

Atti del 6° Censimento Generale dell'Agricoltura

IL SISTEMA INFORMATICO INTEGRATO

3



6° Censimento Generale
dell'Agricoltura

 Istat
Istituto Nazionale
di Statistica

RACCOGLIAMO RISPOSTE, SEMINIAMO FUTURO.

Atti del 6° Censimento Generale dell'Agricoltura

IL SISTEMA INFORMATICO INTEGRATO

Atti a cura di: Giampaola Bellini

Fascicolo 3 a cura di: Marina Venturi, Stefano De Francisci, Filippo Lioy, Maura Giacummo

Atti del 6° Censimento Generale dell'Agricoltura

IL SISTEMA INFORMATICO INTEGRATO

© 2013

Istituto nazionale di statistica

Via Cesare Balbo, 16 - Roma

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.

DISTRIBUITO DA
STEALTH
BY SIMPLICISSIMUS BOOK FARM

Indice

Presentazione dell'opera	5
Il 6° Censimento generale dell'agricoltura	7
Introduzione	11
Capitolo 1 - L'architettura	15
1.1 L'infrastruttura centralizzata	15
1.2 L'architettura logico/fisica.....	16
1.3 La definizione dei requisiti.....	17
1.4 La modellazione dei requisiti funzionali	17
1.5 La progettazione del software	18
1.6 Test delle applicazioni.....	26
1.7 Sicurezza dei dati.....	29
Capitolo 2 - La rilevazione	31
2.1 Il Portale della Rete	31
2.2 Il Sistema di acquisizione dei dati via web	35
2.3 Il Sistema di Gestione della Rilevazione	45
2.4 L'infrastruttura di gestione dei dati.....	61
Capitolo 3 - La diffusione dei risultati	71
3.1 Architettura e componenti generali del sistema di diffusione dei dati	71
3.2 La fornitura dei microdati	73
3.3 La diffusione dei dati aggregati	74
3.4 Visualizzazione e rappresentazione grafica dei dati.....	78
Conclusioni	82
Riferimenti bibliografici	83
Allegati in formato digitale (cfr. cd-rom allegato al fascicolo <i>Il quadro normativo</i>)	
1. Manuale del Sistema di Gestione della Rilevazione (SGR)	
2. Guida alla compilazione on-line del Questionario elettronico	
3. Regole di controllo del Questionario on-line	





Presentazione dell'opera¹

Gli Atti del 6° Censimento generale dell'agricoltura, vista la rilevanza e la complessità degli argomenti trattati, sono stati strutturati in forma di cofanetto, articolato in cinque fascicoli che raccontano il Censimento attraverso l'iter normativo percorso, nonché le principali soluzioni organizzative, tecniche e metodologiche messe a punto e adottate per questa tornata censuaria. L'opera è stata realizzata in maniera unitaria, essendo evidenti le interconnessioni delle fasi e dei processi produttivi, ma ciascun fascicolo è autonomo nella trattazione tematica, ed è stato realizzato per affrontare in maniera esauriente il tema trattato. Il cofanetto è corredato di un cd-rom, collocato all'interno del primo fascicolo, contenente i documenti ufficiali del Censimento, in formato .pdf.

I cinque fascicoli che costituiscono il cofanetto sono i seguenti:

1. *Il quadro normativo;*
2. *L'organizzazione della rilevazione e le informazioni censuarie;*
3. *Il sistema informatico integrato;*
4. *Il Censimento sul territorio;*
5. *La valutazione della qualità.*

Il primo fascicolo *Il quadro normativo* contiene le norme internazionali e quelle nazionali sulla base delle quali si è progettato e realizzato il Censimento. Sono presentati e descritti nei loro aspetti principali la legge di indizione e finanziamento, il DPR relativo al regolamento di esecuzione, una serie di atti, che sono stati adottati per garantire un adeguato funzionamento della complessa macchina organizzativa, nonché le circolari operative emanate dall'Istat per definire e dare attuazione a tutte le fasi censuarie.

Il secondo fascicolo *L'organizzazione della rilevazione e le informazioni censuarie* descrive le linee guida adottate nell'organizzazione della rilevazione, la composizione e le caratteristiche della rete di rilevazione, nonché la modalità di definizione ed erogazione della formazione agli operatori censuari. Vengono altresì descritte le caratteristiche generali e tecniche della rilevazione, percorrendo le principali tappe che hanno portato dalla definizione della lista pre-censuaria di unità agricole e dei contenuti del questionario alla diffusione dei risultati.

Il terzo fascicolo *Il sistema informatico integrato* descrive l'architettura e le soluzioni informatiche adottate per un Censimento che per la prima volta utilizza la modalità di compilazione on-line e la registrazione dei questionari presso gli uffici di rilevazione. Tutti i processi adottati sono connessi e dialoganti tra loro garantendo coerenza logica, velocità e trasparenza al sistema e alle operazioni realizzate. Sono descritti il Sistema di Gestione della Rilevazione (SGR) e l'impianto informatico e gli strumenti di diffusione dei dati.

Il quarto fascicolo *Il Censimento sul territorio* descrive, da una parte, come la novità dell'organizzazione decentrata alle amministrazioni regionali delle operazioni censuarie, seppure sotto l'egida di norme e linee guida dettate dal centro, sia stata supportata dagli uffici territoriali Istat, dall'altra, come la realizzazione di un Censi-

¹ Si desidera ringraziare per l'attento lavoro di revisione e per i preziosi suggerimenti forniti Franco Lorenzini e Massimo Greco.



mento secondo una modalità organizzativa più tradizionale e più centralizzata per una parte delle regioni italiane sia stata gestita con un forte coordinamento dell'Istat a livello locale.

Il quinto fascicolo *La valutazione della qualità* descrive le metodologie adottate e i risultati ottenuti dalle due indagini di qualità: 1) l'indagine di misura è volta alla stima della distorsione e delle varianze di risposta dovute al complesso processo di rilevazione; 2) l'indagine di copertura è volta alla stima del tasso di sotto-copertura del 6° Censimento generale dell'agricoltura.

Il 6° Censimento generale dell'agricoltura

a cura di **Andrea Mancini**

Direttore del Dipartimento per i censimenti e gli archivi amministrativi e statistici dell'Istat

Come avviene tradizionalmente in occasione di ogni censimento, l'Istituto nazionale di statistica redige gli "Atti" del censimento, pubblicazione mediante la quale descrive le caratteristiche tecniche, organizzative e metodologiche delle attività svolte. In occasione del 6° Censimento generale dell'agricoltura si è scelto di pubblicare un cofanetto di cinque fascicoli tematici, corredato da un supporto informatico, collocato all'interno del primo fascicolo *Il quadro normativo*, contenente copia delle norme di riferimento e di altra documentazione prodotta dall'Istat. Gli "Atti del 6° Censimento generale dell'agricoltura" rappresentano lo strumento che garantisce la "memoria storica" del Censimento, punto di riferimento essenziale per chi è interessato a conoscere come è stata progettata ed eseguita una tra le rilevazioni più importanti del Paese. In questa pubblicazione sono descritte tutte le fasi dell'indagine censuaria in modo da consentire al lettore di conoscere le motivazioni delle strategie adottate, le soluzioni organizzative e tecniche, il processo di produzione dei dati.

Il 6° Censimento generale dell'agricoltura ha avuto luogo con riferimento alla data del 24 ottobre 2010 e ha aperto la tornata censuaria del 2010-2011^{II} contraddistinta da un ampio uso di dati amministrativi e da una forte attenzione all'innovazione. Esso è stato svolto con l'obiettivo principale di delineare un quadro informativo statistico, aggiornato e dettagliato della struttura delle aziende agricole e zootecniche operanti in Italia e dei metodi di produzione agricola adottati. La fotografia dell'agricoltura italiana emersa dai dati del censimento ha mostrato le trasformazioni che hanno interessato il comparto nel decennio 2000-2010, caratterizzato dalla crisi economica, dalla riforma della Politica agricola comunitaria (PAC) e dall'enfasi con cui è stata promossa l'adozione di metodi di produzione rispettosi dell'ambiente. Inoltre, il Censimento ha assolto gli obblighi internazionali di produrre statistiche comunitarie comparabili sui fenomeni agricoli, sulle condizioni di sviluppo rurale e di sostenibilità ambientale^{III} - necessarie ai fini di programmazione e valutazione della Politica agricola comunitaria - sulle superfici viticole^{IV}, utilizzando definizioni, concetti e metodi armonizzati a livello europeo. Inoltre, l'estensione delle analisi ai confronti storici fino al Censimento agricolo del 1982 ha permesso di delineare il quadro delle trasformazioni strutturali dell'agricoltura nazionale nell'ultimo trentennio.

Dal punto di vista metodologico e tecnologico sono state introdotte rilevanti innovazioni, in parte riutilizzate con adattamenti per il 15° Censimento generale della popolazione e delle abitazioni e il 9° Censimento generale dell'industria e dei servizi, condotti nel 2011 e 2012.

La principale innovazione è stata l'ampio uso di dati amministrativi sia per la pre-

II 6° Censimento generale dell'agricoltura, 15° Censimento generale della popolazione e delle abitazioni, 9° Censimento generale dell'industria e dei servizi.

III *Regolamento (CE) n. 1166/2008 del Parlamento europeo e del Consiglio del 19 novembre 2008* avente per oggetto l'istituzione di un quadro di riferimento per la produzione di statistiche comunitarie comparabili sulla struttura delle aziende agricole e per un'indagine sui metodi di produzione agricola.

IV *Regolamento (CEE) n. 357/79 del Consiglio del 5 febbraio 1979* e successive modificazioni, concernente le indagini statistiche sulle superfici viticole.



disposizione della lista pre-censuaria sia per il controllo - anche contestuale alla rilevazione - e la correzione dei dati raccolti. 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 lista è stata predisposta mediante integrazione di fonti amministrative a carattere generale e di fonti amministrative e statistiche specifiche del settore tra le quali l'archivio dei Fascicoli aziendali, contenuti nel Sistema informativo agricolo nazionale (Sian), e l'archivio delle Anagrafi zootecniche. Nel complesso sono state utilizzate 17 fonti amministrative o statistiche, dopo aver eseguito nel 2008 la rilevazione sperimentale e di controllo del prototipo di lista pre-censuaria, svolta in collaborazione con gli uffici di statistica delle Regioni e Province autonome.

Salvaguardando la continuità storica rispetto ai Censimenti precedenti, altre significative innovazioni sono state apportate attraverso:

- la nuova rilevazione sui metodi di produzione agricola adottati dalle aziende agricole, svolta dall'Italia in forma censuaria, sebbene per la Unione Europea fosse sufficiente l'esecuzione in forma campionaria;
- il campo di osservazione individuato dal solo Universo Ue e non anche l'Universo Italia come avvenuto nei precedenti censimenti;
- l'individuazione ex ante delle micro unità con segnali amministrativi di possibile attività di coltivazione o allevamento da escludere dalla rilevazione;
- la definizione del metodo per identificare ex ante l'unitarietà tecnico-economica dell'azienda agricola;
- l'intervista effettuata, a cura del rilevatore, presso la residenza o sede legale del conduttore anziché presso il centro aziendale;
- l'individuazione ex ante della possibile localizzazione del centro aziendale elaborando i dati di archivio;
- la pubblicazione di dati riferiti al comune di localizzazione dei terreni/allevamenti^V in aggiunta alla tradizionale pubblicazione di dati riferiti al centro aziendale;
- la richiesta di informazioni volte a consentire la georeferenziazione del centro aziendale e la stima dei volumi d'acqua aziendali consumati per l'irrigazione, nonché notizie relative alla cittadinanza della manodopera, alla presenza di impianti per la produzione di energie rinnovabili, alle superfici dedicate a colture energetiche;
- l'esclusione dal campo di osservazione delle unità esclusivamente forestali.

Ulteriori innovazioni tecniche hanno riguardato:

- la tecnica di rilevazione multicanale (*mixed mode*), con facoltà per le aziende agricole di compilare il questionario direttamente via internet;
- l'architettura informatica interamente on-line, che ha riguardato tutte le macro funzioni del processo produttivo;
- il ricorso, per la prima volta in Italia, ai social network come ausilio alle attività degli addetti al censimento;
- l'utilizzo di strumenti a distanza di e-learning a supporto delle attività formative degli addetti al Censimento;
- un nuovo impianto metodologico del sistema di controllo e correzione dei dati, assai accurato e supportato dal confronto con i dati amministrativi;

^V Riprendendo quanto pubblicato in occasione del Censimento dell'agricoltura del 1990, non replicato in quello del 2000.

- la diffusione assai tempestiva di dati provvisori di ampio contenuto informativo, con disaggregazione dell'informazione senza precedenti rispetto ai passati censimenti;
- la diffusione dei risultati definitivi mediante il sistema di data warehousing dell'Istat, I.Stat.

Il Sistema di Gestione della Rilevazione (SGR), nucleo portante dell'architettura informatica del Censimento, ha consentito di seguire l'andamento di tutte le fasi del processo di produzione per singolo organo di censimento, rilevatore o coordinatore, grazie alla realizzazione di oltre cinquanta funzioni componenti del sistema. SGR è stato il Centro di controllo dell'organizzazione ed è stato utilizzato autonomamente da ciascun organo di censimento, rilevatore o coordinatore, che attraverso proprie *userid* e *password*, ha avuto accesso alle funzioni e alle tabelle di pertinenza.

In generale, la realizzazione dei sistemi informatici a supporto della fase di acquisizione, monitoraggio, controllo, correzione e diffusione dei dati del 6° Censimento generale dell'agricoltura 2010 è stata condotta utilizzando tecniche di progettazione, modellazione e sviluppo software in un'ottica di standardizzazione complessiva. 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 implementazione e presentazione dei questionari elettronici, tanto da essere stati di fatto riutilizzati, con adattamenti, nel 15° Censimento della popolazione e delle abitazioni e nel 9° Censimento dell'industria e dei servizi. Nondimeno SGR ha presentato alcune criticità connesse alla complessità della rete e dell'organizzazione del lavoro. All'avvio del sistema di compilazione on-line da parte delle aziende agricole, il consistente numero di accessi contemporanei ha messo in evidenza un comportamento anomalo dell'applicazione, non evidenziato in fase di test. È stato necessario, a poche ore dalla partenza, sospendere il servizio e sottoporre l'applicazione a test più approfonditi. SGR è stato riaperto alla rete territoriale di rilevazione dopo 48 ore, mentre il sistema di acquisizione on-line è stato riaperto alle aziende dopo due settimane, con conseguente abbassamento delle risposte via web da parte delle aziende agricole rispetto alle aspettative iniziali. Superate le difficoltà iniziali, il sistema ha permesso di gestire una rete articolata composta da più di 20.000 operatori. I questionari relativi alle oltre 2 milioni di unità di rilevazione sono stati registrati on-line, in tutto o in parte: circa 1.352.000 questionari sono stati inseriti via web dagli operatori di censimento delle regioni ad Alta partecipazione (66,1 per cento); circa 636.000 notizie di base riferite alle variabili primarie sono state inserite dagli operatori degli Uffici Comunali di Censimento (UCC) delle regioni a Partecipazione integrativa attraverso l'apposita funzionalità a disposizione in SGR (31,0 per cento) e 61.000 questionari sono stati compilati via web direttamente dalle aziende agricole (2,9 per cento).

