

Atti del 9° Censimento generale  
dell'industria e dei servizi e  
Censimento delle istituzioni non profit

# Il sistema informatico integrato

6



CENSIMENTO  
DELL'INDUSTRIA  
E DEI SERVIZI 2011





Atti del 9° Censimento generale  
dell'industria e dei servizi e  
Censimento delle istituzioni non profit

# Il sistema informatico integrato

6



CENSIMENTO  
DELL'INDUSTRIA  
E DEI SERVIZI 2011



*Atti a cura della:*

Direzione Centrale delle Rilevazioni Censuarie  
e Registri Statistici (DCCR)

Fascicolo a cura di: Marina Venturi

Atti del 9° Censimento generale dell'industria e dei servizi  
e Censimento delle istituzioni non profit

6 - IL SISTEMA INFORMATICO INTEGRATO

© 2015

Istituto nazionale di statistica  
Via Cesare Balbo, 16 - Roma

ISBN 978-88-458-1858-5 (elettronico)

ISBN 978-88-458-1864-6 (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 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit</b>	9
<b>Introduzione</b>	19
<b>PARTE PRIMA</b>	
<b>I SISTEMI INFORMATICI DI ACQUISIZIONE E GESTIONE DELLA RILEVAZIONE SULLE IMPRESE E DEL CENSIMENTO DELLE ISTITUZIONI NON PROFIT</b>	
<b>Capitolo 1 - L'Architettura generale dei sistemi informatici</b>	23
1.1 L'architettura funzionale	23
1.2 L'architettura software	24
1.2.1 <i>L'architettura del sistema di gestione e del questionario</i>	25
1.2.2 <i>Il sistema di versioning</i>	27
1.2.3 <i>L'architettura del Portale della rete</i>	29
1.3 Il modello dei dati	30
1.3.1 <i>La base dati del sistema di gestione</i>	30
1.3.2 <i>La base dati del questionario</i>	31
1.4 L'architettura logico-fisica dei sistemi	35
1.5 I test delle applicazioni e dei sistemi	37
1.5.1 <i>I test funzionali</i>	37
1.5.2 <i>I test automatici</i>	39
1.5.3 <i>I test di sicurezza</i>	41
1.5.4 <i>I test di carico</i>	41
1.6 I sistemi di autenticazione e autorizzazione utenti degli uffici di censimento	45
1.6.1 <i>L'autenticazione</i>	45
1.6.2 <i>L'autorizzazione</i>	47
1.7 La formazione on line (e-learning)	48
<b>Capitolo 2 – Il sistema di acquisizione</b>	53
2.1 La modellazione del questionario	54
2.1.1 <i>La modellazione della struttura del questionario</i>	55
2.1.2 <i>La modellazione dei percorsi di compilazione e il grafo del questionario</i>	57
2.1.3 <i>Come usare il grafo del questionario</i>	62
2.2 Il progetto del sistema di acquisizione	65
2.2.1 <i>L'architettura interna</i>	65
2.2.2 <i>Tipologie di domande</i>	67
2.2.3 <i>L'implementazione del grafo</i>	70
2.2.4 <i>I controlli (client e server)</i>	72
2.2.5 <i>L'accessibilità</i>	74

	Pagina
<b>Capitolo 3 – Il sistema di gestione</b>	77
3.1 La comunicazione con la rete	77
3.2 Ruoli e funzioni della rete di rilevazione nel sistema di gestione	79
3.3 Il workflow di lavorazione e la gestione dei ricicli	82
3.3.1 <i>Il flusso del questionario</i>	82
3.3.2 <i>L'esito come elemento cardine della rilevazione</i>	84
3.4 Il controllo dei dati	85
3.4.1 <i>Il check dei microdati</i>	85
3.4.2 <i>La validazione del questionario</i>	88
3.5 Il monitoraggio aggregato e di dettaglio	90
3.5.1 <i>L'architettura DB</i>	90
3.5.2 <i>L'architettura software</i>	91
3.5.3 <i>Le liste di dettaglio</i>	94
3.5.4 <i>Dalle liste precensuarie alle liste dei pesi</i>	95
3.5.5 <i>Lo scarico dei microdati</i>	96
3.6 La chiusura dell'indagine e la gestione delle eccezioni	97
3.6.1 <i>La gestione delle procedure di accertamento dell'obbligo di risposta da parte della rete</i>	97
3.6.2 <i>Il sistema di scarico delle sanzioni</i>	98
3.6.3 <i>Le operazioni di chiusura</i>	99
3.6.4 <i>L'interazione con il vettore postale per il ritiro della documentazione censuaria</i>	100
3.7 La rendicontazione delle spese censuarie	101
3.8 L'assistenza in fase di esercizio e la documentazione di supporto	102
3.9 Il Sistema di gestione degli accertamenti della violazione dell'obbligo di risposta - SGS	103

## PARTE SECONDA

### IL SISTEMA INFORMATICO DI ACQUISIZIONE E GESTIONE DEL CENSIMENTO DELLE ISTITUZIONI PUBBLICHE

<b>Capitolo 1 – Il sistema di acquisizione e di gestione</b>	109
1.1 Le due fasi del censimento	111
1.2 Il sito web	112
1.3 In-formazione on-line (la piattaforma di e-learning)	114
1.4 Il sistema di gestione	118
1.4.1 <i>Monitoraggio Istat, scuole ed uffici regionali</i>	118
1.4.2 <i>Il monitoraggio delle unità istituzionali</i>	119
1.4.3 <i>La gestione dinamica delle utenze e/o dei questionari</i>	120
1.5 Il questionario elettronico	123
1.5.1 <i>La realizzazione della struttura di data capturing</i>	124
1.5.2 <i>La progettazione dei questionari e la tecnologia utilizzata</i>	124
1.5.3 <i>Il trasferimento dei dati dalla Prima fase del censimento, modello MT1 e popolamento del database</i>	126
1.5.4 <i>La definizione delle procedure di acquisizione dati per le unità istituzionali "complesse" e caricamento dei tracciati offline</i>	127
1.6 L'integrazione tra i sistemi	128
1.6.1 <i>Il modello dei dati</i>	128
1.6.2 <i>Il sistema di accesso</i>	129

	Pagina
1.7 Lo scarico dei dati	130
1.7.1 <i>La modalità online</i>	130
1.7.2 <i>La modalità offline</i>	132
1.8 La documentazione online	133
1.9 I test e prestazioni del sistema	134

## PARTE TERZA

### IL SISTEMA INFORMATICO DI DIFFUSIONE

<b>Capitolo 1 – Il sistema di diffusione</b>	<b>139</b>
1.1 L'architettura complessiva	139
1.2 La progettazione dei cubi dati	140
1.3 Il sistema di aggregazione per la gestione dei casi di multiresponse e le dipendenze funzionali tra le dimensioni di analisi: le istituzioni non profit, le istituzioni pubbliche e le imprese dell'indagine multiscopo	142
1.3.1 <i>Il flusso dei dati</i>	142
1.3.2 <i>Organizzazione della base dati</i>	143
1.3.3 <i>La fase di ETL1: la gestione dei metadati per la generazione automatica delle fact table</i>	145
1.3.4 <i>La fase di ETL2: il software, il passaggio dalla modellazione dei cubi alla generazione dei file dati, le regole di controllo</i>	148
1.4 Il sistema di aggregazione di grandi quantità di dati: il registro delle imprese e dell'occupazione nelle imprese	149
1.4.1 <i>Progettazione del data warehouse primario (DW1)</i>	149
1.4.2 <i>Progettazione del data warehouse secondario (DW2)</i>	151
1.5 Il data warehouse di diffusione	153
1.5.1 <i>L'architettura tecnologica del data warehouse di diffusione</i>	154
1.5.2 <i>L'ambiente di navigazione</i>	156
1.5.3 <i>I metadati</i>	158
1.5.4 <i>L'ambiente di amministrazione ed il caricamento dei dati</i>	160
1.6 Il Single Exit Point per la diffusione dati machine-to-machine	162
1.6.1 <i>L'architettura tecnologica</i>	162
1.6.2 <i>La modellazione dei cubi di dati in formato DSD e la mappatura SDMX dei cubi di dati</i>	164
1.7 La disponibilità di microdati per gli utenti esterni	166
1.7.1 <i>Il sistema di archiviazione ARMIDA</i>	166
1.7.2 <i>La comunicazione dei microdati agli enti Sistan</i>	167
1.7.3 <i>L'accesso ai microdati attraverso il Laboratorio ADELE</i>	168
1.7.4 <i>I file di microdati disponibili per il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit</i>	169
<b>Conclusioni</b>	<b>171</b>
<b>Riferimenti Bibliografici</b>	<b>173</b>
<b>Elenco allegati disponibili in formato digitale</b>	<b>175</b>



## Presentazione dell'opera

Come avvenuto in occasione di tutti i precedenti censimenti, l'Istituto nazionale di statistica ha redatto gli "Atti" del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit. La pubblicazione descrive le caratteristiche tecniche, organizzative e metodologiche delle attività svolte.

Gli Atti sono editi esclusivamente in forma telematica e si articolano in sei fascicoli corredati dalla documentazione tecnica e dai documenti ufficiali del censimento. Considerata la rilevanza e la complessità degli argomenti trattati, l'opera è stata realizzata in maniera unitaria, essendo evidenti le interconnessioni fra le diverse fasi censuarie; tuttavia ciascun fascicolo ha propria autonomia, poiché tratta in maniera esaustiva un tema specifico.

I sei fascicoli che costituiscono gli Atti del 9° Censimento generale dell'industria e dei servizi sono:

1. *Norme e organizzazione;*
2. *Il censimento delle imprese;*
3. *Il censimento delle istituzioni non profit;*
4. *Il censimento delle istituzioni pubbliche;*
5. *Le sezioni di censimento;*
6. *Il sistema informatico integrato.*

Il primo fascicolo *Norme e organizzazione* è strutturato in quattro parti. La prima parte, *Il quadro normativo*, illustra le norme in base alle quali è stato progettato e realizzato il censimento. Sono presentati e descritti nei loro aspetti principali la legge di indizione e finanziamento, il Piano Generale di Censimento, nonché le circolari operative emanate dall'Istat per definire e dare attuazione a tutte le fasi censuarie. La seconda parte riguarda l'impianto organizzativo della *Rilevazione sulle imprese* e della *Rilevazione sulle istituzioni non profit*, mentre la terza è relativa all'impianto organizzativo della *Rilevazione sulle istituzioni pubbliche*. La distinzione è motivata dalle profonde differenze metodologiche, organizzative e tecnologiche che hanno contraddistinto il processo di rilevazione delle tre tipologie di unità economiche. Sia nella seconda che nella terza parte si descrivono le linee guida adottate nell'organizzazione delle rilevazioni, la composizione e le caratteristiche della rete di rilevazione, le modalità di formazione degli operatori censuari, gli strumenti di supporto alle rilevazioni e di gestione delle criticità nel corso del processo produttivo. La quarta parte, *Comunicazione e diffusione*, presenta la campagna di comunicazione integrata realizzata dall'Istat per promuovere la condivisione degli obiettivi da parte dei rispondenti e dei potenziali utilizzatori dei dati. In particolare illustra le azioni e gli strumenti specifici ideati e realizzati per ciascuna rilevazione censuaria, pur nell'unitarietà della strategia comunicativa. In essa è anche descritto il piano di diffusione, incentrato sul sistema di *data warehousing* dell'Istat I.stat e articolato in diversi prodotti.

Il secondo, terzo e quarto fascicolo trattano in dettaglio gli aspetti peculiari di ciascuna delle tre rilevazioni: sulle imprese, sulle istituzioni non profit e sulle istituzioni pubbliche. In ciascun fascicolo sono illustrate le caratteristiche generali e tecniche della rilevazione, percorrendo le principali tappe che hanno portato alla definizione delle liste precensuarie delle unità di rilevazione, dei contenuti informativi dei questionari, del piano di controllo e correzione dei dati, della diffusione dei risultati. Nel secondo fascicolo è inoltre descritto il processo metodologico e produttivo seguito per la realizzazione del sistema di registri sulle unità economiche Asia.

Il quinto fascicolo, *Le sezioni di censimento*, descrive la procedura di attribuzione delle sezioni di censimento alle unità locali, evidenziandone caratteristiche e peculiarità e presentando i risultati finali.

Il sesto fascicolo, *Il sistema informatico integrato*, descrive l'architettura e le soluzioni informatiche adottate per un censimento articolato e complesso che ha utilizzato differenti modalità di compilazione dei questionari e anche differenti modalità di restituzione dei questionari compilati, che ha previsto la registrazione dei questionari acquisiti in formato cartaceo presso gli uffici di rilevazione nonché un *paperless census* per la rilevazione sulle istituzioni pubbliche. Tutti i processi adottati, connessi e dialoganti tra loro, hanno garantito coerenza logica, velocità e trasparenza alle operazioni realizzate. Nel fascicolo sono descritti il Sistema di Gestione della Rilevazione (SGR), l'impianto informatico di acquisizione dati e gli strumenti informatici di diffusione dei dati.

## Il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit

A cura di Andrea Mancini

Il tradizionale campo di osservazione del censimento economico italiano si presenta complesso e molto articolato. In esso sono comprese imprese che agiscono sul mercato, amministrazioni pubbliche e organizzazioni private non profit. Si tratta in ogni caso di unità economiche produttrici di beni e servizi ma operanti con forme giuridiche molto diversificate, scopi plurimi, strutture produttive assai differenti, dimensioni molto variabili. L'informazione statistica relativa a questo universo di riferimento è altrettanto articolata, ma in occasione delle decennali tornate censuarie il quadro strutturale delle unità economiche residenti nel nostro Paese è sempre divenuto oggetto di osservazione statistica generalizzata, simultanea e a grande dettaglio di livello territoriale. Tuttavia la progressiva realizzazione del sistema di registri statistici (Asia-imprese, Asia-gruppi, Asia-unità locali), tutti basati su informazioni desunte da una pluralità di archivi amministrativi, ha consentito di produrre annualmente parti sostanziali del quadro statistico strutturale prima demandato al censimento decennale. Di fatto il sistema dei registri Asia permette di avere annualmente il conteggio delle unità economiche attive e delle loro unità locali, insieme con alcune variabili fondamentali, quali l'attività economica prevalente, il numero di addetti, il fatturato. Ciò almeno per le imprese operanti sul mercato.

In questo contesto l'obiettivo più rilevante e peculiare del 9° Censimento generale dell'industria e dei servizi è stato di completare, con riferimento al 31 dicembre del 2011, il sistema di registri statistici basato su dati di fonte amministrativa e di fornire informazioni strutturali di approfondimento mirato sui singoli e differenti settori istituzionali compresi nell'universo delle unità economiche. A quest'ultimo fine il 9° Censimento ha fatto uso di tre liste precensuarie delle unità economiche (imprese, istituzioni non profit, amministrazioni pubbliche) per raccogliere informazioni mediante varie rilevazioni, sia campionarie che totali, ciascuna specializzata nella raccolta di informazioni e dati idonei a completare il quadro statistico di ciascun settore, tenendo conto delle loro peculiarità giuridico-istituzionali e operative.

Dunque il 9° Censimento ha presentato elementi di omogeneità e sinergia, ma anche differenziazioni tematiche e metodologiche rilevanti, che hanno reso la sua realizzazione particolarmente complessa. Di fatto si è trattato di un censimento suddiviso in differenti rilevazioni tematiche, rese omogenee da una unica *governance* che ha permesso la gestione unitaria sul piano organizzativo, metodologico, informatico e tecnico. Esso è stato caratterizzato da innovazioni, tra le quali rilevanza generale hanno assunto:

- ▶ l'ampio uso di dati amministrativi per la predisposizione delle liste precensuarie, per il controllo e la correzione delle informazioni rilevate (anche contestuali alla raccolta dei dati), per la realizzazione di un censimento virtuale sui dati strutturali delle imprese grazie alle informazioni residenti nel sistema di registri statistici Asia. Grazie all'utilizzo di dati amministrativi, gli obiettivi raggiunti sono stati la maggiore tempestività nel rilascio dei risultati, il contenimento del carico statistico sui rispondenti, il miglioramento della qualità dei dati prodotti, la riduzione dei costi. Gli esiti censuari hanno consentito anche di migliorare i processi metodologici e statistici che portano alla realizzazione dei registri

- statistici di unità economiche. Questi costituiranno il fulcro sul quale ruoterà l'impianto organizzativo e metodologico delle rilevazioni post censuarie sulle unità economiche e dei censimenti economici continui;
- ▶ le novità sul piano dei contenuti informativi: tematiche inedite sono state approfondite, grazie alla realizzazione di specifiche rilevazioni con differenti questionari di settore, di tipo *long form*, che hanno permesso di trattare tematiche di interesse di ciascuno dei tre comparti;
  - ▶ la realizzazione di una complessa e omogenea architettura informatica in grado di gestire sia il sistema di monitoraggio di tutte le operazioni censuarie, sia il sistema di acquisizione online dei questionari. Il nucleo portante dell'architettura informatica del censimento è stato il Sistema di Gestione della Rilevazione (SGR), che ha consentito di seguire l'andamento di tutte le fasi del processo di produzione per singolo organo di censimento e operatore censuario, grazie alla realizzazione di oltre cinquanta funzioni componenti il sistema;
  - ▶ l'uso massivo degli indirizzi di posta elettronica certificata (PEC), non solo nel rispetto di un obbligo dettato dalla normativa vigente ma anche come indispensabile strumento di riduzione dei costi di una indagine statistica. Di questo strumento si è fatto un uso il più ampio possibile nelle varie fasi del processo produttivo, sebbene il quadro complessivo presenti una situazione a "macchia di leopardo". Infatti, tale strumento è stato usato in via esclusiva per l'universo delle imprese e delle istituzioni pubbliche, mentre è stato usato in modo selettivo per l'universo delle istituzioni non profit, che non hanno l'obbligo normativo di dotarsi di indirizzi PEC;
  - ▶ la valorizzazione del sistema di *data warehousing corporate* dell'Istat I.Stat (<http://dati-censimentoindustriaeservizi.istat.it>), grazie al quale l'utente può accedere liberamente e gratuitamente al rilevante patrimonio informativo censuario reso disponibile per la navigazione dei macrodati, per comporre tabelle ed effettuare confronti storici. Il censimento, con la sua volumetria, complessità e dettaglio di dati, ha usufruito della tecnologia adottata dall'Istat per il *data warehouse* I.Stat, seppur in una versione differente e resa più performante;
  - ▶ una strategia di comunicazione e informazione che ha integrato le azioni della pubblicità classica con le attività tipiche delle relazioni pubbliche, con azioni *ad hoc* realizzate per target specifici e con l'uso dei nuovi media, al fine di spiegare l'importanza del censimento, informare sulle modalità di svolgimento delle operazioni, sollecitare la partecipazione, rassicurare sulla tutela della privacy.

Vista l'impostazione generale del 9° Censimento generale, nel seguito vengono illustrati brevemente gli aspetti peculiari delle rilevazioni per singoli comparti.

### **Le imprese**

Consistente l'innovazione che riguarda le imprese per le quali l'Istat ha voluto superare la rilevazione censuaria tradizionale costituita dall'enumerazione esaustiva delle unità giuridico-economiche svolta sul campo, realizzando un "censimento virtuale" grazie all'utilizzo integrato di numerose fonti amministrative. Si è inteso realizzare così un progetto con approccio metodologico misto, che ha portato alla pubblicazione di dati desunti da archivi statistici e da rilevazioni campionarie tematiche assistite da archivi. I dati strutturali di base (natura giuridica, attività economica, occupazione, unità locali, localizzazione, fatturato) sono stati desunti dal sistema di registri Asia. Ciò ha consentito di fornire un quadro informativo statistico sulla struttura del sistema economico relativo alle imprese e alle unità locali fino al livello comunale e di sezione di

censimento e di garantire la confrontabilità dei censimenti economici, in particolare con quello del 2001. Oltre a questi dati strutturali sono state fornite ulteriori informazioni per l'analisi di specifici domini (in termini di settori di attività economica, tematiche di rilievo, eccetera), utilizzando informazioni amministrative finora parzialmente utilizzate e rilevazioni campionarie assistite da archivio. Il patrimonio informativo delle fonti amministrative è valorizzato in chiave statistica, restituendo all'utenza molte più informazioni rispetto a quelle prodotte tradizionalmente dal censimento, senza bisogno di farne oggetto di un questionario da compilare a cura di tutte le imprese italiane.

In particolare la valorizzazione del patrimonio informativo residente negli archivi amministrativi e le innovazioni metodologiche hanno consentito di realizzare un nuovo prodotto, denominato Asia-occupazione, che integra informazioni micro su impresa e occupati, realizzando un sistema informativo di tipo LEED (*Linked Employer Employee Database*). La disponibilità di nuove fonti amministrative con informazioni sul legame tra lavoratore e impresa e lo sviluppo di un quadro concettuale di definizioni e regole per il trattamento di tali dati a fini statistici hanno consentito lo sviluppo di un sistema di microdati integrati, dove è possibile identificare l'unità economica e l'unità lavoratore e dove ciascun individuo viene classificato in base alla tipologia occupazionale all'interno dell'impresa con cui, alla data di riferimento, ha un rapporto di lavoro. Anche queste nuove informazioni saranno d'ora in avanti prodotte e diffuse annualmente dall'Istituto grazie all'aggiornamento continuo dei dati amministrativi.

Per completare il quadro informativo sulle imprese sono state condotte tre rilevazioni campionarie: Indagine sulle unità locali delle grandi imprese (IULGI), Rilevazione multiscopo sulle imprese, Rilevazione sulle unità complesse. Questa struttura di rilevazioni ha consentito di cogliere importanti elementi conoscitivi del sistema produttivo italiano, solo parzialmente rilevabili sulla base del tradizionale impianto delle indagini strutturali sulle imprese.

L'Indagine IULGI è svolta dall'Istat con frequenza annuale per aggiornare l'Archivio statistico delle imprese attive (Asia) e delle unità locali ad esse afferenti (Asia-unità locali). Essa è nata nel 2004 per sopperire alle carenze di dati amministrativi sulle unità locali d'impresa, in particolare per quanto riguarda la distribuzione territoriale degli addetti. Con riferimento al 2011 la rilevazione annuale IULGI è stata ampliata per migliorare il controllo di qualità del registro statistico Asia-unità locali, in termini sia di errore di copertura sia di errore di misura. La rilevazione è stata condotta su un campione probabilistico di circa 40 mila imprese (circa l'1 per cento delle imprese italiane, pari a tre volte la numerosità di IULGI annuale), estratto da Asia-imprese, con probabilità di inclusione pari a 1 per le imprese con almeno 100 addetti.

Le altre due rilevazioni sono state progettate ed eseguite per la prima volta in occasione del 9° Censimento, dando la possibilità di cogliere la trama dei processi decisionali da cui dipendono le strategie aziendali, collocando le scelte su produzione, innovazione, capitale umano e sulle altre variabili competitive nel complesso delle relazioni di collaborazione o competizione che l'impresa intrattiene con il contesto che la circonda. Infatti, i recenti rapporti redatti da istituzioni internazionali, associazioni di categoria, network di ricerca rilevano come, dietro una apparente "staticità strutturale" delle imprese, in termini di dimensione media e settori prevalenti, il sistema industriale italiano nasconde mutamenti di grande rilevanza, che vanno verso una maggiore complessità e articolazione dei modelli di business. Le soluzioni e i percorsi evolutivi intrapresi dalle imprese presentano un'attenzione particolare alla flessibilità produttiva, all'innovazione, al

capitale umano e al controllo della tecnologia, che a sua volta richiede una conoscenza approfondita dell'intera filiera produttiva. Tali caratteristiche, peraltro, si compongono diversamente a formare nuovi modelli di attività e allo stesso tempo accomunano spesso tutte le classi dimensionali d'impresa. Mutamenti di questo tipo sono ampiamente diffusi nel settore produttivo italiano, ma l'informazione statistica ufficiale manca ancora di completezza rispetto a questi aspetti.

Più in particolare, mediante la Rilevazione multiscopo sulle imprese (rivolta a circa 260 mila imprese, pari al 6 per cento delle imprese italiane) si è inteso cogliere alcuni dei suddetti fenomeni, estendendo la disponibilità di dati relativi alle tecnologie ICT e all'Innovazione alle imprese con meno di 10 addetti, oltre a cogliere informazioni, alcune delle quali per la prima volta, come le strategie di *governance*, la gestione delle risorse umane, la competitività e la posizione sul mercato nazionale e internazionale, la capacità di innovazione, le scelte finanziarie, l'internazionalizzazione e le ragioni principali delle scelte di delocalizzazione<sup>1</sup>. Questa rilevazione è stata realizzata sul campo dagli Uffici di censimento costituiti appositamente a livello provinciale o regionale presso le Camere di commercio competenti per territorio e con il supporto a livello nazionale dell'Ufficio di statistica di Unioncamere.

Infine, con la Rilevazione sulle unità complesse, il 9° Censimento ha effettuato un investimento mirato a cogliere le peculiari caratteristiche di queste unità economiche anche nella prospettiva di razionalizzare i loro rapporti con l'Istat nella conduzione delle numerose rilevazioni correnti. Queste particolari unità sono spesso costituite nella forma di gruppi di impresa di livello nazionale o multinazionale ed hanno una complessa organizzazione interna delle attività e funzioni di supporto aziendale; tendono però a non essere adeguatamente rappresentate dalla statistica ufficiale sulla base dei tradizionali criteri e strumenti di classificazione e rilevazione. È stata pertanto realizzata una approfondita attività di progettazione che ha coinvolto anche accademici ed esperti nel settore dell'organizzazione aziendale. L'impianto di questa indagine presenta numerosi aspetti innovativi nella selezione del campione censuario, nei contenuti informativi del questionario, nelle modalità di conduzione della rilevazione. I fenomeni sono stati rilevati rispetto all'intero perimetro delle singole unità economiche complesse e coprono tutti i principali aspetti dell'organizzazione e della performance economica, in modo coerente con l'effettiva articolazione delle attività economiche. La raccolta di informazioni ha riguardato, tra gli altri aspetti, l'articolazione dell'unità economica complessa per linee di attività economica e funzioni aziendali (con riferimento sia alle modalità di internalizzazione e/o esternalizzazione in Italia e/o all'estero), le competenze e le capacità manageriali, la formazione dei processi decisionali, il capitale umano, i rapporti e le scelte strategiche adottate nei confronti dei fornitori e dei clienti, i processi di internazionalizzazione, il ruolo della conoscenza esteso a tutti i fattori intangibili di competitività. La rilevazione ha interessato circa 3.500 vertici di gruppi di impresa di tipo nazionale ed internazionale, incluse anche imprese indipendenti con dimensioni economiche rilevanti, che rappresentano una quota rilevante di imprese ed occupati residenti in Italia e all'estero. La raccolta dei dati è stata condotta sul campo da rilevatori con una elevata esperienza professionale nei rapporti con il top management aziendale, selezionati mediante una gara internazionale.

---

<sup>1</sup> Questi argomenti, unitamente alle nuove informazioni prodotte con il sistema dei registri Asia, sono stati analizzati nel convegno "Check-up delle imprese italiane", tenutosi il 28 novembre 2013 a Milano. Sono stati presentati diversi lavori, articolati in due focus tematici su "Struttura, organizzazione, risorse umane" e "Performance e strategie", oltre a tre sessioni di lavoro parallele, rivolte a studiosi e ricercatori, dal titolo: "Strategie di competizione e cooperazione tra le imprese", "Globalizzazione e mercati", "Imprese, occupazione e territorio". Gli interventi sono disponibili all'indirizzo del sito Istat <http://www.istat.it/it/archivio/103067>.

### **Le istituzioni non profit**

Il Censimento delle istituzioni non profit costituisce un'occasione di particolare rilevanza nel panorama della produzione statistica nazionale e internazionale. Con l'eccezione di alcune indagini annuali condotte dall'Istat tra la fine degli anni '90 e la metà degli anni 2000, riferite ad alcuni specifici comparti del non profit (organizzazioni di volontariato, cooperative sociali, fondazioni), non ci sono nel Paese informazioni di quadro generale come quelle prodotte dal censimento<sup>2</sup>. In questo senso esso permette di colmare un vuoto informativo su quello che è ormai divenuto il "modello non profit" di produzione e di fornitura dei servizi, in grado di creare opportunità di lavoro anche in un periodo di grande difficoltà generale. Oggi molte attività rilevanti legate ai bisogni sociali della collettività sono svolte da soggetti istituzionali che non si pongono fini di lucro, in grado di offrire lavoro retribuito e in larga parte di operare grazie al lavoro volontario dei cittadini. Ancora poco conosciuto, il mondo del non profit in Italia è caratterizzato da diversità sia dei soggetti che lo compongono sia delle loro modalità di azione. Fortemente dicotomico, il settore è caratterizzato dalla presenza contestuale di poche istituzioni di grandi dimensioni con strutture organizzative complesse e - sia in termini di occupazione sia economici - e di tantissime organizzazioni di piccole dimensioni, spesso attive solo grazie all'esclusivo apporto volontario dei cittadini.

Anche a livello internazionale, la rilevazione censuaria colloca l'Italia in posizione di avanguardia nel panorama delle informazioni statistiche ufficiali e affidabili sul settore, in linea con le definizioni e le indicazioni fornite dal *System of National Accounts*, dal Sistema europeo dei conti economici e dal Manuale sulle istituzioni non profit<sup>3</sup>.

Il censimento è stato condotto a partire dalla lista precensuaria delle istituzioni non profit, la cui realizzazione si è rivelata particolarmente complessa. Il settore è, infatti, caratterizzato da una forte eterogeneità nella qualità delle fonti amministrative utili, alcune disponibili a livello nazionale ma con copertura parziale, altre gestite da enti territoriali sub nazionali con maggiori problemi di qualità. Sono state 17 le fonti amministrative utilizzate per realizzare la lista precensuaria, alcune di grandi dimensioni, altre costituite da poche specifiche unità.

La tecnica di rilevazione si è basata sull'invio postale dei questionari long form e sulla loro compilazione su supporto cartaceo o informatico, con intervento mirato dei rilevatori per la compilazione e il recupero dei questionari non ancora pervenuti. Come da tradizione, la rete di rilevazione è stata costituita, in cooperazione con l'Ufficio statistico di Unioncamere, dagli uffici di censimento di livello provinciale o regionale, operanti presso le Camere di commercio.

L'importanza del censimento è ben evidenziata anche dagli obiettivi colti: rispondere alle esigenze informative di *policy maker*, studiosi, esperti e operatori del settore riguardo a dimensioni economiche e sociali, caratteristiche organizzative e peculiarità delle istituzioni non profit nel nostro Paese; definire la consistenza del settore non profit in Italia e pervenire alla sua rappresentazione statistica in termini di grandezze confrontabili con quelle delle precedenti rilevazioni censuarie del 1999 e del 2001; rilevare informazioni utilizzabili per la predisposizione dei conti nazionali e in

<sup>2</sup> Con riferimento al 1999 l'Istat ha effettuato la prima rilevazione censuaria delle istituzioni non profit (Cfr. "Istituzioni non profit in Italia", Istat, Collana Informazioni, n. 50, 2001) seguita dal 8° Censimento generale dell'industria e servizi nel cui campo di osservazione esse erano comprese (Cfr. *data warehouse* all'indirizzo del sito Istat <http://dwcis.istat.it/cis/index.htm>).

<sup>3</sup> United Nations, Department of Economic and Social Affairs – Statistics Division, *Handbook on Non-profit Institutions in the System of National Accounts*, Studies in Methods, Series F., No. 91, New York 2003.

particolare per la costruzione del conto satellite sulle istituzioni non profit; costruire un sistema di statistiche comparabili a livello internazionale; verificare la copertura e qualità della lista precensuaria in vista della realizzazione del registro statistico delle istituzioni non profit come parte integrante ed annualmente aggiornata del sistema dei registri sulle unità economiche.

Il patrimonio informativo raccolto con il censimento consente di analizzare la natura non profit delle unità istituzionali censite e di classificarle rispetto alle principali variabili di natura euristica, quali la tipologia di attività economica svolta (*market/non market*), l'orientamento (mutualistico/di pubblica utilità), la fonte di finanziamento prevalente (pubblico/privato); di rilevare le attività svolte (prevalenti e secondarie), i servizi erogati e i destinatari di tali servizi; di misurare l'entità delle risorse umane impiegate (sia retribuite sia volontarie) e le loro caratteristiche principali (categorie e profili professionali); di conoscere il numero e le caratteristiche dei soci; di rappresentare le dimensioni economico-finanziarie e le varie componenti di bilancio in termini di entrate e uscite.

La ricchezza informativa offerta si è rivelata pienamente in diverse occasioni seminariali e convegnistiche. La principale, organizzata dall'Istat, è stata il convegno nazionale "Il non profit in Italia. Quali sfide e quali opportunità per il Paese" svoltosi a Roma il 16 aprile 2014<sup>4</sup>. A livello internazionale, in occasione della Undicesima Conferenza Internazionale della *International Society for Third Sector Research* (ISTR), tenutasi a Muenster il 22-25 luglio 2014, è stato organizzato un Panel dal titolo "A statistical framework for the analysis of the Third Sector: the findings of the 2011 Italian Census", durante il quale sono stati presentati dei lavori realizzati da studiosi del settore e ricercatori Istat su diversi aspetti del non profit italiano e sull'esperienza statistica italiana nel settore.

### **Le istituzioni pubbliche**

Nell'ambito del 9° Censimento, l'Istat ha condotto una rilevazione *ad hoc* sulle istituzioni pubbliche, con un questionario da compilare esclusivamente online dai referenti di ciascuna amministrazione, senza l'intervento di rilevatori e con l'uso di strumenti di *e-learning* disponibili nel portale del censimento quale supporto alla formazione dei referenti di amministrazione.

Le amministrazioni coinvolte nella rilevazione sono state inserite dall'Istat in una lista precensuaria predisposta sulla base dell'integrazione di 10 archivi amministrativi e di altre fonti statistiche. Come in tutti i precedenti censimenti, anche in occasione di questa edizione non sono rientrate nel campo di osservazione le informazioni relative alle Forze armate, alla Guardia di finanza, ai Corpi di polizia e di sicurezza.

L'impianto organizzativo è stato completamente nuovo. Infatti, la rilevazione è stata articolata in due fasi distinte: la prima fase, svolta tra il 26 luglio e il 20 settembre 2012, aveva lo scopo di aggiornare le informazioni anagrafiche dell'unità istituzionale presente nella lista precensuaria, individuare il nominativo del dirigente o funzionario incaricato del coordinamento della rilevazione per l'unità istituzionale, acquisire l'elenco delle unità locali di ciascuna unità istituzionale e i nominativi dei referenti di ciascuna unità locale. La seconda fase, iniziata il primo ottobre 2012, ha riguardato la compilazione dei questionari di unità istituzionale e di unità locale.

<sup>4</sup> Cfr. gli interventi diffusi sul sito dell'Istat <http://www.istat.it/it/archivio/116695>.

In contenuti informativi del censimento sono stati fortemente innovativi. Numerose le tematiche specifiche che hanno consentito di fornire utili informazioni per una migliore conoscenza dell'organizzazione e funzionamento delle amministrazioni pubbliche italiane. Si evidenziano le informazioni sui percorsi formativi nella PA (quali istituzioni hanno formato il proprio personale, quali tematiche, quante ore, quante risorse umane coinvolte), sul grado di informatizzazione (dotazione e uso di ICT sia nelle unità istituzionali sia nelle singole unità locali), sull'amministrazione sostenibile intesa come tutela e sicurezza sul luogo di lavoro, sostenibilità ambientale, trasparenza e qualità dell'azione amministrativa (*accountability*).

Inoltre, il censimento ha consentito di tracciare la mappatura territoriale delle unità locali della pubblica amministrazione (cioè dei luoghi di lavoro e delle risorse umane ivi impegnate) e quella dei servizi pubblici offerti presso ciascuna unità locale (quali servizi, da chi, dove, quante risorse umane impegnate e il bacino d'utenza raggiunto). Queste informazioni sono rilevabili soltanto attraverso una operazione censuaria, che consente di effettuare rielaborazioni uniche in campo economico, grazie alla profondità informativa che si spinge fino ai dati a livello di categoria di attività economica e comune di ubicazione delle unità locali, cioè, appunto, dei luoghi di lavoro. Le analisi condotte utilizzando la variabile "unità locale" e "addetti alle unità locali" sono le più indicate per le analisi territoriali e settoriali perché l'informazione è quella di pertinenza, cioè quella che "insiste" direttamente sull'economia e sulle caratteristiche infrastrutturali e sociali del territorio di residenza dei cittadini e di ubicazione delle imprese e delle istituzioni.

In considerazione della crescente richiesta di informazione statistica sull'organizzazione e sul funzionamento delle amministrazioni pubbliche è necessario che l'Istituto si doti di strumenti innovati per fornire queste informazioni non più a carattere decennale, come è tradizione dei censimenti italiani, ma con scadenze più ravvicinate nel tempo, al fine sia di aggiornare periodicamente la mappa dei luoghi di lavoro e dei servizi offerti dalle istituzioni pubbliche sia di fornire informazioni statistiche di interesse per i decisori pubblici e gli esperti di settore. Anche da questa esigenza deriva la nuova strategia dell'Istituto volta alla realizzazione di censimenti economici continui, come evidenziato in sintesi nel seguito.

### ***Il sistema informatico censuario***

Centro nevralgico dell'organizzazione censuaria è stato il Sistema di Gestione della Rilevazione (SGR), utilizzato autonomamente da ciascun organo di censimento, rilevatore o coordinatore, che ha avuto accesso alle funzioni e alle tabelle di pertinenza attraverso proprie chiavi di accesso (user id e password).

In generale, i sistemi realizzati per la gestione delle rilevazioni del 9° Censimento sono stati sviluppati utilizzando prevalentemente tecnologie software *open source*, librerie di sviluppo e schemi di progetto standard. Essi si sono basati su un'infrastruttura completamente centralizzata che ne ha consentito l'accesso e l'utilizzo a tutte le persone dotate di un personal computer o tablet, di una connessione a internet e di un browser per la navigazione, senza dover installare software o configurare *ad hoc* il proprio dispositivo elettronico. Di fatto si è costruito un sistema integrato, parametrizzato e standardizzato in grado di dare supporto a tutte le attività legate alla fase di acquisizione dei dati e in alcuni casi anche di loro controllo, minimizzando le attività di sviluppo di codice specifico e massimizzando il riuso dei software già realizzati per il Censimento dell'agricoltura e per il Censimento della popolazione, in particolare per i sistemi di gestione della rilevazione e i sistemi di generazione e presentazione dei questionari elettronici.

Per supportare tutte le attività di raccolta dei questionari compilati e di monitoraggio della rilevazione sono stati realizzati due sistemi web direttamente accessibili attraverso il portale principale del censimento: il primo sistema è stato dedicato alla Rilevazione sulle imprese e al Censimento delle istituzioni non profit; il secondo sistema è stato approntato per il Censimento delle istituzioni pubbliche. I sistemi hanno permesso di gestire una rete di rilevazione articolata composta di 18 Uffici Regionali di Censimento (Istat) e di 103 Uffici Provinciali di Censimento, dei quali 99 costituiti presso le Camere di commercio, 2 presso le Unioni regionali delle Camere di commercio e 2 costituiti presso i Servizi di statistica delle Province autonome di Trento e Bolzano. Gli Uffici Regionali di Censimento hanno impiegato 166 addetti, mentre gli Uffici Provinciali di Censimento hanno creato una rete composta di quasi tremila addetti al censimento, di cui 2.257 rilevatori. Il sistema di acquisizione telematico ha portato alla registrazione online via internet dei questionari di tutte le unità di rilevazione. In particolare, ha scelto di compilare il questionario online il 78,8 per cento delle imprese e il 58,9 per cento delle istituzioni non profit, mentre il resto dei questionari è stato compilato online direttamente dagli Uffici di censimento, che hanno registrato i questionari cartacei compilati dalle imprese e dalle istituzioni non profit tramite l'utilizzo dello stesso sistema di acquisizione predisposto per le unità di rilevazione. Il Censimento delle istituzioni pubbliche si è basato invece su un sistema di compilazione e acquisizione dei dati esclusivamente online (*paperless census*).

### **Conclusioni e prospettive**

Le numerose innovazioni di prodotto e di processo descritte hanno permesso di accrescere la tempestività della diffusione dei dati definitivi anticipando rispetto al passato la restituzione delle informazioni. Il 10 luglio del 2013 sono stati diffusi i principali dati a livello di unità istituzionali, a 4 mesi dalla data di chiusura delle operazioni censuarie, a 11 mesi dall'inizio delle rilevazioni sul campo e a 18 mesi dalla data di riferimento del censimento. Il 30 ottobre 2013 sono stati diffusi analoghi dati a livello di unità locali e i dati tematici sulle imprese desunti dalla Rilevazione multiscopo. Il 21 dicembre 2013 i principali dati tematici sulle istituzioni pubbliche e all'inizio del 2014 i dati relativi all'intero questionario delle istituzioni non profit.

Numerose le occasioni, nazionali e internazionali, nelle quali rappresentanti del mondo della politica, della ricerca, delle imprese e delle istituzioni, pubbliche e non profit, oltre a ricercatori Istat, hanno avuto modo di confrontarsi, discutere e analizzare la struttura produttiva del Paese così come emersa dai dati censuari. Oltre agli eventi citati in precedenza, l'Istat ha organizzato un ciclo di incontri svolti in tutte le regioni italiane, da maggio a luglio 2014, per diffondere sul territorio i risultati del 9° Censimento generale dell'industria e dei servizi. Ciascun appuntamento ha rappresentato un'opportunità di confronto e di dibattito sul contributo offerto dal censimento alla conoscenza del sistema produttivo locale.

Al termine della tornata censuaria 2010-2011 è sorta la necessità di dare continuità alle numerose e rilevanti innovazioni di metodo e di processo realizzate. Il 9° Censimento ha costituito un banco di prova significativo di processi produttivi articolati e complessi che ben si adattano alla fase successiva dei processi statistici connessi alla produzione di informazioni statistiche sulla struttura del sistema produttivo italiano.

L'Istituto nazionale di statistica è oggi impegnato nella progettazione di una nuova stagione dei censimenti continui i cui punti cardine riguardano l'uso integrato di indagini e di informazioni di fonte amministrativa. In particolare, per quanto riguarda le unità giuridico-economiche l'attività può essere divisa in due macro filoni: a) la realizzazione annuale di un sistema integrato di registri statistici per le diverse tipologie di unità economiche (imprese, aziende agricole, istituzioni pubbliche, istituzioni non profit); b) la realizzazione annuale o biennale di indagini necessarie non solo a verificare la qualità delle fonti amministrative e a garantire l'eshaustività dei registri da esse derivanti, ma anche a completare le informazioni in esse non presenti. Le esperienze acquisite e gli investimenti realizzati con il 9° Censimento costituiscono il punto di riferimento sia per l'aggiornamento annuale del sistema dei registri, sia per lo svolgimento delle rilevazioni periodiche.



## Introduzione

L'innovazione tecnologica e la conseguente possibilità di realizzare applicazioni web complesse e di trattare in rete anche grandi moli di dati ha consentito la predisposizione di applicazioni per la gestione delle rilevazioni censuarie e la compilazione online dei questionari, caratterizzate da elevata scalabilità ed efficienza e completamente fruibili via internet.

I sistemi realizzati per la gestione delle tre rilevazioni del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit si basano su un'infrastruttura completamente centralizzata che consente l'accesso tramite personal computer o tablet con una connessione a internet e un browser per la navigazione, senza dover installare software o configurare *ad hoc* il proprio dispositivo elettronico. Di fatto si è costruito un sistema integrato, parametrizzato e standardizzato in grado di supportare tutte le attività legate alla fase di acquisizione e in alcuni casi anche di controllo dei dati minimizzando le attività di sviluppo di codice specifico e massimizzando il riuso di software già realizzato per il Censimento dell'agricoltura e per il Censimento della popolazione, specialmente per quanto riguarda i sistemi di gestione della rilevazione e i sistemi di generazione e presentazione dei questionari elettronici.

I sistemi censuari sono stati sviluppati utilizzando prevalentemente tecnologie software *open source*, 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 e consolidati nel 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit.

Per supportare tutte le attività di raccolta e monitoraggio sono stati realizzati due sistemi web direttamente accessibili attraverso il portale principale del censimento: il sistema dedicato alle imprese e alle istituzioni non profit e il sistema dedicato alle istituzioni pubbliche.

Il primo sistema web è composto da tre sottosistemi web: il portale di documentazione SGRRETE, accessibile dagli operatori del censimento, contenente tutte le informazioni e i documenti utili per lo svolgimento delle operazioni censuarie; il sistema di gestione SGRCIS, concepito come un sistema di *workflow* distribuito, che ha consentito a tutti gli operatori della rete di lavorare autonomamente seguendo un percorso ben delineato, standardizzato e completamente guidato e il sistema di acquisizione QCIS per la compilazione elettronica dei modelli di rilevazione da parte di imprese e istituzioni non profit. Esso si è dimostrato, dal punto di vista applicativo, un sistema robusto e in grado di supportare in modo adeguato le varie fasi tipiche di rilevazioni multicanale, multi questionario, con una rete di rilevazione complessa e distribuita sull'intero territorio nazionale. Il secondo sistema web è composto da due moduli e ha consentito di gestire interamente su internet le due fasi distinte del Censimento delle istituzioni pubbliche: i) la fase di definizione e costruzione dell'universo delle unità locali; ii) la fase di compilazione online di tutti i questionari, sia delle unità istituzionali che delle unità locali, in assenza di una rete di rilevazione.

Il progetto si è articolato nelle seguenti fasi:

- ▶ raccolta dei requisiti utenti e definizione delle specifiche funzionali;
- ▶ individuazione degli strumenti standard di progettazione, sviluppo software, modellazione dei dati, documentazione e *testing*;
- ▶ definizione dei requisiti di dettaglio, in cui sono state elencate e descritte tutte le singole funzioni da implementare;
- ▶ progettazione della base dati, ovvero del modello logico-fisico nel rispetto della metodologia delle basi dati relazionali e predisposizione del database;

- ▶ progettazione architettonica: ortogonalmente e contemporaneamente alle fasi precedenti è stata definita l'architettura del software in termini di componenti di base e *middleware*, standardizzazione delle versioni, *tuning* e ottimizzazione delle interazioni tra i vari componenti;
- ▶ sviluppo dei moduli applicativi, corredato di test e casi d'uso;
- ▶ *testing* sia di funzionalità sia di sicurezza e di carico;
- ▶ esercizio dei sistemi.

Nel fascicolo vengono descritti in modo dettagliato i sottosistemi di acquisizione e gestione e i sistemi realizzati per la diffusione online dei dati e il popolamento del *data warehouse* di diffusione web dell'Istat I.Stat.

In particolare la Parte Prima è dedicata alla Rilevazione sulle imprese e al Censimento delle istituzioni non profit ed è organizzata in tre distinti capitoli: il primo riporta la descrizione dell'architettura generale del sistema e dei suoi componenti (la metodologia di *testing*, i sistemi di autenticazione e di accesso e le funzioni di formazione online per la rete di rilevazione); il secondo riguarda la predisposizione dei questionari elettronici e il terzo è dedicato alla descrizione del Sistema di Gestione della Rilevazione e delle funzioni che lo compongono.

Nella Parte Seconda viene descritto il sistema di acquisizione e di monitoraggio delle due fasi del Censimento delle istituzioni pubbliche.

Nella Parte Terza vengono affrontate le problematiche relative alla diffusione dei dati e illustrate le principali soluzioni adottate.

PARTE PRIMA

I SISTEMI INFORMATICI DI ACQUISIZIONE E GESTIONE DELLA  
RILEVAZIONE SULLE IMPRESE E DEL CENSIMENTO DELLE  
ISTITUZIONI NON PROFIT



## Capitolo 1

### L'Architettura generale dei sistemi informatici<sup>5</sup>

#### 1.1 L'architettura funzionale

Per la conduzione della Rilevazione delle imprese e del Censimento delle istituzioni non profit, è stata predisposta una piattaforma informatica web completamente integrata contenente tre sottosistemi web distinti ed accessibili, mediante credenziali di accesso, attraverso un portale unico all'indirizzo

<http://censimentoindustriaeservizi.istat.it> (Figura 1.1). I sistemi sono stati messi a disposizione di due distinte tipologie di utenza: rispondenti (imprese e istituzioni non profit) e operatori della rete di rilevazione. I sistemi web sono:

- ▶ *Portale della rete* (RETE). Sito dedicato agli operatori della rete di rilevazione (<https://censimentoindustriaeservizi.istat.it/rete>) contenente tutta la documentazione utile relativa alla rilevazione sulle imprese e al Censimento sulle istituzioni non profit;
- ▶ *Sistema di Gestione della Rilevazione* (SGRCIS). Sito dedicato agli operatori della rete di rilevazione (<https://censimentoindustriaeservizi.istat.it/sgr>) per l'effettuazione di tutte le funzioni di acquisizione, controllo, correzione dei dati e monitoraggio delle diverse fasi di lavorazione. SGRCIS è concepito come un sistema di workflow distribuito, che ha permesso a tutti gli operatori della rete di lavorare autonomamente seguendo un percorso ben delineato, standardizzato e completamente guidato. La sua struttura modulare ha consentito inoltre ulteriori implementazioni in corso d'opera senza stop del sistema e modifiche a quanto già rilasciato e funzionante;
- ▶ *Sistema di acquisizione online* (QCIS). Sistema utilizzato sia dalle imprese e dalle organizzazioni non profit per la compilazione web del questionario ma anche dagli operatori della rete per l'inserimento dei dati raccolti attraverso i questionari cartacei (<https://censimentoindustriaeservizi.istat.it/questionario>). Il sistema ha permesso 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.

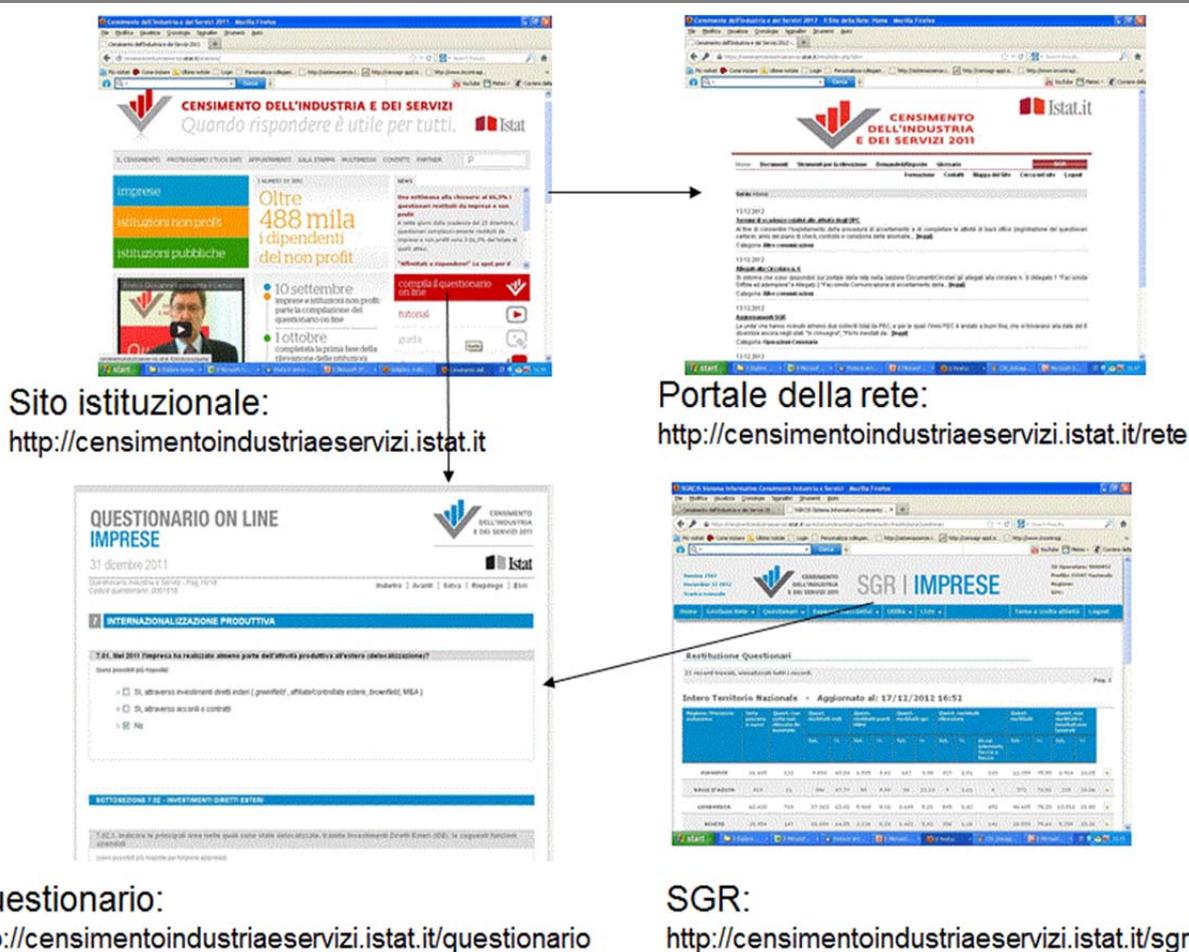
Ogni applicazione è stata realizzata garantendo la massima sicurezza dei dati nelle fasi di trasmissione e di conservazione, secondo le regole standard dell'Istituto. Tutti i siti sono stati sviluppati utilizzando tecnologie *open source*, permettendo un risparmio nei costi rispetto ad una scelta di un software proprietario.

<sup>5</sup> A cura di Marina Venturi. Autori del capitolo:

- Paragrafo 1.1: Marina Venturi;
- Paragrafo 1.2: Marco Silipo (Paragrafo 1.2.1), Mauro Bruno (Paragrafo 1.2.2), Daniele Ingretolli e Rossana Vignola (Paragrafo 1.2.3);
- Paragrafo 1.3: Raffaella Rosati (Paragrafo 1.3.1), Leonardo Tininini (Paragrafo 1.3.2);
- Paragrafo 1.4: Fausto Panicali e Donato Moscara;
- Paragrafo 1.5: Manlio Paterniti (Paragrafo 1.5.1), Ferdinando Ruggiero (Paragrafo 1.5.2), Bernardo Palazzi (Paragrafo 1.5.3), Sara Letardi e Fausto Panicali (Paragrafo 1.5.4);
- Paragrafo 1.6: Mauro Bruno e Marco Silipo;
- Paragrafo 1.7: Antonella Bianchino e Giulia De Candia.

Il sistema realizzato si è basato su una infrastruttura centralizzata che ha consentito l'accesso e l'utilizzo dell'intero sistema a tutte le persone in possesso di un personal computer o tablet di ultima generazione, connessione a internet e un browser a scelta tra i più diffusi sul mercato per la navigazione, senza nessuna installazione aggiuntiva di software o particolari configurazioni del proprio dispositivo elettronico. Sebbene le infrastrutture centralizzate abbiano alcuni inconvenienti, quali la maggiore criticità di un corretto dimensionamento per garantire un'adeguata affidabilità del servizio, la scelta di questo approccio è stata motivata da ragioni organizzative, economiche, di complessità e consistenza.

Figura 1.1 - Architettura funzionale



## 1.2 L'architettura software

La progettazione di sistemi web con un elevato grado di complessità come il software di gestione della rilevazione (SGRCIS) ed il questionario online (QCIS), ha richiesto un'attenta analisi riguardante diversi aspetti: definizione dei casi d'uso, definizione del processo statistico, valutazione dei rischi, progettazione dell'architettura hardware e software, eccetera.

Un ulteriore aspetto di criticità nell'implementazione di un sistema di tali dimensioni e importanza è 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.

Dal punto di vista architetturale, sono stati realizzati sistemi replicati dotati di un meccanismo di bilanciamento del carico. Il *middleware* alla base del sistema era fondato su uno *stack* tecnologico consolidato dalle precedenti esperienze censuarie che non ha richiesto onerose rielaborazioni del codice sorgente.

### 1.2.1 L'architettura del sistema di gestione e del questionario

Come mostrato in Figura 1.2, il Sistema di Gestione delle Rilevazioni sulle imprese e del Censimento delle istituzioni non profit SGRCIS è stato il nucleo dell'architettura complessiva, integrando i suoi moduli con l'applicazione del questionario (QCIS) usata sia dai rispondenti che dagli operatori autorizzati della rete di rilevazione, impegnati nel data entry dei questionari cartacei di loro competenza. L'integrazione tra il questionario online e il Sistema di Gestione della Rilevazione ha consentito un costante monitoraggio in tempo reale dell'andamento delle compilazioni.

La gestione dei dati per SGRCIS e QCIS si è basata su una infrastruttura Oracle molto complessa, accuratamente ottimizzata per gestire la grande mole di accessi prevista da parte di operatori e rispondenti senza stop o sovraccarico dei sistemi.

L'infrastruttura software di SGRCIS e QCIS si è basata sul design pattern Model-View-Controller e sulla tecnologia *Java Enterprise Edition* (JEE). 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 e durante il 15° Censimento della popolazione e delle abitazioni. Il team di sviluppo responsabile della realizzazione del Sistema di Gestione della Rilevazione ha, infatti, migliorato il software in termini di sicurezza e qualità, riutilizzando le componenti principali dell'infrastruttura e al tempo stesso adottando *framework open source* ampiamente diffusi (*Struts2*, *Spring* e *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.

La Figura 1.3 mostra l'integrazione tra i *layer* applicativi e i *framework* utilizzati. Il *layer* di accesso alla base dati ha avuto lo scopo di astrarre la complessità del DBMS facilitando le operazioni di interrogazione attraverso un semplice modello a oggetti Java, chiamati "*beans*"; gli sviluppatori delegavano al *framework* la generazione delle *query* SQL, utilizzando la sintassi Java.

Figura 1.2 – Architettura complessiva del sistema

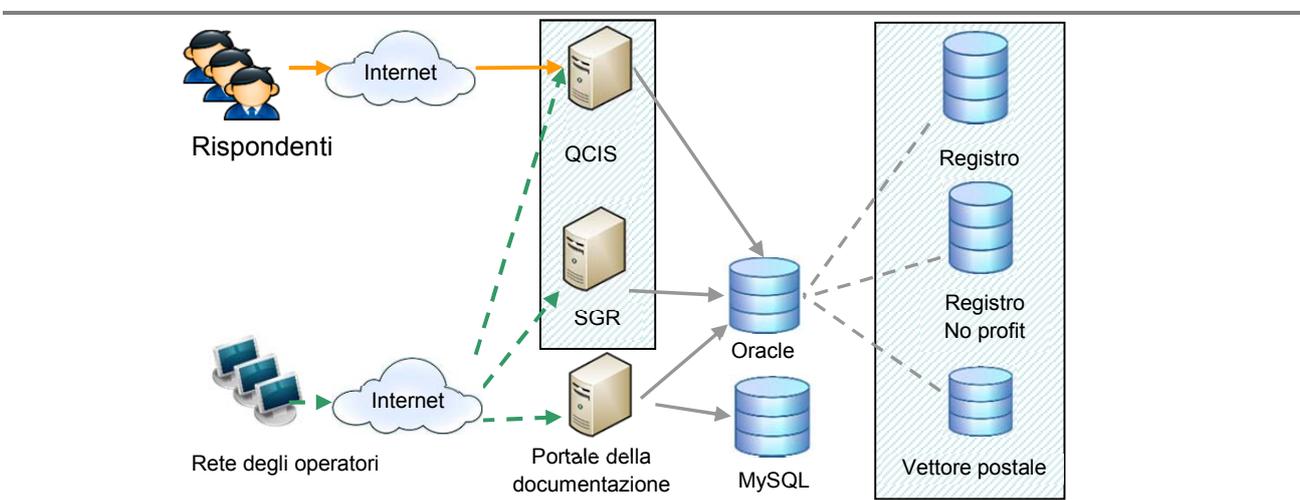
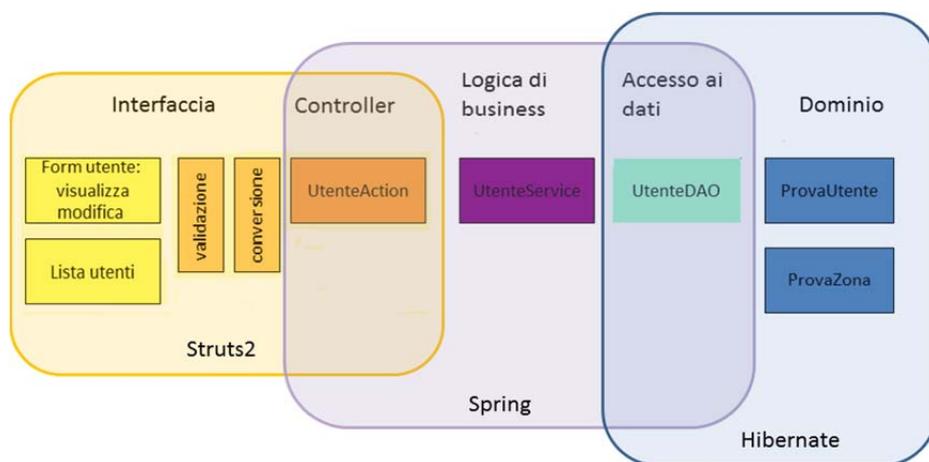


Figura 1.3 – *Stack* tecnologico degli applicativi SGRCIS e QCIS

In particolare il *framework* per l'*Object-relational mapping* (*Hibernate*) ha permesso 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* ha consentito di limitare e semplificare notevolmente il codice per l'accesso alla base dati, che è risultato di conseguenza anche più facilmente manutenibile.

