver configurati i parametri di accesso, permettono di interagire con il web service di Documentum. Per interrogare il repository tramite DCF, si effettuano delle queries che utilizzano una sintassi tipo Sql, ad esempio:

```
String select = "SELECT r_object_id from ist_quest where ist_id_quest = '"  
+idquestionario+"' and ist_tipo_quest='"+tipoQ+"'";  
IDfQuery query = new DfQuery();  
query.setDQL(select);  
IDfCollection collection = query.execute(session, IDfQuery.DF_READ_QUERY);  
IDfId ido = collection.getId("r_object_id");  
IDfDocument docu = (IDfDocument) session.getObject(ido);
```

con la stringa select si fa una ricerca nel repository sul datamodel ist\_quest (modello dati del questionario) i cui campi sono i metadati ist\_id\_quest (per l' ID questionario) e ist\_tipo\_quest (per il tipo modello), il risultato è mappato nell'oggetto java IDfDocument. In questo modo Corinto recupera l' oggetto IDfDocument relativo alla ricerca, che contiene sia il file multitiff, che verrà salvato in una cartella locale del server, e sia i metadati del questionario che andranno a riempire il bean java CatalogoImgDBBean che verrà utilizzato nell'applicazione.

### 7.8.2.2 Recupero file da File-System remoto

Tutti i file immagini sono stati archiviati su una Nas dedicata di 40 terabyte, essi sono raggruppati in file zip, ognuno di questi file contiene il file batch.xml contenente informazioni sul pacchetto zip e, per ogni questionario contenuto, un file immagine in formato multitiff e un file xml contenente i metadati informativi del questionario. Questi pacchetti sono gli stessi utilizzati nella fase di injection per caricare i questionari su Documentum.

Con le informazioni contenute nei file di metadati e nei file batch.xml è stato realizzato su database Oracle il "Catalogo delle Immagini".

Quando su Corinto viene richiesto di recuperare un questionario su filesystem, il sistema prima accede al catalogo sul database, qui mediante idquestionario e tipoModello recupera le informazioni dei meta-dati e il nome del pacchetto zip che lo contiene nello spazio Nas. Con quest'ultima informazione accede al server remoto Nas, trova il pacchetto zip, lo apre e ricerca il file multitiff, dopo averlo trovato lo copia in una cartella locale e lo rende disponibile all'applicazione.

### 7.8.3 Utenti e funzionalità

Per l'utilizzo della web application sono state previste le seguenti classi di utenza con le relative funzionalità ammesse:

- ▶ **utente ricercatore parziale:** Ricerca delle immagini questionari con esclusione delle pagine contenenti liste A e B, lista C e prima pagina del modello CP.2; monitoraggio delle proprie ricerche richieste; accesso ai propri *log* di ricerca.
- ▶ **utente ricercatore completo:** ricerca delle immagini di tutti i questionari senza esclusioni; monitoraggio delle proprie ricerche richieste; accesso ai propri *log* di ricerca.
- ▶ **utente amministratore:** ricerca delle immagini di tutti i questionari senza esclusioni; monitoraggio di tutte le ricerche richieste; accesso a tutti i *log* di ricerca; gestione degli utenti del sistema.

#### **7.8.4 Ricerca questionari**

In Corinto si possono fare ricerche dei questionari in due modalità: tramite interfaccia web o in batch tramite file. Da interfaccia web è sufficiente riempire opportunamente i campi del form di ricerca, il sistema mostrerà i questionari trovati, quindi si può scegliere di fare l'export dei file o di visionare un questionario on line. Il download di più file viene eseguito con un processo in batch. In visualizzazione, invece, è possibile "sfogliare" il questionario scelto. In questa modalità si possono fare ricerche puntuali per codice questionario e tipo, oppure ricerche per localizzazione geografica. La modalità di ricerca da file consente di fare ricerche massive, fornendo al sistema un file di testo in formato csv. Nel tracciato del file di input è necessario indicare, per ogni questionario, l' ID questionario e il tipo modello.

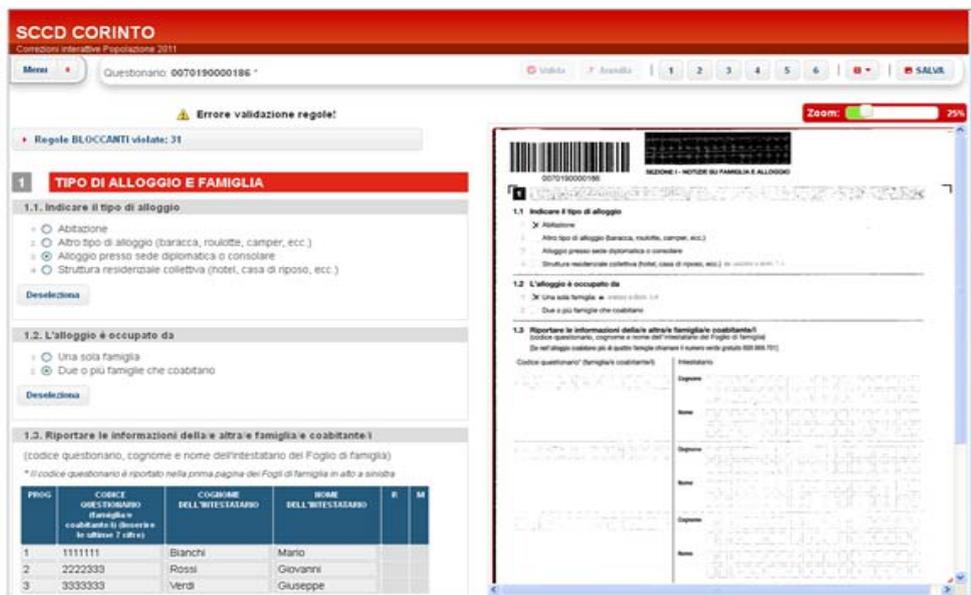
Dopo un controllo formale del file, il sistema avvia un processo in batch, che effettua la ricerca, l'estrazione e il salvataggio del file multitiff su uno spazio disco sul server dedicato all'utente. Infatti ogni utente ha accesso a un'area personale, gestibile dall'applicazione, su cui vengono memorizzati i questionari risultato delle ricerche. Durante una ricerca in batch vengono generati anche file di report in formato csv, per registrare in maniera puntuale l'esito della ricerca di ogni singolo questionario. I processi in batch richiedono un maggior tempo di esecuzione, per questo l'applicazione mette a disposizione una console di monitoraggio per controllarne l'esecuzione.

Tutte le attività di ricerca, sia da web che da file, sono tracciate in appositi file di log. Ogni utente ricercatore può accedere ai propri file di log, mentre l'utente amministratore può accedere a tutti i file di log.

#### **7.8.5 Ulteriori utilizzi**

I servizi offerti da Corinto sono stati utilizzati in varie applicazioni sviluppate ad hoc per le attività del Controllo e Correzione dei questionari, dove si è reso necessario avere, interattivamente, un immediato riscontro tra il dato compilato e il dato presente nel database (figura 7.9).

Figura 7.9 - Confronto tra dati compilati e dati presenti nel database



**SCCD CORINTO**  
Comunicazione Interattiva Popolazione 2011

Mostra Questionario: 0070190000186 \* Valida Stampa 1 2 3 4 5 6 SALVA

**Errore validazione regole!**  
Regole BLOCCANTI violate: 31

**1 TIPO DI ALLOGGIO E FAMIGLIA**

1.1. Indicare il tipo di alloggio.

- Abitazione
- Altro tipo di alloggio (baracca, roulotte, camper, ecc.)
- Alloggio presso sede diplomatica o consolare
- Struttura residenziale collettiva (hotel, casa di riposo, ecc.)

Deselezioni

1.2. L'alloggio è occupato da

- Una sola famiglia
- Due o più famiglie che coabitano

Deselezioni

1.3. Riportare le informazioni delle altre famiglie coabitanti (codice questionario, cognome e nome dell'intestatario del Foglio di famiglia)

\* Il codice questionario è riportato nella prima pagina dei Fogli di famiglia in alto a sinistra

PRIMO	CODICE QUESTIONARIO (famiglie coabitanti) (devono essere 7 cifre)	COGNOME DELL'INTESTATARIO	NOME DELL'INTESTATARIO	F.	M.
1	1111111	Blanchi	Mario		
2	2222333	Rossi	Giovanni		
3	3333333	Verdi	Giuseppe		



## CAPITOLO 8

### Il sistema di validazione dei dati<sup>24</sup>

La produzione dei dati definitivi del censimento è stata gestita grazie ad un sistema costruito ad hoc, le cui funzionalità sono state organizzate in sottosistemi:

- ▶ sottosistema di acquisizione: comprende tutti i processi necessari al caricamento dei dati provenienti dalla lettura ottica e le operazioni di caricamento dei dati introdotti attraverso l'applicazione web;
- ▶ sottosistemi di controllo e correzione: raggruppa tutte le procedure attraverso le quali si imputano i valori mancanti e/o si correggono incoerenze tra variabili;
- ▶ sottosistema di validazione dei dati: permette tutte le attività necessarie a validare l'output delle procedure di correzione nonché ad assicurare la qualità dei dati da passare ai sistemi di diffusione;
- ▶ sottosistema di ponderazione e calcolo della qualità dell'indagine : esegue il calcolo dei pesi da associare ai macrodati di diffusione su cui incidono variabili rilevate solo su base campionaria e calcola gli indicatori di qualità della rilevazione;
- ▶ sottosistemi di diffusione: sono dedicati alla predisposizione dei dati per la diffusione esterna nelle modalità previste.

In particolare, il Sistema di validazione dei dati utilizza un applicativo web avente lo scopo di gestire e monitorare il processo di validazione dei dati definitivi del Censimento della Popolazione. Mediante l'accesso al sistema è possibile sia visualizzare lo stato di avanzamento del processo secondo i vari livelli territoriali, sia gestire i passaggi di stato al superamento dei vari controlli inerenti la validazione.

Il processo di validazione viene monitorato non appena sono terminate le attività di controllo e correzione dei dati acquisiti durante la rilevazione censuaria: i dati vengono caricati sul data warehouse primario e vengono preparati per la validazione a livello comunale. Per i dati campionari è prevista un'attività parallela alla validazione (in questo caso validazione di primo livello) denominata ponderazione, mediante la quale a ciascuna variabile vengono associati pesi per il riporto all'universo. Effettuata la ponderazione e la validazione di primo livello, i comuni vengono sottoposti a una seconda validazione (la cosiddetta validazione di secondo livello o con pesi). Il superamento dei controlli per la validazione deve avvenire per tutti i livelli territoriali (comunale, provinciale, regionale, nazionale). Raggiunto il livello nazionale, si avviano le procedure per la produzione dei file necessari al caricamento dei dati su I.stat per la diffusione italiana e sulla piattaforma Census-hub per la diffusione Eurostat. Il processo di lavorazione dei dati censuari è schematizzato nella figura 8.1.

I dati, una volta acquisiti tramite lettura ottica oppure tramite il Web, sono stati sottoposti ad un complesso processo di controllo, correzione e validazione per arrivare alla loro versione definitiva; tale processo è suddiviso in diverse fasi di elaborazione. Tutta la produzione dei dati è stata gestita all'interno di un sistema costruito ad hoc le cui funzionalità sono organizzate in sottosistemi, rappresentati nella Fig. 8.1 la quale, in particolare, mostra:

- ▶ le possibilità di riciclo tra correzione e validazione;

<sup>24</sup> Il capitolo è a cura di Daniela Casale. Il paragrafo 8.1 è stato redatto da Simona Spirito; il paragrafo 8.2 è stato redatto da Francesco Bosio; il paragrafo 8.3 è stato redatto da Marco Passacantilli; il paragrafo 8.4 è stato redatto da Gianluca Flabiano.

- ▶ il ritorno dei dati in validazione dopo la ponderazione, per assicurare che anche il dato riponderato venga “certificato” dai responsabili dell’indagine.

Figura 8.1 - Flusso di lavorazione dei dati



I sottosistemi hanno potuto disporre di procedure, controlli e database propri, ma hanno condiviso tre importanti elementi trasversali a tutte le attività della produzione:

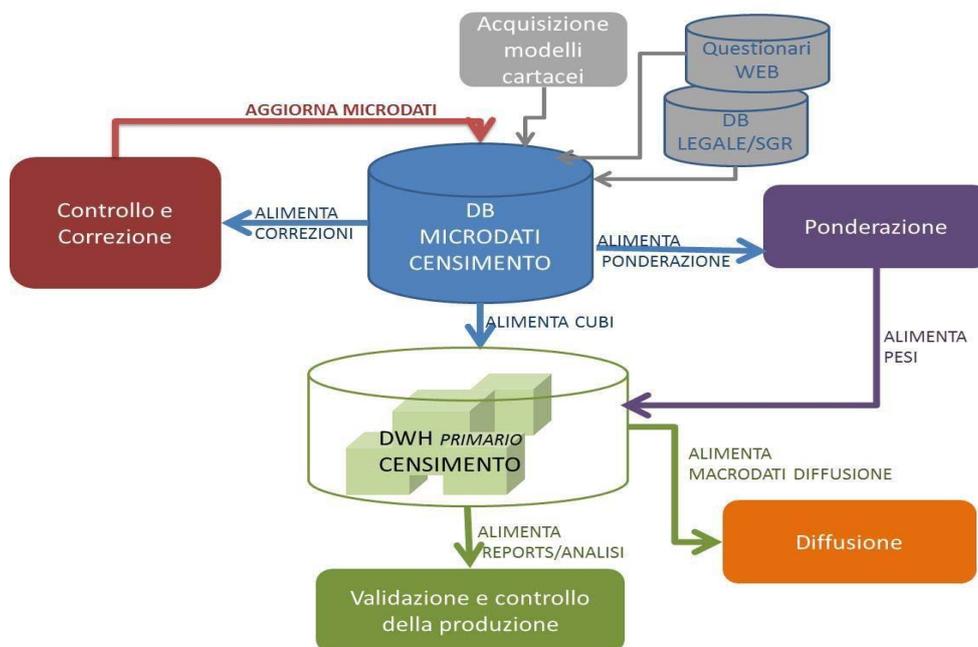
- ▶ **il database di validazione dei microdati del censimento**, cioè la base dati in cui sono stati conservati i microdati di censimento così come acquisiti e successivamente modificati dalle varie correzioni. È stato l’ambiente di riconciliazione dei risultati delle diverse procedure di modifica (correzione/ponderazione);
- ▶ **il DataWarehouse Primario (Dwp)** per l’interrogazione e la validazione, che garantiva la possibilità agli utenti statistici di eseguire analisi e costruire reports sui dati nelle differenti fasi di lavorazione, attraverso interfacce che ne consentivano l’estrazione e l’aggregazione;
- ▶ **il sistema di controllo di processo** per regolare la sequenza delle attività, in quanto i dati venivano lavorati secondo una linea di produzione che li spostava, organizzati in lotti, secondo una sequenza prestabilita, tra i diversi sottosistemi e tra database di validazione e datawarehouse per l’interrogazione. Il controllo del rispetto di tale sequenza, nonché degli scambi tra i vari database, è stato di fondamentale importanza, considerata la mole dei dati, la complessità delle attività da mettere in atto e il numero di utenti del sistema.

La validazione è stata condotta in maniera interattiva, analizzando i reports costruiti per evidenziare anomalie o anticipare le tavole di diffusione, oppure in maniera batch, vale a dire attraverso programmi che automatizzavano i controlli e distinguevano tra lotti che superavano la validazione e altri che necessitavano di maggiori controlli o una successiva validazione interattiva. Al termine del processo i dati caricati nel Dwp venivano sottoposti alle fasi finali di elaborazione a seconda della tipologia di diffusione da effettuare. La figura 8.2 evidenzia il flusso dei dati tra i principali sottosistemi e le basi dati.

Nel caso della produzione dei dati censuari quindi, il sistema di validazione si è basato su ben precise scelte tecnologiche (datawarehouse e software per analisi e reports) e su software

progettati e realizzati ad hoc per consentire all'organizzazione di eseguire passaggi di stato nei dati e verificare lo stato della lavorazione.

Figura 8.2 - Flusso di lavorazione dei dati e database



### 8.1 I requisiti statistici

Il Sistema di validazione dei dati viene definito sulla base di una serie di diagrammi di flusso in cui vengono riportati - per ogni fase del processo - i vari stati di avanzamento, i passaggi di stato, le eventuali attività in parallelo, le funzionalità per tipo di utente.

L'accesso all'applicativo è differenziato in base alle diverse tipologie di utente:

- ▶ il “responsabile di indagine” che può accedere in sola lettura e ha visibilità su tutte le attività e su tutto il territorio;
- ▶ il “supervisore” che ha visibilità su tutte le attività e su tutto il territorio, può svolgere tutte le azioni previste dall'applicazione, ad eccezione di alcune che sono direttamente gestite dal “superutente”;
- ▶ il “revisore”, che può operare solo sul territorio che gli è stato assegnato e può svolgere solo determinate azioni;
- ▶ il “superutente Eurostat”, che ha visibilità su tutte le attività e su tutto il territorio, dà la conferma sulla validità dei dati per la diffusione a Eurostat;
- ▶ il “superutente i.stat”, che ha visibilità su tutte le attività e su tutto il territorio, dà la conferma sulla validità dei dati per la diffusione nazionale.

Nell'applicativo ciascun comune viene monitorato mediante la combinazione di più variabili:

- ▶ “fase”, che indica la fase del processo di produzione. Il processo di validazione viene gestito in più fasi poiché la diffusione dei dati avviene a blocchi di variabili; pertanto ciascuna fase indica un determinato blocco di variabili da validare per poi essere diffuse;
- ▶ “sottofase”, che individua le macroattività in parallelo: lavorazione dei comuni, ok per diffusione i.stat, ok per diffusione Eurostat;

- ▶ “stato”, che indica lo stato di avanzamento del comune. Ad esempio, lo stato “10 – Controllo e correzione terminati” indica che l’attività di controllo e correzione è stata effettuata e il comune è pronto per l’attività successiva. Il passaggio da uno stato all’altro può avvenire tramite comando batch oppure mediante apposita azione da parte dell’utente che accede al sistema.

Dato che determinate attività vengono svolte a blocchi di province o regioni, sono previsti - per ogni stato - oltre al livello comunale anche un livello provinciale e un livello regionale, per indicare che tutti i comuni appartenenti a una determinata provincia o a una determinata regione abbiano terminato una specifica attività: ad esempio, se l’attività di validazione deve essere effettuata per provincia, essa può essere avviata solo nel caso in cui tutti i comuni della provincia abbiano terminato l’attività precedente.

Per le attività del processo in cui è previsto un aggiornamento del data warehouse di produzione, il passaggio di dati dal Db dei Microdati al data warehouse viene gestito tramite la predisposizione di un flag, denominato flag\_dw. Se flag\_dw è pari a 0 significa che i microdati sono pronti per il caricamento nel Dw. Il flag passa da 0 a 1 quando l’aggiornamento del data warehouse è terminato. Anche l’attività di ponderazione (ove prevista), essendo effettuata in parallelo rispetto alla vera e propria attività di validazione, viene gestita tramite flag: il flag\_wi (“wi” sta per weight in) e il flag\_wo (“wo” sta per weight out). Di default il flag\_wi risulta non valorizzato (null), viene posto pari a 0 se un’intera regione è in attesa di caricamento nell’ambiente dove si effettuerà la ponderazione, viene posto uguale a 1 quando i dati relativi all’intera regione sono stati caricati per la ponderazione. Flag\_wo è posto di default pari a 0 e assume valore 1 quando le operazioni relative alla ponderazione sono terminate.

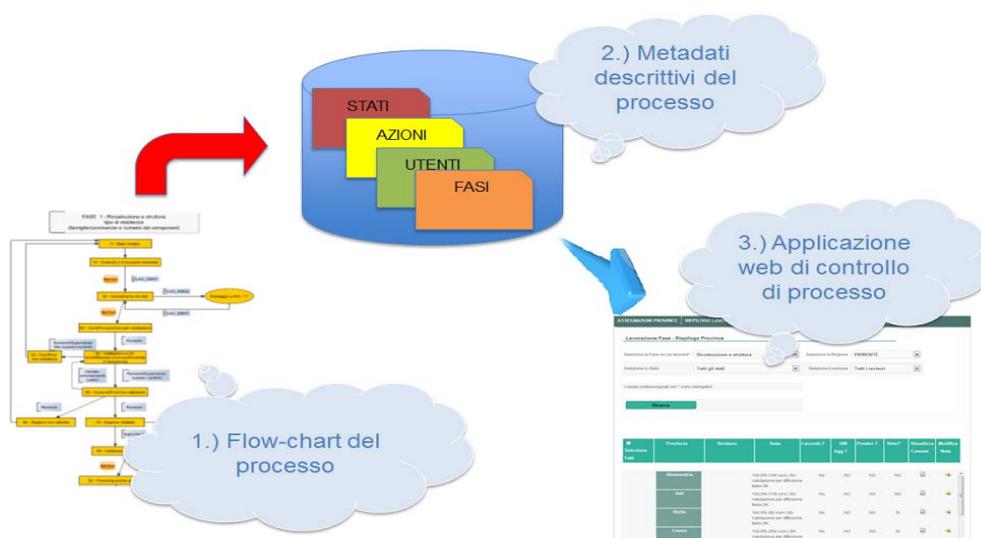
Nella figura 8.3 viene mostrato un esempio di diagramma di flusso, dove sono riportati gli stati riferiti alla fase 4 relativa alla diffusione delle “variabili long.



## 8.2 Il controllo del processo

Tutto il sistema di validazione dei dati definitivi del Censimento è basato su una metodologia di controllo del processo che è realizzata informaticamente a partire dal diagramma di flusso del processo stesso, tramite uno schema completamente parametrico della base dati ed una interfaccia utente flessibile. Il controllo del processo è illustrato nello schema nella figura 8.4.

Figura 8.4 - Il controllo di processo



Il database del sistema contiene i metadati mediante i quali viene descritto il processo di elaborazione dei dati. I principali concetti che vengono mappati con apposite tabelle nel database sono:

- ▶ le fasi di elaborazione dei dati (es. “Ricostruzione e struttura”, “Alloggi e abitazioni short”). Più fasi possono essere anche eseguite contemporaneamente, fatte salve le eventuali propedeuticità tra una fase e l'altra.
- ▶ gli stati di elaborazione (es. “Stato Iniziale”, “Caricamento dati terminato”). Una fase può trovarsi in un determinato stato, sia a livello nazionale, che a livello territoriale più basso.
- ▶ le azioni che possono essere compiute.
- ▶ gli utenti del sistema, divisi in profili (supervisore, revisore, responsabile indagine).
- ▶ gli elementi territoriali (regioni, province e comuni).

Prima della messa in produzione viene effettuato il popolamento iniziale del database, a partire dai flow-chart di ogni singola fase del processo, messi a disposizione dai responsabili statistici del progetto. Tale operazione non è stata automatizzata e deve essere fatta manualmente da parte del responsabile informatico del progetto, mediante la predisposizione di appositi script di popolamento.

Parte del sistema è l'applicazione web che opera sui dati del database appena descritto. Tramite tale applicazione, installata sulla intranet dell'Istituto, i singoli utenti possono monitorare lo stato delle fasi di cui si compone il progetto. Ogni fase attraversa una serie di stati, a partire da quello iniziale fino a quello finale, in cui si può considerare conclusa. Per definizione, una azione è un'operazione mediante la quale una fase passa da uno stato ad un altro, avanzando passo dopo passo verso lo stato finale ma, se necessario, anche tornando indietro verso uno stato precedente. Un'azione può, a seconda del caso, riferirsi all'intero territorio nazionale, ma anche ad entità territoriali minori (regioni, province, comuni).

Gli utenti del sistema effettuano le operazioni previste per il profilo a cui appartengono;

- ▶ i **supervisor** hanno visibilità su tutte le fasi e su tutto il territorio nazionale;
  - possono visualizzare i report che descrivono lo stato di lavorazione di ogni fase;
  - assegnano le province su cui possono operare i singoli revisori;
  - possono effettuare tutte le azioni loro riservate;
- ▶ i **revisori** hanno visibilità sulle province loro assegnate dai supervisor;
  - possono visualizzare i report che descrivono lo stato di lavorazione di ogni fase;
  - effettuano le azioni non riservate agli altri profili utente;
- ▶ i **responsabili dell'indagine** hanno visibilità su tutto il territorio nazionale, limitatamente alle fasi di loro competenza;
  - possono visualizzare i report che descrivono lo stato di lavorazione delle fasi di loro competenza;
  - effettuano tutte le azioni.

Da un punto di vista architetturale, l'applicazione è stata realizzata utilizzando Oracle come Rdbms. Per la realizzazione del client Web è stato utilizzato un ambiente Java/JSP, con l'utilizzo di Java Script. Si tratta di un client leggero, in quanto la parte principale dell'elaborazione dei dati avviene lato server, con un notevole uso di viste di Db e procedure Pl/Sql.

Il sistema realizzato si è rivelato molto flessibile e riutilizzabile in più contesti sia di tipo demografico sia di tipo economico. Come è stato accennato, il popolamento iniziale del database non è stato automatizzato e questo è indubbiamente un limite, in quanto tale fase richiede l'intervento di uno specialista informatico che conosca la struttura del Db. Una delle possibili evoluzioni del sistema è senz'altro quella di creare uno strumento che consenta anche agli utenti statistici la modellazione iniziale del processo all'interno della base dati.

### 8.3 L'interfaccia utente

Tramite l'interfaccia utente è possibile monitorare l'esecuzione delle varie fasi e, per ciascuna di queste, la percentuale di comuni che si trovano in quel momento nei vari stati previsti. Tale monitoraggio può essere eseguito da varie maschere che l'applicativo web mette a disposizione, le quali forniscono una visione dello stato di avanzamento del processo secondo i previsti livelli di dettaglio territoriale. Le maschere presentano diverse modalità di accesso (non accessibile, di sola visualizzazione oppure in lettura/scrittura) in base all'utente connesso, in modo da adeguarsi alle tipologie di utenze.

La maschera principale del programma è costituita da un riepilogo, accessibile unicamente agli utenti di tipo "Supervisore", "Responsabile Indagine" e "Amministratore", il quale visualizza per ogni regione le percentuali di comuni che si trovano nei vari stati all'interno di ogni specifica fase. Tale prospetto si presenta con una forma tabellare simile a quella riportata nella tabella 8.1.

Da questo è possibile poi accedere ad altre maschere di lavorazione per provincia e (successivamente) per comune, le quali non solo forniscono il dettaglio provinciale o comunale delle informazioni del riepilogo per una determinata regione e fase, ma permettono anche ad alcune tipologie di utenti (ad esempio "Amministratore" e "Revisore") di far avanzare di stato province o comuni selezionati qualora siano rispettate le condizioni previste. Viene anche offerta la possibilità agli utenti "Supervisore" e "Revisore" di editare le note da associare ad un comune o ad una provincia, le quali restano accessibili in sola lettura anche a tutti gli altri operatori. In generale, seppur riferite a dettagli territoriali diversi, le due maschere presentano layout molto simili, come si evince anche dalle seguenti schermate. Nella figura 8.5 compare un esempio di maschera di "validazione delle province".

Tabella 8.1 - Esempio di riepilogo delle fasi del processo di validazione

Regione	Fase 1	Fase 2	....	Fase n
Piemonte	10 % stato 10, 20 % stato 20, 70 % stato 30	20 % stato 10, 50 % stato 20, 30 % stato 30	20 % stato 10, 50 % stato 20, 30 % stato 30	10 % stato 10, 81 % stato 20, 9 % stato 30
Valle d'Aosta	20 % stato 10, 50 % stato 20, 30 % stato 30	10 % stato 10, 81 % stato 20, 9 % stato 30	12 % stato 10, 79 % stato 20, 9 % stato 30	10 % stato 10, 20 % stato 20, 70 % stato 30
Lombardia	10 % stato 10, 81 % stato 20, 9 % stato 30	10 % stato 10, 20 % stato 20, 70 % stato 30	12 % stato 10, 79 % stato 20, 9 % stato 30	20 % stato 10, 50 % stato 20, 30 % stato 30
...	...	...	....	...