Sul piano dell'organizzazione il Censimento ha adottato schemi a "geometria variabile", coinvolgendo assai più delle precedenti edizioni gli uffici di statistica e le direzioni di agricoltura delle Regioni e Province autonome. La flessibilità organizzativa ha trovato il suo compimento nella definizione di due modelli organizzativi alternativi, uno ad "Alta partecipazione" e uno a "Partecipazione integrativa", che la singola Regione o Provincia autonoma poteva adottare. Nel primo modello la Regione o Provincia autonoma ha definito in autonomia la struttura della propria



rete territoriale. Nel secondo modello il ruolo della Regione è stato minore e analogo a quello già espletato nel precedente Censimento in quanto l'organizzazione generale è stata a carico dell'Istat ed è spettato ai Comuni il compito di eseguire la rilevazione sul campo.

A conclusione di queste brevi note si vuole sottolineare che la trasmissione ad Eurostat delle informazioni di interesse comunitario raccolte dal Censimento è avvenuta in anticipo rispetto ai termini fissati per l'Italia dal Regolamento, che il rapporto di qualità trasmesso dall'Istituto a Eurostat è stato da questo vagliato positivamente, che i numerosi dati provvisori sono stati diffusi a meno di 9 mesi dalla data di riferimento e a 5 mesi dalla data di chiusura delle operazioni sul campo prefissata dal calendario censuario, che la diffusione dei dati definitivi è iniziata a luglio 2012 e terminata a dicembre 2012. Inoltre, le indagini di controllo della qualità e della copertura del Censimento hanno restituito il quadro della qualità statistica dei dati prodotti in modo trasparente e disponibile nel quinto fascicolo di questi "Atti". I risultati sono stati ottenuti grazie al lavoro svolto da tutte le componenti, interne ed esterne all'Istituto, che hanno operato nella rilevazione, fin dalle attività svolte dal "Comitato consultivo per la preparazione a livello regionale del 6° Censimento generale dell'agricoltura", dove la condivisione delle scelte e la trasparenza delle decisioni ha rappresentato un rilevante fattore di efficacia per il funzionamento della macchina censuaria.

Introduzione

Le innovazioni di metodo, tecniche e organizzazione del 6° Censimento generale dell'Agricoltura hanno avuto un rilevante impatto sulla definizione della struttura informatica di servizio alla gestione della rilevazione.

L'uso di più fonti amministrative e la realizzazione della lista pre-censuaria di aziende agricole ha indotto la predisposizione di funzioni di assegnazione delle unità in lista agli Uffici di censimento di competenza e ai loro coordinatori e rilevatori.

L'adozione di una organizzazione della rete territoriale basata su criteri di flessibilità e sulla diversificazione dei ruoli e delle funzioni assegnate ai vari livelli ha imposto la realizzazione di profili di accesso diversificati al sistema informatico.

L'adozione di una tecnica multicanale di acquisizione delle informazioni ha richiesto la creazione di un questionario elettronico direttamente fruibile dalle aziende agricole via internet per la auto-compilazione, ma utilizzabile anche dagli operatori della rete di rilevazione per l'inserimento dei dati raccolti precedentemente su supporto cartaceo.

Più in generale, le citate innovazioni hanno richiesto l'approntamento di un sistema informatico di gestione della rilevazione in condivisione tra tutti i livelli territoriali della rete organizzativa.

Le caratteristiche della rilevazione censuaria, gli elevati numeri in gioco e la stretta integrazione con gli strumenti di gestione e monitoraggio della rilevazione hanno reso il progetto di sviluppo software del questionario on-line particolarmente complesso. Per offrire un servizio adeguato all'importanza delle applicazioni e alla vastità della platea degli utilizzatori, doveva essere garantito un alto livello di prestazioni, con tempi di risposta rapidi, indipendentemente dal numero di utenti contemporaneamente operativi sul sistema, così come la massima robustezza. Per assicurare un elevato standard di sicurezza, andavano previsti controlli a vari livelli in modo da preservare l'integrità dei dati, anche in presenza di tentativi di attacco al sistema, tesi ad aggirare i controlli previsti nell'interfaccia.

L'interfaccia del sito è stata quindi progettata cercando di rendere la compilazione il più possibile rapida e agevole, bilanciando i numerosi controlli sui dati inseriti tra la necessità di garantire un'alta qualità dei dati, con l'esigenza di non sovraccaricare l'utente di messaggi di errore bloccanti, tutto ciò garantendo la massima compatibilità sui browser più diffusi e rendendo disponibile l'applicazione in tre lingue (italiano, tedesco e sloveno).

Il questionario on-line ha soddisfatto inoltre il criterio di alta affidabilità attraverso l'uso di strategie di ridondanza (*mirroring, clustering e load balancing*) su tutte le componenti essenziali. Sono state anche realizzate procedure di *recovery* finalizzate a minimizzare i tempi di fermo del sistema, dovuti a guasti o attività di manutenzione programmata.

L'adozione di una tecnica di rilevazione multi-canale richiedeva un'organizzazione complessa e strutturata, capace di predisporre e gestire i vari processi integrandoli. Di conseguenza c'era il bisogno di monitorare in tempo reale e costantemente tutte le fasi del processo, con particolare attenzione al recupero dei questionari, per evitare ad esempio che i Rilevatori contattassero aziende che avevano già risposto on-line. Così, se da una parte le aziende dovevano essere in grado di poter compilare un que-



stonario elettronico mediante una semplice connessione ad internet, analogamente doveva essere disponibile a tutti gli operatori della rete censuaria, oltre 20.000, un sistema in grado di monitorare in tempo reale sia l'arrivo dei questionari compilati su web da parte delle aziende agricole, sia l'arrivo dei questionari cartacei in modo da permettere le operazioni di rilevazione sul campo da parte dei Rilevatori solo verso i non rispondenti effettivi, con notevoli guadagni in termini di tempo, efficienza e riduzione dei costi.

Al questionario elettronico è stato affiancato un sistema di gestione della rilevazione, 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. La sua struttura modulare ha consentito ulteriori implementazioni in corso d'opera senza stop del sistema e modifiche a quanto già rilasciato e funzionante.

L'intera operazione censuaria è stata monitorata attraverso specifiche funzioni rese disponibili da un'applicazione web progettata per assicurare che il monitoraggio non interferisse con le altre attività, ma fosse di ausilio alla gestione della rilevazione.

Il sistema ha permesso:

- la gestione delle fasi di rilevazione, revisione, registrazione, controllo e correzione dei questionari e la registrazione delle variabili principali del questionario;
- la gestione della rete di rilevazione (operatori e profili utente);
- il monitoraggio dell'andamento delle operazioni sul campo;
- la gestione in tempo reale dell'acquisizione multicanale dei questionari (web, cartacea).

Il sistema ha funzionato da centro di controllo dell'intera organizzazione ed è stato utilizzato autonomamente da ciascun organo della rete censuaria, che ha avuto accesso alle funzioni peculiari del proprio ruolo e dei propri compiti attraverso un modulo di autenticazione unico. La struttura della rete di rilevazione comportava, infatti, una serie di controlli a cascata con specifiche responsabilità derivanti dai ruoli e compiti assegnati a ciascun attore in modo tale che il numero di funzioni disponibili crescesse al crescere del livello gerarchico nella rete di rilevazione.

Il portale di documentazione della rete è stato realizzato in modo perfettamente integrato con il sistema di gestione della rilevazione, accessibile attraverso le medesime credenziali di accesso, con l'obiettivo di contenere in un unico *repository* tutte le informazioni opportunamente strutturate in aree utili alla corretta esecuzione del censimento da parte della rete e, nel contempo, di fornire anche strumenti di *e-learning* per consentire l'autoformazione da parte degli operatori coinvolti nelle operazioni censuarie.

Si può ragionevolmente affermare che il 6° Censimento generale dell'agricoltura è stato un Censimento completamente assistito da web attraverso la predisposizione di un sistema software complesso articolato in più componenti.

All'interno del portale principale <http://censimentoagricoltura.istat.it> sono stati realizzati tre sottosistemi integrati tra loro e accessibili mediante un sistema protetto da *userid* e *password*, costruito nel rispetto dei criteri stabiliti dalla normativa vigente in materia di sicurezza e tutela della riservatezza:

- <http://censimentoagricoltura.istat.it/ACQUIS> per la compilazione on-line del questionario di rilevazione da parte delle aziende agricole;

- <http://censimentoagricoltura.istat.it/RETE> di tipo documentativo, accessibile dagli operatori del censimento, contenente tutte le informazioni utili per lo svolgimento delle operazioni censuarie;
- <http://censimentoagricoltura.istat.it/SGR> il vero e proprio sistema di gestione della rilevazione, a disposizione della rete territoriale per supportare i vari organi di rilevazione nell'esecuzione del Censimento in tutte le fasi previste per l'acquisizione, la validazione e il monitoraggio della fase di acquisizione.

Nel presente fascicolo viene descritto in modo dettagliato il sistema informatico, sia relativamente all'architettura generale e all'infrastruttura tecnologica, sia con riferimento alle funzionalità realizzate. In particolare nel Capitolo 1 sono illustrate l'architettura logica e fisica generale del sistema, le modalità seguite per definire e modellare i requisiti funzionali dell'intero sistema, la progettazione del software e delle sue componenti, i test delle applicazioni realizzate. Il Capitolo 2 è dedicato alle descrizioni delle tre componenti principali del sistema informatico: il portale documentativo a disposizione della rete di rilevazione, il sistema di acquisizione via web dei dati del questionario e il sistema di gestione della rete di rilevazione. Sempre nella seconda parte è descritta l'infrastruttura centralizzata di gestione e memorizzazione dei dati e dei metadati. Nel Capitolo 3 vengono affrontate le problematiche relative alla diffusione dei dati e illustrate le principali soluzioni adottate.





Capitolo 1

L'architettura¹

1.1 L'infrastruttura centralizzata

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. In particolare, anche se l'uso di dispositivi quali gli *smartphone* è stato considerato non adatto per la compilazione *on-line* del complesso questionario, alcune regioni ad alta partecipazione hanno comunque scelto di dotare i propri Rilevatori sul campo di dispositivi di tipo *tablet* per la compilazione del questionario contestualmente all'intervista somministrata all'azienda agricola.

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, oltre che per i motivi sopra esposti anche da ragioni organizzative, economiche, di complessità e consistenza.

La Figura 1.1 seguente schematizza le componenti del sistema e gli attori coinvolti.

Figura 1.1 - I moduli e gli utenti del sistema

Aziende agricole	Istat	Rilevatori sul campo	Organi censuari
Internet			
Questionario on-line	Portale della Rete		
	Sistema di Gestione della Rilevazione	Sistema di validazione interattiva	Sistema documentale
Gestione dati – DB centralizzato			

I diversi utenti del sistema hanno potuto accedere in modalità differenziata. Le aziende agricole hanno avuto a disposizione il solo questionario elettronico, da compilare autonomamente via web. Tutti gli organi censuari hanno avuto accesso al Sistema di Gestione della Rilevazione (SGR). Le regioni che hanno optato per l'alta partecipazione hanno potuto accedere anche al questionario elettronico, per la registrazione dei dati dei propri questionari cartacei, spesso effettuato da una particolare classe di Rilevatori, denominati "Rilevatori ad alta operatività". Sempre agli operatori e ai Rilevatori ad alta operatività delle regioni ad Alta partecipazione è stato riservato l'accesso al modulo di validazione interattiva. Gli operatori dell'Istat hanno avuto accesso a SGR e al sistema di validazione interattiva con utenze di sola lettura (per esigenze di monitoraggio) e supervisione.

¹ Autori: G. Sindoni §§ 1.1, 1.7; M. Venturi § 1.2; F. Liroy §§ 1.3, 1.4; M. Amato § 1.5; F. Cuccia § 1.5.1; D. Ingretolli § 1.5.2; B. Palazzi § 1.6.1, 1.6.2, 1.6.3; D. Frongia § 1.6.4.



1.2 L'architettura logico/fisica

L'architettura logica del sistema si è basata su tre livelli:

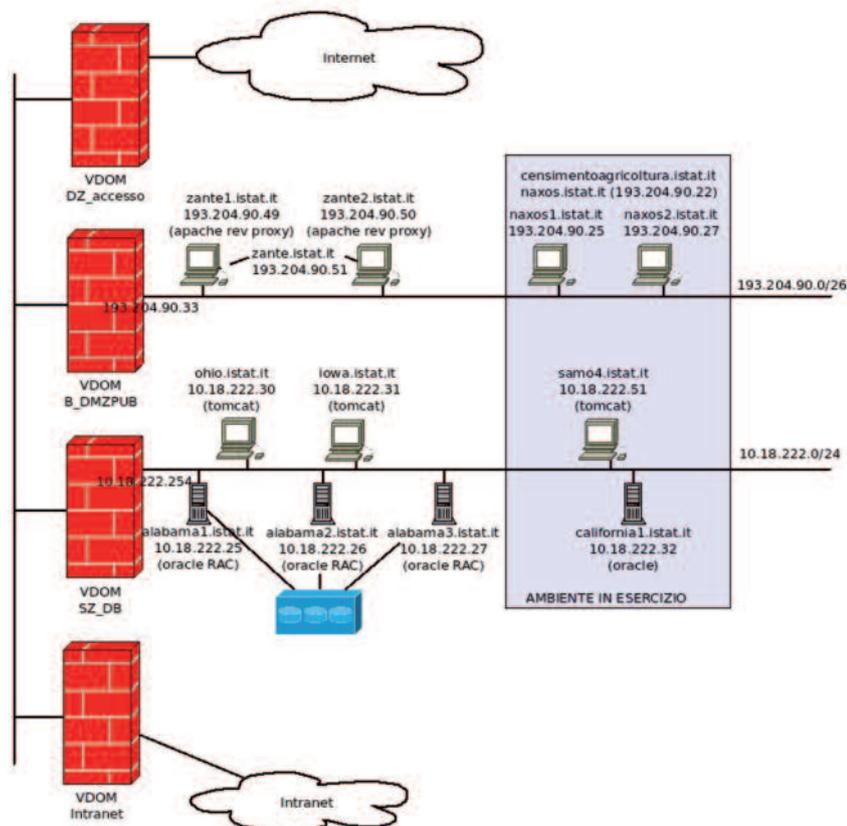
- **FRONTEND LAYER**, batteria di server basati su Sistema Operativo (S.O.) *Redhat Linux*, opportunamente dimensionati per bilanciare gli accessi ai *Web Server* tramite l'utilizzo di *Apache* come *reverse proxy*;
- **APPLICATION LAYER**, batteria di server basati su S.O. *Redhat Linux*, opportunamente dimensionati per sostenere gli accessi alle componenti dinamiche e applicative secondo i requisiti stimati dall'Istituto. Tra gli applicativi *middleware* ospitati è da menzionare *Apache Tomcat*;
- **DATABASE LAYER**, *cluster* di server *Oracle RAC* dimensionati secondo i requisiti dell'Istituto. I server di *backend* ospitavano i seguenti *middleware*:
 - *Oracle Database*;
 - *MySQL Database*;
 - *OpenLDAP*.