Il *framework Spring* ha permesso la gestione dichiarativa delle dipendenze tra le classi: gli oggetti utilizzati nell'ambito di una classe venivano automaticamente creati dal *framework* al momento dell'uso (*autowiring*). Attraverso file di configurazione *xml* *Spring* ha consentito la gestione della connessione alla base dati, rendendola trasparente al programmatore; inoltre il *framework* ha permesso 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, ha ridotto e semplificato il codice da scrivere, rendendolo più robusto, manutenibile e meno soggetto ad errori.

*Struts2* ha permesso l'organizzazione di una applicazione web in livelli, separando le funzionalità dei componenti di presentazione e di logica applicativa. Ha fornito le funzionalità di base per la gestione semplificata dell'acquisizione dei dati delle *form* e la loro validazione. Inoltre ha offerto 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 stato essenziale per la realizzazione di una applicazione web complessa.

Le pagine dell'applicazione sono state rese dinamiche attraverso l'utilizzo della tecnologia AJAX, che ha permesso una stretta interazione tra gli elementi dell'interfaccia utente, ad esempio bottoni, liste, menu a tendina, eccetera, 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 della 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 SGRCIS e il Portale della 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.

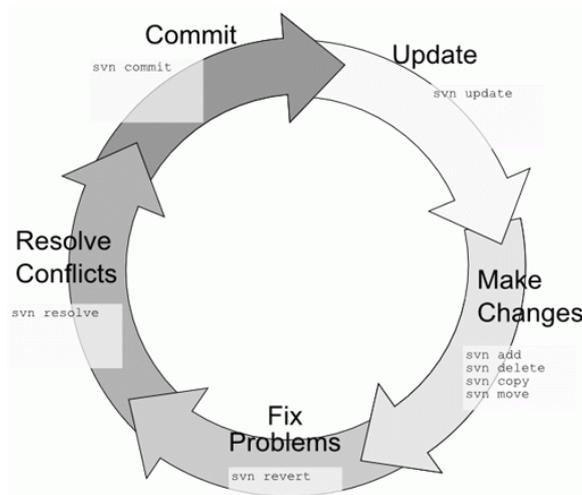
L'implementazione di SGRCIS e QCIS è stata realizzata applicando consolidate regole di sicurezza (OWASP 2013, SAFECode 2013) quali ad esempio: meccanismi di autenticazione, uso di algoritmi di crittografia, livelli di autorizzazioni basati sulla profilazione utente, protezione delle URL per impedire accessi non autorizzati alle varie funzionalità. In particolare 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 memorizzate 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 SGRCIS o di QCIS, loggato al 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 territorio di propria competenza o ai questionari assegnati.

### 1.2.2 Il sistema di versioning

Il lavoro in team degli sviluppatori è stato gestito attraverso strumenti di "software versioning". Si è data continuità allo strumento utilizzato per il Censimento dell'agricoltura e a quello della popolazione, ossia *Subversion (SVN)*.

Attraverso SVN il lavoro di ogni sviluppatore è stato mantenuto sempre allineato con quello degli altri elementi del team, potendo costantemente aggiornare il proprio ambiente di sviluppo locale con i cambiamenti apportati dagli altri colleghi (fase di "update"). Al tempo stesso ogni componente software di responsabilità dello sviluppatore, una volta completato e testato, poteva essere salvato e messo a disposizione del team di progetto (fase di "commit"). Con riferimento alla Figura 1.4, l'aggiunta di nuovi file o la loro cancellazione corrispondono alle operazioni di "add" e "delete" mentre la risoluzione di conflitti tra diverse versioni dello stesso file corrisponde alla operazione di "resolve". Generalmente queste operazioni venivano svolte in maniera trasparente direttamente dall'ambiente di sviluppo, attraverso interfacce grafiche semplificando così il lavoro dello sviluppatore.

Figura 1.4 – Processo di sviluppo collaborativo



La modalità di lavoro prescelta per il team ha seguito i processi più comuni nella comunità di sviluppo, in particolare considerando alcune “*best practices*” che nel tempo sono state discusse e affinate da chi si occupa di ingegneria del software e di strumenti di collaborazione.

Si elencano per brevità le principali direttive che il team di sviluppo si è dato:

- lo strumento di *versioning* non doveva essere utilizzato come un *backup* del proprio lavoro;
- il “*commit*” andava effettuato solo quando tutti i file cambiati costituivano una unità logica, come ad esempio la *fix* di uno specifico *bug* o il completamento dello sviluppo e test di una nuova funzionalità;
- commenti al lavoro fatto andavano specificati al momento del *commit*: essi dovevano essere precisi e descrittivi del task di progetto specifico a cui il caricamento faceva riferimento;
- prima del *commit*, il processo prevedeva di effettuare un update del lavoro altrui, ricompilare l'intero progetto (“*clean & build*”) e assicurarsi che la compilazione andasse a buon fine.

SVN è dotato di un proprio database interno basato su un sistema chiamato FSFS, che memorizza su *filesystem* solo i delta tra una versione e l'altra di ogni file, invece di replicare tutto il contenuto. Questo ha permesso di ottimizzare le prestazioni degli strumenti client.

In termini di organizzazione dei file, è stato possibile definire delle unità logiche di lavoro che rappresentavano delle istantanee scattate sul progetto, chiamate *branch*, mediante dei *tag* da assegnare a particolari attività completate sul codice.

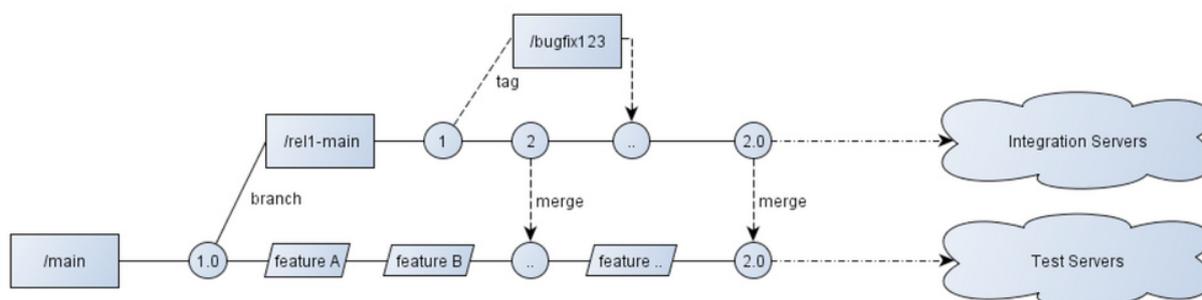
Gli sviluppatori hanno lavorato comunemente su quello che viene chiamato il *trunk* principale, ed era responsabilità di chi si occupava della gestione del progetto di creare delle *release* da inviare al team di test, estraendo e compilando dei *branch* precedentemente creati.

Il team di test poteva validare in questo modo una release e gli sviluppatori potevano risolvere i *bug* trovati direttamente in quella *release*. Al termine dei test veniva condotta una fase chiamata di “*merge*” in cui il contenuto del *branch* veniva riportato nel *trunk* principale.

Lo sviluppo delle nuove funzionalità infatti è continuato ad avvenire sul *trunk* principale.

Il processo era iterativo: nel momento opportuno veniva creato un nuovo *branch* per permettere agli sviluppatori di testare ulteriori nuove funzionalità. Questo processo è sintetizzato nell'immagine sottostante.

Figura 1.5 – Version control attraverso il software SVN



### 1.2.3 L'architettura del Portale della rete

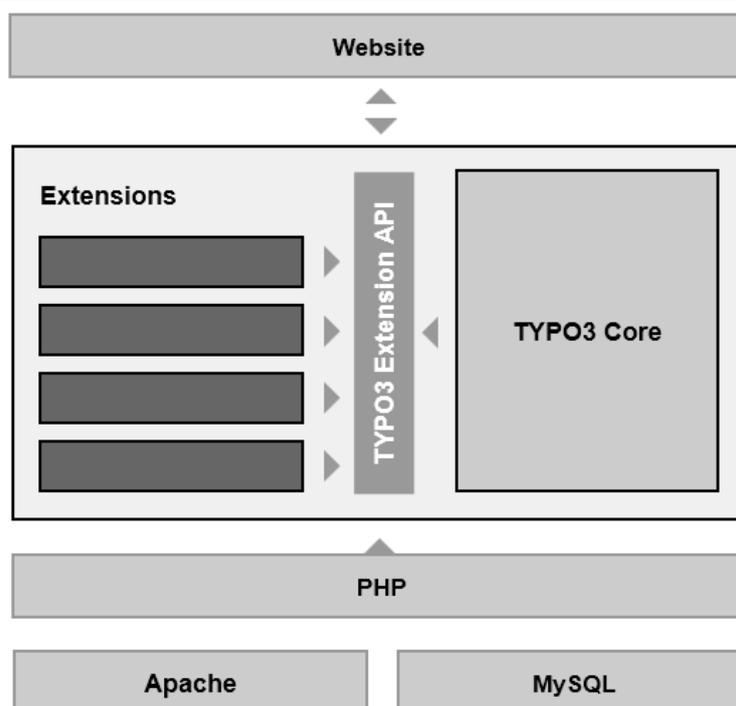
Il Portale della rete, accessibile dagli operatori della rete tramite user id e password personali, è stato realizzato utilizzando un meta ambiente, denominato TYPO3 (Figura 1.6), un CMS (*Content Management System*) per la costruzione e gestione di siti web di qualsiasi tipo, particolarmente indicato per portali.

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 *open source* che viene sviluppato continuamente.

Figura 1.6 – Architettura TYPO3



La gestione del sito si è basata sul *TYPO3 Extension Manager* incluso nel *TYPO3-Core* per installazione, attivazione e disattivazione di estensioni in una determinata installazione TYPO3.

L'installazione standard ha permesso 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 è stato possibile ricorrere a delle estensioni per integrare molte altre funzioni.

La gestione delle utenze è avvenuta in modo parallelo al sistema SGRCIS; infatti, poiché il cms usa una propria tabella per la gestione degli accessi si è dovuto creare una procedura per mantenerle allineate a SGRCIS.

Così ogni volta che veniva creata o eliminata un'utenza SGRCIS questa veniva replicata o eliminata dalla tabella degli utenti di TYPO3. Per tutte le altre funzionalità invece il Portale della rete è stato gestito in modo completamente indipendente dagli altri sottosistemi.

### 1.3 Il modello dei dati

#### 1.3.1 La base dati del sistema di gestione

Il database del sistema di gestione del censimento non profit e della rilevazione sulle imprese (SGRCIS) è di tipo relazionale, ossia le tabelle sono correlate tra loro attraverso dei campi chiave. La base dati, schematizzata in Figura 1.7, può essere raggruppata nelle seguenti strutture-dati, quasi tutte precaricate prima dell'inizio della rilevazione:

- ▶ tabelle territoriali: contenenti le informazioni di decodifica del territorio (regioni, province e comuni italiani);
- ▶ tabelle anagrafiche: contenenti i dati anagrafici, il profilo degli operatori, le utenze per accedere al sistema e le password criptate degli operatori;
- ▶ tabelle di decodifica: contenenti la descrizione delle variabili di SGRCIS, in questo gruppo erano incluse anche le tabelle lette dall'applicativo che permettevano la costruzione dinamica della pagina del sito.

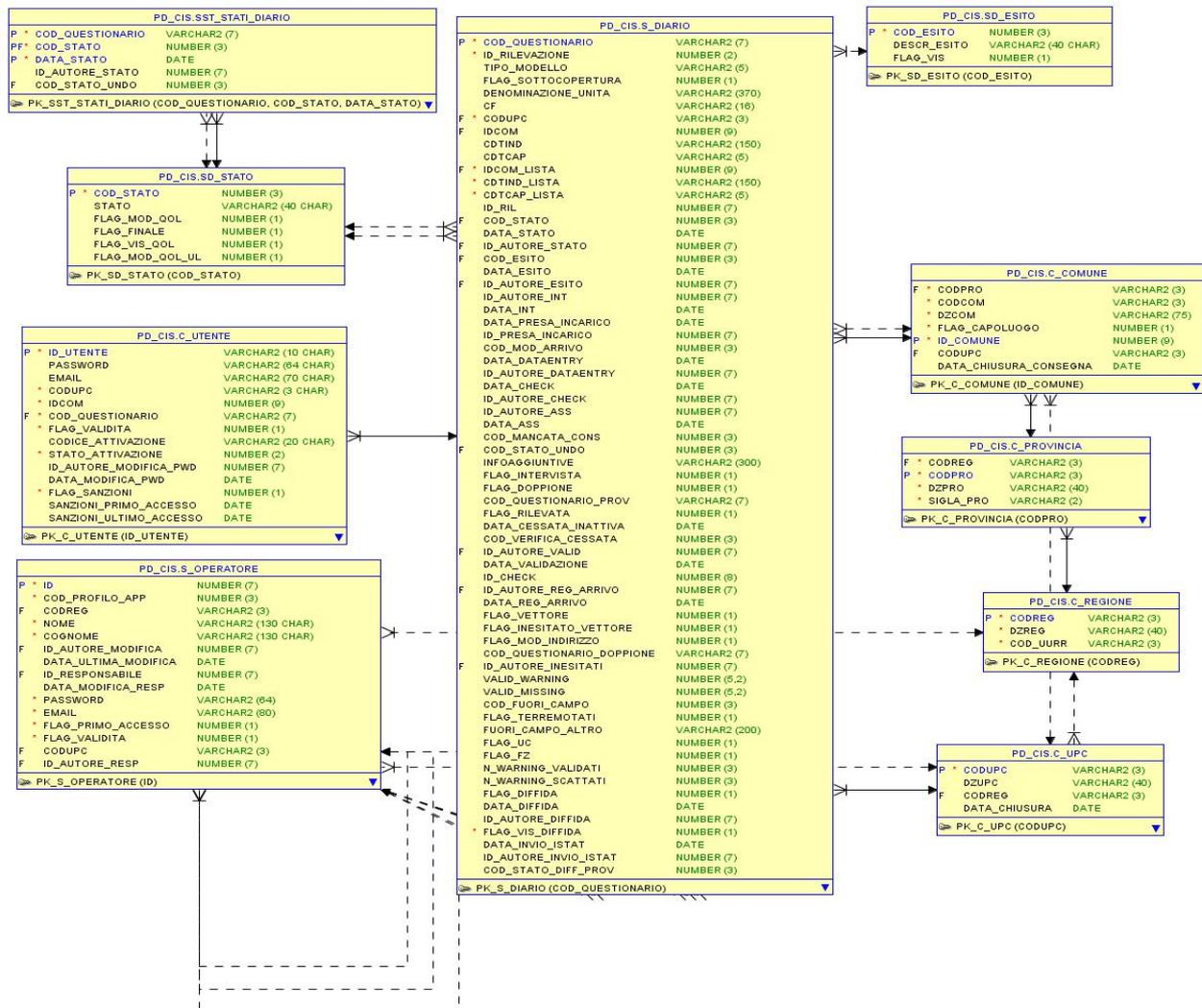
Le tabelle create erano identificabili all'interno della base dati attraverso il prefisso "SD" ed il nome del campo per le quali era descritta la decodifica, inoltre avevano una struttura standard composta di due colonne, la prima contenente i valori che la variabile di SGCIS poteva assumere e la seconda colonna di descrizione; nello specifico un'ulteriore suddivisione logica è stata tra:

- ▶ tabelle dei dati: contenenti i dati raccolti tramite l'applicazione;
- ▶ tabelle di gestione: comprendenti tutte le tabelle che consentivano all'applicazione di funzionare in modo parametrico e dinamico; appartenevano a questa categoria le tabelle con all'interno le informazioni per la creazione dinamica del menu in base all'utente connesso;
- ▶ tabelle storiche: i dati più significativi venivano salvati in tabelle storiche a ogni modifica effettuata. Questo ha consentito di poter ripristinare in caso di problemi dell'applicazione la situazione immediatamente precedente. Sono state storicizzate tutte le tabelle contenenti i microdati.

La base dati di SGRCIS oltre che da tabelle era formata da *trigger* e procedure che consentivano la gestione completa dell'applicazione. In particolare i *trigger* permettevano di automatizzare alcune operazioni senza demandare il controllo solo ed esclusivamente all'applicativo.

Tutte le operazioni di storicizzazione sono state demandate a questi oggetti. Le procedure Oracle, invece sono state sviluppate per il popolamento delle tabelle, e soprattutto per la gestione dei rapporti riassuntivi che consentivano il monitoraggio dei censimenti.

Figura 1.7 – Schema E/R: struttura



### 1.3.2 La base dati del questionario

Prima di analizzare la modellazione, e conseguente memorizzazione nella base dati, del questionario elettronico, è utile una premessa di carattere terminologico. Parlando di “questionario” ci si può infatti riferire sia alla parte strutturale (tipicamente definita componente dei metadati e che in una rilevazione di tipo tradizionale corrisponderebbe alle informazioni desumibili dal modello cartaceo inviato ai rispondenti) sia alla parte di contenuti (tipicamente nota come componente dei microdati e che corrisponde in pratica all’insieme di risposte che uno specifico rispondente fornisce, in forma più o meno codificata, ai vari quesiti proposti). La modellazione descritta nel seguito ha riguardato entrambi gli aspetti, producendo come risultato due sezioni distinte della base dati, strettamente correlate, ma con caratteristiche molto diverse.

La componente dei metadati strutturali è senz’altro la più complessa e verrà solo brevemente accennata in questo paragrafo, in quanto già oggetto di una descrizione più dettagliata nel Paragrafo 2.1. Essa ha adottato una tecnica di modellazione, basata sulla rappresentazione del

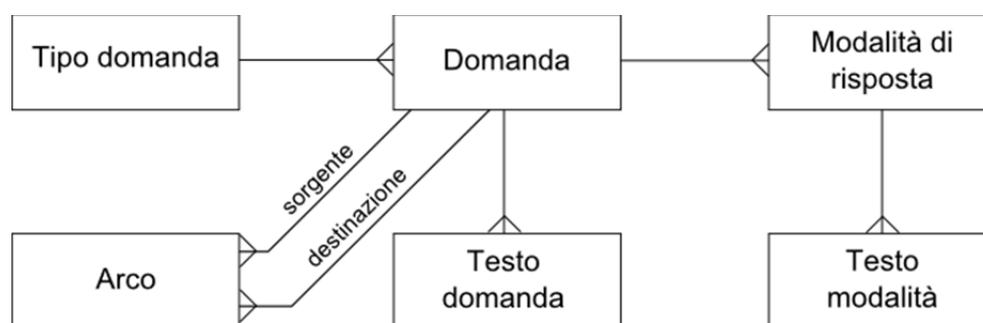
questionario in termini di un grafo diretto aciclico (DAG). In tale grafo i nodi-domande rappresentano i singoli quesiti, mentre gli archi descrivono il flusso di compilazione e le condizioni che ne regolano il funzionamento (in pratica un certo arco dovrà essere attraversato da un percorso di compilazione corretto solo se la condizione associata all'arco stesso è vera). La modellazione tramite grafo è stata utilizzata sia lato server che lato client per determinare l'attivazione e disattivazione delle domande ed eventuale reset di alcune risposte, guidando così l'utente finale nel processo di compilazione.

Il diagramma concettuale semplificato che rappresenta gli elementi fondamentali della modellazione di un questionario in termini di grafo è riportata in Figura 1.8. Viste le cardinalità di relazione presenti, ad ogni entità del modello ER corrispondeva ovviamente una relazione/tabella del diagramma logico-fisico.

Nella base dati era quindi presente una tabella specifica per la rappresentazione dei nodi-domande, corrispondenti sia ai quesiti veri e propri che ad altre componenti accessorie del questionario, quali istruzioni di compilazione, titoli di sezione/sottosezione, e anche componenti ausiliarie di modellazione. Si vedrà infatti, nel già citato Paragrafo 2.1, che in alcuni casi si è resa necessaria l'introduzione di nodi virtuali, non direttamente visibili nell'interfaccia utente e finalizzati unicamente ad esprimere alcune combinazioni di condizioni, cioè, in definitiva, ad ottenere il comportamento desiderato dell'applicazione.

Trattandosi di un grafo, era ovviamente presente nella base dati anche una tabella che rappresentava gli archi (orientati) del grafo stesso, cioè le transizioni da un quesito a un altro, per la descrizione del percorso di compilazione e delle condizioni (tipicamente basate sulle risposte fornite a uno o più quesiti) che determinavano la transizione medesima. Ogni arco era quindi caratterizzato da un nodo-domanda sorgente, da un nodo sorgente destinazione e da una condizione logica (booleana), espressa in un formalismo opportuno, che consentiva di determinare dinamicamente durante la compilazione, se l'arco dovesse/potesse essere attraversato da un percorso di compilazione corretto.

**Figura 1.8 – Il diagramma concettuale semplificato con gli elementi fondamentali della modellazione di un questionario**



A ogni nodo-domanda doveva inoltre essere associato uno specifico tipo di domanda (e i tipi di domanda disponibili erano memorizzati anch'essi in una tabella opportuna) che determinava la visualizzazione del quesito nell'interfaccia utente, il suo comportamento dal punto di vista della logica applicativa, nonché la tipologia di memorizzazione nella base dati della corrispondente

risposta fornita dall'utente. Esempi di tipi di domanda sono "domanda con scelta singola da una lista di risposte", "domanda con scelta multipla da una lista di risposte", "domanda con risposta numerica intera", eccetera. Tutte le domande che condividevano il medesimo tipo facevano riferimento di fatto ai medesimi componenti software per la visualizzazione dei dati e la loro memorizzazione, massimizzando il riuso e la robustezza del codice prodotto, oltre ovviamente a garantire un'elevata modularità ed estensibilità.

In corrispondenza di alcuni tipi (es. "domanda con scelta singola da una lista di risposte") alla domanda erano associate delle modalità di risposta prefissate, memorizzate in un'apposita tabella della base dati (unica e che contiene le modalità di risposta di tutte le domande). La tabella conteneva varie informazioni accessorie che consentivano di ottenere una presentazione nell'interfaccia web il più possibile conforme a quella utilizzata nel modello cartaceo. La stessa tabella delle modalità veniva utilizzata anche per descrivere i dati in testata e fiancata di quesiti espressi in forma matriciale.

Infine un'ulteriore componente fondamentale dei metadati del questionario è stata quella dei testi associati ai vari quesiti, testi che tipicamente dovevano poter essere presentati all'utente secondo "localizzazioni" diverse (cioè lingue: in particolare italiano, tedesco e sloveno) a seconda delle impostazioni presenti nel browser e/o delle scelte specifiche dell'utente. Questi metadati avrebbero potuto essere memorizzati molto semplicemente nella tabella relativa ai nodi-domande (e ovviamente alle modalità di risposta), tuttavia questo approccio comportava una certa rigidità e delle potenziali ridondanze/inconsistenze. È, infatti, evidente che la struttura di un questionario è invariante rispetto alla specifica localizzazione: ad esempio la condizione sulla base della quale la compilazione doveva procedere da una certa domanda X a un'altra domanda Y quando era vera una condizione C non dipende dalla lingua nella quale era visualizzato il questionario. Memorizzare in un'unica tabella informazioni strutturali e testuali avrebbe quindi costretto ad introdurre campi specifici per ciascuna lingua oppure a replicare le informazioni strutturali per ciascuna versione linguistica di una domanda.

I testi associati alle domande sono stati perciò memorizzati in una tabella separata rispetto a quella che rappresentava le domande in senso stretto (e i relativi legami di flusso). In questo modo è stato possibile separare e semplificare il caricamento delle varie versioni localizzate dei quesiti, consentendo correzioni e aggiornamenti anche parziali in modo agevole e poco invasivo, senza richiedere la ricompilazione dei programmi. Stesso discorso vale anche per la componente testuale delle modalità di risposta: ad esempio il fatto che una condizione facesse riferimento ad uno o più indici di modalità non dipendeva dalla lingua nella quale la modalità stessa veniva visualizzata all'utente. Anche in questo caso i testi associati sono stati quindi memorizzati in una tabella separata, semplificando i caricamenti, le correzioni e più in generale la gestione delle problematiche di localizzazione.

In merito ai metadati del questionario elettronico si evidenzia che lo spostamento di parte della logica applicativa (in particolare quella che regola i percorsi di compilazione) nella base dati ha avuto indubbi vantaggi, ma avrebbe potuto causare delle problematiche non banali in termini prestazionali. Per evitare il caricamento *ex novo* dei metadati del questionario a ogni richiesta di pagina web (che avrebbe determinato un significativo e pericoloso sovraccarico a danno del DBMS), è stato quindi realizzato un meccanismo di *caching* a livello applicativo, che ha consentito di evitare il continuo caricamento dei metadati e viceversa il loro costante riutilizzo, richiedendo l'interazione con la base dati solo quando strettamente necessario.

Venendo alla parte della base dati dedicata alla memorizzazione dei microdati, grande attenzione è stata posta ai requisiti di efficienza. Trattandosi, infatti, di dati con un'elevatissima transazionalità (quindi continui e frequenti accessi concorrenti in scrittura e/o lettura), era fondamentale

ottimizzarne la struttura per rendere efficiente l'accesso. I requisiti progettuali che hanno guidato la definizione delle varie tabelle di microdati sono stati quindi i seguenti:

- ▶ minimizzare il numero di record da inserire/aggiornare in corrispondenza di un generico salvataggio di questionario. In generale se un salvataggio richiede l'accesso in scrittura a molti record, aumenta il numero di *lock* sui dati da ottenere, il numero di blocchi su disco da (ri)scrivere, e quindi in definitiva la probabilità di innescare conflitti sulle risorse e avere tempi di elaborazione più lunghi (che a loro volta aumentano ulteriormente la probabilità di conflitto sulle risorse). All'atto pratico questo si è tradotto con la scelta di memorizzare i microdati usando, per quanto possibile, un solo record per ogni questionario compilato. Fanno eccezione alcuni dati particolari come quelli relativi alle unità locali. Poiché la numerosità di tali unità non era nota a priori e poteva essere significativamente alta, ogni sottosezione relativa ad una singola unità locale è stata considerata come un questionario a sé stante e quindi memorizzata con un record separato in una tabella di microdati apposita;
- ▶ rendere il singolo record di microdati il più possibile "leggero". Il fatto di avere un unico record per memorizzare tutti i microdati di un questionario comporta inevitabilmente un numero elevato di colonne, che sarebbe potuto esplodere in corrispondenza di quesiti *multiresponse*. Per la memorizzazione di tali quesiti si è quindi utilizzata una memorizzazione parzialmente "compressa", in cui la combinazione di caselle biffate è stata rappresentata da una stringa testuale con struttura analoga ad un *bitmap*. In pratica, ad esempio, la risposta (multipla) relativa a un quesito con *n* possibili modalità di risposta (non mutuamente esclusive) risulta memorizzata attraverso una stringa testuale di lunghezza *n*, costituita da caratteri 0 e 1, a seconda che l'utente abbia o meno biffato la modalità corrispondente. Si sarebbe potuto spingere ulteriormente la compressione utilizzando una codifica numerica (corrispondente a una vera e propria sequenza di bit). Tuttavia la soluzione adottata è apparsa preferibile, in quanto rappresenta una mediazione ragionevole tra facilità di analisi da parte degli utenti statistici (che con una semplice vista hanno potuto analizzare i dati nella forma, a loro più congeniale, con le singole colonne esplose) e spazio occupato su disco da ogni questionario (pochi KB);
- ▶ automatizzare il *mapping* tra metadati, colonne della tabella di microdati, e componenti Java responsabili della persistenza. Per garantire tale corrispondenza, in fase di progetto sono state specificate delle regole di nomenclatura che hanno garantito una corrispondenza biunivoca tra il codice identificativo di ogni nodo-domanda (presente nella tabella dei metadati di cui sopra), il nome della colonna corrispondente nella tabella di microdati e il nome del componente Java responsabile del suo trattamento nell'applicazione. Le regole di nomenclatura comprendevano anche casi speciali in cui a un singolo quesito corrispondono più colonne nella tabella di microdati (es. quesiti che prevedono matrici di inserimento di valori numerici). Per quanto riguarda i quesiti a risposta chiusa (es. "domanda con scelta singola da una lista di risposte") i codici numerici da inserire nella tabella dei microdati sono stati esattamente gli stessi memorizzati nella tabella delle modalità, garantendo una facile decodifica delle risposte fornite dal compilatore.

La Figura 1.9 mostra un esempio di base dati con una versione semplificata di alcuni dei metadati e microdati relativi a un frammento di questionario. Come già detto, la parte relativa ai percorsi di compilazione sarà descritta in maggiore dettaglio nel Paragrafo 2.1.

Figura 1.9 – Alcuni dei metadati e microdati corrispondenti ad un frammento di questionario

Frammento di questionario del rispondente con identificativo ID_012345			Tabella delle domande		Tabella dei testi delle domande				
2.06 Nel corso del 2011 l'impresa ha effettuato attività di formazione aziendale? 1 <input checked="" type="checkbox"/> Sì 2 <input type="checkbox"/> No • vai al quesito 2.09			ID_domanda		ID_dom	lang	testo		
2.07 Quali attività di formazione aziendale l'impresa ha effettuato nel corso del 2011? (sono possibili più risposte) A <input checked="" type="checkbox"/> Corsi di formazione a gestione interna B <input type="checkbox"/> Corsi di formazione a gestione esterna C <input checked="" type="checkbox"/> Training on the job (apprendimento sul campo) D <input type="checkbox"/> Altro			....	....	....				
			Q206	....	Q206	it	Nel corso del...		
			Q207	....	Q207	it	Quali attività...		
			....		....				
			....		....				
Tabella degli archi-transizioni			Tabella delle modalità			Tabella dei testi delle modalità			
ID_srg	ID_dest	Condizione	ID_dom	Ind_mod		ID_dom	Ind	lang	testo
....			....	....		....	....	....	....
Q206	Q209	Q206 = 2	Q206	1		Q206	1	it	Sì
Q206	Q207	else	Q206	2		Q206	2	it	No
Q207	....	....	Q207	1		Q207	1	it	Corsi...
....			Q207	2		Q207	2	it	Corsi...
			Q207	3		Q207	3	it	Training...
			Q207	4		Q207	4	it	Altro
			....	....		....	....	....	....
Tabella dei microdati									
ID_rispondente			Q206	Q207					
....			....	....					
ID_012345			1	1010					
....			....	....					

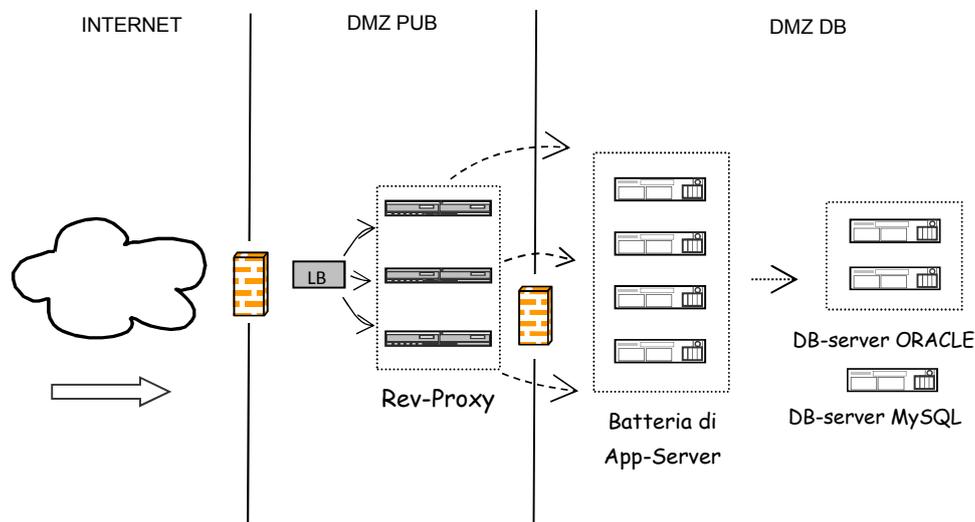
#### 1.4 L'architettura logico-fisica dei sistemi

L'architettura utilizzata si può descrivere con uno schema a tre livelli funzionali: *reverse-proxy*, *application-server*, *DB-server*. Lo schema logico-fisico è quello mostrato nella Figura 1.10.

Le richieste provenienti da Internet sono state gestite da tre server (*reverse-proxy*) che inoltravano le richieste verso i server applicativi della rete interna (*application-server*). Il traffico web dei tre server è stato regolato da un bilanciatore di carico (indicato con la sigla "LB" in Figura 1.10) che garantiva la distribuzione del carico e la continuità di servizio.

I *reverse-proxy* inoltravano le richieste verso quattro *application-server* (Figura 1.10). Anche questo traffico veniva gestito dai *reverse-proxy* mediante un bilanciatore di carico di tipo software (*mod\_proxy\_balancer* di Apache) che distribuiva le richieste verso i quattro server della batteria.

Figura 1.10 – Schema dell'architettura utilizzata in produzione



Ogni modifica effettuata ai file di configurazione veniva eseguita su un solo server (server master) ed estesa agli altri due *reverse-proxy* tramite una procedura di allineamento che eseguiva anche il *restart* del servizio *httpd* sui 3 server.

Come *application-server* sono stati utilizzati quattro server con le caratteristiche hardware mostrate nel Prospetto 1.1. I quattro server condividevano le caratteristiche software (Prospetto 1.2) e i contenuti. Ogni dieci minuti veniva lanciata una procedura che eseguiva l'allineamento dal server master verso gli altri tre server. L'allineamento riguardava le configurazioni (*httpd* e *Tomcat*) e le aree dati di tutte le applicazioni. L'aggiornamento sulle applicazioni che utilizzano Java (*deploy*) veniva invece eseguito direttamente sui server del *pool*.

Il sito <http://censimentoindustriaeservizi.istat.it/rete>, che è stato sviluppato utilizzando TYPO3 come gestore dei contenuti, si è appoggiato ad un database MySQL su server dedicato.

Per la gestione del solo sito <http://censimentoindustriaeservizi.istat.it/istatcens/> (PHP+WordPress) si è fatto uso di un database MySQL che risiedeva all'interno degli application-server. Il DB veniva aggiornato sul server master e mantenuto allineato su tutti i server della batteria.

Per la gestione degli utenti e dei contenuti web dei siti di acquisizione e gestione (QCIS e SGRCIS) sono stati utilizzati due *DB-server* in *cluster*: *alabama1* e *alabama3* (Prospetto 1.1). Sui due server era installata la versione 11.2 di Oracle (Prospetto 1.2).

### Prospetto 1.1 – Caratteristiche tecniche dei server

SERVER	Prodotto	Ram	Processori
Rev-Proxy	HP ProLiant DL585	128GB	16 AMD Opteron 6128
App-Server	HP ProLiant DL585	128GB	32 AMD Opteron 6128
DB-Server	SUN FIRE X4450	128GB	8 Intel Xeon (R) X7350

## Prospetto 1.2 – Caratteristiche software dei server

SERVER	Sistema Operativo	Apache	PHP	MySQL	Tomcat	Java	Oracle
Rev-Proxy	Linux RedHat 6.3	2.2.15	-	-	-	-	
App-Server	Linux RedHat 6.3	2.2.15	5.3.3	5.1.61	6.0.24	1.6.0_24	
DB-Server	Linux RedHat 5.10			5.0.95			11.1.0.6

### 1.5 I test delle applicazioni e dei sistemi

Lo scopo primario delle attività di test delle applicazioni e dei sistemi è stata quella di dotare l'Istituto di un supporto specializzato nella revisione qualitativa e prestazionale dell'architettura hardware e software necessaria per la gestione informatica del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit, in particolare prima che esso venisse messo in esercizio e nel monitoraggio durante le fasi di particolare criticità del servizio.

Le attività di test funzionali e automatici si sono poste come obiettivo il miglioramento della qualità del software sviluppato, tramite un processo continuo di verifica e segnalazione di problemi individuati nelle fasi di sviluppo e *prerelease* del software tramite un processo di comunicazione diretto con il team di sviluppo.

La *code review* si è occupata della ricerca di possibili criticità e vulnerabilità di sicurezza presenti nel codice e nella verifica del rispetto delle “*best practice*” internazionali (OWASP, NIST, eccetera) in materia di sviluppo.

Il servizio di test di carico per il dimensionamento dell'infrastruttura si è posto come obiettivo l'individuazione del carico massimo dell'infrastruttura prima che le prestazioni (es. i tempi di risposta) degradino oppure che i parametri di performance diventino inaccettabili.

Le attività di test per il dimensionamento dell'infrastruttura hanno avuto una forte interazione con il gruppo dei sistemisti per verificare l'adeguatezza delle infrastrutture predisposte all'interno dei carichi stimati di utenti e di tipologia di funzioni richieste durante l'andamento della rilevazione.

La gestione dei test si è rivelata non banale ed essendo in relazione con attività che andavano dall'analisi delle specifiche alla conoscenza dell'architettura hardware e software, dall'analisi del codice all'interazione con i database, dalle problematiche di usabilità a quelle di sicurezza e ha richiesto figure specializzate e con forte esperienza oltre ad una forte interrelazione e coordinamento del processo.

L'esperienza svolta sia in questa che nelle precedenti rilevazioni censuarie, anche con il supporto di società esterne di consulenza, ha permesso di acquisire all'interno dell'Istituto le competenze per la formazione di personale specializzato nelle differenti attività di test.

#### 1.5.1 I test funzionali

I test funzionali hanno come principale obiettivo la verifica della correttezza funzionale di una applicazione rispetto ai requisiti indicati dal committente.

Considerando il processo di sviluppo iterativo del software risulta evidente come ogni rilascio, intermedio o finale, è un naturale punto di verifica per i test. Questi test a ogni iterazione forniscono in aggiunta utili indicazioni inerenti la qualità del software prodotto come ad esempio: affidabilità, usabilità, efficienza, sicurezza e prestazioni del sistema.

Nell'ambito del progetto sono state utilizzate le seguenti tipologie di test:

- ▶ test modulari, cioè orientati su una singola funzionalità dell'applicazione;
- ▶ test di integrazione, focalizzati sulle modalità con cui i vari moduli interagiscono fra loro;
- ▶ test di sistema, per verificare il sistema nella sua interezza, in quanto alcuni malfunzionamenti non si evidenziano prima che il sistema sia completo;
- ▶ test di regressione, ripetizione dei test per verificare la correzione di errori segnalati o l'introduzione di nuovi errori al rilascio di una nuova versione.

Inoltre si possono considerare:

- ▶ i *Black Box Test* quando in esame sono le funzionalità e i comportamenti dell'applicazione visibili all'utente finale e definite dallo scenario principale;
- ▶ i *White Box Test* relativi al comportamento dell'applicazione a più basso livello con una analisi più profonda (tipi di dati, salvataggio dei dati, *trigger* e *batch*, strutture di controllo, flusso degli stati, dipendenze, eccetera).

La metodologia di pianificazione dei test utilizzata per il Censimento dell'industria e dei servizi si riassume brevemente nei seguenti punti (approccio dell'oracolo):

- a. analisi dei documenti delle specifiche funzionali, dei requisiti e dei documenti di progetto con individuazione dei casi rilevanti e delle possibili criticità in modo da testare ogni funzionalità con una selezione dei valori di input più significativi (valori di frontiera, casi speciali, eccetera);
- b. definizione per ogni caso d'uso di un piano di test (*test plan*) composto da uno o più *test case*, definiti tramite:
  - identificazione delle pre-condizioni che devono essere soddisfatte prima dell'esecuzione del test;
  - l'insieme di azioni (*action*) da intraprendere per effettuare il test;
  - il comportamento atteso dall'applicazione in seguito a tali azioni (*expected outcome*) da validare, dopo l'esecuzione del test, con il comportamento effettivo dell'applicazione (*actual outcome*);
- c. avvio della fase iterativa di *testing* e tracciamento dei test effettuati (*test run*) al fine di tracciare l'identificativo del *test case* in esame, e il risultato (positivo o negativo) corredato da alcune informazioni aggiuntive (data, versione del prodotto). Nelle iterazioni successive alla prima si cerca di automatizzare il più possibile la procedura di test;
- d. rilascio della documentazione relativa ai risultati delle campagne di test, con verifica nel caso web della compatibilità con le differenti tipologie e versioni dei browser.

Applicando tale metodologia, nell'arco di sette mesi, sono stati predisposti 27 *test plan*, articolati in 55 *test case*, che hanno permesso di individuare ben 547 *bug*.

Bugzilla, un *bug-tracker open source*, ha supportato i *tester* nell'attività di definizione dei *test plan* e dei *test case*, il tracciamento dei *test run* e la redazione assistita delle segnalazioni in caso di anomalie. In particolare si sono rilevate di grande utilità le funzioni di:

- gestione condivisa online del *workflow* del *bug* delle comunicazioni fra *tester*, analisti e sviluppatori;
- monitoraggio rapido ed efficace dell'andamento delle correzioni delle anomalie: in un'unica schermata è possibile individuare i *test case* (segnalati in rosso) a cui sono associati dei *bug* non ancora risolti.

Riguardo il problema del *cross-browser*, cioè la possibilità di avere difformità di comportamento tra i vari browser, fra gli strumenti utilizzati si segnalano: *IETester*, un programma che simula il comportamento di varie versioni del browser *Internet Explorer*, e soprattutto l'utilizzo di ambienti virtualizzati su cui installare determinate versioni del browser.

È bene infine sottolineare il prezioso contributo apportato dagli *stakeholder* (sedi territoriali Istat, rete di rilevazione, eccetera) per effettuare il *beta testing*<sup>6</sup> nell'ottica di fornire utili suggerimenti per il miglioramento della qualità dell'applicazione in termini di usabilità.

### 1.5.2 I test automatici

Una delle attività che ha contribuito in modo sostanziale a rendere la “*web application* SGRCIS” robusta e priva di errori è stata l'esecuzione dei “test” durante la fase di sviluppo.

A sostegno di tale attività è stata realizzata l'applicazione “*Automatic Test CIS*” che ha provveduto ad automatizzare la maggior parte dei test facilitando e velocizzando, in tal modo, l'esecuzione, soprattutto dei “*regression test*”.

L'applicazione *Automatic Test CIS* è stata in grado di eseguire in modo automatico le funzionalità web di SGRCIS e di segnalare eventuali errori.

Il processo alla base dell'applicazione *Automatic Test CIS* è composto da:

- 1) pianificazione: vengono selezionate le funzionalità della *web application* SGRCIS da testare e vengono progettati i test da eseguire su ognuna di esse;
- 2) esecuzione test: vengono implementati ed eseguiti, mediante *Automatic Test CIS*, i test pianificati nella fase precedente.

L'applicazione *Automatic Test CIS* è stata sviluppata in Java ed ha utilizzato le tecnologie fornite dai *framework Selenium* e *TestNG*.

I componenti del *framework Selenium* utilizzati per lo sviluppo dell'applicazione *Automatic Test CIS* sono i seguenti:

- *Selenium Remote Control*: fornisce i meccanismi in grado di aprire, in automatico, un *browser web* (*Explorer*, *Firefox*, *Opera*) all'URL relativo alla *web application* SGRCIS ed eseguire le sue funzioni;
- *Selenium Test*: è la codifica in linguaggio Java dei test pianificati.

La tecnologia *TestNG* fornisce il meccanismo attraverso il quale è possibile eseguire lo stesso *Selenium Test* uno svariato numero di volte con parametri differenti simultaneamente.

L'architettura è così composta: un insieme *Selenium Test* che vengono dati input ad una griglia di *Selenium Remote Control* che si occupano della loro esecuzione.

Le Figure 1.11 e 1.12 mostrano la struttura generale della griglia e il funzionamento in particolare di ogni *Selenium Remote Control*.

<sup>6</sup> Una terminologia molto diffusa divide i controlli finali in *alfa-test* e *beta-test*. I due termini distinguono i controlli rispetto al soggetto che li esegue. Il primo stadio di *testing* (*alfa*) sarà svolto dal gruppo di *testing* con il supporto dei coordinatori dei progetti di sviluppo. Seguirà il *beta testing* in cui verranno coinvolti attivamente i rappresentanti degli *stakeholder*.

Figura 1.11 – Architettura *Selenium Grid* per l'esecuzione di multipli test automatici in parallelo

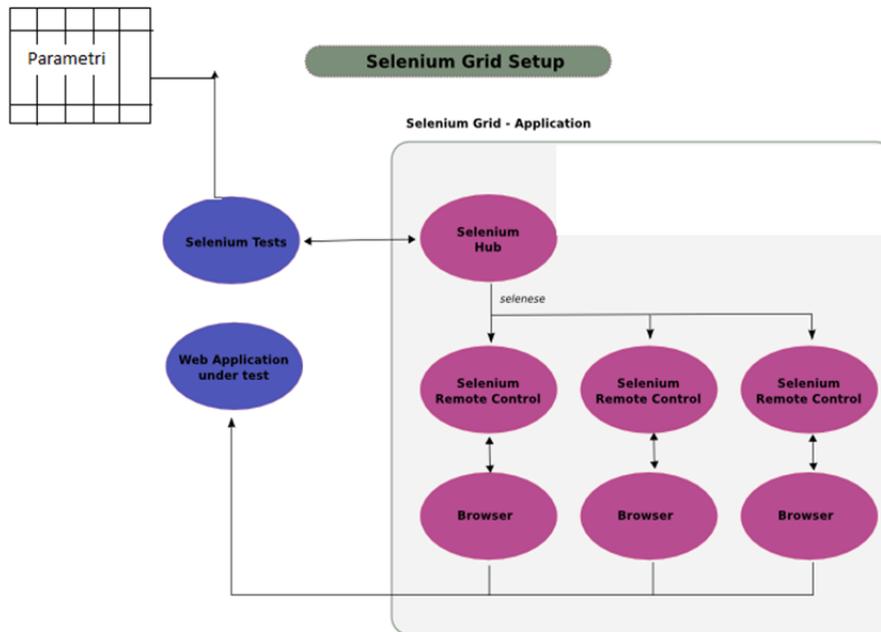
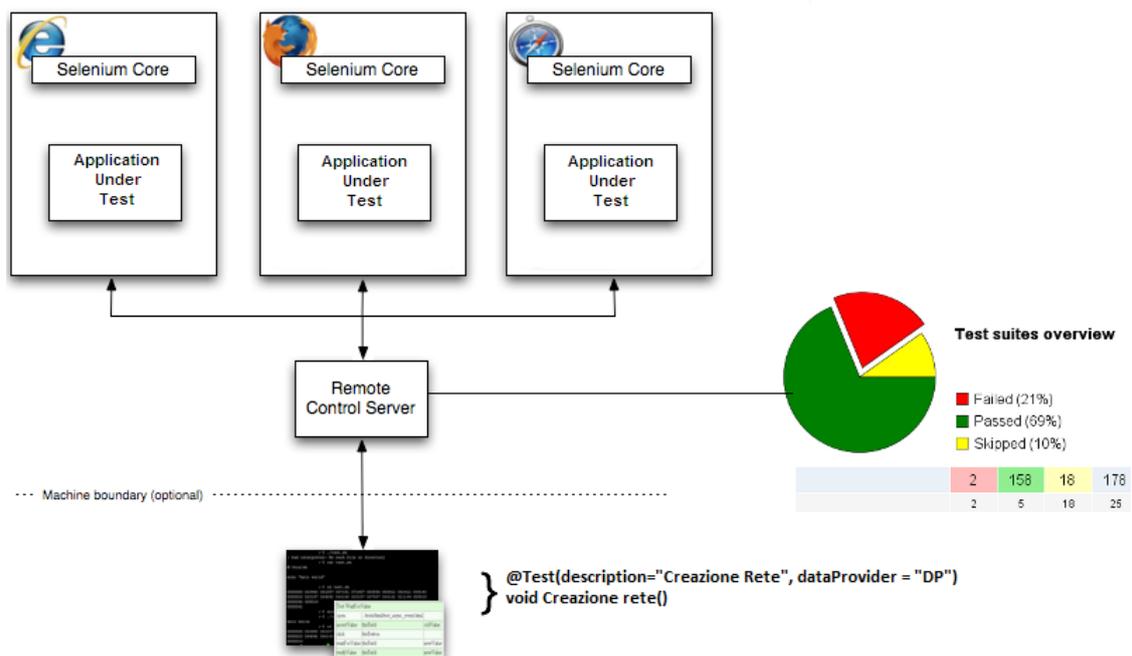


Figura 1.12 – Architettura *Selenium RC* per l'esecuzione di un singolo test automatico



### 1.5.3 I test di sicurezza

Le attività di test di sicurezza hanno previsto una forte interazione con i differenti gruppi per lo sviluppo del codice e si sono giovati della possibilità di avviare la collaborazione fin dalle prime fasi dello sviluppo del software, in particolare per la fase di *code review*, per permettere di ridurre al minimo l'impatto di eventuali richieste di correzione codice. Il processo di certificazione si è articolato nelle seguenti fasi:

- *Code Review*  
attività di analisi del codice (statica e dinamica) in ambiente di test con modalità *white box*, cioè con una conoscenza completa dei dettagli implementativi del sistema (simulazione dell'attività di un attaccante interno):
  1. *Analisi statica*  
analisi del codice sorgente senza necessità di compilare o eseguire il codice sorgente, volta a identificare possibili migliorie nel codice avvalendosi delle best practice note in letteratura.
  2. *Analisi dinamica*  
analisi a *runtime* del codice installato e configurato volta a identificare problemi relativi all'esecuzione del programma.
- *Vulnerability Assessment (VA)*  
analisi tramite strumenti, automatici e non, della presenza di vulnerabilità note in letteratura all'interno del sistema da certificare.
- *Penetration Test (PT)*  
verifica a *runtime*, su un apposito sistema di test, delle criticità individuate utilizzabili per un attacco al sistema da certificare, in modalità *black box*: cioè nella modalità di analisi del sistema da certificare senza alcuna conoscenza pregressa di materiale che non sia pubblicamente disponibile (simulazione dell'attività di un utente finale).

L'esperienza svolta durante l'attività di certificazione è stata utile per acquisire conoscenze ed *expertise* sulle tecniche di programmazione all'interno dell'Istituto che si sono rivelate di grande utilità nelle successive rilevazioni.

### 1.5.4 I test di carico

I test di carico per le applicazioni CIS sono state effettuate tramite un prodotto *open source* che potesse essere considerato affidabile per l'esecuzione dei test e che, allo stesso tempo, potesse accrescere le competenze presenti in Istituto.

Il prodotto scelto è stato *JMeter*, un'applicazione *desktop* sviluppata dalla *Apache Software Foundation* che permette di effettuare test di carico delle proprie applicazioni. Il prodotto presenta diversi vantaggi:

- è possibile scaricarlo gratuitamente da internet (*open source*);
- ha una duplice modalità di gestione: attraverso un'interfaccia grafica o da linea di comando;
- risulta relativamente semplice da installare e configurare;
- utilizza un *proxy* interno per registrare la navigazione attraverso il sito web da sottoporre a controllo;
- può testare diversi protocolli, non solo HTTP;
- si può creare una batteria di client *JMeter* per lanciare test da client multipli.

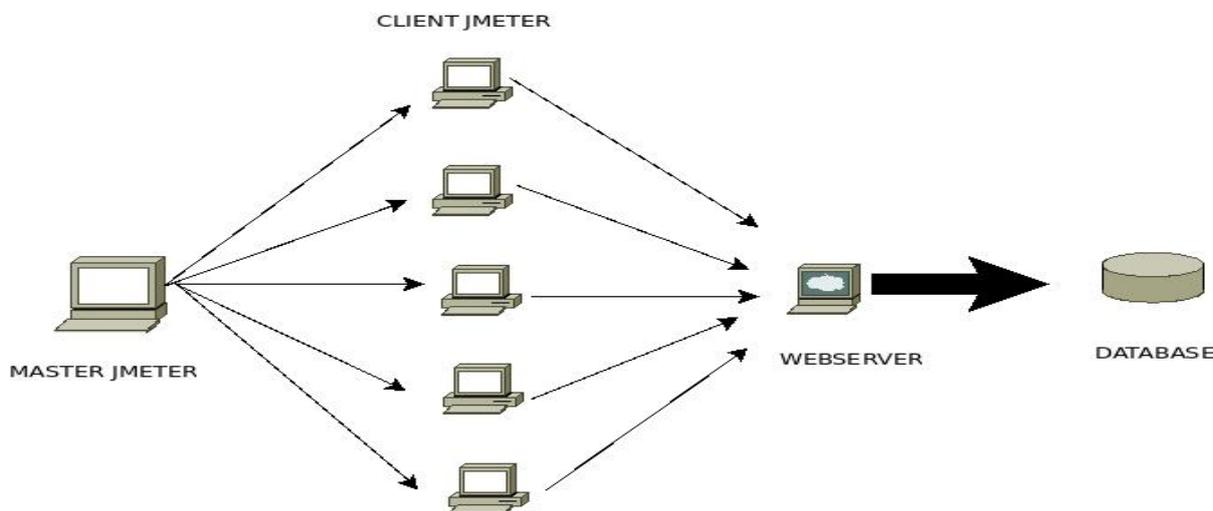
I risultati prodotti da *JMeter* forniscono valori statistici come la media, la mediana, la varianza e il *throughput*.

I test che si possono eseguire con *JMeter* sono essenzialmente di due tipi ed entrambi sono stati utilizzati per validare ed ottimizzare l'applicazione CIS prima della sua messa in esercizio:

- test di prestazioni: con tale tipo di test si valutano le risposte del codice in condizioni di traffico dati e numero di accessi normali;
- test di stress: con questo tipo di test si valutano i tempi di risposta dell'applicazione in condizioni di carico elevate.

L'architettura di test è di tipo *master-client*. Come client sono stati utilizzati venti PC presenti in Istituto nell'apposita "sala corsi", normalmente destinati, quindi, a scopo didattico. Su ciascun PC, dotato di sistema operativo Windows, è stato installato il prodotto *JMeter*. Dell'intera batteria, un PC ha svolto le funzioni di master e gli altri quelle di client. Il master inviava ai client le richieste di esecuzione del test; da questi, a loro volta, partivano le chiamate al sito del CIS. In Figura 1.13 viene mostrato un esempio dell'architettura. Il numero di client utilizzati nel test è variabile e viene controllato dal master. Si è scelto di simulare circa venti utenti per ciascun client quindi, sfruttando la batteria al massimo, si sono potuti simulare 400 utenti concorrenti in accesso al sito CIS.

Figura 1.13 – Preparazione ed esecuzione dei test



Al fine di simulare il carico di utenti concorrenti al sito, *JMeter* è stato configurato nel modo seguente. È stato definito un percorso da seguire all'interno del sito uguale per tutti gli utenti simulati (l'esecuzione completa del percorso all'interno del test è chiamata convenzionalmente "ciclo"). Gli utenti non iniziavano il ciclo contemporaneamente ma seguivano una rampa di accesso definita. Quando il primo utente aveva terminato il primo ciclo iniziava immediatamente un ciclo successivo, trovandosi così in concorrenza con il secondo ed eventualmente il terzo utente che, nel frattempo, aveva iniziato il rispettivo ciclo di test. Il numero degli utenti che accedevano al sito cresceva fino a che non veniva raggiunto il valore desiderato (es. venti utenti concorrenti). A quel punto, tutto il gruppo di utenti continuava a ripetere il ciclo per il periodo di tempo definito in fase di configurazione del test (es. cinque minuti). Quindi, in questo lasso di tempo, veniva simulato un gruppo di N utenti che ripeteva continuamente il proprio percorso di navigazione (tale percorso era lo stesso per tutti). Passato questo periodo temporale gli utenti simulati abbandonavano via via lo scenario del test, fino a che il numero di utenti che accedevano al sito non tornava ad essere pari a zero ed il test veniva considerato concluso.

I tempi di risposta venivano quindi riportati in tabella e visualizzati in un grafico.

Sono stati effettuati diversi test utilizzando l'architettura predisposta ed i risultati hanno fornito importanti informazioni sui tempi di risposta in funzione del carico. Questi risultati sono stati utilizzati per ottimizzare l'applicazione prima della sua messa in esercizio, in particolare, sono stati creati due processi *Tomcat* per tutte le applicazioni ospitate sui *web server*.

Successivamente alla fase di test, il sito è stato messo in esercizio e sottoposto a monitoraggio continuo per valutarne le prestazioni. La fase di monitoraggio permette, infatti, di rilevare quelle criticità sfuggite alla fase di test. Si ricorda, infatti, che nessun test può dirsi completamente esaustivo, quindi, in fase di produzione è opportuno effettuare controlli continui sulle performance del sistema. Come strumento di monitoraggio si è scelto di utilizzare *Splunk*, un prodotto in grado di indicizzare ed analizzare qualunque tipo di dato macchina. In particolare, nel caso del CIS, i dati di interesse sono stati quelli relativi ai *log* di accesso al sito (*log* di *Apache* e *Tomcat*).

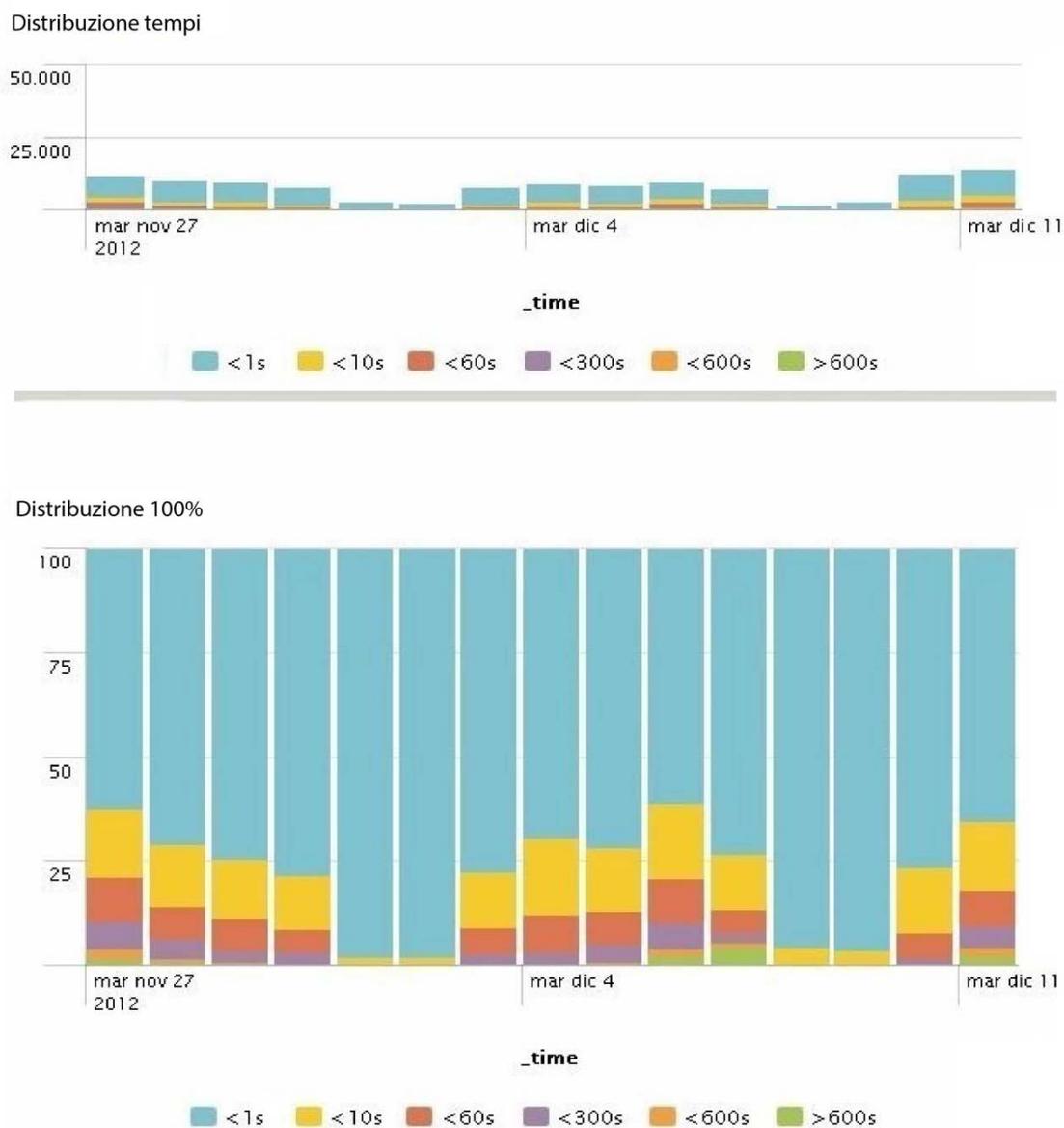
È stato quindi installato e configurato opportunamente un ambiente *Splunk*. Tale ambiente è composto da un *server master* che colleziona i *log* e da *client forwarder* che, a loro volta, inviano al master i *log* di accesso al sito. I *forwarder* sono stati installati e configurati sia sui *web server* della batteria che sui *reverse proxy*.

I dati raccolti hanno permesso di rilevare che, dopo un periodo di funzionamento normale, alcune pagine dell'applicativo mostravano tempi di risposta dell'ordine dei venti secondi e più, rendendo di fatto difficile il completamento del questionario e causando le segnalazioni di alcuni utenti.

Le anomalie venivano rilevate in modo esteso e randomico, rendendo di fatto critica la situazione, visto che circa il 30 per cento degli utenti non otteneva risposte in tempi accettabili. Nella figura seguente mostriamo i tempi di risposta delle pagine del questionario. È da osservare che a tempi di risposta elevati non corrispondeva un numero di accessi elevato, essendo l'evento randomico.