Figura 8.5 - Esempio di maschera di validazione delle province

Validazione Comuni Popolazione Legale

Fase "Controllo struttura" - Validazione province

Regione: Lazio Revisore: Visualizza

Stato:

Selezione tutti Deselezione tutti Note

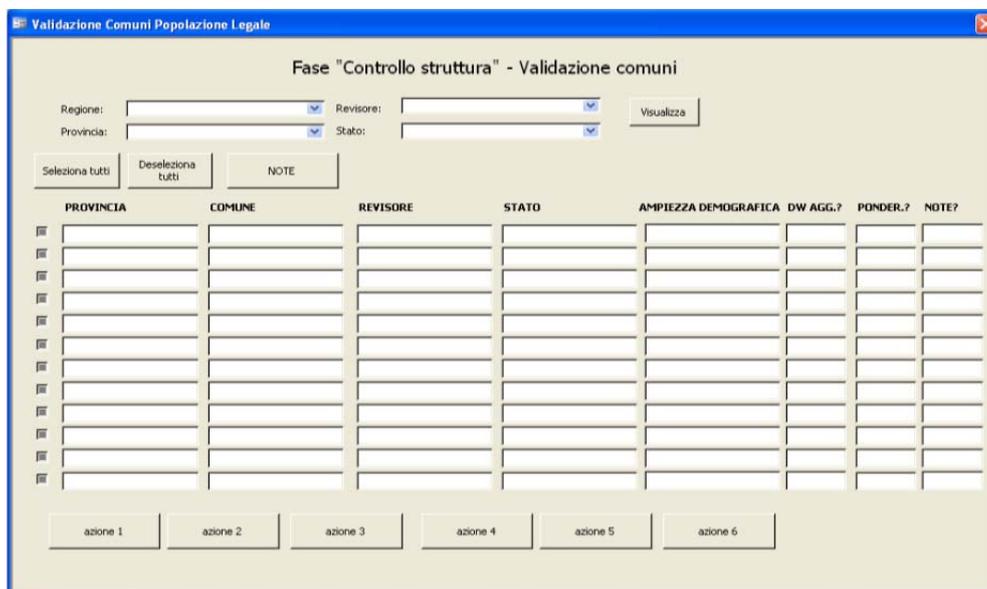
PROVINCIA	REVISORE	STATO	DW AGG.?	PONDER.?	NOTE?
Roma	Spirito	10% stato 10, 90 % stato 20	NO	NO	SI Com.
Latina	Spirito	Non ancora in lavorazione	SI	NO	NO Com.
Frosinone	Spirito	10% stato 10, 90 % stato 20	NO	SI	NO Com.
Viterbo	Spirito	Non ancora in lavorazione	SI	NO	NO Com.
Rieti	Spirito	Stato 10	NO	SI	SI Com.

azione 1 azione 2 azione 3 azione 4 azione 5 azione 6

Nella figura 8.6 compare, invece, un esempio di maschera di "validazione dei comuni".

Una volta conclusa la "Lavorazione province/comuni", il software mette a disposizione un'altra coppia di maschere a cui può accedere il "Supervisore" per far transitare il flusso di esecuzione ad una delle due sotto fasi successive: "OK per diffusione Eurostat" oppure "OK per diffusione Italia". In questo modo l'operatore abilitato può di fatto sancire la validazione finale dei dati lavorati.

Figura 8.6 - Esempio di maschera di "validazione dei comuni"



Validazione Comuni Popolazione Legale

Fase "Controllo struttura" - Validazione comuni

Regione:  Revisore:  Visualizza

Provincia:  Stato:

Selezione tutti Deselezione tutti NOTE

	PROVINCIA	COMUNE	REVISORE	STATO	AMPIEZZA DEMOGRAFICA	DW AGG.?	PONDER.?	NOTE?
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								
<input type="checkbox"/>								

azione 1 azione 2 azione 3 azione 4 azione 5 azione 6

#### 8.4 L'utilizzo del sistema di gestione delle immagini

Il sistema di gestione delle immagini Documentum (Sgi), già illustrato nel capitolo 7, è un sistema hardware e software che permette di richiamare ed estrarre le immagini in formato "multitiff", relative alla digitalizzazione dei modelli di questionario cartacei compilati dai cittadini in occasione del 15° censimento della popolazione e delle abitazioni, per un totale di 42.988.150 file immagine archiviati e disponibili alla consultazione.

Il sistema Sgi consente di ricercare i modelli di questionario in base a vari criteri di ricerca, tra i quali il numero identificativo di questionario, la Regione, la Provincia o il Comune di appartenenza, l'intestatario, la tipologia di questionario. La figura 8.7 mostra la schermata principale per la ricerca dei questionari.

Il sistema Sgi viene utilizzato dal personale Istat idoneamente formato per svolgere molteplici tipi di verifiche, tra le quali si possono annoverare:

- ▶ controlli di coerenza sui dati;
- ▶ le verifiche in fase di determinazione della popolazione legale;
- ▶ le verifiche in caso di sanzioni o contestazione delle sanzioni;
- ▶ i controlli sugli edifici censiti;
- ▶ le analisi su specifici questionari di famiglia o di convivenza, nei casi dubbi.

Il sistema ha supportato la correzione manuale interattiva dei dati nei casi più complessi e ambigui nei quali non è stato possibile applicare alcun algoritmo automatico.

Figura 8.7 - Homepage del sistema di gestione immagini

The screenshot shows the 'Sistema Gestione Immagini' interface on the Istat.it website. The browser address bar displays the URL: <http://postel-as-01.7003/taskspace/taskspace/library/ultraframework/roleSelector.jsp>. The user is identified as 'Gianluca Fiabiano Ruolo: istat\_national\_search'. The page includes a navigation menu with options: 'Ricerca Per Questionario', 'Ricerca Per Pagine', 'Ricerca Da File', 'Ricerche In Corso', 'Aggiorna Sezione Censimento', and 'Apri elementi'. The main content area is titled 'Dati Questionario' and contains several input fields: 'Codice Questionario', 'Codice Distinta', 'Codice blocco di fornitura ad ISTAT', 'Modalità di registrazione del questionario', 'Sezione Censimento', 'Tipo Questionario', 'Sezione', 'Codice Regione', 'Codice Provincia', 'Sigla Provincia', 'Codice Comune', and 'CAP'. There are also four 'Seleziona' buttons and an 'Annulla' button. At the bottom, there are radio buttons for 'Flag sloveno', 'Flag terremotati', 'Flag censimento linguistico', and 'Flag Testo password'. A 'Questionario Aggiuntivo' dropdown menu is visible, showing options like 'Lista A e B' and 'Prima pagina'. The 'Risultati:' section is currently empty.

## CAPITOLO 9

### Il sistema di rendicontazione delle spese censuarie<sup>25</sup>

Il Sistema di Rendicontazione delle Spese censuarie (Srs) è nato con l'obiettivo di permettere agli Uffici Comunali (Ucc) e Provinciali di Censimento (Upc) di adempiere all'obbligo della rendicontazione delle spese per risorse umane e logistiche connesse alle attività censuarie, secondo le disposizioni comunicate attraverso le circolari n.14 prot.7743 del 14 ottobre 2011 e n.14 bis prot. 7743 del 23 novembre 2011, e di fornire ai monitori Istat gli strumenti necessari per la verifica e il supporto alla compilazione del modello e il monitoraggio delle attività di rendicontazione.

Per adempiere all'obbligo della rendicontazione, gli Uffici Comunali e Provinciali di Censimento sono stati tenuti, come comunicato nell'informativa del 26 settembre 2012, a compilare il modello di rendicontazione delle spese censuarie esclusivamente con l'ausilio del Sistema predisposto dall'Istat (Srs).

Il modello ha previsto la compilazione con riferimento alle spese sostenute a partire dall'inizio delle attività censuarie, comprese quelle eventualmente sostenute per la costituzione dell'Ufficio di censimento, fino a conclusione delle attività censuarie e comunque non oltre il 31 dicembre 2012; le eventuali spese previste successivamente a tale data sono state indicate in modo presuntivo.

In riferimento all'erogazione del saldo del contributo variabile, il Sistema è stato arricchito di un'ulteriore funzionalità per permettere ai Comuni di scaricare i riepiloghi delle informazioni contenute su Sgr, per ciascuno dei loro operatori.

Il Sistema è stato pubblicato su web all'indirizzo internet

<https://censimentopopolazione.istat.it/rendicontazione> il giorno 1 settembre 2012.

Il sito del Sistema della rendicontazione può essere suddiviso in tre parti principali: il portale, il modello on line e l'area di monitoraggio. Il modello on line e l'area di monitoraggio sono stati progettati e scritti in Php, Javascript e MySql, con particolare attenzione a realizzare un prodotto di facile fruizione e compatibile con i browser più diffusi (Internet Explorer 7+, Firefox, Chrome, Safari, Opera, ecc.). Per la realizzazione del portale di accesso al sistema è stato invece utilizzato il Cms Typo3, sempre in Php e MySql.

#### 9.1 Modalità di accesso e di utilizzo del sistema

Gli utenti, al momento dell'autenticazione, in caso di primo accesso possono utilizzare inizialmente le credenziali provvisorie allegate alla lettera informativa e sono forzati a creare una nuova password valida per i successivi accessi. L'utente, una volta autenticato, accede al portale dal quale può disporre di una lista di funzioni (figura 9.1) dipendenti dal tipo di profilo a cui fa riferimento (Ucc, Upc o monitore Istat):

- ▶ accesso alla documentazione, quali circolari e informative;
- ▶ accesso alle guide e agli ausili alla compilazione, compresi i fac-simile dei modelli;
- ▶ accesso al modello elettronico di rendicontazione (solo uffici di censimento);
- ▶ accesso all'area di monitoraggio (solo monitori Istat);
- ▶ consultazione delle Faq (risposte alle domande più frequenti);
- ▶ accesso ai riepiloghi per operatore Sgr (vedi *Figura 1*);
- ▶ consultazione dei contatti Istat per il supporto diretto.

<sup>25</sup> Il capitolo è a cura di Simonetta Feliciotti. I paragrafi 9.1, 9.2 e 9.3 sono stati redatti da Emanuele Bolle; il paragrafo 9.4 è stato redatto da Simona Spirito.

Il menù del portale e i suoi contenuti testuali sono stati resi disponibili per tutti i profili utente sia in italiano che in tedesco.

Figura 9.1 - Scarico riepilogo operatori, Portale Srs



## 9.2 La struttura del modello elettronico

Il modello e tutte le funzioni di compilazione e controllo sono state programmati in Php, Javascript e MySQL. La struttura del modello elettronico di rendicontazione delle spese censuarie riproduce il modello cartaceo previsto nelle circolari Istat n. 14 e n. 14 bis, i suoi contenuti sono dipendenti dal profilo utente di riferimento (Ucc, Upc) e sono organizzati in sezioni separate. Indipendentemente dalla sequenza con cui sono numerate le sezioni, il responsabile dell'ufficio di censimento può decidere autonomamente l'ordine di compilazione ed effettuare anche inserimenti parziali delle informazioni, rispettando però i controlli richiesti dal Sistema in ciascuna sezione al momento del salvataggio (controlli di tipo formale, controlli di obbligatorietà e correttezza di alcune voci). Soltanto al momento della verifica finale viene effettuato un controllo sui contenuti dell'intero modello e viene richiesto il rispetto della completezza e della congruenza dei dati inseriti.

Gli errori e le incorrettezze di compilazione vengono evidenziati puntualmente dal sistema in fase di scrittura (controlli di primo livello), mentre tramite funzione specifica l'utente può, in modo più approfondito, validare il contenuto dell'intero modello (controlli di secondo livello). La validazione dei campi e la congruenza dei rapporti tra le voci compilate rilascia in caso di errore una lista puntuale di messaggi descrittivi a cui l'utente può fare riferimento per effettuare le dovute correzioni. In caso contrario l'utente che ha compilato correttamente il proprio modello può stamparlo in *pdf* con l'apposita funzione, firmarlo, protocollarlo e infine inviarlo all'Istat.

Per mantenere un controllo sul flusso di compilazione, il Sistema considera valido solo l'ultimo accesso effettuato da ciascun utente compilatore, impedendo che vengano eseguite modifiche in contemporanea sullo stesso modello e garantendo quindi all'utente compilatore l'attendibilità dei dati visualizzati. Gli utenti monitori tuttavia sono in grado di accedere contemporaneamente al modello, anche se in uso, ma esclusivamente in sola lettura. La compilazione di un modello viene inoltre interrotta dopo un tempo limite di inattività definito dal Sistema (timeout); viene in questo caso richiesto all'utente compilatore di effettuare un nuovo accesso.

Il modello riporta le seguenti informazioni: (figura 9.2):

<p><b>Notizie generali</b> <i>Uffici di censimento comunali e provinciali</i></p>	<p>Dati identificativi dell'Ufficio di censimento e del responsabile.</p>
<p><b>Numero addetti</b> <i>Uffici di censimento comunali e provinciali</i></p>	<p>Numero di addetti per ciascuna tipologia di contratto e per le singole funzioni svolte dal personale coinvolto nelle operazioni censuarie nel periodo di riferimento (vedi <i>Figura 2</i>).</p>
<p><b>Mesi uomo</b> <i>Uffici di censimento comunali</i></p>	<p>Numero di mesi/uomo lavorati per ciascuna tipologia di contratto e per le singole funzioni; ore di straordinario effettuate dai dipendenti dell'amministrazione di riferimento per lo svolgimento delle attività censuarie.</p>
<p><b>Spese di personale</b> <i>Uffici di censimento provinciali</i></p>	<p>Spese per il personale dell'Ufficio di censimento relativamente alle funzioni svolte, con indicazione delle voci di costo del lavoro (retribuzione lorda, oneri sociali obbligatori a carico del datore di lavoro, spese per la formazione, missioni e rimborsi).</p>
<p><b>Altre spese</b> <i>Uffici di censimento comunali e provinciali</i></p>	<p>Numero di mesi/uomo lavorati e relativa spesa complessiva sostenuta; spese per missioni.</p> <p>Spese sostenute per altri fattori di produzione diversi dal lavoro:</p> <ol style="list-style-type: none"> <li>1. Assistenza informatica e software;</li> <li>2. acquisizione di hardware;</li> <li>3. logistica e arredi;</li> <li>4. altre spese di gestione (materiali di consumo, utenze, spese postali e altre spese).</li> </ol>

Figura 9.2 - Numero addetti per funzione svolta e tipo di contratto, Modello on line: versione per gli uffici di censimento comunali

SEZ. 1 NOTIZIE GENERALI		SEZ. 2.1 NUMERO ADDETTI	SEZ. 2.2 MESI UOMO	SEZ. 2.3 COSTO DEL LAVORO	SEZ. 3.1 ALTRE SPESE
HOME   VERIFICA   CREA PDF   AIUTO   ESCI					
<b>SEZIONE 2.1: NUMERO DI ADDETTI PER FUNZIONE SVOLTA E TIPO DI CONTRATTO</b>					
Tipo di contratto	Funzione svolta				
	Responsabile dell'Ufficio Provinciale/Comunale di censimento	Altro Personale dell'ufficio Provinciale/Comunale di censimento	Coordinatori comunali	Rilevatori	
Dipendenti del Comune o della Prefettura	45	0	0	0	0
Addetti assunti secondo le tipologie contrattuali di cui al comma 4, art. 50 L. 122/10 (escluso i contratti di somministrazione di lavoro)	40	0	0	0	0
Somministrazione di lavoro	0	0	0	0	0
<b>Totale addetti all'Ufficio Provinciale/Comunale di censimento</b>	<b>85</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

Istat - Istituto nazionale di statistica  
Via Cesare Balbo 16 00184 - Roma tel. +39 06 46731

### 9.3 Le funzioni di controllo di invio e di supporto

Gli utenti monitori Istat accedono dal portale all'ambiente di monitoraggio del sistema. L'area è suddivisa in tre sezioni, ognuna delle quali fornisce al monitore Istat strumenti di verifica e supporto alla compilazione da parte degli uffici di censimento, di monitoraggio della rendicontazione e di scarico dei microdati.

Dalla sezione "verifica stato utente" è possibile effettuare ricerche sia per codice utente che per denominazione (figura 9.3). A ricerca completata il monitore può verificare lo stato dell'utente ed accedere ad una serie di funzioni specifiche per il supporto alla compilazione:

- 1) la riapertura/chiusura del modello;
- 2) la modifica del nome del responsabile;
- 3) il ripristino della delle credenziali provvisorie;
- 4) la modifica dell'email del responsabile, l'accesso in sola lettura al modello di rendicontazione;

Dalla sezione "Stato rendicontazione" è possibile esaminare lo stato di avanzamento della rendicontazione, corredato di rappresentazione grafica, per profilo utente in valori assoluti e in termini percentuali (figura 9.4).

Dalla sezione "Scarico dati" è possibile, attraverso l'utilizzo di filtri a più livelli (profilo, categoria, voce, ecc,...) ottenere un download personalizzato dei microdati in formato Excel (figura 9.5).

Figura 9.3 - Ricerca per utente e denominazione, monitoraggio

Ricerca per codice utente e denominazione		
CERCA:	Test	
<input type="button" value="MODIFICA EMAIL"/> <input type="button" value="RIPRISTINA PWD"/> <input type="button" value="COMPLETA STATO"/> <input type="button" value="FORZA APERTURA"/> <input type="button" value="ACCEDI MODELLO"/>		
Utente	Test	
Profilo	UCC	
Codice Provincia	000	
Codice Comune	000	
Denominazione	Test	
Descrizione tipo UCC	Singolo	
Descrizione associazione		
Nome responsabile	Mario	
Cognome responsabile	Rossi	
Nuovo nome responsabile	Ugo	
Nuovo cognome responsabile	Bianchi	
Email	rispondente@comune.it	
Codice attivazione		
Ultimo accesso	18/06/2013 15:20	
Data chiusura rendicontazione	08/04/2013 12:33	
Riepilogo stato utente	<b>Login</b>	Cambio password e attivazione completate
	<b>Rendicontazione</b>	Rendicontazione aperta, in lavorazione
User Agent	Internet Explorer 8	

Figura 9.4 - Stato di avanzamento della rendicontazione, monitoraggio

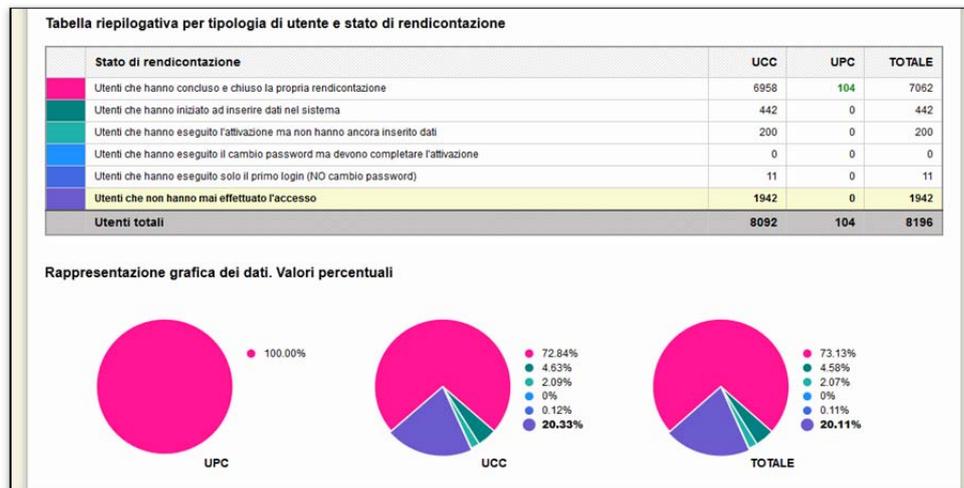


Figura 9.5 - Scarico personalizzato dei microdati, monitoraggio

**Scarico personalizzato**

Selezione profilo:  
 UCC  UPC

CATEGORIA	TIPOLOGIA DI DATO	DATI SELEZIONABILI	SELEZIONATI
INFO GENERALI	MESI X CONTRATTO	MESI C2 F2	MESI C1 F3
ADDETTI	MESI X FUNZIONE	MESI C2 F3	MESI C1 F4
TEMPI	DETTAGLIO MESI	MESI C2 F4	MESI C3 F1
COSTI	DETTAGLIO STRAORDINARI	MESI C3 F1	STRAORD F1
TOTALI		MESI C3 F2	STRAORD F2
		MESI C3 F3	
		MESI C3 F4	
		STRAORD F1	
		STRAORD F2	
		STRAORD F3	
		STRAORD F4	

Azzerare selezione  Scarica selezionati

Ultima sincronizzazione: 04/04/2014, 11:36 Sincronizza

[Guida all'utilizzo dello Scarico Personalizzato - download](#)

## 9.4 Supporto ai comuni per il calcolo del contributo variabile

La funzione “Scarico dei riepiloghi per operatore” ha lo scopo di supportare i comuni nell’attività di calcolo del contributo variabile da erogare a ciascun operatore coinvolto nella rilevazione censuaria. Lo scarico prevede il download di 4 tabelle in formato xls, riferite alle diverse attività svolte dagli operatori durante la rilevazione:

- 1) riepilogo abitazioni non occupate: per ciascun operatore è riportato il conteggio dei questionari, riferiti ad abitazioni non occupate, che sono stati inseriti nel diario;
- 2) riepilogo famiglie: per ciascun operatore è riportato il numero dei questionari di famiglia gestiti nel sistema Sgr, distinti sia per operazione effettuata e sia per canale di restituzione. Le operazioni effettuate dagli operatori riguardano: la registrazione dell’arrivo del questionario cartaceo, l’inserimento del questionario nei modelli riepilogativi, la modifica dello stato del questionario nel diario, il confronto censimento-anagrafe per ogni questionario

registrato. I possibili canali di restituzione del questionario sono: mediante rilevatore, presso il centro comunale di raccolta, presso l'ufficio postale, tramite webRiepilogo convivenze: per ciascun operatore è riportato il numero dei questionari di convivenza gestiti nel sistema Sgr, distinti sia per operazione effettuata e sia per canale di restituzione. Le operazioni e i canali sono analoghi a quelli del punto 2.

- 3) riepilogo sezioni e sottosezioni: per ciascun operatore è riportato l'ambito territoriale di competenza (sezione e sottosezione di censimento) con i relativi conteggi dei questionari di famiglia e di convivenza, dei questionari riferiti ad abitazioni non occupate e di quelli compilati di ufficio.

## CAPITOLO 10

### I sistemi di diffusione dei dati<sup>26</sup>

#### 10.1 Caratteristiche generali

Il 15° Censimento Generale della popolazione e delle abitazioni ha previsto un progetto di diffusione articolato su tre principali tecnologie: data warehousing, business intelligence e navigazione avanzata ed interattiva sul Web.

In particolare la componente di data warehousing ha previsto la realizzazione di due tipologie di data warehouse: un data warehouse primario (Dwp) a supporto della validazione delle variabili di rilevazione e di diffusione; un data warehouse per supportare la presentazione, lo scarico e la navigazione dei dati diffusi da parte dell'utenza. Il primo è stato un nuovo impianto realizzato con una specifica tecnologia mentre per il secondo si è utilizzata la medesima tecnologia adottata in Istat per il data warehouse di diffusione, *I.Stat*, opportunamente evoluta per rispondere alle specifiche esigenze censuarie.

A queste componenti si sono aggiunte, secondo gli standard e gli strumenti di diffusione dell'Istituto, la messa a disposizione dell'utenza esterna della possibilità di effettuare lo scarico dei macrodati in formato Sdmx mediante il sistema Single Exit Point (Sep) e di disporre dei file archiviati centralmente nel sistema Armida (ARchivio di MIcroDATi), accessibili presso il Laboratorio Adele da parte di ricercatori e studiosi di università, istituti o enti di ricerca.

Due ulteriori servizi sono stati, infine, realizzati in occasione della tornata censuaria: il portale dei Linked Open Data per la diffusione dei dati per sezione di censimento; la messa a disposizione dei dati italiani nell'ambito del sistema statistico Europeo, collegato al sistema Census Hub di Eurostat tramite una istanza dedicata del Single Exit Point.

In figura 10.1 sono illustrati il flusso dei dati e i diversi ambienti di trattamento dati e tecnologici:

- ▶ Oltp (on line Transaction processing): La lista dei nominativi della Popolazione Legale ha definito la popolazione di riferimento utilizzata per la selezione dei dati dei questionari dai repository di SAMORPOP e Sgr. Il trattamento dei dati è di tipo Oltp e la tecnologia Rdbms è Oracle;
- ▶ Riconciliazione e Staging Area: i dati sono stati elaborati dal Sistema di Controllo e Correzione Dati (Sccd) che ne ha operato la riconciliazione oltre che la correzione e imputazione delle mancate risposte. La Staging Area ha previsto l'organizzazione dei dati riconciliati in due principali databases: microdati validati e dati di benchmark da utilizzare per la validazione. L'ambiente tecnologico è Oracle;

---

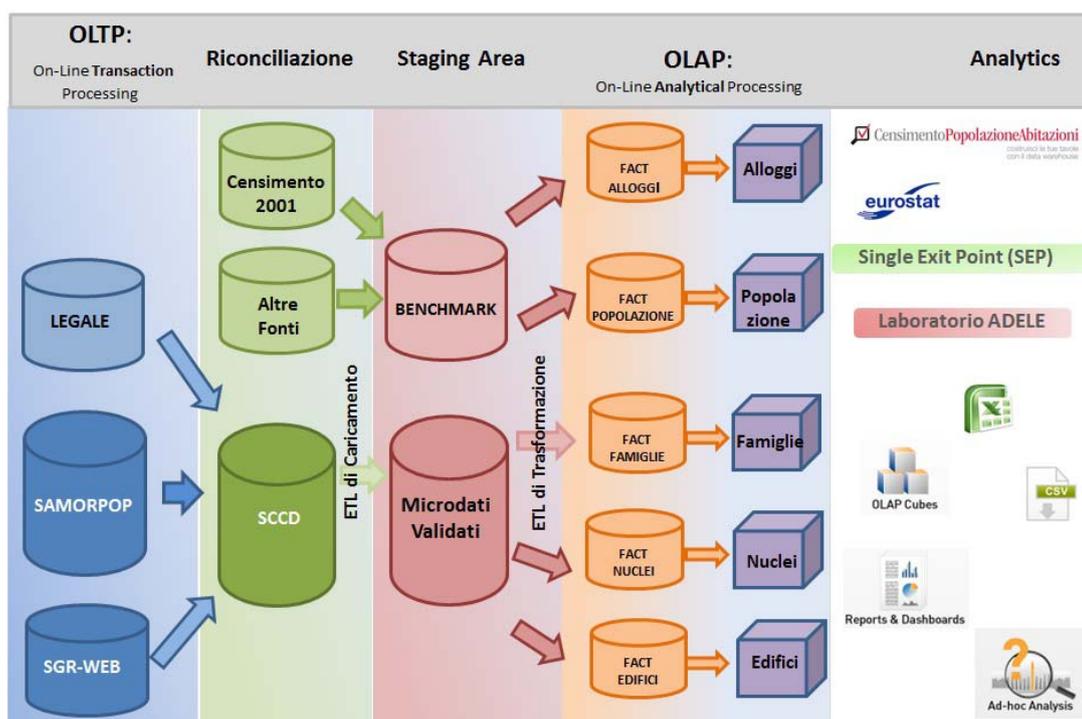
<sup>26</sup> Il capitolo è a cura di Monica Scannapieco, Stefania Bergamasco e Antonino Virgillito. Il paragrafo 10.1 è stato redatto da Monica Scannapieco e Stefania Bergamasco; il paragrafo 10.2 ed il sottoparagrafo 10.4.3 sono stati redatti da Roberta Radini; i sottoparagrafi 10.2.1 e 10.3.3 sono stati redatti da Luca Valentino; il sottoparagrafo 10.3.1 è stato redatto da Mariano Crea; il sottoparagrafo 10.3.2 è stato redatto da Pasquale Testa; i sottoparagrafi 10.4.1 e 10.4.2 sono stati redatti da Annunziata Fiore; il sottoparagrafo 10.4.4 è stato redatto da Gianluca D'Amato; il paragrafo 10.5 ed il sottoparagrafo 10.5.1 sono stati redatti da Stefania Bergamasco; i sottoparagrafi 10.5.2, 10.5.3, 10.5.4 e 10.5.5 sono stati redatti da Enrica Massi; il paragrafo 10.6 è stato redatto da Domenico Scalzo; il sottoparagrafo 10.6.1 è stato redatto da Domenico Scalzo e Raffaella Aracri; il sottoparagrafo 10.6.2 è stato redatto da Domenico Scalzo e Andrea Santilli; il sottoparagrafo 10.6.3 è stato redatto da Rossana Vignola e Daniele Ingretolli, il paragrafo 10.7 ed il sottoparagrafo 10.7.1 sono stati redatti da Alessio Cardacio, il sottoparagrafo 10.7.2 è stato redatto da "Mariangela Verrascina", il paragrafo 10.8 ed il sottoparagrafo 10.8.1 sono stati redatti da Mara Cammarota; il sottoparagrafo 10.8.2 è stato redatto da Marco Caputi; il sottoparagrafo 10.8.3 è stato redatto da Florinda Damiani; il sottoparagrafo 10.9.1 è stato redatto da Stefano De Francisci e Monica Scannapieco; il sottoparagrafo 10.9.2 è stato redatto da Laura Tosco; il sottoparagrafo 10.9.3 è stato redatto da Raffaella Aracri, Andrea Pagano, Paolo Pizzo e Luca Valentino.