I server applicativi ospitavano l'*application server Tomcat* e il CMS *TYPO3*, mentre i server di *frontend* ospitavano il *Web Server Apache*.

L'architettura tecnologica è riportata nella Figura 1.2.

Figura 1.2 - Architettura tecnologica

CENSIMENTO AGRICOLTURA



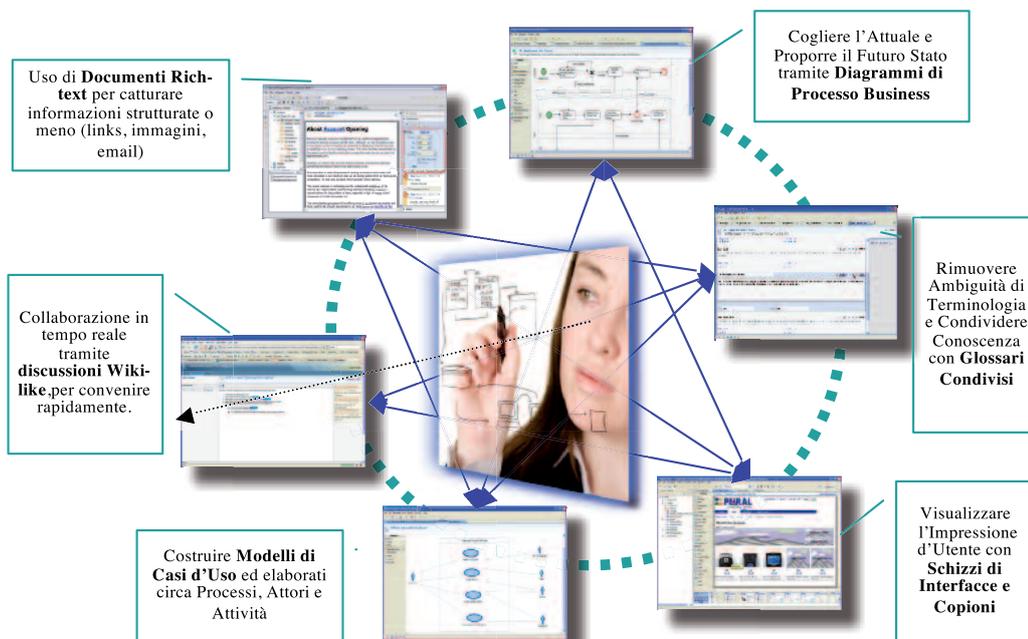
1.3 La definizione dei requisiti

Per la definizione dei requisiti è stato adottato l'ambiente *Rational Requirements Composer*TM di IBM. Esso ha consentito ai soggetti coinvolti di definire e usare i requisiti in modo efficace per tutta la durata del progetto.

Con *Rational Requirements Composer* la documentazione, costituita da documenti, prototipi di interfaccia grafica, casi d'uso, è stata memorizzata in un *repository* condiviso tra tutti gli attori coinvolti: statistici, progettisti, analisti e sviluppatori, Fondamentale è stata la possibilità di collaborare in ambiente web, venendo meno l'esigenza di scambi di documenti via mail.

Nella Figura 1.3 sono indicate le funzionalità dello strumento adottato.

Figura 1.3 - Funzionalità di IBM Rational Requirements ComposerTM



1.4 La modellazione dei requisiti funzionali

L'adozione di diagrammi dei casi d'uso ha rappresentato un modo estremamente utile per comunicare con gli specialisti statistici delle tematiche del settore agricolo circa le modalità di funzionamento del sistema in sviluppo.

Un caso d'uso può essere associato a più requisiti funzionali; un requisito funzionale può dare origine a più casi d'uso.

L'efficacia dei casi d'uso nell'individuazione dei requisiti è riposta nel dialogo che si instaura tra il committente del sistema e chi lo progetta. Vengono infatti analizzati, per ogni caso d'uso, gli scenari concreti di operatività degli utilizzatori nei confronti del sistema. La descrizione del sistema nella prospettiva dell'utente ha certamente consentito di centrare le esigenze, riducendo di conseguenza il tempo di analisi dei requisiti.

I diagrammi sono stati "tradotti" in casi d'uso verbosi (specifiche di caso d'uso), con la descrizione testuale dello scenario, degli attori coinvolti, delle azioni e dell'ordine temporale della loro sequenza, nonché del trattamento degli errori.



In questo modo i casi d'uso sono divenuti le specifiche tecniche di dettaglio per il programmatore, creando un'efficiente continuità tra chi pone i requisiti e il risultato dello sviluppo finale, oltre a divenire un riferimento principale nell'esecuzione dei test da effettuare. In Figura 1.4 è riportato un esempio di "caso d'uso verboso" (o specifica di caso d'uso), relativo all'anagrafica dell'azienda agricola.

Figura 1.4 - Esempio di "caso d'uso verboso" (o specifica di caso d'uso)

Attori coinvolti	Utente (azienda agricola o utente abilitato) Sistema
Sommario	Il caso d'uso descrive come l'utente gestisce i dati relativi al Promemoria della sezione Anagrafica azienda del questionario
Precondizioni	L'utente è stato riconosciuto dal sistema e ha selezionato dal menù la funzione "Promemoria " dalla sezione Anagrafica azienda
Tabelle DB	QUESTIONARIO, STATO_COMPILAZIONE
<i>Percorso base eventi</i>	<p>IL SISTEMA</p> <p>1. Preleva da QUESTIONARIO le informazioni riguardanti il Promemoria e visualizza i pulsanti:</p> <ul style="list-style-type: none"> • "Salva " • "Salva in bozza" (non abilitato) <p>L'UTENTE</p> <p>1. Inserisce i dati nella pagina selezionando i <i>check-box</i> e riempiendo il campo note</p> <p>2. Seleziona il pulsante "salva"</p> <p>IL SISTEMA</p> <p>3. Conferma il corretto inserimento dei dati</p> <p>IL SISTEMA</p> <p>4. Aggiorna la pagina di visualizzazione e salva i dati</p>
Postcondizioni	Il sistema aggiorna i valori nelle tabelle QUESTIONARIO e STATO_COMPILAZIONE con flag_bozza=0 dati completi
<i>Percorso alternativo 1</i>	<p>Dal punto 2 del percorso base</p> <p>IL SISTEMA</p> <p>1. Avvisa l'utente della presenza di dati incompleti o errati e chiede all'utente di correggere i campi segnalati o se vuole salvare in bozza la pagina</p> <p>2. Abilita il pulsante "Salva in bozza"</p> <p>L'UTENTE</p> <p>Seleziona il pulsante "Salva in bozza".</p>
Postcondizioni	Il sistema aggiorna i valori nelle tabelle QUESTIONARIO e STATO_COMPILAZIONE con flag_bozza=1 dati in bozza
Requisiti di dominio	
Note	

1.5 La progettazione del software

Le innovazioni introdotte e la complessità dell'architettura generale hanno implicato la realizzazione di alcune funzioni aggiuntive rispetto a un normale sistema web: sincronizzazione e scambio dati con sistemi esterni; sincronizzazione col sistema di raccolta questionari via internet; maggiore complessità delle funzioni di monitoraggio dovuta alla diversificazione dei questionari e alla multi-canalità della loro restituzione; accessibilità in scrittura da parte di numerose categorie di utenti; gestione esplicita delle informazioni relative allo stato di ciascun questionario; utilizzo delle informazioni per guidare l'attività quotidiana e sistematica dei rilevatori.

In base a queste considerazioni, il sistema di monitoraggio, inteso come sistema distribuito di comunicazione telematica tra i vari attori e di gestione delle operazioni sul campo, diventava elemento portante della macchina censuaria e doveva offrire alla rete di rilevazione tutti i servizi necessari, quali funzioni di gestione e aggiornamento

di facile accesso e utilizzo, nonché prospetti riassuntivi aggiornati “in tempo reale”, in base ai flussi informativi in entrata.

Ciascun sottosistema ha avuto un proprio sito di riferimento, accessibile mediante una URL configurata su un apposito *Web Server*. Era possibile utilizzare diversi tipi e versioni di *browser* per avere accesso alle applicazioni, rispettando solo requisiti minimi di risoluzione e impostazione locale sul *personal computer* dell'utilizzatore. I siti si sono avvalsi del protocollo sicuro HTTPS e sono stati sviluppati utilizzando la tecnologia Java e gli strumenti standard disponibili in Istat, mentre il portale della rete si è avvalso del CMS TYPO3.

Il colloquio tra i sottosistemi è avvenuto a livello dei dati in modo unidirezionale, ossia il sottosistema di acquisizione ha sempre alimentato quello di gestione trasferendo in esso una copia dei dati inseriti in modo definitivo dal rispondente. Questa scelta architettonica ha presentato i seguenti vantaggi:

- indipendenza dei sottosistemi;
- controllo separato delle prestazioni e *tuning* differenziato;
- maggiore tutela della riservatezza dei dati.

Come RDBMS per la gestione del database è stato utilizzato *Oracle* che ha rappresentato il software principale per la gestione dei dati grazie all'alta affidabilità ed alla sua estrema scalabilità.

Per consentire l'aggiornamento dinamico di una pagina web senza esplicito ricaricamento da parte dell'utente, molte funzioni, in particolar modo quelle relative alla territorialità, sono state implementate utilizzando AJAX, acronimo di *Asynchronous JavaScript and XML*, una tecnica di sviluppo per la realizzazione di applicazioni web interattive (*Rich Internet Application*). AJAX è asincrono nel senso che i dati extra sono richiesti al server e caricati in *background* senza interferire con il comportamento della pagina esistente. Normalmente le funzioni richiamate sono scritte con il linguaggio *JavaScript*.

1.5.1 L'architettura software del questionario e del sistema di gestione della rilevazione

Il sistema software si è basato su un'applicazione web sviluppata secondo le specifiche della piattaforma *Java Enterprise Edition* (JEE).

Il sistema presenta un'architettura distribuita composta di tre livelli logico-funzionali tra loro indipendenti, ciascuno dei quali implementa un insieme di funzioni messe a disposizione degli altri livelli, isolandone i dettagli implementativi.

Distinguiamo quindi:

- un livello di *Presentazione* che fornisce l'interfaccia del sistema verso il mondo esterno e si occupa di acquisire dati e visualizzare risultati;
- un livello di *Logica di Business* che contiene la logica applicativa alla base del funzionamento del sistema;
- un livello di *Persistenza* che fornisce le funzioni necessarie per l'interazione con una base di dati.

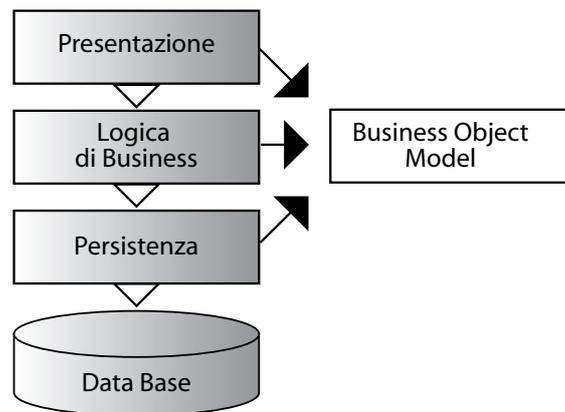
Nella Figura 1.5 sono mostrati i tre livelli logico-funzionali con l'aggiunta di un ulteriore livello, il *Business Object Model*, che contiene la rappresentazione orientata agli oggetti del dominio applicativo ed è condiviso da tutti gli altri livelli a fini dello scambio di dati. Questi tre livelli logico-funzionali sono distribuiti fisicamente su altrettanti livelli hardware:

- il *Client*, che corrisponde all'insieme dei computer dei singoli utenti che interagiscono col sistema tramite un *browser*. L'interazione consiste nell'invio di richieste e nella ricezione di risposte attraverso il protocollo HTTP;



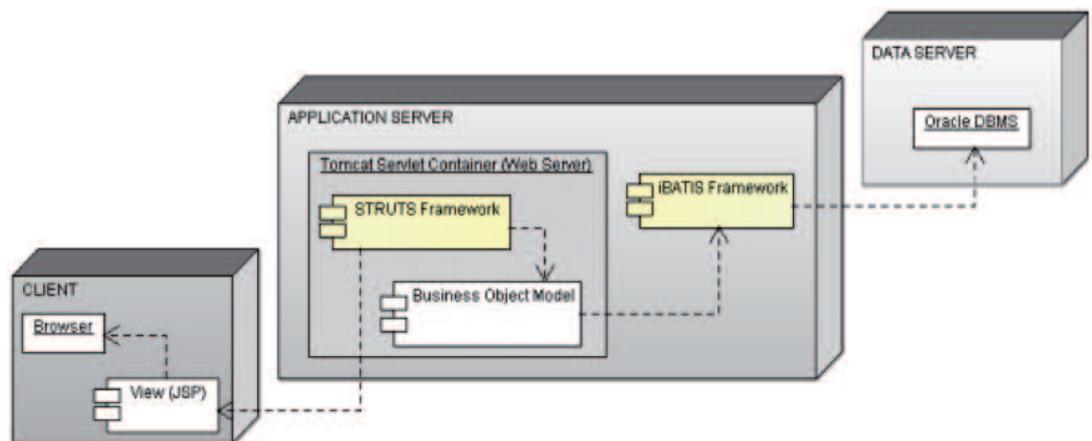
- l'*Application Server*, che rappresenta la parte centrale del sistema in quanto fornisce l'infrastruttura e le funzionalità di supporto necessarie per l'esecuzione dell'applicazione web. Qui risiede il *Web Server* che agisce da intermediario per le richieste e le risposte HTTP;
- il *Data Server*, che fornisce l'infrastruttura e le funzionalità di supporto necessarie per l'accesso a una base dati relazionale tramite i protocolli di rete.