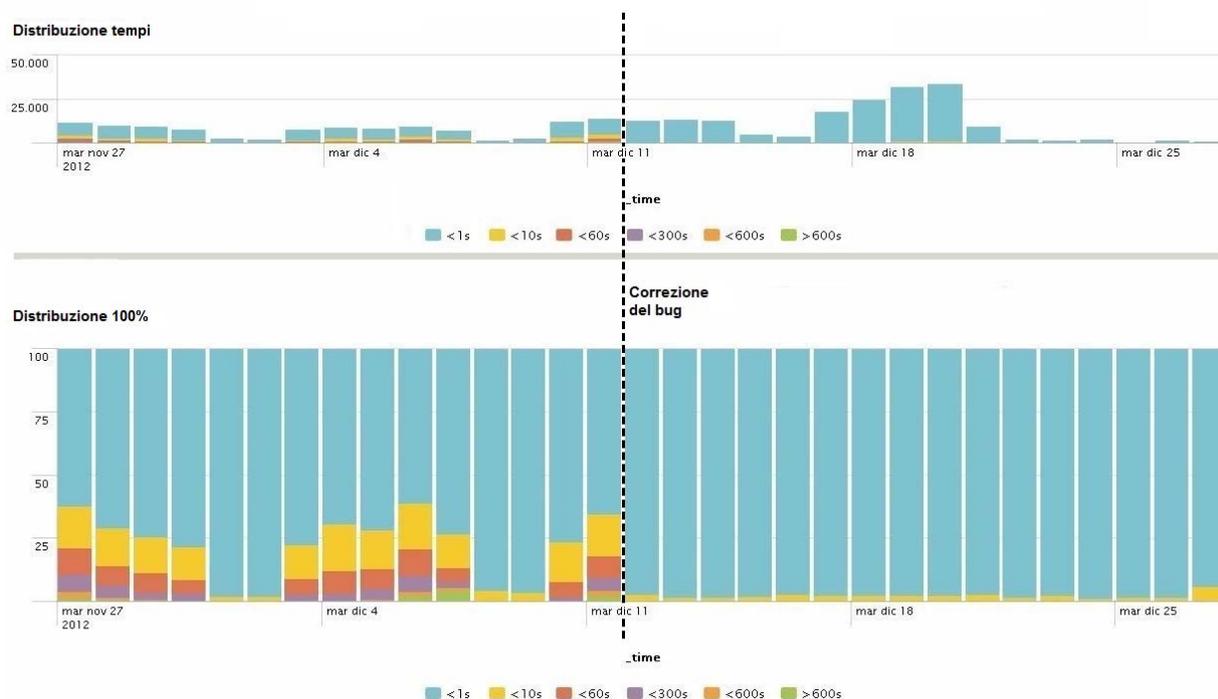
Venivano quindi attivati *log* sempre più "parlanti" che hanno permesso, tramite un'analisi accurata di individuare e rimuovere il "bug". Il problema era particolarmente insidioso poiché legato alla generazione sicura di numeri random e alla entropia necessaria per generarli. L'insidia era dovuta a una risoluzione automatica ma purtroppo temporanea del problema a ogni accesso diretto al server oggetto del disservizio, causata dall'aumento di entropia del sistema che ha reso inefficace qualunque azione di analisi diretta.

Figura 1.14 – Distribuzione dei tempi di risposta del questionario



Il monitoraggio successivo ha confermato l'efficacia della soluzione proposta. Nella Figura 1.15 si riportano i tempi di risposta del questionario prima e dopo la correzione del *bug*. Come si può osservare, dopo la rimozione del *bug*, i tempi di risposta sono stati nettamente inferiori.

**Figura 1.15 – Distribuzione dei tempi di risposta del questionario prima e dopo la rimozione del bug**



## 1.6 I sistemi di autenticazione e autorizzazione utenti degli Uffici di censimento

Per il meccanismo di autenticazione è stato predisposto un sistema di *single-sign-on* che permetteva ad un operatore degli Uffici di Censimento costituiti presso il sistema camerale, autenticato sul sistema SGRCIS e con l'autorizzazione ad operare il *data entry* mediante il sistema di acquisizione dei questionari QCIS, senza necessità di autenticarsi nuovamente per lavorare sui questionari a lui assegnati. Lo stesso meccanismo, basato su tabelle DB, permetteva agli utenti di SGRCIS di autenticarsi sul Portale della rete senza bisogno di un'ulteriore immissione di username e password.

### 1.6.1 L'autenticazione

Per poter operare all'interno del Sistema di Gestione della Rilevazione SGRCIS era necessario essere registrati all'interno dell'applicazione. L'autenticazione veniva effettuata inserendo le credenziali username e password nella pagina di login (Figura 1.16). La username di un operatore era composta da sette numeri.

Figura 1.16 – Pagina di login di SGRCIS



CENSIMENTO DELL'INDUSTRIA E DEI SERVIZI 2011

SGR | SISTEMA GESTIONE RILEVAZIONE

**Autenticazione Operatore**

Username:

Password:

Login

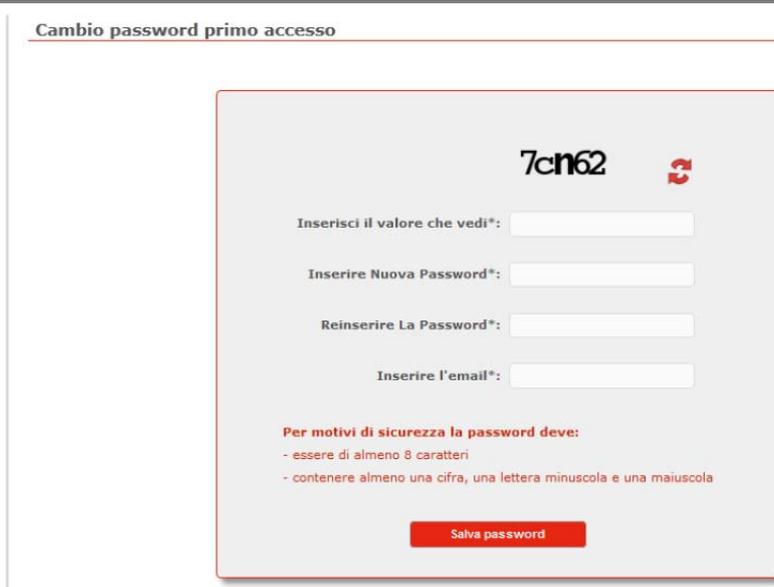
Hai dimenticato la password?  
Clicca qui per recuperarla. 

Il processo di autenticazione prevedeva, inoltre, che al primo accesso al sistema fosse necessario cambiare la propria password e che si dovesse specificare un indirizzo di posta elettronica, possibilmente diverso da quello impostato in fase di creazione o caricamento dell'utenza (Figura 1.17).

Il nuovo indirizzo di posta elettronica veniva poi utilizzato dal sistema per ogni eventuale comunicazione. Inoltre, per motivi di sicurezza, l'utente doveva riportare il contenuto di un'immagine generata casualmente dal sistema nell'apposito campo (denominato "captcha").

La nuova password doveva essere di almeno otto caratteri e contenere almeno una cifra, una lettera minuscola, una lettera maiuscola. Il cambio password o il recupero in caso di smarrimento o dimenticanza poteva essere effettuato dalla pagina di login cliccando sul link "Hai dimenticato la password? Clicca qui per recuperarla".

Figura 1.17 – Pagina di cambio password al primo accesso di SGRCIS



**Cambio password primo accesso**

7cn62 

Inserisci il valore che vedi\*:

Inserire Nuova Password\*:

Reinserire La Password\*:

Inserire l'email\*:

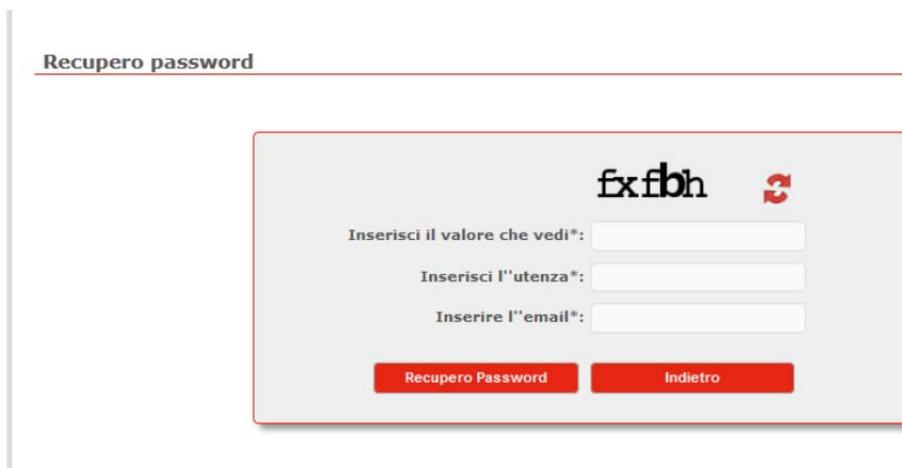
**Per motivi di sicurezza la password deve:**

- essere di almeno 8 caratteri
- contenere almeno una cifra, una lettera minuscola e una maiuscola

Salva password

Se si effettuava questa operazione il sistema chiedeva di inserire il *captcha*, lo username dell'operatore e l'e-mail (Figura 1.18).

Figura 1.18 – Pagina di recupero password di SGRCIS



In questo caso il sistema effettuava il controllo sulla corrispondenza tra utenza ed e-mail all'interno del suo database. In caso affermativo inviava all'operatore una nuova password all'indirizzo email indicato. La nuova password doveva essere modificata al successivo accesso secondo le indicazioni già fornite.

### 1.6.2 L'autorizzazione

Una volta inserite le credenziali di accesso l'operatore accedeva al sistema SGRCIS limitatamente alle funzioni a cui era stato abilitato in base al proprio profilo applicativo.

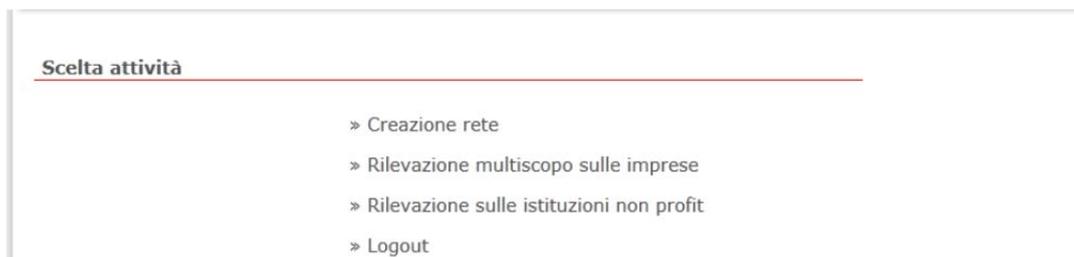
Ad ogni utente del sistema era associato un ruolo e un profilo applicativo, in base al quale il Sistema di Gestione della Rilevazione veniva dinamicamente personalizzato rispetto ai seguenti aspetti:

- ▶ rilevazioni sulle quali si ha la possibilità di operare;
- ▶ insieme di funzionalità accessibili all'interno di ciascuna rilevazione;
- ▶ visibilità territoriale dei dati.

Il sistema permetteva di gestire il Censimento sulle istituzioni non profit e la Rilevazione sulle imprese in modo integrato, ovvero l'utente poteva utilizzare le stesse credenziali di accesso senza doversi autenticare ad ogni cambio di rilevazione, mediante un meccanismo chiamato "fork".

Gli utenti abilitati a lavorare su entrambe le rilevazioni potevano, in qualsiasi momento, cambiare rilevazione semplicemente scegliendo "Torna a scelta attività", presente nella home page delle due rilevazioni.

Dopo l'autenticazione, se l'utente era abilitato a lavorare su entrambe le rilevazioni, il sistema presentava la pagina riportata in Figura 1.19. Gli utenti che potevano operare solo su un'indagine visualizzavano direttamente la home page della rilevazione a loro associata.

**Figura 1.19 – Pagina di selezione dell'indagine (fork) in SGRCIS**

I ruoli gestiti dal sistema sono descritti in dettaglio nel paragrafo 3.2.

### 1.7 La formazione on line (e-learning)

Il piano di formazione del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit è stato articolato in modo da garantire l'acquisizione di abilità e competenze da parte di tutti i soggetti coinvolti nell'indagine ai diversi livelli territoriali – nazionale, regionale e provinciale – con strategie e strumenti innovativi.

Gli obiettivi sono stati:

- 1) fornire le conoscenze del contesto normativo e istituzionale, dell'organizzazione della rete di rilevazione e del processo di indagine, dei questionari di rilevazione;
- 2) fornire le capacità di utilizzare gli strumenti informatici previsti (in particolare SGRCIS);
- 3) sviluppare la consapevolezza della propria funzione nell'ambito del processo censuario e di operare per l'interesse collettivo.

Il primo passo del processo di condivisione delle conoscenze ha visto coinvolto il personale del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat e di Unioncamere nazionale. Successivamente sono stati formati tutti gli operatori impegnati nella rilevazione censuaria: a livello regionale i Responsabili Istat Territoriali (RIT) e i componenti degli Uffici Regionali di Censimento (URC), a livello provinciale i componenti degli Uffici Provinciali di Censimento (UPC), i coordinatori e i rilevatori, per un totale di oltre tremila persone.

La complessità organizzativa e la numerosità dei soggetti coinvolti ha portato a scegliere un metodo di formazione mista, in cui alla tradizionale formazione a cascata effettuata in presenza sono stati affiancati strumenti di formazione a distanza (FAD)<sup>7</sup>.

<sup>7</sup> Metodologia didattica che permette allo studente di dedicarsi all'attività di formazione dal proprio posto di lavoro o da casa. Discente e docente si trovano in spazi fisici diversi e l'attività formativa avviene anche in spazi temporali diversi. La comunicazione può avvenire attraverso differenti tecnologie telematiche. I primi casi di formazione a distanza coincidono con i corsi per corrispondenza, poi lo sviluppo delle tecnologie della comunicazione ha offerto nuove opportunità per la didattica in rete. Oggi, la tendenza a far ricorso a questo tipo di modalità si sta espandendo anche nelle pubbliche amministrazioni, tanto che il regolamento ex art.5 della L. 150/2000 prevede espressamente che una parte della formazione debba essere erogata in modalità FAD (CNIPA, 2007).

La formazione a cascata – dal livello nazionale a quello provinciale – si è svolta mediante formazione diretta “in presenza” ed è stata essenziale ai fini della motivazione degli operatori; inoltre ha consentito di raccogliere i *feedback* della rete e di utilizzarli per migliorare il processo.

Gli strumenti di formazione a distanza hanno completato e arricchito la preparazione acquisita in aula, hanno permesso l’aggiornamento delle competenze degli operatori e la gestione del *turnover* dei rilevatori.

I contenuti della formazione sono stati calibrati in modo da rispondere ai diversi fabbisogni informativi di ogni figura professionale (RIT, componenti URC e UPC, coordinatori e rilevatori).

I nuclei tematici comuni ai percorsi formativi di tutti i soggetti coinvolti sono stati:

- Organizzazione
  - 1.1 La rete: funzioni e compiti
  - 1.2 Il piano delle attività
- I questionari
  - 2.1 Questionari per le imprese
  - 2.2 Questionario per le istituzioni non profit
  - 2.3 Revisione dei questionari cartacei
- La rilevazione
  - 3.1 Le operazioni preliminari alla rilevazione
  - 3.2 La consegna dei questionari
  - 3.3 La restituzione spontanea
  - 3.4 Il completamento della rilevazione sul campo
- La piattaforma
  - 4.1 I questionari online
  - 4.2 SGRCIS
  - 4.3 Il sito della rete di rilevazione
- Profili giuridici e adempimenti
  - 5.1 La formazione del censimento
  - 5.2 L’obbligo di risposta e il segreto statistico

Tali contenuti sono stati diversamente declinati in funzione del profilo dei destinatari, al fine di rispondere in maniera coerente e sistematica ai diversi fabbisogni formativi.

Per sostenere e rafforzare la formazione in presenza sono stati progettati e sviluppati degli strumenti didattici integrativi fruibili online. A tal fine è stata implementata una sezione dedicata alla formazione nel Portale della rete di rilevazione del Censimento dell’industria e dei servizi (Figura 1.20).

Figura 1.20 – Home page dell'area Formazione nel portale della rete censuaria



In questa sezione sono stati realizzati i seguenti prodotti:

1. le linee guida per l'organizzazione e la somministrazione dei percorsi formativi sul territorio;
2. i consigli per gli istruttori: indicazioni di metodo e tecniche per la gestione degli eventi in presenza;
3. i manuali tecnici e i materiali informativi; rielaborati in chiave didattica;
4. i questionari ipertestuali delle rilevazioni sulle imprese e sulle istituzioni non profit.

Altro materiale – la documentazione ufficiale, il Piano Generale di Censimento, le norme, le classificazioni, le lettere informative, le FAQ, i *video tutorial* per la compilazione dei questionari – era fruibile sul portale del Censimento dell'industria e dei servizi, nell'area pubblica.

Tra i vari prodotti realizzati, in particolare i questionari ipertestuali (Figura 1.21) hanno rappresentato un'opportunità importante per veicolare in maniera innovativa i contenuti dei modelli di rilevazione, rendendo facilmente accessibili agli operatori della rete, e ai rispondenti, definizioni, classificazioni, obiettivi, regole di compilazione e altro materiale per agevolare la corretta comprensione e compilazione dei quesiti, garantendo un aumento della qualità dei dati raccolti.

I questionari ipertestuali, nella duplice versione per i rispondenti e per gli operatori, sono stati progettati e realizzati da un gruppo di lavoro delle sedi territoriali dell'Istat<sup>8</sup>.

I questionari ipertestuali sono stati utilizzati dagli operatori censuari come strumento di auto apprendimento a rinforzo della formazione ricevuta in aula, di consultazione nella fase di raccolta e come strumento didattico per la formazione in presenza dei rilevatori. I rispondenti hanno consultato il questionario ipertestuale per acquisire informazioni sull'indagine e indicazioni sulle corrette modalità di compilazione del modello. Gli elementi ipertestuali predisposti sono stati:

<sup>8</sup> Il gruppo di lavoro è stato istituito con delibera del Dipartimento per l'integrazione, la qualità e lo sviluppo delle reti di produzione e di ricerca n. 23 del 21 giugno 2012.

- pop-up con definizioni di parole e concetti chiave – attivabili dall'icona ⓘ. Per compilare correttamente un questionario economico è importante che il rispondente condivida le definizioni e le classificazioni utilizzate. Talvolta, infatti, il significato statistico di alcuni termini non coincide con quello comune o con quello attribuito da una certa sottopopolazione;
- pop-up con le regole di compilazione dei quesiti: istruzioni e, dove necessario, esempi di compilazione dei quesiti - attivabili dall'icona ⚙;
- pop-up con gli obiettivi di ciascuna sezione e di ciascun quesito – attivabili dall'icona 🎯. Conoscere l'obiettivo di un quesito ne agevola la comprensione;
- pop-up con allerta sui punti critici – icona ⚠;
- link alla guida alla compilazione – icona 📖;
- link a tutta la normativa richiamata nel questionario: rendere facilmente accessibili le norme che prescrivono diritti e doveri del titolare dell'indagine, degli operatori e dei rispondenti è un segno di trasparenza e di rispetto del rispondente e degli operatori;
- link al materiale informativo e integrativo sull'indagine (piano di censimento, circolari eccetera).

Inoltre nella versione per gli operatori sono stati predisposti anche:

- link ai riepiloghi dei controlli di sezione, che riportano i controlli di coerenza e compatibilità che devono essere effettuati dal rilevatore;
- link al manuale di istruzione – icona 📖.

I vantaggi dell'ipertestualità per i rispondenti e per gli operatori sono stati molteplici: accesso in maniera non sequenziale all'informazione ricercata; possibilità di trovare in un unico prodotto la documentazione di riferimento dell'indagine; possibilità di fare una ricerca "visiva"; accesso simultaneo a tutte le informazioni relative ad un certo quesito.

Pertanto la maggiore accessibilità del materiale informativo a supporto dell'indagine, agevolando la comprensione degli item del questionario, ha ridotto le ambiguità dei quesiti migliorando l'attendibilità delle risposte e quindi la qualità dei dati.

Figura 1.21 – Una pagina del questionario ipertestuale della rilevazione delle istituzioni non profit

← PAGINA 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 →

**29. Nel corso del 2011 l'istituzione non profit ha erogato servizi a persone?**  1 Sì  2 No ➔ (vai al quesito 30)

**NOTA BENE:** tra i servizi a persone non sono da includere quelli rivolti alla collettività in generale.

**29.1 Indicare il numero delle persone destinatarie dei servizi erogati nel 2011, suddividendole per settore di attività dell'istituzione non profit:**

Settore di attività	Numero
1. Cultura, sport e ricreazione	<input type="text"/>
2. Istruzione e ricerca	<input type="text"/>
3. Sanità	<input type="text"/>
4. Assistenza sociale e protezione civile	<input type="text"/>
5. Ambiente	<input type="text"/>
6. Sviluppo economico e coesione sociale	<input type="text"/>
7. Tutela dei diritti e attività politica	<input type="text"/>
8. Filantropia e promozione del volontariato	<input type="text"/>
9. Cooperazione e solidarietà internazionale	<input type="text"/>
10. Religione	<input type="text"/>
11. Relazioni sindacali e rappresentanza di interessi	<input type="text"/>
12. Altre attività	<input type="text"/>
<b>13. Totale</b>	<input checked="" type="checkbox"/>

Se la somma dei parziali non coincide con il totale, il rilevatore ha il compito di contattare l'istituzione non profit rispondente al fine di verificare l'informazione. Anche un numero elevato di destinatari dei servizi attiva un'incongruenza da risolvere contattando l'istituzione.

**29.2 Nel corso del 2011 l'istituzione non profit ha erogato servizi a persone con specifici disagi?**  1 Sì  2 No ➔ (vai al quesito 30)

**29.3 Indicare il numero delle persone con specifici disagi, destinatarie dei servizi erogati nel 2011, suddividendole per categoria di disagio prevalente:**

Il quesito ammette più risposte. Il rilevatore deve prestare attenzione affinché ogni persona sia considerata e quindi conteggiata una sola volta, cioè in una sola categoria di disagio.

3. Dipendenza (inclusi alcolismo e tossicodipendenza)	<input type="text"/>
4. Disabilità e non autosufficienza (anche parziale)	<input type="text"/>
5. Malattia	<input type="text"/>
6. Immigrazione e nomadismo	<input type="text"/>
7. Povertà ed esclusione sociale	<input type="text"/>
8. Abusi e molestie	<input type="text"/>
9. Prostituzione	<input type="text"/>
10. Altro.....	<input type="text"/>
<small>(specificare)</small>	
<b>11. Totale</b>	<input checked="" type="checkbox"/>

**NOTA BENE:** ogni persona deve essere considerata e conteggiata in una sola categoria di disagio.

**30. Nel corso del 2011 l'istituzione non profit ha utilizzato strumenti di comunicazione?**  1 Sì  2 No ➔ (vai al quesito 31)

**30.1 Indicare le modalità di comunicazione utilizzate (sono possibili più risposte):**

- 1 Newsletter periodica
- 2 Sito internet
- 3 Blog, forum, chat
- 4 Social network
- 5 Comunicati e brochure informative
- 6 Pubblicità
- 7 Bilancio di missione/bilancio sociale
- 8 Altro.....

(specificare)

**31. Nel corso del 2011 l'istituzione non profit ha svolto attività di raccolta fondi?**  1 Sì  2 No ➔ (vai al quesito 32)

**31.1 Indicare le modalità di raccolta fondi utilizzate (sono possibili più risposte):**

- 1 Mezzi di comunicazione di massa
- 2 Internet (siti web, blog, forum, social network)
- 3 Organizzazione di eventi e/o manifestazioni pubbliche
- 4 Vendita di beni e/o prodotti
- 5 Contatto diretto ➔ (vai al quesito 31.2)
- 6 Altro.....

(specificare)

**31.2 Per la raccolta fondi, l'istituzione non profit è entrata in contatto diretto con:**  (sono possibili più risposte)

- 1 Singoli cittadini
- 2 Istituzioni/imprese pubbliche
- 3 Istituzioni non profit
- 4 Imprese private



**DUBBI?  
Consulta  
la guida!**

13

RIEPILOGO CONTROLLI SEZIONE 5

Censimento Generale dell'Industria e dei Servizi e Censimento delle Istituzioni non Profit 2011

## Capitolo 2

### Il sistema di acquisizione<sup>9</sup>

Una tra le caratteristiche innovative nell'organizzazione e conduzione del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit (CIS) è sicuramente rappresentata dall'introduzione di un sistema di acquisizione dati via web (QCIS), che ha affiancato la tradizionale raccolta dati attraverso questionario cartaceo.

Il sistema QCIS è un'applicazione web che è stata utilizzata sia dai rispondenti per compilare il questionario censuario in una modalità perfettamente equivalente al questionario cartaceo sia dagli operatori della rete di rilevazione per inserire i dati precedentemente raccolti su 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 la Rilevazione sulle imprese (*long* e *short*, destinate a imprese con un numero di addetti rispettivamente maggiore/uguale e inferiore a venti) e il questionario per il Censimento delle istituzioni non profit.

Come già sottolineato, anche gli operatori della rete di rilevazione hanno utilizzato QCIS per registrare i questionari cartacei compilati dalle unità di rilevazione attraverso le medesime credenziali di accesso al Sistema di Gestione della Rilevazione (SGRCIS) di cui si parlerà approfonditamente nel Capitolo 3.

QCIS è pienamente integrato con il Sistema di Gestione della Rilevazione: non appena un rispondente completava un questionario in QCIS, lo stato di lavorazione del questionario veniva immediatamente aggiornato nel server centrale e visibile agli operatori della rete di rilevazione attraverso il sistema SGRCIS. Pertanto gli operatori hanno potuto utilizzare QCIS anche per effettuare le operazioni di *data entry*.

Data l'evidente criticità dell'applicazione QCIS nel processo produttivo censuario, 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 rappresenta 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 su un grande numero di utenti contemporanei (il campione era costituito da circa 700 mila rispondenti complessivi), dall'altro garantire la massima robustezza (nessun errore applicativo visualizzato dagli utenti) e la sicurezza dei dati, un aspetto critico in un'applicazione esposta a un pubblico vasto e quindi a potenziali rischi in questo senso.

<sup>9</sup> A cura di Antonino Virgillito. Autori del capitolo:

- Paragrafo 2.1 Leonardo Tininini;

- Paragrafo 2.2: Loredana Giunta (Paragrafo 2.2.1), Sonia Scialanca (Paragrafo 2.2.2), Francesco Cuccia (Paragrafi 2.2.3 e 2.2.4), Cecilia Addone, Antonino Virgillito (Paragrafo 2.2.5).

Sono state adottate tecniche sofisticate nel progetto dell'applicazione sia per ridurre il carico verso la base dati anche in presenza di un alto numero di utenti connessi all'applicazione sia per eseguire la validazione dei dati in diversi punti dell'applicazione garantendo la consistenza dei dati immessi dagli utenti.

QCIS è basato sull'analogo sistema realizzato per la rilevazione del 15° Censimento della popolazione e delle abitazioni. Nonostante questo, è stato necessario estendere in modo significativo il progetto originale per includere da un lato caratteristiche specifiche delle rilevazioni del CIS, dall'altro miglioramenti tecnologici generali. Per quanto riguarda la prima categoria di caratteristiche, nel questionario CIS erano previste tipologie di domande complesse che non erano presenti nel questionario del Censimento della popolazione. Esempi sono matrici di valori numerici con calcolo automatico dei totali o matrici di *check* mutuamente esclusivi. Tali domande presentano notevoli difficoltà sia dal lato della modellazione dei metadati che da quello della visualizzazione, che da quello dei controlli.

I miglioramenti tecnici generali invece riguardano principalmente due aspetti. Il primo è il miglioramento del motore dei controlli, sia dal punto di vista del linguaggio usato per le regole, sia per quanto riguarda la visualizzazione dei messaggi di errore. Il secondo aspetto è l'accessibilità per gli utenti con disabilità. L'interfaccia utente è stata progettata in modo che tutto il questionario potesse essere compilato interamente anche da utenti ipovedenti attraverso *screen reader*, aderendo pienamente alle indicazioni della normativa vigente in materia di accessibilità.

Il capitolo è articolato in due parti: nella prima (Paragrafo 2.1) viene presentata la modellazione del questionario, la seconda (Paragrafo 2.2 e sottosezioni) è dedicata al progetto del sistema.

## 2.1 La modellazione del questionario

Questa sezione è dedicata all'illustrazione della tecnica di modellazione utilizzata per i questionari elettronici del Censimento delle istituzioni non profit e della rilevazione campionaria sulle imprese. Tale tecnica rappresenta un'evoluzione di quella precedentemente utilizzata per il Censimento della popolazione e delle abitazioni, della quale condivide l'approccio formale di base, pur con alcune modifiche di un certo rilievo. Tali modifiche derivano in parte dall'esperienza maturata con il precedente censimento e in parte dall'obiettivo di rendere l'approccio maggiormente generalizzato e quindi utilizzabile in un ventaglio di contesti più ampio. I principi alla base della tecnica di modellazione adottata sono stati i seguenti:

- memorizzazione nella base dati di tutti i testi ("localizzati" nelle versioni italiano, tedesco e sloveno) presenti nell'interfaccia web del questionario, in modo che le modifiche e correzioni potessero essere realizzate semplicemente modificando il contenuto di uno o più campi di una tabella della base dati;
- semplicità di caricamento delle varie versioni localizzate dei testi, che consentisse correzioni e aggiornamenti anche parziali in modo agevole e poco invasivo, senza richiedere la ricompilazione dei programmi;
- modellazione e memorizzazione nella base dati 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'impresa ha effettuato attività di formazione aziendale, passare a compilare la domanda Y";
- sulla base della modellazione della componente strutturale di cui al punto precedente e delle risposte già fornite dall'intervistato, elaborazione di un algoritmo che determinasse in modo totalmente automatico le domande da mostrare come "attive" e "non attive" nell'interfaccia utente. In altre parole si volevano sfruttare le informazioni sulla struttura del

questionario per eseguire (almeno in parte) la validazione delle risposte fornite e, lato client, “accendere” o “spegnere” alcune domande, sulla base delle risposte fornite.

- 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*. In pratica l'idea è che, una volta scritto il frammento di codice di programmazione che realizzava il comportamento di una domanda di un certo tipo (es. una “domanda con scelta multipla semplice”), l'implementazione di un'altra domanda dello stesso tipo potesse essere ottenuta usando il medesimo frammento software parametrico a cui veniva passato solo ciò che era realmente specifico, tipicamente l'identificativo della stessa e altre caratteristiche dipendenti dal tipo (es. lo specifico elenco di modalità di risposta).

### 2.1.1 La modellazione della struttura del questionario

La prima famiglia di dati oggetto della modellazione è quella che potremmo chiamare la parte *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 dati sono:

- memorizzazione della struttura di base del questionario (essenzialmente l'elenco dei quesiti che lo compongono);
- memorizzazione dei testi che compongono l'interfaccia web del questionario, in modo da semplificare eventuali modifiche e correzioni;
- realizzazione semplificata delle versioni “localizzate” (cioè in più lingue diverse) del questionario;
- realizzazione di una corrispondenza automatica tra i vari componenti dell'interfaccia utente visualizzati, i componenti costitutivi dell'applicazione (Java Beans) e i dati scritti/letti nella base dati;
- realizzazione di alcuni controlli di base sull'input dell'utente.

Nel seguito si analizzano brevemente le componenti fondamentali dedicate alla modellazione della parte statica del questionario e le loro principali funzioni.

*Il questionario.* Con questo termine ci si potrebbe riferire sia ad una struttura di questionario (che nel contesto tradizionale potrebbe coincidere con il modello cartaceo da inviare ai rispondenti) che al particolare questionario (corredato di risposte) restituito dallo specifico rispondente. In questo contesto si è interessati alla prima accezione del termine, cioè al questionario inteso come una struttura da riempire, costituita da una serie di quesiti, elencati secondo una sequenza e regole di compilazione ben precise. Ogni questionario è identificato nella base dati da un codice univoco di identificazione e da una descrizione testuale che ne chiarisce brevemente il contenuto. Il codice identificativo viene utilizzato nella tabella delle domande (di cui si parlerà oltre) per stabilire di fatto quali siano i quesiti che lo costituiscono.

*Le domande.* Questa componente fondamentale di modellazione (a cui corrisponde nella base dati una specifica tabella *QS\_DOMANDA*), rappresenta non solo, come è ovvio dal nome, 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 alcuni casi particolari, alla corretta realizzazione del flusso di compilazione. Nella tabella delle domande sono memorizzate le componenti informative più essenziali di un quesito, in particolare il questionario di cui esso è parte, la tipologia (che indica, ad esempio, se si tratta di un quesito che prevede la scelta tra una serie di possibili risposte prefissate, oppure un testo libero, oppure ancora una scelta multipla tra una lista/matrice di risposte possibili, eccetera), l'indice di

ordinamento topologico (che specifica l'ordine esatto secondo il quale i vari quesiti vengono visualizzati all'utente nella pagina web e di cui si parlerà diffusamente più avanti) e varie informazioni su controlli e comportamenti di base, perlopiù dipendenti dalla specifica tipologia di domanda. Ad esempio se il quesito prevede una matrice di opzioni con possibilità di scelta multipla (*checkbox*), è possibile specificare il numero massimo di risposte consentito nel complesso e/o su ogni riga/colonna. Nel caso di matrici di valori numerici è possibile specificare se visualizzare in automatico i totali di riga e/o colonna e così via.

*I testi di domanda.* È importante sottolineare che la componente domanda descritta sopra prescinde dal particolare testo attraverso cui il quesito viene formulato. Il testo o, più precisamente, i vari testi associati a un medesimo quesito sono infatti memorizzati in una componente separata (a cui corrisponde nella base dati una specifica tabella QS\_TESTI\_DOMANDA). Come già detto, i testi associati a un medesimo quesito sono in generale più di uno e questo non solo per via delle diverse lingue, ma anche perché il quesito viene spesso formulato attraverso più testi con finalità e stili di visualizzazione diversi. Ad esempio nel frammento di Figura 2.1, è presente una parte che costituisce il testo vero e proprio del quesito e un'altra parte che costituisce un testo per così dire ausiliario, la cui finalità è fornire delle indicazioni/istruzioni su come rispondere al quesito stesso. Le due tipologie di testi sono distinte a livello di base dati in modo da garantire una visualizzazione corretta e coerente nell'ambito dell'intero questionario: ad esempio una volta stabilito quale sia lo *stile* per visualizzare la componente istruzioni, esso viene utilizzato per tutte le domande che ne prevedono una ed è possibile modificarlo in modo coerente agendo semplicemente sul foglio di stile CSS della pagina HTML.

**Figura 2.1 – Un esempio di quesito con testi che richiedono stili di visualizzazione diversi**

---

**3.01 Indicare se nel 2011 e nel 2012 l'impresa ha intrattenuto le relazioni indicate di seguito**  
(sono possibili più risposte per ciascun tipo di relazione: se in essere sia nel 2011 sia nel 2012, barrare entrambe le caselle)

---

*I tipi di domanda.* Come già detto, 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 componenti software per la visualizzazione dei dati e la loro memorizzazione, massimizzando il riuso e la robustezza del codice prodotto, oltre ovviamente a garantire un'elevata modularità ed estensibilità. Di fatto quindi il catalogo dei tipi di domanda supportati (che sono memorizzati nella tabella specifica QSD\_TIPO\_DOMANDA) 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.

*Le modalità di risposta e i relativi testi.* 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 sono modellate dal concetto di modalità di risposta e memorizzate nella tabella QS\_MODALITA\_RISPOSTA della base dati. Come per QS\_DOMANDA, anche in questa tabella vengono rappresentate le componenti più essenziali di

una specifica risposta (indice di ordinamento, eventuale posizionamento in riga o colonna, stile di visualizzazione, eccetera), mentre i testi corrispondenti (nelle varie lingue disponibili) sono memorizzati in una tabella a parte QS\_TESTI\_MODALITA, facilmente modificabile secondo necessità, senza richiedere ricompilazioni del codice di programmazione.

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

È stato evidenziato che tra le informazioni memorizzate nella parte statica del questionario, per ogni domanda è presente anche un indice di sequenza che specifica l'ordine secondo il quale i vari quesiti verranno visualizzati all'utente nella pagina web. Tuttavia la sequenza di quesiti presenti in un questionario non rappresenta praticamente mai il percorso effettivo di compilazione di un rispondente (lo rappresenterebbe soltanto se tutti i rispondenti dovessero rispondere a tutti i quesiti presenti, indipendentemente dalle risposte date). Nella pratica sappiamo che la compilazione prevede spesso dei "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).

La tecnica utilizzata per rappresentare i percorsi di compilazione del questionario 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 è quello corrispondente al frammento di questionario delle imprese riportato in Figura 2.2. Il sottografo corrispondente a tale frammento è mostrato in Figura 2.3. Per facilitare la lettura del diagramma sono stati utilizzati dei codici corrispondenti ai numeri di quesito nel questionario cartaceo e si è aggiunta una breve descrizione nei blocchi che rappresentano i nodi del grafo (e quindi le domande del questionario). Le etichette condizione sono state espresse in forma discorsiva, anche se nella realtà viene utilizzato un formalismo logico specifico che può essere interpretato dal programma che implementa il questionario online.

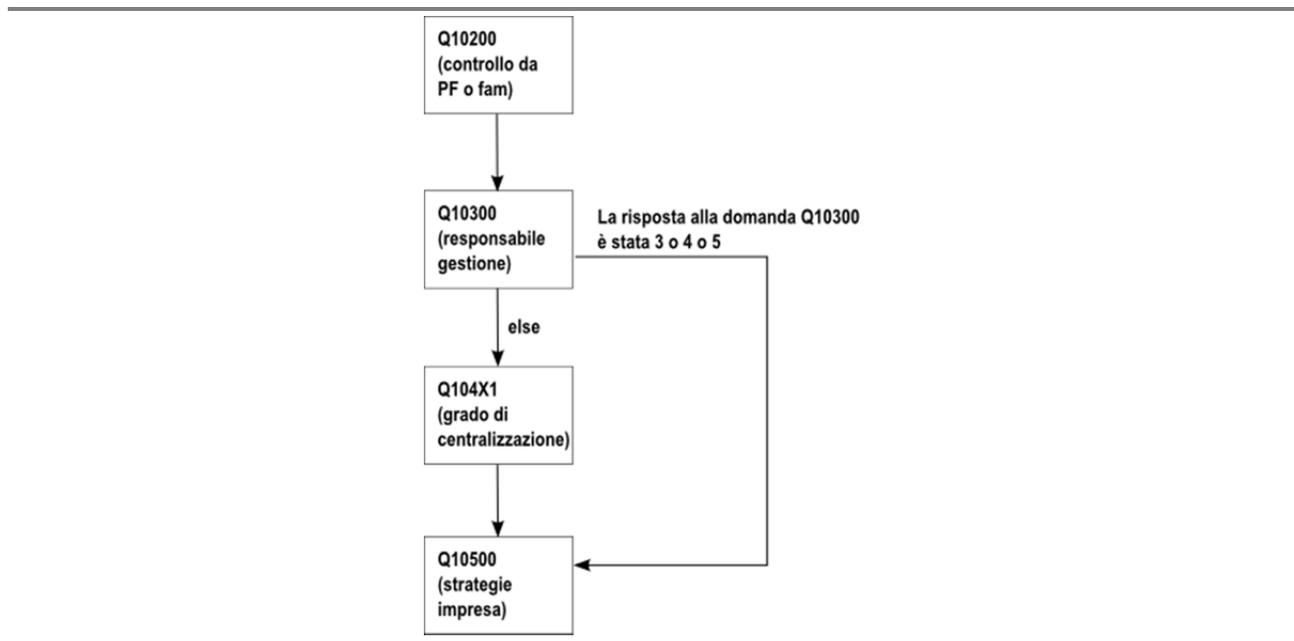
**Figura 2.2 – Un frammento del questionario per le imprese con 10 addetti o più**

<p>1.02 L'impresa era direttamente o indirettamente controllata da una persona fisica o da una famiglia al 31/12/2011?</p> <p>1 <input type="checkbox"/> Sì</p> <p>2 <input type="checkbox"/> No</p>	<p>1.03 Indicare chi aveva la responsabilità della gestione dell'impresa al 31/12/2011</p> <p>1 <input type="checkbox"/> L'imprenditore/il socio principale o unico</p> <p>2 <input type="checkbox"/> Un altro membro della famiglia proprietaria/controlante</p> <p>3 <input type="checkbox"/> Un manager selezionato all'interno dell'impresa → <a href="#">vai al quesito 1.05</a></p> <p>4 <input type="checkbox"/> Un manager selezionato all'esterno dell'impresa → <a href="#">vai al quesito 1.05</a></p> <p>5 <input type="checkbox"/> Nessuna delle precedenti → <a href="#">vai al quesito 1.05</a></p>
--	--

1.04.X1 Indicare, per ciascuna delle seguenti funzioni aziendali, il grado di centralizzazione delle decisioni strategiche (massimo una risposta per funzione aziendale)				
	Grado di accentrimento decisionale			
	Alto	Medio	Basso	Nessuno
A Attività principale	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Funzioni di supporto</b>				
B Approvvigionamento	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
C Marketing, vendita e servizi post-vendita	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
D Distribuzione, trasporto e magazzino	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
E Amministrazione/finanza	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
F Ricerca e sviluppo, innovazione, progettazione	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
G Tecnologie informatiche (ICT)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
H Gestione e formazione delle risorse umane	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

Figura 2.3 – Il (sotto)grafo corrispondente al frammento di Figura 2.2



La transizione dal quesito 1.02 a 1.03 è incondizionata; è infatti presente un arco senza etichette che congiunge il nodo Q10200 al nodo Q10300. Dal quesito 1.03 le possibili uscite sono invece due: verso il quesito 1.05 nel caso in cui la modalità di risposta sia 3, 4 o 5 e verso il quesito 1.04.X1, altrimenti. In uscita dal nodo Q10300 si hanno quindi due archi: uno con etichetta “*la risposta alla domanda Q10300 è stata 3 o 4 o 5*” che porta al nodo Q10500 e uno con etichetta “*else*” che porta al nodo Q104X1. Infine è presente un arco incondizionato da Q104X1 a Q10500 che ricongiunge le due diramazioni.

Come già anticipato, poiché le etichette rappresentano delle espressioni logiche (booleane), è fondamentale stabilire quale linguaggio formale utilizzare per esprimere tali espressioni, tenendo presente che se esso risulta troppo “povero” (in termini di potere espressivo) potrebbe essere difficile o addirittura impossibile esprimere alcune delle condizioni richieste e, di contro, se risulta troppo ricco, il *parser* (cioè quella parte del programma che avrà il compito di analizzare l’espressione logica e tradurla in una procedura che possa produrre di volta in volta un valore vero o falso) potrebbe diventare piuttosto complesso e la valutazione dell’espressione onerosa dal punto di vista elaborativo.

Come nel caso delle tipologie di domanda, anche per le tipologie di condizioni, il sistema prevede un approccio di tipo modulare, in base al quale si possono prevedere condizioni atomiche e meccanismi di combinazione di tali condizioni più o meno complessi, a seconda delle tipologie di condizioni che il questionario prevede. Nello specifico, la notevole complessità delle condizioni necessarie a modellare il questionario del CIS ha fatto sì che le tipologie di condizioni previste per tale questionario fossero significativamente più potenti rispetto a quelle che erano state a suo tempo utilizzate per il questionario del censimento popolazione e abitazioni.

Alcuni esempi di tipiche condizioni atomiche di “salto” sono:

- La risposta (o una delle risposte nel caso di *multiresponse*) dell'utente alla domanda D è stata la modalità di risposta con indice  $i$ ?
- La risposta (o una delle risposte nel caso di *multiresponse*) dell'utente alla domanda D è stata una delle modalità di risposta con indice compreso tra  $i$  e  $j$ ?
- La risposta (o una delle risposte nel caso di *multiresponse*) dell'utente alla domanda D è stata una delle modalità di risposta con indice appartenente all'insieme  $\{i_1, i_2, \dots, i_N\}$ ?
- La data specificata dall'utente in risposta alla domanda D è maggiore (o minore) della data  $D_x$  (oppure compresa tra le due date  $D_x$  e  $D_y$ )?
- Il valore (numerico) specificato dall'utente in risposta alla domanda D è maggiore (o minore) del valore di soglia  $N_x$  (oppure compreso tra i due valori numerici  $N_x$  e  $N_y$ )?
- L'utente ha biffato almeno una delle caselle nella riga (o colonna)  $i$ -esima di una matrice di opzioni?
- L'utente ha biffato tutte le caselle della riga (o colonna)  $i$ -esima di una matrice di opzioni?

Le condizioni dei primi tre tipi sono di norma piuttosto semplici da esprimere e possono essere ulteriormente semplificate prevedendo di mettere in OR condizioni elementari di uguaglianza ad un valore. Qualora il dominio di variabilità della risposta sia potenzialmente infinito (come nel caso di date e valori numerici) si rendono necessarie delle tipologie di condizione specifiche che di fatto si possono ridurre a prevedere una condizione di appartenenza ad un intervallo (in cui uno dei due estremi potrebbe eventualmente essere infinito) e sono anch'esse di norma abbastanza semplici da esprimere. Più complesso il discorso su condizioni come le ultime due degli esempi, che prevedono una forma più o meno esplicita di quantificazione esistenziale e universale. La Figura 2.4 illustra un frammento del questionario delle imprese che prevede una condizione atomica con quantificazione esistenziale (“è stata barrata almeno una casella, eccetera”). In pratica si tratta di verificare se esiste almeno una biffatura nell'intervallo della matrice di risposte indicato.

Figura 2.4 – Un'istruzione di compilazione rappresentabile con una quantificazione esistenziale

#### SOTTOSEZIONE 3.07 - ALTRI ACCORDI

RISPONDERE ALLA SOTTOSEZIONE 3.07 SE L'IMPRESA HA ISTITUITO ACCORDI DI COLLABORAZIONE, ALTRI ACCORDI FORMALI O ALTRO, OSSIA È STATA BARRATA ALMENO UNA CASELLA CON CODICE "2" (SÌ, NEL 2011) E/O CON CODICE "3" (SÌ, NEL 2012) PER I SEGUENTI ACCORDI: CONSORZIO, CONTRATTO DI RETE, ALTRI ACCORDI FORMALI, *FRANCHISING*, ACCORDO INFORMALE E ALTRO.

La Figura 2.5 illustra invece un esempio in cui la condizione è rappresentabile con una quantificazione universale. Si tratta infatti di verificare che non esista nemmeno una biffatura su un determinato intervallo di celle (ovvero, equivalentemente, che tutte le celle dell'intervallo stesso siano non biffate). L'intervallo in questione è quello delimitato da tutte le righe della matrice e dalle sue ultime due colonne. La compilazione dovrà procedere alla domanda 3.09.1 solo se tutte le celle di tale intervallo risultano essere non biffate.

Figura 2.5 – Un'istruzione di compilazione rappresentabile con una quantificazione universale

3.01 Indicare se nel 2011 e nel 2012 l'impresa ha intrattenuto le relazioni indicate di seguito  
(sono possibili più risposte per ciascun tipo di relazione: se in essere sia nel 2011 sia nel 2012, barrare entrambe le caselle)

	No	Sì	
		nel 2011	nel 2012
A Commessa (l'impresa acquista beni e/o servizi)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
B Subfornitura/subappalto (l'impresa vende beni e/o servizi)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
C Consorzio	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
D Contratto di rete	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
E Altri accordi formali ( <i>joint venture</i> , ATI, ecc.)	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
F <i>Franchising</i>	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
G Accordo informale	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>
H Altro	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>

SE SONO STATE BARRATE ESCLUSIVAMENTE CASELLE CON CODICE "1" (NO), VAI AL QUESITO 3.09.1.

Un formalismo che preveda quantificazioni e livelli di nidificazione arbitrari è molto complesso sia da analizzare che da valutare. Tuttavia nel caso specifico si è sfruttato il fatto che i domini da analizzare sono finiti e "simulato" le quantificazioni con delle semplici condizioni numeriche basate su operatori di conteggio che, proprio grazie alla finitezza (e alle ridotte dimensioni degli intervalli) possono essere valutati molto rapidamente e in modo semplice.

In pratica una quantificazione esistenziale viene simulata con un verifica di conteggio maggiore di zero e una quantificazione universale su un dominio di  $N$  variabili con una verifica che il conteggio delle variabili biffate sia esattamente pari a  $N$ . Questa strategia, che sarebbe inefficiente qualora il numero di variabili da verificare fosse elevato, risulta invece vincente nel caso specifico, perché gli elenchi e le matrici da verificare sono di dimensioni molto ridotte e risulta del tutto trascurabile il tempo necessario per effettuare il *parsing* della formula che nel caso più complesso è costituito da un intervallo bidimensionale, all'interno del quale effettuare il conteggio delle biffature, e da una coppia di valori  $valMin$  e  $valMax$ .

In sostanza l'espressione viene valutata *true* se il valore di conteggio nell'intervallo bidimensionale specificato è compreso tra  $valMin$  e  $valMax$  (e valutata *false* altrimenti). La condizione è espressa quindi da una tupla di sei valori ( $rigaMin$ ,  $rigaMax$ ,  $colMin$ ,  $colMax$ ,  $valMin$ ,  $valMax$ ), che si può leggere informalmente come segue: "Il conteggio delle biffature nell'intervallo individuato da ( $rigaMin$ ,  $rigaMax$ ,  $colMin$ ,  $colMax$ ) è compreso tra  $valMin$  e  $valMax$ ?". Ad esempio la tupla corrispondente alla condizione di salto di Figura 2.5 è la seguente (1,8,2,3,0,0) che può essere letta come segue: "Il conteggio delle biffature nell'intervallo compreso tra la prima e l'ottava riga e tra la seconda e la terza colonna è compreso tra 0 e 0, cioè esattamente uguale a 0"

Per quanto riguarda i tre operatori booleani standard:

- la congiunzione (AND) si ottiene mettendo in serie più archi contenenti le varie condizioni da combinare (e utilizzando i nodi virtuali accennati in precedenza (Figura 2.6));
- la disgiunzione (OR) si ottiene molto banalmente, prevedendo più archi in parallelo tra la medesima coppia di nodi (Figura 2.7);
- la negazione (NOT), qualora non sia direttamente presente sulle condizioni atomiche, si ottiene "spingendo" le negazioni fino al livello delle condizioni atomiche (sfruttando ad esempio le leggi di De Morgan) e agendo quindi direttamente su di esse (che sempre per l'ipotesi di finitezza possono sempre esprimere le condizioni e le rispettive negate).

Ciò che rende particolarmente interessante questa soluzione è che la valutazione delle condizioni logiche è assolutamente banale dal punto di vista programmatico e può essere svolta in maniera molto efficiente. Questo 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 rappresenta un requisito fondamentale.

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 (QS\_TRANSIZ\_DOMANDA). In pratica ogni record presente in questa tabella corrisponde biunivocamente ad un arco del grafo che rappresenta i vari percorsi di compilazione del questionario.

Figura 2.6 – Simulazione di condizioni in AND tramite nodo virtuale

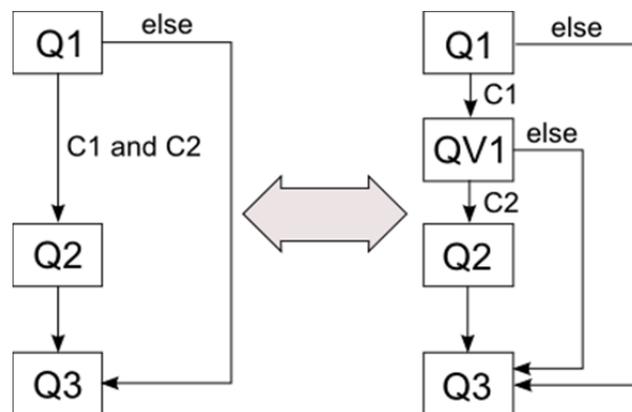
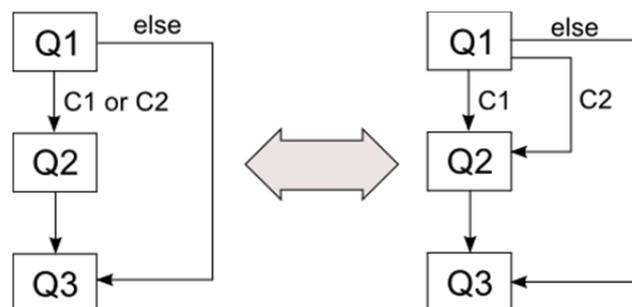


Figura 2.7 – Simulazione di condizioni in OR tramite archi in parallelo



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 costruendo un diagramma concettuale che descrive i percorsi di compilazione e utilizzare per esprimere le condizioni sugli archi il formalismo di alto livello che più si ritiene opportuno. 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, saremmo in presenza di una indeterminatezza di comportamento, perché non sapremmo quale arco/percorso seguire);
- in uscita dal medesimo nodo deve essere presente al più un arco con una condizione di tipo "else";
- 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 porteranno alla sua definizione logico-fisica (il vero e proprio *grafo del questionario*), dipendente essenzialmente dallo specifico formalismo utilizzato per esprimere le condizioni sugli archi. Se il formalismo prevede condizioni complesse espresse con esplicite quantificazioni e operatori logici, le trasformazioni saranno minime, ma si renderà necessario un *parser* sofisticato e complesso per consentire la valutazione delle condizioni durante il normale funzionamento del questionario elettronico.

Se, come nel caso censuario, il formalismo prevede un formalismo *ad hoc* con meccanismi specifici per simulare, ad esempio, le condizioni in AND, si potrebbero rendere necessarie delle trasformazioni più significative, che non altereranno tuttavia il comportamento del questionario così prodotto in termini di percorsi di compilazione. D'altra parte questo sforzo aggiuntivo è ripagato dal fatto che risulterà enormemente semplificata la valutazione delle condizioni logiche. L'esperienza del questionario della Rilevazione sulle imprese e del Censimento delle istituzioni non profit (che, come abbiamo mostrato, prevedevano un numero e tipologie di condizioni tutt'altro che banali) ha richiesto in pochissimi casi l'introduzione di nodi virtuali, come visto sopra, per simulare condizioni in AND, per cui il modello concettuale del questionario è risultato quasi identico a quello logico-fisico.

### **2.1.3 Come usare il grafo del questionario**

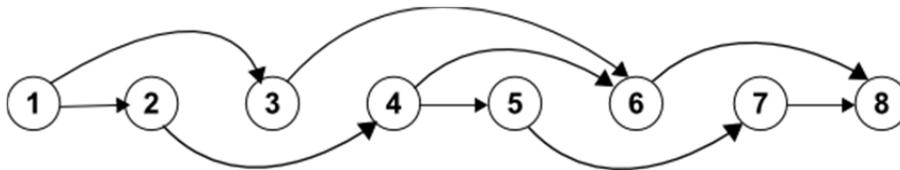
La concettualizzazione dei percorsi di compilazione attraverso la definizione del grafo del questionario è sicuramente interessante di per sé, tuttavia ciò che giustifica realmente l'innegabile sforzo necessario per la sua costruzione e verifica è il fatto che tale modellazione può essere utilizzata 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 dei "reset" di alcune risposte a seguito di modifiche. Come si è visto in precedenza il grafo del questionario non è un grafo qualsiasi, bensì 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  $u < v$ .

In modo equivalente si può supporre di definire una funzione  $s$  che associa 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 è possibile descrivere questa proprietà immaginando di disporre i nodi del grafo lungo una retta secondo l'ordine fissato dalla funzione  $s$  (Figura 2.8). In base alla proprietà descritta, una volta disposti i nodi secondo l'ordinamento topologico, gli archi del DAG saranno tutti orientati "da sinistra a destra".

Figura 2.8 – Ordinamento topologico dei nodi di un DAG



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 all'ovvia 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 destinazione. Ad esempio la lista (ordinata) corrispondente al grafo raffigurato sopra, è quella raffigurata nella Figura 2.9.

Vista la corrispondenza stabilita, nel seguito si parlerà 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 dirà 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 possiamo 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 2.9 – La lista ordinata di coppie di indici che rappresenta il grafo di Figura 2.8

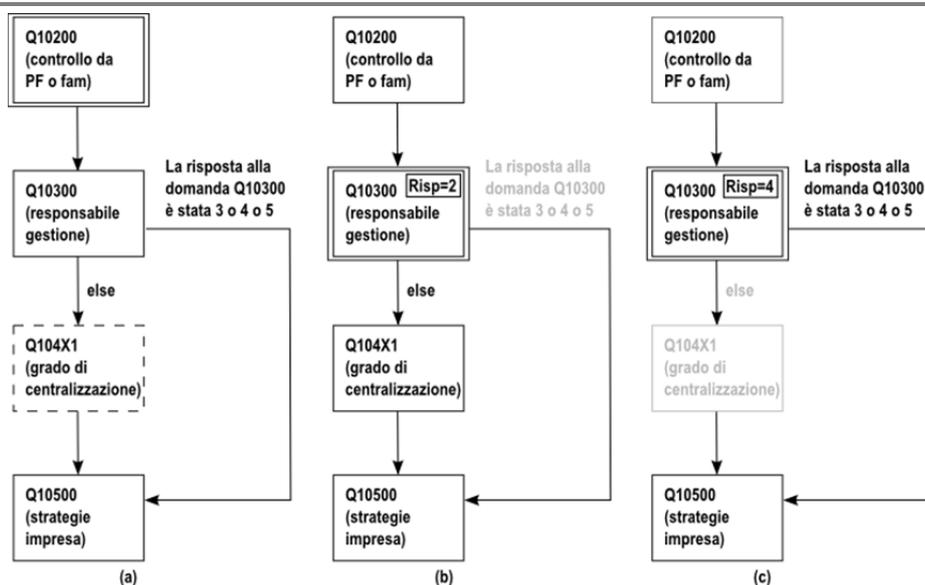
Sorgente	Destinazione
1	2
1	3
2	4
3	6
4	5
4	6
5	7
6	8
7	8

La situazione è illustrata graficamente in Figura 2.10, dove consideriamo nuovamente il sottografo di Figura 2.3 (corrispondente al frammento di questionario di Figura 2.2).

Nella porzione (a) della Figura 2.10 supponiamo che sia stata fornita una risposta al quesito 1.02 (il nodo Q10200 è compilato), ma non sia stata ancora fornita una risposta al quesito 1.03 (nodo Q10300). I nodi Q10300 e Q10500 sono mostrati con tratto continuo, perché saranno *necessariamente raggiunti* dai successivi passi della compilazione. A partire dal nodo attualmente compilato Q10200 non esistono infatti sequenze di risposte (a cui corrispondono naturalmente percorsi di compilazione) che consentono di escludere Q10300 e Q10500 dalla compilazione. Al contrario Q104X1 è mostrato tratteggiato perché è solo *potenzialmente raggiungibile*: esistono infatti alcune sequenze di risposte che ne determinano la compilazione e altre che fanno sì che non debba essere compilato.

Nella porzione (b) della figura si suppone che la compilazione proceda e che venga fornita la risposta 2 al quesito 1.03 (nodo Q10300). Lo stato del nodo Q104X1 cambia e diventa anch'esso necessariamente raggiunto dalla compilazione, perché sulla base delle risposte fornite e della struttura del questionario non c'è modo di escludere la sua compilazione.

Figura 2.10 – Lo stato dei nodi a seconda delle risposte fornite



Infine nella porzione (c) della Figura 2.10 si suppone che la compilazione proceda diversamente a partire dalla domanda Q10200 e venga fornita la risposta 4 al quesito 1.03 (nodo Q10300). Lo stato del nodo Q104X1 cambia anche in questo caso, ma in modo diverso e diventa *irraggiungibile* (questo viene evidenziato nella figura cambiandone i contorni e le scritte in grigio per analogia con quanto accade nell'interfaccia utente dei questionari elettronici): 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 Q10300.

Lo stato di ogni nodo determina quindi se la domanda dovrà essere abilitata alla compilazione e/o lo stile di visualizzazione (o anche, per alcune situazioni specifiche realizzate con il questionario delle imprese, far sì che la domanda venga completamente nascosta dall'interfaccia utente). D'altra parte, come si è visto, lo stato dipende dalle risposte 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 efficiente 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 ed esemplificato in Figura 2.9. 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. Ulteriori dettagli sull'utilizzo del grafo del questionario, sia lato client che lato server, verranno forniti nel Paragrafo 2.2.3.

## 2.2 Il progetto del sistema di acquisizione

In questo paragrafo vengono presentati gli aspetti caratterizzanti del progetto dell'applicazione. Dopo una panoramica dell'architettura interna del sistema ne vengono presentati gli aspetti più innovativi e le soluzioni tecniche adottate.

### 2.2.1 L'architettura interna

QCIS è una applicazione web implementata attraverso la piattaforma *Java2 Enterprise Edition*, organizzata in un'architettura a cinque strati basati sul *design pattern Model-View-Controller*. L'applicazione sfrutta tre diffusi *framework open source*, vale a dire *Struts2*, *Spring* e *Hibernate*. La progettazione di applicazioni a strati è illustrato nella Figura 2.11.

Figura 2.11 – L'architettura a cinque strati di QCIS



La funzionalità di ogni strato è di seguito dettagliata:

- *GUI*: pagine JSP che implementano l'interfaccia utente grafica. 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. Essi 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.

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, eccetera) venivano caricati nella memoria principale. I metadati memorizzati nella *cache* venivano 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 mostrava una pagina del questionario, esso innanzitutto caricava tutti i metadati relativi alle domande di quella pagina. Tali metadati includevano tutti i testi (tenendo conto anche della localizzazione) e la diversa tipologia (*check*, *multiresponse*, testi liberi, eccetera).