- ▶ **Olap (on line Analytical processing):** i dati sono stati organizzati secondo un modello multidimensionale, con creazione di fact table tra loro correlate (Alloggi, Popolazione, Famiglie, Nuclei, Edifici). A partire dalle fact table, sono stati generati un insieme di cubi Olap. Gli ambienti tecnologici di questa fase sono stati Oracle e Microsoft Business Intelligence;
- ▶ **Analytics:** la fase di analytics ha avuto come obiettivo principale la validazione degli aggregati accessibili a partire dai cubi.

Di seguito un elenco dei principali benefici derivanti da un approccio basato su data warehousing per l'analisi e la diffusione di dati statistici:

- ▶ il data warehouse abilita la centralizzazione dei dati, in questo modo può garantire un'alimentazione consistente di diversi canali di diffusione;
- ▶ in virtù della possibilità di interrogare i dati sulla base di un modello dati, quello multidimensionale appunto, l'alimentazione dei canali di diffusione è facile: in altri termini, la dinamicità delle interrogazione al data warehouse consente un'elevata flessibilità nel reperimento dei dati per i canali progettati e nell'aggiunta di nuovi canali. Da un punto di vista applicativo è possibile utilizzare strumenti dedicati di Etl per il reperimento dei dati (figura 10.1);
- ▶ l'esistenza di un unico data warehouse che supporta il processo di produzione non richiede un'organizzazione dei dati ad-hoc per alimentare il sistema di diffusione I.Stat. In particolare, una volta allineato il data warehouse con il sistema di produzione (e.g. con riferimento alle classificazioni), lo stesso sistema può supportare sia le fasi di processo relative all'analisi dei dati che l'alimentazione di I.Stat, utilizzando classificazioni diverse per le varie fasi. Questo è esattamente l'utilizzo del data warehouse primario che è stato fatto per il censimento.

Figura 10.1 - Flusso dati e ambienti tecnologici



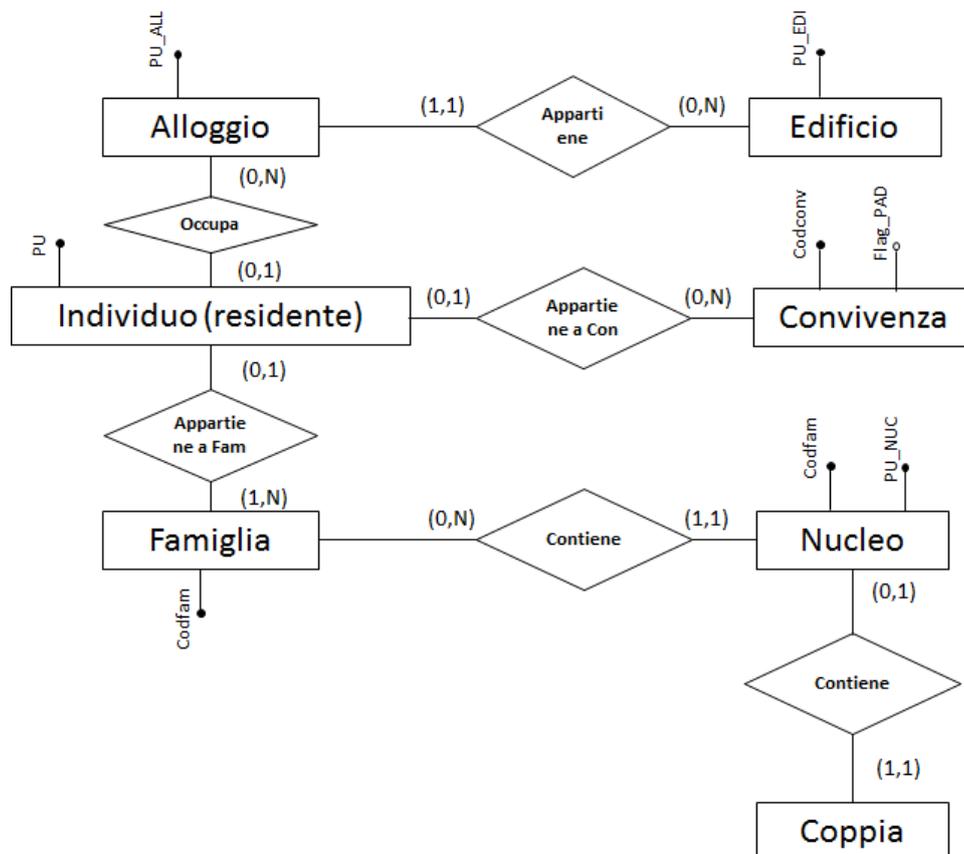
## 10.2 Architettura dei microdati

Nel data base dei microdati sono confluiti i dati provenienti dal “Sistema di Controllo e Correzione Dati”, ossia Sccd, i dati delle fonti di confronto utilizzate dal processo di validazione, la raccolta dei pesi applicati ad ogni singolo record distinto per unità di analisi e i metadati di classificazione.

I dati di confronto sono stati acquisiti principalmente dai microdati del censimento del 2001 dopo aver effettuato operazioni di riconciliazione che rendessero i dati confrontabili con quelli del censimento 2011, in particolare: gli individui (D\_IND\_2001\_CONF\_2011), gli alloggi (D\_ALLOGGI\_2001\_CONF\_2011), le convivenze (D\_CONV\_2001\_CONF\_2011), gli edifici (D\_EDI\_2001\_CONF\_2011), le famiglie (D\_FAM\_2001\_CONF\_2011) e i nuclei (D\_NUCLEI\_2001\_CONF\_2011). Inoltre sono state acquisite fonti di confronto dedicate, come ad esempio i dati ACI per il numero di automobili per comune (D\_ACI\_AUTO) e il numero di connessioni internet per tipologia (D\_CITTADINI\_NUOVE\_TECNOLOGIE).

Sono state identificate 6 entità di analisi (figura 10.2): la popolazione, gli alloggi, le famiglie, i nuclei, le convivenze e gli edifici. Ognuna di esse è stata caratterizzata da innumerevoli attributi, alcuni definiti dalle variabili di rilevazione altri ottenuti da una combinazione di queste secondo regole definite. Gli individui sono stati caratterizzati da circa 190 attributi. Tra le entità, il fulcro è rappresentato dall'individuo che è in relazione con quasi tutte le altre entità.

Figura 10.2 - Schema Entità e Relazioni dei microdati del Censimento



Per ogni entità è stata definita un tabella:

- ▶ D\_INDIVIDUI per i residenti distinti in residenti in convivenza o in famiglia, occupanti un alloggio o no (senza tetto);
- ▶ D\_ALLOGGI per gli alloggi abitativi e non, occupati da: residenti, non residenti, o vuoti; appartenenti ad edifici residenziali;
- ▶ D\_EDIFICI per tutti gli edifici censiti;
- ▶ D\_CONVIVENZE per tutte le convivenze anagrafiche, ossia a cui appartiene almeno un residente, o non anagrafiche (composte da soli non abitualmente residenti, Nad);
- ▶ D\_FAMIGLIE per tutte le famiglie contenenti i riferimenti all'alloggio da esse occupato;
- ▶ D\_NUCLEI per i nuclei che compongono le famiglie, e la definizione della coppia se presente.

Il flusso di alimentazione del Data Base (Db) è stato guidato dai rilasci del sistema Scdd che operava su un set di variabili di rilevazione distinte per temi, ossia i dati anagrafici, dell'alloggio, della famiglia, del lavoro, dello studio, del pendolarismo, delle convivenze, ecc.

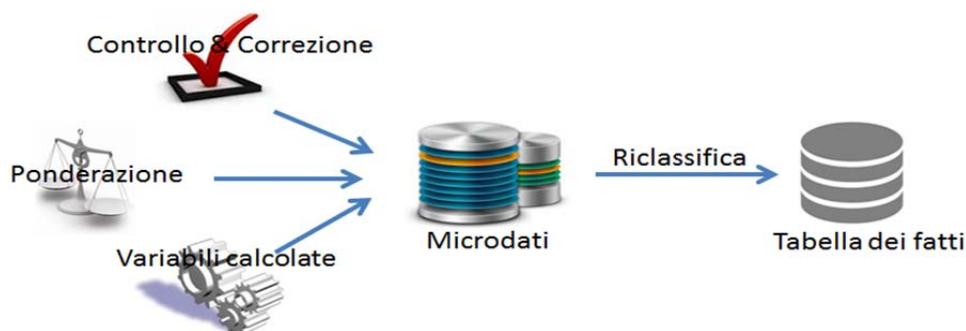
Inoltre, per le unità di analisi degli individui in famiglia e per gli alloggi la rilevazione ha previsto modelli Long e Short (che contenevano un sotto insieme delle domande del modello Long) e per questi è stata effettuata una ponderazione del campione delle famiglie, vincolata attraverso pesi di riporto all'universo degli individui e delle abitazioni. Quindi per ogni individuo, per ogni alloggio e per ogni famiglia sono stati acquisiti i pesi forniti dalle operazioni di ponderazione.

### 10.2.1 Etl di caricamento e riconciliazione

Nella tabella di microdati, ogni record rappresenta una unità con tutte le variabili rilevate e calcolate. I flussi di alimentazione delle tabelle dei microdati sono stati essenzialmente di 6 tipi (figura 10.3):

- ▶ due aventi come input i risultati dei processi di controllo e correzione; di questi uno massivo in caso di rilascio di un set di variabili del questionario, e uno puntuale in caso di correzioni successive (ricicli);
- ▶ due aventi come input i risultati delle procedure di ponderazione; di questi uno massivo per il caricamento dei pesi, ed uno puntuale per le imputazione di questionari che da short diventavano long per risolvere problematiche di convergenza;
- ▶ due con input variabili calcolate con procedure particolari (es: la tipologia familiare che dipende dai dati dei componenti della famiglia) o variabili ereditate da altre unità di analisi (es: tipo di alloggio abitato per gli individui); anche per questi uno massivo in caso di rilascio di una o più variabili, e uno puntuale in caso di correzioni successive;

Figura 10.3 - Processo di caricamento Etl1



I flussi di alimentazione delle tabelle di microdati sono stati di solo aggiornamento, quindi non hanno mai modificato la numerosità delle unità ma solo quella delle variabili.

L'implementazione di tali flussi è avvenuta tramite procedure ad hoc eseguite direttamente sul database. Tali procedure sono state sviluppate modificando due procedure template: una prevista per alimentazioni di tipo massivo (i dati vengono aggiornati a blocchi con chiave territoriale a livello di comune) e una per alimentazioni di tipo puntuale (i dati vengono aggiornati per singola unità).

### 10.3 Architettura dei macrodati

#### 10.3.1 Il modello Multidimensionale: il Data Warehouse Primario

Per supportare le attività di validazione e successivamente di diffusione dei dati del censimento è stato sviluppato un Data Warehouse Primario (Dwp). Tale tecnologia è stata utilizzata per condurre l'analisi dei dati ed infine per attivare i processi di validazione e, in alcuni casi, al ritorno di un set di dati o di variabili alla fase di controllo e correzione. Al termine delle attività di validazione il Dwp è stato utilizzato per l'estrazione dei dati di diffusione.

Con il termine Data Warehouse Primario si intende un modello multidimensionale che si basa su ambienti di data base relazionali modellati in schemi dati a stella (star schema) o a snowflake e loro varianti.

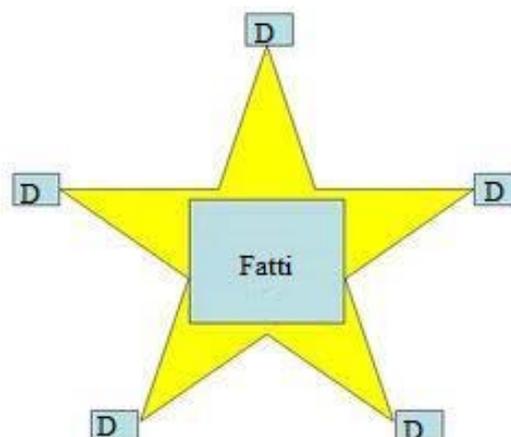
Tali schemi dati sono formati esclusivamente da due tipi di tabelle:

- ▶ le tabelle delle “*dimensioni*”, che sono le entità intorno alle quali si articolano le analisi, le variabili qualitative;
- ▶ la tabella dei “*fatti*”, che contiene le informazioni relative alle analisi, quindi i valori numerici detti “*misure*”, le variabili quantitative, e i collegamenti alle tabelle delle dimensioni.

Le dimensioni rappresentano gli attributi dei fatti di cui si vuole misurare la quantità, in particolare alcune di esse sono strutturate in maniera gerarchica. Un esempio esplicativo è il territorio italiano, il quale è strutturato in ripartizioni, regioni, province e comuni.

Si è scelto di strutturare le dimensioni in maniera de-normalizzata, tale da avere le gerarchie in una unica tabella, in questo caso si parla di schema a forma di stella (star schema) (figura 10.4); altrimenti, nel caso in cui le dimensioni gerarchiche mantengono le relazioni tra livelli gerarchici su tabelle distinte, lo schema viene detto a fiocco di neve (snowflake schema).

Figura 10.4 - Schema a stella



Quindi l'analisi dei dati avviene rappresentando i dati in forma multidimensionale, ed i concetti rilevanti sono:

- ▶ il fatto, concetto sul quale centrare l'analisi;
- ▶ la misura, proprietà del fatto da analizzare;
- ▶ la dimensione, descrive una prospettiva lungo la quale effettuare l'analisi.

In termini intuitivi un modello multidimensionale è una matrice di tabelle e l'individuazione di un certo evento si ottiene attraverso uno spazio n-dimensionale i cui assi sono le dimensioni di analisi. Si deduce che un singolo evento è identificato univocamente dalle dimensioni scelte ed è descritto attraverso le misure. Le dimensioni del modello multidimensionale sono formate da gerarchie di aggregazione e sono strutturate in gerarchie di livelli. Questi livelli, che compongono una o più gerarchie, sono chiamati attributi dimensionali e hanno granularità differenti. Tale modello può essere strutturato in cubi n-dimensionali detti anche ipercubi.

Quindi nel Data Warehouse Primario è stato modellato:

- ▶ un cubo per ogni unità di analisi: popolazione, famiglie, convivenze, nuclei, alloggi, edifici;
- ▶ le dimensioni di analisi sono state progettate in modo tale da modellare le variabili sia di validazione, spesso legate alle modalità della singola variabile, sia quelle di diffusione;
- ▶ ogni cubo, oltre ad avere la tabella dei fatti inerente ai dati censuari 2011 hanno ulteriori tabelle di fatti per il confronto sui dati del censimento 2001 e di altre fonti.

L'analisi multidimensionale eseguita navigando attraverso i livelli di aggregazione delle dimensioni del fatto è stata implementata attraverso tecnologie Olap messe a disposizione dalla suite di Business Intelligence (Bi) Microsoft in particolare i servizi di Business Analytics implementati attraverso Reports, Dashboards, ed estrazioni personalizzate tramite Power Pivot.

Questa tecnologia ha consentito l'elaborazione di operazioni per il supporto alle decisioni, operazioni complesse e non previste, che hanno coinvolto una notevole quantità di dati, sia in forma micro, sia aggregata e storicizzata. Le analisi con strumenti Olap hanno la caratteristica della visione multidimensionale e logica delle informazioni, vi è la possibilità di analisi interattive delle informazioni con aggregazioni per ogni intersezione di ogni dimensione.

Le operazioni basilari dell'analisi multidimensionale che sono state utilizzate con questi strumenti, sono le seguenti:

- ▶ drill down: disaggrega i dati e cioè mostra i dati ad un minor livello di aggregazione rispetto alla visione corrente, cioè consente di scendere nel dettaglio lungo una o più dimensione gerarchica;
- ▶ roll up: è l'operatore duale del drill down, in quanto consente di risalire lungo una o più dimensione gerarchica aggregando i dati, cioè mostra i dati ad un maggior livello di aggregazione rispetto alla visione corrente
- ▶ drill across: è l'estensione dell'operatore di drill down, che consente di scendere nel dettaglio contemporaneamente su più dimensioni;
- ▶ slice: è l'operatore che permette di vedere il cubo trasversalmente (letteralmente "a fette"), fissando un valore per almeno una delle dimensioni e analizzando i dati relativamente a tutte le altre, cioè concentrando l'attenzione su un iper-cubo di dimensione minore del cubo originario;
- ▶ dice: è l'operatore per cui fissato un intervallo su ciascuna dimensione, si analizza una riduzione volumetrica, senza contrazioni del numero di dimensioni;
- ▶ pivot: è l'operazione che consente di "girare" la vista multidimensionale dei dati.

Inoltre sono stati eseguiti sviluppi particolari nei casi in cui le misure non erano comparabili con la misura principale, il conteggio delle occorrenze, poiché erano il risultato di domande di questionario con risposta multipla. In questi casi sono state create ulteriori tabelle dei fatti che

mantenevano le dimensioni in comune con la misura enumerativa, ed erano arricchite con ulteriori dimensioni relative alle risposte multiple.

Successivamente verranno definiti i dettagli delle strutture delle dimensioni e dei fatti, e gli strumenti Olap utilizzati per l'analisi dei dati.

### 10.3.2 I Fatti e le Dimensioni

Il modello multidimensionale del Dwp del censimento è composto da un cubo per ogni unità di analisi. Inoltre, poiché ci sono state due diffusioni, una per Eurostat e una per l'Italia, differenti per tipologia di dimensioni e di entità di interesse, sono stati realizzati cubi ad hoc per le due diffusioni. Sono stati prodotti in totale 10 cubi: 4 per Eurostat (Individui, Nuclei, Famiglie e Alloggi) e 6 per la diffusione italiana (Individui, Nuclei, Famiglie, Alloggi, Convivenze ed Edifici). Ogni cubo è composto di fatti e dimensioni valorizzati a partire da tabelle del database dei microdati rilevati.

Nelle figure 10.5 e 10.6 si riporta un esempio di schema Dimensional Fact Model, (Dfm).

Figura 10.5 - Modello concettuale del cubo alloggi

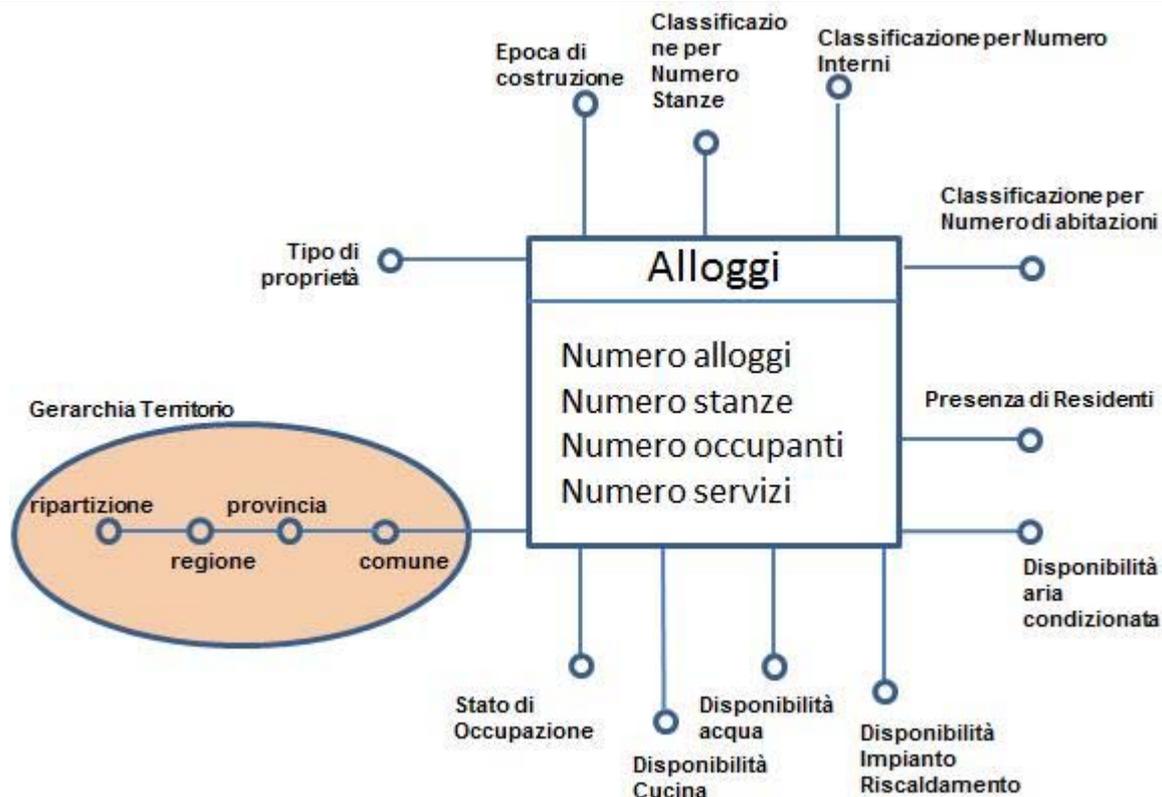
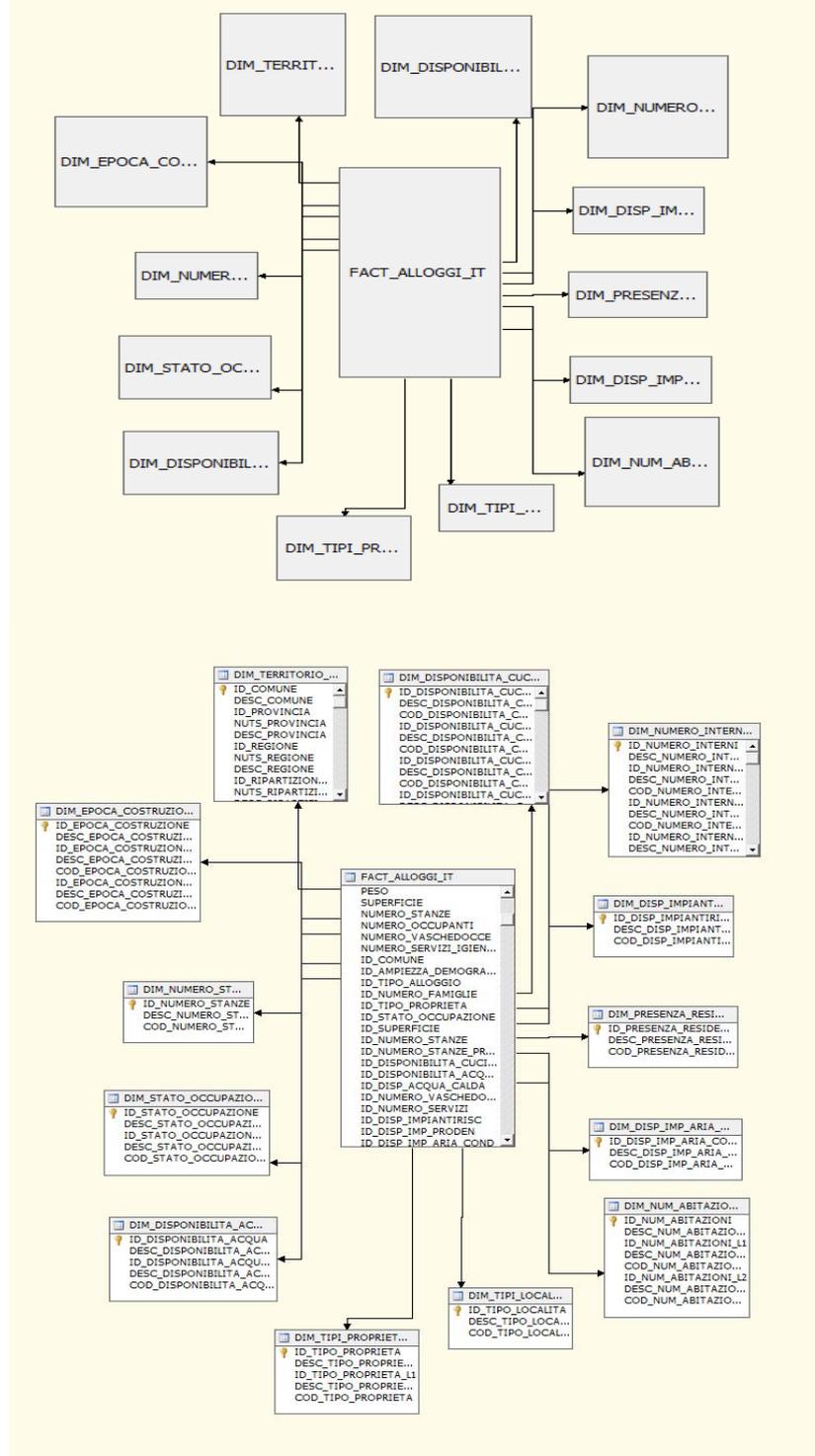


Figura 10.6 - Modello logico e fisico del cubo alloggi



Le tabelle dei microdati di ogni unità sono state caratterizzate dalle variabili di rilevazione, mentre nelle tabelle dei fatti, ossia le unità di analisi, i record sono stati classificati secondo le dimensioni di diffusione e di validazione.

Nella tabella 10.1 si riportano le numerosità dei cubi, delle dimensioni e dei livelli di gerarchie.

**Tabella 10.1 - Numerosità delle dimensioni dei cubi**

CUBI	Numero dimensioni	Numero livelli gerarchici
Popolazione IT	98	130
Popolazione EU (population)	19	37
Alloggi IT	28	50
Alloggi EU (dwellings)	15	17
Famiglie IT	38	52
Famiglie EU (households)	4	9
Nuclei IT	27	43
Nuclei EU (families)	3	8
Edifici IT	19	26
Convivenze IT	11	18

Alcuni quesiti del questionario prevedevano la possibilità di una risposta multipla (ad esempio il quesito sugli impianti di riscaldamento dell'abitazione) e, in diffusione, su queste domande sono stati previsti ipercubi che misuravano le entità derivate, ossia i servizi. Per rispondere a questo tipo di esigenza sono state create delle fact specifiche, come nel caso degli impianti di riscaldamento, che hanno consentito di aggiungere misure al cubo degli alloggi, per misurare i servizi degli impianti di riscaldamento degli alloggi.