Figura 1.5 - Livelli logico-funzionali del sistema software



Nella Figura 1.6 è schematizzata la distribuzione delle componenti software del sistema sui tre livelli fisici. In particolare sono evidenziati i due *framework*, *Struts* e *iBATIS*, scelti come supporto allo sviluppo. Tutte queste componenti sono descritte nella restante parte del paragrafo.

Figura 1.6 - Livelli fisici e distribuzione delle componenti software del sistema



Sia il questionario on-line che il sistema di gestione della rilevazione sono costituiti da un'applicazione web realizzata sulla base di due tecnologie caratteristiche della piattaforma JEE: le *Servlet Java* e le *Java Server Pages (JSP)*.

Una *Servlet* è una classe Java il cui compito è:

- gestire le richieste HTTP;
- effettuare eventuali elaborazioni;
- fornire opportune risposte al client sotto forma di *JSP*.

Le *JSP* sono pagine di testo contenenti tipici *tag* HTML per la definizione dei contenuti statici e speciali *tag* con cui possono essere invocate funzioni predefinite o codice Java per la generazione di contenuti dinamici.

La combinazione di queste due tecnologie consente la separazione del processo di disegno e realizzazione delle pagine HTML dallo sviluppo della logica applicativa permettendo, tra l'altro, di distribuire meglio le responsabilità di sviluppo secondo le competenze.

Le *Servlet* sono localizzate nel *Web Server* e per essere eseguite necessitano di un *Servlet container* che le ospiti. Si tratta di un componente del *Web Server* che fornisce un ambiente di esecuzione alle *Servlet* gestendone il ciclo di vita. In particolare, per il sistema realizzato è stato adottato come *Servlet container* *Apache Tomcat* che è una implementazione Java del protocollo HTTP. *Tomcat* di fatto ha le stesse funzioni di un *Web Server* tradizionale che include anche una funzione di *Servlet container* e rappresenta uno standard per le applicazioni Web su piattaforma JEE.

Nella realizzazione dell'applicazione Web per il questionario on-line sono stati tenuti presenti tre obiettivi essenziali:

- massimizzare la possibilità di riutilizzo del codice;
- facilitare il lavoro del team di sviluppo suddividendo il progetto in componenti indipendenti da assegnare a gruppi di lavoro differenti;
- minimizzare i costi di manutenzione del codice e di adattamento a nuove esigenze inizialmente non previste.

Tali obiettivi hanno reso naturale la scelta di *Model-View-Controller (MVC)*, un modello di architettura software molto diffuso nella realizzazione di applicazioni web complesse perché consente una maggiore strutturazione del codice attraverso una chiara separazione tra la logica di business, la presentazione dell'output e l'elaborazione delle richieste del client.

Come suggerito dal suo nome, questo modello prevede tre componenti fondamentali:

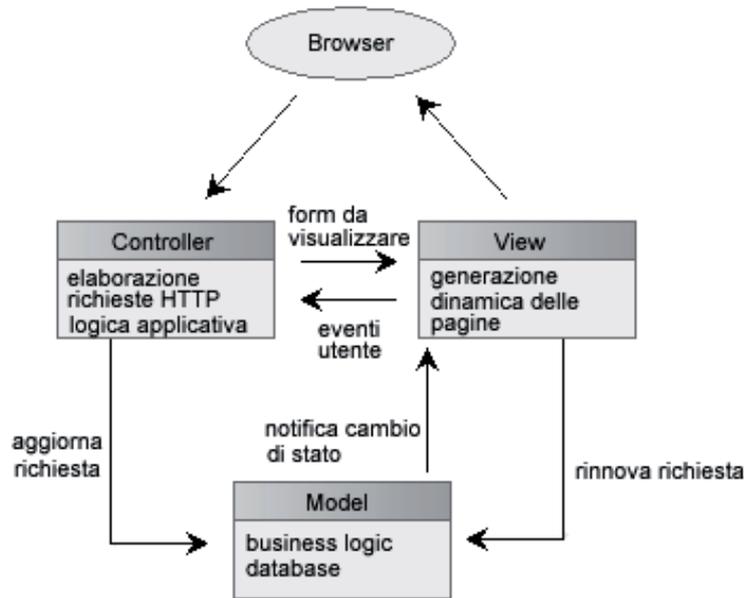
- il Modello (*Model*) che fornisce una rappresentazione dei dati ed è responsabile della gestione degli stessi. Nell'ambito dello *sviluppo object oriented* su piattaforma Java è costituito da oggetti Java popolati con dati tipicamente persistenti su un database;
- la Vista (*View*) che gestisce la presentazione dei dati. In ambito Java è costituita da pagine HTML e JSP;
- il Controllore (*Controller*) che gestisce il flusso di dati tra Vista e Modello intercettando le richieste HTTP, traducendo ogni singola richiesta in una operazione per il Modello e selezionando la prossima Vista da mostrare. La sua implementazione Java corrisponde a una *Servlet*.

Nella Figura 1.7 è rappresentata l'architettura tipica di un'applicazione web basata sul modello MVC.

Nel caso specifico è stata adottata l'implementazione del modello MVC fornita dal *framework* Struts. Questo è basato sulla tecnologia delle *Servlet Java* e delle *Java Server Pages* e fornisce un insieme di classi e interfacce che compongono l'infra-



Figura 1.7 - Architettura web basata sul modello MVC



struttura di base per costruire applicazioni web su piattaforma JEE conformi alle specifiche del modello MVC. Esso si integra nel *Web Server* e necessita di un *Servlet container* per poter funzionare. Struts si fa carico di gestire tutte le richieste *client* e di smistare il flusso applicativo in base alle regole configurate, consentendo agli sviluppatori di concentrarsi soprattutto sulla costruzione della logica di business dell'applicazione. Tutta la configurazione dell'applicazione è contenuta all'interno di uno specifico file XML in cui sono definite le modalità di interazione tra le varie componenti del sistema.

L'utilizzo di Struts ha dato vantaggi significativi alla realizzazione del progetto:

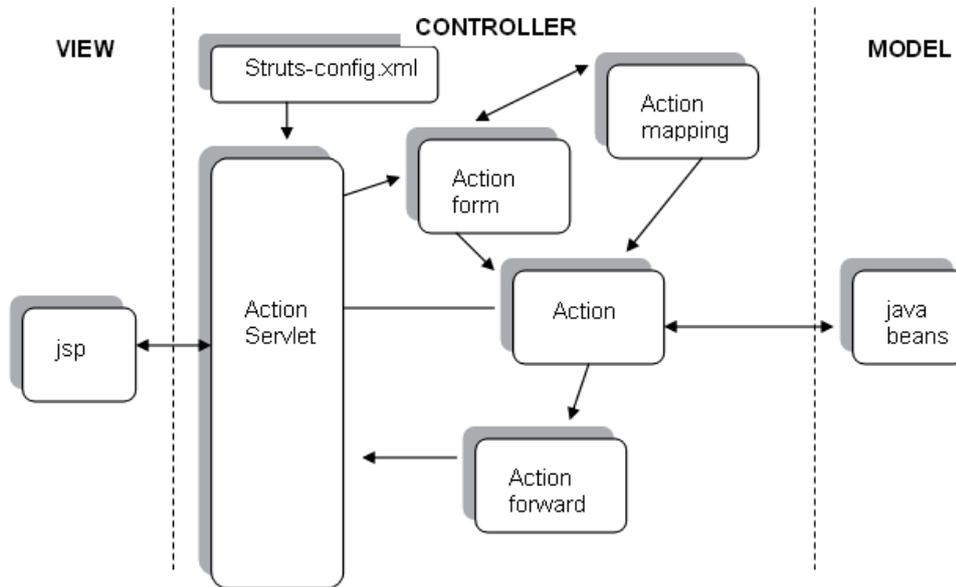
- modularità e riusabilità, in quanto i diversi ruoli dell'applicazione sono affidati a diverse componenti, cosicché è possibile sviluppare un codice modulare più facilmente riutilizzabile;
- facilità di manutenzione, in quanto l'applicazione è costituita da livelli logici ben distinti, cosicché una modifica in uno dei livelli non comporta modifiche negli altri;
- rapidità di sviluppo, in quanto è possibile sviluppare in parallelo le varie parti dell'applicazione, logica di *business* e di *view*.

Nella Figura 1.8 è descritta l'implementazione di Struts del modello MVC.

I componenti fondamentali di Struts sono:

- l'*ActionServlet*, *servlet* che funziona da *Controller* del *pattern* MVC. Gestisce il flusso di tutte le richieste dell'applicazione delegando alle singole istanze della classe *RequestProcessor* il controllo delle specifiche richieste/risposte;
- il *RequestProcessor*, classe che gestisce le singole richieste/risposte invocando le opportune classi *Action* in base a quanto definito all'interno del file *struts-config.xml*;
- il file *struts-config.xml*, file di configurazione di tutta l'applicazione. In questo file vengono definiti gli elementi dell'applicazione e le loro associazioni determinando l'azione da intraprendere durante la richiesta di una risorsa (*action mapping*);
- l'*Action*, classe a cui è delegata l'elaborazione della richiesta. Riceve i dati prove-

Figura 1.8 - Implementazione Struts del modello MVC



nienti dalla richiesta utente, interagisce col *Model* e seleziona le opportune JSP della *View* per la presentazione dei risultati ottenuti;

- l'*ActionMapping*, contiene il riferimento alla *Action* - che il controller deve richiamare in base ad una determinata richiesta - e tutte le informazioni necessarie a soddisfare una richiesta.
- le *ActionForm*, ossia classi contenitori di dati. Vengono popolate automaticamente dal *framework* con i dati contenuti nelle *request* HTTP che vengono passati a una *Action*;
- l'*ActionForward*, classe che contiene l'informazione relativa al nodo successivo del flusso di esecuzione, che il controller deve attivare una volta che la *Action* ha terminato l'elaborazione, in base alla logica dell'applicazione.

Il **flusso elaborativo** nella logica di Struts può essere così schematizzato:

1. in fase di *startup* tutti i dati di configurazione vengono letti dal file XML di configurazione *struts-config.xml*;
2. il client invia una richiesta HTTP;
3. la richiesta viene ricevuta dall'*ActionServlet* di Struts che popola l'*ActionForm* associato alla richiesta con i dati della request e l'*ActionMapping* con gli oggetti relativi all'*Action* associata alla richiesta;
4. l'*ActionServlet* delega l'elaborazione della richiesta alla *Action* associata al path della richiesta passandole in input *request* e *response* HTTP, l'*ActionForm* e l'*ActionMapping* precedentemente valorizzati;
5. l'*Action* si interfaccia con lo strato di business che implementa la logica applicativa. Al termine dell'elaborazione restituisce alla *ActionServlet* un *ActionForward* contenente l'informazione del *path* della vista da fornire all'utente;
6. l'*ActionServlet* esegue il *forward* alla vista specificata nell'*ActionForward*.

Per la gestione della persistenza dei dati è stato adottato il *framework* iBATIS, costruito su SQL e Java. Esso si inserisce al livello di "persistenza" gestendo l'interfacciamento



con basi di dati di tipo relazionale. La sua funzione è di fornire una integrazione tra software orientato agli oggetti e database relazionali, consentendo agli sviluppatori di scrivere meno codice rispetto all'uso diretto della tecnologia *Java DataBase Connectivity* (JDBC) e di un driver JDBC.

iBATIS si fa carico di:

- stabilire una connessione con la base dati;
- settare i parametri;
- eseguire l'istruzione;
- prelevare i risultati;
- rilasciare tutte le risorse;

tutte funzioni realizzate con una quantità di codice da scrivere significativamente ridotto, più conciso e quindi più semplice da mantenere. Infatti l'integrazione tra oggetti e database relazionale avviene in maniera semplice attraverso opportune istruzioni SQL inserite all'interno di un file di configurazione XML. iBATIS più che un data mapper si può definire un *SQL mapper*, strumento che consente di trasferire dati tra oggetti e database relazionali conservando l'indipendenza tra i due mondi. Infatti, a differenza degli *object/relational mapper* (come ad esempio *Hibernate*), iBATIS non lega le classi a tabelle o colonne di tabelle, ma mappa l'input e l'output di una istruzione SQL in una classe, inserendosi come uno strato intermedio che isola il disegno della base dati dal modello a oggetti. Questo consente maggiore flessibilità nel modo in cui classi e tabelle sono mappate, senza richiedere particolari vincoli nel disegno del modello dei dati o del modello a oggetti.

Ogni istruzione SQL è vista come un insieme di input e output: l'input è dato dai parametri che si trovano tipicamente nella clausola WHERE dell'istruzione SQL; l'output è dato dalle colonne che si trovano nella clausola SELECT. Il vantaggio di questo approccio è che allo sviluppatore è dato uno strumento molto flessibile con il quale può facilmente manipolare i dati per adattarli al modello a oggetti, senza dover modificare il sottostante disegno della tabella. Inoltre, è possibile facilmente introdurre tabelle multiple o risultati provenienti da funzioni o "*stored procedure*" predefinite all'interno del database. È possibile così sfruttare direttamente la piena potenzialità dell' SQL: iBATIS mappa l'input e l'output usando le indicazioni contenute in un semplice file XML.