Sono stati definiti più di venti diversi tipi di domanda, ognuno associato ad un differente *template*. Un *template* specificava come un tipo di domanda doveva essere costruita in termini di HTML ed era associato ad uno specifico *Java bean* per implementare il corrispondente trattamento dei dati (Paragrafo 2.2.2). L'utilizzo 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 dovevano superare i controlli di qualità. Tali controlli venivano effettuati innanzitutto lato client attraverso dei controlli *JavaScript* con una correzione immediata del dato errato (Paragrafo 2.2.4). Gli stessi controlli venivano poi ripetuti lato server per una verifica finale prima del salvataggio finale dei dati. Questo passo era fondamentale per motivi di sicurezza, un utente malintenzionato avrebbe potuto aggirare i controlli sulla pagina modificando i dati inviati attraverso la richiesta HTTP. Questo avrebbe potuto portare all'inconsistenza dei dati o addirittura a minacce alla sicurezza (es. iniezione di codice SQL malizioso che poteva rilevare dati sensibili dal database).

L'operazione di salvataggio ha sfruttato ampiamente le peculiarità dei *framework* utilizzati, con l'ausilio della *Reflection* di Java. I dati inviati tramite HTTP venivano salvati automaticamente da *Struts2* in un oggetto Java (l'entità). A quel punto, *Hibernate* era 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 stata 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 un'applicazione più robusta.

### 2.2.2 Tipologie di domande

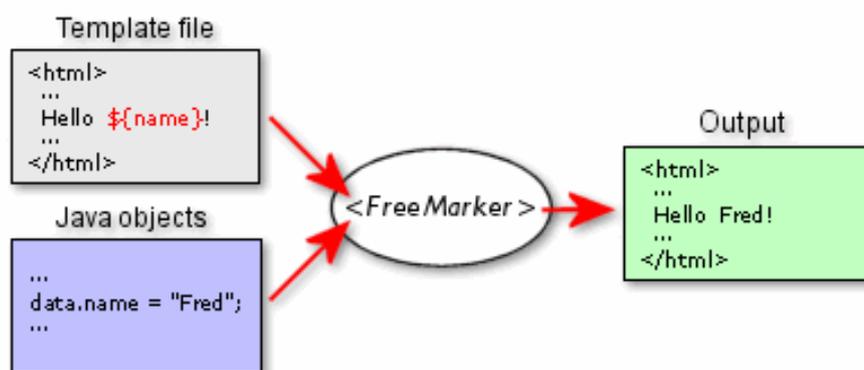
Per la riproduzione su web delle domande previste per il questionario del Censimento industria e servizi (QCIS) si è deciso di utilizzare uno strumento che consentisse piena flessibilità nella definizione grafica delle pagine: *FreeMarker* (Figura 2.12).

*FreeMarker* è una libreria *open source*, basata su Java, che si innesta perfettamente nell'architettura MVC delle applicazioni web *servlet-oriented*, come quella del questionario del Censimento industria e servizi.

Questo strumento è in grado di generare contenuti dinamici di pagine web sulla base di *template* generici o personalizzati: in pratica, fornisce una serie di documenti con una parte di codice HTML statico, che viene ripetuto per ogni pagina web in cui è incluso il *template*, e una parte di codice dinamico in cui possono essere richiamate delle variabili attraverso dei *placeholder*.

Tali *placeholder* permettono ad ogni *template* di accedere agli oggetti Java utilizzati dalle stesse pagine JSP per modellare i dati, costruendo così una struttura HTML adattabile al contesto.

Figura 2.12 – Utilizzo dei template in FreeMarker



Oltre alla lettura degli oggetti Java, *FreeMarker* consente di lavorare con dei costrutti particolari per eseguire iterazioni, assegnazioni o segmenti condizionali, in modo da permettere una selezione programmatica della parte di codice HTML che viene inserita nella pagina web finale.

*FreeMarker* mette a disposizione dei *template* di default per la gestione dei singoli elementi HTML più semplici, come *checkbox*, *radio button*, *text area*, *combobox*, *label*, eccetera.

Questi semplici pacchetti di codice predefinito, però, non si sono rivelati sufficienti per coprire le esigenze specifiche dell'applicazione QCIS, dove erano presenti tipologie di domande particolarmente complesse e non riconducibili agli elementi normalmente presenti nelle *form web* canoniche.

Si è reso necessario, quindi, sviluppare dei *template* personalizzati che venivano richiamati di volta in volta a seconda delle tipologie delle domande.

Sono state identificate 26 tipologie di domande:

1. titolo di punto (domanda virtuale);
2. lista di risposte (radio);
3. data completa (giorno/mese/anno);
4. selezione di provincia e comune;
5. selezione di stato estero;

6. mese e anno;
7. istruzioni di compilazione (domanda virtuale);
8. lista di risposte multiple (checkbox);
9. domanda su famiglie coabitanti;
10. valore numerico intero;
11. lista di risposte (menu a tendina);
12. selezione di Titolo di studio per esteso;
13. testo libero (es. indirizzo per esteso);
14. ore e minuti;
15. domanda virtuale generica priva di testo associato;
16. matrice di risposte multiple (checkbox);
17. selezione stato estero di nascita;
18. inizio di domanda contenitore di altre domande;
19. fine di domanda contenitore di altre domande;
20. tabella di input numerici (interi);
21. inizio di domanda iteratore;
22. fine di domanda iteratore;
23. matrice di risposte con al massimo una risposta attiva per ogni colonna;
24. matrice di risposte con al massimo una risposta attiva per ogni riga;
25. domanda da non visualizzare a schermo con lista di risposte associate;
26. lista di risposte (menu a tendina) selettivo.

Ad ogni tipologia è stato associato un *template*, che eventualmente ne poteva contenere altri, ad esempio nel caso in cui si rendesse necessario l'esecuzione di codice *JavaScript* per il controllo di coerenza di risposte e vincoli associati alla stessa pagina web, cioè non deducibili da precedenti salvataggi sul database, ma dagli input dell'utente al momento, oppure nel caso in cui una stessa domanda richiedeva più risposte di tipologie diverse, come *checkbox* e caselle di testo ad esse associate.

Inoltre il *layout* della pagina veniva costruito dinamicamente, oltre che in base al tipo di domanda, anche in base al contenuto: numero di risposte, lunghezza e numero di testi, formattazioni particolari.

Alcune domande in QCIS prevedevano la compilazione di matrici di input numerici (Figura 2.13).

Ad esempio veniva richiesto all'utente di indicare valori percentuali, per cui il *template* doveva essere in grado di agganciare alle caselle di testo uno script per calcolare la somma dei valori immessi, in modo da verificare il totale di tali valori corrispondesse al 100 per cento. Il calcolo doveva essere immediato, quindi effettuato con *JavaScript* e legato agli eventi di input dell'utente.

Inoltre, tali somme potevano essere fatte per riga, per colonna, ma anche per cella, cioè l'incrocio tra riga e colonna, che doveva essere coerente con gli altri totali.

Per fare ciò è stato costruito un *template* che generava una matrice di caselle di testo, le quali consentivano l'inserimento solo di caratteri numerici. Ogni casella era identificata univocamente dal codice della domanda, concatenato ad un indice di riga e uno di colonna.

Il codice HTML che generava la matrice veniva iterato a seconda del numero di risposte possibili.

Il calcolo dei totali veniva agganciato all'evento di modifica delle caselle e se si cercava di salvare dei valori che non corrispondevano ad un corretto valore percentuale, la modifica veniva bloccata attraverso un messaggio di errore che compariva vicino alla casella non coerente.

Figura 2.13 – Esempio di tabella di input numerici

6.12. Indicare, con riferimento al 2011, l'incidenza percentuale degli eventuali crediti commerciali vantati verso i seguenti soggetti

▲ Famiglie	<input type="text"/>
● Imprese	50
○ Ministeri	<input type="text"/>
○ Comuni	30
■ Asl e ospedali	<input type="text"/>
▲ Altre amministrazioni pubbliche	10
○ Enti/imprese/famiglie estere	<input type="text"/>
▲ Altro	<input type="text"/>
<b>Totale</b>	

La somma delle percentuali dei crediti commerciali è diversa da 100. Verifica la correttezza dei valori inseriti.

Questionario Industria e Servizi - Pag 15/18 Errori da correggere: 1 Successivo Indietro | Avanti | Salva | Riepilogo | Esci

In caso di una matrice di risposte multiple (Figura 2.14) è stato necessario l'utilizzo di un *template* personalizzato che costruisce dinamicamente una matrice di *checkbox*.

Grazie alle proprietà dell'oggetto Java che modella le domande sul database è stato possibile definire dei gruppi di compatibilità per le risposte, sono stati associati, cioè, degli identificativi per ogni gruppo di *checkbox* che prevedeva la stessa risposta per domande diverse.

In questo modo risposte non coerenti venivano segnalate all'utente direttamente al momento della compilazione, bloccando qualsiasi salvataggio di errori.

Figura 2.14 – Esempio di matrice di risposte multiple

3.03. Indicare, per ciascuna tipologia di relazione, le funzioni per le quali l'impresa intrattiene rapporti con altri soggetti (sono possibili più risposte)

\* Hai indicato la tipologia di relazione "Altro" nel quesito 3.03, ma non nel quesito 3.01. Verifica la correttezza dei dati inseriti.

	1 Attività principale	2 Ricerca e sviluppo, innovazione, progettazione	3 Tecnologie informatiche (ICT)	4 Approvvigionamento	5 Distribuzione, trasporto e magazzino	6 Marketing, vendita e servizi post-vendita	7 Servizi legali e/o finanziari	8 Altro
A. Commessa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
B. Sottofornitura/subappalto	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
C. Consorzio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
D. Contratto di rete	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
E. Altri accordi formali (joint venture, ATI, ecc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
F. Franchising	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
G. Accordo informale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
H. Altro	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Questionario Industria e Servizi - Pag 6/18 Errori da correggere: 1 Successivo Indietro | Avanti | Salva | Riepilogo | Esci

Un'altra tipologia è la matrice di risposte con al massimo una risposta attiva per ogni colonna/riga (Figura 2.15).

Come per il *template* precedente, sono state costruite matrici di *checkbox*, ma in questo caso con l'aggiunta di vincoli sul numero di risposte: non più di una per riga o per colonna.

Per fare ciò è stato agganciato del codice *JavaScript* all'identificativo delle risposte, formato dal codice della domanda e dagli indici di riga e colonna.

Alla selezione dell'utente questo tipo di controllo faceva scattare la deselezionare automatica di tutte le altre *checkbox* presenti sulla stessa riga/colonna, simulando così il funzionamento dei *radio button*. È stato scelto di non utilizzare direttamente i semplici radio HTML perché non avrebbero permesso l'inserimento di più di un valore per lo stesso input, mentre in questo caso l'oggetto risposta doveva essere modellato da una sequenza di valori.

**Figura 2.15 – Esempio di matrice di risposte con al massimo una risposta attiva per colonna/riga**

**1.04.X1. Indicare, per ciascuna delle seguenti funzioni aziendali, il grado di centralizzazione delle decisioni strategiche**  
(sono possibili più risposte, massimo una risposta per funzione aziendale)

	Grado di accentrimento decisionale			
	1 Alto	2 Medio	3 Basso	4 Nessuno
A Attività principale	1 <input checked="" type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
<b>Funzioni di supporto</b>				
B Approvvigionamento	1 <input checked="" type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
C Marketing, vendita e servizi post-vendita	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input checked="" type="checkbox"/>
D Distribuzione, trasporto e magazzinaggio	1 <input checked="" type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
E Amministrazione/finanza	1 <input checked="" type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
F Ricerca e sviluppo, innovazione, progettazione	1 <input type="checkbox"/>	2 <input type="checkbox"/>	3 <input checked="" type="checkbox"/>	4 <input type="checkbox"/>
G Tecnologie informatiche (ICT)	1 <input type="checkbox"/>	2 <input checked="" type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>
H Gestione e formazione delle risorse umane	1 <input checked="" type="checkbox"/>	2 <input type="checkbox"/>	3 <input type="checkbox"/>	4 <input type="checkbox"/>

### 2.2.3 L'implementazione del grafo

Al fine di facilitare la compilazione del questionario web è stato realizzato un sistema in grado di generare dinamicamente un percorso guidato in base alle risposte date. Le relazioni tra le varie domande sono state rappresentate tramite un grafo di navigazione i cui nodi sono le singole domande e gli archi, orientati, rappresentano i possibili percorsi di compilazione. Ciascun arco è etichettato con una condizione che deve essere verificata affinché sia possibile la transizione da un nodo all'altro.

A ogni risposta data, il grafo viene a trovarsi in un nuovo stato che si riflette nell'attivazione o disattivazione di determinati percorsi di compilazione in base alle condizioni che risultano verificate al momento.

Lato client il grafo di navigazione è stato implementato tramite un *array* associativo avente come indici gli identificativi delle domande. Una procedura *JavaScript* prendeva in input questo *array* e apportava in maniera dinamica le dovute modifiche alle pagine web visualizzate. La libreria *jQuery* è stata utilizzata per l'accesso dinamico ai contenuti delle pagine.

Ogni elemento dell'*array* è un oggetto che rappresenta un nodo del grafo e ha i seguenti attributi:

- un *nome* che è l'identificativo della domanda corrispondente;
- un'*espressione booleana* che rappresenta la condizione da verificare;
- il nome del *nodo successivo*.

A ogni cambiamento di stato si attraversa il grafo aggiornando la vista del questionario.

L'aggiornamento della vista avviene:

- all'accesso di una pagina del questionario;
- alla selezione di una data risposta.

Nel primo caso si eseguiva una visita del grafo partendo dal nodo iniziale. Nel secondo caso invece il nodo di partenza era quello che corrispondeva alla domanda corrente.

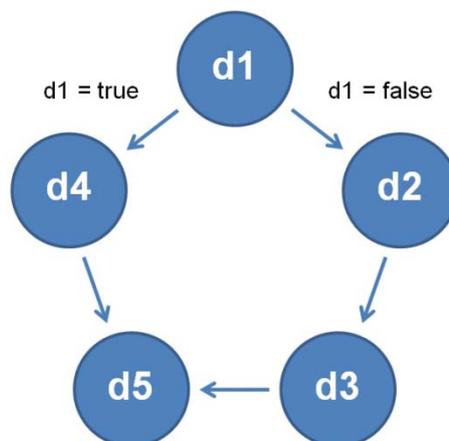
Un *array* ausiliario è stato utilizzato per memorizzare lo stato finale di ciascun nodo (attivato o disattivato).

Inizialmente tutti i nodi si trovavano nello stato disattivato. A ogni risposta data si procedeva inizialmente ad aggiornare le informazioni relative allo stato di tutti i nodi che si trovavano sul cammino determinato dalle condizioni associate agli archi a partire dal nodo corrente (quello corrispondente alla risposta data).

Quindi si scorreva l'*array* degli stati così aggiornato e si attivavano o disattivavano di conseguenza gli elementi della pagina del questionario.

Ad esempio, supponendo di avere cinque nodi d1, d2, d3, d4 e d5 connessi come nella Figura 2.16, a partire dal nodo d1 sono possibili due percorsi verso d5 in base al valore assunto dalla condizione associata a d1. Se assume valore *true* il percorso passa per i nodi d1, d4 e d5, altrimenti il percorso passa per i nodi d1, d2, d3, d5.

Figura 2.16 – Esempio di grafo di questionario



Inizialmente  $stato[d]$  è uguale a *false* per ogni nodo  $d=\{d1, d2, d3, d4, d5\}$ .

Se in base alla risposta data  $d1$  risulta *true* allora  $stato[d]$  sarà *true* per  $d=\{d1, d4, d5\}$  e *false* per  $d=\{d2, d3\}$ .

Se invece  $d1$  risulta *false* allora  $stato[d]$  sarà *true* per  $d=\{d1, d2, d3, d5\}$  e *false* per  $d=\{d4\}$ .

A questo punto iterando sull'*array* di  $stato$  è possibile procedere all'attivazione e disattivazione degli elementi della pagina.

Come per le regole di validazione, anche le specifiche del grafo sono state memorizzate all'interno di un database e le strutture software utilizzate lato client sono state generate dinamicamente lato server.

### 2.2.4 I controlli (client e server)

Per la validazione lato client dei dati inseriti nel questionario web, è stato realizzato un sistema di verifica in *JavaScript*, facendo uso della libreria *jQuery* per l'accesso dinamico ai contenuti del documento web.

Il sistema ha consentito di associare regole di validazione ai campi di input di una *form* in maniera non invasiva; l'unica richiesta era quella di attribuire ai campi da validare la classe "*validate*".

Il cuore del sistema è costituito da un motore di validazione che esponeva un insieme di funzioni predefinite per le verifiche più comuni (campi obbligatori, numerici, data, e-mail, eccetera) e al quale potevano essere date in input le regole di validazione specifiche da applicare, scritte secondo una precisa sintassi.

È stato possibile distinguere tra due tipi di regole: *hard* e *soft*.

Le *regole hard* corrispondono a vincoli stretti che devono essere rispettati in fase di inserimento dei dati: la mancata osservazione di tali vincoli impedisce il completamento della compilazione. Nella Figura 2.17 è mostrato un esempio di segnalazione di errore *hard*.

Figura 2.17 – Esempio di segnalazione di errore *hard*

1 PROPRIETÀ, CONTROLLO, GESTIONE

1.01. Indicare la tipologia, la nazionalità e la quota di capitale dei tre principali soci/proprietari al 31/12/2011

**Primo socio**

\* Non hai indicato la tipologia del primo socio

[Selezionare] ▼

Italiana ▼

Quota di capitale detenuta

**Secondo socio**

Per: \*

\* Hai indicato la tipologia del secondo socio, ma non la nazionalità

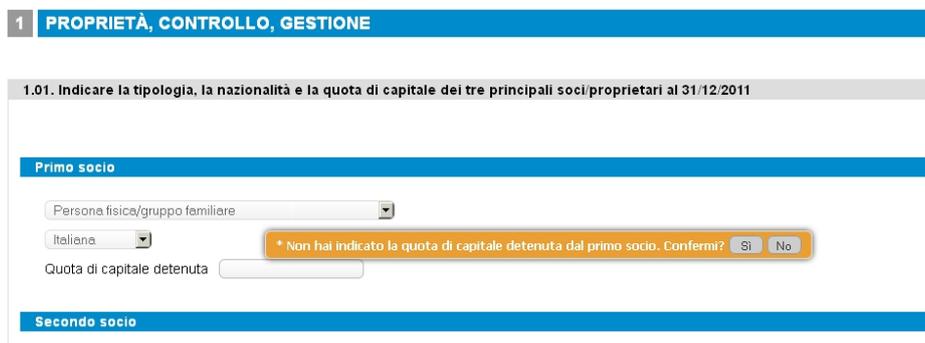
[Selezionare] ▼

Quota di capitale detenuta

I campi "tipologia di socio" e "nazionalità del socio" erano due campi obbligatori: la mancata risposta non consentiva di procedere nella compilazione.

Le *regole soft* invece sono meno stringenti e possono essere violate: il sistema comunque ne rilevava la violazione e chiedeva semplicemente una conferma della mancata osservanza. Nella Figura 2.18 è mostrato un esempio di segnalazione di errore *soft*.

Figura 2.18 – Esempio di segnalazione di errore soft



Non è stata riportata la quota di capitale detenuta dal socio, ma la risposta non era vincolante ai fini della continuazione della compilazione.

Ogni regola di validazione era delimitata da una coppia di parentesi quadre. Nel caso più semplice veniva richiamata una sola funzione: ad esempio con la regola `[required]` si diceva al motore di validazione di verificare che un campo obbligatorio fosse stato valorizzato.

Era possibile chiamare più funzioni da eseguire in cascata: le singole funzioni erano separate da una virgola. Ad esempio la regola `[required, date]` diceva al motore di validazione di verificare innanzitutto che il campo obbligatorio fosse stato valorizzato e quindi, se valorizzato, di verificare che il dato inserito fosse una data valida.

Alcune funzioni richiedevano il passaggio di parametri come nel caso della regola `[length[3,10]]` che imponeva che una variabile di testo fosse di almeno tre caratteri e al massimo di dieci: i parametri seguivano immediatamente il nome della funzione racchiusi tra parentesi quadre e separati da virgola.

Le regole potevano essere condizionate.

Ad esempio la regola `[if[d141r1c1 > d1311r1c1],noMatch]` diceva al motore di validazione di rilevare la non corrispondenza dei valori inseriti nei campi "d141r1c1" e "d1311r1c1" solo se il primo era maggiore del secondo.

Le condizioni sono espressioni booleane in cui possono comparire operatori di confronto, costanti e nomi di campi di input, da validare attraverso il motore di validazione.

Il sistema ha messo a disposizione una funzione `call` che consentiva di estendere l'insieme di funzioni predefinite del motore di validazione con nuove funzioni. Un caso in cui questo si sarebbe potuto rendere necessario era quello in cui il vincolo da applicare era particolarmente complesso da non potersi esprimere mediante una espressione booleana. Allo scopo era sufficiente definire una funzione che restituisse `true` o `false` e farla richiamare dal motore di validazione passandola come parametro alla funzione predefinita `call`.

Ad esempio la regola `[call[verificaModificaIndirizzo]]` diceva al motore di validazione di chiamare la funzione `verificaModificaIndirizzo`.

La definizione delle regole da dare in input al motore di validazione consisteva nella creazione di oggetti JSON secondo una sintassi specifica in cui ogni oggetto JSON ha tre campi:

- *regola*: definisce il vincolo che deve essere soddisfatto;
- *etichettaCampo*: è il nome che nel messaggio di errore individua il campo sottoposto a validazione;

- *testo*: è il testo, diverso da quello predefinito, da mostrare nel messaggio di errore.

Al momento dell'invio della *form* il motore di validazione verificava le regole definite iniziando con le regole *hard*. Solo se tutte le regole *hard* erano soddisfatte, si passava a verificare le regole *soft*. È da osservare che era possibile definire anche solo regole *hard* o solo regole *soft*.

Le regole, essendo parte dei metadati del questionario, sono state memorizzate nel database, al fine di rendere più agevole eventuali aggiunte o variazioni senza la necessità di modifiche al software. All'accesso ad ogni pagina web le regole di validazione specifiche per quella pagina erano ottenute tramite richiesta AJAX sotto forma di un oggetto JSON con le regole estratte dal database.

### 2.2.5 L'accessibilità

La norma che regola gli obblighi delle amministrazioni pubbliche in tema di accessibilità dei servizi da queste messe a disposizione dei cittadini tramite strumenti telematici era, al momento di realizzare il software, la L. 4/2004, "Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici", nota come "Legge Stanca". In particolare l'art. 4 ("Obblighi per l'accessibilità") definisce al comma 2 gli obblighi relativi alla fornitura di servizi su internet, stabilendo che una pubblica amministrazione nel realizzare un sito o applicazione web deve prevedere il rispetto dei requisiti di accessibilità stabiliti dalla stessa legge all'art. 11.

L'insieme di tali requisiti tecnici è specificato in un documento di linee guida ispirate da analoghe raccomandazioni in materia definite dal W3C, il consorzio internazionale che si occupa di definire gli standard relativi al web. L'insieme di queste raccomandazioni prende il nome di *Web Content Accessibility Guidelines* (WCAG) ed è stato pubblicato in una prima versione nel 1999. La rapida evoluzione tecnologica sia nel campo del web che in quello delle tecnologie assistive ha portato il W3C a definire un sostanziale aggiornamento delle WCAG, la cui versione 2.0 è stata rilasciata nel 2008.

Un documento del 2010 del Ministero della Funzione Pubblica<sup>10</sup> recepisce la seconda versione delle WCAG, aggiornando i requisiti tecnici per l'accessibilità in modo da ricalcare esattamente le linee guida WCAG 2.0. Tale documento, sostituendo l'allegato originale della L. 4/2004, costituisce pertanto la normativa attualmente vigente in materia. Gli elementi significativi dell'applicazione dal punto di vista dell'accessibilità sono i seguenti:

- tutte le pagine hanno superato la validazione di accessibilità (*Wave toolbar*);
- il sito è interamente navigabile attraverso la tastiera, anche per le funzioni di richiesta di conferma dei valori inseriti;
- gli elementi delle *form* sono stati strutturati in modo tale che, quando un controllo di una *form* era evidenziato, uno *screen reader* poteva leggere in modo significativo in sequenza sia le domande che le modalità delle risposte, per tutte le tipologie di domande, comprese quelle tabellari;
- considerato che gli errori di compilazione venivano visualizzati dinamicamente sulla pagina in corrispondenza delle domande, e pertanto non letti dagli *screen reader*, è stata introdotta una "modalità accessibile", attivabile da una voce di menu nella pagina principale del questionario, nella quale gli errori di compilazione venivano presentati in una sequenza significativa in pop-up modali che venivano segnalati e letti dagli *screen reader*.

<sup>10</sup> Criteri e metodi per la verifica tecnica e requisiti tecnici di accessibilità per i contenuti e i servizi forniti per mezzo di applicazioni basate su tecnologie Web (Revisione dell'allegato A del D.M. 8 luglio 2005) - versione 26 aprile 2010.

Il test di accessibilità è stato effettuato attraverso i seguenti metodi:

- validazione dell'HTML attraverso lo strumento *Wave toolbar* disponibile sul sito <http://wave.webaim.org/>;
- lettura di tutte le pagine mediante lo *screen reader open source* NVDA (disponibile sul sito <http://www.nvda-project.org/>);
- revisione di tutti i criteri di accessibilità corrispondenti al livello A.



## Capitolo 3

### Il sistema di gestione<sup>11</sup>

La gestione contemporanea di più indagini comporta per la rete di rilevazione un grosso sforzo in termini di organizzazione e di monitoraggio; gli Uffici Provinciali di Censimento (UPC) si sono trovati ad affrontare, in un tempo relativamente breve, la gestione di due indagini, diverse in termini di contenuto informativo e di lavorazione. Per poter coadiuvare e supportare i suddetti organi è stata messa a loro disposizione un'applicazione web ad accesso protetto che ha permesso la gestione integrata di tutte le indagini, dalla fase di definizione degli operatori fino alle operazioni di chiusura con la conseguente tracciatura della consegna dei materiali. Gli Uffici Provinciali di Censimento, inoltre, sono stati coinvolti nell'intero processo di produzione del dato ed hanno partecipato in modo attivo alle fasi di controllo e correzione dei questionari. Questo sistema, denominato "Sistema di Gestione della Rilevazione" (SGRCIS) è stato sviluppato e utilizzato sia per il Censimento dell'agricoltura che per quello della popolazione, procurando notevoli benefici in termini di rapidità di diffusione dei dati e facilità di gestione dell'indagine. Per il Censimento dell'industria e dei servizi è stato quindi possibile riutilizzare tale sistema, sviluppando funzioni *ad hoc* per le esigenze specifiche delle nuove indagini, ma soprattutto customizzando quelle esistenti. Il sistema è stato sviluppato integrando le indagini in un unico cruscotto telematico pur conservando la gestione separata delle indagini. Di seguito sono illustrate in modo approfondito le principali funzionalità del sistema.

#### 3.1 La comunicazione con la rete

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

La struttura delle pagine del portale (Figura 3.1) era composta da due menu orizzontali posizionati nella parte alta e da un'area dei contenuti posta in primo piano.

<sup>11</sup> A cura di Maura Giacommo. Autori del capitolo:

- Paragrafo 3.1: Rossana Vignola e Daniele Ingretolli;
- Paragrafo 3.2: Eleonora Sibilio;
- Paragrafo 3.3: Maura Giacommo (Paragrafo 3.3.1), Giulia Vaste (Paragrafo 3.3.2);
- Paragrafo 3.4: Linda Corradi e Paola Giorgetti (Paragrafo 3.4.1), Rolando Duma (Paragrafo 3.4.2);
- Paragrafo 3.5: Manuela Marrone (Paragrafo 3.5.1), Enrico Orsini (Paragrafo 3.5.2), Annalisa Cesaro (Paragrafi 3.5.3 e 3.5.5), Manlio Paterniti (Paragrafo 3.5.4);
- Paragrafo 3.6: Lucia Toti (Paragrafo 3.6.1), Guido Drovandi (Paragrafo 3.6.2), Marco Amato (Paragrafo 3.6.3), Giulia Vaste (Paragrafo 3.6.4);
- Paragrafo 3.7: Eleonora Sibilio e Simonetta Feliciotti;
- Paragrafo 3.8: Cristiano Francone;
- Paragrafo 3.9: Rolando Duma e Lucia Toti.

Figura 3.1 – Homepage del portale



**CENSIMENTO DELL'INDUSTRIA E DEI SERVIZI 2011**

**Istat.it**

---

Home Documenti Strumenti per la rilevazione Domande&Risposte Glossario **SGR**

Formazione Contatti Mappa del Sito Cerca nel sito Logout

---

**Sei in:** Home

09.07.2013  
**Presentazione dei primi risultati alla stampa**  
 L'11 luglio 2013 alle ore 10.30, presso la Sala Polifunzionale della Presidenza del Consiglio dei Ministri, saranno presentati alla stampa i primi risultati del 9° Censimento generale dell'industria e dei servizi e Censimento... **[leggi]**  
 Categoria: **Operazioni Censuarie**

---

09.07.2013  
**Contestazioni dell'obbligo di risposta**  
 Le unità di rilevazione che hanno ricevuto una lettera di contestazione dall'Istat possono chiedere spiegazioni sull'iter amministrativo scrivendo una e-mail alla casella dedicata contestazionici@istat.it o contattando il numero... **[leggi]**  
 Categoria: **Operazioni Censuarie**

---

09.07.2013  
**Accertamenti**  
 Si sollecitano i Responsabili UPC a completare le attività di accertamento della violazione dell'obbligo di fornire dati statistici e l'invio all'Istat delle relative documentazioni. Alla data odierna sono 6246 i casi risultanti... **[leggi]**  
 Categoria: **Operazioni Censuarie**

---

09.07.2013  
**Chiusura Censimento ed erogazione del saldo del contributo variabile**  
 Il 28 giugno 2013 è stato completato l'invio ad Istat del Report Chiusura e del Report Pesi da parte di tutti gli UPC. L'Istat sta ultimando i controlli e le verifiche delle informazioni in essi contenute in modo da calcolare i... **[leggi]**  
 Categoria: **Operazioni Censuarie**

---

09.07.2013  
**Comunicazione dei dati**  
 Agli enti del Sistan la comunicazione dei dati censuari è effettuata dall'Istat nel rispetto dei principi stabiliti dalla vigente normativa in materia di riservatezza, in conformità dell'articolo 50 del decreto legge 31 maggio... **[leggi]**  
 Categoria: **Altre comunicazioni**

---

**Archivio News**

---

**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 poteva navigare tra le varie aree del portale (Home, Documenti, Strumenti per la rilevazione, Domande&Risposte, Glossario, SGRCIS).

Nella homepage erano presenti delle *news*, aggiornate costantemente, per tenere informati gli utenti riguardo l'andamento della operazioni censuarie, il rilascio di strumenti utili alla rilevazione, la pubblicazione di circolari o di avvisi e comunicazioni varie; in fondo alla pagina era presente un link che rimandava all'archivio contenente tutte le *news* pubblicate.

L'area *Documenti* raccoglieva la documentazione di riferimento del censimento; oltre al Piano Generale di Censimento, articolato in paragrafi navigabili, nelle altre sottosezioni erano disponibili le comunicazioni ufficiali dell'Istat, le circolari, le principali norme nazionali e comunitarie di interesse per il censimento, le *Newsletter* periodiche e i documenti connessi ad alcuni aspetti di progettazione tecnica o ai contenuti delle rilevazioni.

Nell'area *Strumenti* erano contenuti i materiali che gli operatori censuari usavano per raccogliere i dati e altri materiali utili per lo svolgimento delle attività. Erano infatti presenti i fac-simile dei materiali contenuti nel plico di rilevazione (lettere informative, questionari e guide alla

compilazione), altri strumenti di aiuto per la compilazione e infine, il manuale della rilevazione navigabile.

In *Domande&Risposte* erano contenute alcune domande frequenti sul censimento e le relative risposte. Per facilitarne la ricerca le domande erano organizzate in categorie e sottocategorie.

Nel *Glossario* erano riportate le definizioni delle principali variabili del censimento.

Nell'area SGRCIS era presente un pulsante che attivava il link al Sistema di Gestione della Rilevazione.

Il menu secondario (*Formazione, Contatti, Mappa del Sito, Cerca nel sito, Logout*) invece conteneva una serie di link utili. L'area *Formazione* conteneva i documenti a supporto dell'attività di formazione realizzata dagli organi della rete di rilevazione sul territorio come ad esempio le linee guida, i consigli per gli istruttori e le presentazioni in *PowerPoint*.

Nell'area *Contatti* erano presenti i numeri di telefono e gli indirizzi di posta elettronica dei Responsabili degli Uffici Provinciali di Censimento e dei referenti territoriali Istat.

### 3.2 Ruoli e funzioni della rete di rilevazione nel sistema di gestione

Per collegarsi a SGRCIS l'utente doveva essere definito all'interno del sistema e disporre delle credenziali di accesso. A ogni utente era associato un ruolo e un profilo applicativo, in base al quale il sistema veniva dinamicamente personalizzato rispetto ai seguenti aspetti:

- rilevazioni sulle quali si aveva la possibilità di operare;
- insieme di funzionalità accessibili all'interno di ciascuna rilevazione;
- visibilità territoriale dei dati.

Tale controllo degli accessi ha permesso ad ogni operatore di accedere alle sole funzioni di competenza del suo ruolo/profilo e di effettuare tagli orizzontali in fase di interrogazione del DBMS in base alla visibilità territoriale dell'operatore.

I ruoli gestiti sono stati i seguenti:

- 1) *Istat nazionale*: ha svolto funzioni di monitoraggio dell'andamento delle rilevazioni a livello nazionale. Gli utenti con tale ruolo potevano accedere alle funzionalità di monitoraggio e di utilità generale;
- 2) *Ufficio di statistica Unioncamere*: ha avuto a disposizione le stesse funzionalità e la stessa visibilità territoriale dei dati riconosciute a Istat nazionale;
- 3) *Responsabile Ufficio Regionale di Censimento (URC)*: ha avuto la possibilità di creare/modificare i responsabili Istat territoriali e ha avuto visibilità sulla regione di pertinenza oltre a svolgere funzioni di monitoraggio;
- 4) *Responsabile Istat Territoriale (RIT)*: creati dagli URC o precaricati, potevano avere una visibilità regionale o provinciale, corrispondente agli UPC di competenza territoriale, anche questo ruolo ha avuto il compito di monitorare l'andamento della rilevazione;
- 5) *Responsabile Ufficio Provinciale di Censimento (UPC)*: si occupava di monitorare le rilevazioni sul campo in ogni loro fase. Più in particolare, questa tipologia di utente ha avuto, ad esempio, il compito di creare la rete di rilevazione, di monitorare il lavoro dei rilevatori e dei relativi coordinatori, di validare il lavoro di *data entry* effettuato dai rilevatori. La visibilità territoriale dei dati era possibile solo per l'ambito territoriale di pertinenza;

- 6) *Coordinatore Provinciale (COP)*: ha avuto la possibilità di creare/modificare i profili relativi ai rilevatori, assegnare ad essi le unità di pertinenza, monitorare il lavoro dei rilevatori assegnati. La visibilità territoriale dei dati era legata alle unità assegnate ai rilevatori di cui il COP era responsabile;
- 7) *Rilevatore*: si occupava del lavoro sul campo e/o di attività di *back office*. La visibilità territoriale variava in funzione delle mansioni che svolgeva.

Per la gestione della multi indagine e per agevolare le diverse realtà organizzative i ruoli di Coordinatore, Responsabile Istat Territoriale e Rilevatore sono stati ulteriormente affinati. In particolare, i coordinatori sono stati suddivisi in base alle rilevazioni su cui potevano lavorare, mentre i RIT sono stati diversificati rispetto alla loro visibilità territoriale accedendo comunque sempre ad ambedue le rilevazioni. Per i rilevatori è stata prevista un'analoga suddivisione in base alle rilevazioni, come per i coordinatori, con una ulteriore ripartizione in base alla tipologia di lavoro svolto:

- *Rilevatore*: svolgeva il "lavoro sul campo", in particolare gestiva l'intero flusso di lavorazione delle unità di rilevazione che gli venivano assegnate (registrazione dell'esito della rilevazione, *data entry*, controllo dei microdati attraverso il *check*). Questo profilo aveva visibilità esclusivamente sui dati relativi alle unità che gli erano state assegnate;
- *Operatore back office*: gestiva le attività all'interno dell'Ufficio di censimento. Era abilitato soltanto alla registrazione dell'arrivo dei questionari cartacei e alla registrazione dei plichi inesitati da parte del vettore postale oltre che alle funzioni di ausilio alla rilevazione. Questo profilo aveva visibilità sull'intero Ufficio Provinciale di Censimento.

Tutte le funzioni del sistema che si occupavano della creazione della rete sono state inserite nella pagina "scelta attività" presente subito dopo la pagina di autenticazione dell'utente (Figure 3.2 e 3.3); tali funzioni consentivano di creare in modo dinamico ed indipendente la propria rete di rilevazione assegnando, gestendo o modificando le figure operative facenti parte dell'Ufficio di censimento pertinente.

Figura 3.2 – Creazione rete



Le funzionalità offerte sono state le seguenti:

- 1) *Variazione UPC*: permetteva di modificare alcuni dati dell'Ufficio Provinciale di Censimento, questa funzione era a disposizione del solo responsabile e permetteva la variazione dei dati del responsabile senza dover dare alcuna comunicazione cartacea all'Istat;

- 2) *Gestione COP*: prevedeva funzionalità quali “*Inserisci*”, per registrare un nuovo operatore con profilo applicativo COP e di “*Modifica/Cessazione*”, per modificarne o eliminarne uno già esistente;
- 3) *Gestione Rilevatori*: anche per questo requisito erano previste funzionalità quali “*Inserisci*” e “*Modifica/Cessazione*”;
- 4) *Assegnazione dei Rilevatori ai COP*: le operazioni di assegnazione erano indispensabili per permettere un corretto andamento della rilevazione e un attento monitoraggio. Questa funzione aveva lo scopo di assegnare al Coordinatore Provinciale i rilevatori di competenza, creando una struttura gerarchica della rete, modificabile in un qualunque momento a secondo delle esigenze della rilevazione;
- 5) *Ricerca Operatori*: questa funzione permetteva all’operatore di effettuare una ricerca degli operatori facenti parte della rete di rilevazione dell’ufficio provinciale di competenza.

I responsabili per consentire ai rilevatori di lavorare dovevano assegnare loro delle unità. L’insieme delle funzioni che collegavano e gestivano tale aspetto erano caratteristiche della rilevazione e pertanto si trovavano all’interno della gestione delle singole indagini. Le funzioni inoltre, a parte poche eccezioni, erano analoghe per le due rilevazioni.

Figura 3.3 – Gestione rete



L’area denominata “Gestione rete” racchiudeva le seguenti funzioni:

- *Assegnazioni Unità ai Rilevatori*: aveva lo scopo di assegnare a ciascun rilevatore una o più unità di rilevazione dell’ufficio provinciale di competenza. Per facilitare l’utilizzo di tale funzione ed evitare che il lavoro di assegnazione fosse troppo oneroso sono state previste delle funzionalità quali la selezione di interi blocchi di record, l’ordinamento dei campi e l’impostazione di filtri opzionali per la ricerca; la Figura 3.4 mostra un esempio di interfaccia per le assegnazioni;
- *Ricerca Operatore*: permetteva all’operatore di effettuare una ricerca degli operatori facenti parte della rete di rilevazione;
- *Composizione Rete*: permetteva all’operatore di visualizzare la composizione rete di rilevazione dell’ufficio provinciale di competenza;
- *Elenco Unità Assegnate*: permetteva all’operatore di visualizzare l’elenco delle unità di rilevazione assegnate ai rilevatori.

Figura 3.4 – Assegnazioni unità

Gestione assegnazione unità al rilevatore: 2001178 – Ilaria Di Luigi

Comune:  CAP:  Denominazione Unità:

Questionario:  Modalità d'Arrivo:

Tipo Unità:  Assegnate al Rilevatore  Da assegnare  De-assegnabili

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

Cod. Quest.	Denom. Unità	Codice Fiscale	Comune	CAP	Ind. da lista	Rilevatore	Responsabile	Nuovo Indirizzo
0647076	A C P S GLI ANIMALI I NOSTRI TESORI	01953350681	MONTE SILVANO	65015	VIA BARI 7	di luigi ilaria	galante selene	Invariato
0647973	HORIZON II		PESCARA	65122	CORSO UMBERTO I 94	di luigi ilaria	galante selene	Invariato
0647978	HORIZON 2 SOCIETA' COOPERATIVA SOCIALE	01495710681	PESCARA	65122	CORSO UMBERTO I 94	di luigi ilaria	galante selene	Invariato
0648024	ISTITUTO DEL NASTRO AZZURRO	91001470680	PESCARA	65122	PIAZZA SANTA CATERINA DA SIENA 4	di luigi ilaria	galante selene	Invariato

### 3.3 Il workflow di lavorazione e la gestione dei ricicli

Quasi tutte le funzioni del sistema sono finalizzate alla gestione del questionario. Si è reso quindi necessario descrivere lo stesso in modo dettagliato per poi comprendere il funzionamento del sistema nel suo complesso. Inoltre particolare attenzione è stata data alla registrazione dell'esito del questionario in quanto tale funzione è stata di particolare rilevanza sia per la mole di informazioni che gestiva sia per il compenso agli organi di Censimento.

#### 3.3.1 Il flusso del questionario

Il Sistema di Gestione della Rilevazione, come è stato detto, doveva sia monitorare l'andamento della rilevazione che permettere la gestione del lavoro sul campo. Per far convergere entrambe le necessità è stato associato uno stato ad ogni questionario, che identificava in modo chiaro in che fase di lavorazione si trovava il questionario e le operazioni necessarie per giungere alla chiusura dello stesso. Un questionario veniva giudicato chiuso se il responsabile dell'UPC validava i risultati della fase di *check*; in questo caso lo stato di lavorazione veniva posto a "validato".

Il cambiamento dello stato e quindi l'avanzamento del flusso, inoltre, poteva avvenire in modo manuale oppure in modo automatico. Nel primo caso la modifica era legata ad operazioni che l'utente svolgeva tramite SGRCIS (come ad esempio la registrazione dell'esito della rilevazione, piuttosto che il controllo delle regole di *check*), mentre la variazione automatica era legata ad informazioni esterne che venivano catturate dal sistema (esempio il vettore postale che comunicava l'avvenuta consegna del questionario piuttosto che l'unità che compilava il questionario). Nel caso di gestioni manuali il sistema registrava per ogni operazione l'utente che l'aveva eseguita e l'orario di esecuzione. Il tutto è stato gestito come un sistema di *workflow* distribuito, nel quale ogni utente poteva muoversi in modo autonomo all'interno di un percorso ben delineato. In un'ottica di questo tipo è stato necessario gestire anche le eventuali eccezioni e i possibili errori che si potevano verificare durante la rilevazione sul campo; ogni operazione manuale poteva essere annullata e il questionario di conseguenza veniva riportato allo stato di lavorazione precedente. Solo la validazione finale del questionario da parte del responsabile dell'UPC bloccava in modo definitivo la lavorazione dello stesso. Alcune funzioni considerate di particolare rilievo, inoltre, potevano essere annullate dal solo responsabile dell'UPC oppure da Istat.

Il flusso di lavorazione del questionario è stato lo stesso per entrambe le rilevazioni, con la sola differenza che nel caso del censimento delle istituzioni non profit era possibile inserire una nuova unità, non presente nella lista di partenza ed è stato quindi necessario disegnare un flusso di lavorazione *ad hoc* per questa casistica. La nuova unità poteva essere trovata dal rilevatore durante il recupero sul campo delle mancate risposte, oppure l'unità stessa poteva segnalare la sua presenza all'Istat e/o all'UPC, in ogni caso era compito della rete di rilevazione inserire la nuova istituzione non profit all'interno del sistema.

Di seguito sarà illustrato l'intero flusso del questionario suddiviso per fasi di lavorazione.

La prima fase che il sistema ha monitorato è quella di consegna dei plichi (contenenti oltre ai questionari anche altri documenti a supporto) alle unità di rilevazione (imprese e istituzioni non profit). Entrambe le rilevazioni, infatti, sono state guidate da lista e i plichi sono stati spediti dal vettore postale all'indirizzo dell'unità rispondente. Le informazioni inerenti alla consegna del plico sono state comunicate dal vettore postale tramite file, con periodicità bisettimanale. Il vettore ha dato comunicazione solo sui plichi non consegnati (inesitati) e sulla fine delle operazioni di consegna presso un comune. La rete di rilevazione, in questo modo, ha potuto monitorare in modo tempestivo l'andamento della consegna dei plichi. Tuttavia, nonostante l'informazione acquisita dal vettore postale la rete di rilevazione doveva registrare sul sistema i plichi inesitati nel momento in cui gli stessi arrivavano fisicamente presso l'Ufficio Provinciale di Censimento.

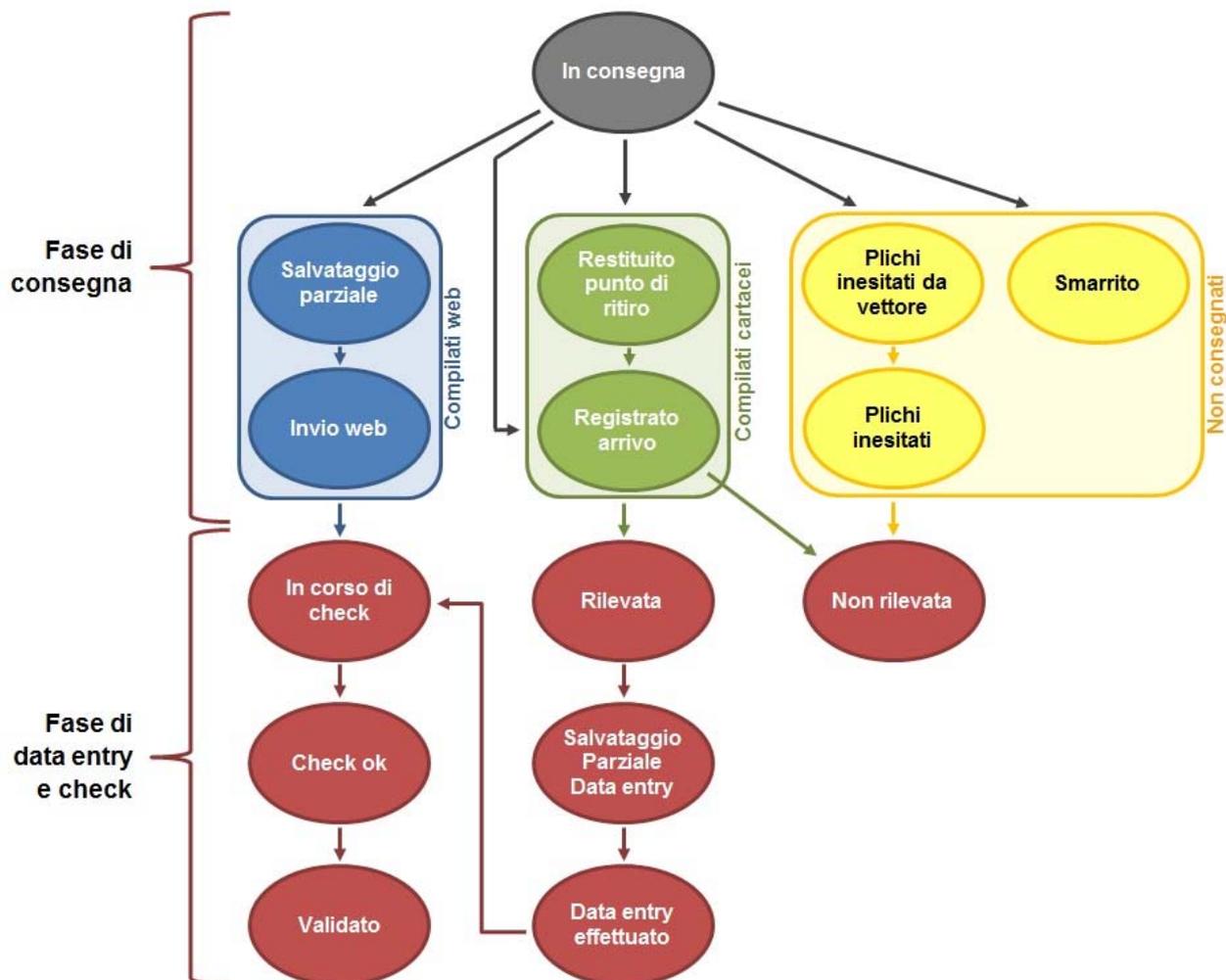
La seconda fase che il sistema ha monitorato è il recupero sul campo dei questionari che prevedeva la visita dei rilevatori presso le unità per le quali il questionario non risultava pervenuto in nessuno dei canali previsti per la restituzione. Si tenga presente che la tracciatura della restituzione è avvenuta in tempo reale per quanto riguarda il questionario online e in modo asincrono per la restituzione presso il vettore postale. In modo analogo a quanto fatto nella fase di consegna, infatti, anche in questo caso, il vettore postale ha inviato l'informazione circa la consegna del questionario cartaceo ai propri sportelli con cadenza bisettimanale, tramite file. I questionari cartacei venivano materialmente consegnati presso l'UPC successivamente. La rete, quindi, ha dovuto inserire anche l'avvenuta ricezione fisica del questionario presso la propria sede; questa operazione avveniva manualmente a cura di operatori dell'UPC, utilizzando l'apposita funzione del sistema.

Per ogni questionario arrivato in forma cartacea doveva essere registrato l'esito della rilevazione, in questa fase, come si vedrà nel paragrafo successivo, veniva inserita una grossa mole di dati, che servivano per diversi scopi, tra cui l'esito vero e proprio; solo nel caso in cui l'unità risultava "rilevata" si proseguiva la lavorazione mediante la registrazione del questionario: per le unità "non rilevate" il flusso terminava con la registrazione dell'esito e la validazione finale da parte del responsabile dell'UPC.

La terza e ultima fase di lavorazione, che interessa le sole unità attive, riguarda il *data entry* ed il *check* dei questionari; in questo contesto la rete di rilevazione ha dovuto inserire le informazioni contenute nei questionari cartacei all'interno del sistema e controllare che tutti i dati pervenuti rispettassero alcune regole di coerenza. Se quest'ultime non erano rispettate, il questionario veniva riaperto in modifica per essere nuovamente lavorato; se invece il processo terminava in modo corretto lo stato del questionario veniva posto a "Check OK" e poteva quindi essere validato dal responsabile UPC. La descrizione delle regole che sottendono il funzionamento del grafo è stata registrata in apposite strutture tabellari di tipo relazionale nel database.

Nella Figura 3.5 è riportato l'intero flusso di lavorazione

Figura 3.5 – Flusso di lavorazione del questionario



### 3.3.2 L'esito come elemento cardine della rilevazione

Come visto nel precedente paragrafo, un passaggio fondamentale nella lavorazione dei questionari cartacei<sup>12</sup> era la registrazione dell'esito della rilevazione: questa operazione ha consentito infatti, ancor prima di effettuare il *data entry*, di disporre di una panoramica tempestiva e puntuale dell'andamento della rilevazione.

La registrazione dell'esito della rilevazione aveva un duplice scopo: consentire di avere in anticipo le informazioni basilari sui questionari compilati restituiti in formato cartaceo, che solo in un secondo tempo sarebbero stati sottoposti a *data entry*, e al contempo tenere traccia di tutte le unità che non erano state rilevate, per quale motivo e con le relative informazioni. La registrazione dell'esito della rilevazione provocava quindi due possibili passaggi di stato del questionario: l'unità poteva risultare "rilevata" o "non rilevata".

<sup>12</sup> Nel caso di questionari telematici l'esito della rilevazione era attivato di default in base dal contenuto dei dati registrati.

Nel caso di unità rilevate, per le quali quindi si disponeva del cartaceo compilato, l'esito della rilevazione veniva ottenuto solo all'invio definitivo del questionario, in fase di *data entry* a cura degli operatori dell'UPC. I possibili esiti erano: "Rilevata Attiva", "Rilevata Cessata", "Rilevata Inattiva", "Rilevata Fuori campo di osservazione". La funzione di registrazione dell'esito in questo caso consentiva solo di indicare che l'unità era stata rilevata e, in caso di restituzione del cartaceo al rilevatore, di specificare se era stata condotta un'intervista diretta, da quale rilevatore e in quale data. Tali informazioni sono risultate fondamentali per il calcolo dei contribuiti agli organi di rilevazione.

Nel caso di unità non rilevate, per le quali non si disponeva del questionario, la funzione di registrazione dell'esito prevedeva le seguenti casistiche:

- "Inattiva";
- "Cessata/Ceduta";
- "Unità Irreperibile";
- "Fuori campo di osservazione";
- "Rifiuto";
- "Duplicata".

In base allo specifico esito della rilevazione, venivano richieste ulteriori informazioni di dettaglio, quali la data della cessazione/inattività, la modalità con cui era stato verificato l'esito della rilevazione, ovvero tramite verifica sul campo o controlli di *back office*, l'eventuale codice questionario del duplicato.

Le informazioni relative alla modalità con cui era stata effettuata la verifica dell'esito della rilevazione è stata tenuta in considerazione per il calcolo dei contribuiti agli organi di rilevazione.

I rilevatori e i coordinatori potevano utilizzare la funzione solo per questionari di propria competenza e il loro lavoro veniva tracciato dal sistema.

La funzione ha pertanto consentito la gestione delle unità non rilevate, per mancata consegna o per mancata risposta, che venivano verificate puntualmente dagli operatori della rete e che venivano tracciate nel sistema con le informazioni relative necessarie. La funzione di esito della rilevazione, pertanto, ha consentito la gestione le unità non rilevate, le unità rilevate tramite compilazione del questionario cartaceo e, in maniera automatica, le unità che hanno utilizzato la compilazione web.

### 3.4 Il controllo dei dati

#### 3.4.1 Il check dei microdati

Lo scopo della fase di controllo denominata "*Check*" è quello di appurare la correttezza dei questionari pervenuti dalla compilazione web o compilati tramite *data entry* dagli operatori dell'UPC. I controlli sono stati eseguiti tramite procedure offline dagli operatori abilitati della rete di rilevazione. Questi controlli di coerenza sono stati concepiti dalla struttura statistica responsabile dell'indagine che li ha forniti in forma descrittiva; a ogni regola da verificare veniva associato un identificativo, un messaggio di errore ed una tipologia di errore: *Warning* nel caso di un accertamento non bloccante (regola *soft*) oppure *Errore* nel caso di un controllo bloccante (regola *hard*).

Ogni indagine ha avuto il proprio set di regole; per il Censimento delle istituzioni non profit 10 *Errori* e 79 *Warning* e per la rilevazione delle imprese 16 *Warning*.

Per la gestione delle regole di controllo sono state definite delle strutture tabellari dedicate e integrate nello schema dati SGRCIS. In esse sono state archiviate le regole tradotte in “linguaggio informatico” sotto forma di codice Oracle SQL, i relativi messaggi di errore ed infine la storicizzazione di tutte le esecuzioni della procedura di *Check* relative ad ogni questionario. Per ogni richiesta di *Check* veniva memorizzato il codice dell'operatore che aveva eseguito l'operazione, la data, l'esito di ogni singolo controllo ed un riepilogo del numero di *Errori* e/o *Warning*.

Una *Oracle Procedure*, cuore del *Check*, gestiva l'intero processo in maniera dinamica ricevendo in partenza la lista delle imprese o delle istituzioni non profit, selezionate dall'utente tramite l'applicazione web, da sottoporre a verifica, e fornendo come output l'esito dello stesso.

Ogni questionario sottoposto a verifica veniva bloccato per impedirne l'accesso in *data entry*; per ognuno venivano letti ed eseguiti tutti i controlli previsti secondo il flusso della compilazione del questionario, ad esclusione dei *Warning* eventualmente validati in un *check* precedentemente eseguito; veniva poi registrato ogni esito negativo delle regole. Alla fine della procedura per ogni questionario veniva calcolato il riepilogo del numero di *Errori* e/o *Warning* e veniva sbloccato l'accesso al *data entry* solo nel caso fossero state riscontrate delle anomalie.

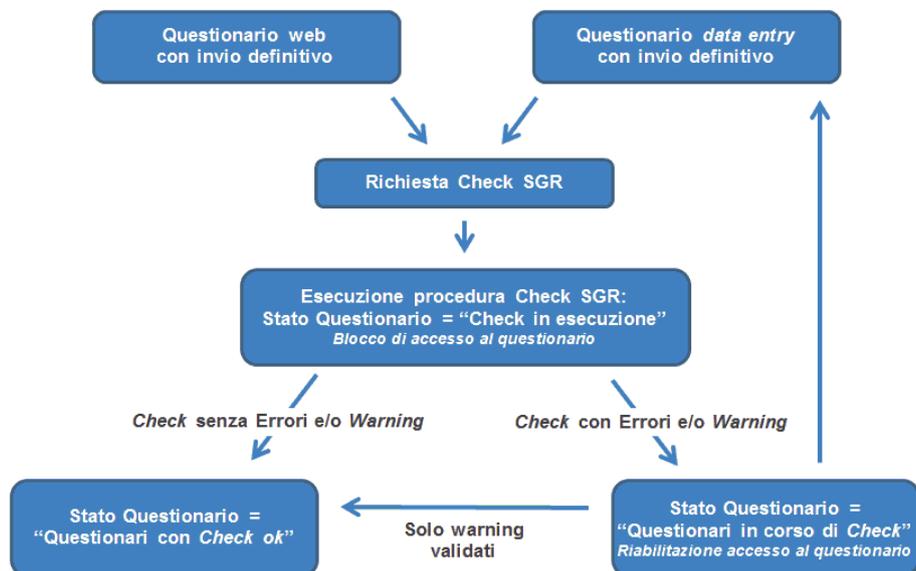
La registrazione dell'esito sulla base dati e la conseguente visualizzazione dell'andamento del *Check* sull'applicazione, per ogni singola regola, permetteva all'operatore la verifica dell'andamento del *Check* e la correzione degli *Errori* del questionario o la validazione a conferma del dato inserito o la correzione dei *Warning*. Una volta validati o corretti tutti i *Warning*, in assenza di *Errori bloccanti*, il questionario passava nello stato di “*Check OK*” proseguendo il suo percorso. Se c'erano *Errori bloccanti* questi dovevano essere corretti ed il *Check* doveva essere eseguito nuovamente fino alla risoluzione di ogni *Errore*.

La *Oracle Procedure* che gestiva le regole di controllo dei microdati era autonoma e completamente indipendente dalle regole di coerenza imposte ai questionari in quanto, esistendo come codice SQL memorizzato in una struttura tabellare, poteva essere modificato, implementato o cancellato, senza la necessità di apportare modifiche alla procedura generalizzata e all'applicativo web.

Grazie alla progettazione modulare il software è stato fortemente riusato in quanto “ereditato” da altre indagini censuarie precedenti; ne è, quindi, derivato un importante miglioramento delle performance in termini di giornate/uomo necessarie allo sviluppo delle funzionalità. Solo gli algoritmi delle regole sono stati riscritti sulla base alle esigenze specifiche del Censimento industria e servizi.

Attraverso la storicizzazione dei *Check* eseguiti per ogni impresa e/o istituzione non profit si sono ottenuti due importanti risultati: da un lato si preservavano le scelte dell'operatore relativamente alle validazioni dei *Warning* e dall'altro si poteva monitorare e verificare il lavoro svolto dai vari operatori della rete di rilevazione. Il flusso del *Check* è rappresentato nella Figura 3.6.

Figura 3.6 – Flusso Check SGRCIS



Il *Check* dei microdati poteva essere eseguito su questionari che risultavano inviati definitivamente e per gli utenti con profili di Rilevatore e Coordinatore per i soli questionari di loro competenza. La validazione dei *Warning* poteva essere effettuata da Rilevatori, Coordinatori e Responsabili dell'Ufficio Provinciale di Censimento. Gli altri profili accedevano alla funzione in sola visualizzazione.

La funzione di *Check* poteva essere applicata ad un singolo questionario oppure ad un insieme di questionari selezionati dall'utente mediante dei filtri (Figura 3.7).

L'utente indicava i filtri desiderati e utilizzando il bottone "ricerca" veniva indirizzato alla successiva interfaccia grafica dove poteva scegliere i singoli questionari oppure eseguire il *Check* su tutti i questionari ottenuti dalla ricerca.

Figura 3.7 – Ricerca questionari per l'esecuzione del Check SGRCIS

Era possibile impostare dei filtri di ricerca anche per la visualizzazione dell'esito del *Check* e dell'eventuale numero di *Errori* e/o *Warning* di ogni questionario. In caso di risultato positivo veniva mostrato l'esito "Eseguito correttamente senza Errori o Warning" e posto lo stato "Check OK" altrimenti l'esito era "Eseguito correttamente con Errori o Warning" e il questionario transitava nello stato "In corso di check".

Cliccando sul numero di *Errori* e/o *Warning* ne veniva visualizzato l'elenco (Figura 3.8). Gli *Errori* bloccavano il processo di elaborazione del questionario e dovevano essere corretti accedendo in modifica ai dati. I *Warning* rappresentavano, invece, delle segnalazioni di possibili anomalie non necessariamente bloccanti per il processo di lavorazione del questionario. L'operatore poteva quindi, effettuate le necessarie verifiche, confermare la correttezza dei dati inseriti validando i *Warning*, oppure procedere alla loro correzione.

Una volta che i *Warning* erano stati validati non venivano più mostrati nei successivi *Check* SGRCIS dello stesso questionario.

In presenza di soli *Warning* tutti validati dall'operatore, il questionario veniva posto automaticamente in stato "Check OK".