Nella base dati ogni dimensione di analisi è stata rappresentata da un'unica tabella de-normalizzata, anche in presenza di diversi livelli di gerarchia. Ogni livello è rappresentato da tre campi: uno per la chiave, uno per la descrizione e uno per la codifica da inviare a I.Stat. La chiave del livello foglia di ogni dimensione è il valore che si ritrova nelle varie tabelle dei fatti (molte dimensioni sono presenti su più cubi e dunque su più tabelle dei fatti).

I livelli di aggregazione sono poi stati ricreati nel modello multidimensionale in modo che i dati possano essere interrogati con ogni singolo livello della dimensione. Inoltre, sul modello sono state create anche gerarchie con le quali possono essere navigati i livelli delle dimensioni che sono legate da relazioni padre/figlio.

A scopo esemplificativo riportiamo qualche dato sulla dimensione più complessa: la dimensione delle età. La dimensione età ha 26 livelli di aggregazione che vengono ulteriormente raccolte nel modello multidimensionale in 22 gerarchie. Ci sono dunque 48 modalità per poter interrogare i dati di popolazione per classe di età ed ognuna di queste può essere incrociata con una o più delle 98 dimensioni del cubo individui.

Le tabelle delle dimensioni sono state caricate (e aggiornate) con un processo automatico. Le tabelle dei fatti, sono state invece caricate con un Etl parametrico che ha come input le tabelle di microdati, tale processo è descritto nel dettaglio nel prossimo paragrafo.

### 10.3.3 Etl di caricamento delle tabelle dei fatti

Le tabelle dei fatti hanno le stesse numerosità delle tabelle dei microdati (ogni record rappresenta una unità) e per rendere più efficienti le attività di caricamento sono state partizionate per provincia. Il caricamento (Load) è stato effettuato tramite una procedura generalizzata di Etl che ha letto (Extract) i microdati e li ha trasformati (Trasformation) secondo regole definite.

Il flusso di elaborazione dei dati è stato strutturato per blocchi di variabili di rilevazione su sottoinsiemi di province. Per ogni flusso è stata definita una fase e un insieme di dimensioni. Per ottenere la riclassificazione dei dati secondo le dimensioni delle tabelle dei fatti sono state definite appropriate regole di trasformazione.

Per ottemperare a queste esigenze, in archivio sono state definite diverse tabelle documentative del processo che contengono:

- ▶ le regole di riclassificazione del microdato, DOC\_REGOLE;
- ▶ le fasi di elaborazione dei dati a blocchi di variabili previste nel workflow, DOC\_PARAMETRI\_Etl; per ogni fase sono descritte: le variabili di rilevazione, le dimensioni e le tabelle di fatti che devono essere ricaricate.

Quindi l'elaborazione per blocchi di dati a livello provinciale è stata strutturata nel seguente modo:

- ▶ cancellazione della partizione per provincia;
- ▶ riclassifica delle variabili previste dalla fase secondo le "regole";
- ▶ caricamento dei dati riclassificati nelle tabelle dei fatti previste dalla fase.

Con questa modalità i tempi di popolamento delle tabelle dei fatti sono stati approssimativamente quelli riportati nella tabella 10.2.

Tabella 10.2 - Tempi medi di caricamento delle tabelle dei fatti

Entità	IT/EU	Numero unità	Numero regole	Risposte multiple	Totale caricati	recordTempo medio (in minuti)
Alloggi	IT	31.295.822	36	4	98.739.011	70
Alloggi	EU	31.295.822	23	-	31.295.822	20
Convivenze	IT	24.029	17	-	24.029	1
Edifici	IT	14.515.795	22	-	14.515.795	10
Famiglie	IT	24.611.766	43	1	33.410.486	50
Famiglie	EU	24.583.190	9	-	24.583.190	15
Individui	IT	59.433.744	77	1	60.696.053	190
Individui	EU	59.433.744	23	-	59.433.744	110
Nuclei	IT	16.648.813	25	-	16.648.813	30
Nuclei	EU	16.646.376	2	-	16.646.376	10

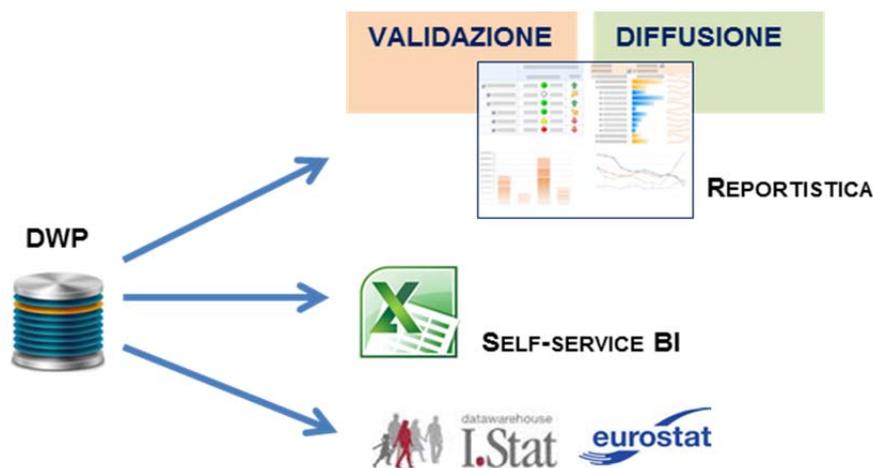
## 10.4 I servizi offerti

### 10.4.1 I servizi offerti dal Data Warehouse Primario

Il Data Warehouse Primario ha supportato le fasi di validazione, analisi e diffusione dati grazie alle seguenti funzionalità rese disponibili (figura 10.7):

- ▶ accesso diretto ai dati mediante un linguaggio di interrogazione per formulare “dinamicamente” le richieste necessarie (client Olap);
- ▶ fruizione di reportistica con presentazione dei dati in una varietà di formati: semplici tabelle, tabelle complesse con subtotali di riga e colonna, grafici di vario genere, immagini, ecc...;
- ▶ utilizzo di indicatori di performance (Key performance Indicator, scorecard, dashboard).

Figura 10.7 - Utilizzo del Data Warehouse Primario



Come illustrato brevemente nella figura 10.7, nell’ambito specifico della diffusione dei dati del censimento Popolazione 2011, tali funzionalità hanno reso possibile da parte del Dwp la disponibilità dei seguenti servizi principali:

- ▶ reportistica di supporto alla fase di validazione e diffusione;
- ▶ self-service Bi di supporto all’attività di analisi esplorativa sui dati: è stata data solo ad utenti esperti la possibilità di estrarre direttamente dati non pre-strutturati dal Dwp e dal Db dei microdati anche per esigenze sporadiche e/o puntuali;
- ▶ estrazione di sottoinsiemi di dati specifici finalizzati alla diffusione verso I.Stat e Eurostat secondo un piano di spoglio stabilito dai piani di diffusione.

#### 10.4.2 La reportistica

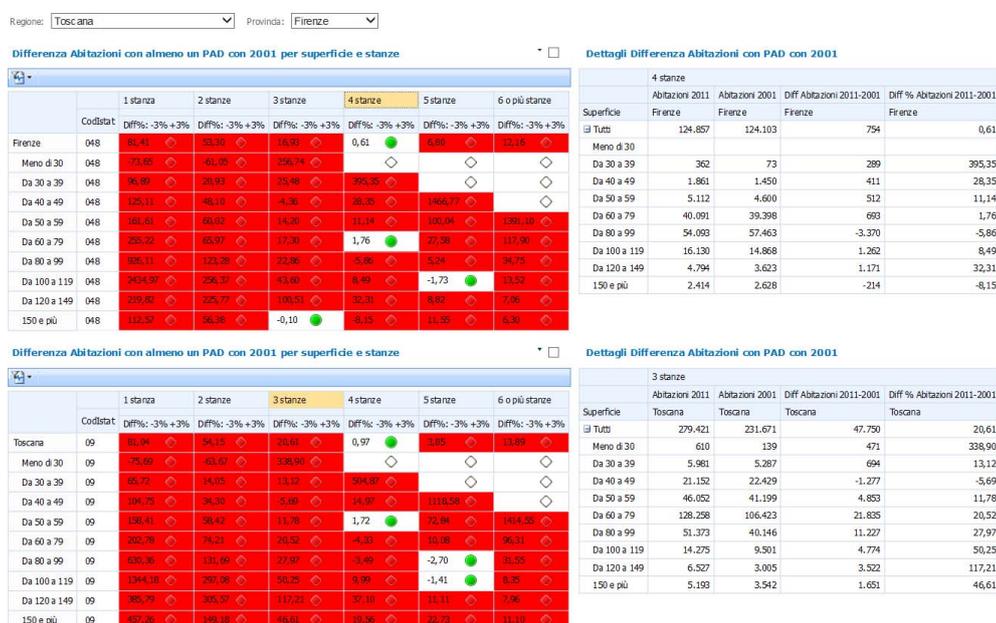
A supporto della fase di validazione dei dati prima e della successiva fase di diffusione, sono stati prodotti report riassuntivi sulle varie tematiche (Popolazione, Famiglie, Alloggi, Convivenze, Edifici, Nuclei) utilizzando principalmente i servizi di PerformancePoint messi a disposizione dalla suite di Bi utilizzata. Questi, sono stati resi fruibili agli utilizzatori via Web in maniera condivisa, completamente integrati con il sistema di collaborazione interno all’Istituto, nel rispetto dei vincoli di autorizzazione e visibilità per utenti specifici.

Durante la fase di validazione, tali report sono stati utilizzati dai referenti tematici e revisori per verificare eventuali anomalie nei dati, attraverso un confronto con diverse fonti di benchmark (dati del censimento del 2001, dati della popolazione legale o dati raccolti da altre indagini dell’Istituto). I report sono stati predisposti principalmente come un insieme organizzato di misure (Kpi), atte ad evidenziare scostamenti di valori in relazione a soglie prefissate. L’obiettivo è stato quello di effettuare una prima scrematura dei valori fuori soglia per evidenziare ai revisori quelle situazioni che realmente hanno avuto necessità di una ulteriore revisione più accurata e dettagliata, tramite analisi esplorative sui dati stessi. Un esempio è il dashboard mostrato nella figura 10.8, che

evidenzia le differenze del numero di abitazioni con almeno una persona dimorante del 2011 rispetto a quelle censite nel 2001, classificate per superficie e numero di stanze con un dettaglio territoriale a livello di regione e di provincia selezionata (Toscana e Firenze in questo caso specifico). Il dashboard è composto da due scorecard (nella parte sinistra) che evidenziano in rosso le situazioni che presentano una differenza  $<-3$  per cento o  $>3$  per cento (soglia prefissata), in verde quelle entro la soglia e in bianco le non definite. Nella parte destra invece, sono riportati i report di dettaglio, in termini di numero di abitazioni (del 2011 e del 2001), corrispondenti allo specifico numero di stanze selezionato (4 stanze o 3 stanze).

Figura 10.8 Esempi di dashboard con scorecard

Figura 10.8 - Esempi di dashboard con scorecard



Per supportare invece la fase di diffusione dei dati, sono stati predisposti report a matrice aventi una struttura fissa, con la stessa visualizzazione delle tavole che successivamente sono state diffuse su I.Stat. A titolo di esempio, la figura 10.9 mostra un report delle Famiglie in abitazione classificate per Titolo di godimento e Superficie dell'abitazione, con il dettaglio territoriale della ripartizione Sud d'Italia.

Sia i report di validazione che quelli di diffusione sono stati realizzati in modo che agli utilizzatori fosse sempre consentito di poter effettuare operazioni di drill-down, sia per verificare casi anomali (in fase di validazione), che per visualizzare dati di maggior dettaglio. Ad esempio si è data la possibilità di poter navigare i dati all'interno dell'intero territorio italiano, partendo dal livello Italia (di massima aggregazione), attraverso quello ripartizionale, regionale e provinciale fino a livello comunale.

Complessivamente sono stati predisposti circa 70 oggetti di tipo PerformancePoint Kpi, 80 Scorecard, 270 Report, 40 Filter necessari alla composizione delle 25 dashboard di presentazione.

Figura 10.9 - Esempio di report di diffusione

Territorio: **Sud** ▼

**Famiglie in abitazione per Titolo di Godimento**

		☑ Tutti		
		Proprietà, Usufrutto o Riscatto	Affitto	Altro Titolo
Territorio, Superficie	Famiglie 2011	Famiglie 2011	Famiglie 2011	Famiglie 2011
+ Sud				
Meno di 30	10.154	5.256	2.679	2.219
Da 30 a 39	153.417	69.772	44.451	39.194
Da 40 a 49	330.488	158.588	101.337	70.563
Da 50 a 59	445.060	236.715	125.023	83.322
Da 60 a 79	1.407.160	844.149	351.880	211.131
Da 80 a 99	2.038.480	1.405.578	390.490	242.412
Da 100 a 119	1.555.892	1.200.044	183.771	172.077
Da 120 a 149	1.054.240	859.221	84.889	110.130
150 e più	864.519	726.920	53.008	84.591

### 10.4.3 Self-Service Business Intelligence

Con il termine Self-Service Business Intelligence (Self-Service Bi) si individua un insieme di strumenti software di semplice utilizzo, che consentono all'utente finale di costruire rapidamente: report, analisi e nuovi modelli di dati (Data Model). Lo scopo principale della Self-Service Bi è di svincolare l'analista-statistico dal personale It, e di consentire la gestione autonoma della:

- ▶ modellazione dei dati;
- ▶ predisposizione di estrazioni dati e di report;
- ▶ analisi autonoma dei dati.

Gli strumenti di analisi messi a disposizione dalla suite Bi Microsoft possono essere affiancati da un componente aggiuntivo di Excel (già presente per la versione 2010), Power Pivot per Excel. Tale add-in consente, all'interno di un foglio di calcolo Excel, di connettersi ad una qualsiasi fonte dati e di costruire una struttura dati tabellare che può essere trasformata in una struttura multidimensionale. Il Tool ha consentito a quanti si occupavano di analisi e validazione dei dati del censimento sia di accedere alla struttura multi dimensionale del Dwp sia di accedere ai microdati del censimento e/o a dati aggiuntivi. In questo modo gli utenti hanno potuto effettuare analisi personalizzate in cui la possibilità di combinare tutte le fonti, consentiva di costruire un modello dati ad hoc.

Lo strumento ha consentito di accedere:

- ▶ ai dati multi dimensionali del Dwp, impostando filtri, e consentendo la libera navigazione delle misure e delle dimensioni d'analisi semplicemente trascinandole nel riquadro della selezione (in figura 10.10 denominato "Dimensioni di analisi");
- ▶ ai microdati del data base, consentendo di costruire un nuovo modello dati (figura 10.11) e con esso una nuova struttura multi dimensionale (figura 10.12).

Figura 10.10 - Esempi di analisi dei dati del Dwp

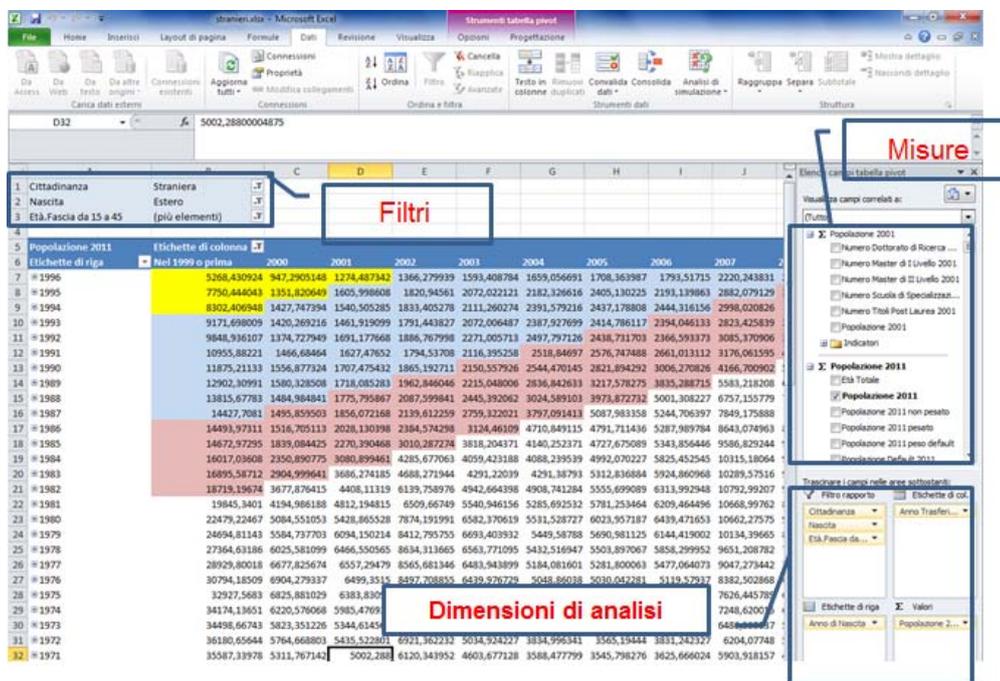


Figura 10.11 - Esempi di analisi dei microdati con PowerPivot

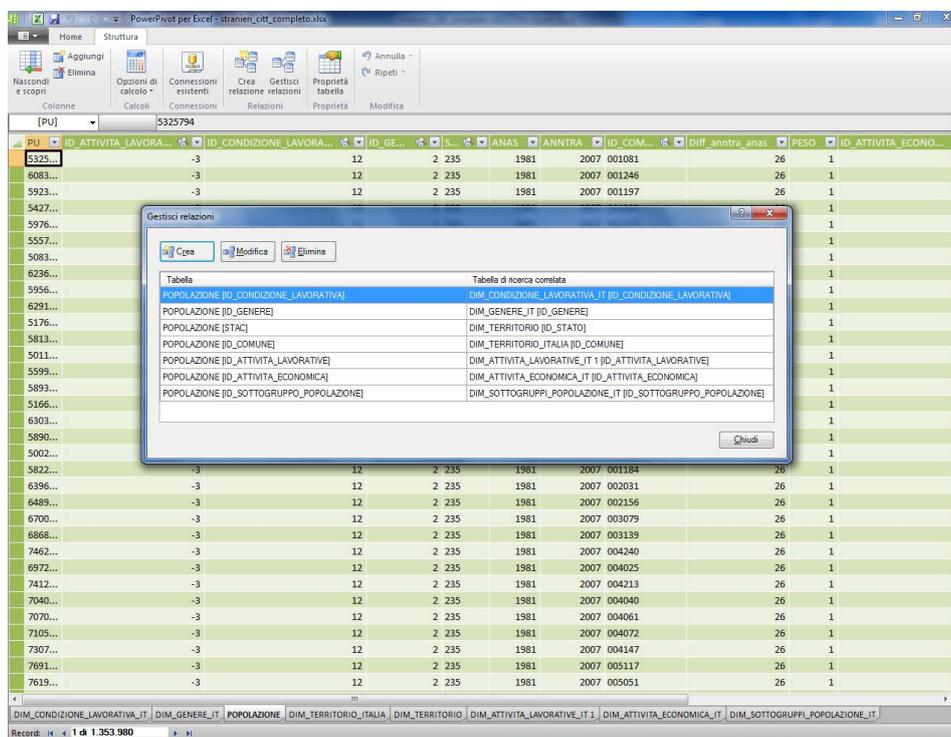
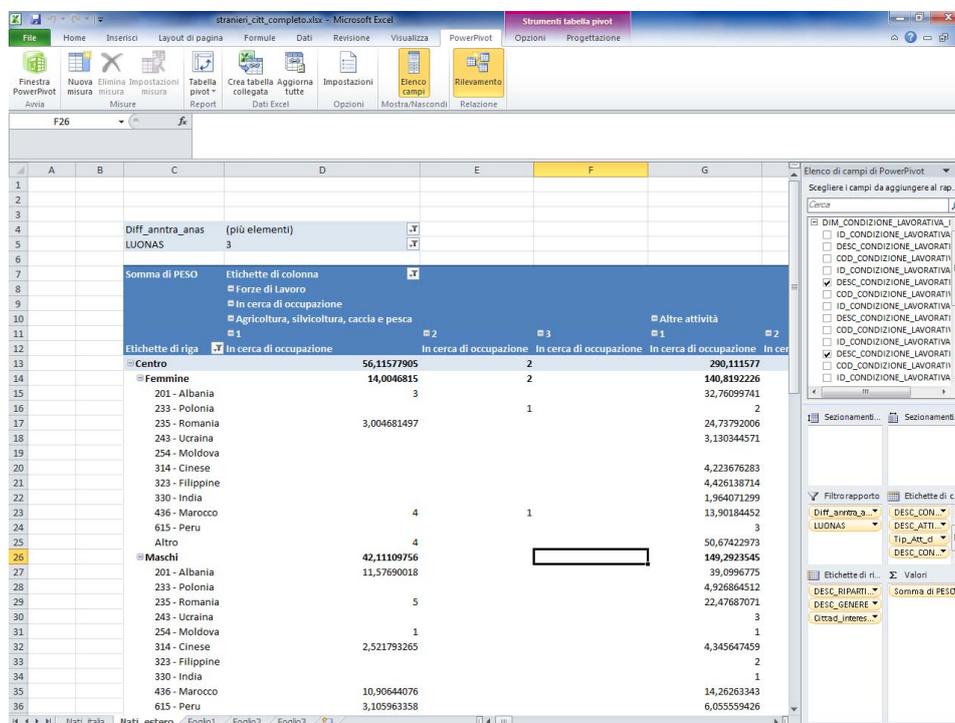


Figura 10.12 - Esempi di creazione di una nuova struttura multidimensionale



	Etichette di colonna			
	Forze di Lavoro			
	In cerca di occupazione			
	Agricoltura, silvicoltura, caccia e pesca			
	Altre attività			
	1	2	1	2
	Etichette di riga	In cerca di occupazione	In cerca di occupazione	In cerca di occupazione
	Centro	56,11577905	2	290,111577
	Femmine	14,0046815	2	140,8192286
	201 - Albania	3		32,78695741
	233 - Polonia		1	2
	235 - Romania	3,004661497		24,73792006
	243 - Ucraina			3,130344571
	254 - Moldovia			
	314 - Cinese			4,223676283
	323 - Filippine			4,426138714
	330 - India			1,964071299
	436 - Marocco	4	1	13,90184452
	615 - Perù			3
	Altro	4		50,67422973
	Maschi	42,11109756		149,2923545
	201 - Albania	11,57690018		39,0996775
	233 - Polonia			4,926864512
	235 - Romania	5		22,47687071
	243 - Ucraina			3
	254 - Moldovia			1
	314 - Cinese	2,521793265		4,345647459
	323 - Filippine			2
	330 - India			1
	436 - Marocco	10,90644076		14,26263343
	615 - Perù	3,105963358		6,055559426

Inoltre Power Pivot, consentendo di salvare la struttura dati costruita, ha permesso agli analisti, con un semplice aggiornamento dei dati, di visualizzarne gli aggiornamenti nelle strutture costruite.

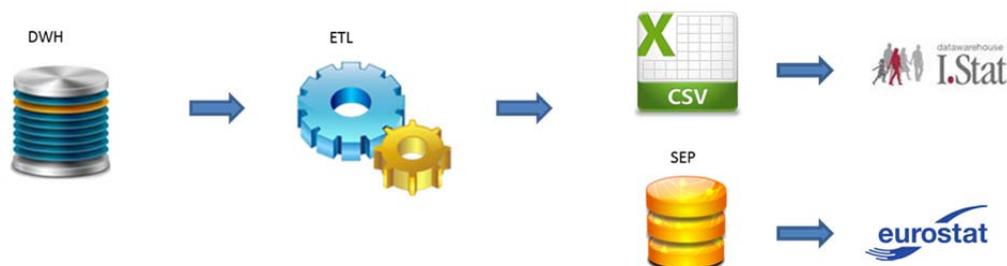
#### 10.4.4 Procedure per l'estrazione dei dati

I.Stat, la banca dati on line delle statistiche prodotte dall'Istat, richiede che il caricamento dei macrodati sia effettuato a partire da file in formato csv. La struttura del file csv dei macrodati deve rispettare, oltre ad una serie di convenzioni proprie del sistema, gli incroci stabiliti nel Piano di Spoglio (PdS), le classificazioni di diffusione e le singole modalità.

È nata quindi l'esigenza di progettare una soluzione informatica (figura 10.13) che risolvesse principalmente le seguenti problematiche:

- ▶ estrarre i dati dal Dwp tramite query di interrogazione (effettuate tramite il linguaggio Mdx, proprio della soluzione BIWin);
- ▶ eseguire le trasformazioni sui dati, in accordo ai vincoli delle specifiche I.Stat e dei PdS;
- ▶ riclassificare i dati in presenza di possibili incongruenze e/o anomalie;
- ▶ unire i risultati ottenuti dall'estrazione considerando tutti gli incroci del PdS in un unico file csv;
- ▶ progettare un sistema di convalida dei dati ottenuti e della correttezza dell'estrazione. In presenza di errori (ad esempio, righe duplicate nel risultato dell'estrazione) notificare le informazioni utili ad una veloce correzione, ai fini della correzione e ripetizione dell'estrazione;
- ▶ considerare la riusabilità della soluzione per diversi punti di destinazione dei dati, come ad esempio nel caso del sistema Single Exit Point (Sep) dell'Istat, utilizzato in questo progetto per diffondere i dati a utenti esterni in modalità machine-to-machine.

Figura 10.13 - Processo di Etl di diffusione



Sono state realizzate due procedure di Etl, una per l'estrazione e un'altra per la sua convalida. Alcuni aspetti significativi che sono stati presi in considerazione per la realizzazione delle procedure Etl sono stati:

- ▶ l'Etl di estrazione deve prendere un insieme di query Mdx in input, una per ogni incrocio del PdS;
- ▶ le query devono essere realizzate in modo da consentire il debug in fase di convalida;
- ▶ l'Etl di convalida deve escludere la presenza di informazioni duplicate del file Csv;
- ▶ in caso di dati duplicati detti "bloccanti" (ad esempio se in corrispondenza di una stessa chiave si hanno due valori associati e diversi delle misure), occorre segnalare via email le query che hanno causato l'anomalia, per poter rieseguire l'estrazione.

Il tool di sviluppo utilizzato per il workflow e gli Etl è stato Sql Server Data Tools 2012 di Microsoft, con la realizzazione di due Package Ssis (Sql Server Integration Services). Nella figura 10.14, è illustrato un esempio del flusso dell'Etl di convalida. A livello di implementazione, le procedure Etl sono state realizzate come componente control flow Ssis che includono passi di estrazione e trasformazione, alcuni realizzati come script in linguaggio C#. Ad esempio, uno di questi script ("Normalizza Decimali", in figura 10.14) si occupa della predisposizione dei dati nel formato a due cifre decimali richiesto da I.Stat.

Le procedure Etl sono state anche usate per alimentare, a partire dalla sorgente Mdx, il Db utilizzato per il Sep, nel contesto del progetto Census Hub di Eurostat. Visti i grandi volumi di dati da trasferire, per questa fase è stato scelto di realizzare un tool aggiuntivo (scritto in linguaggio Java) per effettuare la scrittura dei dati nel database in maniera più efficiente rispetto a un caricamento diretto da Ssis. Più precisamente il tool permette di effettuare caricamenti in modalità batch e in parallelo, riducendo notevolmente i tempi di esecuzione rispetto a un'esecuzione seriale di Ssis.

Figura 10.14 - Esempio del flusso dell'Etl di convalida



## 10.5 La progettazione dei cubi dati per la diffusione

In occasione del 15° Censimento generale della popolazione e delle abitazioni l'Istat ha scelto, come tecnologia per la diffusione dei dati, la medesima utilizzata per il data warehouse dell'Istituto: I.Stat. Obiettivo del presente paragrafo è la presentazione della metodologia utilizzata per il passaggio dal Piano di diffusione dati alla modellazione e progettazione di cubi dati navigabili all'interno del data warehouse, con particolare riferimento alle caratteristiche dei dati di tipo censuario. Si riporta di seguito una sintesi dei passi realizzati per la modellazione ed alcune esemplificazioni, giungendo ad una breve sintesi.