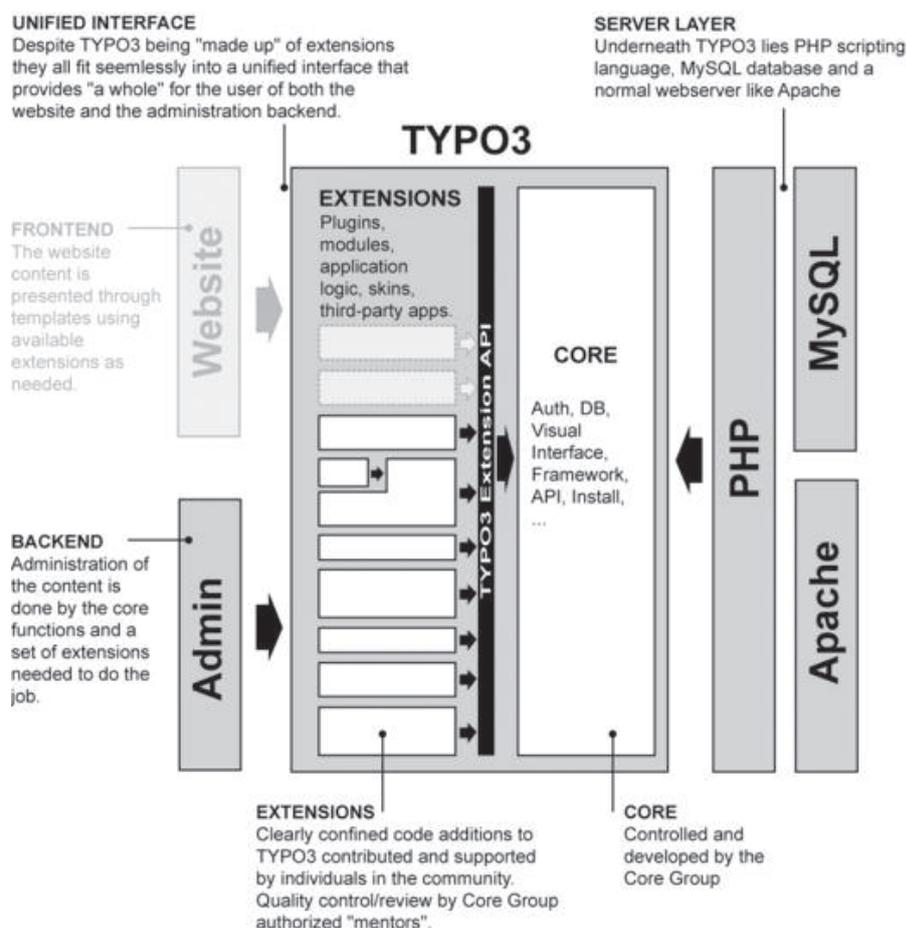
iBATIS garantisce il mantenimento di una architettura software stratificata gestendo tutti gli elementi dello strato relativo alla persistenza dei dati (connessione al db, query e result set) e mantenendo le istruzioni SQL separate dal codice sorgente dell'applicazione.

Ciò consente agli sviluppatori della logica applicativa di lavorare sempre con oggetti senza dover mai interagire direttamente con la base dati relazionale e di delegare eventualmente la realizzazione dell'interfaccia verso la base dati a chi ha maggiore competenza in ambito SQL.

1.5.2 L'architettura software del portale della rete

Il Portale della rete è stato realizzato utilizzando un meta ambiente, denominato TYPO3, un CMS (*Content Management System*) per la costruzione e gestione di siti web di qualsiasi tipo, particolarmente indicato per portali. TYPO3 è scritto completamente in PHP mentre tutti i contenuti e le configurazioni di un sito sono gestiti in un database MySQL. MySQL è un sistema di gestione di database relazionali (RDBMS - *Relational Database Management System*), composto da un client - con interfaccia a riga di comando - e un server, entrambi disponibili sia per sistemi Unix sia per Windows.

Figura 1.9 - Architettura di sistema dell'ambiente TYPO3



Nella Figura 1.9 è rappresentata l'architettura di sistema di TYPO3. TYPO3 è un ambiente Open Source, licenziato come GNU *General Public License* (GNU GPL), che assicura all'utente libertà di utilizzo, copia, modifica e distribuzione. Tale ambiente è da anni il CMS di riferimento all'interno dell'Istat. L'installazione standard permette di gestire le funzionalità di base di un sito quali ad esempio l'inserimento di testi o di link a dei documenti. Per funzionalità più complesse è possibile ricorrere a dei plug-in (o estensioni) che permettono di integrare molte altre funzioni senza necessità di nuove programmazioni in PHP. Molte estensioni possono essere scaricate collegandosi al *repository* di TYPO3 ma possono anche essere create in autonomia, in funzione delle proprie specifiche necessità. Nella presente realizzazione sono state aggiunte alcune funzioni standard, disponibili nel *repository* di TYPO3, quali quella per la gestione delle "Notizie in evidenza" (*News*) e per la gestione delle "Risposte a domande frequenti" (FAQ - *Frequent Asked Questions*). Per la gestione degli accessi è stata utilizzata un' estensione per la gestione degli utenti LDAP (*Lightweight Directory Access Protocol*), in modo da importare le utenze e password dal server Istat nella tabella degli utenti di TYPO3. Il personale interno Istat è infatti riconosciuto in generale sui sistemi informativi dell'Istituto tramite il sistema LDAP, protocollo standard che permette di gestire degli elenchi di utenze, ovvero di accedere alle informazioni degli utenti di una rete attraverso dei protocolli TCP/IP.



Un'altra caratteristica di TYPO3 risultata utile è stata quella di poter gestire l'accesso ai contenuti del sito in funzione dell'appartenenza di un utente a uno specifico gruppo.

1.6 Test delle applicazioni

Prima di mettere in produzione gli applicativi del 6° Censimento generale dell'agricoltura sono state svolte attività esplorative di test informatici.

Le attività di test sono state inizialmente pianificate prevedendo tre direzioni principali: i) test funzionali; ii) test di sicurezza; e iii) test di carico. Tuttavia, mentre i test funzionali sono stati condotti regolarmente, i test sulla sicurezza, sulla robustezza e tenuta dell'applicazione in presenza di un elevato numero di utenti contemporanei sono stati condotti parzialmente in quanto i tempi a disposizione sono stati ristretti e non hanno consentito l'applicazione completa di tutte le fasi programmate. Anche i test di carico sono stati eseguiti solo parzialmente e senza poter simulare l'ambiente reale di esercizio. Pertanto non è stato possibile verificare eventuali comportamenti anomali degli applicativi nei momenti di picco.

Le attività sono state svolte principalmente attraverso le professionalità e le strutture interne all'Istituto.

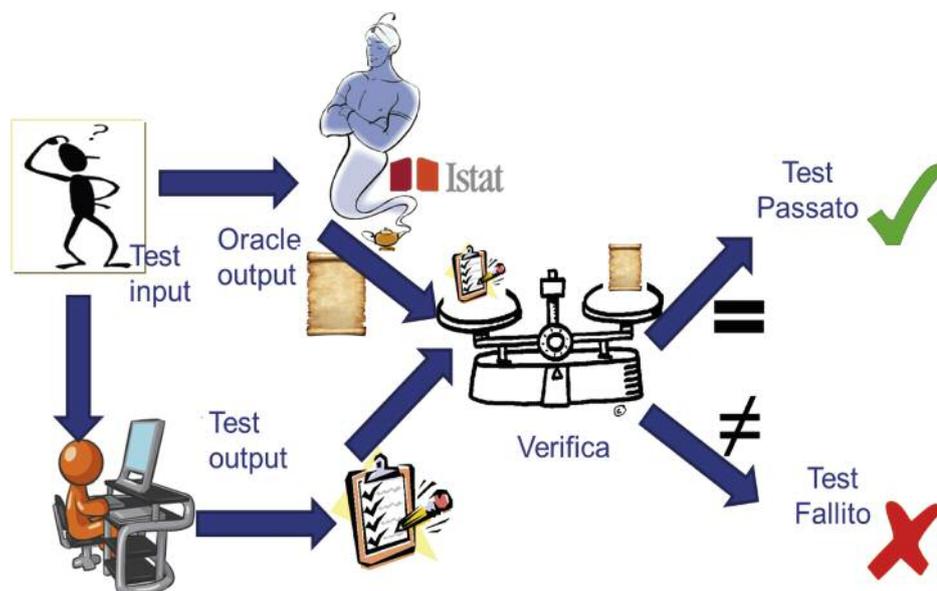
1.6.1 Test funzionali

Le attività di test funzionali hanno previsto una verifica puntuale della rispondenza tra:

- i requisiti richiesti dagli *stakeholder* e descritti nei documenti di progettazione degli applicativi a supporto del censimento;
- le funzioni disponibili a valle del processo dello sviluppo informatico da parte dei programmatori.

L'attività di test funzionali si è basata sul modello dell'oracolo, come indicato nella Figura 1.10. Per ogni test di input un *tester* verifica la compatibilità tra l'*Oracle output*, cioè quello previsto nei documenti di specifica funzionale, e *Test output*, cioè il risultato

Figura 1.10 - Modello dell'oracolo per i test funzionali



ottenuto dal programma sviluppato. Se il risultato del test è difforme, il test è considerato fallito e si comunicano agli sviluppatori le azioni correttive.

L'attività di test funzionale è stata suddivisa in tre fasi:

1. *alfa-testing*, verifica delle funzioni appena rilasciate dal gruppo di sviluppo ed effettuato dagli informatici che compongono il gruppo di test. Tutte le attività sono state svolte poche settimane dopo l'inizio dello sviluppo; sono state ripetute per ogni *milestone* e nuovo rilascio (*deploy*) in produzione;
2. *beta-testing* interno, verifica funzionale da parte degli *stakeholder* statistici appartenenti al servizio DICA/SCE e alla rete territoriale Istat, attività a valle dell'*alfa-testing* e che ha permesso una verifica più approfondita della semantica delle funzioni, rispetto all'*alfa-testing*;
3. *beta-testing* esterno, verifica funzionale effettuata dagli utenti finali, in particolare dai responsabili degli Uffici Regionali di Censimento (URC) (*early adopters*) che ha permesso di verificare eventuali problemi di *User eXperience* dei prodotti sviluppati.

Questa fase ha previsto una segnalazione dei *bug* o di possibili miglioramenti tramite *email* indirizzate al gruppo di test da parte della rete territoriale dell'Istituto.

1.6.2 Test di sicurezza

La metodologia predisposta per testare la sicurezza delle applicazioni prevedeva le seguenti attività ed analisi:

- a. **Code Review**, che include l'attività di analisi del codice (statica e dinamica), in ambiente di test con modalità *white e black box*. L'analisi statica consiste nell'analisi del codice sorgente, senza necessità di compilare o eseguire il codice sorgente, volta a identificare possibili migliorie avvalendosi delle *best practice* note in letteratura. L'analisi dinamica, detta anche a *runtime*, del codice installato e configurato, è volta a identificare problemi relativi all'esecuzione del programma;
- b. **Vulnerability Assessment (VA)**, ossia l'analisi tramite strumenti, automatici e non, della presenza di vulnerabilità note in letteratura all'interno del sistema da certificare;
- c. **Penetration Test (PT)**, la verifica a *runtime*, su un apposito sistema di test, delle criticità individuate, utilizzabili per un attacco al sistema da certificare.

Per i motivi cui si è accennato in precedenza, queste attività, sono state svolte solo parzialmente prima dell'avvio del censimento. Si è reso invece necessario effettuarle durante la fase iniziale di esercizio del sistema di acquisizione on-line tramite il supporto di una società di consulenza esterna. L'esperienza svolta in questa attività è stata utile per acquisire all'interno dell'Istituto tecniche di programmazione che si sono rivelate di grande utilità nei successivi censimenti.

1.6.3 Test di carico

Per test di carico (*stress test*) si intende una serie di attività idonee ad individuare il limite del sistema. Durante uno *stress test* il carico utente viene simulato e incrementato, finché il sistema non diventa indisponibile, oppure i parametri di performance diventano inaccettabili.

Anche in questo caso i test di carico sul sistema da certificare sono stati condotti in modo parziale prima della fase di esercizio. A causa di comportamenti anomali degli applicativi, verificatisi il primo giorno di apertura del questionario on-line alla compilazione da parte delle aziende agricole, in corrispondenza dell'elevato numero di accessi contemporanei, è stato necessario sospendere la fase di esercizio e sottoporre i sistemi ad opportuni e più approfonditi test di carico.



1.6.4 Gestione operativa dei test funzionali

La gestione dei test funzionali è stata attuata distinguendo le attività svolte su SGR nel suo complesso e sul questionario elettronico.

Per la realizzazione dei test di funzionalità del sistema di gestione della rilevazione è stato predisposto un ambiente dedicato del tutto identico all'ambiente di esercizio e direttamente raggiungibile su internet all'indirizzo

<https://censimentoagricoltura.istat.it/sgrtest>.

Il database di riferimento è stato pre-caricato utilizzando esclusivamente dati di prova, non corrispondenti a situazioni reali.

Individuati i tester sulla base di adesioni volontarie, è stato predisposto un calendario di esecuzione durante il periodo aprile-settembre 2010 ed è stato definito un *template* per la documentazione degli esiti, riportato in Figura 1.11, mentre in Figura 1.12 sono riportati alcuni esempi. Ai tester è stato richiesto di documentare i soli esiti negativi, in cui, a fronte di una determinata richiesta, non si otteneva la risposta attesa.

Inoltre è stato predisposto un breve manuale di istruzioni.

Figura 1.11 - Template per la documentazione esiti test

Macrofunzione	Funzione	Eseguito da:	Nome Cognome (Profilo)	In data: ...
	Versione SGR Descrizione del test: Esito: Note aggiuntive:	v. 1.1.0		

Figura 1.12 - Documentazione esiti test - esempi

Operatori	Gestione URC - inserimento	Eseguito da:	Mario Rossi (URC)	In data: 23/06/2010
	Versione SGR Descrizione del test: Esito: Note aggiuntive:	v. 1.1.0 Inserimento di un URC nuovo. Negativo. Per ogni nuovo operatore inserito deve essere automaticamente inviata una mail all'operatore contenente login e password di accesso al sistema. L'email di comunicazione login e password al nuovo URC non è arrivata.		
Operatori	Gestione URC - inserimento	Eseguito da:	Bianca Verdi (CIC)	In data: 23/06/2010
	Versione SGR Descrizione del test: Esito: Note aggiuntive:	v. 1.1.0 Inserimento di un URC nuovo. Negativo. La funzione deve essere a disposizione del solo responsabile URC (RURC). Un CIC (coordinatore) NON può inserire un URC mentre il sistema lo consente. Questa funzione non dovrebbe essere visibile o cliccabile		

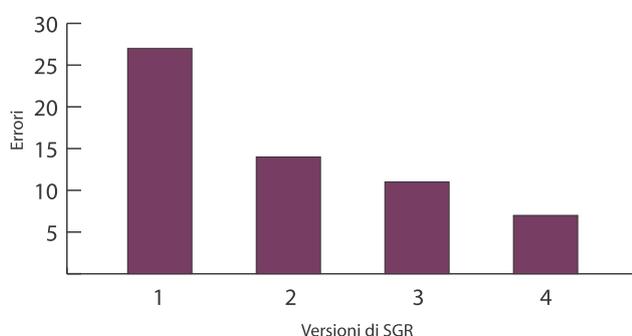
I RIT Veneto, Lombardia, Friuli-Venezia Giulia e Bolzano, hanno partecipato alla fase di test di un primo sottoinsieme di funzioni del sistema di gestione, denominata Beta test 1. I RIT Liguria, Veneto, Sicilia, Puglia, Sardegna, Friuli-Venezia Giulia, Basilicata, Marche, Campania e Lombardia hanno invece partecipato alla seconda fase di test sulla totalità delle funzioni del sistema, denominata Beta test 2.