Figura 3.8 – Visualizzazione esito del Check SGRCIS e validazione Warning

The screenshot displays the 'Accertamenti questionario: 0300101' window. It features a table with the following columns: 'Regola', 'Warning', and 'Valida'. The table lists 12 items, each with a unique ID and a description of the warning. A summary box on the right indicates 'N° errori: 4' and 'N° warning: 10'. The interface also includes a search bar, a dropdown for 'Selezionare il comune', and a dropdown for 'Selezionare l'esito Check'.

Regola	Warning	Valida
37009	Verificare: non è stato indicato l'indirizzo e-mail dell'istituzione nonprofit	Valida
37011	Verificare: non è stato indicato il telefono dell'istituzione nonprofit	Valida
33099	Verificare: il numero di dipendenti è significativamente diverso da quello registrato nelle fonti d'archivio	Valida
34019	Verificare: l'istituzione ha soci ma non risultano nel bilancio contributi annui degli aderenti	Valida
34073	Verificare: l'istituzione ha soci ma non risultano nel bilancio contributi annui degli aderenti	Invalida
34083	Verificare: la differenza tra entrate e uscite è piuttosto elevata	Valida
34103	Verificare: rispetto al numero di dipendenti impiegati, le spese per i dipendenti sono troppo basse	Valida
34071	Verificare: la percentuale della voce "altre entrate di fonte privata" è piuttosto elevata	Valida
34009	Verificare: è stato indicato un ammontare dei proventi inferiore a 1000 euro	Valida
34031	Verificare: è stato indicato un ammontare degli oneri inferiore a 1000 euro	Valida

### 3.4.2 La validazione del questionario

La fase di validazione dei questionari rappresenta il passo conclusivo del processo di controllo dei dati implementato in SGRCIS. A fronte dell'operazione di validazione di un questionario, che era eseguibile solo su questionari nello stato di "Check OK" o "Non rilevato", il questionario passava nello stato "Validato". Tale operazione poteva essere effettuata dal solo responsabile dell'Ufficio di censimento. Inoltre da questo punto in poi non era possibile effettuare ricicli sulla lavorazione del questionario.

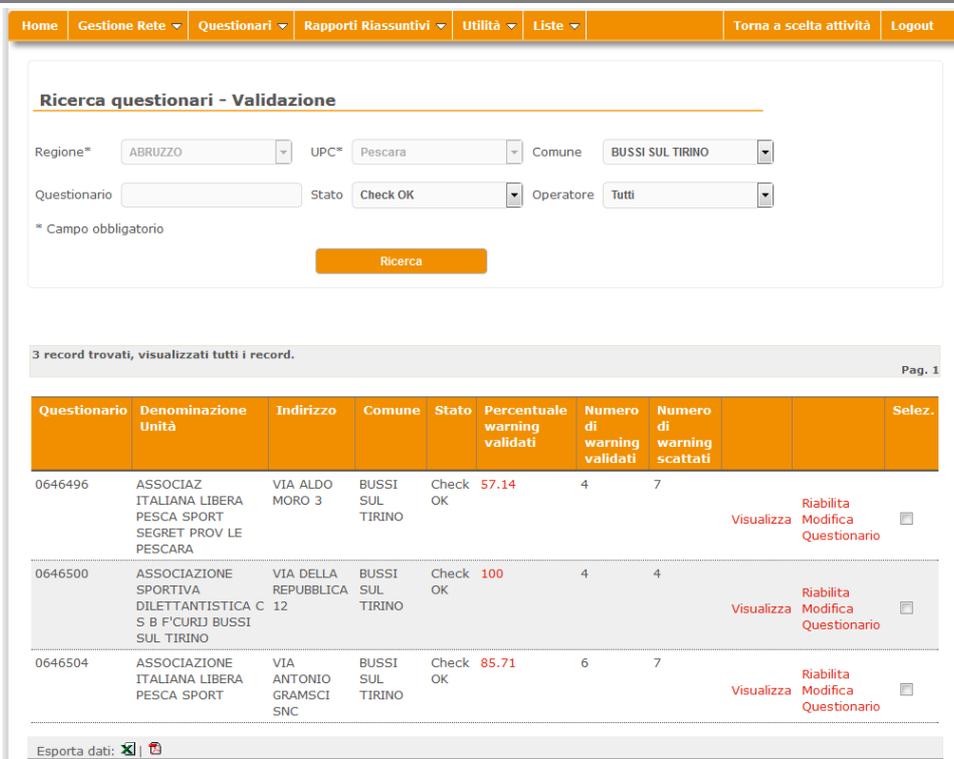
L'implementazione del *workflow* consisteva di una pagina di ricerca dei questionari di pertinenza dell'operatore collegato che si trovavano negli stati di partenza previsti, con possibilità di restringere la ricerca a un solo questionario o a un territorio specifico (in questo caso il singolo

comune). Sulla lista dei questionari individuati dalla ricerca era possibile eseguire una selezione multipla per eseguire la validazione in blocco, il cambio di stato quindi veniva eseguito mediante un' opportuna transazione che evitava il verificarsi di problemi legati a *dirty reads* sul questionario. L'esito della validazione veniva quindi mostrato all'utente per ciascun questionario sul quale aveva eseguito la validazione mediante l'utilizzo di *Ajax*.

Accanto ad ogni questionario, inoltre, sono stati inseriti i seguenti indicatori: percentuale di *warning* validati dall'operatore, numero di *warning* validati e numero di *warning* scattati al momento dell'esecuzione del *check*; la Figura 3.9 mostra un esempio di maschera di validazione. Queste informazioni avevano lo scopo di coadiuvare il responsabile nella validazione del questionario in quanto potevano essere utilizzate come *proxy* per valutare sia la bontà del questionario iniziale sia il lavoro del rilevatore. Da un punto di vista tecnico, per evitare tempi di attesa lunghi, gli indicatori non venivano calcolati al momento del caricamento della pagina ma venivano memorizzati in tabelle Oracle e calcolati da un *trigger* che si attivava quando il questionario superava la fase del *check*.

Nel caso di dubbi da parte del responsabile sulla correttezza del questionario o sulle modifiche effettuate dall'operatore, l'interfaccia grafica consentiva la riapertura dello stesso in modalità "modifica". In questo caso il rilevatore poteva rientrare direttamente in registrazione sul questionario (Figura 3.9).

**Figura 3.9 – Validazione in SGRCIS**



The screenshot shows the 'Ricerca questionari - Validazione' interface. At the top, there is a navigation menu with items: Home, Gestione Rete, Questionari, Rapporti Riassuntivi, Utilità, Liste, Torna a scelta attività, and Logout. Below the menu, the search filters are set to: Regione\* (ABRUZZO), UPC\* (Pescara), Comune (BUSSI SUL TIRINO), Questionario (empty), Stato (Check OK), and Operatore (Tutti). A 'Ricerca' button is located below the filters. Below the search area, it indicates '3 record trovati, visualizzati tutti i record.' and 'Pag. 1'. The main content is a table with the following columns: Questionario, Denominazione Unità, Indirizzo, Comune, Stato, Percentuale warning validati, Numero di warning validati, Numero di warning scattati, and Selez. The table contains three rows of data:

Questionario	Denominazione Unità	Indirizzo	Comune	Stato	Percentuale warning validati	Numero di warning validati	Numero di warning scattati	Selez.
0646496	ASSOCIAZ ITALIANA LIBERA PESCA SPORT SEGRET PROV LE PESCARA	VIA ALDO MORO 3	BUSSI SUL TIRINO	Check OK	57.14	4	7	Visualizza Riabilita Modifica Questionario
0646500	ASSOCIAZIONE SPORTIVA DILETTANTISTICA C S B F'CURIJ BUSSI SUL TIRINO	VIA DELLA REPUBBLICA 12	BUSSI SUL TIRINO	Check OK	100	4	4	Visualizza Riabilita Modifica Questionario
0646504	ASSOCIAZIONE ITALIANA LIBERA PESCA SPORT	VIA ANTONIO GRAMSCI SNC	BUSSI SUL TIRINO	Check OK	85.71	6	7	Visualizza Riabilita Modifica Questionario

At the bottom of the table, there is an 'Esporta dati:' button with icons for Excel and PDF.

### 3.5 Il monitoraggio aggregato e di dettaglio

I rapporti riassuntivi hanno consentito di monitorare sia l'andamento del Censimento non profit e della Rilevazione sulle imprese fino al livello di dettaglio, sia l'attività dei rilevatori e degli operatori coinvolti.

Nella macro area ad essi dedicata gli utenti avevano a disposizione diversi report di sintesi che permettevano di analizzare l'andamento della rilevazione, il lavoro degli operatori e i risultati della fase di controllo e correzione dei dati. Tutti i report presenti nel sistema visualizzavano macrodati a diversi livelli di dettaglio, territoriale e operativo, memorizzati in tabelle Oracle e aggiornati periodicamente. Accanto ai report di sintesi sono state, inoltre offerte delle liste di dettaglio che hanno permesso il monitoraggio puntuale dei dati. Ogni lista aveva le proprie peculiari caratteristiche in termini di dati offerti, di utilità e visibilità. In tale contesto con lista si intende un elenco di componenti, come i questionari lavorati da ogni operatore, ovvero l'elenco dei dati. In ultimo, il sistema permetteva di effettuare lo scarico delle cosiddette liste di microdati, che, per ciascun questionario registrato, aveva superato la fase di controllo e verifica della consistenza e di correttezza dei dati immessi.

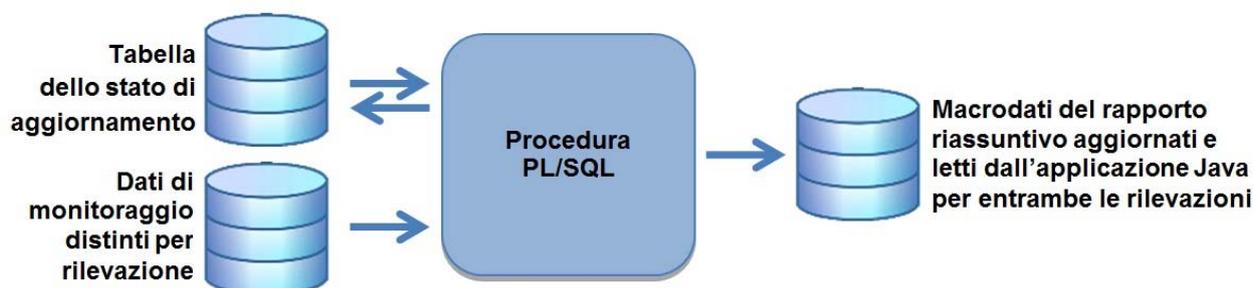
#### 3.5.1. L'architettura DB

Per la realizzazione dei rapporti riassuntivi è stata implementata una particolare infrastruttura tecnica (Figura 3.10), allo scopo di ridurre i tempi di attesa nella visualizzazione dei report da parte dell'applicazione.

La quantità di dati da elaborare nell'ambito delle due rilevazioni non era talmente elevata da non consentire la visualizzazione di report calcolati sui dati di monitoraggio, e quindi sempre aggiornati. Tuttavia, ciò ha implicato dover adottare tutti gli accorgimenti necessari per far sì che il tempo di creazione dei prospetti si mantenesse entro un limite accettabile per gli utenti.

Il più importante di tali accorgimenti è stato quello di effettuare tutte le *query* ed i calcoli necessari per ricavare i dati di un rapporto riassuntivo all'interno di un'unica procedura PL/SQL residente sul database Oracle.

Figura 3.10 – Flusso di calcolo dei dati aggregati per ciascun rapporto riassuntivo



Una procedura PL/SQL leggeva distintamente i dati di monitoraggio del Censimento non profit e della Rilevazione sulle imprese ed effettuava per ciascuna rilevazione tutte le elaborazioni necessarie per produrre i macrodati del rapporto riassuntivo, che venivano poi inseriti distinti per rilevazione all'interno di una tabella Oracle.

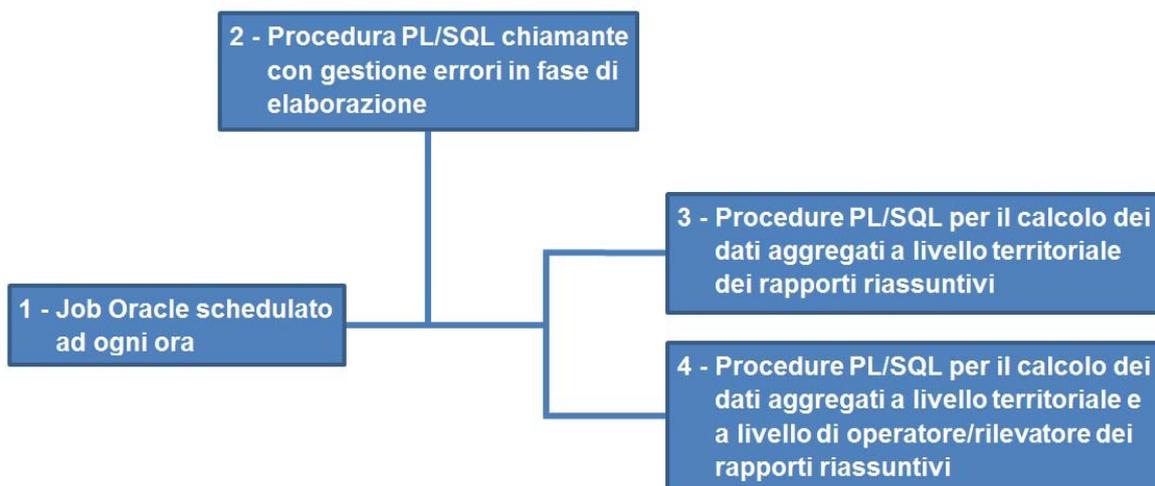
È importante sottolineare che, al fine di sincronizzare la fase di aggiornamento del macrodato con quella di visualizzazione dello stesso nel report, sono state create per ciascun rapporto riassuntivo

due tabelle Oracle identiche per struttura. Nel momento in cui una delle due era in fase di aggiornamento dei dati, l'applicazione Java leggeva l'altra tabella per visualizzare il report. Questo meccanismo è stato gestito all'interno della procedura PL/SQL leggendo da una tabella dei metadati la modalità di lettura prima di effettuare l'inserimento del macrodato calcolato su una o l'altra tabella e aggiornando la modalità di lettura stessa al termine dell'inserimento.

Seguendo in linea di massima il flusso di calcolo descritto in precedenza, sono state sviluppate 11 procedure PL/SQL per la visualizzazione dei corrispondenti rapporti riassuntivi. Queste procedure erano richiamate all'interno di un'unica procedura PL/SQL di avvio, che gestiva anche gli errori in fase di elaborazione con opportune tabelle di *log* (Figura 3.11). L'esecuzione della procedura chiamante era effettuata ad intervalli di tempo regolari di circa un'ora mediante la schedulazione di un *Job Oracle*.

In questo modo era possibile aggiornare periodicamente i macrodati per ciascun rapporto riassuntivo in concomitanza con l'andamento delle due rilevazioni.

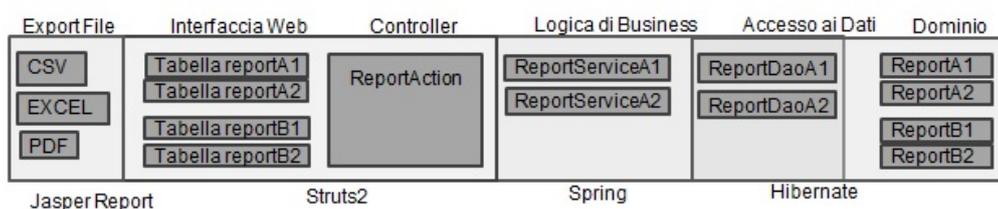
**Figura 3.11 – Step di esecuzione delle procedure PL/SQL per l'aggiornamento periodico di tutti i rapporti riassuntivi**



### 3.5.2 L'architettura software

L'infrastruttura tecnologica è stata ideata come un modulo software, perfettamente integrato nell'architettura generale del Sistema di Gestione della Rilevazione, contraddistinto quindi dall'adesione ai *design pattern* informatici quali MVC (*Model-View-Controller*), DAO (*Data Access Object*) e IoC (*Inversion of Control*), come schematizzato in Figura 3.12.

**Figura 3.12 – Schema Architettura sw delle funzioni di monitoraggio**



I dati visualizzati, erano letti da tabelle di macrodati popolate mediante il sistema descritto al paragrafo precedente. Poiché per ogni rapporto riassuntivo erano presenti due tabelle di macrodati, di cui alternativamente, una aggiornata e l'altra in fase di aggiornamento, il modulo è stato progettato in modo da selezionare automaticamente la tabella opportuna, mediante l'implementazione nell'architettura delle API *Java Reflection*. L'uso della libreria ha permesso di eterizzare l'accesso alla base dati automatizzando la selezione *on demand* del report in relazione alla definizione di ruolo, livello e ambito territoriale dell'operatore della rete di rilevazione.

L'interfaccia web, costituita a partire da strumenti propri dell'architettura generale, come *Struts2*, *jQuery* e le librerie *DisplayTag*, è stata personalizzata per permettere di rappresentare tabelle elaborate, navigabili tra i livelli previsti mediante la tecnica del *drill-down* e del *roll-back*.

Inoltre, il sistema permetteva di scaricare i dati e gestire file di grandi dimensioni per un elevato numero di utenti, quasi in *real-time* rispetto alle operazioni. La funzionalità di scarico dei dati è stata implementata aggiungendo la libreria Java *JasperReports* e il relativo *plugin* di *Struts2*. Queste tecnologie hanno permesso di snellire tutta la gestione di esportazione dei differenti formati dei file di dati, strutturandoli secondo *template* costruiti a partire dal tool designer *IReport*.

Era 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 reinterrogazione delle tabelle del database e ottimizzando le prestazioni del download. Contemporaneamente, si evitavano 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 era automatizzata l'intercettazione delle eccezioni con relativo invio e-mail di segnalazione errori.

I rapporti riassuntivi previsti per il monitoraggio della rilevazione sono stati 11, come mostra la Figura 3.13, ognuno associato ad una determinata fase del processo di rilevazione.

Figura 3.13 – Menu a tendina con le funzioni di monitoraggio disponibili



La struttura è analoga tra tutte le funzioni di monitoraggio ed un esempio è riportato in Figura 3.14. La navigazione tra i possibili livelli territoriali si effettuava cliccando sull'icona a destra, per la fase di *drill-down*, e sul tasto indietro per il *roll-back* in funzione del livello territoriale di competenza dell'operatore.

I livelli previsti nel sistema partivano dalla regione fino ad arrivare al massimo livello di dettaglio possibile, ovvero il comune o l'operatore di rete (rilevatore) a seconda del tipo di rapporto.

Se la lista dei dati visualizzata era costituita da un elevato numero di record, tale da non permetterne la visualizzazione agevole nella pagina web, la lista stessa veniva paginata mediante la barra di navigazione evidenziata in alto, in Figura 3.14.

Per tutti i rapporti era possibile salvare i dati in due differenti formati cliccando sulle apposite icone:

- file *pdf*: che riportava i dati visualizzati nella pagina corrente (salvataggio per pagina);
- file *xls*: che riportava i dati visualizzati nella pagina corrente (salvataggio per pagina).

Inoltre, per i rapporti “Stato Consegna da Vettori”, “Motivo Mancata Consegna”, “Restituzione Questionari”, “Riepilogo per Pesi”, “Stato Lavorazione Questionari Per Territorio”, “Rapporto Esito Territorio”, “Attività unità Rilevate” e “Rapporto Diffide” era possibile salvare i dati in un unico file *csv* (*comma separator value*) così organizzato:

- tutti i comuni d'Italia (utenza nazionale);
- tutti i comuni della regione (utenza regionale).

Mentre per i rapporti “Riepilogo per Pesi” e “Rapporto Diffide” era possibile anche scaricare un unico file *csv* così organizzato:

- tutti i comuni d'Italia (utenza nazionale);
- tutti i comuni della regione (utenza regionale);
- tutti gli UPC d'Italia (utenza provinciale).

Figura 3.14 – Esempio di Struttura di un rapporto riassuntivo

Restituzione Questionari															
21 record trovati, visualizzati tutti i record.													Pag. 1		
Intero Territorio Nazionale - Aggiornato al: 25/09/2013 12:01															
Regione/Provincia autonoma	Lista processi e nuovi	Quest. con esito non rilevato da inasitato	Quest. restituiti web		Quest. restituiti punti ritiro		Quest. restituiti upc		Quest. restituiti rilevatore		Quest. restituiti		Quest. non restituiti o inasitati non lavorati		
			Tot.	%	Tot.	%	Tot.	%	Tot.	%	Tot.	%	Tot.	%	
PIEMONTE	16.405	9	3.175	19,36	484	2,95	63	0,38	14	0,09	0	3.736	22,79	12.660	77,21
VALLE D'AOSTA	819	0	84	10,26	35	4,27	11	1,34	0	0,00	0	130	15,87	689	84,13
LOMBARDIA	60.420	51	11.916	19,74	1.637	2,71	453	0,75	22	0,04	1	14.028	23,24	46.341	76,76
VENETO	25.954	1	5.219	20,11	526	2,03	133	0,51	18	0,07	6	5.896	22,72	20.057	77,28
FRIULI VENEZIA GIULIA	5.312	2	952	17,93	129	2,43	49	0,92	10	0,19	7	1.140	21,47	4.170	78,53
LIGURIA	5.489	6	840	15,32	139	2,54	43	0,78	2	0,04	1	1.024	18,68	4.459	81,32
EMILIA-ROMAGNA	24.255	16	4.571	18,86	833	3,44	220	0,91	14	0,06	3	5.638	23,26	18.601	76,74
TOSCANA	19.233	25	2.960	15,41	360	1,87	175	0,91	17	0,09	1	3.512	18,28	15.696	81,72
UMBRIA	3.877	1	567	14,63	113	2,92	41	1,06	3	0,08	1	724	18,68	3.152	81,32
MARCHE	7.860	2	1.215	15,46	185	2,35	35	0,45	7	0,09	0	1.442	18,35	6.416	81,65
LAZIO	23.776	1	2.819	11,86	340	1,43	72	0,30	2	0,01	0	3.233	13,60	20.542	86,40
ABRUZZO	5.440	3	571	10,50	93	1,71	21	0,39	2	0,04	0	687	12,64	4.750	87,36
MOLISE	1.259	0	124	9,85	21	1,67	28	2,22	1	0,08	1	174	13,82	1.085	86,18
CAMPANIA	18.282	3	1.901	10,40	115	0,63	34	0,19	3	0,02	0	2.053	11,23	16.226	88,77
PUGLIA	12.024	7	1.570	13,06	108	0,90	100	0,83	5	0,04	0	1.783	14,84	10.234	85,16
BASILICATA	1.921	1	199	10,36	33	1,72	7	0,36	1	0,05	0	240	12,50	1.680	87,50
CALABRIA	4.694	3	428	9,12	48	1,02	32	0,68	11	0,23	2	519	11,06	4.172	88,94
SICILIA	11.927	3	1.162	9,75	128	1,07	73	0,61	8	0,07	1	1.371	11,50	10.553	88,50
SARDEGNA	5.114	3	571	11,17	85	1,66	27	0,53	1	0,02	0	684	13,38	4.427	86,62
BOLZANO-BOZEN	3.334	1	828	24,84	52	1,56	34	1,02	1	0,03	0	915	27,45	2.418	72,55
TRENTO	2.715	4	595	21,95	176	6,49	25	0,92	0	0,00	0	796	29,36	1.915	70,64
<b>Totale</b>	<b>280.110</b>	<b>142</b>	<b>42.267</b>	<b>16,26</b>	<b>5.640</b>	<b>2,17</b>	<b>1.676</b>	<b>0,64</b>	<b>142</b>	<b>0,05</b>	<b>24</b>	<b>49.725</b>	<b>19,13</b>	<b>210.243</b>	<b>80,87</b>

Tutti i comuni del territorio nazionale 

In definitiva, le funzioni di monitoraggio gestivano un'elevata mole di dati quasi in *real time* (nell'ordine di 105 record ed *export* online di file fino a 5 MB) garantendo accessibilità e fruibilità ad un elevato numero di utenti (numero di accessi stimato dell'ordine di dieci su un totale di undici rapporti riassuntivi).

### **3.5.3 Le liste di dettaglio**

Ciascun operatore, durante lo svolgimento delle proprie funzioni, ha avuto a disposizione degli elenchi in base ai permessi e alla visibilità a lui concessi nel sistema. Tali liste offerte hanno accompagnato lo sviluppo delle operazioni censuarie in tutte le sue fasi. In particolare esistono liste che supportavano le suddette attività primariamente nella fase iniziale del censimento, quali, ad esempio, le liste precensuarie, e successivamente erano messi a disposizione degli utenti diversi elenchi basati su dati specifici di uno o più questionari.

Ciascuna delle suddette liste presenta caratteristiche peculiari in termini di dimensione, tipo di dati registrati ed esigenza d'uso e questo ha permesso di giustificare scelte progettuali differenti in termini di metodi e di tecnologie di trattamento.

Tutte le informazioni di dettaglio sulla lavorazione dei questionari sono state gestite attraverso un Diario, cioè attraverso un'entità concettuale che accorpa e struttura tutte le principali informazioni che sostenevano i rilevatori ed i loro coordinatori nello svolgimento delle proprie attività, quali, i dati identificativi dell'unità economica, il tipo di rilevazione, lo stato, gli autori e le date di alcuni passaggi di stato rilevanti a fini di monitoraggio. Il Diario rappresenta, quindi, uno degli strumenti principali con cui i rilevatori e i loro responsabili hanno interagito all'interno del sistema SGRCIS, nell'ambito dello svolgimento delle proprie operazioni censuarie e, di fatto, schematizza e struttura le principali informazioni utili all'avanzamento delle attività censuarie nelle sue diverse fasi. Esso è stato quindi, in definitiva, uno strumento utilizzato per effettuare un continuo monitoraggio dello stato dei singoli questionari, nonché per l'individuazione di date ed autori caratterizzanti gli stati maggiormente rilevanti in cui il questionario transitava durante la lavorazione, e tale da guidare ed assistere gli operatori nelle varie operazioni da compiere sui singoli questionari, compatibilmente col flusso di lavoro disegnato. In dettaglio, il Diario ha, altresì, permesso ai rilevatori:

- di verificare i possibili ritardi da parte delle unità da rilevare in modo da attivarsi, all'occorrenza, per un sollecito;
- di pianificare le proprie uscite sul campo per effettuare le operazioni di consegna, di controllo e di verifica sui dati del questionario;
- di ottimizzare i percorsi sul territorio, portando con sé l'elenco delle unità da contattare.

A tal proposito si evidenzia come il sistema web SGRCIS possa essere descritto come un *metadata-driven system*. Il Diario ha rappresentato un utile strumento di lavoro per operatori con il profilo di Responsabile Ufficio Regionale di Censimento (URC), Responsabile Istat Territoriale (RIT), Responsabile Ufficio Provinciale di Censimento (UPC), Coordinatore Provinciale (COP) e Rilevatore (RIL). Le varie aree di lavoro presenti nel sistema web SGRCIS che si interfacciavano con il suddetto strumento prevedevano generalmente differenti modalità di ricerca sul Diario: dai filtri territoriali, in cui alcuni campi sono stati pre-valorizzati a seconda dell'operatore ed al ruolo da esso ricoperto, alla ricerca per stato del questionario o per "Rilevatore assegnatario", fino alla ricerca puntuale per identificativo del questionario.

Gli scenari di lavoro che sono stati individuati, in sede di progettazione tecnica e tecnologica del sistema, sono stati sostanzialmente due.

In prima istanza si è rilevata l'esigenza di effettuare ricerche mirate sul Diario, imponendo quindi vincoli di obbligatorietà su specifici incroci di filtri di ricerca. In tal modo si è mantenuta entro limiti

precalcolati la dimensione degli insiemi di dati da trattare e quindi i tempi di risposta del sistema. Si è potuto quindi ricorrere a chiamate sincrone verso il sistema server per rispondere alle diverse richieste degli utenti. Si ricorda che la gestione di tali tipi di chiamate nell'ambito delle applicazioni web richiede un onere inferiore in termini di difficoltà progettuale. Nello specifico, sono state utilizzate librerie, appositamente definite per la gestione del *rendering* di contenuti web online in forma tabellare e per la relativa esportazione in formati diversi, che hanno gestito, contestualmente a ogni singola richiesta degli utenti, sia l'interrogazione del DBMS, al fine di recuperare i dati di interesse, sia la costruzione del file richiesto, nei formati *csv*, *xls* e *pdf*.

Si è rilevata, altresì, l'esigenza per gli operatori di ampliare lo spazio di ricerca e di poter esportare ampie sezioni del Diario. La soluzione progettuale sopra delineata sarebbe stata caratterizzata da eccessivi tempi di risposta del sistema e sarebbe comunque incorsa in problemi di timeout in vari punti dell'architettura web disegnata; test applicativi hanno permesso di orientare le scelte progettuali verso l'utilizzo di chiamate asincrone, la definizione di apposite procedure PL/SQL sul server DBMS per l'interrogazione della base dati e la costruzione del file nel formato *csv* e l'utilizzo di strutture dati appositamente dedicate alla gestione di tali esportazioni, così da avere comunque bassi tempi di risposta ed evitare operazioni ridondanti o ripetute. Nello specifico, alla richiesta dell'utente di esportazione di una determinata lista veniva attivata una procedura sul server DBMS Oracle e non veniva tenuto in attesa l'utente, il quale poteva richiedere lo scarico effettivo del file, solo al termine della costruzione del file richiesto e relativa memorizzazione in una struttura tabellare del DBMS. Il sistema rigenerava gli stessi file, a seguito di una nuova richiesta, solo se il disallineamento temporale era tale da giustificare la riattivazione della procedura Oracle sopra richiamata.

In sintesi a seconda della specifica esigenza degli utenti è stata resa possibile l'esportazione, in file con diversi formati, sia del Diario, e quindi della lista di questionari, con la specifica valorizzazione degli attributi annessi, espressamente richiesta, sia delle liste contenenti informazioni storiche relativamente all'evoluzione di ogni questionario rispetto agli stati caratterizzanti il flusso di lavorazione.

#### **3.5.4 Dalle liste precensuarie alle liste dei pesi**

Uno degli obiettivi tenuti presenti nella ideazione del portale SGRCIS è la sua struttura modulare che ha consentito di inserire accanto a strumenti previsti sin dagli inizi, come le "liste precensuarie" anche altre funzionalità le cui caratteristiche di dettaglio sono emerse in corso d'opera come la "Visualizzazione Dettaglio Pesi", utilizzata nelle fasi finali delle attività censuarie per la quantificazione dei contributi agli organi di censimento.

Le liste precensuarie rappresentano la base per l'identificazione, la localizzazione e la descrizione strutturale di imprese e istituzioni non profit coinvolte nella rilevazione censuaria e quindi destinatarie dei questionari. Tali liste sono state create utilizzando le informazioni provenienti da varie fonti amministrative e statistiche proprie dei rispettivi settori di cui si cita, uno per tutti, l'Archivio statistico delle imprese attive, (Asia), gestito e mantenuto dall'Istat, che rappresenta la principale fonte informativa sulla struttura delle imprese.

Per quanto riguarda le imprese, le liste precensuarie comprendevano i dati delle imprese con almeno venti addetti ed un campione per le imprese con meno di venti addetti.

Per quanto riguarda le istituzioni non profit le liste precensuarie comprendevano i dati delle unità individuate e classificate in relazione alle diverse tipologie istituzionali (dalle associazioni culturali o sportive alle associazioni di volontariato).

Le liste precensuarie sono state raffinate con operazioni di normalizzazione, georeferenziazione degli indirizzi ed estrazione della sezione di censimento.

Si sono rilevate particolarmente utili, nelle varie fasi del censimento:

- durante la fase di progettazione del censimento, perché hanno consentito di differenziare le indagini censuarie attraverso questionari mirati alle diverse tipologie di imprese;
- per la spedizione dei questionari a imprese e istituzioni non profit;
- per i solleciti inviati tramite posta elettronica certificata o posta tradizionale;
- nelle fasi di correzione dei dati raccolti e di verifica della qualità dei dati.

Le suddette liste sono state rese esportabili, in formato *csv*, da un'area applicativa specifica di SGRCIS, progettata tenendo presenti alcuni importanti obiettivi:

- sicurezza e salvaguardia della privacy. Ogni operatore doveva accedere solo alle funzionalità e ai dati strettamente connessi al proprio ambito operativo; le liste precensuarie in particolare erano prerogativa del Responsabile dell'Ufficio Regionale di Censimento (URC), del Responsabile Istat Territoriale (RIT) e del Responsabile dell'Ufficio Provinciale di Censimento (UPC);
- ottimizzazione del lavoro. I filtri territoriali nella pagina di ricerca venivano precaricati automaticamente in base al livello territoriale di competenza dell'operatore (regione o provincia);
- ottimizzazione delle performance dei download. In considerazione della natura dei dati delle liste precensuarie, noti sin dall'inizio e sostanzialmente non soggetti a cambiamenti durante le attività del censimento, si è deciso di generare preliminarmente dei file in formato *csv* contenenti ciascuno i dati di interesse di ognuno degli Uffici Provinciali di Censimento (UPC) e di memorizzare tali file direttamente in base dati. È evidente, infatti, che è molto più performante, in termini di carico di lavoro per i server e di tempi di attesa per l'operatore, scaricare direttamente dal DB, dei dati già formattati in modo definitivo in un file *csv*, anziché doverli ricostruire e "confezionarli" a ogni richiesta.

Nelle ultime fasi della rilevazione censuaria si è reso necessario avere un quadro sintetico e completo delle attività da loro svolte dagli operatori censuari.

A tal fine sono stati definiti dei particolari algoritmi che con delle procedure *batch*, eseguite in maniera massiva direttamente sul DB, hanno calcolato per ciascun questionario un "peso", cioè un indicatore numerico funzione delle varie casistiche di lavorazione del questionario stesso, utilizzato per la corresponsione dei contributi agli organi di censimento.

Attraverso una funzionalità sviluppata *ad hoc* sul portale SGRCIS, gli utenti con profilo di Responsabile Ufficio Regionale di Censimento (URC), Responsabile Istat Territoriale (RIT) e Responsabile Ufficio Provinciale di Censimento (UPC) hanno avuto la possibilità di visualizzare le liste dei questionari di propria competenza con i dettagli dei pesi associati. Anche in questo caso erano presenti dei filtri di ricerca, precaricati in base al livello territoriale dell'utente.

Le liste che riportavano i pesi erano esportabili in formato *pdf* o *xls*. Le loro dimensioni hanno giustificato il ricorso a chiamate sincrone al server e l'utilizzo di tecnologie Java per gestire, nell'ambito della stessa richiesta HTTP, sia l'interrogazione al DBMS, che la costruzione della lista nel formato di interesse.

### **3.5.5 Lo scarico dei microdati**

In questa sezione si presenta una particolare funzionalità del sistema web SGRCIS, utile in particolar modo a valle delle operazioni censuarie: lo scarico dei microdati. Per microdati, si intende l'insieme delle risposte degli utenti, coinvolti nella rilevazione censuaria, alle domande dei questionari.

Si rileva che l'esportazione di tali dati poteva essere effettuata per i questionari registrati, che avevano superato il *check*. In particolare, essa era possibile in ambito di rilevazione sulle imprese, ovvero sulle istituzioni non profit. In quest'ultimo caso erano ricomprese anche le unità locali.

Nello specifico, la suddetta funzionalità è stata messa a disposizione degli operatori con i profili di Responsabile Ufficio Regionale di Censimento (URC), Responsabile Istat Territoriale (RIT) e Responsabile Ufficio Provinciale di Censimento (UPC).

Nell'area di lavoro, appositamente predisposta per offrire tale funzionalità agli utenti, l'operatore poteva decidere se effettuare tale operazione su tutti i dati del territorio di competenza o se impostare un filtro in base a variabili territoriali e/o all'esito della rilevazione. La presenza di tali filtri garantiva che la dimensione delle liste da esportare fosse controllata e limitata. Data la natura delle informazioni da esportare e la quantità di contenuto informativo associato a ogni singolo questionario, costituente un componente della lista, sono state scelte librerie definite appositamente per supportare attività di *reporting*, quali quelle *open source* offerte dalla comunità *JasperSoft*.

### 3.6 La chiusura dell'indagine e la gestione delle eccezioni

Gli strumenti offerti dal sistema per guidare le operazioni di chiusura del censimento, hanno permesso la razionalizzazione delle attività e il risparmio di tempo in quanto hanno facilitato il controllo delle informazioni ed il colloquio tra la rete di rilevazione e l'Istat. Tramite il sistema, infatti, il responsabile UPC ha potuto gestire le attività connesse all'accertamento della violazione dell'obbligo di risposta e l'invio della relativa documentazione e contemporaneamente ha potuto visionare dei report di sintesi che mostravano la situazione del suo territorio al momento della chiusura. Di seguito sono descritte tutte le funzionalità messe a disposizione della rete e dell'unità rispondente per concludere le operazioni.

#### 3.6.1 La gestione delle procedure di accertamento dell'obbligo di risposta da parte della rete

Le unità di rilevazione che, al termine della rilevazione, non avevano effettuato la restituzione del questionario nei canali previsti hanno subito un'ulteriore lavorazione che prevedeva l'invio da parte degli UPC delle diffide ad adempiere. Trascorso un lasso di tempo dalla data di invio della diffida, se il questionario non veniva restituito, tutta la documentazione acquisita riguardo l'unità doveva essere inviata all'Istat che provvedeva alle attività connesse alla violazione dell'obbligo di risposta.

Per facilitare l'interazione tra Istat e la rete e per permettere un primo monitoraggio sul numero di diffide e sulle sanzioni il sistema consentiva la registrazione della data di invio della diffida ad adempiere da parte dell'UPC e della successiva data di invio dell'incartamento dall'UPC all'Istat. Il sistema prevedeva un insieme di regole che gestivano l'inserimento di tali date, in particolare:

- la data di invio della diffida ad adempiere poteva essere inserita solo se lo stato del questionario era compatibile, ossia non risultava pervenuto non solo all'UPC ma neanche al vettore postale (Figura 3.15);
- la data di invio ad Istat poteva essere inserita solo se era stata inserita la data dell'invio della diffida ad adempiere.

Figura 3.15 - Registrazione diffida

Analogamente per quanto fatto per il flusso del questionario era sempre possibile eliminare una registrazione precedentemente effettuata riportando il questionario allo stato in cui si trovava precedentemente alla registrazione della data di invio della diffida, mentre per eliminare l'informazione circa l'invio della diffida all'Istat il responsabile dell'UPC doveva necessariamente contattare l'Istat.

### 3.6.2 Il sistema di scarico delle sanzioni

Nell'infrastruttura informatica del censimento è stata sviluppata un'applicazione di supporto alla gestione delle procedure di accertamento della violazione dell'obbligo di risposta. La finalità dell'applicazione era fornire un sistema per l'accesso alla documentazione relativa alle suddette procedure simile nell'aspetto e nel funzionamento al sistema di compilazione del questionario online (Figura 3.16).

Figura 3.16 – Home page di accesso al sistema relativo alla documentazione

Gli utenti previsti potevano accedere al sistema dopo essersi registrati tramite una procedura semplificata rispetto a quella del questionario online:

- 1) l'utente effettua una richiesta di password indicando il suo codice utente (lo stesso del questionario online) e un indirizzo email valido;
- 2) il sistema invia all'utente una password casuale per l'accesso alla documentazione.

L'utente poteva richiedere una password indicando un indirizzo differente finché non effettuava un accesso al sistema, a quel punto l'e-mail dell'utente veniva assegnata definitivamente e non poteva più essere modificata; l'utente, in caso di necessità, poteva richiedere una nuova password indicando l'indirizzo e-mail con cui aveva effettuato il primo accesso.

Una volta effettuato l'accesso, l'utente poteva scaricare la documentazione personale e il modello specifico (modello F3). La documentazione era composta da uno o più file *pdf* forniti dal servizio che si occupava della gestione delle procedure di accertamento della violazione dell'obbligo di risposta e resi disponibili all'applicazione in una cartella condivisa.

L'utilizzo di una applicazione web per lo scarico della documentazione, piuttosto che l'invio della stessa tramite posta ha permesso un risparmio nei costi ed una tempestiva tracciatura delle operazioni svolte.

### 3.6.3 Le operazioni di chiusura

È stata predisposta nel sistema una apposita funzione per la gestione delle operazioni di chiusura del censimento. Questa funzione ha permesso la stampa dei report previsti e la chiusura dell'indagine censuaria, secondo le dispositivi dell'Istat descritte nella Circolare n. 7, paragrafo 2 del 19 dicembre 2012 esposta nel Portale della rete. Le operazioni di chiusura erano di pertinenza del solo responsabile dell'UPC, che dopo aver controllato i report di sintesi riferiti ad entrambe le rilevazioni poteva effettuare la chiusura e stampare i documenti da firmare ed inviare all'Istat (Figura 3.17). Le operazioni di chiusura valevano per entrambe le rilevazioni e non consentivano alcuna ulteriore operazione sui questionari.

Il primo report riassume la situazione finale delle unità di pertinenza dell'UPC per singolo tipo di lavorazione, secondo quanto previsto delle circolari attuative dell'Istat.

Il secondo report è una sintesi di quanto viene mostrato, nel dettaglio, dal rapporto riepilogo per i pesi (Figura 3.17). I pesi, sono stati descritti nei paragrafi precedenti e sono calcolati in base alla Circolare n. 5 del 23 ottobre 2012 e della Circolare n. 7 del 19 dicembre 2012.

Figura 3.17 – Chiusura censimento per UPC

REPORT CHIUSURA						
Tipologia Indagine	Unità nello stato P1cchi Investiti, P1cchi Investiti da Vettore o per le quali non è possibile attribuire un peso	Unità rilevate da UPC	Unità con diffida da adempiere effettuata	Unità con diffida da adempiere non effettuata		
Imprese	817	1620	5	5918		
Non Profit	5521	1692	5	10941		

REPORT PESI							
Tipologia Indagine	Totale Questionari in lista nuovi	Peso 0.70	Peso 0.35	Peso 0.20	Peso 2.40	Peso 1.10	Peso 0
Imprese	8358	112	1488	19	0	7	6739
Non Profit	18160	229	1089	373	0	5	16466

### **3.6.4 L'interazione con il vettore postale per il ritiro della documentazione censuaria**

SGRCIS ha consentito la gestione del ritiro dei questionari cartacei effettuato dal vettore postale presso gli UPC. Affinché fosse possibile il ritiro dei questionari, questi dovevano essere impacchettati in scatole e accompagnati da specifica documentazione richiesta dal vettore. La presenza di scatole pronte per il ritiro doveva inoltre essere segnalata al vettore perché potesse pianificare e realizzare il ritiro. SGRCIS ha quindi gestito l'interazione con il vettore postale attraverso due gruppi di funzioni: funzioni dedicate all'UPC per il supporto alla predisposizione del ritiro e funzioni dedicate al vettore postale per la gestione del ritiro stesso.

Gli UPC dovevano inserire i questionari in tre tipologie di scatole: imprese, non profit e miste. Ogni scatola doveva essere accompagnata da una distinta di trasporto, ovvero un modulo, stampato in duplice copia, contenente alcune informazioni quali il numero di modelli contenuti nella scatola, il tipo di scatola, un codice a barre specificamente richiesto dal vettore postale per il tracciamento della stessa, eccetera. La spedizione delle scatole doveva essere accompagnata da uno specifico documento, il documento di trasporto (DDT), recante le informazioni complessive sulla spedizione, quali il numero di scatole per tipologia, la data e l'indirizzo presso cui effettuare il ritiro. Il sistema consentiva la predisposizione di tali documenti, inclusa la generazione del codice a barre per le distinte, e la loro stampa.

In particolare le funzioni previste per gli UPC erano le seguenti:

- creazione del DDT;
- gestione del DDT;
- chiusura del DDT;
- stampa del DDT e distinte.

Il processo che l'UPC doveva svolgere per preparare la documentazione di trasporto prevedeva:

- 1) la creazione del DDT attraverso l'omonima funzione;
- 2) la predisposizione delle distinte (creazione, eventuali modifiche e cancellazione), contestualmente alla creazione del DDT o attraverso la funzione "Gestione DDT";
- 3) la chiusura del DDT: operazione fondamentale, che doveva essere effettuata solo al termine del lavoro di predisposizione delle scatole. Con tale operazione il DDT e le relative distinte non erano più modificabili e veniva prenotato il ritiro, nella data selezionata, da parte del vettore postale;
- 4) la stampa della documentazione di trasporto, ovvero del DDT e delle relative distinte: operazione indispensabile per il ritiro delle scatole da parte del vettore postale.

Il vettore postale effettuava i ritiri un giorno infrasettimanale, all'interno di un periodo concordato con l'Istat. Il giorno precedente erano reperibili su SGRCIS le informazioni relative alle richieste di ritiro, con le quantità delle relative scatole. L'operatore postale, che disponeva di uno specifico profilo applicativo all'interno del sistema, aveva a disposizione le seguenti funzioni:

- scarico dei documenti di trasporto: la funzione consentiva lo scarico di un file riguardante tutti i documenti di trasporto dei vari UPC a livello nazionale, regionale o di singolo UPC, attinenti a una certa data di ritiro; tale scarico consentiva al vettore di sapere in quali UPC erano stati predisposti e richiesti i ritiri;
- scarico delle distinte: la funzione consentiva lo scarico di un file riguardante tutte le distinte, collegate a DDT inseriti dai vari UPC, a livello nazionale e facenti riferimento ad una certa data di ritiro; tale scarico consentiva al vettore di avere un'informazione quantitativa sulle scatole da ritirare;

- stampa della documentazione di trasporto.

Nella gestione dell'interazione tra UPC e vettore postale SGRCIS ha avuto un ruolo cruciale in quanto ha offerto un unico punto di accesso al vettore per reperire tutte le informazioni necessarie ai ritiri, consentendone una gestione efficace e tempestiva, e ha fornito agli UPC gli strumenti necessari alla predisposizione della documentazione e alla prenotazione dei ritiri.

### 3.7 La rendicontazione delle spese censuarie

Tra le varie opzioni riguardanti la scelta iniziale dell'attività è presente la voce "Rendicontazione". Le funzioni interne a quest'area consentivano agli Uffici Provinciali di Censimento (UPC) di adempiere all'obbligo di effettuare la rendicontazione delle spese censuarie, attraverso la compilazione di un apposito modello.

I profili abilitati a tale area erano: il profilo UPC (solo il responsabile), che poteva svolgere le attività di compilazione, visualizzazione e stampa del proprio modello, i profili Istat e Unioncamere, che potevano accedere a tutti i modelli in sola visualizzazione, in modo da effettuare una sorta di monitoraggio sull'andamento della compilazione o su richiesta effettuare in casi particolari, se lo ritenevano necessario, la riapertura del modello per permettere al responsabile UPC ulteriori modifiche dopo la chiusura del modello.

Il modello di rendicontazione era unico per tutte e due le rilevazioni e aveva la stessa struttura del modello cartaceo predisposto dall'Istat. Inoltre la sua navigazione era libera, ovvero non esisteva una gerarchia di compilazione tra le sezioni. Pertanto, indipendentemente dalla sequenza con cui erano state numerate, l'UPC poteva decidere autonomamente l'ordine di compilazione ed effettuare anche inserimenti parziali delle informazioni, rispettando solo 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 veniva invece effettuato un controllo sui contenuti dell'intero modello e richiesto il rispetto della completezza e della congruenza dei dati inseriti complessivamente.

Sono state previste diverse sezioni; ciascuna di esse presentava delle maschere di inserimento o modifica dati, simili a quella riportata in Figura 3.18 che riguarda nello specifico la Sezione 2.1. All'interno di ciascuna sezione erano previsti due tasti, uno di salvataggio e uno di annullamento. Il primo consentiva di salvare le informazioni inserite sino a quel momento all'interno della singola sezione di riferimento (ossia quella visualizzata sulla pagina). Tali informazioni rimanevano disponibili e modificabili sino al momento della creazione del documento in *pdf*. L'altro tasto permetteva di annullare tutti i dati immessi all'interno di una sezione e non ancora salvati, ripristinando così la situazione relativa all'ultimo salvataggio effettuato.

La Sezione 1 "Notizie Generali" era in sola visualizzazione: il responsabile aveva la possibilità in questa sezione di visualizzare diverse informazioni riguardanti il proprio ufficio di censimento come denominazione UPC, nominativo responsabile, stato del modello di rendicontazione, indirizzo e-mail del responsabile.

La Sezione 2 comprendeva tre sottosezioni distinte ("Numero addetti", "Mesi uomo", "Costo del lavoro") in ognuna delle quali si potevano gestire informazioni ricollegabili al lavoro come il numero degli addetti, i tempi e i costi ripartiti per funzione svolta e tipologia di contratto. Più nello specifico, il responsabile UPC aveva la possibilità, per ciascuna tipologia di contratto e funzione svolta, di inserire e modificare:

- il numero addetti coinvolto nelle operazioni censuarie (Sezione 2.1);
- il numero di mesi uomo lavorati dagli addetti (Sezione 2.2);

- le ore di straordinario effettuate dai dipendenti dell'amministrazione di riferimento per lo svolgimento delle attività censuarie (Sezione 2.2);
- le voci di costo del lavoro, come retribuzione lorda, gli oneri sociali obbligatori a carico del datore del lavoro, le spese sostenute per la formazione del personale e quelle sostenute per eventuali missioni e per eventuali rimborsi spese (Sezione 2.3).

Nella Sezione 3 "Altre Spese" il responsabile UPC aveva la possibilità di inserire o modificare le seguenti informazioni, tutte necessarie all'espletamento delle attività censuarie:

- spesa complessiva sostenuta per assistenza informatica e software;
- spesa complessiva sostenuta per acquisto di attrezzatura hardware;
- spesa complessiva sostenuta per arredamento, l'utilizzo o l'eventuale affitto dei locali;
- spesa complessiva sostenuta per le spese di gestione dell'UPC (materiali di consumo, utenze, spese postali, eccetera).

Figura 3.18 – Rendicontazione

Tipo di contratto	Funzione svolta			
	Responsabile	Altro Personale	Coordinatori	Rilevatori
Dipendenti CCIAA/URCCIAA/PA	5	0	0	9
Addetti assunti secondo le tipologie contrattuali di cui al comma 4, art. 50 L. 122/2010 (escluso i contratti di somministrazione di lavoro)	0	0	0	0
Somministrazione lavoro	0	0	0	0
Totale addetti dell'Ufficio di censimento	5	0	0	9
Composizione di rete Sgr	1	0	4	37

### 3.8 L'assistenza in fase di esercizio e la documentazione di supporto

Per garantire agli operatori tutta l'assistenza necessaria nella gestione di tutti i sistemi web (acquisizione, monitoraggio e Portale della 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 in formato online e direttamente in formato *pdf*, dalla homepage del SGRCIS.

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 (SGRCIS), 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 *helpdesk* 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 avevano delle funzioni dedicate, che permettevano in molti 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 del personale interno che avesse conoscenza sia del sistema informatico sia 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.

### **3.9 Il Sistema di gestione degli accertamenti della violazione dell'obbligo di risposta - SGS**

Nell'infrastruttura informatica del censimento è stato progettato e sviluppato anche un componente a supporto della gestione del protocollo informatico per la gestione degli accertamenti della violazione dell'obbligo di risposta.

Questa componente è stata progettata sulla base di un'applicazione *legacy* precedentemente utilizzata per lo stesso scopo; la reingegnerizzazione si è resa necessaria per garantire l'accesso multiutente e la gestione della concorrenza. È stata sviluppata un'applicazione web in linguaggio Java, conforme agli standard ed ai *framework* utilizzati per lo sviluppo di SGRCIS.

L'applicazione era fruibile agli utenti Istat ed ha consentito la gestione condivisa della documentazione cartacea relativa agli accertamenti della violazione dell'obbligo di risposta.

L'applicazione inoltre ha fornito supporto a più generiche funzioni, ad esempio la gestione della documentazione dei vettori postali coinvolti nelle attività del censimento (Poste Italiane, POSTEL). Di seguito (Prospetto 3.1) uno schema dei ruoli previsti per gli utenti con accesso al sistema.

### Prospetto 3.1 – Tabella ruolo funzioni abilitate

RUOLO UTENTE	Descrizione delle funzioni abilitate
Amministratore	Gestione delle utenze per tutte le tipologie di indagini censuarie
Operatore di segreteria	Segreteria
Operatore Sanzioni	Sanzioni Censimento industria e servizi
	Sanzioni Censimento popolazione
	Sanzioni Censimento agricoltura
Supervisore	Sanzioni Censimento industria e servizi
	Sanzioni Censimento popolazione
	Sanzioni Censimento agricoltura
	Segreteria

La principale finalità del sistema SGS è stata quindi l'archiviazione e la gestione dei documenti che all'interno del sistema erano logicamente catalogati in due tipologie:

- Protocolli: documenti di segreteria;
- Fascicoli: documenti relativi alle sanzioni.

Dopo una fase preliminare, consistente nella scansione del documento cartaceo e la sua conversione in formato *pdf* (attività svolta offline esternamente al contesto dell'applicativo web), i documenti venivano poi archiviati sul sistema tramite una funzione applicativa di *upload*.

A ciascun documento inserito nel sistema SGS venivano poi associate le informazioni necessarie ad una classificazione degli stessi per facilitarne l'individuazione tramite una funzione di ricerca. Ad esempio veniva memorizzato la tipologia di documento (es. ingresso o uscita), la provenienza (es. interna od esterna all'Istituto), le indicazioni sul mittente e la modalità di ricezione (es. fax, PEC, raccomandata, posta ordinaria, eccetera).

Alle attività connesse alle procedure di accertamento veniva anche associata l'indagine di riferimento (CIS, Censimento popolazione, eccetera) e, esclusivamente per il CIS, venivano associate anche le informazioni anagrafiche del questionario in lavorazione, recuperate dalla base dati dell'indagine.

I documenti, una volta inseriti nel sistema, erano resi disponibili agli utenti per iniziare il processo di lavorazione che era definito e controllato da una variabile di stato associata a ciascun documento. Il Prospetto 3.2 mostra gli stati previsti per il controllo della lavorazione.

### Prospetto 3.2 – Descrizione stati

STATI INIZIALI	STATI INTERMEDI	STATI FINALI
Inserito	In lavorazione, Lavorato KO	Annullato, Lavorato OK, Sospeso, Inoltrato, Annullato per nuovo accertamento

Ogni stato rappresentava una fase di lavorazione e il flusso era definito e guidato dalle varie funzioni del sistema. Tra le altre:

- Correzione e Annullamento Protocollo;
- Lavorazione Fascicoli;
- Invio Fascicoli.

Di particolare interesse è la funzione Invio Fascicoli: questa funzione (abilitata ai soli supervisori) procedeva alla creazione di archivi nel formato *zip* composti dai fascicoli relativi agli accertamenti

la cui lavorazione era terminata. Ciascun archivio generato era corredato da un file *x/s* contenente informazioni di riepilogo dei file in esso contenuti; le informazioni erano:

- 1) Nr. Pratica;
- 2) Cod. Questionario;
- 3) Codice Utente;
- 4) Legale Rappresentante;
- 5) Ragione Sociale;
- 6) Indirizzo;
- 7) CAP;
- 8) Comune;
- 9) Provincia;
- 10) Protocollo Diffida;
- 11) Data Diffida;
- 12) Modalità Invio;
- 13) Codice Ufficio Territoriale;
- 14) Data Accertamento;
- 15) Data Scadenza;
- 16) Tipologia Rispondente.

Per ciascun archivio creato, il sistema SGS associava il documento al proprio archivio, memorizzava la posizione su *filesystem* dell'archivio, la data di creazione. Inoltre, il sistema consentiva di inviare gli archivi agli utenti che si sarebbero occupati delle procedure di notifica di contestazione della sanzione; l'invio dell'archivio comportava il cambio di stato dello stesso da "non inviato" a "inviato". L'invio consisteva nel depositare l'archivio in un *filesystem* condiviso in modo sicuro tra diverse strutture.



## PARTE SECONDA

### IL SISTEMA INFORMATICO DI ACQUISIZIONE E GESTIONE DEL CENSIMENTO DELLE ISTITUZIONI PUBBLICHE



## Capitolo 1

### Il sistema di acquisizione e di gestione<sup>13</sup>

In questo capitolo vengono descritti gli aspetti principali del sistema informatico a supporto del Censimento delle Istituzioni pubbliche 2011 (CensIP), le cui linee guida sono stabilite nel Piano Generale del Censimento (PGC).

Il CensIP è inserito nell'ambito del Censimento dell'industria e dei servizi 2011 (CIS) e presenta alcune scelte organizzative originali per un censimento.

In questo capitolo saranno evidenziate alcune di queste scelte, ed in particolare: (a) la rilevazione è formata da due fasi distinte e sequenziali, (b) la compilazione del questionario è possibile solo tramite interfaccia web, (c) non esiste una rete di rilevatori. Questi punti attribuiscono al CensIP delle caratteristiche uniche rispetto alle altre rilevazioni del CIS ed anche la definizione e sviluppo del sistema informatico ne risulta influenzata.

Considerando il flusso principale delle informazioni acquisite, la rilevazione è articolata in due distinte fasi; la Prima fase raffina la lista precensuaria delle unità di rilevazione (unità istituzionali – UI) e crea la lista iniziale delle loro unità locali (UL), mentre la Seconda fase considera come unità di rilevazione sia le UI sia le UL e ha un duplice obiettivo: (1) raffinamento della lista delle UL e (2) compilazione dei questionari di UI ed UL.

Il vincolo (b) è quello che definisce meglio il ruolo del SI nell'ambito del CensIP; i censimenti tradizionali prevedono una molteplicità di canali per la compilazione e l'invio dei questionari da parte dei rispondenti (es. questionario cartaceo vs. questionario elettronico). Questo implica che il sistema informatico rappresenta uno degli strumenti messi a disposizione ma non l'unico possibile. Il vincolo (b) afferma la centralità del SI che richiede elevate caratteristiche di robustezza e affidabilità. Inoltre, non potendo fare alcuna ipotesi sul livello minimo di aggiornamento delle piattaforme di lavoro delle istituzioni pubbliche, sono stati considerati limiti tecnologici molto blandi (es. versioni dei sistemi operativi); infine i test sul sistema sono stati effettuati sia su piattaforme proprietarie sia *open source*. Nella Prima fase, è stato chiesto alle UI coinvolte nel CensIP di compilare un unico questionario, mentre nella Seconda fase, le UI hanno avuto in carico la compilazione di più questionari; per questo motivo, il sistema della Seconda fase permette la compilazione di un questionario UI e di n questionari UL dove  $n \geq 1$ , e questo numero dipende dalla singola UI e non è costante:  $n = n(t, UI)$ . Questo implica che ogni UI debba poter svolgere attività di monitoraggio sullo stato di avanzamento della compilazione degli  $n+1$  questionari ad essa assegnati.

<sup>13</sup> A cura di Vincenzo Spinelli. Autori del capitolo:

- Paragrafo 1.1: Vincenzo Spinelli;
- Paragrafo 1.2: Andrea Nunnari;
- Paragrafo 1.3: Stefania Taralli;
- Paragrafo 1.4: Francesco Cuccia (Paragrafi 1.4.1 e 1.4.2), Vincenzo Spinelli (Paragrafo 1.4.3);
- Paragrafo 1.5: Stefano De Santis;
- Paragrafo 1.6: Vincenzo Spinelli (Paragrafo 1.6.1), Francesco Cuccia (Paragrafo 1.6.2);
- Paragrafo 1.7: Rolando Duma (Paragrafo 1.7.1), Stefano De Santis (Paragrafo 1.7.2);
- Paragrafo 1.8: Cristiano Francone;
- Paragrafo 1.9: Bernardo Palazzi, Manlio Paterniti.

Il vincolo (c) viene descritto utilizzando la figura dell'*Incaricato del coordinamento della rilevazione* che rappresenta il referente della UI per il CensIP e che può delegare ad altri referenti la compilazione dei singoli questionari UL nel caso di enti di grandi dimensioni:  $n = O(10^k)$ ,  $k \in \{2,3\}$ .

Il sistema informatico è stato progettato utilizzando un ciclo di vita evolutivo e strumenti non proprietari (es. Java e PHP); questo ha comportato lo sviluppo di funzioni a complessità crescente (sia come numero sia come flessibilità) mentre lo stesso sistema informatico era già operativo. In altre parole è possibile descrivere una relazione d'ordine dei sistemi in cui il più recente (Seconda fase) eredita funzioni e dati da quello precedente (Prima fase). Il sistema informatico a supporto della Prima fase (SI1) include alcune funzioni di base: sistema di monitoraggio, questionario elettronico e autenticazione degli utenti del sistema. Queste funzioni vengono arricchite ed ampliate nel sistema informatico della Seconda fase (SI2): monitoraggio esteso, gestione dinamica delle utenze/questionari, caricamento online/offline dei microdati. Poiché SI2 ha ereditato (per continuità) le informazioni acquisite con SI1, si può affermare che vale la relazione di inclusione "SI1  $\subset$  SI2".