### 10.5.1 Le fasi del processo, le logiche e la metodologia seguite

Il Piano di diffusione, realizzato dal settore responsabile dell'operazione censuaria, ha avuto il compito di elencare, enumerare e dettagliare gli aggregati statistici ed i relativi incroci che si intendeva diffondere. In fase di progettazione dei cubi dati si sono realizzate tre fasi:

- 1) **primo passo - individuazione degli aggregati statistici di diffusione e relativa organizzazione tematica.** A partire dal piano di diffusione si sono identificati gli *aggregati statistici* di diffusione (denominati *oggetti* di diffusione o *tipi dato*) e si è proceduto alla loro suddivisione in sotto-insiemi secondo una logica di congruenza in termini di *contenuto* e di *dimensioni di analisi* utilizzate. I sotto-insiemi ottenuti hanno rappresentato la prima stesura dei cubi di diffusione. A loro volta essi sono stati raggruppati in macro-argomenti ciascuno dei quali ha rappresentato un ramo dell'albero dei temi di navigazione (figura 10.15). Ad esempio nel Tema "Famiglie" sono stati inclusi i cubi così denominati: le "Caratteristiche

delle famiglie” – le “Famiglie per tipologia di famiglia” – le “Famiglie con almeno uno straniero residente”;

Figura 10.15 - Albero dei temi del data warehouse di diffusione

The screenshot shows a web interface titled 'Esplora Temi' with a sub-tab 'Tabelle più richieste'. It features a search bar with the text 'Cerca nei temi', a red search button with a double arrow, and an 'Annulla' button. Below the search bar is a dropdown menu currently set to 'Tutti i temi'. A list of topics follows, each with a red square icon containing a white 'x' and a text label:

- Primi Risultati
- Popolazione legale
- Popolazione
- Famiglie
- Convivenze (Ist. religiosi, assistenziali, di cura, ecc.)
- Alloggi ed edifici
- Istruzione e formazione
- Condizione professionale o non professionale
- Caratteristiche dell'attività lavorativa
- Pendolarismo
- Serie storiche

- 1) **secondo step - riorganizzazione delle dimensioni di analisi.** Il processo è consistito innanzitutto nel suddividere, come sarà meglio descritto nel paragrafo successivo, il concetto di *dimensione* di analisi in due concetti distinti: *variabile di diffusione*, *code list* delle modalità utilizzate. Tale suddivisione ha consentito, da un lato, di individuare esattamente gli assi di analisi utilizzati in fase di diffusione per ciascun fenomeno esplicitando i concetti di base, e dall'altro lato di creare code list integrate e riusabili tra diverse variabili di analisi. È stato stimato che da un iniziale numero di 177 code list previsto dal piano di diffusione, attraverso il processo di concettualizzazione ed integrazione, sono state prodotte e gestite circa 45 code list integrate. Un risultato che ha consentito un notevole alleggerimento del processo di implementazione del data warehouse e un positivo effetto anche nella gestione dei flussi di lavoro complessivi. In particolare a volte è stato necessario operare in termini di suddivisione dei concetti presenti nel piano, al fine di individuare esplicitamente i contenuti di ciascuno, oppure di generalizzazione ed astrazione degli stessi al fine di individuarne gli elementi comuni che potessero essere riusati in diversi contesti;
- 2) **terzo step – modellazione di dettaglio.** Questa fase è consistita nel rappresentare gli incroci contenuti all'interno del singolo cubo dati e per ciascun incrocio nel rappresentare le specifiche modalità, per ciascuna dimensione di analisi utilizzata. Lo strumento utilizzato in questa fase è stato un foglio *xls* denominato *Piano di spoglio*. Non è oggetto del presente paragrafo definirne la struttura e la logica. È di rilievo, invece, sottolineare, che esso è stato lo strumento che ha consentito a persone con diverse professionalità - statistiche, informatiche e di diffusione - di colloquiare tra di loro con chiarezza.

Nei paragrafi successivi si forniscono alcuni esempi chiarificatori.

### 10.5.2 Caso 1: Famiglie per numero di componenti – livello comunale

Si consideri l'incrocio del “Piano di diffusione” con il dettaglio della relativa classificazione utilizzata e del livello territoriale (figura 10.16).

Il primo passo di progettazione del cubo, come sopra detto, è stato quello di estrapolare tutti gli aggregati statistici di diffusione afferenti ad un medesimo contenuto e di racchiuderli all'interno di

uno stesso cubo dati di navigazione. Nel caso in questione si sono individuati tutti gli oggetti relativi alle caratteristiche delle famiglie che sono stati collezionati nel cubo denominato *Caratteristiche delle famiglie*.

Figura 10.16 - Famiglie per numero di componenti - livello comunale - Piano delle diffusioni

N. Incrocio	Oggetto	Dettaglio territoriale	Dimensione di analisi	Numero di componenti (8)	
F4.9	Famiglie	Comunale	Numero di componenti (8)	Totale	
				1 persona	
				di cui: non in coabitazione	
				2 persone	
				3 persone	
				4 persone	
				5 persone	
				6 o più persone	

Il secondo passo di modellazione ha riguardato le dimensioni di analisi. Nel caso specifico l'incrocio è caratterizzato dalla classificazione *Numero di componenti (8)* che si compone di otto modalità, con la particolarità che la modalità "1 persona" ha un *di cui* "non in coabitazione". Tale dimensione di analisi, al fine di individuare i concetti di base utilizzati, è stata scissa in due *variabili di classificazione* e relative *code list* opportunamente generalizzate:

- ▶ *numero di componenti* - specifica la numerosità dei componenti della famiglia ed utilizza la *code list* generalizzata *Numerosità* (cod. NUMEROSITA);
- ▶ *condizione abitativa* - dettaglia la condizione abitativa utilizzando specificatamente la voce "non in coabitazione" relativamente alle famiglie unipersonali ed utilizza la *code list* *Condizione abitativa* (cod. COND\_ABIT). È da sottolineare che, poiché la specifica condizione abitativa è definita esclusivamente per le famiglie unipersonali, l'individuazione di due dimensioni separate ha comportato la conseguente definizione delle *dipendenze funzionali* tra di esse. In tal senso per ciascuna dimensione si sono creati gruppi di modalità (figura 10.17), e la dipendenza è stata successivamente mappata all'interno del foglio degli incroci del Piano di Spoglio all'interno del passo tre di lavorazione (figura 10.18). Si precisa che, qualora nel foglio degli incroci non sia presente la x su di una dimensione, significa che si deve automaticamente considerare la relativa voce di *totale* che, di fatto, esclude la dimensione dall'incrocio stesso.



Figura 10.19 - Famiglie per numero di componenti - visualizzazione on line

Caratteristiche delle famiglie : Famiglie per numero di componenti		Metadati						
Tipologia della famiglia	famiglie	NPHH_AV.HH.9.TOTAL.99.N1.SINGLEOCC.ALL.ALL						
Sesso della persona di riferimento della famiglia	totale	Caratteristiche dei dati						
Classe di età della persona di riferimento della famiglia	totale	Altre caratteristiche dei dati						
Stato civile della persona di riferimento della famiglia	totale	la modalità "non in coabitazione" include solo le famiglie in abitazione e in altro tipo di alloggio, non sono conteggiate le famiglie in strutture residenziali collettive e senza tetto per le quali non è stata rilevata l'informazione.						
Situazione abitativa della famiglia	tutte le voci							
Specie di alloggio	tutte le voci							
Titolo di godimento	tutte le voci							
Anno di Censimento	2011							
Tipo dato	numero di famiglie (valori assoluti)							
Numero di componenti	1	2	3	4	5	6 e più	totale	
Condizione abitativa (in/non in coabitazione)	non in coabitazione	tutte le voci						
Territorio	i							
Italia	7 177 537	7 667 305	6 665 800	4 892 316	3 977 401	1 060 350	348 594	24 611 71
Nord-ovest	2 209 388	2 335 543	2 029 394	1 351 914	935 829	208 843	66 446	6 927 91
Piemonte	638 637	671 901	591 709	374 659	246 850	52 204	16 037	1 953 31
Valle d'Aosta / Vallée d'Aoste	22 545	23 509	16 758	9 888	7 276	1 552	387	59 31
Liguria	288 241	309 801	218 633	133 401	75 671	15 674	4 981	758 11
Lombardia	1 259 965	1 330 332	1 202 294	833 966	606 032	139 413	45 041	4 157 01

### 10.5.3 Caso 2: Famiglie per numero di componenti e situazione abitativa della famiglia – livello provinciale

Si consideri ora l'incrocio del Piano di Diffusione *Famiglie per numero di componenti e situazione abitativa della famiglia – livello provinciale* (figura 10.20). Anche tale incrocio, come il precedente, è stato fatto confluire all'interno del cubo *Caratteristiche delle famiglie* (primo passo di modellazione).

Figura 10.20 - Famiglie per numero di componenti e situazione abitativa della famiglia - livello provinciale - Piano delle diffusioni

N. Incrocio	Oggetto	Dettaglio territoriale	Dimensione di analisi	Numero di componenti (7)	Situazione abitativa della famiglia (9)
F3.9	Famiglie	Provinciale	Numero di componenti (7) X Situazione abitativa della famiglia (9)	Totale 1 persona 2 persone 3 persone 4 persone 5 persone 6 o più persone	Totale FAMIGLIE CHE OCCUPANO UN'ABITAZIONE In proprietà In affitto Ad altro titolo FAMIGLIE CHE OCCUPANO UN ALTRO TIPO DI ALLOGGIO* FAMIGLIE IN STRUTTURA RESIDENZIALE COLLETTIVA FAMIGLIE COABITANTI FAMIGLIE SENZA TETTO

Relativamente alle dimensioni di analisi (secondo passo di modellazione), a partire dalle due indicate nel piano di diffusione, *Numero di componenti* e *Situazione abitativa della famiglia*, si sono individuate tre *variabili di classificazione* e tre *code list* di cui una già precedentemente definita (figura 10.21):

*numero di componenti* - specifica la numerosità dei componenti della famiglia ed utilizza la *code list* *Numerosità* (cod. NUMEROSITA);

*situazione abitativa della famiglia* - descrive la struttura dell'abitazione familiare ed utilizza la *code list* - *Abitazione familiare* (cod. ABITAZ\_FAM);

*titolo di godimento* - descrive, nel caso esclusivamente delle famiglie che occupano abitazione, il titolo di possesso del bene ed utilizza la *code list* *Titolo possesso* (cod. TIT\_POSSESSO). Anche

in questo caso si sono dovuti definire dei raggruppamenti per la descrizione delle dipendenze funzionali scaturite dalla individuazione concettuale e separata delle dimensioni di analisi.

Figura 10.21 - Famiglie per numero di componenti e situazione abitativa della famiglia - le code list associate alle dimensioni di analisi

Code list			
Nome	Codice		
Numerosità	NUMEROSITA		
Modalità della Code list			
Codice	Descrizione in italiano	Descrizione in inglese	1 e più (7)
N1	1	1	x
N2	2	2	x
N3	3	3	x
N4	4	4	x
N5	5	5	x
N6_GE	6 e più	6 and over	x
TOT	totale	total	x

Classificazione pubblica di riferimento			
Nome	Codice		
TITOLO POSSESSO	TIT_POSSESSO		
Modalità della classificazione			
Codice	Descrizione in italiano	Descrizione in inglese	prop+aff+altr (3)
OWN	proprietà	ownership	x
RENT	affitto	rent	x
OTH_OR	altro titolo diverso da proprietà, affitto	other form of tenure different from ownership, rent	x

Classificazione pubblica di riferimento					
Nome	Codice				
Abitazione famiglia	ABITAZ_FAM				
Modalità della classificazione					
Codice	Descrizione in italiano	Descrizione in inglese	Ordine di visualizzazione	SIT_ABIT (6)	sit_abit fam in ab (1)
PH_IN_CDW	famiglia in abitazione	private household living in a conventional dwelling	10	x	x
OTHER1	famiglia in altro tipo di alloggio	private household living in other housing unit	20	x	
PH_IN_CLQ	famiglia in struttura residenziale collettiva	private household living in a collective living quarter	30	x	
PH_COHAB	famiglia coabitante	cohabiting private household	40	x	
PH_NO_HOU	famiglia senza tetto	homeless private household	50	x	
ALL	tutte le voci	all items	60	x	

raggr. 1      raggr. 2

Nella figura 10.22 è riportato l'output del terzo passo di modellazione e nella figura 10.23 il risultato on line.

Figura 10.22 - Famiglie per numero di componenti e situazione abitativa della famiglia - foglio degli incroci

Variabile che si vuole pubblicare	Territorio	Tipo dato	Tipologia della famiglia	Sesso della persona di riferimento della famiglia	Classe di età della persona di riferimento della famiglia	Stato civile della persona di riferimento della famiglia	Numero di componenti della famiglia	Condizione abitativa (in/non in coabitazione)	Situazione abitativa della famiglia	Specie di alloggio	Titolo di godimento	Anno di Censimento	Variabili di Classificazione (VCI)																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
													VC1	VC2	VC3	VC4	VC5	VC6	VC7	VC8	VC9	VC10	VC11	VC12																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
													Reg/Rip/Italia	Prov/GC/Reg/Rip/Italia	Com/Prov/Reg/Rip/Italia	numero di famiglie (valori assoluti)	numero di famiglie (valori percentuali)	numero di componenti per famiglia (valori med)	numero di famiglie in alloggio per titolo di godimento (valori)	TIPFAM_L (6)	TIPFAM_tot (1)	TIPFAM_A (1)	TIPFAM_L (1)	TIPFAM_L (2)	TIPFAM_L (3)	TIPFAM_L (4)	TIPFAM_L (5)	TIPFAM_L (6)	TIPFAM_L (7)	TIPFAM_L (8)	TIPFAM_L (9)	TIPFAM_L (10)	TIPFAM_L (11)	TIPFAM_L (12)	TIPFAM_L (13)	TIPFAM_L (14)	TIPFAM_L (15)	TIPFAM_L (16)	TIPFAM_L (17)	TIPFAM_L (18)	TIPFAM_L (19)	TIPFAM_L (20)	TIPFAM_L (21)	TIPFAM_L (22)	TIPFAM_L (23)	TIPFAM_L (24)	TIPFAM_L (25)	TIPFAM_L (26)	TIPFAM_L (27)	TIPFAM_L (28)	TIPFAM_L (29)	TIPFAM_L (30)	TIPFAM_L (31)	TIPFAM_L (32)	TIPFAM_L (33)	TIPFAM_L (34)	TIPFAM_L (35)	TIPFAM_L (36)	TIPFAM_L (37)	TIPFAM_L (38)	TIPFAM_L (39)	TIPFAM_L (40)	TIPFAM_L (41)	TIPFAM_L (42)	TIPFAM_L (43)	TIPFAM_L (44)	TIPFAM_L (45)	TIPFAM_L (46)	TIPFAM_L (47)	TIPFAM_L (48)	TIPFAM_L (49)	TIPFAM_L (50)	TIPFAM_L (51)	TIPFAM_L (52)	TIPFAM_L (53)	TIPFAM_L (54)	TIPFAM_L (55)	TIPFAM_L (56)	TIPFAM_L (57)	TIPFAM_L (58)	TIPFAM_L (59)	TIPFAM_L (60)	TIPFAM_L (61)	TIPFAM_L (62)	TIPFAM_L (63)	TIPFAM_L (64)	TIPFAM_L (65)	TIPFAM_L (66)	TIPFAM_L (67)	TIPFAM_L (68)	TIPFAM_L (69)	TIPFAM_L (70)	TIPFAM_L (71)	TIPFAM_L (72)	TIPFAM_L (73)	TIPFAM_L (74)	TIPFAM_L (75)	TIPFAM_L (76)	TIPFAM_L (77)	TIPFAM_L (78)	TIPFAM_L (79)	TIPFAM_L (80)	TIPFAM_L (81)	TIPFAM_L (82)	TIPFAM_L (83)	TIPFAM_L (84)	TIPFAM_L (85)	TIPFAM_L (86)	TIPFAM_L (87)	TIPFAM_L (88)	TIPFAM_L (89)	TIPFAM_L (90)	TIPFAM_L (91)	TIPFAM_L (92)	TIPFAM_L (93)	TIPFAM_L (94)	TIPFAM_L (95)	TIPFAM_L (96)	TIPFAM_L (97)	TIPFAM_L (98)	TIPFAM_L (99)	TIPFAM_L (100)	TIPFAM_L (101)	TIPFAM_L (102)	TIPFAM_L (103)	TIPFAM_L (104)	TIPFAM_L (105)	TIPFAM_L (106)	TIPFAM_L (107)	TIPFAM_L (108)	TIPFAM_L (109)	TIPFAM_L (110)	TIPFAM_L (111)	TIPFAM_L (112)	TIPFAM_L (113)	TIPFAM_L (114)	TIPFAM_L (115)	TIPFAM_L (116)	TIPFAM_L (117)	TIPFAM_L (118)	TIPFAM_L (119)	TIPFAM_L (120)	TIPFAM_L (121)	TIPFAM_L (122)	TIPFAM_L (123)	TIPFAM_L (124)	TIPFAM_L (125)	TIPFAM_L (126)	TIPFAM_L (127)	TIPFAM_L (128)	TIPFAM_L (129)	TIPFAM_L (130)	TIPFAM_L (131)	TIPFAM_L (132)	TIPFAM_L (133)	TIPFAM_L (134)	TIPFAM_L (135)	TIPFAM_L (136)	TIPFAM_L (137)	TIPFAM_L (138)	TIPFAM_L (139)	TIPFAM_L (140)	TIPFAM_L (141)	TIPFAM_L (142)	TIPFAM_L (143)	TIPFAM_L (144)	TIPFAM_L (145)	TIPFAM_L (146)	TIPFAM_L (147)	TIPFAM_L (148)	TIPFAM_L (149)	TIPFAM_L (150)	TIPFAM_L (151)	TIPFAM_L (152)	TIPFAM_L (153)	TIPFAM_L (154)	TIPFAM_L (155)	TIPFAM_L (156)	TIPFAM_L (157)	TIPFAM_L (158)	TIPFAM_L (159)	TIPFAM_L (160)	TIPFAM_L (161)	TIPFAM_L (162)	TIPFAM_L (163)	TIPFAM_L (164)	TIPFAM_L (165)	TIPFAM_L (166)	TIPFAM_L (167)	TIPFAM_L (168)	TIPFAM_L (169)	TIPFAM_L (170)	TIPFAM_L (171)	TIPFAM_L (172)	TIPFAM_L (173)	TIPFAM_L (174)	TIPFAM_L (175)	TIPFAM_L (176)	TIPFAM_L (177)	TIPFAM_L (178)	TIPFAM_L (179)	TIPFAM_L (180)	TIPFAM_L (181)	TIPFAM_L (182)	TIPFAM_L (183)	TIPFAM_L (184)	TIPFAM_L (185)	TIPFAM_L (186)	TIPFAM_L (187)	TIPFAM_L (188)	TIPFAM_L (189)	TIPFAM_L (190)	TIPFAM_L (191)	TIPFAM_L (192)	TIPFAM_L (193)	TIPFAM_L (194)	TIPFAM_L (195)	TIPFAM_L (196)	TIPFAM_L (197)	TIPFAM_L (198)	TIPFAM_L (199)	TIPFAM_L (200)	TIPFAM_L (201)	TIPFAM_L (202)	TIPFAM_L (203)	TIPFAM_L (204)	TIPFAM_L (205)	TIPFAM_L (206)	TIPFAM_L (207)	TIPFAM_L (208)	TIPFAM_L (209)	TIPFAM_L (210)	TIPFAM_L (211)	TIPFAM_L (212)	TIPFAM_L (213)	TIPFAM_L (214)	TIPFAM_L (215)	TIPFAM_L (216)	TIPFAM_L (217)	TIPFAM_L (218)	TIPFAM_L (219)	TIPFAM_L (220)	TIPFAM_L (221)	TIPFAM_L (222)	TIPFAM_L (223)	TIPFAM_L (224)	TIPFAM_L (225)	TIPFAM_L (226)	TIPFAM_L (227)	TIPFAM_L (228)	TIPFAM_L (229)	TIPFAM_L (230)	TIPFAM_L (231)	TIPFAM_L (232)	TIPFAM_L (233)	TIPFAM_L (234)	TIPFAM_L (235)	TIPFAM_L (236)	TIPFAM_L (237)	TIPFAM_L (238)	TIPFAM_L (239)	TIPFAM_L (240)	TIPFAM_L (241)	TIPFAM_L (242)	TIPFAM_L (243)	TIPFAM_L (244)	TIPFAM_L (245)	TIPFAM_L (246)	TIPFAM_L (247)	TIPFAM_L (248)	TIPFAM_L (249)	TIPFAM_L (250)	TIPFAM_L (251)	TIPFAM_L (252)	TIPFAM_L (253)	TIPFAM_L (254)	TIPFAM_L (255)	TIPFAM_L (256)	TIPFAM_L (257)	TIPFAM_L (258)	TIPFAM_L (259)	TIPFAM_L (260)	TIPFAM_L (261)	TIPFAM_L (262)	TIPFAM_L (263)	TIPFAM_L (264)	TIPFAM_L (265)	TIPFAM_L (266)	TIPFAM_L (267)	TIPFAM_L (268)	TIPFAM_L (269)	TIPFAM_L (270)	TIPFAM_L (271)	TIPFAM_L (272)	TIPFAM_L (273)	TIPFAM_L (274)	TIPFAM_L (275)	TIPFAM_L (276)	TIPFAM_L (277)	TIPFAM_L (278)	TIPFAM_L (279)	TIPFAM_L (280)	TIPFAM_L (281)	TIPFAM_L (282)	TIPFAM_L (283)	TIPFAM_L (284)	TIPFAM_L (285)	TIPFAM_L (286)	TIPFAM_L (287)	TIPFAM_L (288)	TIPFAM_L (289)	TIPFAM_L (290)	TIPFAM_L (291)	TIPFAM_L (292)	TIPFAM_L (293)	TIPFAM_L (294)	TIPFAM_L (295)	TIPFAM_L (296)	TIPFAM_L (297)	TIPFAM_L (298)	TIPFAM_L (299)	TIPFAM_L (300)	TIPFAM_L (301)	TIPFAM_L (302)	TIPFAM_L (303)	TIPFAM_L (304)	TIPFAM_L (305)	TIPFAM_L (306)	TIPFAM_L (307)	TIPFAM_L (308)	TIPFAM_L (309)	TIPFAM_L (310)	TIPFAM_L (311)	TIPFAM_L (312)	TIPFAM_L (313)	TIPFAM_L (314)	TIPFAM_L (315)	TIPFAM_L (316)	TIPFAM_L (317)	TIPFAM_L (318)	TIPFAM_L (319)	TIPFAM_L (320)	TIPFAM_L (321)	TIPFAM_L (322)	TIPFAM_L (323)	TIPFAM_L (324)	TIPFAM_L (325)	TIPFAM_L (326)	TIPFAM_L (327)	TIPFAM_L (328)	TIPFAM_L (329)	TIPFAM_L (330)	TIPFAM_L (331)	TIPFAM_L (332)	TIPFAM_L (333)	TIPFAM_L (334)	TIPFAM_L (335)	TIPFAM_L (336)	TIPFAM_L (337)	TIPFAM_L (338)	TIPFAM_L (339)	TIPFAM_L (340)	TIPFAM_L (341)	TIPFAM_L (342)	TIPFAM_L (343)	TIPFAM_L (344)	TIPFAM_L (345)	TIPFAM_L (346)	TIPFAM_L (347)	TIPFAM_L (348)	TIPFAM_L (349)	TIPFAM_L (350)	TIPFAM_L (351)	TIPFAM_L (352)	TIPFAM_L (353)	TIPFAM_L (354)	TIPFAM_L (355)	TIPFAM_L (356)	TIPFAM_L (357)	TIPFAM_L (358)	TIPFAM_L (359)	TIPFAM_L (360)	TIPFAM_L (361)	TIPFAM_L (362)	TIPFAM_L (363)	TIPFAM_L (364)	TIPFAM_L (365)	TIPFAM_L (366)	TIPFAM_L (367)	TIPFAM_L (368)	TIPFAM_L (369)	TIPFAM_L (370)	TIPFAM_L (371)	TIPFAM_L (372)	TIPFAM_L (373)	TIPFAM_L (374)	TIPFAM_L (375)	TIPFAM_L (376)	TIPFAM_L (377)	TIPFAM_L (378)	TIPFAM_L (379)	TIPFAM_L (380)	TIPFAM_L (381)	TIPFAM_L (382)	TIPFAM_L (383)	TIPFAM_L (384)	TIPFAM_L (385)	TIPFAM_L (386)	TIPFAM_L (387)	TIPFAM_L (388)	TIPFAM_L (389)	TIPFAM_L (390)	TIPFAM_L (391)	TIPFAM_L (392)	TIPFAM_L (393)	TIPFAM_L (394)	TIPFAM_L (395)	TIPFAM_L (396)	TIPFAM_L (397)	TIPFAM_L (398)	TIPFAM_L (399)	TIPFAM_L (400)	TIPFAM_L (401)	TIPFAM_L (402)	TIPFAM_L (403)	TIPFAM_L (404)	TIPFAM_L (405)	TIPFAM_L (406)	TIPFAM_L (407)	TIPFAM_L (408)	TIPFAM_L (409)	TIPFAM_L (410)	TIPFAM_L (411)	TIPFAM_L (412)	TIPFAM_L (413)	TIPFAM_L (414)	TIPFAM_L (415)	TIPFAM_L (416)	TIPFAM_L (417)	TIPFAM_L (418)	TIPFAM_L (419)	TIPFAM_L (420)	TIPFAM_L (421)	TIPFAM_L (422)	TIPFAM_L (423)	TIPFAM_L (424)	TIPFAM_L (425)	TIPFAM_L (426)	TIPFAM_L (427)	TIPFAM_L (428)	TIPFAM_L (429)	TIPFAM_L (430)	TIPFAM_L (431)	TIPFAM_L (432)	TIPFAM_L (433)	TIPFAM_L (434)	TIPFAM_L (435)	TIPFAM_L (436)	TIPFAM_L (437)	TIPFAM_L (438)	TIPFAM_L (439)	TIPFAM_L (440)	TIPFAM_L (441)	TIPFAM_L (442)	TIPFAM_L (443)	TIPFAM_L (444)	TIPFAM_L (445)	TIPFAM_L (446)	TIPFAM_L (447)	TIPFAM_L (448)	TIPFAM_L (449)	TIPFAM_L (450)	TIPFAM_L (451)	TIPFAM_L (452)	TIPFAM_L (453)	TIPFAM_L (454)	TIPFAM_L (455)	TIPFAM_L (456)	TIPFAM_L (457)	TIPFAM_L (458)	TIPFAM_L (459)	TIPFAM_L (460)	TIPFAM_L (461)	TIPFAM_L (462)	TIPFAM_L (463)	TIPFAM_L (464)	TIPFAM_L (465)	TIPFAM_L (466)	TIPFAM_L (467)	TIPFAM_L (468)	TIPFAM_L (469)	TIPFAM_L (470)	TIPFAM_L (471)	TIPFAM_L (472)	TIPFAM_L (473)	TIPFAM_L (474)	TIPFAM_L (475)	TIPFAM_L (476)	TIPFAM_L (477)	TIPFAM_L (478)	TIPFAM_L (479)	TIPFAM_L (480)	TIPFAM_L (481)	TIPFAM_L (482)	TIPFAM_L (483)	TIPFAM_L (484)	TIPFAM_L (485)	TIPFAM_L (486)	TIPFAM_L (487)	TIPFAM_L (488)	TIPFAM_L (489)	TIPFAM_L (490)	TIPFAM_L (491)	TIPFAM_L (492)	TIPFAM_L (493)	TIPFAM_L (494)	TIPFAM_L (495)	TIPFAM_L (496)	TIPFAM_L (497)	TIPFAM_L (498)	TIPFAM_L (499)	TIPFAM_L (500)	TIPFAM_L (501)	TIPFAM_L (502)	TIPFAM_L (503)	TIPFAM_L (504)	TIPFAM_L (505)	TIPFAM_L (506)	TIPFAM_L (507)	TIPFAM_L (508)	TIPFAM_L (509)	TIPFAM_L (510)	TIPFAM_L (511)	TIPFAM_L (512)	TIPFAM_L (513)	TIPFAM_L (514)	TIPFAM_L (515)	TIPFAM_L (516)	TIPFAM_L (517)	TIPFAM_L (518)	TIPFAM_L (519)	TIPFAM_L (520)	TIPFAM_L (521)	TIPFAM_L (522)	TIPFAM_L (523)	TIPFAM_L (524)	TIPFAM_L (525)	TIPFAM_L (526)	TIPFAM_L (527)	TIPFAM_L (528)	TIPFAM_L (529)	TIPFAM_L (530)	TIPFAM_L (531)	TIPFAM_L (532)	TIPFAM_L (533)	TIPFAM_L (534)	TIPFAM_L (535)	TIPFAM_L (536)	TIPFAM_L (537)	TIPFAM_L (538)	TIPFAM_L (539)	TIPFAM_L (540)	TIPFAM_L (541)	TIPFAM_L (542)	TIPFAM_L (543)	TIPFAM_L (544)	TIPFAM_L (545)	TIPFAM_L (546)	TIPFAM_L (547)	TIPFAM_L (548)	TIPFAM_L (549)	TIPFAM_L (550)	TIPFAM_L (551)	TIPFAM_L (552)	TIPFAM_L (553)	TIPFAM_L (554)	TIPFAM_L (555)	TIPFAM_L (556)	TIPFAM_L (557)	TIPFAM_L (558)	TIPFAM_L (559)	TIPFAM_L (560)	TIPFAM_L (561)	TIPFAM_L (562)	TIPFAM_L (563)	TIPFAM_L (564)	TIPFAM_L (565)	TIPFAM_L (566)	TIPFAM_L (567)	TIPFAM_L (568)	TIPFAM_L (569)	TIPFAM_L (570)	TIPFAM_L (571)	TIPFAM_L (572)	TIPFAM_L (573)	TIPFAM_L (574)	TIPFAM_L (575)	TIPFAM_L (576)	TIPFAM_L (577)	TIPFAM_L (578)	TIPFAM_L (579)	TIPFAM_L (580)	TIPFAM_L (581)	TIPFAM_L (582)	TIPFAM_L (583)	TIPFAM_L (584)	TIPFAM_L (585)	TIPFAM_L (586)	TIPFAM_L (587)	TIPFAM_L (588)	TIPFAM_L (589)	TIPFAM_L (590)	TIPFAM_L (591)	TIPFAM_L (592)	TIPFAM_L (593)	TIPFAM_L (594)	TIPFAM_L (595)	TIPFAM_L (596)	TIPFAM_L (597)	TIPFAM_L (598)	TIPFAM_L (599)	TIPFAM_L (600)	TIPFAM_L (601)	TIPFAM_L (602)	TIPFAM_L (603)	TIPFAM_L (604)	TIPFAM_L (605)	TIPFAM_L (606)	TIPFAM_L (607)	TIPFAM_L (608)	TIPFAM_L (609)	TIPFAM_L (610)	TIPFAM_L (611)	TIPFAM_L (612)	TIPFAM_L (613)	TIPFAM_L (614)	TIPFAM_L (615)	TIPFAM_L (616)	TIPFAM_L (617)	TIPFAM_L (618)	TIPFAM_L (619)	TIPFAM_L (620)	TIPFAM_L (621)	TIPFAM_L (622)	TIPFAM_L (623)	TIPFAM_L (624)	TIPFAM_L (625)	TIPFAM_L (626)	TIPFAM_L (627)	TIPFAM_L (628)	TIPFAM_L (629)	TIPFAM_L (630)	TIPFAM_L (631)	TIPFAM_L (632)	TIPFAM_L (633)	TIPFAM_L (634)	TIPFAM_L (635)	TIPFAM_L (636)	TIPFAM_L (637)	TIPFAM_L (638)	TIPFAM_L (639)	TIPFAM_L (640)	TIPFAM_L (641)	TIPFAM_L (642)	TIPFAM_L (643)	TIPFAM_L (644)	TIPFAM_L (645)	TIPFAM_L (646)	TIPFAM_L (647)	TIPFAM_L (648)	TIPFAM_L (649)	TIPFAM_L (650)	TIPFAM_L (651)	TIPFAM_L (652)	TIPFAM_L (653)	TIPFAM_L (654)	TIPFAM_L (655)	TIPFAM_L (656)	TIPFAM_L (657)	TIPFAM_L (658)	TIPFAM_L (659)	TIPFAM_L (660)	TIPFAM_L (661)	TIPFAM_L (662)	TIPFAM_L (663)	TIPFAM_L (664)	TIPFAM_L (665)	TIPFAM_L (666)	TIPFAM_L (667)	TIPFAM_L (668)	TIPFAM_L (669)	TIPFAM_L (670)	TIPFAM_L (671)	TIPFAM_L (672)	TIPFAM_L (673)	TIPFAM_L (674)	TIPFAM_L (675)	TIPFAM_L (676)	TIPFAM_L (677)	TIPFAM_L (678)	TIPFAM_L (679)	TIPFAM_L (680)	TIPFAM_L (681)	TIPFAM_L (682)	TIPFAM_L (683)	TIPFAM_L (684)	TIPFAM_L (685)	TIPFAM_L (686)	TIPFAM_L (687)	TIPFAM_L (688)	TIPFAM_L (689)	TIPFAM_L (690)	TIPFAM_L (691)	TIPFAM_L (692)	TIPFAM_L (693)	TIPFAM_L (694)	TIPFAM_L (695)	TIPFAM_L (696)	TIPFAM_L (697)	TIPFAM_L (698)	TIPFAM_L (699)	TIPFAM_L (700)	TIPFAM_L (701)	TIPFAM_L (702)	TIPFAM_L (703)	TIPFAM_L (704)	TIPFAM_L (705)	TIPFAM_L (706)	TIPFAM_L (707)	TIPFAM_L (708)	TIPFAM_L (709)	TIPFAM_L (710)	TIPFAM_L (711)	TIPFAM_L (712)	TIPFAM_L (713)	TIPFAM_L (714)	TIPFAM_L (715)	TIPFAM_L (716)	TIPFAM_L (717)	TIPFAM_L (718)	TIPFAM_L (719)	TIPFAM_L (720)	TIPFAM_L (721)	TIPFAM_L (722)	TIPFAM_L (723)