Gli esiti sono stati analizzati singolarmente e classificati in "errori reali", "falsi errori", "chiarimenti".

Tutti i test sono stati ri-eseguiti sulle versioni successive di SGR.

La distribuzione degli errori reali, nelle 4 versioni di SGR precedenti la versione di esercizio priva di errori, è riportata nella Figura 1.13.

Figura 1.13 - Numero degli errori per rilascio di SGR



Per la realizzazione dei test del questionario elettronico sono state definite e consegnate ai tester una serie di utenze di prova per l'accesso al questionario. In questo caso è stata utilizzata direttamente l'applicazione in esercizio. La documentazione degli esiti è avvenuta attraverso la predisposizione di documenti word riportanti le pagine del questionario che presentavano degli errori.

1.7 Sicurezza dei dati

Il sistema ha messo in atto le misure di sicurezza necessarie a garantire:

- la **disponibilità dei dati**, che è stata garantita stabilendo livelli di disponibilità del servizio tali da conciliare le esigenze di accessibilità degli utenti con la minimizzazione del rischio di downtime e accesso illecito ai dati;
- l'**integrità dei dati**, che è stata protetta minimizzando il rischio di cancellazione illecita o modifica delle informazioni a seguito di errori e azioni da parte di attori non autorizzati, o perdita d'informazione dovuta a rotture del sistema, incendi, allagamenti, eccetera;
- la **riservatezza dei dati**, che è stata salvaguardata minimizzando il rischio che un attore non autorizzato (umano o software) potesse deliberatamente o accidentalmente accedervi, assicurandosi che tutti gli accessi fossero controllati da misure di protezione adeguate.

Queste misure si sono aggiunte a due caratteristiche strutturali del sistema: la verificabilità e la controllabilità delle transazioni sui dati, che permettevano l'identificazione univoca (attraverso file di log, di audit, eccetera) di chi utilizzasse il sistema e di quali operazioni venissero portate a termine.





Capitolo 2

La rilevazione²

Il 6° Censimento generale dell'agricoltura ha introdotto l'adozione di sistemi di interazione web con tutti gli attori coinvolti nella rilevazione: le aziende agricole, i Rilevatori, nonché gli addetti alla gestione della rilevazione.

I tre sottosistemi e precisamente:

- il Portale della Rete, rivolto agli operatori della rete di rilevazione;
- il Questionario elettronico, rivolto sia alle unità di rilevazione sia agli operatori della rete di rilevazione appositamente designati;
- il Sistema di gestione della rilevazione, rivolto a tutti gli operatori della rete di rilevazione addetti al censimento;

sono stati realizzati per cercare di soddisfare in modo esaustivo le varie esigenze dei diversi utenti.

Nel Paragrafo 2.1 viene descritto il **Portale della Rete**. Nel successivo Paragrafo 2.2 è descritto il **Sistema di acquisizione dei dati via web** o **Questionario elettronico**. Infine, nel Paragrafo 2.3 viene descritto il **Sistema di gestione della rilevazione**.

2.1 Il Portale della Rete

Il **Portale della Rete** censuaria sintetizza nel suo nome l'obiettivo e le funzioni che ha inteso soddisfare:

- “Portale”, perché è il luogo ove reperire le informazioni dedicate al 6° Censimento generale dell'agricoltura;
- “della Rete” censuaria, proprio perché ha consentito l'accesso controllato alle informazioni di supporto a tutti gli operatori della rete organizzativa censuaria, ovvero:
 - alle Sedi Territoriali Istat, nelle proprie specifiche funzioni, a cominciare dai Referenti Istat Territoriali (RIT);
 - agli Uffici di censimento, a tutti i livelli costituiti;
 - ai Responsabili nei vari profili;
 - ai Rilevatori;
 - alla Direzione centrale dei censimenti generali dell'Istat.

2.1.1 Il sistema di accesso

Il portale è stato reso accessibile a personale interno Istat e a personale esterno, appartenente agli Uffici territoriali di censimento opportunamente costituiti presso i Comuni e/o le Regioni.

Il personale interno Istat coinvolto nella rilevazione è stato riconosciuto tramite utenze di tipo LDAP, in modo di poter accedere al portale della rete con la stessa utenza e password utilizzata per l'accesso ai sistemi interni dell'Istituto. Le utenze esterne all'Istat sono state create all'interno del Sistema di gestione della rilevazione.

² Autori: F. Lioy §§ 2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5; D. Ingretolli § 2.1.1; R. Vignola §§ 2.1.2, 2.1.3; F. Cuccia § 2.2.1; M. Giacomo § 2.3.1; M. Bruno, G. Vaste, § 2.3.2; M. Amato §§ 2.3.3, 2.3.5; R. Duma, § 2.3.4; M. Silipo § 2.3.6; R. Rosati, M. Marrone, § 2.4.1; L. Corradi, P. Giorgetti, §§ 2.4.2, 2.4.4; R. Rosati, § 2.4.3.



2.1.2 Definizione delle diverse aree del sito e loro relazioni

Una volta eseguito l'accesso la pagina principale del portale, appariva strutturata in 3 parti fondamentali (Figura 2.1):

- un'area orizzontale, posta superiormente, ove erano presenti i pulsanti per navigare all'interno del portale, suddivisa in una barra di menù principale e una secondaria;
- un'area verticale per la navigazione di maggiore dettaglio;
- un'area di lavoro in cui sono state pubblicate le informazioni di interesse.

Figura 2.1 - Organizzazione pagina principale del portale



2.1.3 Barra dei menù

La barra del menù si è articolata su due livelli secondo i contenuti:

- una prima barra di comandi (**principale**) per tutto ciò che riguardava strettamente le operazioni censuarie, indicante, da sinistra a destra:
 - Home (pagina iniziale)
 - Organizzazione
 - Strumenti
 - Documenti
 - Formazione
 - Glossario
 - Domande&Risposte
- una seconda barra di comandi di utilità (**secondaria**), indicante, da destra a sinistra:
 - Cerca, con text-box ove inserire il termine da ricercare ed icona 🔍, per avviare la ricerca
 - Logout
 - Link utili
 - Contatti
 - Mappa

Infine, in alto a destra si trovava il link a **SGR**, per accedere al **Sistema di Gestione della Rilevazione**, sottosistema dedicato alla gestione delle operazioni censuarie.

Di seguito si descrivono nel dettaglio le due barre di menù menzionate.

Barra del menù principale

La selezione di Home rimanda alla pagina iniziale che veniva aggiornata in tempo reale con notizie riguardanti l'andamento della rilevazione, informazioni sull'aggiornamento e la creazione di strumenti a supporto dell'indagine, nonché pubblicazione di documenti ufficiali relativi all'indagine e di documenti utili allo svolgimento della stessa (Figura 2.2).

Figura 2.2 - Home Page



The screenshot shows the Istat.it website interface for the 6th General Agricultural Census 2010. At the top, the Istat.it logo is on the left, and the title '6° Censimento Generale dell'Agricoltura 2010' is centered, with the subtitle 'Il sito per la rete censuaria' below it. A horizontal navigation bar contains the following links: Home, Organizzazione, Strumenti, Documenti, Formazione, Glossario, Domande&Risposte, and SGR. Below this bar, there are additional links: Mappa, Contatti, Links, and Logout, followed by a search box. The main content area is divided into two columns. The left column contains two news items: one dated 23.03.2010 about e-commerce in food retail, and another dated 18.03.2010 about the start of operations in Friuli Venezia Giulia. The right column features a map of Italy with regional boundaries and labels for various regions.

Il pulsante **Organizzazione** consentiva l'accesso ai documenti che illustravano l'organizzazione della rete di raccolta dei dati per regione. Attraverso una navigazione cartografica veniva illustrata la modalità di partecipazione della regione ai lavori del censimento (Alta Partecipazione o Partecipazione Integrativa) e le informazioni sulle strutture interessate, con i relativi riferimenti.

Sotto la voce **Strumenti** si trovava tutto ciò che era utile ai Rilevatori per lo svolgimento dell'indagine, dal questionario, alle guide, alle tabelle per la decodifica, quali quelle relative alle coltivazioni o ai vitigni.

Il pulsante **Documenti** rimandava alla documentazione ufficiale prodotta ai fini della realizzazione del Censimento, dalla lettera informativa, alle circolari, al Regolamento, al Piano generale di censimento e agli altri documenti a valore giuridico.

Selezionando il pulsante **Formazione** si otteneva il collegamento alla piattaforma di e-learning (formazione interattiva on-line, realizzata in ambiente *open source Dokèos*). Tale piattaforma è stata integrata nel portale, oltre che graficamente, anche dal punto di vista dell'accreditamento, infatti l'utente che aveva già acceduto al portale non doveva riaccreditarci alla piattaforma di formazione.



Sotto la voce **Glossario**, diviso in 6 macro pagine/aree distinte per lettera iniziale dell'argomento, sono state riportate tutte le definizioni dei termini utili nel contesto del Censimento (Figura 2.3).

Figura 2.3 - Pagina del Glossario

Istat.it
6° Censimento Generale dell'Agricoltura 2010
il sito per la rete censuaria

Home Organizzazione **Strumenti** Documenti Formazione **Glossario** Domande&Risposte SGR
Mappa Contatti Links Logout

Sei in: Glossario | A

A	Agrumi Arancio, mandarino, clementina e i suoi ibridi, limone, altri agrumi (arancio amaro o melangolo, bergamotto, cedro, chinotto, kumquat, limetta, pompelmo).
B - C	Allevamenti I dati sulla consistenza degli allevamenti fanno riferimento a tutti i capi di bestiame che al 24 ottobre si trovano presso l'azienda, sia che si tratti di bestiame in dotazione dell'azienda stessa, sia che si tratti di bestiame affidato o da essa allevato. Sono inclusi i capi temporaneamente assenti per transumanza, pascolo, ecc., sono esclusi gli animali di passaggio (es: femmine presenti per la monta).
D - F	Allevamenti avicoli Polli da carne – compresi i giovani capi destinati alla produzione di carne. Galline da uova – Galline che hanno già cominciato a deporre uova e giovani capi destinati alla produzione di uova.
G - L	Allevamenti (altri) Allevamenti di anatre, lumache, lombrichi, rane, cervidi.
M - R	Artigianato Attività artigianali esercitate dal conduttore, dai membri della sua famiglia o da manodopera non familiare, a condizione che essi siano addetti anche ai lavori agricoli.
S - Z	

Il pulsante **Domande & Risposte** rimandava ad un elenco di soluzioni ai problemi più frequenti, elenco che è stato ampliato anche in fase di rilevazione in base alle richieste pervenute ai numeri verdi o alle caselle di posta elettronica a supporto della stessa (Figura 2.4).

Figura 2.4 - Pagina delle Domande & Risposte

Istat.it
6° Censimento Generale dell'Agricoltura 2010
il sito per la rete censuaria

Home Organizzazione Strumenti Documenti Formazione **Domande&Risposte** SGR
Mappa Contatti Links Logout

Sei in: Domande&Risposte

Privacy, tutela della riservatezza e obbligo di risposta	A chi ci si può rivolgere in caso di difficoltà? Categoria: Assistenza ai rispondenti
Informazioni sull'indagine: obiettivi e caratteristiche	Che cosa contiene il Foglio di Famiglia? Categoria: Informazioni sull'indagine: obiettivi e caratteristiche
Il questionario	Quando va restituito il questionario compilato? Categoria: Il questionario
Assistenza ai rispondenti	

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

Barra del menù secondario

Con il pulsante **Cerca** era possibile operare ricerche a testo libero all'interno del sito. **Logout** era il pulsante per chiudere la sessione di collegamento al sito. Con la selezione di **Link** si visualizzava la pagina contenente link ad altri siti, che potevano essere utili nel corso delle operazioni censuarie. La selezione di **Contatti** portava ad una navigazione cartografica per l'individuazione di informazioni utili a contattare i referenti del Censimento, per i vari livelli territoriali (Figura 2.5).

Figura 2.5 - Pagina Contatti



Istat.it
6° Censimento Generale dell'Agricoltura 2010
il sito per la rete censuaria

Home Organizzazione Strumenti Documenti Formazione Glossario Domande&Risposte **SGR**
Mappa Contatti Links Logout

Sei in: Contatti | Piemonte

Regione Piemonte
Direzione Regionale dell'Agricoltura
Settore programmazione in materia di agricoltura e di sviluppo rurale
Dott. Mario Perosino
c.so Stati Uniti, 21
10128 Torino
tel: +39 011 4324 369
fax: +39 011 4324 339
email: psr@regione.piemonte.it

Ufficio Regionale ISTAT
Dott. Claudio Moriani (*)
Via A. Volta, 3
10121 Torino
tel: +39 011 5166 711
fax: +39 011 535 800
email: urto@istat.it

(*)indicare riferimento RIT

La selezione di **Mappa** visualizzava la struttura del sito, fornendo una visione rapida e immediata di tutti i contenuti.

2.2 Il Sistema di acquisizione dei dati via web

2.2.1 Il sistema di accesso

L'accesso vero e proprio al questionario on-line dipendeva dalla corretta esecuzione di una fase iniziale di registrazione che si componeva di due sottofasi. Nella prima l'utente entrava nel sistema di acquisizione dati on-line tramite il proprio codice identificativo, che aveva ricevuto insieme alla lettera d'invito alla rilevazione. In questo modo accedeva a una pagina in cui gli veniva chiesto di indicare un indirizzo di posta elettronica al quale inviare una password temporanea. Tale pagina è mostrata nella Figura 2.6 dove è visibile anche l'utilizzo di un *CAPTCHA*³ per verificare che l'utente fosse un essere umano.