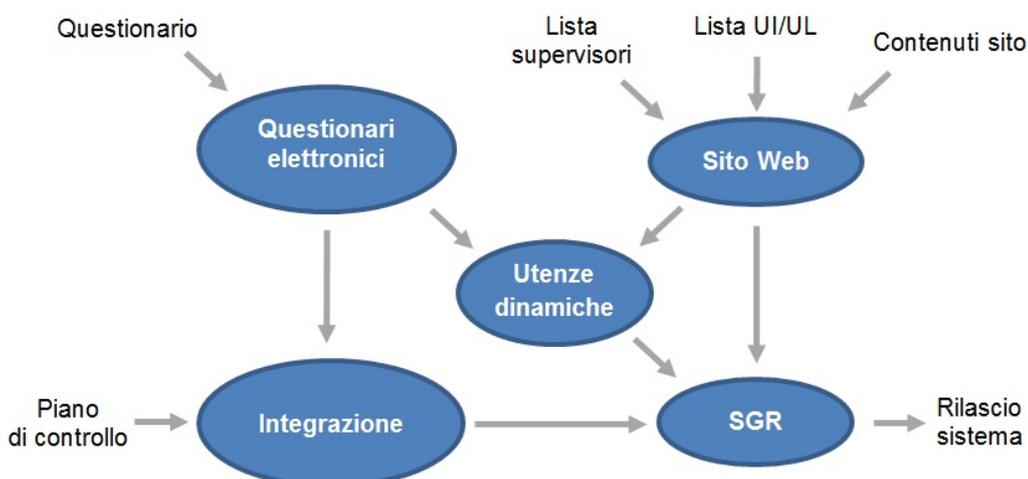
Per garantire la riservatezza delle informazioni inserite nel sistema sono stati adottati vari livelli di sicurezza, alcuni dei quali sono da considerarsi standard (es. utenze con password criptate e protocolli di sicurezza "*Hypertext Transfer Protocol Secure*" o HTTPS). Per avere un quadro oggettivo sul livello di sicurezza e prestazioni del sistema, sono state pianificate varie sessioni di test specifici, utilizzando strumenti dedicati e non proprietari (es. *Apache JMeter*).

La progettazione di un sistema informatico a supporto di attività censuarie deve tener conto delle rigide scadenze previste. A questo si aggiunge che il CensIP è caratterizzato da due fasi ovvero due rilevazioni integrate ma distinte, separate da un breve periodo di transizione (circa dieci giorni).

L'utilizzo di tecniche di tipo CPM (*critical path method*) nelle prime fasi di progettazione del sistema ha permesso l'individuazione di "*logical end point*" e *milestone* sia nelle attività per lo sviluppo del sistema informatico sia nell'organizzazione complessiva del CensIP. In Figura 1.1 sono descritti i blocchi di livello più alto che sono stati individuati per definire i GANTT del progetto. Questi schemi sono al netto delle attività di migrazione da SI1 ad SI2 e di misurazione del livello di sicurezza e prestazioni del sistema.

Per valutare le possibili criticità dinamiche del sistema informatico è stata formulata una modellazione statistica delle code di accesso al sistema; questi modelli hanno quindi influenzato lo sviluppo di SI1. Le verifiche sulla dinamica degli accessi durante lo svolgimento della Prima fase ha permesso di raffinare il modello e di estenderlo alla progettazione di SI2.

Figura 1.1 – Schema CPM per il Censimento delle istituzioni pubbliche



Il resto del capitolo si articola come segue: SI1 viene descritto brevemente nel Paragrafo 1.2, mentre nel Paragrafo 1.3 vengono presentati la piattaforma ed il *layout* utilizzato per le funzioni esterne del sistema (es. autenticazione e documentazione online) per SI1 e SI2. I paragrafi successivi sono rivolti alla descrizione esclusiva di SI2: nel Paragrafo 1.4 viene descritto il sistema di formazione online messo a disposizione su una piattaforma esterna e nei Paragrafi 1.5, 1.6 e 1.7 vengono presentati le principali funzioni interne di SI2. Nel Paragrafo 1.8 vengono introdotte alcune funzioni non pubbliche di SI2 utili per lo scarico dei microdati e le successive analisi e processi di consolidamento. Infine, il Paragrafo 1.9 descrive l'organizzazione dei documenti (normativi, statistici e tecnologici) presenti sul sito web ed il Paragrafo 1.10 illustra le attività relative alla misurazione delle prestazioni del sistema e del suo livello di sicurezza.

## 1.1 Le due fasi del Censimento

Il sistema informatico del CensIP è stato sviluppato nell'ambito della piattaforma ufficiale dell'Istituto per l'acquisizione dei dati (*Indata*). Questo ha permesso di ereditare alcune funzioni di *Indata* e, quindi, di non doverle progettare *ex novo*.

La scelta di questa piattaforma è dovuta anche a ragioni di continuità, in quanto alcune rilevazioni sugli enti pubblici già utilizzavano questa tecnologia ed è stato quindi naturale riproporla nell'ottica dei censimenti. La funzione di login e cambio password e il *layout* del sito sono stati definiti in continuità con queste rilevazioni. In particolare, la gestione delle utenze (login, password) segue la politica standard di *Indata*: il sistema obbliga a modificare la password iniziale al primo accesso.

Altre funzioni, originali per il CIS, sono state sviluppate e integrate in questo ambiente standard.

Potendo considerare SI1 come il nucleo di base per SI2, le funzioni sviluppate per SI1, al netto di quelle già presenti in *Indata*, sono il sistema di monitoraggio (M1) e il questionario elettronico (Q1). Queste funzioni sono state sviluppate utilizzando *Struts*, la piattaforma più popolare per lo sviluppo di applicazioni web basate su Java. La piattaforma *Struts* si basa su un'architettura "Model-View-Controller" (MVC).

Gli utenti del sistema sono stati classificati in due classi non sovrapposte: monitori e rispondenti. I rispondenti erano le singole UI, mentre i monitori erano i dipendenti dell'Istituto (sia della Sede centrale sia degli Uffici territoriali) preposti alle operazioni di controllo e supporto delle fasi censuarie.

I rispondenti della Prima fase erano le sole UI ed il questionario in questa fase era composto da due sezioni, di cui una statica (anagrafica UI e incaricato UI) ed una dinamica (lista UL). Il questionario poteva essere compilato solo in modalità online, quindi non lo si poteva scaricare localmente sul PC del rispondente; comunque era sempre possibile effettuare salvataggi parziali che consentivano una compilazione diluita nel tempo. Il valore aggiunto del questionario è la sezione dinamica delle UL, dove veniva richiesto l'elenco puntuale delle singole sedi e dei loro referenti in ottica Seconda fase. Non potendo prevedere a priori le dimensioni della lista con sufficiente precisione, la compilazione di questa sezione è stata resa possibile in due modalità:

- Online – ogni UL veniva inserita in modo interattivo; questa modalità risultava essere utile nei casi di UI con un limitato numero di UL (es.  $O(10^2)$ ); inoltre la compilazione guidata (campi con risposte vincolate: es. provincia/comune) e i controlli interattivi (controllo del formato e/o sintassi dei valori inseriti: es. telefono/e-mail) permettevano di minimizzare gli errori di digitazione.
- Offline – la lista delle UL poteva essere preparata localmente sul PC del rispondente e caricata nel questionario come un file di testo avente la struttura "comma separated value" (csv).

La modalità offline presupponeva che il rispondente disponesse di strumenti automatici per la creazione della lista (es. UI di grandi dimensioni con una gestione logistica centralizzata) ed escludeva i controlli previsti nel caricamento online.

Ogni questionario era caratterizzato da uno stato di lavorazione tra tre possibili: “*mai aperto*”, “*in compilazione*” e “*chiuso*”. La transizione da uno stato all'altro veniva gestita automaticamente dal sistema in base all'attività svolta, secondo una serie di regole formali:

- mai aperto - tutti i questionari erano in questo stato all'avvio della rilevazione;
- in compilazione - se veniva effettuato un salvataggio di dati (parziale) il questionario veniva portato in questo stato intermedio. Per poter salvare le informazioni era comunque necessario che fossero coerenti con le regole di controllo implementate nel questionario;
- chiuso - era lo stato finale del questionario e implicava che nessuna altra modifica poteva essere effettuata sul questionario stesso. Il sistema permetteva la chiusura solo quando erano stati compilati tutti i campi vincolanti e i dati immessi erano coerenti con le regole di controllo. La chiusura di un questionario veniva certificata con la creazione di una scheda di fine compilazione.

Una transizione non esplicitamente prevista dal sistema (es. dallo stato “chiuso” allo stato “in compilazione”) veniva considerata una forzatura ed era effettuabile solo dagli utenti monitori del sistema.

L'ambiente di lavoro dei rispondenti era ristretto al solo questionario elettronico, mentre i monitori accedevano a un ambiente più articolato che permetteva di effettuare analisi sull'andamento della rilevazione. Il sistema distingueva tra monitori di carattere locale (Uffici territoriali Istat), che avevano una visibilità regionale dei rispondenti, e globale (Sede centrale Istat), che avevano la visione su tutti i rispondenti. Questa differenza (impostazione libera/bloccata dei filtri regionali) era l'unica prevista nell'ambiente di monitoraggio. Sono stati implementati una batteria di filtri sulla lista dei rispondenti per permettere di calcolare le liste di UI che soddisfacevano i criteri impostati dai monitori (es. le UI che (a) non hanno mai aperto il questionario, (b) appartengono ad una specifica regione/provincia/comune e (c) appartengono ad una specifica tipologia di ente). Sulla base della lista risultante venivano effettuate le normali operazioni di sollecito/ritorno sulle UI. I monitori avevano la possibilità di accedere ai questionari nel caso vi fossero problemi puntuali non risolvibili attraverso i normali canali di assistenza agli utenti (es. e-mail e telefono).

Con l'avvio della Seconda fase del censimento, l'URL della Prima fase veniva disattivato e reso operativo l'URL della Seconda fase (<https://indata.istat.it/cis2011ipfase2/>); per ridurre al minimo i disagi dei rispondenti, l'URL della Prima fase rimandava automaticamente al nuovo sistema censuario SI2 e le informazioni acquisite nella Prima fase risultavano precaricate nella banca dati di SI2, in questo modo i rispondenti trovavano alcuni campi dei questionari della Seconda fase già precompilati. Per motivi di sicurezza e per sottolineare l'inizio di una nuova rilevazione tutte le password correnti degli utenti sono state rigenerate nel nuovo sistema SI2, mantenendo l'utenza dove possibile: le scuole statali di ogni ordine e grado e le UL erano rispondenti della sola Seconda fase, quindi l'utenza (login, password) non veniva ereditata dalla Prima fase.

## 1.2 Il sito Web

Il sistema di acquisizione dati è stato ospitato sulla piattaforma standard di *Web Data Collection* Istat (*Indata*). Seguendo il modello funzionale di *Indata* sono stati implementati sia il sito web di *front end* per entrambe le fasi del processo di raccolta (agli indirizzi <https://indata.istat.it/cis2011ipfase1/> e <https://indata.istat.it/cis2011ipfase2/>), sia il sito di *back-end* per il monitoraggio e l'amministrazione dei *front end* (<https://indata.istat.it/cis2011Ges/>).

Tutti e tre i nodi implementano sicurezza a livello di canale tramite crittografia del protocollo di trasferimento (HTTPS) e sicurezza a livello applicativo attraverso il processo di autenticazione *two-steps* basato su credenziali logiche:

I rispondenti e i monitori ricevevano dall'Istat un codice utente numerico di lunghezza variabile e una password iniziale di dieci caratteri alfanumerici (composta da un prefisso alfabetico, identificativo dell'indagine, di tre caratteri più sette caratteri alfanumerici generati randomicamente). La password iniziale era valida soltanto per il primo accesso, che si risolveva nella procedura di registrazione. In questa fase al rispondente veniva chiesto di inserire le credenziali ricevute e di scegliere una password personale (minimo otto, massimo venti caratteri alfanumerici) che da quel momento sostituiva quella iniziale. La password personale veniva registrata nel sistema con criptazione basata su *hashing* a sedici caratteri esadecimali (64 bit).

Ogni successivo accesso avveniva attraverso la nuova coppia di credenziali.

Sul sito di *front end*, rivolto a rispondenti e monitori, le funzioni offerte ricadono sostanzialmente in due classi:

- 1) funzioni pubbliche: tutte le funzionalità accessibili senza previa identificazione. Si tratta generalmente di funzioni informative, descrittive e di comunicazione con i gestori dell'indagine e la rete territoriale:
  - Indagine - home page contenente la descrizione della rilevazione;
  - Documentazione - risorse documentali (legali o informative) legate alla rilevazione e al processo di raccolta, disponibili in formato *pdf*;
  - Istruzioni - risorse informative sia statiche (file) che ipertestuali (link a risorse web di *e-learning*);
  - Contatti - liste *pdf* di referenti con informazioni di contatto (telefono, e-mail). In questa pagina era presente anche un *modulo di comunicazione* per l'invio controllato di richieste di assistenza per problemi relativi all'utilizzo del sito, alla compilazione del questionario o problemi relativi ai rispondenti delle istituzioni scolastiche. Il modulo era accessibile senza effettuare login, quindi veniva richiesto l'inserimento di informazioni base sull'utente, quali utenza e indirizzo e-mail, per confrontarli successivamente con quelle già presenti nella base dati del sito. Alle diverse tipologie di richiesta erano associati diversi indirizzi e-mail a seconda delle competenze necessarie per l'assistenza;
  - Glossario - la pagina metteva a disposizione un file *pdf* con le definizioni dei termini contenuti nel questionario;
  - Domande e Risposte - FAQ in formato *pdf*.
- 2) funzioni private: funzionalità accessibili solo tramite autenticazione. Si tratta di tutte le funzionalità legate al processo di raccolta vero e proprio (compilazione del questionario, monitoraggio) o all'aggiornamento di dati legati all'utenza:
  - Accesso - attraverso questo modulo era possibile eseguire i due step dell'autenticazione per poi accedere ai moduli del questionario. Il passaggio dal contesto del sito web (sviluppato su piattaforma LAMP (*Apache/PHP*) ai moduli del questionario (*Tomcat/Java*) veniva effettuato tramite POST HTML delle variabili necessarie alla condivisione virtuale della sessione fra i due contesti e al riconoscimento del profilo attribuito all'utente (livello e privilegi di lettura/scrittura sulle risorse);

- Password - l'utente poteva effettuare alcune semplici operazioni che riguardavano la propria password:
  - ripristino password iniziale: nel caso di smarrimento della password personale o per avvicendamento fra persone fisiche afferenti alla stessa utenza, era possibile richiedere al gestore dell'indagine il ripristino della password iniziale oppure effettuarlo direttamente tramite procedura disponibile sul sito dell'indagine. La procedura di ripristino, con relativa ripetizione della fase di registrazione, si rendeva necessaria in quanto le password personali venivano registrate nei database in formato criptato e non erano quindi recuperabili.
  - cambio password personale: funzione web standard che permetteva di cambiare la password personale precedentemente inserita.

Sul sito di *back-end*, non esposto e rivolto agli amministratori della piattaforma *Indata*, le funzioni offerte ricadevano in due classi:

- viste sulle tabelle dei metadati e dei dati di processo della rilevazione:
  - dati anagrafici degli utenti;
  - lista delle utenze e dello stato di autenticazione;
  - motore di ricerca sulle tabelle;
- procedure di modifica dei metadati e dei dati di processo della rilevazione:
  - caricamento lista utenti e relativi dati anagrafici;
  - creazione batch delle password iniziali;
  - ripristino delle password iniziali degli utenti.

### 1.3 In-formazione on-line (la piattaforma di e-learning)

Nella Seconda fase del censimento, tra le funzionalità del sito web è stata inserita una piattaforma di *e-learning*, accessibile ai rispondenti e alla rete di rilevazione mediante autenticazione, e rispondente alla duplice esigenza di veicolare le indicazioni metodologiche e operative essenziali per la qualità della rilevazione e di consentire il monitoraggio delle attività di auto-istruzione degli utenti.

Come è ampiamente dimostrato, la formazione e l'informazione dei soggetti coinvolti nelle rilevazioni sono due leve fondamentali per la prevenzione dell'errore non campionario. Nel caso della Rilevazione sulle istituzioni pubbliche, le innovazioni nei contenuti informativi, nella tecnica di raccolta dei dati e nell'organizzazione della rete censuaria hanno ulteriormente rafforzato l'esigenza da parte dell'Istat di utilizzare anche queste leve per garantire la qualità della rilevazione.

Infatti, nonostante l'assenza di rilevatori ed il ricorso generalizzato alla auto-compilazione online, aspetti vincolanti rispetto alla numerosità e complessità degli item, i questionari contenevano ampie batterie di quesiti articolati e inediti e richiedevano l'applicazione di classificazioni piuttosto estese e complesse<sup>14</sup>. Inoltre nelle istituzioni funzionalmente o territorialmente più articolate un "Incaricato del coordinamento della rilevazione presso l'istituzione pubblica" ha avuto il compito di assicurare la completezza e correttezza della rilevazione e di coordinare il personale interno

<sup>14</sup> Ad esempio la Classificazione delle attività economiche – Ateco 2007 o la Classificazione dei servizi – Cofog – 2009.

coinvolto nella fornitura dei dati<sup>15</sup>. A fronte dei maggiori compiti dei rispondenti, e in assenza di rilevatori e di organi censuari intermedi, le funzioni di monitoraggio e di assistenza tecnico-metodologica degli Uffici Regionali di Censimento dell'Istat (URC) hanno assunto importanza strategica.

Per ridurre l'onere di risposta delle Istituzioni pubbliche e sostenere le funzioni degli URC si sono predisposte strategie di informazione e formazione – “*InFormazione IP*” - diversificate in base al *target* da raggiungere: *blended learning* (formazione in presenza e a distanza) rivolto alle unità istituzionali complesse; *e-learning* puro per le unità meno complesse<sup>16</sup>.

Gli strumenti e i servizi di *e-learning* sono stati implementati sulla piattaforma *open source Dokeos*, da tempo riconosciuta dal CNIPA come affidabile e ad elevata usabilità. Per tutti i rispondenti, inoltre, i principali metadati di supporto alla corretta e completa compilazione sono stati integrati direttamente nel questionario web. Nell'ambiente *e-learning* sono stati pubblicati per il download o la consultazione online i seguenti *learning objects* (Figura 1.2):

- questionari e guide ipertestuali: versioni elettroniche e navigabili dei materiali illustrativi, arricchite da link alla documentazione di indagine, alla normativa, alle definizioni e classificazioni ufficiali;
- brevi moduli di auto-formazione sugli aspetti salienti della rilevazione, auto-consistenti e corredati da ipertesti, fruibili in maniera modulare;
- *video tutorial* illustrativi della navigabilità e delle funzioni degli ambienti di lavoro web, basati sull'utilizzo integrato di immagini, grafica illustrativa e testi esplicativi;
- linkografia e *repository* per la consultazione e il download di tutto il materiale illustrativo e di supporto alla rilevazione e di numerosi documenti di approfondimento.

<sup>15</sup> Piano del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit; Istat, Circolare n. 1/IPS del 16 luglio 2012.

<sup>16</sup> L'Info&For per la Rilevazione delle istituzioni pubbliche è stato pianificato dalla Task Force interdipartimentale per la progettazione e organizzazione della formazione agli organi di censimento del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit (Del. DICA 2/2012). L'*e-learning* è stato progettato e realizzato da un Gruppo di Lavoro costituito a tale scopo nell'ambito della DCSR (Del. 23/DIQR/2012).

Figura 1.2 – Home page del corso e-learning per la rilevazione delle istituzioni pubbliche



La progettazione dei contenuti e dell'architettura del corso si è improntata a criteri di utilità, rilevanza e parsimonia ed ha cercato di assicurare caratteristiche di granularità fine e di modularità dei contenuti, in accordo ai principi basilari della formazione agli adulti e alle linee guida per l'*e-learning* nella Pubblica Amministrazione.

Tra i prodotti realizzati un breve cenno di approfondimento va rivolto ai questionari ipertestuali e ai *video tutorial*, che grazie alla multimedialità e alla comunicazione visiva semplificano la comprensione e la memorizzazione delle informazioni.

I primi sono un strumento utile alla formazione d'aula e all'autoformazione, ma si prestano anche alla consultazione rapida e sono risultati tra i prodotti e-learning più apprezzati anche nell'ambito di altre rilevazioni<sup>17</sup>, soprattutto perché permettono di esaminare in maniera agile, integrata e selettiva tutta la documentazione disponibile anche con riferimento ad un singolo quesito. Nel caso in esame queste potenzialità informative dello strumento sono state valorizzate con un accurato lavoro di progettazione concettuale che ha portato ad individuare le aree informative e gli item dei questionari di indagine su cui richiamare l'attenzione dell'utente: contenuto informativo delle sezioni, variabili di diffusione *core*, quesiti filtro, quesiti critici, variabili di classificazione. A ciascuno di questi elementi, corrispondente graficamente ad una "parola calda" o ad un'area "sensibile" del questionario ipertestuale, è stato associato un diverso simbolo grafico, indicante la specifica funzione della meta-informazione collegata (Figura 1.3)

La stessa progettazione logica ha permesso di individuare il sottoinsieme dei metadati da integrare nel questionario elettronico, sotto forma di *tool tips* o di link a risorse internet, quali ausili essenziali alla compilazione online.

Figura 1.3 – Questionario ipertestuale

<sup>17</sup> Per esempio il 6° Censimento dell'agricoltura, come documentato dall'indagine di valutazione dell'*e-learning* effettuata *ex post*.

PAGINA H 2 3 4 5 6 7 8 9 

### MODELLO PER LE UNITÀ ISTITUZIONALI

**1 IDENTIFICAZIONE DELL'UNITÀ ISTITUZIONALE E DELL'INCARICATO DEL COORDINAMENTO DELLA RILEVAZIONE**

1.1 Controllare i dati riportati nel riquadro superiore (Quadro A) e indicare in quello sottostante (Quadro B) le integrazioni e le eventuali variazioni:

Verificare che le informazioni di carattere anagrafico contenute nel riquadro, riferite alla data della compilazione, siano quelle della sede unica o centrale dell'istituzione pubblica; in caso contrario apportare le modifiche necessarie.

Codice meccanografico

Codice meccanografico (Solo per le scuole)

Forma giuridica

Acronimo

**legenda dei simboli**

-  breve spiegazione di una definizione o di un quesito
-  breve spiegazione della regola di compilazione per un quesito
-  link alla Guida alla compilazione, al glossario, ad altri documenti illustrativi. Apre i documenti nel punto in cui trattano il quesito (in un'altra finestra per agevolare la consultazione)
-  breve spiegazione degli obiettivi informativi della sezione
-  link a materiale ufficiale (classificazioni, regolamenti, normativa) di riferimento per la compilazione, o a risorse esterne (necessario collegamento a internet)

I *video tutorial* sono stati realizzati per avvicinare gli utenti del sito web e del sistema di acquisizione dati, in particolare i rispondenti, all'utilizzo delle funzioni informative e operative di interesse, presentando una panoramica generale del sistema *Indata*, le funzioni di compilazione del questionario online, il monitoraggio e la gestione delle unità locali, la gestione dei controlli e l'invio definitivo del questionario, la gestione delle password. Gli *storyboard* sono stati pensati per guidare *step by step* alla navigazione dei menu, all'individuazione, comprensione e corretto utilizzo delle funzioni, dei comandi e dei controlli presenti nel sistema, e all'agevole gestione delle maschere. A questo scopo in fase di *video editing* le sequenze video catturate nel corso di sessioni di lavoro dimostrative sono state integrate con elementi grafici e testuali volti a richiamare l'attenzione sugli aspetti chiave delle procedure e ad illustrarli (Figura 1.4).

**Figura 1.4 - Fotogramma del videotutorial sulla compilazione del questionario**

SEZIONE 1 - IDENTIFICAZIONE DELL'UNITÀ ISTITUZIONALE E DELL'INCARICATO DEL COORDINAMENTO DELLA RILEVAZIONE

1.1 Controllare i dati anagrafici riportati ed apportare le integrazioni e le eventuali variazioni

Codice Fiscale  Partita IVA

Codice Meccanografico

Denominazione

Acronimo

Forma giuridica

Indirizzo

I campi in grigio non sono compilabili: si attivano se le risposte fornite ai quesiti precedenti lo richiedono.

La fruizione dell'*e-learning* è stata monitorata settimanalmente attraverso il registro della piattaforma *Dokeos*<sup>18</sup>. I report sono stati trasmessi sistematicamente agli URC che hanno potuto avvalersene, insieme ai report di monitoraggio della rilevazione, per valutare *in itinere* la qualità della rilevazione e per individuare eventuali azioni di controllo sui dati o di ritorno sui rispondenti. Sulla stessa piattaforma di *e-learning* è stato allestito anche un *repository* per gli URC contenente tutti i supporti d'aula e le utilities necessarie ad organizzare, realizzare e monitorare le riunioni di *Info&for* a livello territoriale<sup>19</sup>.

L'*e-learning* è stato progettato e implementato interamente *in house*. Il progetto non ha generato spese per acquisto di software e licenze, o costi diretti per attività di messa in esercizio e diffusione mentre ha garantito una maggiore e più accessibile offerta di formazione e informazione a supporto della qualità della rilevazione.

## 1.4 Il sistema di gestione

Il sistema di monitoraggio prevedeva tre livelli di accesso cui corrispondevano altrettante aree territoriali con ampiezza che andava da tutto il territorio nazionale al singolo territorio regionale alle singole unità locali. Questi livelli corrispondevano a tre categorie di utenti:

- l'Istat;
- gli Uffici territoriali dell'Istat;
- le unità istituzionali.

### 1.4.1 Monitoraggio Istat, scuole ed uffici regionali

All'Istat competeva il livello di controllo più ampio che consentiva un monitoraggio dello stato di avanzamento della compilazione per le unità istituzionali e le unità locali su tutto il territorio nazionale. Le informazioni potevano essere filtrate per territorio e stato di compilazione.

Sono stati previsti i seguenti filtri (Figura 1.5):

- specifico codice utente/questionario;
- territorio (regione, provincia, comune);
- stato di compilazione dei questionari per le unità istituzionali e relative unità locali (mai aperto, in compilazione, chiuso);
- stato di completamento di tutti i questionari afferenti all'unità istituzionale;
- forma giuridica;
- tipo istituzione (istituto scolastico o non scolastico).

La ricerca produceva un elenco che per ciascuna istituzione pubblica riportava le seguenti informazioni:

- il codice della unità istituzionale;
- la denominazione della unità istituzionale;
- la provincia sede della unità istituzionale;
- il comune sede della unità istituzionale;
- lo stato del questionario della unità istituzionale;

<sup>18</sup> L'amministrazione della piattaforma e i report di monitoraggio sono stati curati dal DICA/IAC.

<sup>19</sup> Il "Kit d'aula" conteneva: Locandina e programma "tipo" della riunione; fogli di firma e attestati; brevi presentazioni personalizzabili sui punti essenziali e sugli aspetti critici della rilevazione; questionari ipertestuali; dispense per i partecipanti. Nel periodo ottobre-novembre 2012, gli Uffici Regionali di Censimento hanno tenuto almeno 128 riunioni in presenza o in *web conference* in 17 diverse regioni.

- il numero di unità locali associate alla unità istituzionale;
- il numero di questionari delle unità locali nello stato di mai aperto, in compilazione e chiuso.

L'Istat aveva inoltre la possibilità di ottenere un maggiore dettaglio delle informazioni assumendo il ruolo di una qualsiasi unità istituzionale sul territorio nazionale e monitorare in questo modo lo stato di compilazione delle singole unità locali fino ad accedere ai singoli questionari.

Agli Uffici territoriali Istat competeva il monitoraggio dello stato di avanzamento della compilazione per le unità istituzionali e le unità locali appartenenti al territorio regionale corrispondente. Essenzialmente disponevano delle stesse funzionalità dell'Istat ma ristrette all'ambito regionale. Potevano quindi filtrare le informazioni per territorio e stato di compilazione e anche scendere ad un maggior dettaglio assumendo il ruolo di una unità istituzionale fino ad accedere ai singoli questionari.

**Figura 1.5 – Filtri di selezione Istat e Uffici territoriali Istat**

**DATA ULTIMO AGGIORNAMENTO DELLE INFORMAZIONI: 14/03/2013 08:36:48**

**Sistema di monitoraggio ISTAT/UR**

Benvenuto/a Utente Nazionale Livello 4 (Istat4)

[Gestione richieste unità locali](#) | 
 [Identificativi UI e password](#) | 
 [Rapporto riassuntivo](#) | 
 [Esci](#)

**Monitoraggio dello stato di avanzamento di compilazione delle unità istituzionali**

Codice utente

Regione    
 Provincia    
 Comune

Stato Questionario Unità Istituzionale    
 Stato Questionari Unità Locali

Stato Completamento Unità Istituzionale

Forma giuridica

Tipo istituzione

**747 record trovati - visualizzati 1 - 25** [Prima/Prec.] 1, 2, 3, 4, 5, 6, 7, 8 [Succ./Ultima]

Utente	Denominazione	Provincia	Comune	Stato UI	#UL	#UL mai aperte	#UL in compilazione	#UL chiuse	Apri	Log
CHIS012006	* L. SAVOIA*	Chieti	Chieti	mai aperto	4	2	0	2		
PEIS00600B	* A. VOLTA*	Pescara	Pescara	chiuso	3	0	0	3		
11116181	Abruzzo Promozione Turismo - Azienda di Promozione Turistica Regionale	Pescara	Pescara	chiuso	25	0	0	25		
11150810	AGENZIA NAZIONALE PER LO SVILUPPO DELL'AUTONOMIA SCOLASTICA - EX IRRE ABRUZZO - GESTIONE COMMISSARIALE	L'Aquila	L'Aquila	chiuso	1	0	0	1		

### 1.4.2 Il monitoraggio delle unità istituzionali

Alle unità istituzionali competeva il monitoraggio dello stato di compilazione per le unità locali che ne facevano parte. Anche per questo livello era possibile filtrare le informazioni per territorio per stato di compilazione e stato dell'unità locale.

Erano previsti i seguenti filtri (Figura 1.6):

- codice utente/unità locale;
- territorio (regione, provincia, comune);
- stato questionario (mai aperto, in compilazione, chiuso);
- fase di accesso (fase1, fase2);

- stato unità locale (attiva, disattivata).

La ricerca produceva un elenco che per ciascuna unità locale riportava le seguenti informazioni:

- codice utente;
- denominazione;
- comune;
- stato compilazione questionario;
- fase di ingresso;
- referente;
- numero errori rilevati.

L'unità istituzionale poteva inoltre accedere al questionario della singola unità locale. Un'ulteriore funzionalità di cui disponevano le unità istituzionali era quella di richiedere l'attivazione di nuove utenze per nuove unità locali non previste al momento dell'avvio della rilevazione. Le richieste erano memorizzate in una coda per essere sottoposte al vaglio dell'Istat.

Figura 1.6 – Filtri di selezione Unità Istituzionali

DATA ULTIMO AGGIORNAMENTO DELLE INFORMAZIONI: 14/03/2013 11:36:45

**Sistema di monitoraggio dell'unità istituzionale**

Benvenuto/a Identificativi UL e password | Stato compilazione | Esci

**Richiesta nuove unità locali**

Numero nuove unità locali

\* Non ci sono richieste in attesa di esecuzione

**Questionari unità istituzionale**

Nota: si ricorda che l'Unità Istituzionale deve compilare sia il questionario per l'Unità Istituzionale che il questionario per l'Unità Locale Principale indicati in questa sezione

**Monitoraggio dello stato di avanzamento di compilazione delle altre unità locali**

Codice utente

Regione

Provincia

Comune

Stato Questionario

Fase di accesso

Stato Unità

3 record trovati, visualizzati tutti i record Pag. 1

Riapri	Rimuovi/Integra	Utente	Denominazione	Comune	Stato UL	Fase	Referente	#Errori	Apri	Log
<input type="button" value="Riapri"/>	<input checked="" type="checkbox"/>	CHTF01201P	L. SAVOIA	Chieti	chiuso	2	Pica Giuseppe	0	<input type="button" value="Apri"/>	<input type="button" value="Log"/>
<input type="button" value="Riapri"/>	<input checked="" type="checkbox"/>	CHTF012514	L. SAVOIA CORSO SERALE	Chieti	mai aperto	2		0	<input type="button" value="Apri"/>	<input type="button" value="Log"/>

### 1.4.3 La gestione dinamica delle utenze e/o dei questionari

Nella Prima fase del censimento gli utenti abilitati ad accedere in SI1 sono classificabili in (a) monitori (locali e non) e (b) rispondenti UI; ad ogni rispondente UI era chiesto di compilare un questionario. Siano LM ed LR le liste di utenti di queste due classi: nella Prima fase esse si possono considerare statiche e vale la relazione  $O(10^2) = |LM| \ll |LR| = (10^4)$ .

Gli utenti della Seconda fase di SI2 risultavano essere classificabili in modo più articolato, ed in particolare: (a) monitori (locali e non), (b) rispondenti UI e (c) rispondenti UL. Questa classificazione ammetteva una eccezione per un importante sottogruppo di utenti di tipo (c), equiparata alla classe (b). Infatti le scuole statali di ogni ordine e grado, anche se definite come unità locali del Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), sono state considerate di tipo UI vista la loro numerosità e specificità. In particolare, gli istituti scolastici sono stati considerati di tipo (b) e le sedi delle scuole afferenti a ciascun istituto scolastico sono state considerate rispondenti di tipo (c) e quindi le due liste della Prima fase, sono state affiancate, nella Seconda fase, da altre due liste: LU (lista unità locali) ed LS (lista scuole). Al contrario della Prima fase, queste liste non sono da considerarsi statiche; poiché l'obiettivo della Seconda fase non era ristretto solo alla compilazione dei questionari ma anche al raffinamento delle liste ottenute dalla Prima fase, SI2 è stato progettato con funzioni di aggiornamento automatizzato di queste liste utenti e relativi questionari.

È possibile definire su queste liste una relazione d'ordine parziale ('<') che permette di attribuire quali funzioni accessibili in SI2 siano effettivamente utilizzabili da uno specifico utente, vedi Paragrafi 1.5.1 e 1.5.2. Partizionando LM nelle due liste LM1 (monitori locali) e LM2 (monitori non locali) si ottengono le due sequenze:

$$\begin{aligned} LU < LR < LM1 < LM2 \\ LU < LS < LM2 \end{aligned}$$

Siano  $u, v$  utenti di SI2 e sia  $f(u)$  (rispettivamente  $f(v)$ ) l'insieme delle funzioni assegnate ad  $u$  (rispettivamente a  $v$ ); allora valgono le relazioni:

- 1)  $u < v \Rightarrow f(u) \subset f(v)$ ;
- 2)  $u$  e  $v$  appartengono alla stessa lista  $\Rightarrow f(u) = f(v)$ .

In altre parole, ogni anello di queste catene è caratterizzato da funzioni specifiche/locali e da tutte le funzioni definite negli anelli sottostanti; ogni utente in LM2 poteva accedere a tutte le funzioni di SI2, mentre gli altri utenti avevano una visibilità sempre più ridotta. Gli utenti in LU avevano l'insieme di funzionalità minimo: compilazione di un questionario. Le relazioni di cardinalità tra un livello e quello sottostante è di tipo 1:n; in altre parole ad ogni utente di un livello corrispondevano (ha la visibilità di) più utenti del livello sottostante. In termini di funzioni di SI2 si può anche dire che ogni utente disponeva delle funzioni locali al livello e di tutte le altre dei livelli sottostanti. Questa organizzazione gerarchica implicava che i monitori non solo potessero controllare lo svolgimento delle operazioni censuarie ma potessero anche entrare nel merito del censimento forzando le modifiche ai questionari se necessario. Il legame tra utenti non si limitava alla visibilità delle funzioni, in quanto ogni utente dipendeva da almeno un utente che lo precedeva nella relazione d'ordine e valevano i seguenti vincoli:

- 1) gli utenti di LM2 non hanno utenti di livello superiore, mentre gli utenti in LU non hanno utenti di livello inferiore (estremi delle due catene);
- 2) ogni utente in LU è sempre e solo legato ad un utente in  $LR \cup LS$ ;
- 3) ogni utente  $r$  in  $LR \cup LS$  ha almeno un utente in LU legato ad esso; questo utente  $u$  in LU viene definito UL principale di  $r$  ed è sempre e solo legato ad  $r$ . L'utente  $r$  può ampliare l'insieme di utenti UL ad esso collegati, richiedendo di attivare nuove utenze di tipo UL; analogamente può eliminare gli utenti di tipo UL ad esso collegato, tranne l'utenza UL principale;

- 4) ogni utente può eliminare gli utenti collegati ad esso nella relazione d'ordine; la cancellazione si propaga a cascata ai livelli inferiori. Gli utenti in LM1 e LM2 sono statici e non possono essere cancellati;
- 5) ogni utente in  $LM1 \cup LS$  dipende da tutti gli utenti in LM2.

Ad ogni utente in  $LR \cup LS$  è stato associato (direttamente) un questionario di tipo UI, mentre ad ogni utente in LU un questionario di tipo UL. Ogni utente di SI2 era responsabile della compilazione di tutti i questionari ad esso legati direttamente e/o indirettamente (tramite la relazione d'ordine); è utile notare che solo i monitori, in  $LM1 \cup LM2$ , non avevano responsabilità diretta di compilazione ma solo indiretta (dovevano controllare i rispondenti).

I vincoli descritti implicano che SI2 doveva poter attivare/disattivare utenze dinamicamente e conseguentemente il questionario ad esso collegato. Per ottenere ciò l'insieme degli utenti in SI2 doveva essere definito in modo più ampio: sia L l'insieme degli utenti riconosciuti in SI2 allora L è l'unione di tre sottoinsiemi disgiunti  $L = Lna \cup La \cup Lx$ . Questi tre insiemi sono così definiti:

- 1) Lna – lista degli utenti non assegnati;
- 2) La – lista utenti attivi;  $La = LM \cup LR \cup LS \cup LU$ ;
- 3) Lx – lista utenti cancellati.

Gli utenti in L sono gli unici riconosciuti in SI2, quindi L è l'universo di SI2, ma solo gli utenti in La possono entrare nel sistema (funzione login nel sito web) ed aprire un questionario. Durante le operazioni censuarie questi tre insiemi venivano continuamente aggiornati da SI2 mantenendo costante L: un utente poteva solo spostarsi da un sottoinsieme all'altro. Questo spostamento di utenti (transizione di stato) avveniva secondo i seguenti criteri:

- 1) Lna è da considerarsi di lunghezza potenzialmente infinita ed è possibile solo togliere utenti che devono essere messi in La;
- 2) gli utenti in La possono essere spostati (solo) in Lx; se l'utente non è un monitore, quindi non in LM, allora anche tutti gli utenti ad esso collegati devono essere spostati in Lx;
- 3) analogamente, gli utenti in Lx possono essere spostati (solo) in La, insieme a tutti gli utenti collegati.

L'insieme di questi criteri descrive un semplice grafo delle transizioni di stato:  $Lna \rightarrow La \leftrightarrow Lx$ .

Le funzioni disponibili in SI2 per la gestione effettiva di queste transizioni di stato sono:

- 1) l'attivazione/disattivazione di UI veniva gestita fuori da SI2, a seguito di specifiche richieste dei monitori, ed impattava direttamente sulla banca dati;
- 2) l'attivazione/disattivazione di UL già assegnate era una operazione in tempo reale implementata come un interruttore ON/OFF posto su ogni elemento della lista UL visualizzata nell'ambiente del rispondente;
- 3) l'attivazione di utenze UL, a seguito di richieste di utenti UI, era una operazione potenzialmente pericolosa ed è stata gestita attraverso un meccanismo di *hand-shake* in cui l'utenza UI "prenotava"  $n \geq 1$  nuove utenze (e questionari) UL, ed uno dei suoi monitori approvava/rifiutava la prenotazione. In caso di approvazione le nuove utenze venivano attivate.

In S2 sono state previste altre liste di utenti utili per le attività di test e formazione sul sistema; questo insieme di utenti di tipo UI ed UL erano reali in quanto accedevano a tutte le funzioni dei rispondenti, ma erano allo stesso tempo virtuali in quanto esterni alle funzioni di monitoraggio.

La formazione su SI2 è stata erogata dagli URC mentre SI2 era operativo ed accessibile, quindi questi utenti realizzavano un'architettura *sandbox* anche se non a fini di sviluppo. Inoltre, essendo

integrati in S2 e non in macchine virtuali sono stati utilizzati durante le sessioni di test e sicurezza, vedi Paragrafo 1.10, fornendo risultati reali.

## 1.5 Il questionario elettronico

Una delle caratteristiche più innovative del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit, è stata la scelta di effettuare la raccolta dei dati, relativamente al Censimento delle istituzioni pubbliche facendo ricorso in via esclusiva ad una applicazione web che ha permesso alle istituzioni di compilare i questionari direttamente online. Complici ormai diffusi livelli di cultura generale ed informatica, nonché la progressiva digitalizzazione della società e della Pubblica Amministrazione in particolare, i tempi sono stati ritenuti maturi per una raccolta di dati totalmente demandata a mezzi elettronici. La scommessa era che i potenziali problemi di copertura legati all'utilizzo dei soli mezzi informatici fossero allo stato attuale più che compensati da tutta una serie di vantaggi rispetto ai tradizionali questionari cartacei usati in precedenza (minori costi, immediata disponibilità dei dati al salvataggio parziale o definitivo, maggiore coerenza dei dati raccolti). I tassi di risposta ottenuti hanno confermato la bontà delle scelte effettuate.

Per soddisfare tale fine è stato realizzato un sistema web integrato, interamente progettato e sviluppato dall'Istat, a sostegno di tutte le attività del processo di raccolta e di cui i questionari online costituiscono solo una parte. La necessità di una forte integrazione con il sistema di gestione del censimento da un lato, l'esigenza di tempi di sviluppo rapidi dall'altro, ha portato alla scelta di un impianto misto in parte sviluppato *ad hoc*, in parte basato sull'utilizzo/estensione di un software generalizzato per l'acquisizione di dati via web (*GINO++ Gathering Information Online*) anch'esso sviluppato internamente all'Istituto con il supporto di consulenti esterni.

Il sistema web del Censimento delle istituzioni pubbliche è perciò un sistema complesso articolato in diversi componenti essenziali:

- 1) il Sistema di Gestione della Rilevazione utilizzato dalla rete territoriale dell'Istat e dal personale di coordinamento per gestire le varie attività;
- 2) il sito web di accesso e supporto alla rilevazione, con tutta la documentazione e il materiale a disposizione dei compilatori e degli operatori;
- 3) la struttura di *data capturing*, articolata in una duplice fase di rilevazione (la prima per definire informazioni sulle istituzioni a livello *frame*; la seconda di raccolta delle informazioni rilevanti).

Prima di esporre gli strumenti di acquisizione, si espongono in sintesi le principali caratteristiche della struttura di *data capturing*, illustrando in particolare l'insieme delle soluzioni per migliorare la qualità dei dati e la forte integrazione con il sistema di gestione del censimento.

Come ricordato in precedenza, l'intera architettura del sistema è stata progettata e sviluppata in Istat. Questa esigenza era motivata principalmente dalla stretta integrazione richiesta tra il processo di acquisizione dei dati (questionari) e il sistema di gestione relativo al censimento, una cooperazione funzionale che sarebbe stata praticamente impossibile da attuare attraverso il solo riutilizzo/ampliamento degli strumenti esistenti per gestione indagini e *survey software tool*. La stretta integrazione tra le due applicazioni ha consentito il monitoraggio in tempo reale di tutto il processo di raccolta, riducendo notevolmente il divario tra la raccolta dei dati e le ulteriori elaborazioni, in particolare il processo di confronto e di riallineamento tra il questionario dell'istituzione e i questionari delle sue unità locali.

Per ogni questionario compilato, tutte le informazioni fornite erano immediatamente disponibili per l'ulteriore elaborazione e controlli di varia natura. Il questionario è perfettamente integrato con il sistema di gestione del censimento: a ogni cambiamento, lo stato del rispondente veniva subito aggiornato nella banca dati centrale e reso visibile (con una latenza massima di trenta minuti) agli operatori. Gli stessi operatori potevano usare direttamente il questionario per eseguire operazioni di inserimento/correzione dati, aggiornando così lo stato del questionario, così come potevano intervenire direttamente sugli stati di lavorazione per riaprire/chiedere i questionari. Per ciascun utente erano resi disponibili funzioni di monitoraggio (tempi di svolgimento dell'indagine, qualità dei dati, indicatori sintetici sugli errori, eccetera) e funzioni di riepilogo (informazioni su registrazioni e anagrafiche dei rispondenti; lo stato di avanzamento di ogni questionario; la percentuale dei questionari inviati per ciascuna unità istituzionale e dei campi compilati; altre informazioni di dettaglio).

Soluzioni sofisticate sono state adottate nella progettazione interna per ridurre il carico al database anche in presenza di un elevato numero di utenti contestualmente collegati all'applicazione, nonché per la convalida dei dati a garanzia della coerenza dell'informazione (all'interno di ciascun questionario e fra questionari delle unità istituzionali e le sue unità locali). La raccolta di dati online ha portato a una significativa riduzione del lavoro su campo, con evidenti vantaggi economici ed organizzativi. Inoltre, i controlli effettuati sui dati inseriti dagli utenti nel momento del salvataggio ha notevolmente migliorato la qualità dei dati raccolti.

### **1.5.1 La realizzazione della struttura di data capturing**

La struttura di *data capturing* del Censimento delle istituzioni pubbliche è un complesso di strumenti utilizzati nelle due fasi di indagine.

Fase prima:

- il questionario online della prima fase;

Fase seconda:

- i due questionari online della Seconda fase (unità istituzionali e unità locali);
- il modello MT1, somministrato ai non rispondenti della Prima fase;
- il tracciato record per il caricamento dei dati delle unità locali appartenenti a unità istituzionali "complesse".

Di seguito ci si concentra sulla descrizione degli strumenti di *data capturing* relativi alla Seconda fase.

### **1.5.2 La progettazione dei questionari e la tecnologia utilizzata**

I questionari online delle unità istituzionali e delle unità locali rappresentano senza dubbio la parte più importante della struttura di *data capturing*, il core del sistema di acquisizione. Il loro sviluppo è stato effettuato avvalendosi di *GINO++* (*Gathering Information Online*), software generalizzato per l'acquisizione di dati via web, realizzato in PHP come strumento *general purpose*. Robustezza e flessibilità del software sono aumentate nel tempo con l'estendersi dei suoi ambiti di utilizzo; la notevole disponibilità di opzioni di disegno del quesito e di funzioni nella realizzazione del questionario, ha consentito grande libertà di manovra.

*GINO++* supporta il disegno del questionario mediante interfaccia grafica. È così possibile definire e verificare *in itinere* il layout del questionario per eventuali controlli ed apportare modifiche, nonché l'inserimento dei controlli (*soft* e *hard*) definiti dal piano di *check* dell'indagine. Il componente di base del disegno del questionario è la singola variabile che può:

- assumere varie forme (es. *radio button*, *checkbox*, eccetera) e prevedere l'inserimento dell'informazione con immissione libera o obbligata (tramite lista di item o completamento automatico);
- essere soggetta a regole (quali controlli di coerenza, calcolo automatico, vincoli di obbligatorietà, salti);
- essere corredata da aiuti alla compilazione (suggerimenti, *box note*, *tooltip*, link a fonti esterne).

Il diverso utilizzo di questi elementi, in particolar modo i primi due, ha una grande influenza sulla qualità dei dati raccolti ed è direttamente responsabile della maggior qualità ottenibile mediante acquisizione web.

Una volta definite, le variabili si possono aggregare in quesiti secondo varie configurazioni topologiche: riga, colonna, tabella (ogni colonna è una variabile), matrice (ogni cella è una variabile). I quesiti sono raccolti in Sezioni al fine di:

- consentire la compilazione sequenziale o a salti;
- visualizzare il questionario su pagine web distinte o su un'unica pagina;
- permettere il salvataggio e/o il controllo delle regole al cambio di sezione o a richiesta.

Le ripetizioni di sezione sono gestite mediante i Moduli. Tali strutture modulari hanno trovato utilizzazione nel questionario delle unità locali riguardo le schede relative ai servizi offerti. Poiché questi potevano essere presenti in numero variabile (da zero a cinque) non noto a priori, il disegno del questionario ha richiesto la definizione dinamica del numero dei moduli ("schede di servizio") in base alle risposte fornite da ciascun rispondente (una scheda per ciascun servizio indicato). In questo modo le unità locali erano chiamate a fornire informazioni relativamente ad un numero di moduli pari alla tipologia di servizi offerti.

Con *GINO++* la libertà di disegno è enfatizzata dal numero di funzioni disponibili, che aumenta le possibilità dell'utente nel disegno del questionario (precaricamento di dati statistici da fonti esterne o da precedenti rilevazioni, riutilizzo di metadati e classificazioni ufficiali, monitoraggio continuo). Lo *storing* di tutti i metadati in uno schema relazionale garantisce la consistenza e la immediata disponibilità per qualunque successivo utilizzo. In particolare modifiche al questionario e correzioni ai testi possono essere eseguiti in qualsiasi momento, anche dopo la distribuzione dell'applicazione, senza intaccarne la consistenza e il funzionamento. In particolare lo *storing* dei dati raccolti viene effettuato ad ogni salvataggio, sia parziale che definitivo, sui server dell'Istituto rendendoli immediatamente disponibili. L'acquisizione dati è effettuata con registrazione normalizzata in ambiente Oracle, consentendo di supervisionare costantemente ogni rispondente nella fase di compilazione del questionario, nonché di monitorare in tempo reale l'andamento della rilevazione nel suo complesso. La tabella dei microdati usata per la registrazione normalizzata presenta una consueta forma verticalizzata (Prospetto 1.1), nella quale l'informazione non assume una forma  $n * m$  (individui \* variabili), ma in cui ogni riga individua univocamente una variabile. Tale tabella assume la seguente forma compatta: l'unione dei campi non costanti (di fatto la *primary key* della tabella) corrisponde 1-1 con una variabile statistica.

### Prospetto 1.1 – Schema della tabella verticalizzata dei microdati

NOME_CAMPO	Tipo	Descrizione
ID_RISPONDENTE_RILEV	NUMBER(6) [.....]	identificativo univoco del rispondente
ID_MODULO	NUMBER(1)	identificativo delle sezioni modulari
ID_VARIABILE	NUMBER(3) [.....]	identificativo di variabile
PROGRESSIVO	NUMBER(1)	identificativo di modalità - variabili <i>multi-choice</i>
VALORE	VARCHAR2(400)	campo di <i>storing</i> dell'informazione
ID_CLASSIFICAZIONE	NUMBER(2)	identificativo di classificazione

La libertà di disegno garantita dall'applicativo ha consentito l'implementazione dei due questionari nel sostanziale rispetto formale (*layout* del questionario) e sostanziale (qualità dei dati e piano di *check*) delle esigenze informative prospettate dal censimento.

#### 1.5.3 Il trasferimento dei dati dalla Prima fase del censimento, modello MT1 e popolamento del database

Il Censimento delle istituzioni pubbliche si è articolato in due fasi. La Prima fase tendeva a dare una prima risposta ai problemi di lista che caratterizzano il settore pubblico (ovvero definire esattamente l'universo di osservazione) ed effettuare l'aggiornamento dei dati relativi alla sezione anagrafica delle istituzioni coinvolte (ragione sociale, forma giuridica, indirizzo e dati di contatto; individuazione dell'incaricato del coordinamento della rilevazione) nonché ad acquisire l'elenco delle unità locali afferenti ciascuna unità istituzionale.

Tale fase è stata assolta con il ricorso a un primo questionario online somministrato a tutte le unità della lista precensuaria. Per tutte le unità rispondenti si è proceduto poi all'elaborazione dei dati raccolti, all'esclusione delle unità fuori campo osservazione e al successivo popolamento del questionario relativo alla Seconda fase con i dati raccolti nella prima. Le unità non rispondenti sono invece state interessate da una ulteriore fase di raccolta dei dati attraverso la somministrazione di un apposito questionario elettronico sviluppato su *xls* e scaricabile dal sito del Censimento in formato *xls* e *xlsx*, (modello MT1).

Il modello MT1 è un *Excel UserForm*, un questionario offline del tutto identico al modello della prima fase (ossia con i medesimi controlli di formato sui campi, compilazione automatica, menu a tendina, convalida dinamica dei menu a tendina, eccetera) che le istituzioni non rispondenti dovevano prelevare dal sito Istat, compilare e inviare ad un apposito indirizzo e-mail. Qui una procedura scritta *ad hoc* eseguiva il recupero massivo dei dati puntando alla *repository* dei *form xls*, inseriva l'informazione nei database Oracle (valorizzando i questionari della Seconda fase) e infine produceva una reportistica utile a monitorare lo stato di avanzamento della compilazione del bacino dei non rispondenti della prima fase.

Indipendentemente dalla modalità di acquisizione, la raccolta dei dati nella Prima fase ha comportato la loro organizzazione per la successiva fruizione nella Seconda fase. Inoltre, per quello che riguarda le Scuole (che costituivano *de facto* un censimento a sé stante nell'ambito del Censimento delle istituzioni pubbliche), sia le liste come le informazioni anagrafiche provenivano da fonte amministrativa (cedolini paga e 770 del servizio buste paga del MEF, fonti degli enti locali per le scuole di competenza delle regioni/province a statuto speciale).

La necessità è stata perciò quella di partire da *dataset* simili, ma eterogenei, per definire un

*dataset* omogeneo volto a soddisfare le seguenti finalità:

- creazione delle anagrafiche in *GINO++*;
- un file per il precaricamento dei dati raccolti nella Prima fase (o provenienti da fonte amministrativa) nei questionari dei rispondenti tematici (unità istituzionali e unità locali, laddove presente l'informazione), richiedendo perciò la normalizzazione e verticalizzazione dei dati in una forma coerente con la registrazione dei microdati in *GINO++*.

Anche in questo caso, poiché il task è stato assolto a più riprese, si è optato per la scrittura di una procedura che automaticamente generasse l'informazione richiesta a partire dagli eterogenei *dataset* di partenza.

#### **1.5.4 La definizione delle procedure di acquisizione dati per le unità istituzionali "complesse" e caricamento dei tracciati offline**

Benché il *core* dell'acquisizione dei dati fosse rappresentato dai questionari online, di fatto l'impianto di *data capturing* della Seconda fase proponeva una compilazione *mixed mode*, in cui veniva offerta una scelta alternativa nella compilazione alle unità istituzionali "complesse", quelle cioè caratterizzate da un così elevato numero di unità locali da rendere necessario l'utilizzo di forme di acquisizione differenti: è stato il caso dei Ministeri, di grandi enti locali (come il Comune di Milano o di Roma), gli enti di previdenza e assistenza sociale, eccetera.

La necessità era dunque rappresentata dalla gestione di rispondenti con una notevole mole di dati. Come nei casi precedenti, si è definito un *form* di acquisizione e una relativa procedura automatica che gestisse il recupero dei dati per tutte le unità locali indicate sul tracciato record, il caricamento (*upload*) dei dati direttamente da file a livello di intero questionario, o più selettivamente a livello di singola sezione o di quesito. In fase di acquisizione dei file nel DB, la procedura effettuava un controllo dei dati sia a livello formale (coerenza con la specifica della variabile – lunghezza e tipo, numerica o alfanumerica), sia con le stesse regole di coerenza utilizzate per l'immissione mediante digitazione. Venivano infine aggiornate tutte le informazioni necessarie alle funzioni di monitoraggio e controllo.

Il tracciato record è stato definito in maniera tale da rispecchiare fedelmente la rappresentazione delle variabili nell'omologo questionario online (ordine, *label*, definizione, lunghezza, eccetera), ma in forma compatta (Prospetto 1.2).

#### **Prospetto 1.2 – Schema dei questionari ricostruiti**

IDENTIFICATIVO UNIVOCO DEL RISPONDENTE (UNITÀ LOCALE)	Variabile 1	[...]	Variabile <i>n</i>
1			
[...]			
m			

La procedura prevedeva inoltre il flusso reciproco di informazioni tracciato-database (cioè da tracciato a database e viceversa da database a tracciato). Era perciò possibile non solo valorizzare il database tramite il tracciato, ma anche l'opposto, cioè il tracciato con le informazioni precedentemente immesse nel database. In tal modo l'unità istituzionale poteva controllare e revisionare i dati avendo una visione d'insieme dei dati immessi per ciascuna unità locale.

È stata così possibile la gestione di notevolissime masse di dati, che in taluni casi hanno riguardato la compilazione di diverse centinaia di unità locali per ciascuna unità istituzionale

interessata. In questo caso alla riduzione del *burden* statistico ha corrisposto senz'altro un aumento della qualità dell'informazione raccolta, evitando problemi di integrità legati al *data-entry* e consentendo una efficace revisione dei dati inseriti.

## 1.6 L'integrazione tra i sistemi

I blocchi fondamentali di SI2 sono sostanzialmente tre: login, monitoraggio e questionario elettronico. Questi blocchi sono stati pensati come insieme di dati e funzioni (classi) indipendenti ma integrate tra loro. L'integrazione è avvenuta sia a livello dei dati sia a livello funzionale. Il resto del paragrafo fornisce una breve descrizione di questa integrazione.

### 1.6.1 Il modello dei dati

L'infrastruttura tecnologica di base è composta dal DBMS Oracle integrato da *Hibernate*, una soluzione ORM *open source* per applicazioni Java.

Questa scelta è stata motivata dal supporto che *Hibernate* fornisce nell'interrogazione e lettura dei microdati e questo diminuisce sensibilmente i tempi di sviluppo delle applicazioni.

Il sistema informatico del censimento si basa su un modello relazionale dei dati formato da tre blocchi fondamentali, che rispecchiano le funzionalità di base del sistema: accesso/login, monitoraggio e questionari.

Questi blocchi sono caratterizzati da un nucleo comune e da strutture locali (quindi informazioni locali non condivise). Questo nucleo comune è formato dalla lista di base delle utenze del sistema; questa lista è costituita in prima approssimazione dalle terne (utente, tipo utente, password iniziale) ed è stata resa non modificabile per tutto il censimento. In ogni blocco questa lista veniva arricchita da molte altre informazioni locali che potevano essere aggiornate dalle normali attività svolte da e nel sistema.

Alcuni esempi di informazioni locali sono:

- accesso/login: la lista di base era integrata da informazioni non condivise quali la password corrente, la data di ultima modifica e di ultimo accesso al sistema, vedi Paragrafo 1.2;
- monitoraggio: la politica di visibilità delle informazioni del sistema è stata implementata descrivendo l'insieme delle utenze come vertici di alcuni grafi. Essendo questi grafi orientati, il legame tra i vertici espresso dalle coppie (utente, utente genitore) è un esempio di informazione locale al blocco del monitoraggio. Utilizzando una terminologia standard, l'utente genitore (da cui dipende ogni singolo utente) era anche definito *root*. In alcuni casi, *root* non era definito (es. monitori LM2) ma era possibile forzare un valore introducendo un'utenza virtuale a cui si potevano ricondurre tutte le altre utenze (es. Istat);
- questionario: ad ogni utenza veniva associato sempre e solo un questionario e lo stato del questionario (es. mai aperto, in compilazione) era una tipica informazione locale a questo blocco.

Il blocco del monitoraggio ha svolto in prima approssimazione due funzioni: (a) aggiornare tutte le informazioni utili alla sincronizzazione delle attività nel sistema e (b) fornire le liste di base per le attività di monitoraggio (a vari livelli) sullo stato di avanzamento delle attività del censimento. In particolare la funzione (a) ha permesso di realizzare la gestione dinamica dei questionari (es. aggiungere, cancellare, ripristinare) che si rifletteva sulle relative utenze di UI e/o UO. Le tabelle utenti avevano come campi principali la coppia (utenza, *root*) che permetteva di realizzare le operazioni tra utenti come operazioni su grafi. In particolare si riconosceva la struttura di grafi orientati non ciclici (DAG) e, essendo presenti relazioni di ordine parziale, si sono ottenuti grafi di

comparabilità rendendo espliciti gli spigoli latenti nella struttura.

Il blocco dei questionari è stato strutturato per il sistema *G/NO++*, e conteneva tutti i microdati inseriti nel sistema. Questi microdati venivano memorizzati secondo una struttura verticalizzata e visti come locali al blocco.

Osservando che la normale attività di compilazione dei questionari portava alla modifica di informazioni utili al monitoraggio (es. anagrafica enti e/o referenti), si è reso necessario prevedere un sistema di *feedback* tra le tabelle dei questionari e quelle del monitoraggio avente lo scopo di aggiornare lo stato di avanzamento delle maschere del monitoraggio. L'operazione inversa (il monitoraggio aggiorna i questionari) non era prevista tranne in un unico caso: essendo i questionari relativi alle istituzioni scolastiche gli unici aventi il campo "CODICE MECCANOGRAFICO" attivo, quando venivano attivati nuovi questionari (funzione attiva solo nel monitoraggio) le tabelle di *G/NO++* venivano modificate dai processi del monitoraggio specificando la tipologia di istituzione.

Poiché le attività di sincronizzazione delle informazioni richiedono processi onerosi, con una complessità polinomiale rispetto al numero di questionari compilati, allora si è deciso di creare processi temporizzati. Con il progredire delle attività del censimento, questo ha avuto come effetto di minimizzare la durata dei possibili tempi di attesa per gli utenti finali e di offrire ai monitori informazioni aggiornati con tempi di latenza noti (generalmente due/tre ore).

### 1.6.2 Il sistema di accesso

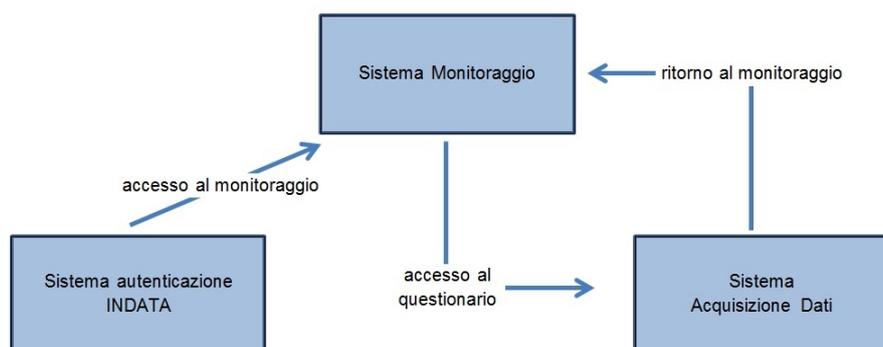
Il sistema di acquisizione e gestione per il Censimento delle istituzioni pubbliche è composto di tre sottosistemi indipendenti che cooperano tra di loro:

1. il sistema di autenticazione di *Indata*
2. il sistema di monitoraggio
3. il sistema di acquisizione dati di GINO

Si tratta di strutture software completamente autonome sviluppate con tecnologie differenti che sono state integrate mediante opportune interfacce al fine di dare all'utente la sensazione di interagire con unico sistema fornito di un unico punto di accesso e con passaggio da un componente all'altro del tutto trasparente.