Figura 10.23 - Famiglie per numero di componenti e situazione abitativa della famiglia - visualizzazione on line

Caratteristiche delle famiglie <sup>1</sup> : Famiglie per situazione abitativa

Personalizza Esportazioni Grafici La tua interrogazione

Page: 1 / 2

Tipo dato		numero di famiglie (valori assoluti)	
Tipologia della famiglia	famiglie		
Sesso della persona di riferimento della famiglia	totale		
Classe di età della persona di riferimento della famiglia	totale		
Stato civile della persona di riferimento della famiglia	totale		
Numero di componenti	totale		
Condizione abitativa (in/non in coabitazione)	tutte le voci		
Specie di alloggio	tutte le voci		
Anno di Censimento	2011		
Situazione abitativa della famiglia		famiglia in abitazione	
		famiglia in altro tipo di alloggio	
		famiglia in struttura residenziale collettiva	
		famiglia coabitante	
		famiglia senza tetto	
		tutte le voci	
Titolo di godimento		proprietà	
		affitto	
		altro titolo diverso da proprietà, affitto	
		tutte le voci	
Territorio		Italia	
		Nord-ovest	
		Piemonte	
		Torino	
		VerCELLI	
		Novara	
		Cuneo	

### 10.5.4 Caso 3: Stanze in abitazioni occupate da persone residenti per numero di stanze e numero di stanze ad uso professionale – livello provinciale

Si consideri, infine, l'incrocio del Piano di diffusione che riguarda le *Stanze in abitazioni occupate da persone residenti* (figura 10.24) con il dettaglio delle sue classificazioni e del livello territoriale. Tale incrocio è stato descritto all'interno del cubo dati *Stanze in abitazioni occupate da persone residenti* all'interno del quale sono stati fatti confluire tutti gli incroci attinenti tale argomento (primo passo di modellazione).

Figura 10.24 - Stanze in abitazioni occupate da persone residenti - Piano delle diffusioni

N. Incrocio	Oggetto	Dettaglio territoriale	Dimensione di analisi	Numero di stanze (7)		Numero di stanze ad uso professionale (6)	
A3.52	Stanze in abitazioni occupate da persone residenti	Provinciale	Numero di stanze (7) X Numero di stanze ad uso professionale (6)	Totale		Totale	
				1		Senza stanze ad uso professionale	
				2		Con stanze ad uso professionale	
				3			di cui: 1
				4			2
				5			3 e più
				6 o più			

A partire da queste due dimensioni di analisi si sono individuate e gestite, sul data warehouse, tre *variabili di diffusione* e due *Code List* di cui una già definita in precedenza (figura 10.25):

- ▶ *presenza stanze ad uso professionale* - precisa se vi siano o meno stanze ad uso professionale e che utilizza la Code List *Si\_NO* (SI\_NO);
- ▶ *numero di stanze* - specifica la numerosità delle stanze presenti e che utilizza la code list *Numerosità* (NUMEROSITA);
- ▶ *numero di stanze ad uso professionale* – per descrivere il numero di stanze specificatamente adoperate con un uso di carattere professionale e che utilizza nuovamente la code list *Numerosità* (NUMEROSITA).

È da precisare che questo esempio, avendo suddiviso i concetti in dimensioni differenti presenta, come per i casi precedenti, la necessità di esplicitare le dipendenze funzionali generate. Inoltre, è importante notare che le dimensioni *Numero di stanze* e *Numero di stanze ad uso professionale*, pur utilizzando la medesima Code List – *Numerosità* - ne utilizzano modalità differenti e che tali modalità sono ulteriormente diverse da quelle utilizzate negli esempi precedenti. Poiché, però per



**Figura 10.27 - Stanze in abitazioni occupate da persone residenti – visualizzazione on line**

Stanze in abitazioni occupate da persone residenti <sup>1</sup> : Stanze in abitazioni occupate da persone residenti per presenza di stanze ad uso professionale, numero di stanze ad uso professionale e numero di stanze

Personalizza Esportazioni Grafici La tua interrogazione

Territorio: Italia

Figura giuridica del proprietario: tutte le voci

Titolo di godimento: tutte le voci

Numero di famiglie: totale

Classi di superficie dell'abitazione (mq): tutte le voci

Tipo di servizi: tutte le voci

Epoca di costruzione: tutte le voci

Anno di Censimento: 2011

Tipo dato: numero di stanze in abitazioni occupate da persone residenti (valori assoluti)

	numero di stanze in abitazioni occupate da persone residenti (valori assoluti)					totale
	no	1	2	3 e più		
Presenza di stanze ad uso professionale						
Numero di stanze ad uso professionale	▲▼	▲▼	▲▼	▲▼	▲▼	▲▼
Numero di stanze						
1	488 463					488 463
2	4 906 213	44 391			44 391	4 949 605
3	14 711 856	187 238	10 159		197 396	14 909 252
4	29 136 761	451 733	33 851	22 485	508 070	29 644 831
5	24 977 629	661 986	42 775	136 604	841 366	25 818 995
6 e più	24 815 623	1 357 257	239 888	273 835	1 870 980	26 686 603
totale	99 035 546	2 702 606	326 673	432 925	3 462 204	102 497 749

### 10.5.5 Considerazioni conclusive

Si conclude dicendo che, rispetto agli esempi illustrati, a partire da cinque dimensioni di analisi:

- ▶ numero di componenti(7);
- ▶ numero di componenti(8);
- ▶ situazione abitativa della famiglia;
- ▶ numero di stanze;
- ▶ numero di stanze ad uso professionale;

e da cinque Classificazioni:

- ▶ numero componenti (7 modalità);
- ▶ numero componenti (8 modalità);
- ▶ situazione abitativa;
- ▶ numero di stanze (7 modalità);
- ▶ numero di stanze ad uso professionale (8 modalità);

il processo di concettualizzazione e progettazione cubi ha consentito l'individuazione esatta degli assi di analisi del fenomeno mediante la definizione di sei variabili di classificazione:

- ▶ numero di componenti;
- ▶ condizione abitativa della famiglia;
- ▶ situazione abitativa della famiglia;
- ▶ titolo di godimento (del bene);
- ▶ numero di stanze;
- ▶ presenza stanze ad uso professionale;
- ▶ numero di stanze ad uso professionale;

e la creazione di cinque Code List integrate e riusabili

- ▶ numerosità;
- ▶ condizione abitativa;
- ▶ abitazione familiare;
- ▶ titolo di godimento;
- ▶ Si\_No.

## 10.6 Il data warehouse di diffusione

La piattaforma utilizzata per data Warehouse di diffusione dei macrodati si basa sulla tecnologia .Stat sviluppata dall'Oecd alla cui evoluzione partecipa, in sede internazionale, anche l'Istat. In occasione dei censimenti però si sono implementate ed utilizzate alcune funzionalità aggiuntive, nonché una nuova architettura tecnologica.

### 10.6.1 L'architettura tecnologica del data warehouse di diffusione<sup>27</sup>

Da un punto di vista architettonico e del software di base (figura 10.28), sinteticamente, l'applicazione è costituita: da un unico data base server, collocato in Dmz, che funge sia da data base interno che esterno; da un web server interno per la consultazione dei dati in ambiente Intranet; da due web server esterni *in balancing* per la consultazione dei dati in ambiente Internet, configurati in maniera tale da garantire l'eventuale scalabilità nel tempo; dallo storage. La pubblicazione dei dati avviene grazie ad un sistema di permessi basato sul sistema di autorizzazione di Windows *Active Directory*, mentre la possibilità offerta agli utenti di registrarsi al fine di salvare le proprie tavole ed interrogazioni, è realizzata mediante il collegamento al sistema di *Single Sign-on* (Sso) dell'Istat. Le funzioni di back-up lato db sono state impostate in modalità giornaliera, *a caldo* e *differenziale* a cui si aggiunge un back-up *settimanale* di tipo full; quelle lato software di base è stato realizzato tramite gli *snapshot* dei server; quelle dello storage mediante le funzionalità della tecnologia *Tivoli Storage Management* (Tsm).

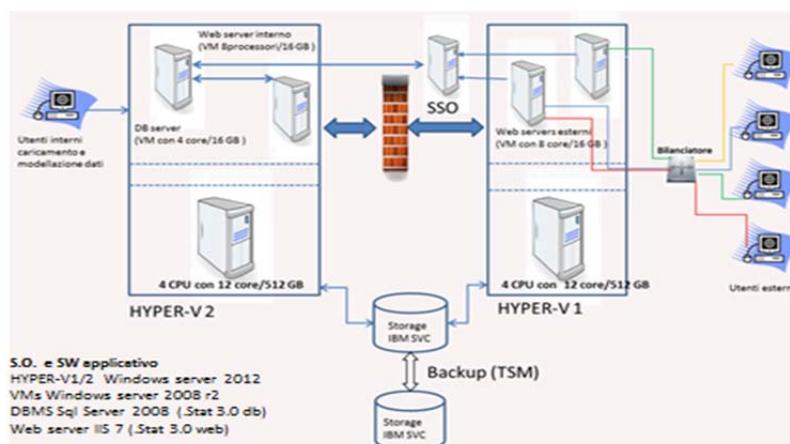
Il livello interno si basa su due macchine virtuali, una dedicata al database ed una all'interfaccia web di amministrazione e caricamento dei dati (Data Provider Interface) che utilizzano la Storage Area Network dell'Istituto per la memorizzazione dei dati.

La componente di navigazione Internet è costituita da due macchine virtuali, eventualmente scalabili in caso di accresciute necessità prestazionali, il cui carico è ripartito tramite bilanciatore.

Il sistema gestisce l'autenticazione degli utenti esterni integrandosi con il Single Sign On (Sso) e gli standard di sicurezza dell'Istituto.

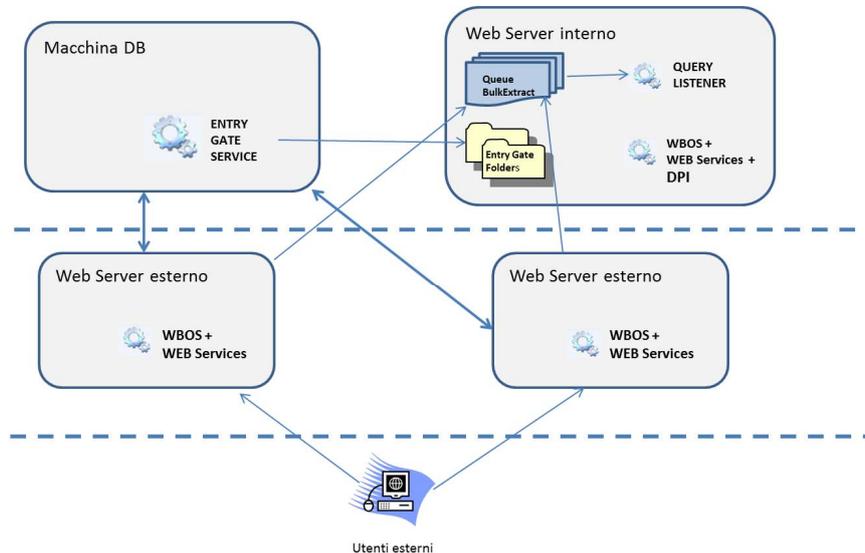
L'architettura software, invece, è sintetizzata nella figura 10.29.

Figura 10.28 - L'architettura tecnologica e software di base



<sup>27</sup> Hanno collaborato alla realizzazione dell'architettura: Gramigni Luca, Magarò Mario, Marchionne Francesco, Montefiori Emiliano, Sardone Angelo

Figura 10.29 - L'architettura software



Le principali innovazioni utilizzate per i Censimenti, rispetto all'impianto realizzato per il data warehouse di diffusione dei macrodati dell'Istat - *I.Stat*, hanno riguardato tre aspetti, migliorativi sia dell'interazione con l'utente che della gestione fisica dei sistemi. In particolare si è trattato:

- 1) dell'ottimizzazione di alcuni indici del db, al fine di innalzare le performance del sistema in fase di navigazione, in considerazione della mole di dati;
- 2) della virtualizzazione di tutti i server, al fine di consentire un più facile governo delle macchine fisiche ed una più veloce eventuale procedura di re-installazione in caso di disaster recovery;
- 3) della messa in esercizio di una funzionalità che gestisce l'associazione dei cubi ad uno o più ambienti di consultazione. In tal modo, è stato possibile utilizzare il db realizzato per il Censimento dell'Industria e Servizi, nonché il software per il caricamento dei dati e la loro navigazione in ambiente Intranet, anche per i dati del Censimento della Popolazione ed Abitazioni. Collegando ciascun specifico cubo al contesto di riferimento è stato contestualmente possibile realizzare due distinti ambienti di navigazione Internet con una ulteriore ottimizzazione delle macchine e delle procedure di gestione dei sistemi.

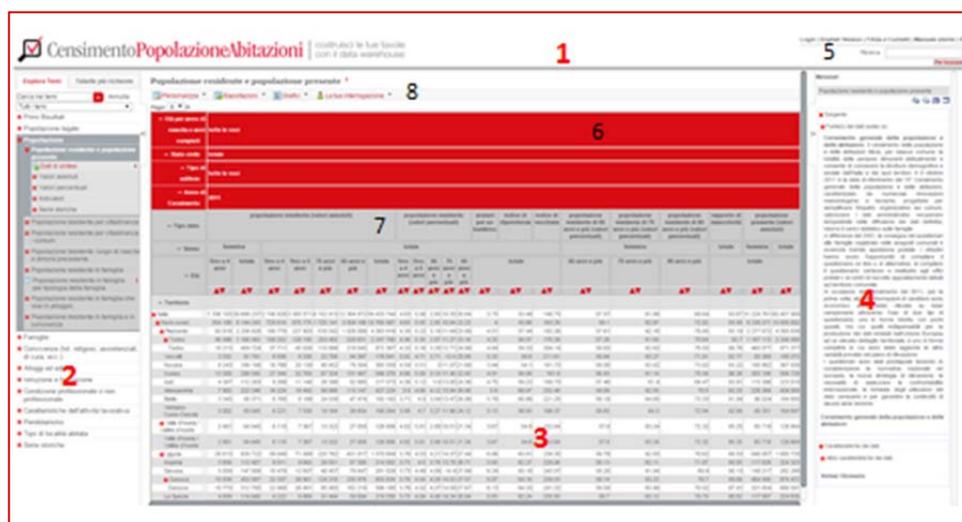
### 10.6.2 L'ambiente di navigazione

L'ambiente di navigazione presenta, come già anticipato, l'interfaccia web del *data warehouse* di Istituto (I.Stat). In occasione della tornata censuaria si è reso necessario lavorare sul miglioramento della sezione dedicata alla selezione dei contenuti da parte dell'utenza: l'albero dei temi (figura 10.30 – sezione 2).

Tale sezione è organizzata nativamente in temi e sotto-temi, ed all'interno di questi vengono presentati i differenti cubi dati messi a disposizione. Al fine di migliorare l'interazione con l'utenza ed offrirle una immediata comprensione delle informazioni contenute all'interno di ciascun cubo, per ciascuno di essi, si sono predisposte una serie di visualizzazioni (*query*) pre-costituite. La numerosità delle medesime ha imposto l'esigenza di organizzarle secondo una logica ad indice. È stato, dunque necessario mettere in esercizio una funzionalità che offrisse la possibilità di raggruppare tali *query* in argomenti e sotto-argomenti, all'interno del singolo cubo dati. Tale

funzionalità è stata utilizzata anche per ridurre i titoli delle singole visualizzazioni rendendole più leggibili e immediatamente comprensibili per l'utenza.

Figura. 10.30 - L'interfaccia web e la sua divisione in sezioni



L'interfaccia è suddivisa in quattro sezioni: quella superiore (figura 10.30 - sezione 1), la parte dell'albero dei temi a sinistra (figura 10.30 - sezione 2), la parte dei dati al centro (figura 10.30 - sezione 3) e la parte con i metadati a destra (figura 10.30 - sezione 4); l'albero dei temi ed i metadati sono dei pannelli a scomparsa che consentono all'utente di visualizzare i dati a schermo intero.

La parte superiore presenta delle funzioni per il controllo dell'account dell'utente (è infatti possibile registrarsi per salvare in remoto le proprie query) e per ricevere assistenza in modo immediato, o attraverso la guida d'uso ed una lista di domande frequenti, oppure contattando direttamente gli amministratori del sito web compilando un'apposita form. Viene anche data la possibilità di ricevere delle indicazioni per iniziare a lavorare, attraverso una finestra di dialogo introduttiva. È possibile passare alla versione del sito in inglese e viceversa, grazie ad un apposito collegamento ipertestuale (figura 10.30 - sezione 5). Tutto il sistema, sia sul fronte dei contenuti che delle interfacce di navigazione è, infatti, in doppia lingua.

Sempre nella stessa sezione, appare una casella per la ricerca del testo tra i contenuti del sito, quali dati, etichette e metadati.

Nella parte sinistra i temi possono essere suddivisi in sottotemi, gruppi e query. È la sezione che organizza le informazioni presenti nel sistema..

In questa sezione è presente una casella di ricerca per facilitare la selezione delle informazioni mediante il titolo dei cubi dati messi a disposizione.

La sezione centrale, quella più estesa, è dedicata alla visualizzazione dei dati.

La sua parte superiore è suddivisa in una tre sotto-sezioni: un insieme di dimensioni selezionabili (figura 10.30 - sezione 6) la tavola dati con la sua testata ( figura 10.30 - sezione 7) e i menù per consentire all'utente la personalizzazione della tavola medesima, l'export in diversi formati, il salvataggio delle proprie interrogazioni (figura 10.30 - sezione 8).

Nel pannello di destra (figura 10.30 - sezione 4) vengono visualizzati i metadati associati direttamente al dataset, mentre con un collegamento posto direttamente sulle celle dei dati o sulle modalità è possibile visualizzare i metadati a loro associati. Tale sezione è dettagliata nel paragrafo successivo.

Sinteticamente, nelle toolbar della sottozione 8 è offerta all'utente la possibilità di:

- ▶ disporre gli assi (figura 10.31 – sezione 9);
- ▶ impostare alcune opzioni di visualizzazione (figura 10.31 – sezione 10);
- ▶ esportare i dati in differenti formati (csv, xls, ...) (figura 10.31 – sezione 11);
- ▶ realizzare grafici (figura 10.31 – sezione 12);
- ▶ gestire le interrogazioni ad esempio salvandole o combinando dati (figura. 10.31 – sezione 13).

Figura 10.31 - Il menu della toolbar



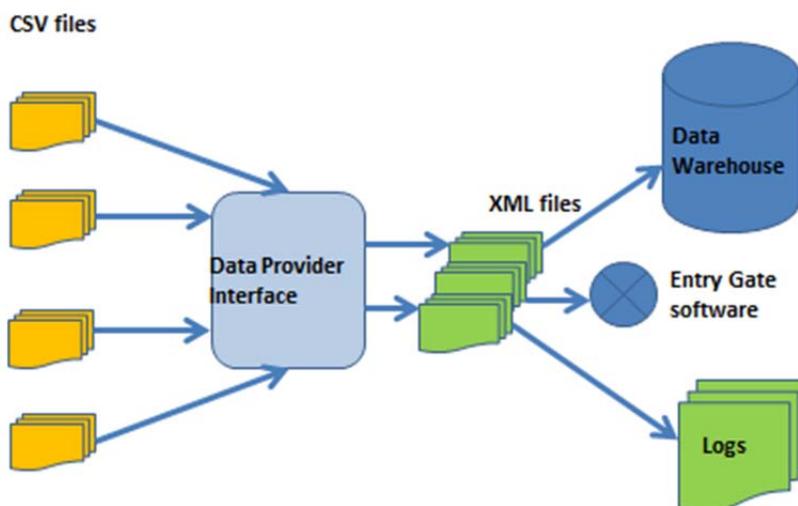
Il sistema di diffusione, oltre ai macrodati quantitativi, mette a disposizione un ampio set di metadati che accompagnano ogni cubo. La loro presenza viene evidenziata tramite la visualizzazione sull'interfaccia del simbolo “i” alla cui pressione si apre una opportuna pop up. I metadati sono di tipo referenziale, per la descrizione della fonte dei dati, e, qualora necessario, di tipo strutturale per la descrizione del contenuto delle singole celle e/o delle modalità delle dimensioni di analisi.

### 10.6.3 L'ambiente di amministrazione ed il caricamento dei dati

Le attività di caricamento dati sono avvenute importando files csv mediante l'ausilio di un ambiente intranet di amministrazione denominato *Data Provider Interface*. La Data Provider Interface complessivamente, consente la gestione (creazione e caricamento) di tutte le strutture necessarie alla navigazione (dataset, dimensioni di analisi, metadati).

La Data Provider Interface per caricare i dati nel data warehouse genera files in formato xml che referenziano i files csv. Tali file xml vengono presi in carica da una componente software che agisce da manager, l'Entry Gate Service (figura 10.32). Tale componente gestisce grandi moli di dati e la concorrenzialità degli utenti amministratori tramite la suddivisione in blocchi di 100 mila righe dei files csv e l'utilizzo delle tecnologia delle code.

Figura 10.32 - L'architettura del sistema di caricamento dati



## 10.7 - Il Single Exit Point per la diffusione dati machine-to-machine

Il Single Exit Point per il Censimento della Popolazione è stato implementato per:

- ▶ la diffusione via web service in formato Sdmx dei dati del Censimento della Popolazione memorizzati all'interno del corporate datawarehouse dei Censimenti;
- ▶ la resa in disponibilità, tramite una specifica istanza del servizio connessa al sistema Census Hub di Eurostat, dei dati del censimento Popolazione italiana nell'ambito del Sistema Statistico Europeo.