3 Con l'acronimo inglese CAPTCHA si denota un test fatto di una o più domande e risposte per determinare se l'utente sia un umano e non un computer o, più precisamente, un programma. L'acronimo deriva dall'inglese "Completely Automated Public Turing test to tell Computers and Humans Apart".



Figura 2.6 - Pagina di fornitura dell'indirizzo e-mail

The screenshot shows a web form for providing an email address. At the top, it reads "OLEIFICIO COLTIVATORI DIRETTI COOPERATIVA AGRICOLA". Below this, the instruction "Indicare di seguito l'e-mail ove desidera ricevere password provvisoria:" is followed by a text input field. Underneath is the label "Ripetere la E-mail:" with another text input field. A CAPTCHA image with the characters "N5D XCB" is displayed. Below the CAPTCHA is the instruction "Inserire la sequenza di caratteri:" followed by a text input field. At the bottom of the form are two buttons: "Continua" and "Disconnetti".

Una volta ricevuta la password temporanea, era possibile completare la fase di registrazione. All'utente veniva chiesto di inserire la password temporanea e quindi di scegliere una propria password personale (Figura 2.7).

Figura 2.7 - Pagina scelta password personale

The screenshot shows a web form for selecting a personal password. At the top, it reads "OLEIFICIO". Below this is a "Disconnetti" button. The instruction "Password ricevuta tramite E-mail:" is followed by a text input field. Below that is the instruction "Nuova Password (minimo 8 caratteri di cui uno numerico):" followed by a text input field. Underneath is the instruction "Ripetere la nuova Password:" followed by a text input field. At the bottom of the form is a "Continua" button.

Da questo momento l'utente poteva accedere effettivamente al questionario on-line per la compilazione usando il suo codice identificativo e la password personale.

L'accesso al questionario on-line è stato gestito in maniera tale da impedire accessi contemporanei sia da parte dello stesso utente, attraverso l'apertura di più finestre o *tab* del browser sullo stesso computer, sia da parte di utenti diversi da computer differenti.

Per ogni questionario aperto, infatti, il sistema memorizzava l'identificativo della sessione di lavoro corrispondente. Quando un utente accedeva a un questionario il sistema verificava se quel questionario era già aperto. In caso affermativo si potevano verificare due casi.

Il primo è quello in cui la sessione di lavoro dell'utente che accedeva al questionario aveva lo stesso identificativo memorizzato. Questo significava che il questionario era già aperto sullo stesso computer in un altro browser o in un altro *tab* dello stesso browser: l'accesso veniva inibito, invitando l'utente a chiudere le altre finestre aperte prima di proseguire.

Il secondo caso è quello in cui la sessione di lavoro dell'utente che stava accedendo al questionario aveva un identificativo diverso da quello memorizzato. Questo significava che un altro utente (o lo stesso utente) stava accedendo allo stesso questionario da un altro computer: a questo secondo utente veniva consentito l'accesso, ma la sessione di lavoro del precedente utente veniva invalidata. Non appena il vecchio utente provava a interagire col server, gli appariva un messaggio di notifica dell'acquisizione del controllo del questionario da parte di un altro utente.

La scelta di questa procedura è stata resa necessaria dal fatto di non aver trovato un modo affidabile per rilevare l'uscita dal questionario, nel caso in cui si potesse verificare un evento accidentale (errore di sistema) o l'utente chiudesse direttamente il browser senza utilizzare l'apposito pulsante di chiusura previsto dall'applicazione. Questo induceva a considerare erroneamente una sessione di lavoro abbandonata come ancora aperta.

In ogni caso questa modalità di gestione degli accessi metteva al riparo da possibili compilazioni incongruenti del questionario.

2.2.2. L'interfaccia grafica

La progettazione grafica del questionario elettronico ha richiesto una cura particolare per risolvere le difficoltà insite nel riportare un questionario articolato su quindici pagine in un formato che risultasse semplice per una compilazione via web.

La prima risposta a tale complessità è stata data con l'organizzazione della barra principale di navigazione (Figura 2.8).

Figura 2.8 - Banner e Menù principale



Questi elementi inseriti nella parte alta della pagina rimanevano presenti e attivi durante tutta la fase di compilazione.

In alto a sinistra, accanto ai pulsanti per la scelta della lingua alternativa di compilazione (tedesco e sloveno), è stato inserito il logo del 6° Censimento Generale dell'Agricoltura.

La barra di navigazione ha ricalcato la suddivisione in sezioni del questionario carta-



ceo, accessibili tutte indipendentemente l'una dall'altra senza un ordine prestabilito. Questa scelta, intenzionalmente voluta, pur comportando una maggiore complessità di sviluppo, ha consentito all'utente di muoversi liberamente tra le varie sezioni, come se effettivamente potesse sfogliare il questionario cartaceo scegliendo quale compilare per prima. La compilazione, quindi, non è stata necessariamente sequenziale.

Inoltre, l'utente poteva decidere in ogni momento di interrompere la compilazione e riprenderla successivamente dalla sezione interrotta, a condizione che ad ogni interruzione provvedesse a salvare i dati immessi cliccando sull'icona , posizionata in alto a destra della pagina.

In questa parte dello schermo erano presenti anche altri pulsanti di utilità.

Accanto al tasto per il salvataggio vi erano due icone con le frecce   (**Avanti e Indietro**), che consentivano di scorrere le pagine senza agire direttamente sulle etichette del menu.

Era presente infine l'icona  per tornare alla pagina iniziale di riepilogo.

Nel caso in cui l'utente non avesse terminato la compilazione della scheda visualizzata sullo schermo, poteva comunque decidere di interrompere momentaneamente la compilazione. Nel caso in cui i dati fossero stati inseriti correttamente, era possibile uscire dalla sessione selezionando l'icona .

La selezione di un'etichetta consentiva di visualizzare, nel caso delle Sezioni dalla II alla VI, un'ulteriore sotto barra di navigazione, come illustrato a titolo di esempio in Figura 2.9.

Figura 2.9 - Esempio di navigazione di secondo livello

INFO	Sez. 1 Generali	Sez. 2 Terreni	Sez. 3 Allevamenti	Sez. 4 Ubicazione	Sez. 5 Lavoro	Sez. 6 Altro				
		Seminativi	Legnose	Altro	Vitigni	Metodi				

Sezione II - NOTIZIE PARTICOLARI SULLA VITE (annata agraria 2009 - 2010)

21 NATURA DELLA PRODUZIONE														
21.1 Uva per la produzione di vini a denominazione di origine controllata (vini DOC) e controllata e garantita (vini DOCG)														
VITIGNO (denominazione)	Codice	SUPERFICIE TOTALE A VITE		SUPERFICIE INVESTITA A VITE SECONDO L'ANNO DI IMPIANTO										
				Posteriore ad agosto 2007		Da settembre 2004 ad agosto 2007		Da settembre 2000 ad agosto 2004		Da settembre 1990 ad agosto 2000		Da settembre 1980 ad agosto 1990		A set
		Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	
FALANCHINA BIANCA	1 079	2		1		1								
ABBONDOSA	1													
ABBUNDAS	1 999	2		1		1								
ABBUOTO N.														
ABOTO														
ABROSTINO														
ABRUSCO														
ABRUSCO N.														
ACCHIAPPALMENTO														
ACQUI														
AGLIANICO DEL VULturne N.														
altri vini														
VITIGNO (denominazione)	Codice	SUPERFICIE TOTALE A VITE		SUPERFICIE INVESTITA A VITE SECONDO L'ANNO DI IMPIANTO										
				Posteriore ad agosto 2007		Da settembre 2004 ad agosto 2007		Da settembre 2000 ad agosto 2004		Da settembre 1990 ad agosto 2000		Da settembre 1980 ad agosto 1990		A set
		Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	

2.2.3 L'interazione con il sistema

L'interazione dell'utente con il sistema è avvenuta secondo le seguenti modalità, in funzione del quesito:

- selezione da lista tramite bottone circolare (Figura 2.10);
- selezione Sì/No tramite bottone circolare (Figura 2.11);
- selezione multipla tramite casella di spunta (Figura 2.12 e Figura 2.13);
- selezione da lista (Figura 2.14);
- inserimento dati numerici (Figura 2.15);
- inserimento testo (Figura 2.16);
- inserimento testo da lista come ad esempio comuni, denominazione vitigni (Figura 2.17);
- inserimento riga aggiuntiva in tabella (Figura 2.18 e Figura 2.19);
- cancellazione riga da tabella (Figura 2.20 e Figura 2.21).

Selezione da lista tramite bottone circolare (radio button)

Figura 2.10 - Esempio di selezione da lista tramite bottone circolare (radio button)

1 FORMA GIURIDICA

1.1 Azienda individuale

1.2 Società semplice

1.3 Altra società di persona (S.n.c., S.a.s., ecc.)

1.4 Società di capitali (S.p.a., S.r.l., ecc.)

1.5 Società cooperativa

1.6 Amministrazione ed Ente pubblico (Stato, Regioni, ecc.)

1.7 Ente (Comunanze, Università, Regole, ecc.) o Comune che gestisce proprietà collettive

1.8 Ente privato senza fini di lucro

1.9 Altra forma giuridica

Specificare

Selezione Sì/No tramite bottone circolare

Figura 2.11 - Esempio di selezione Sì/No tramite bottone circolare

6 INFORMATIZZAZIONE DELL'AZIENDA

6.1 L'azienda dispone di personal computer per fini aziendali? Sì No

6.1.1 L'Azienda usa normalmente proprie attrezzature informatiche per:

a) Servizi amministrativi Sì No

b) Gestione informatizzata di coltivazioni Sì No

c) Gestione informatizzata degli allevamenti Sì No

6.2 L'azienda utilizza normalmente la rete internet per le proprie attività? Sì No

6.3 L'azienda ha un sito web oppure una o più pagine su internet? Sì No

6.4 L'azienda fa uso di commercio elettronico per:

a) La vendita di prodotti e servizi aziendali Sì No

b) L'acquisto di prodotti e servizi Sì No



Selezione multipla tramite casella di spunta

Figura 2.12 - Esempio di selezione multipla tramite casella di spunta

5 ELEMENTI DEL PAESAGGIO AGRARIO			
Indicare la presenza di elementi lineari del paesaggio:			
Elementi	Codice	Sottoposti a manutenzione durante gli ultimi tre anni	Di nuova realizzazione negli ultimi tre anni
5.1 Siepi	01	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5.2 Filari di alberi	02	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
5.3 Muretti	03	<input type="checkbox"/>	<input type="checkbox"/>

Figura 2.13 - Altro esempio di selezione multipla tramite casella di spunta

7 SOSTEGNO ALLO SVILUPPO RURALE	
7.1 Indicare se l'azienda ha beneficiato di una o più delle seguenti misure nel corso del 2008-2009-2010	
a)	<input type="checkbox"/> Insediamento di giovani agricoltori (misura 112)
b)	<input type="checkbox"/> Utilizzo di servizi di consulenza (misura 114)
c)	<input type="checkbox"/> Ammodernamento delle aziende agricole (misura 121)
d)	<input type="checkbox"/> Accrescimento del valore aggiunto dei prodotti agricoli e forestali (misura 123)
e)	<input type="checkbox"/> Cooperazione per lo sviluppo di nuovi prodotti, processi e tecnologie nel settore agricolo e alimentare e in quello forestale (misura 124)
f)	<input type="checkbox"/> Rispetto delle norme basate sulla legislazione comunitaria (misura 131)
g)	<input type="checkbox"/> Partecipazioni degli agricoltori ai sistemi di qualità alimentare (misura 132)
h)	<input type="checkbox"/> Indennità a favore degli agricoltori delle zone montane (misura 211)
i)	<input type="checkbox"/> Indennità a favore degli agricoltori delle zone caratterizzate da svantaggi naturali diverse da zone montane (misura 212)
l)	<input type="checkbox"/> Indennità Natura 2000 (misura 213)
m)	<input type="checkbox"/> Indennità connesse alla Direttiva Quadro 2000/60/CE sulle acque (misura 213)
n)	<input type="checkbox"/> Pagamenti agro-ambientali (misura 214)
- n1	<input type="checkbox"/> di cui nel quadro dell'agricoltura biologica
- n2	<input type="checkbox"/> di cui nel quadro dell'agricoltura integrata
o)	<input type="checkbox"/> Pagamenti per il benessere degli animali (misura 215)
p)	<input type="checkbox"/> Sostegno agli investimenti non produttivi (misura 216)
q)	<input type="checkbox"/> Diversificazione in attività non agricole (misura 311)
r)	<input type="checkbox"/> Incentivazione di attività turistiche (misura 313)

Selezione da lista

Figura 2.14 - Esempio di selezione da lista

43 FAMIGLIA DEL CONDUTTORE E PARENTI							
43.1 Conduttore (16 anni e più - responsabile giuridico ed economico dell'azienda)							
Codice	Sesso	Anno di nascita	Cittadinanza	Codizione professionale	GIORNATE DI LAVORO SVOLTE IN AZIENDA (attività agricole e connesse)		% del tempo di lavoro aziendale dedicato ad attività connesse (quesito 48 a pagina 13)
					Numero giorni	Media ore giornaliera	
101	<input type="text"/>	<input type="text"/>	<input type="text"/> <ul style="list-style-type: none"> Italiana Altro paese Unione Europea Paese extra-Unione Europea 	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Inserimento dati numerici

Figura 2.15 - Esempio di inserimento dati numerici

8 SEMINATIVI			
	Codice	Superficie	
		Coltivazione principale	
		Ettari	Are
8.1 Cereali per la produzione di granella			
a) Frumento tenero e spelta	01	12	20
b) Frumento duro	02	14	00
c) Segale	03		
d) Orzo	04		
e) Avena	05		
f) Mais (escluso mais in erba ed a maturazione cerosa)	06		

Inserimento testo

Figura 2.16 - Esempio di inserimento testo

E UBICAZIONE DEL CENTRO AZIENDALE	
Nota: Questo riquadro deve essere compilato solo se l'ubicazione del centro aziendale è diversa dalla residenza o dalla sede legale del conduttore.	
Per 'centro aziendale' si intende il complesso dei fabbricati connessi all'attività aziendale situato entro il perimetro dei terreni aziendali oppure, in assenza di fabbricati, il luogo che identifica la maggior parte della superficie aziendale	
Indirizzo (Via/Piazza/Località e numero civico del centro aziendale)	C.A.P.
<input type="text"/>	<input type="text"/>

Inserimento testo da lista

Figura 2.17 - Esempio di inserimento testo da lista

21 NATURA DELLA PRODUZIONE																		
21.1 Uva per la produzione di vini a denominazione di origine controllata (vini DOC) e controllata e garantita (vini DOCG)																		
VITIGNO (denominazione)	Codice	SUPERFICIE TOTALE A VITE		SUPERFICIE INVESTITA A VITE SECONDO L'ANNO DI IMPIANTO										+	X			
				Posteriore ad agosto 2007		Da settembre 2004 ad agosto 2007		Da settembre 2000 ad agosto 2004		Da settembre 1990 ad agosto 2000		Da settembre 1980 ad agosto 1990				Anteriore al settembre 1980		
		Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are	Ettari	Are			Ettari	Are	
agl	1																	
AGLIANICO DEL VULTURE	1 999																	
AGLIANICO N.																		
AGLIANICONE N.																		
AGLIATICA																		
AGLIATICO																		



Inserimento riga aggiuntiva in tabella

In alcuni casi si è richiesta la compilazione di tabelle in cui i dati si riferivano a una o più unità presenti in azienda (es. persone della famiglia, vitigni coltivati, eccetera). Si è reso necessario quindi dare la possibilità di aggiungere, o eventualmente eliminare, righe di tabella.