I sottosistemi sono stati tra loro sincronizzati in maniera tale che dopo essere stato autenticato un utente avesse accesso solamente alle funzionalità riconosciute al proprio ruolo e potesse accedere in modo appropriato ai dati e ai questionari. L'interazione fra i sottosistemi è avvenuta come indicato in Figura 1.7.

Figura 1.7 – Interazione fra i sottosistemi di acquisizione



Ciascun utente era fornito di proprie credenziali (identificativo e password) di accesso al sistema. La prima fase di autenticazione era gestita dal sistema *Indata* che in caso di esito positivo reindirizzava l'utente al sistema di monitoraggio. Il passaggio del controllo comportava la comunicazione dell'identificativo dell'utente interno al sistema di monitoraggio, il suo ruolo e i relativi permessi di lettura e/o scrittura affinché fossero attivate le corrette funzionalità.

Il sistema di monitoraggio rappresenta il cuore dell'intero sistema. Oltre a consentire ai monitori l'acquisizione delle informazioni sullo stato di avanzamento della compilazione dei questionari a vari livelli di dettaglio, è stato anche il punto di accesso ai questionari stessi da parte dei rispondenti.

Tramite il sistema di monitoraggio, tutti gli utenti hanno avuto la possibilità di accedere ai singoli questionari secondo una ampiezza territoriale decrescente al decrescere del livello attribuito. L'utente di livello più alto (Istat) poteva accedere ai questionari delle unità istituzionali e relative unità locali di tutta Italia. L'utente di livello successivo (Ufficio territoriale Istat) poteva accedere ai questionari delle unità istituzionali e relative unità locali della regione di appartenenza. Ogni unità istituzionale poteva accedere al proprio questionario e a quelli delle unità locali corrispondenti. Ogni unità locale poteva accedere solo al proprio questionario. L'Istat, gli Uffici territoriali Istat e le unità istituzionali hanno avuto anche funzione di monitoraggio rispettivamente a livello nazionale, regionale e per le unità locali afferenti. Quando veniva richiesto l'accesso ad un singolo questionario, il sistema di monitoraggio reindirizzava l'utente al sistema di acquisizione dati. Il passaggio comportava la comunicazione degli identificativi dell'utente e del questionario oltre che del punto di ritorno interno al sistema di monitoraggio cui restituire il controllo all'uscita dal questionario.

Al sistema di acquisizione dati spettava il compito di visualizzare il questionario e acquisire i dati inseriti.

## **1.7 Lo scarico dei dati**

La fase di analisi e il consolidamento statistico delle informazioni acquisite in SI2 hanno richiesto la realizzazione di interfacce non pubbliche per lo scarico dei microdati. Le funzioni di scarico realizzate sono riconducibili sostanzialmente a due distinte modalità: (a) online ed (b) offline. La prima modalità permette di selezionare e filtrare a priori le informazioni da analizzare e questo è possibile farlo tramite interfacce web disponibili sulla rete locale dell'Istituto (intranet); la seconda modalità permette di vedere i microdati direttamente nell'ambiente relazionale di SI2. La scelta di utilizzare una modalità rispetto all'altra dipende dal tipo di analisi statistica che si sceglie: lo scarico interattivo era rivolto ad analisi limitate nel numero di questionari (es. una singola provincia/regione) da effettuare in locale, mentre la seconda ha permesso di avere la massima visibilità sia dei microdati sia dei metadati ed è stata quindi utile in presenza di procedure non interattive. Il resto del paragrafo è dedicato alla descrizione delle due modalità di scarico.

### **1.7.1 La modalità online**

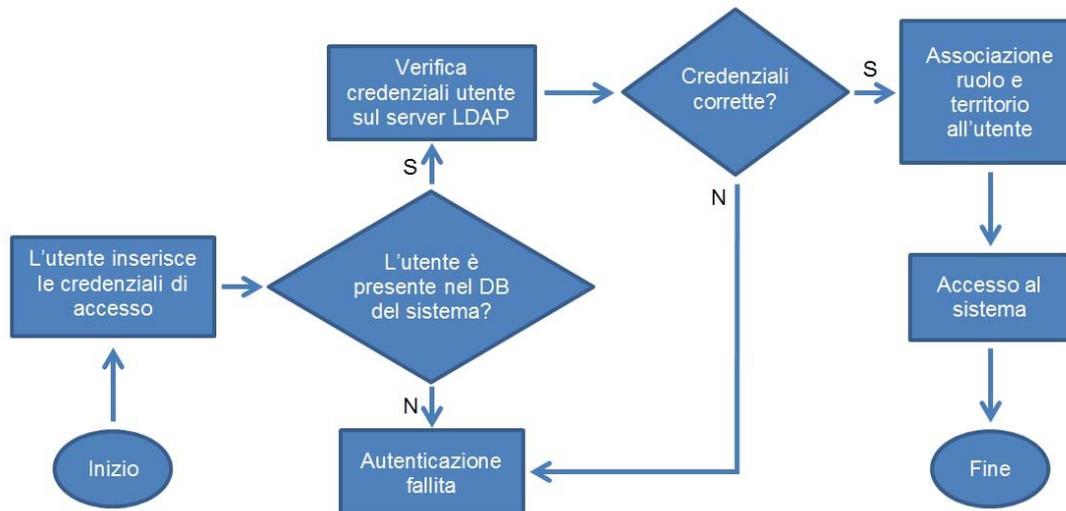
Lo scarico dei dati per le istituzioni pubbliche ha introdotto alcuni spunti di interesse nell'infrastruttura informatica di gestione del censimento. Questi sono derivati principalmente da due aspetti che si differenziavano rispetto alla generale architettura del SGRCIS: la base di utenti interessata allo scarico e la differente architettura informatica utilizzata per l'acquisizione dei dati.

Per quanto riguarda la base di utenti essa era identificabile in due categorie: utenti delle sedi territoriali Istat interessati allo scarico dei soli dati relativi al proprio territorio e utenti dell'Istituto interessati ai dati dell'intero territorio nazionale. Oltre a questa differenziazione essendo tutti gli utenti interessati già presenti nei sistemi di autenticazione dell'Istituto è stato scelto per l'autenticazione di utilizzare tale infrastruttura e più precisamente di implementare un nuovo modulo software di autenticazione mediante protocollo LDAP. In questa implementazione è stato in particolare anche aggiunto un controllo di autorizzazione sulla base dati dedicata all'applicativo

per restringere gli utenti abilitati allo scarico rispetto all'intera base degli utenti presente nei sistemi di autenticazione LDAP dell'Istituto e contestualmente individuare il territorio di interesse da associare all'utente che accedeva all'applicativo.

La Figura 1.8 illustra il *workflow* di autenticazione implementato:

Figura 1.8 – *Workflow di autenticazione Indata*



Il secondo aspetto rappresentato dalla differente architettura informatica si può suddividere essenzialmente in due novità rispetto all'architettura di SGRCIS, la prima riguarda il *front end* di acquisizione che si differenziava per l'utilizzo di un componente software (*GINO++*) sviluppato in linguaggio PHP ed il secondo per un differente schema utilizzato per il salvataggio dei dati.

Proprio quest'ultimo ha introdotto nuove problematiche in quanto il predetto software utilizza una struttura dei dati di tipo verticale con i dati del singolo questionario distribuiti su molteplici tuple. Lo scarico dei dati era consentito anche per i dati parziali/provvisori quindi è stata scartata l'opzione di realizzare dei processi *batch* per costruire una base dati *ad hoc* per lo scarico, optando invece per la progettazione e l'utilizzo di opportune viste sul RDBMS. L'elevato numero di variabili del questionario ha posto però un problema di performance nella costruzione delle viste che trasponevano i dati da una struttura verticale ad una orizzontale, la scelta quindi è stata quella di separare lo scarico in più parti, una per ciascuna sezione del questionario e progettare per ciascuna parte una vista di trasposizione dei dati. Lo scarico quindi è stato possibile con prestazioni più che accettabili attraverso la generazione di file nel formato *csv* con carattere | ("pipe") come separatore.

La realizzazione del progetto, in linguaggio Java è conforme agli standard ed ai *framework* utilizzati in Istituto, ha consentito di realizzare nuovi componenti software (autenticazione LDAP) e di affrontare la tematica dello scarico dei dati acquisiti dal sistema *GINO++*, consentendo di costruire una base software che certamente potrà essere riutilizzata in futuro per altri applicativi di interesse per l'Istituto.

### 1.7.2 La modalità offline

L'ultima parte del lavoro riguarda il passaggio dalla raccolta dei dati alla loro fruizione da parte degli utenti statistici. Il problema dell'inserimento dei dati nei database in tabelle normalizzate/verticalizzate e quello della loro fruizione in tabelle ridondanti del tipo individui – variabili è stato anticipato dal paragrafo precedente con il flusso reciproco di informazioni tracciato-database.

Tale problema costituisce in realtà un duale che può essere risolto con delle viste che rappresentino in maniera opportuna la medesima informazione come relazione individui – variabili, ovvero in forma verticalizzata. Il problema è perciò di tipo logico prima che computazionale. In questo paragrafo viene trattato il primo problema, descrivendo i diversi passi eseguiti per la creazione di una o più viste Oracle nelle quali le variabili, che erano presenti in forma verticalizzata nella tabella dei microdati, sono state sottoposte a una rielaborazione al fine di renderne possibile la fruizione da parte dei software di analisi statistica.

La procedura seguita è stata quella di creazione di oggetti del database che consentissero la manipolazione delle informazioni contenute nella tabella dei microdati in forma verticalizzata, passando a una o più tabelle in forma ridondante dove l'informazione assumeva una forma  $n * m$  (individui \* variabili). La deverticalizzazione dei dati è stata effettuata, per ciascuna delle tre tabelle menzionate (unità istituzionali, unità locali, schede di servizio delle unità locali) facendo ricorso ad una serie di viste (in linea di principio una per ciascuna sezione di questionario), riunite infine in tre viste materializzate. Ciascuna vista prodotta era un oggetto del DBMS che esisteva solo da un punto di vista logico, senza contenere quindi materialmente dei dati, anche se poteva essere acceduta in lettura come una normale tabella. La vista, infatti, nella tecnologia DB, non contiene dati, ma come suggerisce il nome, rappresenta solo un modo speciale di "vedere" i dati che sono presenti nella/e tabella/e da essa referenziata/e. La vista così creata poteva essere interrogata liberamente (Prospetto 1.3).

#### Prospetto 1.3 – Schema della vista dei microdati non verticalizzata

IDENTIFICATIVO UNIVOCO DEL RISPONDENTE (UNITÀ LOCALE)	Nuova variabile 1	[...]	Nuova variabile n
1			
[...]			
m			

La creazione delle nuove variabili ha comportato un lavoro differenziato secondo casistiche precise. Nel caso più semplice la colonna è stata frutto di un'espressione che costruiva, a partire da un insieme di righe della tabella di partenza una nuova variabile. Il ricorso a tipologie di quesiti più complessi e/o il ricorso a porzioni di questionario riprodotte in sezioni modulari ha comportato l'identificazione dell'insieme di righe interessate attraverso un insieme opportuno di campi della tabella di partenza (identificativi di unità, variabile, di classificazione, eccetera). Nei casi più complessi all'identificazione del sottoinsieme di righe deputato a produrre la variabile finale si è sovrapposta l'esigenza di produrre un nuovo valore in transcodifica della variabile finale, eliminando le variabili che possedevano una funzione puramente strumentale nel database (identificazione delle righe chiamate a produrre la variabile finale), ma non contenevano informazione utile ai fini statistici.

Lo scopo finale di questa operazione era fornire un rapido accesso ai dati raccolti, in particolar modo ai soli quesiti per i quali risultasse necessaria una consultazione in tempo reale dal DB in produzione. Tali viste hanno costituito l'input di una applicazione dotata di un *front end* gestibile anche dall'utente generico (che è stato oggetto di uno sviluppo *ad hoc*), oltre ad essere direttamente consultabili dall'utente esperto (da sole o combinate fra loro) tramite opportuna sintassi in ambiente SAS attraverso una *pass-through facility*.

L'informazione trasposta nelle viste relative alle diverse sezioni è stata poi raccolta tematicamente nei tre questionari menzionati (unità istituzionali, unità locali, schede di servizio delle unità locali) attraverso tre viste materializzate. Anche la vista materializzata (*materialized view*) è un oggetto del DBMS, che può essere considerato un elemento a metà fra una vista e una tabella. Si tratta, infatti, di viste che vengono scritte fisicamente su disco per consentirne una lettura più rapida, ma i cui dati sono aggiornati dal DBMS secondo regole precise (a intervalli temporali regolari o al verificarsi di un particolare evento) e sono perciò definite sia logicamente (il codice che genera e aggiorna i dati come in una vista), sia fisicamente (l'insieme dei dati raccolti in una tabella). Queste viste sono in genere utilizzate per applicazioni di *data warehousing* come nel caso censuario, dove il *refresh* totale con una frequenza di sei ore assicura la piena fruibilità dei dati raccolti attraverso una procedura *batch* che riduce notevolmente i costi temporali della consultazione dei dati (ridotti da qualche minuto a pochi secondi) al prezzo di un *lag* temporale di alcune ore rispetto all'acquisizione dei dati. Rispetto alle normali viste, le viste materializzate realizzate hanno riportato anche una serie di informazioni come stato del questionario, ultimo accesso/salvataggio e generalità del rispondente che si trovavano riassunte in una ulteriore vista, utili nelle fasi successive dell'indagine (*recall*, controllo e correzione, eccetera).

## 1.8 La documentazione online

La Rilevazione sulle istituzioni pubbliche, nell'ambito del 9° Censimento generale dell'Industria e dei servizi e Censimento delle istituzioni non profit si è svolta esclusivamente online. In particolare, i rispondenti hanno fatto riferimento al portale del censimento che aveva come URL <https://indata.istat.it/cis2011ipfase2>.

Per garantire agli operatori tutta l'assistenza necessaria nelle varie fasi della rilevazione, sono stati resi disponibili diversi documenti, visualizzabili in linea o direttamente scaricabili in formato *pdf* dal portale sopraindicato. La documentazione presente sul sito è stata suddivisa in quattro classi di documenti:

- Documentazione;
- Istruzioni;
- Glossario;
- Domande e Risposte.

Nella sezione Documentazione erano presenti i documenti ufficiali del censimento di interesse generale per i rispondenti; questi documenti si riferivano ad entrambe le fasi del censimento. I documenti facevano parte di una lista ordinata temporalmente, dal testo di legge di indizione del CIS al testo inerente le lettere di sollecito.

Nella sezione Istruzioni sono stati messi a disposizione dei rispondenti tutta una serie di documenti fondamentali per una corretta compilazione dei questionari. In questa sezione erano presenti documenti relativi ad ambedue le fasi del censimento, inoltre era possibile visualizzare, esclusivamente online, una serie di informazioni come il *video tutorial* e le normative. Per quanto riguarda la Prima fase, sono stati messi a disposizione degli utenti i seguenti documenti: il questionario, la guida alla compilazione, gli aspetti tecnologici e l'esempio di lista unità locali, in

formato csv. Per la Seconda fase sono stati altresì messi a disposizione: il questionario, la guida alla compilazione, le istruzioni per la compilazione del questionario sulle scuole statali, le istruzioni per individuare le unità locali, i riferimenti normativi, il catalogo dei servizi, gli aspetti tecnologici, il modello MT1 e la classificazione delle attività economiche Ateco 2007.

Nella sezione Glossario era disponibile una raccolta di definizioni dei termini specifici della Rilevazione sulle istituzioni pubbliche.

Nella sezione Domande e risposte era disponibile un ulteriore mezzo di assistenza e supporto: le *Frequently Asked Questions* (FAQ). Le FAQ sono una serie di risposte stilate come soluzione alle domande più frequenti. Con questo canale si è cercato di fornire indicazioni di rapida consultazione alle problematiche più comuni. Le FAQ indicate erano presenti per entrambe le fasi del censimento.

Il portale ha permesso tramite la funzione "Contatti" di visualizzare e scaricare in formato pdf i riferimenti dei contatti specifici per le richieste di chiarimento e/o risoluzione delle problematiche della rilevazione.

## 1.9 I test e le prestazioni del sistema

L'attività di test è stata svolta solo per la Seconda fase del Censimento delle istituzioni pubbliche, poiché sono stati coinvolti più enti rispetto alla Prima fase.

Il sistema di acquisizione e gestione per il Censimento delle istituzioni pubbliche è stato realizzato riutilizzando e integrando sistemi preesistenti con un componente sviluppato *ex novo*:

- *Indata*: per l'autenticazione degli utenti e il controllo degli accessi;
- *GINO++* (*Gathering Information Online*) per il disegno del questionario tramite una interfaccia di amministrazione e l'acquisizione di dati attraverso un questionario esposto sul web;
- Il componente sviluppato *ad hoc* (in Java, *Struts2* e *Hibernate* e DB Oracle) per integrare i primi due moduli e fornire delle funzionalità aggiuntive, fra cui il monitoraggio della rilevazione in tempo reale e la gestione di profili utente differenziati in base al ruolo e all'appartenenza territoriale.

Il riuso e l'integrazione di componenti preesistenti, nel sistema software oggetto di test, presenta alcuni vantaggi, in quanto, i due componenti preesistenti si possono considerare come moduli *black box* e quindi si può assumere che già siano stati testati e collaudati.

Partendo da queste considerazioni per pianificare i test in generale si sono seguite due diverse strategie: per la parte costruita *ex novo* è stata applicata la metodologia di test standard già descritta al Paragrafo 1.5 della Parte Prima e, limitatamente alle componenti preesistenti, sono stati eseguiti in particolare i test di integrazione e di sistema tra i diversi componenti.

I test di integrazione hanno l'obiettivo di verificare la corretta interazione tra i vari moduli che costituiscono il sistema. Essendo il sistema da sottoporre a test costituito da moduli già esistenti e da un componente costruito *ex novo*, si è proceduto provando le tre possibili tecniche descritte in letteratura:

- *top-down*: vengono integrati e testati inizialmente i moduli ad alto livello;
- *bottom-up*: si inizia a testare un insieme limitato dei moduli e progressivamente si aggiungono gli altri;
- *sandwich testing*: una tecnica mista fra le prime due che prevede la pianificazione di test a partire dal percorso funzionale dei dati e del flusso di controllo (*business process oriented*).

A seguito delle prove effettuate su un sottoinsieme di funzionalità, è stato scelto di avvalersi dell'ultima tecnica descritta.

Particolare importanza ha rivestito l'attività di test prestazionali, poiché l'utilizzo di sistemi eterogenei, non appositamente sviluppati per essere integrati tra loro, ha comportato un deterioramento delle performance. In particolare ci si è trovati di fronte a una duplicazione dei dati, poiché *GINO++* ha un suo DB *embedded* in cui i dati sono memorizzati in tabelle a struttura "verticale" costituita da una tripla (chiave, valore, tabella di riferimento), non compatibile con gli applicativi censuari e per cui si sono dovuti "deverticalizzare" i dati, trasformandoli in tabelle standard dei database relazionali. Inoltre, ogni singolo sistema disponeva di un proprio insieme di oggetti software disgiunto per cui era necessario un continuo scambio di informazioni.

L'attività di test prestazionali si è avvalsa anche di un sistema di monitoraggio, basato sul software *Splunk*, che ha permesso di monitorare la fase di produzione e di effettuare un "*fine tuning*" per le prestazioni e per il funzionamento degli applicativi.

Il carico previsto di accessi giornaliero al sistema non è stato elevatissimo, rispetto alla Rilevazione sulle imprese e al Censimento delle istituzioni non profit, in quanto i questionari stimati ammontavano a 200 mila e a fine rilevazione i questionari raccolti sono stati 12.183 per le unità giuridico-economiche e 95.611 per le unità locali. Inoltre ci sono stati 3,8 milioni di accessi dal primo ottobre 2012, fino a gennaio 2013 da parte di circa 40 mila utenti.



PARTE TERZA

IL SISTEMA INFORMATICO DI DIFFUSIONE



## Capitolo 1

### Il sistema di diffusione<sup>20</sup>

#### 1.1 L'architettura complessiva

Il Censimento dell'Industria e dei servizi del 2011, ha visto il coinvolgimento di quattro fonti statistiche:

1. Archivio statistico delle imprese attive (Asia);
2. Rilevazione sulle istituzioni non profit;
3. Rilevazione sulle istituzioni pubbliche;
4. Rilevazione campionaria sulle imprese.

A fronte del trattamento di processi statistici così differenti tra loro, la componente tecnica ha dovuto operare in modo tale da bilanciare contemporaneamente due differenti aspetti:

- armonizzazione complessiva del sistema per la componente di interazione con l'utenza;
- rispetto delle caratteristiche del singolo processo di produzione.

Ciò ha portato, da un lato, ad una messa a punto della metodologia di definizione di cubi di dati armonizzati tra diversi processi statistici (Paragrafo 1.2) e, dall'altro, alla scelta di un'architettura modulare, come di seguito illustrato.

Il sistema di diffusione dei dati per il Censimento dell'industria e dei servizi, infatti, è stato realizzato secondo una logica modulare: a supporto di ciascuna fase si sono sviluppati o utilizzati uno o più moduli software specifici. In particolare la componente di aggregazione, responsabile di tutti i processi per il passaggio dai microdati ai macrodati, ha previsto uno sviluppo software mirato, mentre per le componenti di messa a disposizione dell'utenza sia dei macrodati che dei microdati sono stati utilizzati o opportunamente riutilizzati i sistemi *corporate* dell'Istituto.

Relativamente alla fase di aggregazione, in considerazione delle diverse caratteristiche dei dati, sono state implementate due differenti soluzioni tecnologiche:

- una per le istituzioni non profit, le istituzioni pubbliche e l'indagine multiscopo sulle imprese, i cui processi di rilevazione hanno previsto l'utilizzo di quesiti *multiresponse* ed un certo numero di flussi di risposta all'interno dei questionari tradottisi in dipendenze funzionali tra le dimensioni di analisi (Paragrafo 1.3);
- una per il registro delle imprese, a supporto di una grande quantità di dati (Paragrafo 1.4).

<sup>20</sup> A cura di Stefania Bergamasco. Autori del capitolo:

- Paragrafo 1.1: Stefania Bergamasco;
- Paragrafo 1.2: Stefania Bergamasco e Andrea Bruni;
- Paragrafo 1.3: Enrico Orsini e Eleonora Sibilio (Paragrafo 1.3.1), Paola Giorgetti, Manuela Marrone e Aureliano Secchiero (Paragrafo 1.3.2), Paola Giorgetti e Manuela Marrone (Paragrafo 1.3.3), Enrico Orsini, Aureliano Secchiero e Eleonora Sibilio (Paragrafo 1.3.4);
- Paragrafo 1.4: Massimiliano Amarone;
- Paragrafo 1.5: Raffaella Aracri e Domenico Scalzo (Paragrafo 1.5.1), Andrea Santilli e Domenico Scalzo (Paragrafo 1.5.2), Andrea Bruni e Concetta Pellegrini (Paragrafo 1.5.3), Andrea Bruni e Manlio Paterniti (Paragrafo 1.5.4);
- Paragrafo 1.6: Alessio Cardacino (Paragrafo 1.6.1), Paolo Rondinone (Paragrafo 1.6.2);
- Paragrafo 1.7: Florinda Damiani (Paragrafi 1.7.1 e 1.7.4), Mara Cammarrota (Paragrafo 1.7.2), Marco Caputi (Paragrafo 1.7.3).

Relativamente, invece, alla fase di diffusione le soluzioni hanno riguardato tre differenti canali:

- il *data warehouse* (<http://dati-censimentoindustriaeservizi.istat.it>) per la navigazione dei macrodati da parte dell'utenza. In questo caso è stata utilizzata la medesima tecnologia presente in Istat per il *data warehouse* di Istituto (I.Stat) seppur in una versione differente resa anche più performante. Tale tecnologia si basa su una soluzione realizzata in sede OECD (.Stat) al cui sviluppo l'Istat partecipa in sede internazionale in collaborazione con l'OECD ed altri istituti nazionali di statistica. I dettagli sono forniti all'interno del Paragrafo 1.5;
- il sistema di messa a disposizione dei macrodati in formato SDMX per i colloqui *machine-to-machine*. Per rispondere a tale esigenza si è scelto di riutilizzare, creando un'istanza *ad hoc*, la soluzione già adottata in Istat a supporto del *data warehouse* di Istituto, il sistema di *Single Exit Point*. Tale tecnologia, come illustrato all'interno del Paragrafo 1.6, è parzialmente basata su tecnologie sviluppate in Eurostat e del cui sviluppo l'Istat è leader in sede internazionale;
- il sistema di archiviazione e messa a disposizione dei microdati. Anche per questa fase si è scelto di utilizzare le soluzioni corporate dell'Istituto: il sistema ARMIDA, il laboratorio ADELE, i file per lo scambio dei dati agli Enti Sistan o per la ricerca (Paragrafo 1.7).

Parte integrante della soluzione tecnologica è stata la progettazione degli ipercubi o cubi di diffusione realizzata congiuntamente tra le professionalità responsabili del *data warehouse* ed il gruppo statistico che ha guidato i processi di indagine. L'interscambio e la stretta collaborazione hanno consentito di trovare soluzioni di disegno degli ipercubi che tenessero in debita considerazione sia le problematiche tecniche di performance per l'elaborazione dei dati e la navigazione su web, sia le esigenze di natura statistica. All'interno del Paragrafo 1.2 si fornisce una sintesi dello strumento e di alcuni aspetti metodologici utilizzati per la progettazione dei cubi.

## 1.2 La progettazione dei cubi di dati

La progettazione dei cubi di dati ha comportato, in primo luogo, la necessità di definire opportune regole metodologiche per assicurare l'integrazione dei dati provenienti da processi statistici differenti, nonché strumenti di lavoro che rappresentassero debitamente il piano delle diffusioni definito in ambito statistico e fossero contemporaneamente adeguati alle professionalità informatiche per la gestione di tutte le fasi dovute.

In riferimento alle metodologie realizzate, il principale aspetto ha riguardato le regole di costruzione di liste integrate di classificazioni in modo che ciascun processo statistico potesse utilizzare le specifiche modalità di interesse sul singolo cubo di dati di diffusione e contemporaneamente che le descrizioni delle modalità e le relative codifiche fossero uguali, laddove statisticamente necessario, tra processi differenti. La soluzione adottata, mutuando la scelta già realizzata in Istituto, per il *corporate datawarehouse* I.Stat, è stata quella di creare liste uniche includenti, data una dimensione di analisi, tutte le modalità necessarie a tutti i processi statistici. In questo modo si è potuto consentire a ciascun processo di selezionare e utilizzare le specifiche modalità di interesse pur nel governo complessivo dei contenuti diffusi. Le diverse tempistiche di rilevazione e diffusione dei singoli processi di indagine, hanno contestualmente imposto un modello ricorsivo per la compilazione delle liste uniche su menzionate. La soluzione adottata, per consentire l'inserimento nel tempo di modalità differenti all'interno delle liste senza ricodifica di quanto già progettato, è stata quella di utilizzare codifiche di tipo acronimo rispetto alle descrizioni, scorporando dal codice il concetto di ordinamento e gestione della gerarchia (informazioni che si sono governate come attributi delle singole modalità). Il riuso delle modalità di diffusione tra una indagine e l'altra ha



Nella figura sono presenti otto sezioni:

- la sezione 1 è la sezione del prospetto. In questa sezione le colonne rappresentano le dimensioni di analisi e le righe gli incroci possibili tra le diverse dimensioni in base al livello territoriale considerato. Le eventuali dipendenze funzionali tra le dimensioni di analisi vengono evidenziati con l'indicazione dei sottogruppi di modalità specificatamente utilizzate sul singolo incrocio (sezione 5);
- la sezione 2 è la sezione in cui per ciascuna riga o gruppo di righe si progetta la visualizzazione che si desidera fornire all'utente;
- la sezione 3 è costituita da tanti fogli di lavoro quanti sono le dimensioni di analisi. In tali fogli i settori riportano la classificazione che si intende utilizzare (sezione 6), di questa indicano le voci di cui necessitano (sezione 7), e gli eventuali sottogruppi di modalità necessari per il governo delle dipendenze funzionali (sezione 8). Per ciascuna modalità, all'interno di questi fogli, viene indicata la domanda di riferimento sul questionario o la regola di calcolo a partire dallo stesso;
- la sezione 4 è dedicata alla descrizione dei metadati.

In ultimo, al fine di minimizzare la numerosità dei cubi e, dunque, di ottimizzare i flussi per la produzione, si è scelto di inserire all'interno di ciascun cubo, forzando la metodologia dei *data warehousing*, una dimensione denominata "Tipo dato", le cui modalità rappresentano gli aggregati statistici che sono messi a disposizione dell'utenza all'interno del singolo cubo dati e per i quali è possibile analizzare i dati in base alle dimensioni associate al cubo. Un esempio è dato dal cubo di sintesi dei dati del censimento all'interno del quale il *Tipo dato* è costituito da:

- numero unità attive;
- numero addetti;
- numero lavoratori esterni;
- numero lavoratori temporanei;
- numero volontari.

### **1.3 Il sistema di aggregazione per la gestione dei casi di multiresponse e le dipendenze funzionali tra le dimensioni di analisi: le istituzioni non profit, le istituzioni pubbliche e le imprese dell'indagine multiscopo**

#### **1.3.1 Il flusso dei dati**

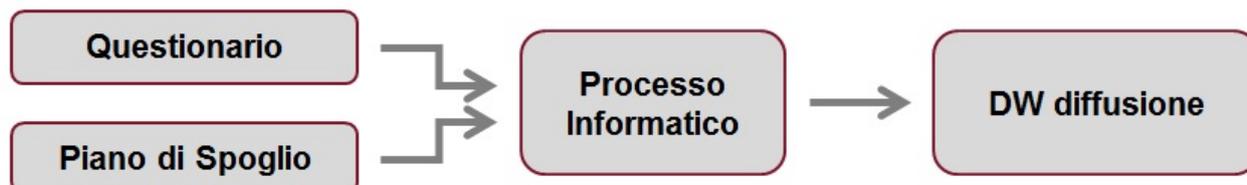
La diffusione dei dati è un processo finalizzato alla generazione di un patrimonio informativo (cubo) disponibile alla consultazione esterna. I punti di partenza sono costituiti dai dati acquisiti durante un processo di rilevazione (es. mediante questionario), che rappresentano i microdati del sistema informativo, e da un piano di spoglio, che costituisce l'input statistico alla parte informatica. Nello specifico, il processo informatico trasforma e aggrega in un cubo le variabili di analisi, in base alle modalità descritte nel piano di spoglio. L'output di tale processo consiste nella trasformazione del cubo in file *csv* da fornire al *data warehouse* di diffusione che si occupa della loro pubblicazione e consultazione.

Lo schema completo del processo è riportato in Figura 1.2 e la componente di aggregazione può essere schematizzato in tre fasi distinte:

1. la fase di ETL (*Extract, Transform, Load*);
2. la fase di controllo dati;
3. la generazione del file *csv* di input al *data warehouse* di diffusione.

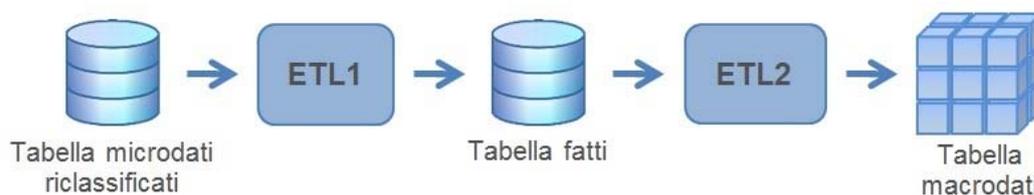
La fase di ETL si occupa di estrarre i dati da un sistema sorgente (DBMS, file e/o sistemi informatici diversi), trasformarli e quindi caricarli in un sistema derivato (DBMS, *data mart* o *data warehouse*).

Figura 1.2 – Flusso dati complessivo



La soluzione tecnologica adottata nel caso della diffusione del censimento, relativamente alle rilevazioni sulle imprese, istituzioni non profit e istituzioni pubbliche, è basata sull'utilizzo di database relazionali (DBMS Oracle) sia per il sistema sorgente che per quello derivato. Il processo di generazione di macrodati a partire dai microdati è stato suddiviso in due fasi come mostrato nello schema di Figura 1.3.

Figura 1.3 – Flusso dati – dettaglio



Le due fasi differiscono per la struttura e per il tipo di trasformazione che effettuano sul dato e sono univoche per ogni singolo piano di spoglio, poiché per ognuno di essi viene generata una tabella dei fatti e una tabella dei macrodati.

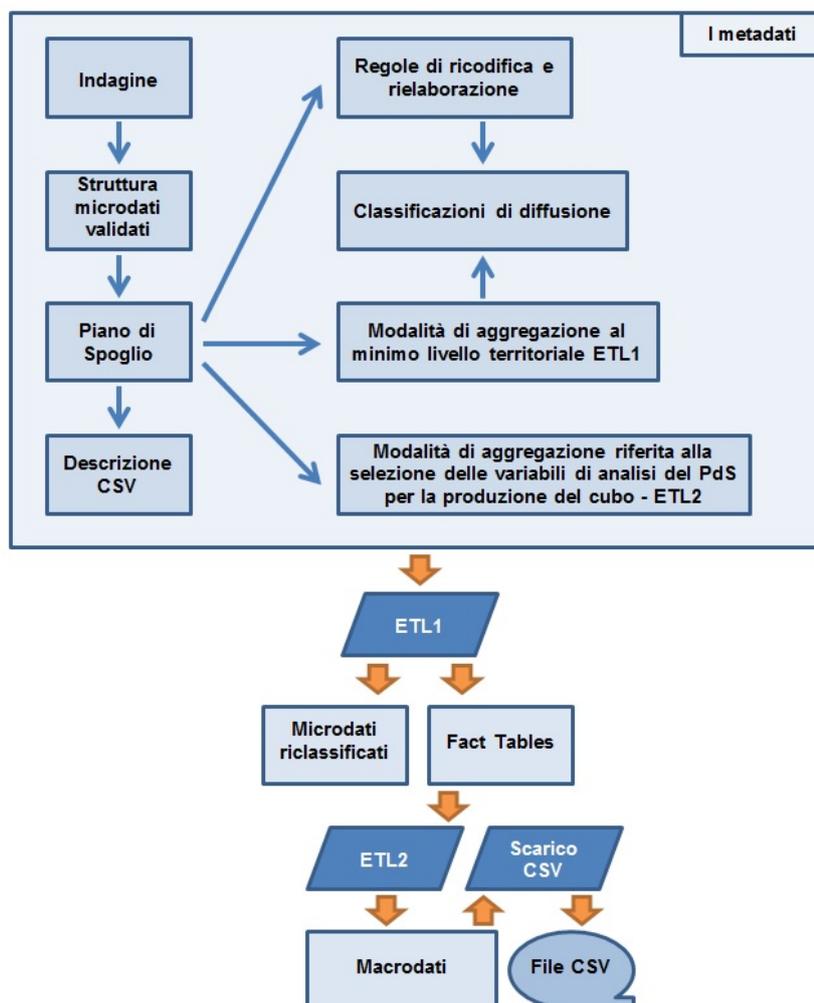
### 1.3.2 Organizzazione della base dati

Per la produzione dei cubi e dei file di dati da fornire al sistema di diffusione, si è scelto di utilizzare metadati memorizzati in strutture tabellari Oracle e procedure PL/SQL che elaborando i microdati validati forniti dai proprietari dell'indagine, secondo le regole descritte nei metadati, creano le aggregazioni richieste per ogni piano di spoglio. I metadati contengono informazioni relative a: struttura di microdati validati, struttura del piano di spoglio, regole di ricodifica dei microdati validati secondo le classificazioni di diffusione fornite insieme al piano di spoglio e generalmente differenti da quelle di rilevazione, modalità di aggregazione delle variabili di analisi. Le procedure PL/SQL si dividono in più gruppi:

- *ETL1*, che prevede la riclassificazione dei microdati validati e l'aggregazione al minimo livello territoriale;
- *ETL2*, che aggrega i dati secondo i criteri di diffusione del piano di spoglio;

- Scarico csv, che produce e controlla il file csv da fornire al sistema di diffusione.

Figura 1.4 - I metadati presenti all'interno della base dati



Si riporta di seguito la suddivisione, di dettaglio, realizzata per i metadati:

- *Indagine*, contiene l'elenco delle indagini di diffusione;
- *Struttura microdati validati*, contiene per ogni indagine la descrizione delle tabelle in cui sono contenuti i microdati validati;
- *Piano di spoglio*, comprende tutte le informazioni descrittive di ogni piano di spoglio: la descrizione, i tipi dato, i nomi dei cubi che devono essere prodotti;
- *Regole di elaborazione e ricodifica*, in questa parte vengono elencate le variabili previste in ogni piano spoglio, compresi i tipi dato, i filtri e le variabili di diffusione, e la loro tipizzazione (da calcolare, gerarchica, *multiresponse*, da riclassificare, eccetera) che ne determina le eventuali regole di calcolo e/o trasformazione/riclassificazione secondo le specifiche del piano di spoglio;

- *Classificazioni di diffusione*, sono tutte le tabelle relative alle classificazioni di diffusione. Contengono sia il codice di diffusione che quello di rilevazione per la ricodifica;
- *Modalità di aggregazione al minimo livello territoriale*, qui sono elencate le variabili di diffusione del piano di spoglio che devono essere aggregate al minimo livello territoriale; una gestione particolare viene riservata alle *multiresponse* perché ogni questionario è conteggiato tante volte quante sono le risposte date al suo interno per lo specifico quesito preso in considerazione;
- *Modalità di aggregazione riferita alla selezione delle variabili di analisi del piano di spoglio per la produzione del cubo*: in questa area le tabelle possono essere suddivise logicamente in tre differenti tipologie:
  - la prima tipologia è relativa alle tabelle di metadati che consentono la configurazione di alcuni parametri necessari al sistema di ETL2 per produrre i dati. Rientrano in questa categoria la denominazione delle tabelle dei fatti e quella dei macrodati nonché la denominazione delle variabili e le relative misure che si intende ottenere;
  - la seconda tipologia di tabelle consente la trasposizione in forma tabellare (matrice) del piano di spoglio. Nelle colonne vengono rappresentate le variabili di analisi e in ogni riga gli incroci delle stesse raggruppate in base al territorio e al tipo dato. I dati presenti in queste tabelle vengono utilizzati da procedure Oracle PL/SQL per estrapolare i dati presenti dalla tabella dei fatti, elaborarli e inserirli nella tabella dei macrodati in base alle regole dei piani di spoglio e rappresentate nelle tabelle descritte.

Poiché il sistema di diffusione del *data warehouse* non effettua aggregazioni online è stato necessario conservare nelle tabelle dei macrodati anche i record riferiti alle voci di *totale* presenti sulle singole dimensioni di analisi. Questa operazione comporta la generazione di record duplicati. Si è, quindi, deciso di definire una chiave primaria nella tabella dei macrodati costituita da tutti i campi relativi alle variabili di analisi. Grazie ad una funzione Oracle, tali record non vengono inseriti nelle tabelle dei macrodati ma sono scartati automaticamente e registrati in una tabella di *log*;
  - la terza tipologia di tabelle è relativa ai metadati necessari per effettuare i controlli (congruità, coerenza, validità, eccetera) dei dati presenti nel cubo finale;
- *Descrizione csv*, contiene la struttura del *csv* così come deve essere rilasciato al sistema di diffusione.

### **1.3.3 La fase di Etl1: la gestione dei metadati per la generazione automatica delle fact table**

La fase di ETL1 prevede due step: gestione Pre\_ETL1 e gestione ETL1 vera e propria.

Per gestione Pre\_ETL1 s'intende la normalizzazione e riclassificazione delle variabili del questionario per riportarle ai valori del piano di spoglio. Questa normalizzazione si rende necessaria per allineare le informazioni registrate nei microdati validati alle logiche e regole di diffusione. A tal fine devono essere descritte, in forma di metadati, la tipologia delle variabili da ricodificare e le regole di ricodifica. Tali metadati vengono letti da procedure PL/SQL che generano istruzioni parametrizzate per creare e popolare la tabella di microdati riclassificati che sarà l'input dello step di ETL1. Le diverse tipologie di variabili e le elaborazioni che vengono eseguite per la riclassificazione/ricodifica sono:

- variabili che devono essere utilizzate senza subire nessuna elaborazione nella tabella riclassificata;
- variabili gerarchiche come ad esempio "Territorio" o "Ateco", per le quali nei microdati è presente l'informazione al minimo livello di gerarchia previsto (es. codice comune), mentre nella tabella microdati riclassificati viene esplicitata tutta la gerarchia (es. "Ripartizione", "Regione", "Provincia", "Comune");

- variabili che devono essere ricodificare perché la classificazione di rilevazione è differente da quella di diffusione: la ricodifica è esplicitata nelle tabelle delle classificazioni;
- variabili che devono essere classificate perché rilevate come variabili quantitative, come ad esempio il numero di addetti che deve essere rielaborato in classi di addetti. Nelle tabelle di classificazione è presente l'intervallo di classe;
- variabili da calcolare e classificare; un esempio è l'età del titolare nel caso in cui nei microdati originali ci sia solo l'anno di nascita, in questo caso deve essere prima eseguito il calcolo dell'età e poi la sua classificazione;
- variabili per le quali nei microdati è prevista una risposta multipla e che vengono gestite diversamente a seconda di come è registrato il microdato e delle modalità di classificazione di diffusione. Esistono due possibili modi di registrazione delle *multiresponse*: *multiresponse registrata in differenti variabili sui microdati validati* (in questo caso si deve solo procedere alla ricodifica e nei microdati riclassificati si generano tante colonne quante sono le modalità di classificazione); *multiresponse registrata in una unica stringa composta da 0 (risposta mancante) e 1 (risposta data) sulla tabella dei microdati validati* - nel questionario la *multiresponse* può essere rappresentata come matrice oppure come elenco di risposte. Nel primo caso la sequenza dei caratteri della stringa rispetta l'ordine di righe e colonne della matrice, nel secondo caso la sequenza è l'ordine delle risposte. Nei metadati specifici del *data warehouse* deve essere descritto come suddividere la stringa per la ricodifica. In entrambi i casi nei metadati deve essere specificato se le modalità di classificazione sono esclusive ("Solo Risposta A", "Solo Risposta B", "Risposta A e B insieme"), *multiresponse* semplici ("Risposta A", "Risposta B", "Risposta C") oppure complesse ("Risposta A ma non B,C", "Risposta B ma non A", "Risposta A e B e C insieme"). Le *multiresponse* per le quali nelle classificazioni sono previste modalità esclusive generano nella tabella dei microdati riclassificati una sola colonna, tutte le altre generano invece tante colonne quante sono le modalità della classificazione.

Per ciascun piano di spoglio lo step di ETL1 è responsabile della creazione di una tabella (*fact table*) che contiene i dati aggregati per tutte le variabili di analisi al minimo livello territoriale previsto. A tal fine devono essere descritti, in forma di metadati e in base a quanto indicato nel piano di spoglio: le misure, cioè l'aggregato statistico che si intende diffondere, con le relative regole di calcolo e le variabili di aggregazione, con l'indicazione della tipologia; sono state sviluppate delle procedure PL/SQL che leggendo questi metadati generano l'istruzione parametrizzata che mediante apposite funzioni di aggregazione Oracle, crea e popola la tabella dei fatti, input per la fase di ETL2.

Le misure si differenziano a seconda della regola di calcolo in:

- *Conteggio*, se devono essere contate le unità di rilevazione. (es. ANPIN - numero istituzioni non profit attive);
- *Somma*, se deve invece essere sommata una particolare variabile quantitativa dell'unità di rilevazione (es. ANPIEO - numero dipendenti delle istituzioni non profit attive).

Per entrambe le tipologie potrebbe essere necessario applicare un filtro che limiti il conteggio o la somma ad alcune caratteristiche dell'unità di rilevazione (es. ANPINE - numero istituzioni non profit attive con dipendenti).

Le variabili di analisi si differenziano secondo la loro tipologia e le elaborazioni che vengono eseguite per la generazione della *fact table* in:

- variabili *multiresponse*;

- *variabili non multiresponse.*

Per quanto riguarda le *variabili multiresponse* nella tabella dei microdati riclassificati vengono generate più colonne, una per ogni voce della classificazione; nella *fact table* la *multiresponse* deve invece essere presente in una sola colonna in modo tale che essa diventi una delle dimensioni di analisi. Si necessita, dunque, di un'operazione di verticalizzazione. Nella Figura 1.5 se ne fornisce un esempio.

Qualora nel piano di spoglio siano presenti più *multiresponse* che si incrociano tra di loro, nella *fact table* deve essere generato un prodotto cartesiano delle voci di classificazione (dettaglio e totale) di tutte le *multiresponse* incrociate. Se al contrario ci sono più *multiresponse* che non si incrociano tra di loro, allora nella *fact table* per ogni *multiresponse* sono presenti tutte le voci di dettaglio della sua classificazione e le sole voci di totale per tutte le altre. Per queste variabili devono essere inoltre calcolati i valori totali, che, non corrispondendo alla somma delle singole voci di classificazione, non possono essere calcolati nella fase di ETL2, come per tutte le altre variabili di analisi. È dunque necessario prevedere, tra le voci di modalità delle classificazioni considerate, anche la modalità "Tutte le voci" ed effettuare il calcolo in fase di ETL1.

Per le variabili *non multiresponse*, comprese quelle territoriali e quelle per le quali nella tabella dei microdati riclassificati è stata esplicitata la gerarchia, non è previsto un trattamento particolare, ma un semplice raggruppamento, quindi ad ogni colonna della tabella dei microdati riclassificati corrisponde una colonna nella *fact table*.

**Figura 1.5 – Esempio di verticalizzazione dei microdati**

**Classificazione di diffusione: "Missioni" (*multiresponse*)**

CODICE_CLASSIFICAZIONE	DESCRIZIONE
ADVOC	promozione e tutela dei diritti
COMM	cura dei beni collettivi
SUPASS	sostegno e supporto di soggetti deboli e/o in difficoltà
NOIND	nessuna indicazione
ALL	tutte le voci

**Microdati riclassificati: variabile "Missioni" su più colonne, una per ogni modalità**

QUESTIONARIO	TERRITORIO	MISSIONI_ADVOC	MISSIONI_COMM	MISSIONI_SUPASS	MISSIONI_NOIND
111	.....			SUPASS	
222	.....	ADVOC			
333	.....	ADVOC		SUPASS	
444	.....	ADVOC	COMM	SUPASS	
555	.....	ADVOC	COMM		
666	.....				NOIND
777	.....		COMM		

**Fact Table: variabile "Missioni" su una sola colonna e righe distinte per ogni modalità**

ANPIN (N° IMPRESE)	TERRITORIO	FORMA_GIURIDICA	MISSIONI
4	.....	.....	ADVOC
3	.....	.....	COMM
3	.....	.....	SUPASS
1	.....	.....	NOIND
7	.....	.....	ALL

### 1.3.4 La fase di Etl2: il software, il passaggio dalla modellazione dei cubi alla generazione dei file dati, le regole di controllo

La fase di ETL2 è la parte del processo informatico in cui sono effettuate le aggregazioni e le normalizzazioni sulle variabili di analisi previste nel piano di spoglio. In funzione delle specifiche dettate in quest'ultimo, l'output della fase è la generazione della tabella dei macrodati a partire dalla tabella dei fatti. Ogni riga del piano di spoglio è associata a uno o più tipi dato e a una specifica combinazione di variabili di analisi; in base alla natura del tipo dato, alla tipologia e alle modalità di incrocio previste tra le variabili di analisi, a partire dalla *fact table* viene effettuato il calcolo per gli aggregati parziali e totali raggruppati in base al territorio e al tipo dato. La registrazione dei metadati, a partire dal piano di spoglio, per la fase di ETL2, è stata supportata dall'utilizzo di una applicazione web specificatamente sviluppata (*DwETL2*).

Il software, infatti, consente la registrazione delle singole righe del piano di spoglio, riportando gli incroci di variabili peculiari per ciascuna riga. Fanno parte di *DwETL2* le procedure che, leggendo i metadati, per ogni singola riga prevista, dinamicamente generano codice SQL che consente l'estrazione dei dati dalla tabella dei fatti, la loro successiva elaborazione mediante apposite funzioni di aggregazione Oracle (*group by, grouping, rollup*, eccetera) e il successivo inserimento sulla tabella dei macrodati. In Figura 1.6 è riportato un esempio di piano di spoglio e la relativa configurazione in cui si mostra il territorio, il tipo dato e tutti gli incroci delle variabili di analisi previste.

Figura 1.6 – Visualizzazione schermata DwEtl2

#### Piano di Spoglio

Territorio	Tipo dato	Settore di attività prevalente non profit		Forma giuridica	Tipo di attività economica	Tipologia di finanziamento prevalente	Orientamento dell'istituzione	Tipologia di enti con cui sono stati stipulati patti o intese
		liv1	liv2					
reg. prov., com	x	x		x				x
reg. prov., com	x	x			x			x
reg. prov., com	x	x				x		x
reg. prov., com	x	x					x	x

#### Pagina web di configurazione

Territorio	Tipo Dato	lcnp		Forma g	Market	Tipo_f	Mut_pub	Tipo enti
		Liv1	Liv2					
Regione	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Provincia	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Comune	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Trento bo	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Regione	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Provincia	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Comune	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Trento bo	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Regione	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Provincia	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Comune	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Trento bo	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Regione	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Provincia	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Comune	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Trento bo	Anpin	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Al termine della compilazione viene attivata, in maniera automatica dal sistema, una procedura Oracle che provvede all'elaborazione dei dati generando i macrodati. Il sistema consente di controllare il risultato di ogni singola configurazione, verificando il numero dei record inseriti nella

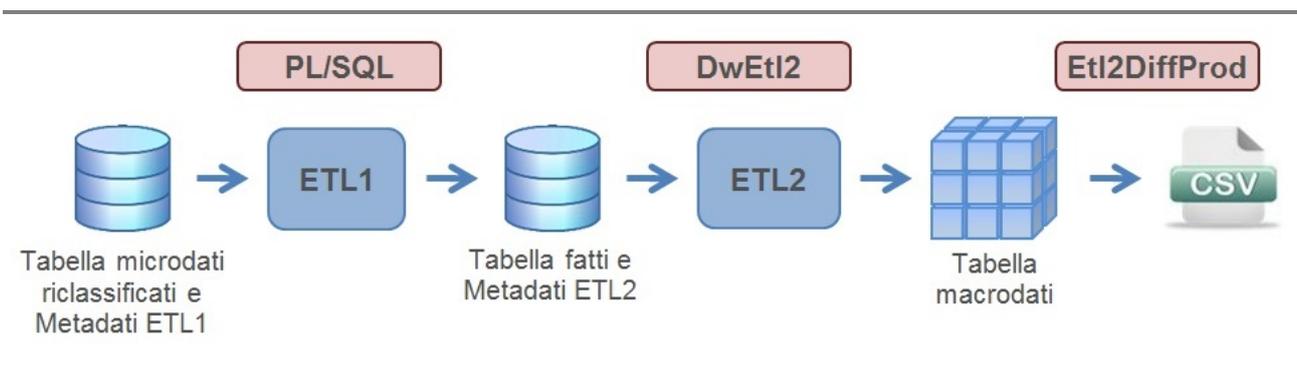
tabella del cubo per ogni singola riga; viene inoltre data la possibilità di rieseguire l'elaborazione, sia parziale sia totale, in caso di interruzione causata da problemi tecnici e/o da errori.

La fase finale del processo consiste nel controllo e nella validazione dei macrodati, nello scarico del file in formato csv e nel relativo controllo.

Il controllo dei dati è effettuato in modo non automatizzato direttamente sulla base dati, mediante la costruzione di opportune *query*. Per ogni *tipo dato*, tale controllo viene effettuato verificando che la somma delle misure calcolate per le modalità di non totale corrisponda alla misura della modalità di totale. In caso di variabili *multiresponse*, essendo possibili più risposte contemporaneamente, il controllo risponde, però, ad una logica differente: ci si accerta che la somma delle misure calcolate per le modalità di non totale sia maggiore (o uguale) alla misura della modalità di totale.

La verifica della struttura del file *csv* è eseguita successivamente alla sua creazione. Tali due ultimi processi sono stati automatizzati mediante un'applicazione web generalizzata (*ETL2DiffProd*), ideata per effettuare lo scarico di file *csv* con diverse modalità, in funzione del piano di spoglio. Ad esempio, si possono estrarre dati filtrandoli per territorio o per tipo dato. L'applicazione web utilizza il sistema dei metadati della fase di ETL1 integrato con alcune informazioni del sistema di metadati della fase di ETL2. In particolare, sul file *csv*, è eseguita la validazione della struttura e delle modalità delle variabili di analisi previste dal piano di spoglio. In conclusione, tutto il processo, con le soluzioni tecnologiche attualmente adottate, è schematizzato in Figura 1.7.

Figura 1.7 – Rappresentazione delle soluzioni tecnologiche all'interno del flusso dei dati



## 1.4 Il sistema di aggregazione di grandi quantità di dati: il registro delle imprese e dell'occupazione nelle imprese

I dati del registro statistico Asia divulgati per il Censimento dell'industria e dei servizi hanno riguardato la struttura delle imprese, delle unità locali e la loro occupazione. Queste informazioni sono state prodotte partendo dall'estrazione di tabelle di notevoli dimensioni in particolare caratterizzate da: un basso numero di misure (variabili numeriche da analizzare), un elevato numero di dimensioni (variabili di classificazione) e un dettagliato livello territoriale.

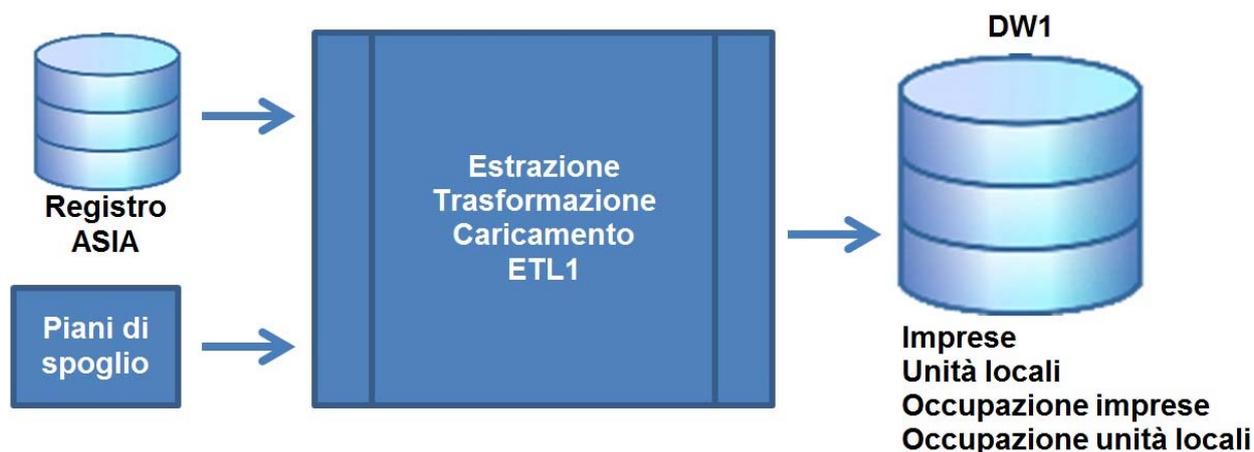
Per ridurre la complessità e favorire le successive attività di verifica e controllo dei risultati, si è preferito suddividere la progettazione in due fasi: progettazione del *data warehouse primario* (DW1) e progettazione del *data warehouse secondario* (DW2).

### 1.4.1 Progettazione del data warehouse primario (DW1)

La fase di progettazione del *data warehouse* primario è partita dall'analisi dei piani di spoglio (PDS) dei dati dell'archivio Asia. Quest'attività ha definito la quantità di variabili da estrarre dal

database dei micro dati, la loro organizzazione e il loro massimo livello di dettaglio. Dall'archivio dei microdati Asia sono state estratte le informazioni riguardanti i lavoratori, le imprese, le unità locali e quelle della loro occupazione (Figura 1.8).

Figura 1.8 – Flusso dati per il popolamento del *data warehouse* primario DW1



Per l'implementazione del *DW* primario si è utilizzato il sistema *ROLAP* (*Relational OnLine Analytical Processing*) in associazione con lo schema a stella. Usando lo schema relazionale la collezione dei fatti è memorizzata in una tabella che ha come attributi le dimensioni e le misure. Ogni dimensione ha attributi propri che sono memorizzati in una tabella distinta (Figura 1.9).

Figura 1.9 - Modello dimensionale



L'interconnessione con il *repository* dei microdati è stata realizzata mediante viste materializzate. Attraverso di esse si opera la fase di estrazione dei microdati. Sono state scelte le viste materializzate rispetto a quelle logiche per le loro migliori prestazioni nella fase d'interrogazione. Inoltre, opportune procedure PL/SQL consentono l'eventuale aggiornamento dei microdati estratti.

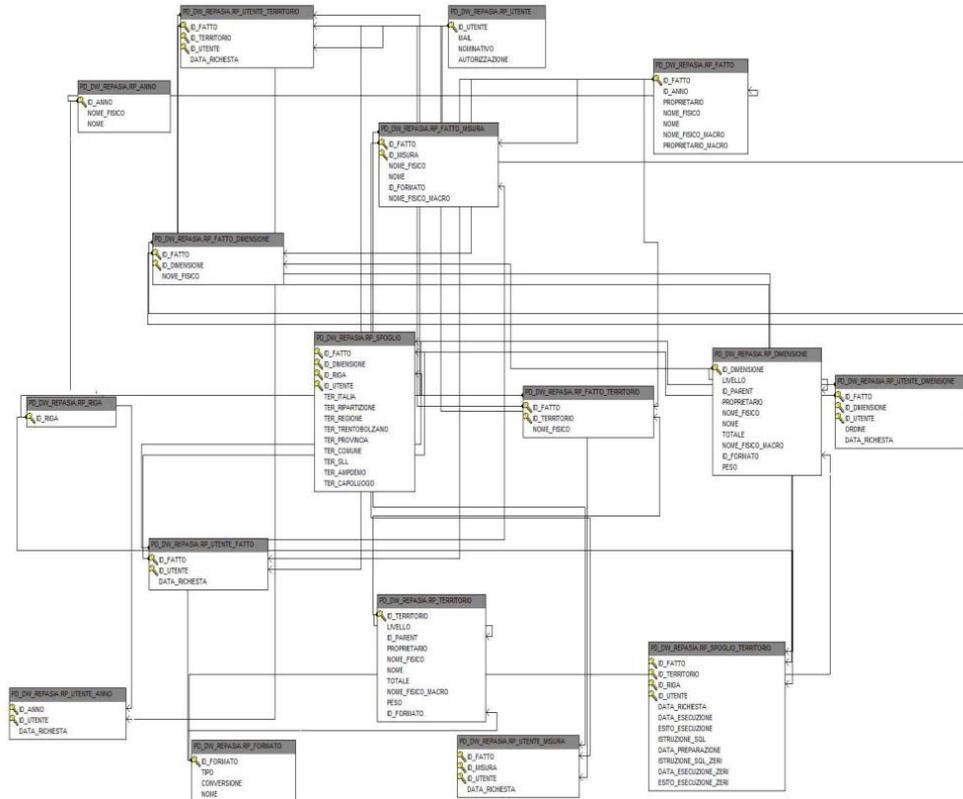
I dati contenuti nelle viste sono aggregati al massimo del dettaglio con il quale si vogliono diffondere, in accordo con il piano di spoglio. In questa fase, per i dati aggregati, si conserva la stessa codifica delle classificazioni della fonte dati per permettere un più semplice confronto sui risultati intermedi ottenuti. In seguito sono state create le rispettive tabelle dei fatti, una per ogni tema di analisi (impresa, unità locale, demografia lavoratore, eccetera), con le rispettive classificazioni opportunamente transcodate. Per escludere eventuali errori di codifica delle classificazioni, le tabelle dei fatti sono dotate di vincoli d'integrità referenziale con le rispettive tabelle delle dimensioni. Al fine di migliorare le prestazioni per le successive elaborazioni, riducendo il numero di *join* tra tabelle, le tabelle dei fatti sono volutamente denormalizzate rispetto le viste precedenti, hanno, cioè, un numero di colonne maggiore per rappresentare le classificazioni gerarchiche come: il territorio, l'attività economica, la forma giuridica, eccetera. Questo migliora le prestazioni per le successive elaborazioni poiché riduce il numero di *join* tra tabelle.

All'interno del DW1 sono, inoltre, presenti delle tabelle per la transcoda dei codici di classificazione al fine di rendere i dati omogenei e conformi con le regole di codifica in fase di diffusione. Per popolare le tabelle dei fatti sono state implementate e raccolte in un unico *package* opportune procedure PL/SQL con le relative funzioni di transcoda, derivazione e omogeneizzazione delle classificazioni. Date le dimensioni delle tabelle dei micro dati (circa 5 milioni per le imprese e le unità locali e 25 milioni per gli occupati) è stato necessario utilizzare opportune tecniche di caricamento. Per migliorare le performance e ridurre l'utilizzo delle risorse del RDBMS la lettura, il raggruppamento e la scrittura dei dati su tabella è stata suddivisa in blocchi più piccoli (*buffered bulk collect*).