### 10.7.1 L'architettura tecnologica ed il workflow

Il sistema si basa sulla piattaforma Sdmx Reference Infrastructure (Sdmx-RI) prodotta da Eurostat che permette di rendere disponibili i dati in formato Sdmx via web service, senza alcuna trasformazione preventiva ma utilizzando a runtime le "mappature" delle strutture dati (colonne delle tabelle o delle viste di database) sulle relative Sdmx Data Structure Definition o Dsd, i cui componenti sono catalogati da un punto di vista semantico in dimensioni, attributi e misure. Lo strumento per effettuare le mappature è il tool Mapping Assistant Il web service Sdmx accetta query di tipo Soap o Rest.

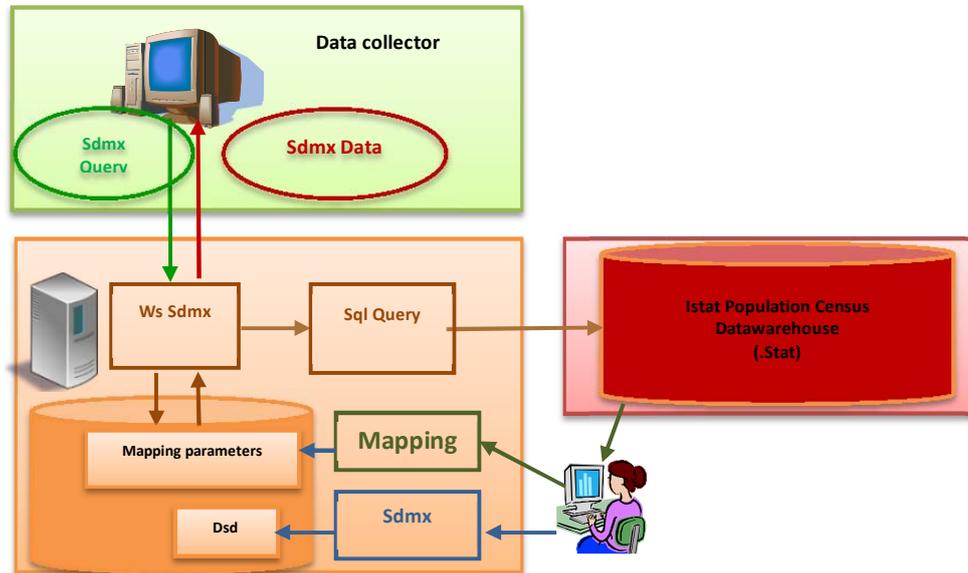
L'ambiente tecnologico server su cui è impiantata la piattaforma Sdmx-RI è il seguente:

- ▶ sistema operativo Windows Server: 2012;
- ▶ web Server: IIS8;
- ▶ database server (per il database contenente le mappature e le Dsd Sdmx): Microsoft Sql Server 2012.

Il Sep per la diffusione dei dati del Censimento Popolazione rende disponibili i dati memorizzati nel datawarehouse dei Censimenti dell'Istat che è implementato in tecnologia .Stat (prodotta da Oecd). Il flusso di lavoro per rendere disponibili i dati in formato Sdmx in questo caso ha comportato le fasi di modellazione (creazione delle Dsd) e mappatura che sono descritte nel successivo paragrafo.

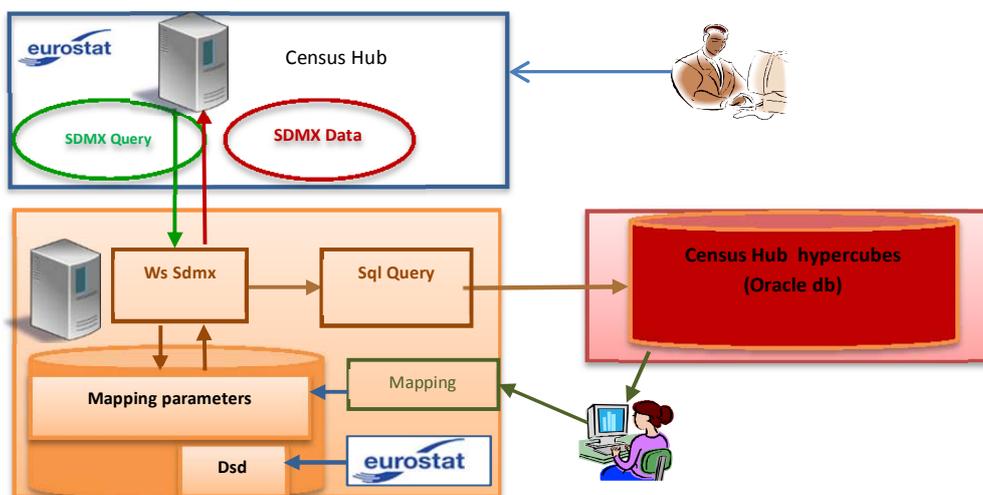
Lo schema dell'architettura, con il flusso funzionale, è riportato all'interno delle figura 10.33.

Figura 10.33 - Architettura funzionale del Sep per il Censimento Popolazione



L'istanza Sep per il Census Hub di Eurostat rende disponibili a runtime a tale sistema web i dati del Censimento della Popolazione italiana contenuti in ipercubi tematici memorizzati in un database Oracle e generati a partire dal Datawarehouse primario: in questo caso il flusso di lavoro per la resa in disponibilità dei dati tramite la Sdmx-Ri ha comportato esclusivamente la fase di mappatura, mentre la modellazione Sdmx non è risultata necessaria in quanto gli artefatti Sdmx (Dsd, Codelist, ConceptScheme, CategoryScheme, Dataflow) sono stati forniti direttamente da Eurostat e, tranne le codelist nazionali del territorio, sono gli stessi per tutti i paesi membri della UE.

Figura 10.34 - Architettura funzionale del Sep per il Census Hub



### 10.7.2 La modellazione dei cubi e la mappatura Sdmx

Il processo attraverso il quale i dati del Censimento della Popolazione e delle Abitazioni sono resi disponibili alla diffusione machine-to-machine in formato Sdmx consta di due fasi:

- ▶ la *modellazione Sdmx* dei cubi di dati da cui si generano le strutture dati Sdmx (*Data Structure Definition* o Dsd):
- ▶ la *mappatura* dei dati del datawarehouse sulle strutture dati Sdmx precedentemente modellate

La *modellazione* ovvero la creazione della struttura dati (Dsd) nello standard Sdmx avviene tramite un software ad hoc "*Data Structure Wizard*" prodotto da Eurostat. A partire dai piani di spoglio, per ogni cubo sono stati individuati tutti i componenti della struttura dati, ciascuno associato in Sdmx a un concetto (*Concept*). I Concetti, classificati in appositi contenitori denominati *ConceptScheme*, sono associati ai componenti delle *strutture* dati Sdmx. I componenti della Dsd sono costituiti da: *Dimensioni*, *Attributi* e *Misure*. Le *Dimensioni* sono combinazioni di valori che identificano univocamente le singole osservazioni; gli *Attributi* descrivono le caratteristiche dell'osservazione; le *Misure* rappresentano i valori osservati. Una volta che vengono definiti i componenti della *Data Structure Definition* e poi associati ai relativi concetti, la modellazione dei singoli cubi è completata. La *mappatura Sdmx* dei cubi riguarda l'associazione delle strutture dati Sdmx (Dsd) con gli ipercubi del Dw .Stat del Censimento della Popolazione e delle Abitazioni. Per effettuarla si è utilizzato il tool "*Mapping Assistant*" (appartenente alla Sdmx-Ri prodotta da Eurostat). Le fasi di predisposizione della mappatura riguardano la creazione del *CategoryScheme*, ovvero dell'albero gerarchico dei temi e sottotemi in formato Sdmx corrispondente a quello definito per il Dw .Stat del Censimento e la definizione dei set di dati denominati *DataFlow*. Ogni dataflow, riferito a una determinata Dsd, è stato associato a una opportuna *category* del *CategoryScheme*. Per ogni cubo di dati del Dw .Stat del Censimento della Popolazione e delle Abitazioni è stato definito e memorizzato nel Mapping Store il *Dataset* corrispondente, ovvero la struttura dati locale da mappare sulla corrispondente Dsd Sdmx. Si è quindi proceduto a mappare ciascun *Dataset* sulla corrispondente Dsd, mettendo in relazione tra loro i campi del dataset con i componenti della Dsd. Per ogni dimensione o attributo codificato sono stati messi in relazione i codici locali utilizzati all'interno del cubo con i codici associati alle relative *Codelist*: tale operazione è detta *Transcoding*. terminate le fasi appena descritte, la mappatura Sdmx è conclusa. L'intero processo è stato eseguito per ogni cubo di dati del Dw .Stat del Censimento della Popolazione e delle Abitazioni. Per rendere disponibili i dati sul sistema Census Hub di Eurostat tramite il servizio web del Sep, non è stato necessario svolgere la fase di modellazione Sdmx dei dati, in quanto gli artefatti Sdmx (*Dsd*, *Codelist*, *ConceptScheme*, *Dataflow*, *CategoryScheme*) sono stati forniti da Eurostat a tutti i Stati Membri del Sistema Statistico Europeo, come previsto dai Regolamenti Europei. È stata svolta da parte di Istat esclusivamente la fase di mappatura Sdmx dei cubi di dati da predisporre per il *Census Hub* generati tramite estrazioni dal Dw Primario.

### 10.8 La disponibilità di microdati per gli utenti esterni

L'Istat garantisce, nel rispetto della legislazione vigente, l'accesso e il rilascio dei microdati validati derivanti dalle proprie rilevazioni.

L'utilizzo dei microdati da parte di esterni può avvenire attraverso due modalità: quella prevista per gli enti Sistan e il Laboratorio Adele.

Per quanto riguarda il primo canale, il decreto legislativo n. 322 del 1989 ed il decreto legislativo n. 196 del 2003 prevedono l'interscambio di dati individuali all'interno del Sistema statistico nazionale, solo se necessario alle esigenze statistiche previste dal programma statistico nazionale oppure per consentire il perseguimento degli scopi istituzionali dell'ente di appartenenza.

Per gli utenti che necessitino di elaborazioni particolari per specifiche esigenze di ricerca, l'Istat mette a disposizione, anche il Laboratorio per l'Analisi dei Dati ELEMENTARI (Laboratorio Adele). Tale laboratorio, situato presso la sede centrale dell'Istat e le sue sedi regionali, è un Research Data Center (Rdc), cioè un luogo sicuro dove si possono recare ricercatori e studiosi di università, istituti o enti di ricerca per effettuare direttamente analisi statistiche sui microdati delle indagini dell'Istituto.

Tutti i microdati disponibili per gli utenti esterni, inclusi quelli dei censimenti, vengono archiviati centralmente nel sistema Armida (ARchivio di MIcroDATi) che viene utilizzato per rispondere alle richieste degli enti del Sistema statistico nazionale (Sistan) ed è accessibile presso il Laboratorio Adele.

Si fornisce di seguito una panoramica sintetica del sistema di diffusione dei microdati dell'Istat e la specifica soluzione adottata per il Censimento della popolazione e delle abitazioni 2011.

### **10.8.1 La comunicazione dei microdati agli enti Sistan**

La direttiva "Criteri e modalità per la comunicazione dei dati personali nell'ambito del Sistema statistico nazionale" del Comstat del 20 aprile 2004 (Direttiva n. 9/Comstat), prevede che un ente o ufficio di statistica facente parte del Sistan possa richiedere ad un altro soggetto del Sistema dati personali già acquisiti per finalità statistiche.

In base a tale direttiva l'Istat è tenuta a comunicare dati personali rilevati attraverso le proprie indagini agli enti Sistan che ne facciano richiesta, purché si tratti di richieste necessarie per la realizzazione delle rilevazioni, delle elaborazioni e degli studi progettuali compresi nel Programma statistico nazionale oppure per l'esecuzione di trattamenti per scopi statistici, strumentali al perseguimento delle finalità istituzionali del soggetto richiedente. Le forniture di dati personali corredati di identificativi è, comunque, limitata ai casi di assoluta e stretta necessità ovvero l'impossibilità di raggiungere l'obiettivo prefissato senza i dati identificativi.

Gli enti Sistan devono formulare la richiesta di dati attraverso il Cont@ctCentre dell'Istat, ovvero il sistema web per l'acquisizione ed il trattamento on line delle richieste di informazioni statistiche e dei servizi di diffusione (<http://contact.istat.it/>). La richiesta deve essere redatta su apposito modello ([http://www.istat.it/dati/microdati/file\\_microdati.html#file\\_sistan](http://www.istat.it/dati/microdati/file_microdati.html#file_sistan)) e contenere indicazioni dettagliate in ordine alle motivazioni, alle finalità perseguite e alla pertinenza e non eccedenza dei dati richiesti rispetto alle finalità dichiarate, nonché, qualora siano richiesti anche dati identificativi, la stretta necessità dei medesimi.

### **10.8.2 L'accesso ai microdati attraverso il Laboratorio Adele**

Il Laboratorio Adele è il Research Data Centre (Rdc) dell'Istituto Nazionale di Statistica, un luogo dedicato a ricercatori e studiosi per effettuare analisi statistiche su microdati nel rispetto delle attuali norme sulla riservatezza statistica. Esso si colloca, nel contesto internazionale, al fianco dei Rdc degli altri istituti nazionali di statistica che offrono, nei rispettivi Paesi, la stessa tipologia di servizio in modo sostanzialmente condiviso e comunque secondo un processo di armonizzazione, in particolare in ambito europeo.

Tramite il Laboratorio Adele, l'Istat mette a disposizione dei ricercatori i dati elementari delle rilevazioni di propria titolarità privati degli identificativi diretti (nome, cognome, codice fiscale etc.), senza alcuna variazione al contenuto informativo degli stessi.

Presso la sede centrale di Roma e le altre sedi territoriali dell'Istat, con la sola eccezione della provincia autonoma di Bolzano, sono a disposizione dei ricercatori postazioni di lavoro provviste di terminali, collegati ad una sottorete dedicata, che permettono l'accesso da remoto ai server della sede centrale, dove sono collocati i dati elementari delle indagini e dove avviene, effettivamente, l'elaborazione degli stessi.

L'accesso ai dati elementari per finalità di ricerca è consentito solo nel rispetto della normativa nazionale in materia di segreto statistico e di protezione dei dati personali, secondo le condizioni e le garanzie stabilite dall'art. 7 del "Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e di ricerca scientifica" effettuati nell'ambito del Sistema statistico nazionale" (Allegato A.3 al d.lgs. n. 196 del 2003). È necessario che il ricercatore presenti un progetto di ricerca e richieda l'autorizzazione all'accesso ai dati al Presidente dell'Istat, sottoscrivendo un contratto che lo obbliga al mantenimento del segreto statistico. Se l'autorizzazione viene concessa, il ricercatore può svolgere le elaborazioni sui dati richiesti, concluse le quali il risultato viene controllato sotto il profilo della tutela della riservatezza ed, eventualmente, rilasciato. L'accesso, in ogni caso, può essere concesso solo qualora non siano adeguati allo svolgimento della ricerca gli altri canali dell'Istituto, il servizio delle Elaborazioni Personalizzate e i file di microdati opportunamente rielaborati (filemicro.Stat, standard e per la ricerca). La tutela della riservatezza dei dati presso il Laboratorio Adele avviene quindi sotto tre punti di vista:

- ▶ fisico, poiché le postazioni dedicate al servizio sono collocate all'interno delle sedi dell'Istituto e l'accesso ad esse è vigilato e consentito ai soli utenti autorizzati. Le postazioni, inoltre, non permettono l'immissione o il prelievo di dati e sono isolate dalla rete Internet;
- ▶ legale, in quanto il ricercatore si impegna per iscritto, rendendosi passibile di denuncia alla competente autorità giudiziaria, a rispettare le norme in materia di segreto statistico e di protezione dei dati personali;
- ▶ della riservatezza statistica, a causa del controllo al quale sono sottoposti i risultati delle elaborazioni del ricercatore prima del loro eventuale rilascio.

### **10.8.3 I file di microdati disponibili per il 15° Censimento della popolazione e delle abitazioni 2011**

Come tutti i microdati delle rilevazioni condotte dall'Istat, anche i microdati relativi al 15° Censimento della popolazione e delle abitazioni 2011 sono stati archiviati in Armida.

Attualmente i file di microdati, oggetto di diffusione, disponibili per rispondere alle richieste degli enti del Sistan e accessibili presso il Laboratorio Adele da parte di ricercatori e studiosi di università, istituti o enti di ricerca sono relativi a:

- ▶ *popolazione legale*: file di microdati sulle caratteristiche demografiche della popolazione residente (italiana e straniera) censita al 9 ottobre 2011 e pubblicata in Gazzetta Ufficiale n. 209 del 18 dicembre 2012;
- ▶ *individui residenti con variabili anagrafiche, istruzione, condizione professionale e pendolarismo*: file di microdati sulle caratteristiche demografiche, sull'istruzione, sulla condizione professionale o non professionale, caratteristiche dell'attività lavorativa e sul pendolarismo (spostamenti per motivi di studio o di lavoro in base al luogo di destinazione) della popolazione residente sul territorio nazionale;
- ▶ *alloggi*: file di microdati sulle caratteristiche delle abitazioni e degli altri tipi di alloggio occupati da persone residenti e non residenti;
- ▶ *edifici*: file di microdati sulle caratteristiche, sul tipo d'uso e di utilizzo degli edifici e dei complessi di edifici;
- ▶ *famiglie*: file di microdati sulla dimensione e sulla tipologia della famiglia, sulle caratteristiche dei nuclei familiari e delle coppie.

La fornitura dei file di microdati attraverso Armida può essere realizzata in diversi formati: formato fisso (.dat) o delimitato (.txt), Access, Excel 2003, Excel 2007, Xml.

La fornitura avviene attraverso un package (un archivio compresso) contenente oltre ai microdati nel formato richiesto, anche tutti i metadati utili ad una corretta interpretazione dei dati: tracciato record, classificazioni, altri allegati. Nel caso del Censimento della popolazione e delle abitazioni 2011 gli ulteriori allegati disponibili sono:

- ▶ questionario famiglia short form da 3 persone;
- ▶ questionario famiglia long form da 3 persone;
- ▶ questionario famiglia short form da 6 persone;
- ▶ questionario famiglia long form da 6 persone;
- ▶ questionario famiglia short form aggiuntivo;
- ▶ questionario famiglia long form aggiuntivo;
- ▶ questionario famiglia scorta;
- ▶ questionario convivenza;
- ▶ questionario convivenza aggiuntivo;
- ▶ questionario edifici;
- ▶ linee guida di utilizzo de microdati del censimento popolazione 2011.

## 10.9 Pubblicazione dei dati in formato Linked Open Data

### 10.9.1 Introduzione

La diffusione dei dati del 15° Censimento generale della popolazione e delle abitazioni in formati aperti è stata caratterizzata, oltre che dalla tradizionale produzione di dati in formati non proprietari e leggibili da programmi software (come ad esempio i formati txt e csv) anche, per la prima volta all'Istat, dal rilascio di output in formato Linked Open Data (Lod).

Linked Open Data è un paradigma di rappresentazione strutturata dei dati che consente di collegare tra loro dati provenienti da diverse fonti, basandosi su tecnologie e standard del web semantico. I dati Lod sono accessibili sia da utenti umani sia da macchine e possono essere direttamente interrogati da applicazioni esterne in modo indipendente dai linguaggi e dalle tecnologie adottate. Elemento centrale della pubblicazione dei dati in formato Lod è costituito dalle ontologie, cioè rappresentazioni formali, condivise ed esplicite di una concettualizzazione di un dominio di interesse.

Per rendere disponibili in tale formato i dati del censimento è stata realizzata una piattaforma sperimentale, disponibile all'indirizzo **[datiopen.istat.it](http://datiopen.istat.it)**, costituita da un ambiente di navigazione e interrogazione relativo sia alle basi territoriali sia alle variabili censuarie a livello di sezioni di censimento, località abitate, aree di censimento e aree sub-comunali.

La piattaforma consente sia interrogazioni libere attraverso uno Sparql endpoint, sia navigazioni guidate attraverso interfacce grafiche tramite le quali è possibile predisporre automaticamente interrogazioni sui dati ed esportare i risultati in vari formati.

La pubblicazione di dati in formato Lod, oltre che aderire alle direttive per la valorizzazione del patrimonio informativo pubblico, pubblicate dall'Agenzia per l'Italia Digitale <sup>(28)</sup>, risponde alle esigenze sempre più vaste espresse dalle comunità di utilizzatori di disporre di dati standardizzati e interoperabili su scala nazionale.

---

<sup>28</sup> (2014) "Le Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico" sono disponibili all'indirizzo: [http://www.agid.gov.it/sites/default/files/linee\\_guida/patrimoniopubblicolg2014\\_v0.7finale.pdf](http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7finale.pdf)

### 10.9.2 Ontologie di Dominio

Un'ontologia è una rappresentazione formale, condivisa ed esplicita di una concettualizzazione di un dominio di interesse. Per la pubblicazione dei dati del 15° Censimento della Popolazione e delle abitazioni a livello di sezione di censimento in formato Lod sono state definite due distinte ontologie: (i) l'ontologia delle basi territoriali per la descrizione e rappresentazione formale del territorio italiano, analizzandolo sia dal punto di vista amministrativo che statistico-geografico; (ii) l'ontologia dei dati censuari per la formalizzazione e descrizione dei metadati dell'indagine censuaria.

La qualità del dato è stata garantita utilizzando la meta-ontologia Prov-O che certifica la provenienza del dato. In maggior dettaglio, i dati pubblicati sono corredati dai seguenti metadati che ne descrivono e certificano la provenienza nonché la qualità:

- ▶ entità, attività e agenti partecipanti al processo di generazione del dato;
- ▶ responsabile del dato;
- ▶ possessore dei diritti sui dati;
- ▶ chi pubblica il dato;
- ▶ data di ultima modifica dei dati;
- ▶ titolo dei dati pubblicati;
- ▶ periodo di riferimento dei dati;
- ▶ licenza di pubblicazione;
- ▶ descrizione dei dati;
- ▶ riferimento spaziale dei dati.

#### 10.9.2.1 Ontologia delle Basi Territoriali

L'ontologia delle Basi Territoriali descrive le principali caratteristiche del territorio italiano, che è stato analizzato sia da un punto di vista amministrativo che statistico-geografico.

Dal punto di vista amministrativo il territorio è stato suddiviso in:

- ▶ *stato*: entità politica che governa ed esercita il potere sovrano su un determinato territorio e sui soggetti ad esso appartenenti;
- ▶ *ripartizione geografica*: Nord, Sud, Est, Ovest;
- ▶ *regione*: ente territoriale presente nell'ordinamento giuridico Italiano;
- ▶ *provincia*: ente locale avente una competenza su un gruppo di comuni, non necessariamente contigui;
- ▶ *comune*: ente amministrativo determinato da precisi limiti territoriali sui quali insiste una porzione di popolazione;
- ▶ *area sub-comunale*: municipi, quartieri, ecc. nei 34 comuni di maggiore dimensione demografica e con popolazione non inferiore a 100 mila abitanti.

Dal punto di vista statistico-geografico, il territorio è partizionato in:

- ▶ *sezioni di censimento*: porzione di territorio su cui sono effettuate le rilevazioni dell'Istat in occasione dei censimenti. La popolazione media è di circa 170 individui;
- ▶ *aree di censimento*: raggruppamenti di sezioni di censimento, tra loro contigue, intermedie tra sezioni e località di centro abitato, appartenenti ai centri principali;
- ▶ *località*: area più o meno vasta di territorio, conosciuta di norma con un nome proprio, sulla quale sono situate una o più case raggruppate o sparse.

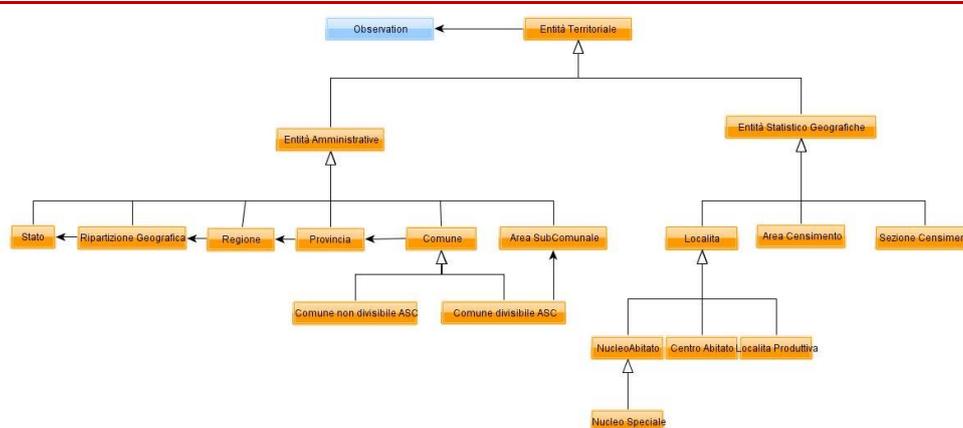
Nell'ontologia vengono inoltre caratterizzate le aree speciali, quali le entità geomorfologiche (stagni, valli da pesca, fiume, laguna,...), le isole amministrative e le zone in contestazione; i nuclei speciali quali ad esempio asl, abbazie, centri di accoglienza, per un totale di 45 nuclei speciali.

Tutte le entità così individuate sono collegate tra loro sia da relazioni di gerarchia che da altre relazioni quali ad esempio la relazione “eAssegnataA” che lega una zona in contestazione con il comune a cui tale zona è stata assegnata, mentre la relazione “eContesaDa” esplicita il comune che rivendica l'appartenenza della zona in contestazione.

Infine, sono stati esplicitati i collegamenti con i concetti equivalenti presenti nella ontologia internazionale Geonames, ad esempio per i fiumi descritti come area speciale nonché entità geomorfologica è stata esplicitata la relazione di equivalenza con la codifica H.Stm presente nell'ontologia Geonames.

La figura 10.35 riporta la rappresentazione grafica dell'ontologia delle basi territoriali.

Figura 10.35 - Visualizzazione dell'Ontologia delle Basi Territoriali (livello alto di raffinamento)



### 10.9.2.2 Ontologia delle Variabili Censuarie

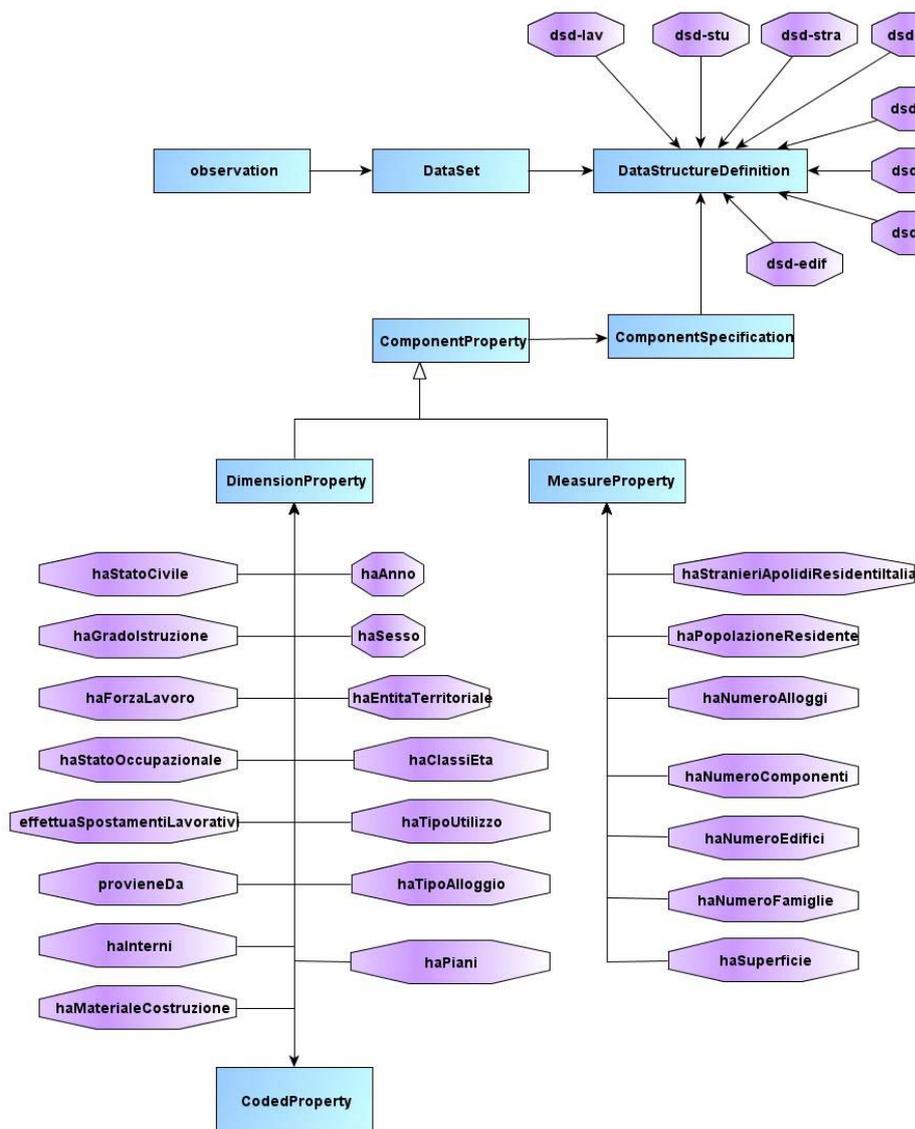
L'ontologia dei Dati Censuari formalizza e descrive i metadati delle variabili censuarie. Per la sua implementazione, trattandosi di dati statistici, è stata utilizzata la meta ontologia Data Cube Vocabulary, raccomandazione W3C del 16 Gennaio 2014, definita per la rappresentazione di dati multidimensionali.