#### **1.4.2 Progettazione del data warehouse secondario (DW2)**

Anche la fase di progettazione del *data warehouse* secondario si basa sull'analisi dei piani di spoglio, con tutti gli incroci tra le classificazioni specificati al suo interno e l'estrazione di tutte le tipologie di classificazione: gerarchiche, normali ed esclusive. Esempi di dimensioni gerarchiche sono il territorio (Italia, ripartizione, regione, provincia, comune, sistema locale lavoro), l'attività economica (dalla sezione alla categoria). Un tipo di dimensione esclusiva è ad esempio la classe di addetti (nei suoi vari raggruppamenti). Per la produzione dei macrodati sono state realizzate delle tabelle di *repository* e sviluppate particolari procedure in PL/SQL raccolte in un unico *package*. Si riporta nella Figura 1.10 lo schema dati.

Le tabelle contengono e descrivono tutte le informazioni necessarie, opportunamente desunte dai piani di spoglio: i fatti, le dimensioni, le misure con la loro tipologia di formato (numerico con o senza decimali). L'applicazione genera la struttura delle relative tabelle dei macrodati, la creazione degli indici di tipo *bitmap* ed esegue le complesse istruzioni SQL per il caricamento di tali tabelle. In aggiunta la procedura genera le corrispondenti istruzioni SQL per l'estrazione dei file *csv* in accordo con il formato e sintassi richiesto dal *data warehouse* di diffusione.

Figura 1.10 – Schema dati del *data warehouse* secondario DW2

Le istruzioni di raggruppamento e inserimento sono generate dinamicamente dall'applicazione. Esse sfruttano la combinazione tra il comando di analisi dei dati *rollup* con quello di *group by* le funzioni *grouping* e *case* disponibili su Oracle.

Le tabelle dei macrodati sono dotate di colonne atte a descrivere i pesi della variabile di *grouping* associati alle variabili di classificazione. È su queste colonne che sono associati gli indici di tipo *bitmap*. Si è preferito l'utilizzo di questa tipologia di indice per le sue dimensioni ridotte e le elevate prestazioni per la realizzazione di interrogazioni con i filtri utilizzate, in particolar modo, durante le fasi di controllo dei dati. Inoltre per migliorare le prestazioni di caricamento ed evitare di inserire delle righe duplicate nella tabella dei macrodati si è creato un vincolo di chiave primaria contenente tutte le colonne di classificazione con associata una procedura (*DBMS\_ERRLOG*) ed una tabella per il *log* degli errori. Dagli incroci predisposti nei vari piani di spoglio, e soprattutto per le variabili di classificazione di tipo gerarchico, si genera un'alta possibilità d'inserimento di righe duplicate. Questa tecnica di esclusione degli errori SQL per comandi di DML ed in particolare per la violazione dei vincoli d'integrità referenziale è suggerita per le sue elevate prestazioni in confronto ad altri metodi (Figura 1.11).

Figura 1.11 – Esempio di istruzione PL/SQL

```

INSERT INTO MT_IMPRESA_2011
SELECT
31,
grouping_id(ID_ATE_SEZIONE, ID_ATE_DIVISIONE, ID_ATE_GRUPPO,
ID_ATE_CLASSE, ID_ATE_CATEGORIA),
grouping_id(ID_ADD_DETT),
grouping_id(ID_FG_AGG),
...,
ID_TER_ITALIA,
case
when grouping_id(ID_ATE_SEZIONE) = 1 then '0010'
when grouping_id(ID_ATE_DIVISIONE) = 1 then ID_ATE_SEZIONE
when grouping_id(ID_ATE_GRUPPO) = 1 then ID_ATE_DIVISIONE
when grouping_id(ID_ATE_CLASSE) = 1 then ID_ATE_GRUPPO
when grouping_id(ID_ATE_CATEGORIA) = 1 then ID_ATE_CLASSE
else ID_ATE_CATEGORIA
end,
case
when grouping_id(ID_ADD_DETT) = 1 then 'TOTAL'
else ID_ADD_DETT
end,
case
when grouping_id(ID_FG_AGG) = 1 then 'TOT'
else ID_FG_AGG
end,
...,
sum(NUM_IMPRESA) NUM_IMPRESA,
sum(NUM_UNITALOCALI) NUM_UNITALOCALI,
sum(TOT_ADDETTI) TOT_ADDETTI
from FT_IMPRESA_2011
group by
ID_TER_ITALIA,
rollup(ID_ATE_SEZIONE, ID_ATE_DIVISIONE, ID_ATE_GRUPPO, ID_ATE_CLASSE,
ID_ATE_CATEGORIA),
rollup(ID_ADD_DETT),
rollup(ID_FG_AGG),
...
log errors into MT_IMPRESA_2011_ERRORI reject limit unlimited

```

Le dimensioni del database per la conservazione dei macrodati sono molto elevate. Dal numero di incroci presenti nel piano di spoglio e dal livello territoriale con il quale si vuole diffondere i dati si determinano le dimensioni delle tabelle dei macrodati. Per avere un'indicazione, dalla tabella dei fatti delle imprese che conta circa 2 milioni di record se ne generano circa 20 milioni. Il trattamento di tabelle di tali dimensioni ha consigliato di rivolgere particolare attenzione a tutta la progettazione, con specifico riferimento alle performance. Entrambe le procedure per il popolamento del *data warehouse* primario e secondario sono sufficientemente elastiche da permettere facili e rapide modifiche per un eventuale nuovo caricamento dei dati.

## 1.5 Il data warehouse di diffusione

La piattaforma di software e servizi è stata sviluppata al fine di consentire il rilascio su web di dati statistici multidimensionali secondo un approccio di *data warehouse*. Come evidenziato in precedenza, essa si basa sulla tecnologia .Stat sviluppata dall'OECD e alla cui evoluzione l'Istat partecipa in collaborazione con l'OECD e altri istituti nazionali di statistica. Tale tecnologia è utilizzata all'interno dell'Istat per la diffusione dei dati di Istituto ma, in occasione dei censimenti, si sono implementate ed utilizzate alcune funzionalità aggiuntive, nonché una nuova architettura

tecnologica. Di seguito, si presentano brevi panoramiche di insieme e gli elementi innovativi introdotti in occasione del censimento.

### **1.5.1 L'architettura tecnologica del data warehouse di diffusione<sup>21</sup>**

Da un punto di vista architetturale (Figura 1.12), sinteticamente, l'applicazione è costituita: da un unico *database server*, collocato in DMZ, che funge sia da database 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 backup *settimanale* di tipo *full*; quello lato software di base è stato realizzato tramite gli *snapshot* dei server; quello 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.

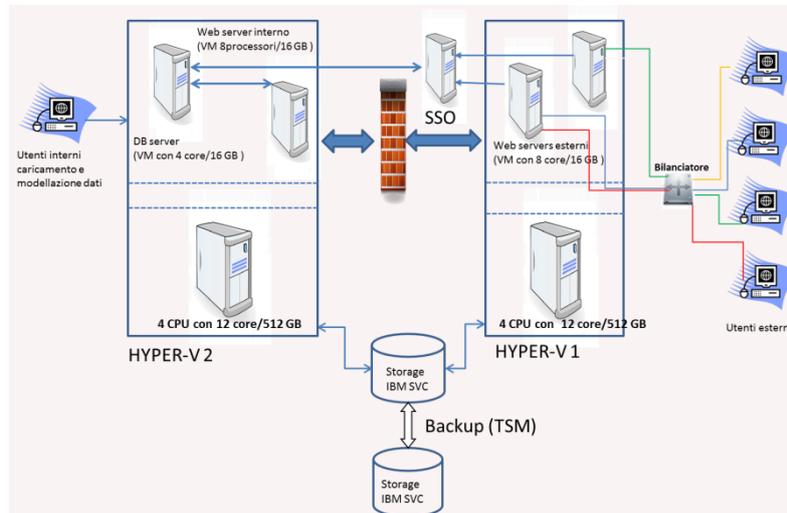
Il sistema gestisce l'autenticazione degli utenti esterni integrandosi con il *Single Sign On* (SSO) e gli standard di sicurezza dell'Istituto.

Il software di base e la configurazione hardware possono essere sintetizzate come riportato nel Prospetto 1.1. L'architettura software, invece, è sintetizzata nella Figura 1.13.

---

<sup>21</sup> Hanno collaborato alla realizzazione dell'architettura: Luca Gramigni, Mario Magarò, Francesco Marchionne, Emiliano Montefiori, Angelo Sardone.

Figura 1.12 – L'architettura tecnologica

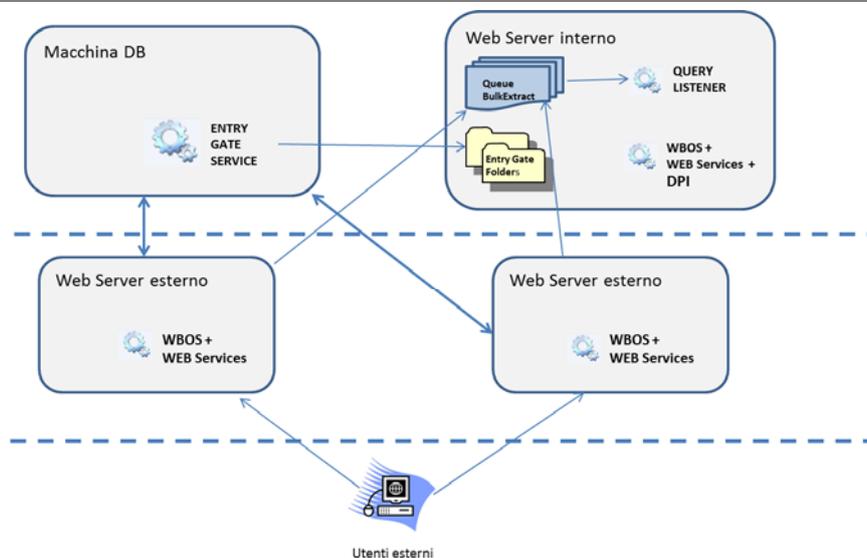


Prospetto 1.1 – Sintesi software di base

MACCHINE	Hw	Os	Web Server	Dbms	Sw Applicativo
HYPER-V (1,2)	4 processori da 12 Core/256 GB	Windows Server 2012			
VMs esterne	8 Core/16GB	Windows Server 2008 r2	IIS 7		.Stat 3.0 web
VM interna (DB)	16 Core/48 GB	Windows Server 2008 r2		SQL Server 2008	.Stat 3.0 DB Server
VM interna (web server)	8 Core/16GB	Windows Server 2008 r2	IIS 7		.Stat 3.0 web/DPI

Sulla macchina riservata alla base dati oltre al DBMS è installato il servizio Windows dedicato al caricamento dei file dei dati sul database. Sui *web server* interno ed esterni sono installati i *web services* dedicati all'estrazione ed esportazione dei dati (in formato *csv*, *xls*, *PC-Axis* ed *SDMX*), alla connessione verso il database ed alla generazione dei grafici oltre alla *web application* dedicata alla navigazione. Sul *web server* interno sono inoltre presenti le *directories* contenenti i dati organizzati secondo i temi visualizzati sul browser ed il servizio Windows dedicato alla gestione delle richieste dei file in formato *csv*.

Figura 1.13 – L'architettura software



Le principali innovazioni utilizzate per il Censimento dell'industria e dei servizi rispetto all'impianto realizzato per il *data warehouse* 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 dei 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.

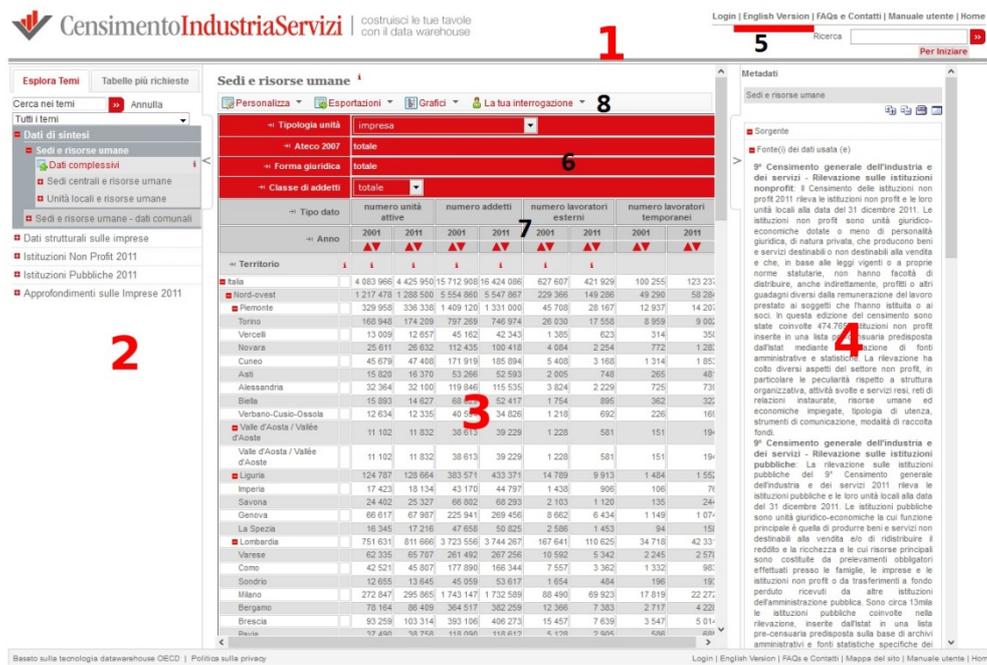
### 1.5.2 L'ambiente di navigazione

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 1.14 – sezione 2).

Tale sezione è organizzata nativamente in temi e sottotemi, 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*) precostituite. 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 sottoargomenti, 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'utente.

Figura 1.14 – L'interfaccia web e la sua divisione in sezioni



L'interfaccia è suddivisa in quattro sezioni: quella superiore (Figura 1.14 - sezione 1), la parte dell'albero dei temi a sinistra (Figura 1.14 - sezione 2), la parte dei dati al centro (Figura 1.14 - sezione 3) e la parte con i metadati a destra (Figura 1.14 - 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'apposito *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 1.14 - 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 sottosezioni: un insieme di dimensioni selezionabili (Figura 1.14 - sezione 6) la tavola dati con la sua testata (Figura 1.14 - sezione 7) e i menu per consentire all'utente la personalizzazione della tavola medesima, l'*export* in diversi formati, il salvataggio delle proprie interrogazioni (Figura 1.14 - sezione 8).

Nel pannello di destra (Figura 1.14 - 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 sottosezione 8 è offerta all'utente la possibilità di:

- disporre gli assi (Figura 1.15 – sezione 9);
- impostare alcune opzioni di visualizzazione (Figura 1.15 – sezione 10);
- esportare i dati in differenti formati - *csv*, *xls*, eccetera (Figura 1.15 – sezione 11);
- realizzare grafici (Figura 1.15 – sezione 12);
- gestire le interrogazioni ad esempio salvandole o combinando dati (Figura. 1.15 – sezione 13).

Figura 1.15 – Il menu della toolbar



### 1.5.3 I metadati

Il sistema di diffusione, oltre ai macrodati quantitativi, rende consultabile un ampio set di metadati forniti al fine di agevolare la comprensione dei contenuti trattati.

Ogni cubo di dati viene accompagnato da metadati referenziali, che descrivono il contenuto dei dati da un punto di vista semantico, la fonte dei dati e, qualora necessario, da metadati di tipo strutturale. Questi ultimi possono essere associati a livello:

- di dimensione di analisi;
- di incrocio di più dimensioni;
- di singola cella.

Selezionando, all'interno dell'albero dei temi presente sulla sinistra, un cubo di dati, viene proposta sulla destra una pop-up con i citati metadati referenziali (Figura 1.16), in cui vengono riportate le informazioni relative alla fonte dei dati utilizzata, in termini di nome del processo e della sua descrizione. Ulteriori informazioni sono reperibili dall'utente cliccando su di un link di collegamento con il sistema *corporate* di Istituto SIDI/SIQual<sup>22</sup>. A tale pop-up l'utente può accedere in ogni

<sup>22</sup> Il Sistema Informativo per la qualità SIDI/SIQual è il sistema trasversale di Istituto per la gestione, l'analisi e la diffusione agli utenti dei metadati e degli indicatori standard di qualità relativi ai processi produttivi statistici di Istituto. Esso è composto da due ambienti: quello di gestione, denominato SIDI-Gestione e quello di navigazione, rappresentato da SIQual.

momento selezionando la “i” posta vicino al nome del cubo.

Relativamente ai metadati di tipo referenziale, il sistema presenta, qualora il loro caricamento si sia reso necessario, una “i” cliccando sulla quale si ha accesso alle informazioni, come presentato nella Figura 1.16.

Relativamente ai metadati di tipo strutturale, il sistema presenta, qualora il loro caricamento si sia reso necessario, una “i” cliccando sulla quale si ha accesso alle informazioni, come presentato nella Figura 1.17.

Figura 1.16 – Schermata relativa ai metadati referenziali

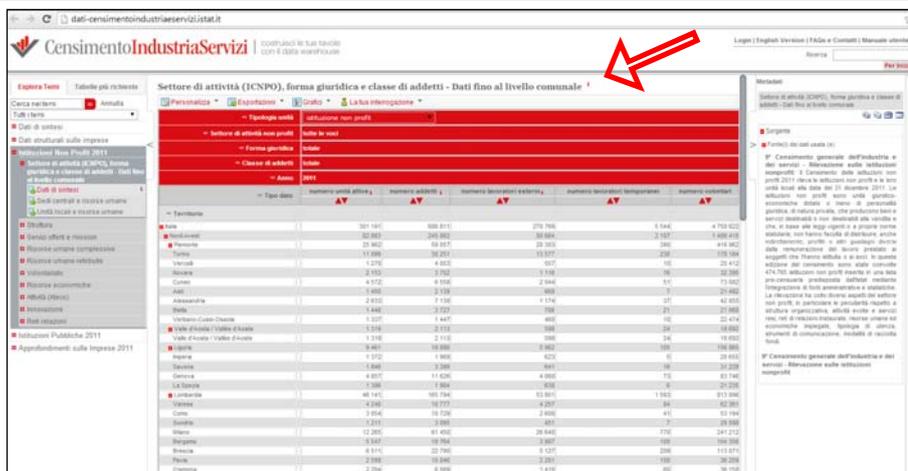
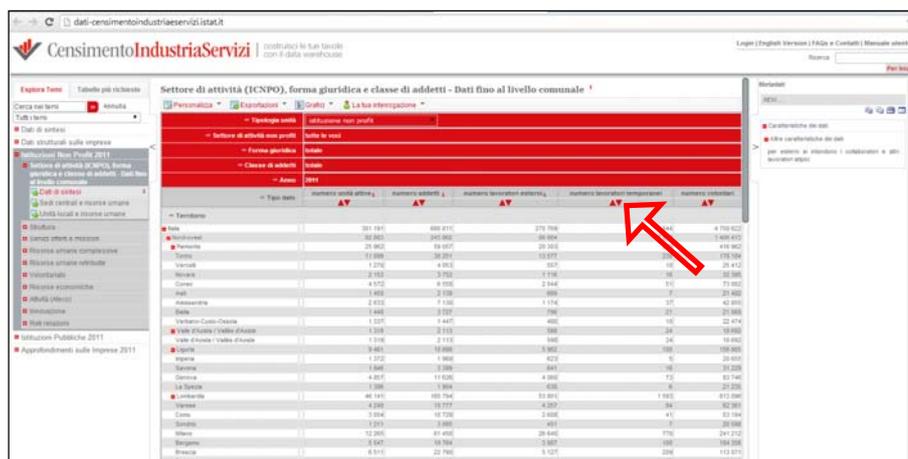


Figura 1.17 – Schermata relativa ai metadati strutturali



Ha fatto parte integrante della progettazione dei cubi di dati la definizione dei relativi metadati referenziali al fine di governare differenti situazioni. Si riportano di seguito alcune casistiche:

- dati di sintesi. È risultato necessario pubblicare, oltre che ai cubi relativi ai singoli processi statistici (registro delle imprese, istituzioni pubbliche, istituzioni non profit, indagine multiscopo sulle imprese) anche alcuni cubi di sintesi, trasversali ai differenti processi. Al fine di creare tale sintesi si sono adottate delle terminologie uguali per tutti i processi (es. per “addetti” e “lavoratori esterni”), avendo cura, contestualmente, di precisarne l’esatto contenuto informativo per ciascun processo;
- riferimenti temporali. Al fine di rendere la lettura della tavola fluida, pur conservando l’esattezza delle informazioni, si è deciso di inserire tutti i riferimenti temporali come metadati referenziali (es. “al 31 Dicembre dell’anno di riferimento”, “nel biennio di riferimento”, “nell’anno di riferimento o stipulati negli anni precedenti”, eccetera);
- informazioni di dettaglio. La voce “altro” è stata generalmente accompagnata da un metadato ausiliario. Ad esempio:
  - per “altro lavoratore esterno” delle imprese attive di Asia si intende ad esempio lavoratori con contratto a progetto, lavoratori casuali, associati in partecipazione;
  - per “altre risorse umane” delle istituzioni non profit si intendono lavoratori temporanei (ex interinali), lavoratori distaccati e/o comandati e religiosi;
  - per “altra modalità di gestione” del servizio erogato dalle istituzioni pubbliche si intendono tutte le forme di servizio non effettuate direttamente;
- voci di sintesi. Nei casi in cui si impone la diffusione di una voce che non sia di tipo semplice, ma un insieme di singole modalità è stato specificato cosa comprenda la voce in questione. Ad esempio, per le categorie professionali delle risorse umane retribuite delle istituzioni non profit dove la voce “dirigenti e imprenditori” include amministratori, dirigenti e/o responsabili di istituzioni sanitarie, scolastiche e di ricerca, di organizzazioni d’interesse nazionale e sovranazionale come partiti politici e organizzazioni sindacali, organizzazioni umanitarie, sportive, culturali e simili.

#### **1.5.4 L'ambiente di amministrazione ed il caricamento dei dati**

Il sistema di diffusione dei dati è un *data warehouse* nella cui architettura si possono sinteticamente distinguere tre componenti:

- un *repository* con i dati della diffusione (*Microsoft SQL Server*);
- delle componenti di *back office* (per la gestione dei dati e per la creazione delle cosiddette *core view*);
- una interfaccia di *front office* (fornita sempre dalla piattaforma .Stat) per offrire all'utente generico la possibilità di ricerche personalizzate.

Il popolamento del *repository* si colloca, all'interno del processo, a conclusione delle operazioni ETL (*Extract, Transform, Load*): i dati a partire dai vari sistemi sorgente, quali ad esempio DB transazionali (OLTP), attraverso varie elaborazioni (selezione dei soli dati di interesse, normalizzazione, eliminazione dei duplicati, codifica, aggregazioni, calcolo di dati derivati, eccetera) vengono memorizzati nelle tabelle del sistema di sintesi (*load*) per essere resi disponibili alla utenza.

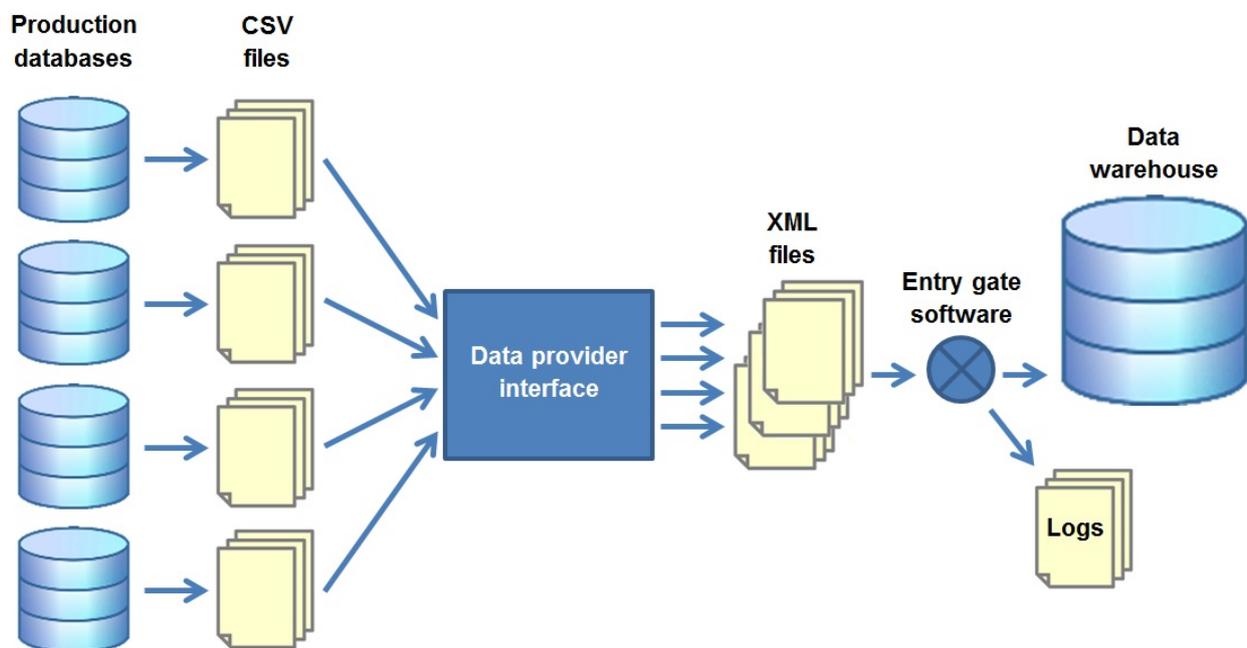
Fisicamente l'operazione di caricamento dati, che comprende sia le operazioni di popolamento dei dati che i successivi aggiornamenti, viene svolta esportando i dati da un DB OLTP, in formato

testuale *csv*, ed importandoli e organizzandoli opportunamente in *dataset* sul *data warehouse* con l'ausilio di un ambiente intranet di amministrazione e caricamento, la *Data Provider Interface*, una *web application* con varie funzionalità specifiche.

La *Data Provider Interface* è un *tool* completo che, oltre al mero caricamento dei dati dai file *csv* di input, consente di gestire la struttura dei *dataset* e delle dimensioni, dei *data content*, la *data presentation*, i *metadata*, le *Core Data view* (delle visualizzazioni predefinite) e la *security* di accesso.

La definizione dei *dataset* e delle dimensioni preliminari al caricamento dei dati permette un verifica a priori dei dati da importare (Figura 1.18).

Figura 1.18 - L'architettura del sistema di caricamento dati



Il caricamento dei dati può essere abbastanza oneroso, a causa della mole dei dati.

La *Data Provider Interface* risolve in modo ottimale tale problematica attraverso un suo componente software che agisce da *manager*, l'*Entry Gate Service*.

Non è la *Data Provider Interface* in sé a caricare dati nel *data warehouse*: essa genera file in formato *xml* che referenziano i file *csv*; i file *xml* generati dalla *Data Provider Interface* sono inviati ad una coda che viene letta dalla applicazione *Entry Gate* (Figura 1.18).

L'*Entry Gate* legge le istruzioni del file *xml*, processa i file *csv* secondo tali istruzioni, aggiorna i dati nel *data warehouse* e produce un corrispondente file di *log*.

In questo modo, se più utenti stanno lavorando al caricamento dei *csv*, l'*entry gate* crea una coda di lavoro partizionando i file *csv* da processare e importando 100 mila righe per volta, in modo da ottimizzare i tempi di *import* e non bloccare l'attività di nessuno. Sono inoltre previsti, mediante posta elettronica, dettagliati servizi di informazione riguardo all'esito di ogni tipologia di operazione condotta.

Il *layout* della presentazione dei dati è realizzato attraverso le "Core Data view", visualizzazioni parziali dell'intero *dataset*, che vengono configurate una volta per tutte. Quando viene fatto l'update dei dati anche le "Core Data view" sono automaticamente aggiornate.

Esse sono organizzate in temi e sottotemi e hanno corrispondenza uno a uno con i *dataset*. Si possono considerare come delle visualizzazioni di default, che gli utenti finali possono utilizzare come punto di partenza per la definizione di visualizzazioni personalizzate per i propri scopi, che poi possono anche esportate in *Excel* o mediante un *Bulk Export*.

## 1.6 Il Single Exit Point per la diffusione dati machine-to-machine

### 1.6.1 L'architettura tecnologica

Il *Single Exit Point* (SEP) consente a chiunque disponga di adeguati strumenti informatici (imprese, istituzioni, altre organizzazioni) di acquisire i relativi dati *machine-to-machine* e di renderli immediatamente fruibili per i propri sistemi informativi, base dati, portali web, eccetera.

Il sistema sviluppato per il Censimento dell'industria e dei servizi e per il Censimento delle istituzioni non profit (SEP CIS) si basa sulla medesima architettura applicativa del *Single Exit Point* utilizzato dall'Istituto per la diffusione *machine-to-machine* dei dati del *corporate data warehouse* (I.Stat), che a sua volta è strutturata sulla piattaforma *SDMX Reference Infrastructure (SDMX-RI)* prodotta da Eurostat. La caratteristica principale di tale infrastruttura applicativa è che, al fine di rendere disponibili i dati in formato SDMX, non si opera alcuna trasformazione preventiva dei dati stessi, ma le relative strutture dati (colonne delle tabelle o viste di database) sono "mappate" da un punto di vista logico, tramite il *tool Mapping Assistant* sulle relative strutture dati SDMX denominate *Data Structure Definition* o DSD, i cui componenti sono catalogati da un punto di vista semantico in dimensioni, attributi e misure.

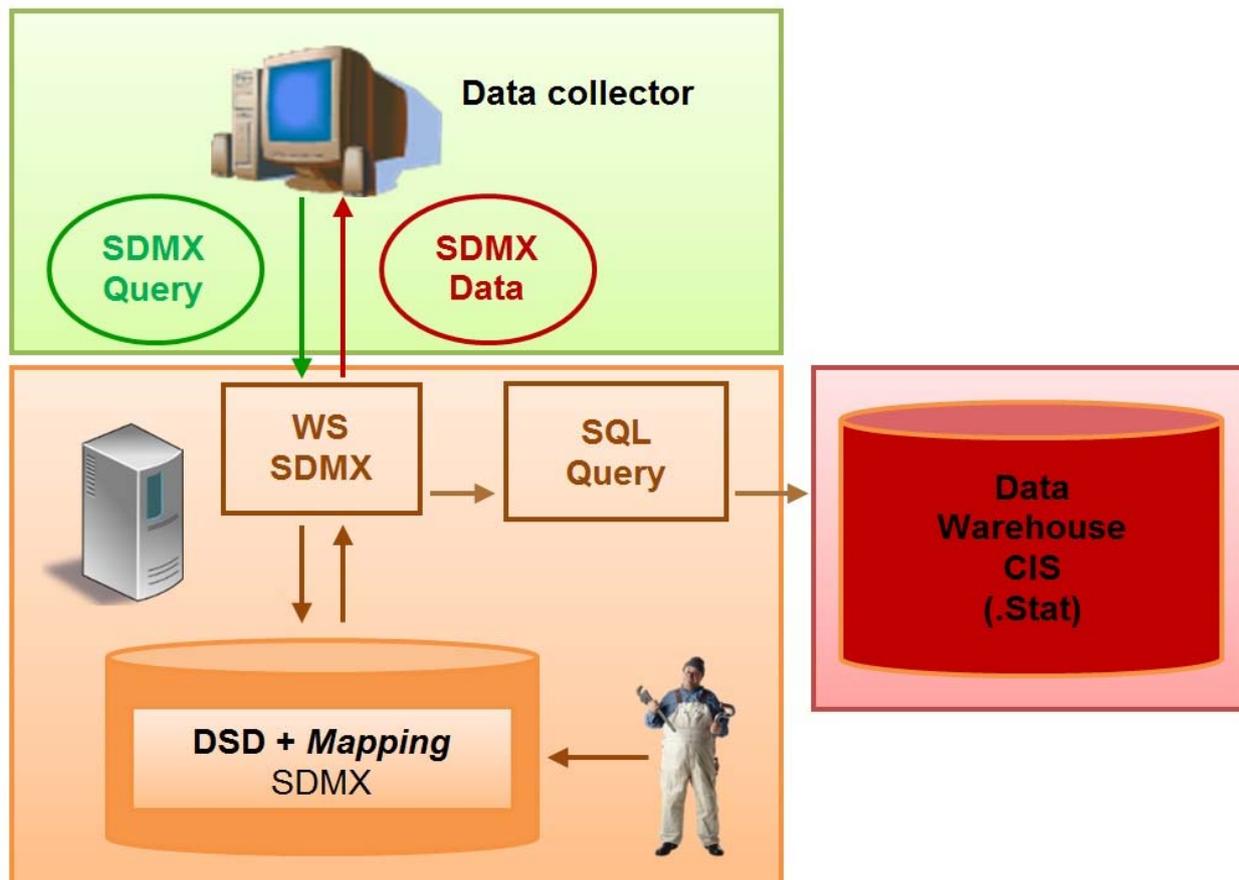
I dati sono messi a disposizione da Istat tramite il *web service* SDMX, che restituisce a *runtime* i dati richiesti dalle macchine degli utenti tramite apposite *query* SDMX. Il *web service* è di tipo SOAP.

Il SEP CIS si connette al *data warehouse* del Censimento dell'industria e dei servizi. Al fine di trasformare la *query* SDMX degli utenti in una *query* SQL al database che contiene i dati da estrarre, il *web service* interroga il database di registro della SDMX-RI denominato *Mapping Store* che contiene:

- metadati strutturali SDMX (o artefatti) relativi ai *dataset* disponibili: nello specifico *Codelist*, *ConceptScheme*, *Data Structure Definition (DSD)*, *CategoryScheme*, *Dataflows*;
- parametri di connessione ai database di diffusione e di mappatura dei relativi *dataset*.

Una volta sottomessa la *query*, il *web service* trasforma in formato SDMX il set di dati restituiti dal *data warehouse* CIS utilizzando anche in questo caso i parametri di mappatura presenti nel *Mapping Store*.

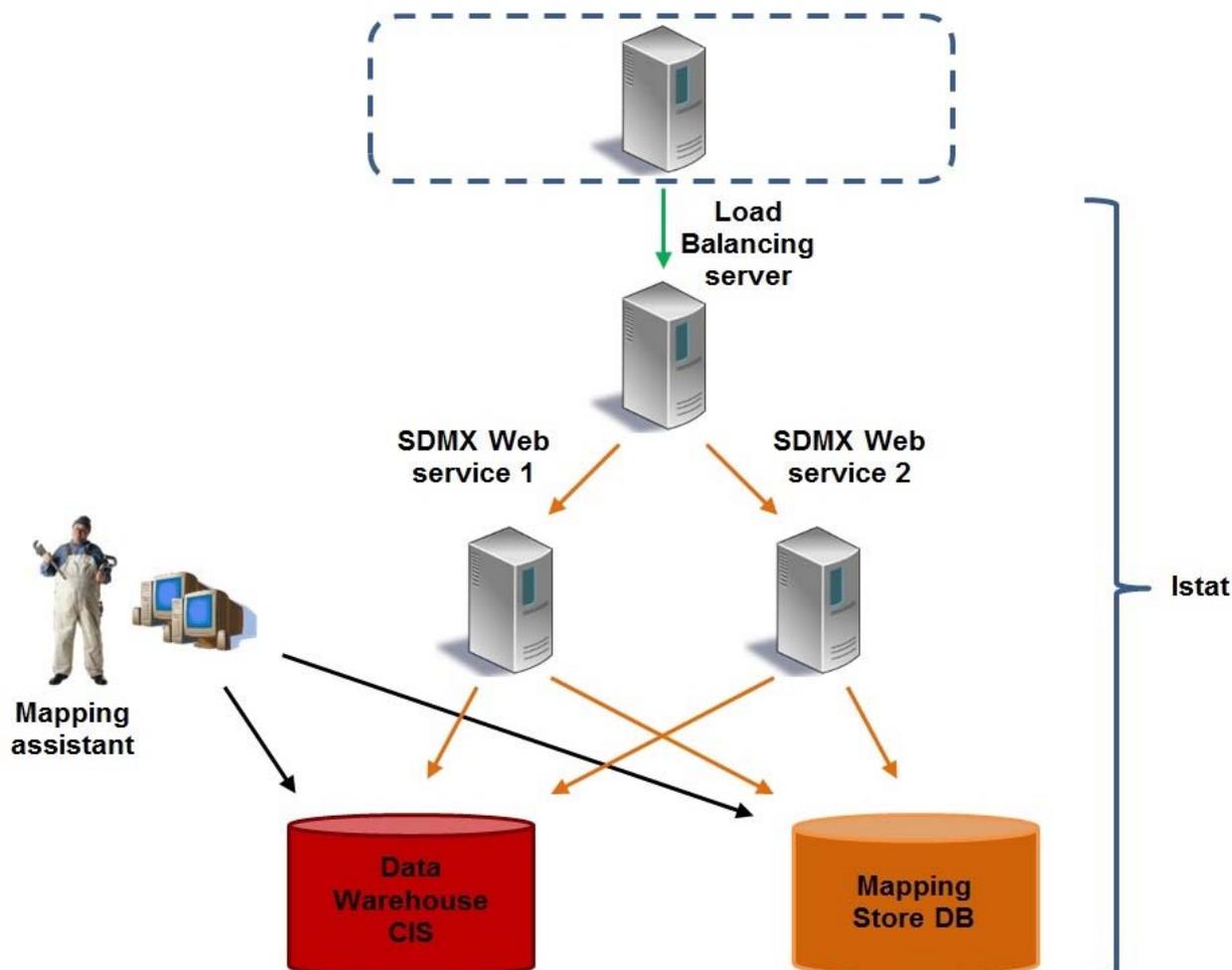
Lo schema dell'architettura, con il flusso funzionale di tipo *query/response* SDMX è riportato all'interno della Figura 1.19.

Figura 1.19 - Architettura funzionale del *Single Exit Point*

In termini di dispiegamento, l'architettura del SEP si basa sui seguenti componenti hardware e software (Figura 1.20):

- due server applicativi con sistema operativo *Windows Server 2008* in configurazione *load/balancing*, ognuno dotato di 8 GB di RAM, in cui i *web services SDMX* sono installati sotto il *web server Internet Information Services (IIS) v. 7.0*;
- un *database server* con 32 GB di RAM e sistema operativo *Windows Server 2008* contenente il *database Mapping Store* installato sotto *Microsoft SQL Server 2008 R2*;
- Il *data warehouse CIS* basato su tecnologia *.Stat*, anch'esso installato sotto *Microsoft SQL Server*;
- I client degli operatori Istat su cui è stato installato il *tool Mapping Assistant*, dotati di sistema operativo *Windows XP* oppure *Windows 7* con una dotazione di RAM a partire da 1 fino a 4 GB.

Figura 1.20 - Architettura di dispiegamento del SEP CIS



### 1.6.2 La modellazione dei cubi di dati in formato DSD e la mappatura SDMX dei cubi di dati

Il processo di resa in disponibilità dei dati del censimento ai fini della diffusione *machine-to-machine* in formato SDMX è stato effettuato tramite due fasi sequenziali distinte, ma strettamente collegate fra loro:

1. la modellazione SDMX dei cubi di dati da cui si generano le strutture dati SDMX (*Data Structure Definition* o DSD);
2. la mappatura dei dati del *data warehouse* sulle strutture dati SDMX precedentemente modellate.

Il primo step è relativo alla modellazione SDMX dei cubi di dati. La modellazione è servita a creare la struttura dati (DSD) nello standard SDMX, per questo ci si è avvalsi di un software per la definizione dei metadati strutturali (o artefatti) SDMX: il *Data Structure Wizard*, anch'esso prodotto da Eurostat. Per effettuare la modellazione SDMX dei cubi di dati del Censimento dell'industria e dei servizi, si è partiti dai relativi piani di spoglio che ne descrivono la struttura (Paragrafo 1.2). Dai piani di spoglio per ogni cubo sono stati individuati tutti i componenti della struttura dati ciascuno

dei quali in SDMX è associato ad un concetto (*Concept*). I Concetti sono classificati all'interno di appositi "contenitori" che in SDMX sono definiti "*Concept Scheme*". Sono stati definiti due *Concept Scheme*: uno per i concetti statistici specifici del Censimento (es. Attività economica, forma giuridica, eccetera) ed uno per i concetti statistici generali oppure per concetti di utilizzo operativo (es. lo stato delle osservazioni, l'unità di misura, il numero di decimali nel valore dell'osservazione, eccetera). I Concetti sono associati ai componenti delle strutture dati SDMX, che possono essere di tre tipi: Dimensioni, Attributi e Misure. Le Dimensioni sono componenti le cui combinazioni di valori identificano univocamente le singole osservazioni; gli Attributi sono componenti che descrivono le caratteristiche dell'osservazione; le Misure rappresentano i valori osservati. Le Dimensioni e gli Attributi codificati sono associati alle *codelist* (CL) relative, che contengono gli insiemi di codici (con le relative descrizioni nelle varie lingue utilizzate) che possono essere associati a tali componenti. In alcuni casi si sono riutilizzate CL già create per altri domini statistici (es. la *codelist* territoriale), in altri casi tali CL sono state create appositamente: in entrambi i casi vi è stato un minuzioso controllo dei codici al fine di verificare la presenza, all'interno delle *codelist*, di tutte le modalità utilizzate nei cubi di dati. Una volta definiti i componenti della *Data Structure Definition* e la loro associazione ai relativi concetti (referenziando opportunamente gli item dei *ConceptScheme*) ed alle *codelist*, la modellazione dei singoli cubi è completata.

Il secondo step è relativo alla mappatura SDMX dei cubi, in altre parole all'associazione delle strutture dati (*Data Structure Definition* o DSD) SDMX preparate con il DSW e le strutture dati degli ipercubi dei dati del *data warehouse* del Censimento dell'industria e dei servizi (*data warehouse* CIS). Per effettuare tale mappatura dati si è utilizzato il *tool Mapping Assistant* (appartenente alla SDMX-RI prodotta da Eurostat). Le fasi di predisposizione della mappatura sono le seguenti:

- 1) creazione in formato SDMX dell'albero gerarchico dei temi (*CategoryScheme*). Tale albero rappresenta gli argomenti in cui si sono suddivisi i contenuti ai fini della navigazione per gli utenti e corrisponde a quello definito per il *data warehouse* CIS. Esso è costituito da un insieme di temi definiti come *Categories* in SDMX, che sono in rapporto gerarchico fra loro (vi sono i temi principali e i relativi sottotemi a cui sono collegati i cubi di dati);
- 2) caricamento delle strutture dati (o DSD) SDMX degli ipercubi di dati, definite nella fase di modellazione precedentemente descritta, all'interno del database di configurazione delle mappature denominato *Mapping Store*, alla base di tutti i *tool* della SDMX-RI. Sono stati inoltre definiti in ambito SDMX i set di dati che si avvalgono di tali strutture dati (cioè associati alle relative DSD), ovvero i *dataflow*: ogni *dataflow*, riferito ad una determinata DSD, è stato associato ad una opportuna *category* del *CategoryScheme* CIS indicato al punto 1;
- 3) per ogni cubo di dati del *data warehouse* CIS è stato definito e memorizzato nel *Mapping Store* il *dataset*, ovvero la struttura dati locale da mappare sulla corrispondente DSD SDMX. Le strutture dati locali sono state ricavate tramite la definizione di apposite viste sui cubi ottenute tramite *query* SQL filtrate ciascuna a livello regionale per ragioni di performance: di fatto per ogni cubo del *data warehouse* sono stati creati 20 cubi regionali per la diffusione *machine-to-machine* in formato SDMX.
- 4) si è mappato ciascun cubo dati sulla corrispondente DSD, mettendo in relazione tra loro i campi del *dataset* con i componenti (Dimensioni, Attributi e Misure) della DSD. Inoltre per ogni dimensione o attributo codificato da mappare obbligatoriamente (o *mandatory*), si è proceduto a mettere in relazione i codici locali utilizzati all'interno del cubo con i codici associati alle relative *codelist* (CL): tale operazione è denominata *Transcoding*.

## 1.7 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.

Per quanto riguarda i dati censuari, l'Istat consente l'utilizzo dei microdati da parte di esterni, attraverso due modalità: quella prevista per gli enti Sistan e il Laboratorio ADELE.

Per quanto riguarda il primo canale, il D.Lgs n. 322/1989 e il D.Lgs n. 196/2003 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 del Sistan.

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 Centre*, 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 da parte di ricercatori e studiosi di università, istituti o enti di ricerca.

Si fornisce di seguito una panoramica sintetica del sistema di diffusione dei microdati dell'Istat e la specifica soluzione adottata per il Censimento dell'industria e dei servizi.

### 1.7.1 Il sistema di archiviazione ARMIDA

Le indagini statistiche misurano uno o più fenomeni sulle unità di una popolazione, che possono essere rappresentate da individui, famiglie e unità giuridico-economiche. Le misure effettuate su ciascuna unità della popolazione (osservazione) costituiscono le variabili rilevate; l'insieme di tali osservazioni sono raccolte in file di microdati.

ARMIDA (ARchivio di MicroDATi) è il sistema nel quale vengono memorizzati e documentati i file di microdati validati prodotti dalle rilevazioni effettuate dall'Istat, per due principali finalità:

- conservare i microdati validati e i metadati delle rilevazioni condotte dall'Istat;
- favorire il riutilizzo dei microdati per finalità statistiche da parte di utenti esterni. I microdati archiviati in ARMIDA vengono attualmente utilizzati per rispondere alle richieste di utenti esterni presentate presso il laboratorio ADELE e alle richieste che pervengono all'Istat da parte di enti facenti parte del Sistan.

Tutti i settori di produzione dell'Istat, prima di procedere alla diffusione di dati relativi a rilevazioni Istat (attraverso volumi, comunicati stampa, tavole di dati, eccetera), devono procedere all'archiviazione in ARMIDA dei microdati dai quali originano le statistiche diffuse. I dati sono, quindi, elaborati presso i settori di produzione e poi memorizzati e documentati centralmente in ARMIDA in base a standard predefiniti (dal 2003).

I file di microdati presenti in ARMIDA sono memorizzati con codifica ASCII con struttura

sequenziale con campi a lunghezza fissa leggibili anche tramite strumenti non proprietari<sup>23</sup>.

Ogni file di microdati è corredato da una serie di informazioni necessarie per descrivere e rendere comprensibile il contenuto di un file di microdati. Tali informazioni, o metadati, sono rappresentati da:

- **tracciato record:** descrive il significato e il formato di ciascuna variabile contenuta nel file di microdati;
- **variabile:** descrive un carattere osservato sull'unità rilevata dall'indagine. Per ogni variabile presente nel tracciato record sono indicate un insieme di attributi quali ad esempio il nome, la descrizione, la lunghezza in caratteri, la tipologia continua o discreta, il formato alfabetico/numerico, l'eventuale quantità di decimali, la scala di misura, eccetera. Tra le informazioni associate viene riportato anche lo stato di riservatezza conseguente alla sua natura (identificativa, sensibile, giudiziaria oppure variabile di lavoro). In ottemperanza alla normativa vigente, i microdati vengono archiviati separando le informazioni riguardanti i dati identificativi e sensibili dal resto dei dati e posti in archivi fisicamente diversi. Inoltre i dati identificativi vengono criptati;
- **classificazione:** ad ogni variabile che nel tracciato è definita qualitativa, viene associata la relativa classificazione in cui sono descritte le modalità, ovvero il codice e la relativa descrizione;
- **allegati:** per ogni tracciato record è possibile memorizzare dei documenti che illustrano le metodologie statistiche di rilevazione, elaborazione, correzione e diffusione dei dati statistici, ovvero il questionario, la nota metodologica, un file di esempio e dei programmi di importazione in ambienti software statistici (es. *R*, *Sas*, *Spss*).

### 1.7.2 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.

<sup>23</sup> In ARMIDA vengono archiviati, in ambienti distinti e dedicati, anche le tipologie di file resi anonimi prodotte dall'Istat per gli utenti esterni, ovvero collezioni di dati elementari d'indagine, nei quali l'anonimità delle unità statistiche, ovviamente già prive di identificativi diretti, viene tutelata tramite l'applicazione di diverse metodologie statistiche che rendono altamente improbabile l'identificazione indiretta delle unità statistiche.

Le richieste provenienti dagli enti Sistan vengono soddisfatte attraverso il sistema ARMIDA.

### **1.7.3 L'accesso ai microdati attraverso il Laboratorio ADELE**

Il Laboratorio ADELE è il *Research Data Centre* dell'Istat, un luogo specificamente pensato per consentire a ricercatori e studiosi di effettuare analisi statistiche su dati elementari nel rispetto delle attuali norme sulla riservatezza statistica. Il Laboratorio ADELE si colloca, nel contesto internazionale, al fianco dei *Research Data Centre* degli altri istituti nazionali di statistica che offrono la stessa tipologia di servizi.

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.

Tramite il Laboratorio, l'Istat mette a disposizione i dati elementari delle rilevazioni di propria titolarità senza modificarne il contenuto informativo, con la sola eccezione di tutti gli identificativi diretti (nome, cognome, codice fiscale, eccetera) che sono eliminati prima di rendere i dati stessi accessibili.

L'accesso al Laboratorio avviene solo a condizione che le analisi del singolo progetto di ricerca non possano essere ottenute tramite altri canali offerti dall'Istat, come lo svolgimento di elaborazioni specifiche da parte dell'Istituto commissionate dal ricercatore stesso (servizio delle elaborazioni personalizzate).

Presso la sede centrale di Roma e le altre sedi territoriali dell'Istituto (è prevista l'attivazione del servizio anche presso le province autonome di Trento e Bolzano) sono a disposizione dei ricercatori terminali che permettono l'accesso da remoto ai server della sede centrale, dove sono collocati i dati e dove avvengono, effettivamente, le elaborazioni.

In generale, le condizioni di utilizzo dei dati nei *Research Data Centre*, le modalità di accesso e le regole di rilascio dei risultati sono condivise, nei tratti essenziali, tra i diversi Paesi e comunque soggette ad un processo di armonizzazione a livello internazionale ed in particolare europeo.

Per quanto riguarda il Laboratorio ADELE, l'accesso avviene 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" (Allegato A.3 al D.Lgs n. 196/2003).

Per fruire del servizio, 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 obbliga al mantenimento del segreto statistico. Se l'autorizzazione viene concessa, il ricercatore può condurre le elaborazioni sui dati elementari richiesti. Una volta concluse le elaborazioni, il risultato viene controllato sotto il profilo della tutela della riservatezza ed, eventualmente, rilasciato al ricercatore. In sintesi, quindi, la riservatezza dei dati presso il Laboratorio ADELE è garantita sotto tre aspetti.

Dal punto di vista fisico, poiché le postazioni per l'accesso al servizio del Laboratorio sono collocate all'interno di sedi dell'Istat. L'accesso alle postazioni è vigilato ed è consentito ai soli utenti autorizzati. I ricercatori, dalle proprie postazioni di lavoro, non hanno accesso alla rete internet ed è fisicamente impedito, nonché formalmente vietato, prelevare o immettere dati.

Dal punto di vista 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.

Dal punto di vista della riservatezza statistica, infine, grazie al controllo al quale sono sottoposti i risultati delle elaborazioni prodotte dal ricercatore prima del loro eventuale rilascio.

### **1.7.4 I file di microdati disponibili per il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit**

Come tutti i microdati delle rilevazioni condotte dall'Istat, anche i microdati relativi al 9° Censimento generale dell'industria e dei servizi (CIS) sono stati archiviati in ARMIDA.

Attualmente i file di microdati<sup>24</sup>, 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:

- 1) Asia-CIS Imprese e occupazione: contiene dati sugli occupati delle imprese analizzati secondo le variabili strutturali censite;
- 2) Asia-CIS Unità locali e occupazione: il file strutturale si riferisce alla totalità delle unità locali rilevate;
- 3) Asia-CIS Unità locali e occupazione: il file disponibile è stato utilizzato per il confronto dei passati censimenti 2001-2011, reso confrontabile dopo avere equiparato il campo di osservazione dei due ultimi censimenti;
- 4) confronto 2001-2011 imprese, istituzioni pubbliche e non profit: il file di microdati, equiparato rispetto al campo di osservazione dei due ultimi censimenti, si riferisce alle imprese, istituzioni pubbliche e non profit del 2001 utilizzate per i confronti 2001-2011;
- 5) istituzioni non profit – unità locali;
- 6) istituzioni non profit – unità istituzionali;
- 7) istituzioni pubbliche – unità istituzionali;
- 8) istituzioni pubbliche – unità locali;
- 9) MPS1: almeno 10 addetti: il file di microdati si riferisce alla rilevazione campionaria sulle imprese del Censimento generale dell'industria e dei servizi 2011;
- 10) MPS1: 3-9 addetti: il file di microdati si riferisce alla rilevazione campionaria sulle imprese del Censimento generale dell'industria e dei servizi 2011.

La fornitura dei file di microdati attraverso ARMIDA può essere realizzata in diversi formati: formato fisso (*dat*) o delimitato (*txt*), *Access*, *xls* 2003, *xls* 2007, *xml*.

Inoltre, la fornitura avviene attraverso un *package* (un archivio compresso) contenente oltre ai microdati nel formato richiesto, corredati da tutti metadati corrispondenti utili a una corretta interpretazione dei dati: tracciato record, classificazioni, altri allegati. Nel caso del 9° Censimento generale dell'industria e dei servizi gli allegati disponibili sono:

- Questionario-per-le-istituzioni-pubbliche;
- Questionario-per-le-istituzioni-non-profit;
- Questionario-per-le-imprese-con-meno-di-10-addetti;
- Questionario-per-le-imprese-con-almeno-10-addetti;
- Nota metodologica Asia\_CIS.

<sup>24</sup> I file di microdati sono da riferirsi a quelli archiviati fino al mese di dicembre 2014.



## Conclusioni

I sistemi informatici a supporto delle fasi di acquisizione gestione monitoraggio e diffusione dei dati del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit (CIS) sono stati realizzati adottando metodologie e modelli standard per lo sviluppo delle applicazioni informatiche, introdotti in occasione del Censimento dell'agricoltura, perfezionati con il Censimento della popolazione e consolidati con il CIS, in grado di ottimizzare tempi e risorse, di produrre software di qualità, di minimizzare le fasi di manutenzione, di tenere sotto controllo tutte le fasi del progetto, di catalogare le applicazioni e di produrre la documentazione appropriata. Sono state inoltre introdotte tecniche di *testing* funzionale, di individuazione delle vulnerabilità e di carico che hanno consentito la messa in produzione di applicazioni ad alto livello di sicurezza minimizzando i rischi di intrusione e di attacco, o di stop per inadeguatezza della infrastruttura tecnologica, in momenti di picco.

I sistemi realizzati, pur non essendo "generalizzati", sono stati progettati in modo da consentire un elevato grado di riutilizzo dei metodi e del software, specialmente per quanto riguarda i sistemi di gestione della rilevazione e i sistemi di costruzione e presentazione dei questionari elettronici, ma anche con riferimento ai sistemi per la produzione automatica dei file per la diffusione elettronica standard sul *data warehouse* I.Stat.

Si riportano alcune informazioni significative dell'usabilità dei vari sistemi a supporto del CIS.

Relativamente al Censimento delle istituzioni non profit e alla Rilevazione sulle imprese, il Sistema di Gestione della Rilevazione (SGR) ha consentito la costruzione di una rete articolata composta da oltre 3.200 operatori, a fronte di oltre 497.192 unità di rilevazione presenti nella lista precensuaria delle istituzioni non profit e di 266.350 unità di rilevazione presenti nel campione delle imprese.

I questionari, utilizzando l'apposita applicazione sono stati tutti registrati online (direttamente dalle unità di rilevazione oppure dagli operatori della rete); in particolare, il 59 per cento direttamente dalle istituzioni non profit e il 79 per cento direttamente dalle imprese.

Nell'ambito del sistema SGR sono state messe a disposizione della rete oltre 60 funzioni per la gestione e il monitoraggio di tutte le fasi di rilevazione.

Relativamente al Censimento delle istituzioni pubbliche il sistema ha permesso una rilevazione totalmente *paperless* e un monitoraggio fino al dettaglio di ciascuna unità, in assenza di una rete di rilevazione esterna. Sono stati inseriti nella Prima fase del censimento, 13 mila unità istituzionali e 46 mila unità locali, mentre nella Seconda fase, circa 24 mila questionari di unità istituzionali e 119 mila questionari di unità locali. La differenza di numero tra le due fasi è dovuta alla rilevazione delle scuole statali di ogni ordine e grado, considerate solo nella Seconda fase.

Con riferimento alla fase di diffusione standard su I.Stat, il sistema realizzato ha consentito la produzione di 71 ipercubi corrispondenti a 110 file csv.



## Riferimenti Bibliografici

- 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.*
- D. Lgs 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.
- D. Lgs 30 giugno 2003, n. 196, recante *Codice in materia di protezione dei dati personali*.
- 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.
- Eurostat. 2012. *Analysis of the Mapping Assistant Generated SQL*.
- Hall, T. 2006. *Oracle PL/SQL Tuning: Expert Secrets for High Performance Programming*. Oracle ACE.
- Brancato, G., R. Carbini, C. Pellegrini, M. Signore, G. Simeone. 2006. *Assessing Quality through the Collection and Analysis of Standard Quality Indicators: The ISTAT Experience*. Proceedings of European Conference on Quality in Survey Statistics. Cardiff UK.
- Brancato, G., R. Carbini, C. Pellegrini, M. Signore, G. Simeone. 2007. *Il sistema SIDI: uno strumento per le analisi di qualità*. Seminario sulla qualità: l'esperienza dei referenti del sistema informativo SIDI – 2a giornata. Contributi Istat.
- Golfarelli, M., Rizzi S. 2006. *Teoria e pratica della progettazione di Data Warehouse*. The McGraw-Hill.
- Istat. 2012. *Guida alla compilazione del questionario – Censimento dell'industria e dei servizi 2011 rilevazione sulle istituzioni pubbliche. Prima fase – aspetti tecnologici*.
- Istat. 2012. *Guida alla compilazione del questionario – Censimento dell'industria e dei servizi 2011 rilevazione sulle istituzioni pubbliche. Prima fase – esempio lista UL*.
- Istat. 2012. *Guida alla compilazione del questionario – Censimento dell'industria e dei servizi 2011 rilevazione sulle istituzioni pubbliche. Prima fase – modello per l'acquisizione elenco unità locali*.
- Istat. 2012. *Guida alla compilazione del questionario – Censimento dell'industria e dei servizi 2011 rilevazione sulle istituzioni pubbliche. Seconda fase – guida al sistema web*.
- Istat. 2012. *Guida alla compilazione del questionario – Censimento dell'industria e dei servizi 2011 rilevazione sulle istituzioni pubbliche. Seconda fase – modello per le Unità Istituzionali e modello per le Unità Locali*.
- Boudy, J.A., Murty U.S.R. 2008. *Graph Theory. Graduate texts in Mathematics*. Springer.

- Santiago J., D. Magallon. 2009. *Critical Path Method*. CEE 320 – VDC Seminar, Stanford University 4 febbraio.
- Eurostat. 2013. *Mapping Assistant user manual v.1.4*.
- Golumbic, M.C. 1980. *Algorithmic Graph Theory and Perfect Graphs*. Academic Press. ISBN 0-12-289260-7.
- OECD. 2013. *OECD.Stat Web Browser User Guide*. Parigi.
- Balestrino, R., A. Gaucci. 2009. *Tecniche di cattura dei dati nei processi di produzione statistica*. Documenti Istat n. 2/2009.
- Malizia, R., F. Lorenzini. 2012. *Il censimento delle istituzioni pubbliche: una grande opportunità per il Sistan*. Convegno Usci "I numeri del futuro", Firenze 6-7 settembre.
- Torelli, R. 2011. *A Generalized System for Aided Development and Monitoring of Web Surveys*. International Methodology Symposium. Statistics Canada, Ottawa 1-4 novembre.
- Torelli, R. 2012. *GINO++, un sistema generalizzato per indagini statistiche via web*. Newsstat 4 febbraio.
- Hennebrueder, S. 2013. *Guide to Java Persistence and Hibernate*.  
<http://www.laliluna.de> (26 giugno 2016).
- SDMX 2.0 standard documentation*.  
[www.sdmx.org](http://www.sdmx.org) (26 giugno 2016).
- SDMX 2.0 XML Schemas (Section 3B – SDMX-ML).  
[www.sdmx.org](http://www.sdmx.org) (26 giugno 2016).
- Struts Quick Start Tutorial.  
<https://www.java.net/today/2004/05/20/Struts.pdf> (26 giugno 2016).
- Venturi, M. (a cura di). 2013. *6° Censimento dell'Agricoltura: Il sistema informatico integrato. Atti del 6° Censimento dell'Agricoltura*. Volume n. 3 Istat.
- Venturi, M., F. Crescenzi, G. Sindoni. 2013. *The role of technology in the 2010-2012 Italian censuses*. NTTS - New Techniques and Technologies for Statistics Conference, Bruxelles 5-7 marzo.
- Venturi, M., G. Bellini, C. Conti, M. Giacommo, G. Mottura. 2012. *Web use for the 6th General Agriculture Census in Italy*. *Journées de Méthodologie Statistique Insee*. Paris Janvier 2012, par. 3, 3.2, 4.

## Elenco allegati disponibili in formato digitale

- Quadro normativo di riferimento e disposizioni tecniche
- Newsletter
- Questionari e materiali per le rilevazioni
- *Frequently asked questions*