Le variabili censuarie sono state suddivise in quattro gruppi: (i) popolazione/individui, (ii) famiglie, (iii) edifici, (iv) alloggi. Per ognuno di tali gruppi, di seguito vengono elencate le misure e le dimensioni formalizzate nell'ontologia:

- ▶ popolazione/individui;
- ▶ famiglie;
- ▶ edifici;
  - misure: popolazione residente, stranieri e apolidi residenti in italia;
  - dimensioni: sesso, classi di età, stato civile, grado di istruzione, forza lavoro, stato occupazionale, spostamenti lavorativi, paesi di provenienza;
  - misure: numero famiglie, numero componenti;
  - dimensioni: componenti della famiglia, condizione di godimento dell'alloggio;
  - misure: numero edifici;
  - dimensioni: numero interni dell'edificio, materiale di costruzione, periodo di costruzione, numero di piani, tipo di utilizzo dell'edificio.
- ▶ alloggi:
  - misure: numero alloggi, superficie;
  - dimensioni: stato di occupazione, tipo di alloggio.

La figura 10.36 riporta la rappresentazione grafica dell'ontologia delle variabili censuarie.

Figura 10.36 - Visualizzazione dell'Ontologia delle Variabili Censuarie (livello alto di raffinamento)

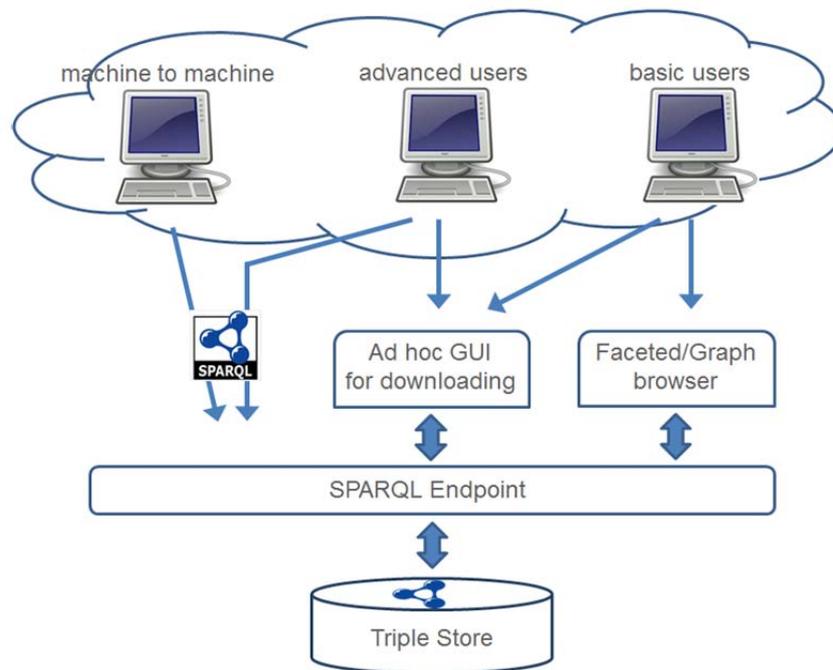


### 10.9.3 Schema Architeturale e Implementazione

I dati censuari a livello sub-comunale (sezione di censimento, aree sub-comunali, località e aree di censimento) in formato Lod sono stati resi disponibili sul sito datiopen.istat.it.

La figura 10.37 descrive lo schema architeturale del sito.

Figura 10.37 - Architettura del sito web



Nell'architettura del sito sono presenti tre componenti principali per accedere ai dati:

- (i) lo Sparql endpoint;
- (ii) un'interfaccia generalizzata per la navigazione dei dati (browsing del grafo RDF);
- (iii) un'interfaccia ad-hoc per il download di data set.

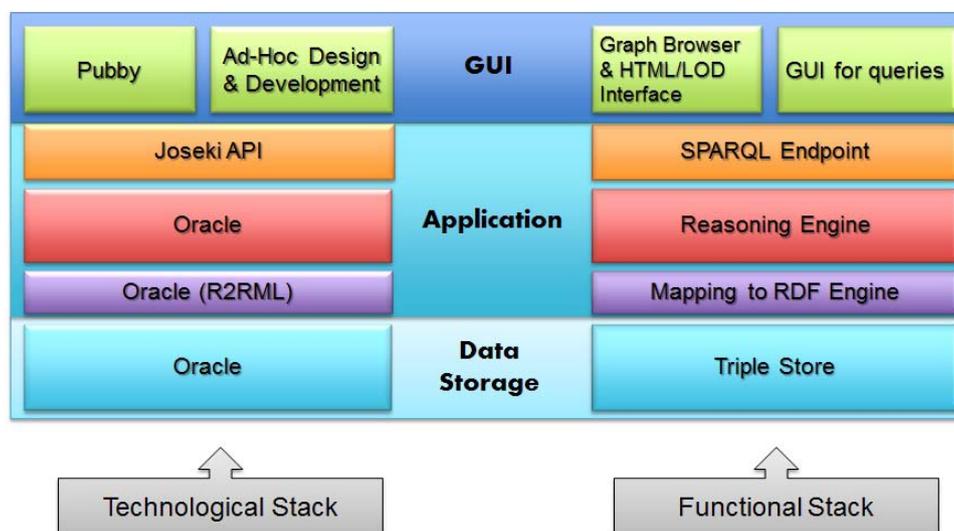
Gli utenti che hanno conoscenza del linguaggio Sparql possono direttamente accedere all'endpoint (entry point per il servizio di interrogazione libera dei dati), che può anche essere utilizzato per la comunicazione machine-to-machine. Gli utenti non tecnici possono sfruttare l'interfaccia generalizzata per navigare i dati oppure l'interfaccia ad-hoc che fornisce la possibilità di costruire interrogazioni guidate in modo semplice.

I dati sono memorizzati in formato RDF in una base dati di triple (Triple Store). La soluzione tecnologia da noi utilizzata è la componente "Spatial & Graph" di Oracle che fornisce sia il Triple Store che il motore Sparql. Tale scelta è motivata dal livello di scalabilità molto alto, fino a miliardi di triple consentito dal triple store.

Tutte le scelte tecnologiche relative alla realizzazione del canale di diffusione Lod sono riassunte in figura 10.38.

Il popolamento del Triple Store avviene partendo dai dati memorizzati in una base dati relazionale e traducendo il formato dei dati secondo quanto specificato dall'ontologia del dominio attraverso una procedura ad hoc di Oracle che mappa i dati in base a regole espresse con il linguaggio standard R2Rml. In generale, l'implementazione del mapping può essere *on-demand* o fatta attraverso la *data materialization*, nel primo caso le triple sono virtuali e create come risposta ad una query, nel secondo caso sono fisicamente memorizzate nell'archivio. Nel nostro Triple Store le triple sono materializzate (*data materialization*) e corredate con nuove triple risultato della fase di *reasoning* in cui vengono dedotte dall'ontologia le triple non esplicitate.

Figura 10.38 - Stack tecnologico dell'applicazione Lod



Lo Sparql Endpoint é Joseki, un framework open di Apache che usa la libreria java Jena integrata con l'Oracle Sparql Engine. Joseki permette l'esposizione delle triple come un endpoint accessibile via http.

L'interfaccia generalizzata per la navigazione dei dati è Pubby, un altro framework open (Java), che permette sia la de-referenziazione delle Uri, che vengono automaticamente convertite in una interrogazione Sparql, che la navigazione vera e propria attraverso una visualizzazione standard personalizzabile.

Contenitore e punto di accesso ai vari strumenti è il sito datiopen.istat.it realizzato col framework responsive "*Bootstrap*". *Bootstrap* oltre a rispettare gli standard dell'Istituto assicura la fruibilità delle pagine sui diversi dispositivi (pc, tablet, smartphone, ecc).

Dalla pagina di accoglienza del sito (figura 10.39) è possibile accedere alla documentazione sul progetto e alle varie componenti.

Figura 10.39 - Home page del sito web

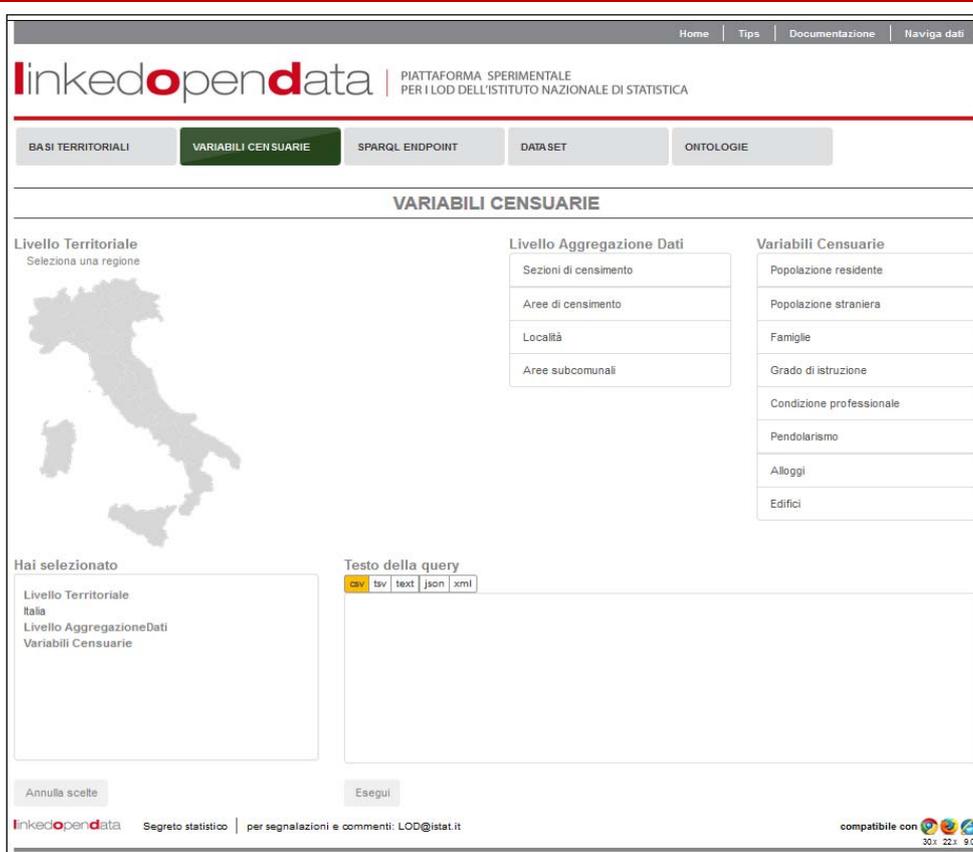
La home page del sito **linkedopendata** (PIATTAFORMA SPERIMENTALE PER I LOD DELL'ISTITUTO NAZIONALE DI STATISTICA) presenta un'interfaccia con un mappe di sfondo e quattro sezioni principali:

- BASI TERRITORIALI**: Interrogazioni guidate: Dati territoriali dell'anno 2011
- VARIABILI CENSUARIE**: Interrogazioni guidate: Indicatori del 15° censimento generale della popolazione e delle abitazioni 2011 aggregati per partizioni del territorio
- SPARQL ENDPOINT**: Interrogazioni libere: accesso diretto all'Endpoint SPARQL
- DATASET**: Download: Dataset censuari
- ONTOLOGIE**: Download: Ontologie del territorio e del censimento

Il footer include il logo **linkedopendata**, il testo "Segreto statistico" e "per segnalazioni e commenti: LOD@istat.it", oltre a icone di compatibilità con i browser Internet Explorer 9.0, Firefox 22.0 e Google Chrome 30.0.

In particolare, facendo riferimento all'architettura di progetto rappresentata in figura 10.37, troviamo lo Sparql endpoint raggiungibile dal terzo dei quattro box in evidenza (di colore marrone), l'interfaccia per la navigazione (realizzata con Pubby) raggiungibile da un link nella parte superiore a destra della pagina, e più accessi per l'interfaccia ad hoc per il download secondo che si scelga l'interrogazione dei data set delle sole basi territoriali o interrogazione dei data set relativi agli indicatori censuari. L'interfaccia per interrogare i data set delle basi territoriali è raggiungibile dal primo box blu, quella per gli indicatori censuari dal secondo box verde (figura 10.40), questi ultimi sono scaricabili in formato compatto a livello regionale (da file di testo pre-elaborati) utilizzando il quarto box rosso.

Figura 10.40 - Interfaccia per l'interrogazione guidata del triple store



Tecnicamente, per il disegno delle pagine del sito è stata adottata una struttura (*impalcatura/scaffolding*) costituita da una griglia (*grid system*). In questa sono state definite righe e colonne, in cui poi sono stati inseriti i contenuti propri del sito. Sono stati utilizzati componenti grafici (*pulsanti, gruppi di pulsanti, barre di navigazione, menu a discesa, liste...*), stili predefiniti per realizzare effetti grafici e visivi e Javascript per il controllo dei componenti stessi. Inoltre è stato utilizzato il framework "jQuery" per gestire in modo completo la manipolazione del DOM e dei processi lato client. Lato server sono state implementate funzioni in Php per l'esecuzione delle richieste lato client al server.

Anche la pagina dello Sparql endpoint è stata disegnata con tali modalità, sostituendo la pagina di default disponibile con Joseki.

Il risultato delle richieste inoltrate sia dallo Sparql endpoint sia dalle interfacce guidate viene salvato su un file il cui formato è variabile e selezionabile dall'utente. Solo nel caso in cui si scelga come output l'Html la visualizzazione dei risultati è direttamente sul browser (figura 10.41).

Figura 10.41 - Risultato Sparql Endpoint

Endpoint SPARQL		linkedopendata				
Comune	Codice_Istat_Comune	Area_Subcomunale	Osservazione	Numero_Edifici	Utilizzo	Periodo_Costruzione
"Firenze" @it	"017"	"Gavinana-Galluzzo" @it	"Edifici ad uso residenziale costruiti prima del 1919 per Area SubComunale Gavinana-Galluzzo del Comune di Firenze"	"1267" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Residenziale" @it	"Anni prima del 1919" @it
"Firenze" @it	"017"	"Campo di Marte" @it	"Edifici ad uso residenziale costruiti prima del 1919 per Area SubComunale Campo di Marte del Comune di Firenze"	"2141" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Residenziale" @it	"Anni prima del 1919" @it
"Firenze" @it	"017"	"Rifredi" @it	"Edifici ad uso residenziale costruiti prima del 1919 per Area SubComunale Rifredi del Comune di Firenze"	"1696" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Residenziale" @it	"Anni prima del 1919" @it
"Firenze" @it	"017"	"Centro Storico" @it	"Edifici ad uso residenziale costruiti prima del 1919 per Area SubComunale Centro Storico del Comune di Firenze"	"4856" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Residenziale" @it	"Anni prima del 1919" @it
"Firenze" @it	"017"	"Isolotto- Legnaia" @it	"Edifici ad uso residenziale costruiti prima del 1919 per Area SubComunale Isolotto- Legnaia del Comune di Firenze"	"1072" ^^<http://www.w3.org/2001/XMLSchema#decimal>	"Residenziale" @it	"Anni prima del 1919" @it

Questa pagina mostra informazioni ottenute via SPARQL Endpoint <http://datiopen.istat.it/sparql>

## Conclusioni

Sul 15° Censimento della popolazione e delle abitazioni, che si è rivolto a più di 25 milioni di famiglie e ha coinvolto circa 90 mila operatori della rete di rilevazione, composta da uffici di censimento a livello comunale, provinciale e nazionale, si è concentrato il maggior numero di nuovi strumenti tecnologici. Le numerose applicazioni descritte nei capitoli precedenti sono state tutte sviluppate in ambiente web in modo da essere totalmente indipendenti dai dispositivi hardware, dai sistemi operativi e dalle configurazioni di base nel rispetto dei requisiti sulla sicurezza e l'integrità dei dati richiesti dalla normativa in materia di protezione dei dati personali.

Il punto di partenza è stata l'acquisizione delle liste anagrafiche comunali (Lac) e la costruzione di una unica lista di base ottenuta integrando le 8092 Lac con altri archivi amministrativi, come ad esempio i dati sulle abitazioni non occupate o sui permessi di soggiorno degli stranieri, in modo da supportare gli operatori comunali e facilitare le operazioni di raccolta dati.

Il questionario elettronico, ad eccezione del primo giorno in cui si è verificato un sovraccarico del sistema che ha di fatto impedito alle famiglie di compilare il questionario, problematica risolta attraverso il potenziamento dell'infrastruttura hardware nel giro di 24 ore, è stato utilizzato da più di 8 milioni di famiglie che hanno scelto la compilazione on line e da 9 grandi comuni che hanno optato per la registrazione autonoma dei dati. Per accedere al questionario on line era sufficiente una connessione internet e un browser. L'interfaccia grafica del sistema guidava le famiglie nella compilazione abilitando e disabilitando automaticamente i percorsi alternativi ed evidenziando lo stato di compilazione di ciascuna sezione del questionario. È stato inoltre realizzato un sistema di controlli per prevenire l'inserimento di dati incoerenti, senza per questo appesantire eccessivamente l'interazione tra l'utente e il sistema.

Il sistema Sgr realizzato per il censimento della popolazione è stato messo a disposizione degli operatori dislocati sull'intero territorio nazionale ed ha permesso di gestire tutte le fasi della rilevazione, dal monitoraggio della consegna postale dei questionari alle famiglie, fino alla riconsegna all'Istat da parte dei comuni dei questionari compilati. Attraverso Sgr gli uffici comunali di censimento hanno potuto gestire autonomamente le utenze dei propri operatori, monitorare da vari punti di vista il progresso delle operazioni, inserire i dati riassuntivi dei questionari di famiglia, convivenza e degli edifici, gestire la fase di invio all'Istat dei questionari compilati. Particolare importanza ha avuto la funzione di confronto tra censimento e anagrafe, con la quale gli uffici comunali di censimento hanno effettuato il confronto tra gli individui censiti e quelli presenti in anagrafe, producendo i bilanci da utilizzare nella successiva fase di revisione anagrafica. I sistemi censuari sono stati sviluppati utilizzando prevalentemente tecnologie software open source e librerie di sviluppo e schemi di progetto standard. L'esperienza iniziale del 6° Censimento dell'agricoltura ha infatti permesso di mettere a punto alcuni standard progettuali specifici che sono stati mantenuti negli altri censimenti.

Queste scelte hanno consentito un elevato grado di riuso dei metodi e del software, specialmente per quanto riguarda i sistemi di gestione della rilevazione e i sistemi di generazione e presentazione dei questionari elettronici.

Il controllo della raccolta dei dati censuari si è potuto avvalere di strumenti innovativi di integrazione di fonti cartografiche di primario interesse quali le foto aeree aggiornate del territorio, i limiti delle sezioni di censimento 2011, i grafi stradali di fonte commerciale, la cartografia catastale dei fabbricati. Più in generale sono stati sviluppati strumenti tecnologici innovativi per migliorare il geo-riferimento dei dati. Attraverso la rilevazione dei numeri civici sono state poste le basi per la realizzazione, in collaborazione con l'Agenzia del territorio, dell'Archivio nazionale dei numeri civici

delle strade urbane (Ancsu), che diventerà l'archivio unico di riferimento per le Pubbliche Amministrazioni i cittadini e le imprese.

La registrazione dei questionari cartacei, assegnata in outsourcing, è avvenuta tramite lettura ottica e registrazione tradizionale. I blocchi di fornitura, di dati e immagini, sono stati trasferiti dalla ditta di registrazione all'Istat tramite Vpn su server presso il Ced Istat.

Per l'acquisizione dei dati e il monitoraggio del processo di registrazione è stata sviluppata un'applicazione software generalizzata (Samor-Pop) che, oltre a verificare il corretto trasferimento dei file, ha permesso di controllare la conformità ai tracciati record descritti nel manuale di registrazione fornendo i risultati dei controlli formali su appositi file di *log*. Le immagini via via acquisite sono state catalogate nel database del Catalogo delle Immagini, che ha fornito informazioni e servizi indispensabili per il processo di controllo di qualità.

Per i controlli di qualità della fornitura è stata sviluppata la procedura software All-In-One che ha permesso di verificare, su campioni rappresentativi di modelli, la leggibilità delle immagini e l'accuratezza dei dati. Il software permetteva la selezione del campione di controllo, la digitazione dei dati dei questionari campione, il controllo delle discordanze tra i dati della fornitura e i dati digitati per il controllo, il conteggio degli errori e la verifica di accettazione rispetto alla soglia massima di errore.

Le applicazioni sviluppate per l'acquisizione, il monitoraggio e la verifica della qualità sono parametrizzabili per differenti tracciati record e tipologie di controlli, permettono un monitoraggio immediato e una tempestiva analisi dei dati.

Il sistema per la diffusione dei dati ha previsto un'architettura di frontiera con l'utilizzo di tecnologie avanzate di integrazione dati, accesso a dati multidimensionali, servizi di analisi dei dati e di accesso remoto. Tale sistema ha pienamente supportato le fasi di analisi e produzione dei dati da diffondere, nonché la diffusione vera e propria su molteplici canali di accesso. In particolare, gli ipercubi di diffusione per Eurostat sono 60, accompagnati da altri 20 aggiuntivi relativi ai quality report. Gli ipercubi per la diffusione italiana sono più di 300. I file di microdati sono suddivisi per provincia ed hanno tracciati differenti per ciascuna tipologia di unità rilevata: individui, famiglie, alloggi, edifici. I dati aggregati per sezione di censimento ed altre aree sub-comunali comprendono anche i dati aggregati per area di censimento e per località abitata.

## RIFERIMENTI BIBLIOGRAFICI

- Abbatini D., Di Pede F. 2012. L'archivio dei numeri civici: strumento di supporto alla rilevazione e nuova infrastruttura informativa per il Paese. In Atti del convegno Ais-Met.
- Abbatini D., Di Pede F. e Patruno E. 2009. Rilevazione Numeri Civici. Manuale per le rilevazioni sul campo. Retrieved from [https://mc.istat.it/DOCUMENTI/Manuale\\_per\\_le\\_operazioni\\_sul\\_campo.pdf](https://mc.istat.it/DOCUMENTI/Manuale_per_le_operazioni_sul_campo.pdf)
- Aracri R., De Francisci S., Pagano A., Scannapieco M., Tosco L., Valentino L. 2014. "Publishing the 15th Italian Population and Housing Census in Linked Open Data". In Second International Workshop on Semantic Statistics.
- Benassi F., Bruno M., Giacummo M., Silipo M., Vaste G., Zindato D. 2014. "Managing census complexity through highly integrated web systems". In Rivista di Statistica Ufficiale, n. 3.
- Bianchi G., Borrelli F., Filippini R., Francescangeli P., Lipsi R.M., Maione C., Polizzi M., Tomeo V., Ricci G.M.A. 2014. "Il processo di acquisizione dei dati del censimento generale della popolazione e delle abitazioni del 2011: guida alla registrazione dei modelli di rilevazione". In Istat working papers, n.2.
- Bruno M., Giacummo M., Silipo M., Vaste G. 2013. "A web based management system for addressing census complexity: the Italian experience". New Techniques and Technologies for Statistics, Bruxelles 5-7 marzo.
- Bruno M., Giacummo M., Silipo M., Vaste G. 2013. "Le tecnologie web come strumento fondamentale per un censimento di qualità: l'esperienza italiana". Undicesima conferenza nazionale di statistica, Roma 20-21 febbraio.
- Cavaness C. 2004. Programming Jakarta Struts. O'Reilly Media
- Di Pede F. 2012. "L'archivio dei numeri civici, un'infrastruttura informativa per il Paese". Forum PA: L'eredità del censimento della popolazione per le Pubbliche Amministrazioni.
- Golfarelli M., Rizzi S. 2005. Data Warehouse – Teoria e pratica della progettazione. McGraw Hill.
- Kurniawan B. 2008. Struts 2 Design and Programming. BrainySoftware.
- Rezzani A. 2012. Business Intelligence – Processi, metodi, utilizzo in azienda, Apogeo.
- Venturi M. (a cura di). 2013. 6° Censimento dell'Agricoltura: Il sistema informatico integrato, fascicolo 4. In Atti del 6° Censimento dell'Agricoltura. Istat.
- Venturi M., Crescenzi F., Sindoni G. 2012, "Strumenti innovativi e ruolo delle tecnologie nei Censimenti 2010-2012". In NewStat n.6.
- Venturi M., Crescenzi F., Sindoni G. 2013. "The role of technology in the 2010-2012 Italian censuses". NTTS (New Techniques and Technologies for Statistics) 2013 Conference. Bruxelles 5-7 March.
- Venturi M., Bellini G., Conti C., Giacummo M., Mottura G. 2012. "Web use for the 6th General Agriculture Census in Italy". Journées de Méthodologie Statistique 2012 - Insee – Paris, Janvier.
- Venturi M., Giacummo M., Tininini L., Virgillito A. 2012. "The web-based information system of Italian Population Census". Joint UNECE-Eurostat meetings on population and housing censuses. General Meeting on Censuses 2.b. Internet data collection, Ginevra, 24-25 Maggio.

## Per saperne di più

“Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e di ricerca scientifica effettuati nell’ambito del Sistema statistico nazionale, Allegato A.3 al decreto “Decreto legislativo 30 giugno 2003, n. 196, recante Codice in materia di protezione dei dati personali”

“Decreto legislativo 6 settembre 1989, n. 322, e successive modifiche e integrazioni, recante Norme sul Sistema statistico nazionale e sulla riorganizzazione dell’Istituto nazionale di statistica, ai sensi dell’articolo 24 della legge 23 agosto 1988, n. 400”

“Direttiva del Comstat n. 9 del 20 aprile 2004 che definisce i Criteri e modalità per la comunicazione dati personali in ambito SISTAN”, Gazzetta ufficiale della Repubblica Italiana del 23 dicembre 2004, n. 300

“Mapping Assistant user manual v.1.4”, Eurostat 2013.

“Measuring Population and Housing - Practices of UNECE countries in the 2010 round of censuses (2013)” - <http://www.unece.org/stats/census.html>

Organisation for Economic Co-operation and Development, “Oecd.Stat - Web Browser User Guide”, e successive integrazioni, Oecd 2013

“Sdmx 2.0 standard documentation”, [www.sdmx.org](http://www.sdmx.org)

“Sdmx 2.0 Xml Schemas (Section 3B – Sdmx-ML) 2010-2012”, [www.sdmx.org](http://www.sdmx.org)