Figura 2.18 - Esempio di tabella

SEZIONE V - LAVORO ED ATTIVITA' CONNESSE (annata agraria 2009 - 2010)

43.3 Altri componenti della famiglia (16 anni e più) che lavorano in azienda

Codice	Sesso	Anno di nascita	Cittadinanza	Codizione professionale	GIORNATE DI LAVORO SVOLTE IN AZIENDA (attività agricole e connesse)		% del tempo di lavoro aziendale dedicato ad attività connesse (quesito 48 a pagina 13)	ALTRE ATTIVITA' REMUNERATIVE EXTRA-AZIENDALI				
					Numero giorni	Media ore giornaliera		Tempo dedicato	Settore di attività prevalente	Posizione		
301												 

Qualora una riga non fosse stata sufficiente ad esaurire i casi relativi all'azienda,

era possibile aggiungere una nuova riga selezionando l'icona  della riga dopo la quale si voleva aggiungerne una nuova.

Figura 2.19 - Esempio di inserimento riga aggiuntiva in tabella

43.3 Altri componenti della famiglia (16 anni e più) che lavorano in azienda

Codice	Sesso	Anno di nascita	Cittadinanza	Codizione professionale	GIORNATE DI LAVORO SVOLTE IN AZIENDA (attività agricole e connesse)		% del tempo di lavoro aziendale dedicato ad attività connesse (quesito 48 a pagina 13)	ALTRE ATTIVITA' REMUNERATIVE EXTRA-AZIENDALI			
					Numero giorni	Media ore giornaliera		Tempo dedicato	Settore di attività prevalente	Posizione	
301	Mas	1955	Italiana	Occupato	300	8	100	Nessun ter	Agricoltura	Imprenditor	 
302											 

Cancellazione riga da tabella

È stato possibile cancellare una riga selezionando la relativa icona .

Figura 2.20 - Esempio di cancellazione riga in tabella - prima della cancellazione (1)

43.3 Altri componenti della famiglia (16 anni e più) che lavorano in azienda

Codice	Sesso	Anno di nascita	Cittadinanza	Codizione professionale	GIORNATE DI LAVORO SVOLTE IN AZIENDA (attività agricole e connesse)		% del tempo di lavoro aziendale dedicato ad attività connesse (quesito 48 a pagina 13)	ALTRE ATTIVITA' REMUNERATIVE EXTRA-AZIENDALI			
					Numero giorni	Media ore giornaliera		Tempo dedicato	Settore di attività prevalente	Posizione	
301	Mas	1955	Italiana	Occupato	300	8	100	Nessun ter	Agricoltura	Imprenditor	 
302	Fem	1953	Italiana	Occupato	300	8	100	Nessun ter	Agricoltura	Imprenditor	 

Figura 2.21 - Esempio di cancellazione riga in tabella - dopo la cancellazione (2)

43.3 Altri componenti della famiglia (16 anni e più) che lavorano in azienda

Codice	Sesso	Anno di nascita	Cittadinanza	Codizione professionale	GIORNATE DI LAVORO SVOLTE IN AZIENDA (attività agricole e connesse)		% del tempo di lavoro aziendale dedicato ad attività connesse (quesito 48 a pagina 13)	ALTRE ATTIVITA' REMUNERATIVE EXTRA-AZIENDALI			
					Numero giorni	Media ore giornaliera		Tempo dedicato	Settore di attività prevalente	Posizione	
301	Mas	1955	Italiana	Occupato	300	8	100	Nessun ter	Agricoltura	Imprendito	+ X

2.2.4 Messaggi di errore e salvataggio parziale

Se i dati inseriti contenevano degli errori, dopo aver cliccato sul tasto "Salva", si apriva una finestra di attenzione circa l'operazione di salvataggio dati (Figura 2.22).

L'utente poteva quindi scegliere se risolvere l'errore subito o in seguito. In quest'ultimo caso bastava premere il pulsante "Salva bozza" che compare in fondo alla pagina.

Figura 2.22 - Messaggio di errore

Denominazione Comune
ERRORE. (007) Comune errato
 San Michele di Ganzaria

Numero

E-mail

Indirizzo

Esito salvataggio dati

ATTENZIONE: sono stati riscontrati degli errori di compilazione.

E' possibile correggere subito i dati inseriti oppure salvare temporaneamente i dati senza alcuna correzione cliccando su "SALVA BOZZA" in fondo alla pagina

Contemporaneamente al messaggio di attenzione - circa l'operazione di salvataggio dati - compariva, ad indicare il punto della pagina in cui era stato riconosciuto l'errore, un messaggio che specificava quale errore fosse stato commesso. I messaggi potevano essere di due colori, giallo e azzurro.

Il messaggio su fondo giallo indicava un errore che consentiva il salvataggio in bozza, ma non quello definitivo. Ciò vuol dire che, se al termine della compilazione gli errori di questo tipo non venivano risolti, non era possibile inviare il questionario.

Il messaggio su fondo azzurro indicava invece un possibile errore che, anche se non risolto, consentiva il salvataggio della pagina (Figura 2.23), anche in forma definitiva.

Figura 2.23 - Messaggio di possibile errore

Indirizzo (Via/Piazza/Località e numero civico)
ATTENZIONE: (011) E' stato modificato il comune ma non l'indirizzo
 VIA SOTTOTENENTE LA ROSA,45

Denominazione Provincia
 Catania

Denominazione Comune
 San Michele di Ganzaria
 Acireale

Nel caso in cui l'utente avesse aperto una nuova scheda, senza procedere al salvataggio dei dati della precedente pagina corrente, un messaggio di avviso ricordava all'utente che ciò avrebbe comportato la perdita delle informazioni inserite (Figura 2.24).

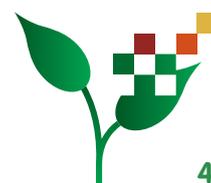
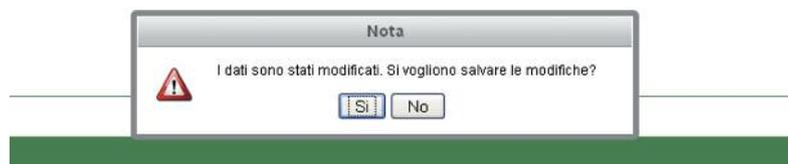


Figura 2.24 - Avviso su rischio di perdita dati



2.2.5 Pagina di riepilogo sullo stato della compilazione

La pagina iniziale riportava le informazioni di riepilogo circa lo stato di avanzamento della compilazione. Per ogni sezione del questionario era infatti segnalato se la compilazione risultava completa, incompleta o assente (Figura 2.25).

Per segnalare come completa la compilazione di pagine del questionario on-line ove non fosse stato necessario inserire alcun dato, era necessario che l'utente avesse eseguito l'operazione di salvataggio in tali pagine.

L'operazione di salvataggio comportava il suo passaggio nello stato "completo" (contraddistinto dal segno di spunta verde).

Dalla *homepage* era inoltre possibile, grazie ai pulsanti presenti nell'area utilità, nella colonna di destra della pagina: i) inviare il questionario (una volta terminata la compilazione); ii) stampare il questionario (sia il fac-simile che la versione compilata); iii) stampare la ricevuta di avvenuto invio. Tra i pulsanti dell'*homepage* vi era, inoltre, il pulsante di accesso al glossario. Cliccando sul pulsante si apriva una nuova pagina in cui erano riportati, in ordine alfabetico, i termini chiave, utili alla compilazione del questionario.

Una volta inviato il questionario compilato da parte dell'azienda agricola, il sistema restituiva una ricevuta di avvenuta compilazione.

Figura 2.25 - Pagina di riepilogo

QUESTIONARIO ONLINE

INFO Sez. 1 Generali Sez. 2 Terreni Sez. 3 Allevamenti **Sez. 4 Ubicazione** Sez. 5 Lavoro Sez. 6 Altro

Livello di completamento: ✓ = assente ✓ = incompleto ✓ = completo

NOTA: come conferma di presa visione, è richiesto il salvataggio di ciascuna pagina anche se non necessariamente da compilare.

• INFO	
◦ Notizie Anagrafiche	✓
◦ Stato Dell'Azienda	✓
◦ Centro Aziendale	✓
• Sez. 1 - Generali	✓
• Sez. 2 - Terreni	✓
◦ Seminativi	✓
◦ Legnose	✓
◦ Altro	✓
◦ Vitiigni	✓
◦ Metodi	✓
• Sez. 3 - Allevamenti	✓
◦ Consistenza	✓
◦ Metodi	✓
• Sez. 4 - Ubicazione	✓
• Sez. 5 - Lavoro	✓
◦ Manodopera	✓
◦ Connesse	✓
• Sez. 6 - Altro	✓

- Nota introduttiva
- Glossario
- Facsimile questionario
- Stampa questionario compilato
- Invio questionario compilato
- Stampa ricevuta avvenuto invio
- Modifica Password

2.3 Il Sistema di Gestione della Rilevazione

2.3.1 Caratteristiche generali

Il Sistema di Gestione della Rilevazione (SGR) è nato con lo scopo di coadiuvare i vari organi della rete di rilevazione nella conduzione delle operazioni sul campo e nel contempo consentire il monitoraggio della rilevazione. Le innovazioni metodologiche e organizzative adottate dall'Istat, quali l'utilizzo della lista pre-censuaria, l'acquisizione multicanale dei dati e la gestione distribuita delle fasi di *check*, non sarebbero state possibili senza un sistema informatico idoneo a supportarle. Tali innovazioni, infatti, implicando una maggiore complessità di gestione dell'indagine, necessitavano di una applicazione che ne facilitasse l'esecuzione e il controllo durante tutte le fasi, dall'acquisizione del dato fino alla sua validazione finale.

In generale l'applicazione SGR, accessibile tramite web, ha consentito una riduzione dei costi di gestione dell'indagine, una maggiore qualità dei dati prodotti e soprattutto la disponibilità in tempo reale dei dati elementari.

Più in particolare, SGR ha permesso agli operatori di lavorare da una qualunque postazione con accesso a internet, senza costringerli a tornare alla sede di lavoro. Quest'aspetto è molto importante, perché, per sfruttare al massimo le potenzialità delle innovazioni, era necessario che il rilevatore avesse le informazioni inerenti la rilevazione aggiornate in tempo reale; questo ha consentito di comprimere i tempi di raccolta dei dati e di diminuire il disturbo statistico verso le unità di rilevazione.

Inoltre SGR ha garantito un'elevata qualità del dato, resa possibile dal continuo scambio di informazioni tra l'organo centrale e la rete di rilevazione per l'esecuzione di tre tipi di controllo sui dati:

- *Check* in fase di acquisizione;
- *Check* micro;
- *Check* macro.

Nel Paragrafo 2.3.4 viene descritto in modo approfondito il processo di controllo e correzione adottato.

L'operatore si è potuto muovere in SGR in modo totalmente autonomo seguendo sempre un flusso di lavorazione ben delineato nel rispetto delle regole definite nel Piano generale di censimento. Inoltre, il sistema consentiva il riciclo all'interno del processo per la gestione di eventuali errori di lavorazione da parte degli organi della rete. Di fatto, in molte situazioni, l'operatore ha potuto scegliere di annullare o modificare le operazioni effettuate sempre compatibilmente con il workflow di riferimento. Questo è stato fondamentale per evitare che eventuali anomalie potessero essere risolte solo centralmente, diventando così un ostacolo per lo svolgimento dell'attività da parte della rete.

Particolare attenzione nello sviluppo degli applicativi è stata data alla diversificazione delle abilitazioni all'uso delle diverse funzioni del sistema in relazione al ruolo da ciascuno ricoperto nella complessa organizzazione della rete di rilevazione. Nel modello organizzativo ad Alta partecipazione gli Uffici Regionali di Censimento (URC) che hanno scelto di registrare in proprio i questionari compilati in forma cartacea, hanno gestito le fasi di *check* fino al consolidamento del dato provvisorio. Nel modello organizzativo a Partecipazione integrativa, invece, gli Uffici Comunali di Censimento (UCC) dovevano registrare soltanto alcune informazioni relative a variabili principali del questionario⁴, sulle quali era previsto un apposito piano di controllo, mentre venivano sottoposti allo stesso *check* adottato nel modello ad Alta partecipazione i soli questionari acquisiti tramite il web.

⁴ Registrazione del "Modello riepilogativo per singolo questionario" contenente le seguenti informazioni: stato di attività, ubicazione, coltivazioni (sette modalità), allevamenti (dieci modalità).



Da un punto di vista informatico tutto ciò ha comportato la necessità di predisporre e gestire due *workflow* diversi ma convergenti in molteplici punti. Ad esempio l'inserimento delle variabili primarie previste nel "Modello riepilogativo per singolo questionario" registrato dagli UCC e obbligatorio per le regioni a Partecipazione integrativa, è stato effettuato, anche per le regioni ad Alta partecipazione ma in modo automatico, desumendo i dati dal questionario inserito tramite data entry dall'ufficio di censimento ovvero dal questionario on line auto compilato dall'azienda agricola.

Inoltre, nel modello ad Alta partecipazione, l'URC poteva scegliere la struttura territoriale della rete di rilevazione, mentre la definizione della rete, per le regioni a Partecipazione integrativa era meno flessibile. La possibilità di scelta tra due modelli organizzativi e l'aggregazione territoriale, diversa tra Alta partecipazione e Partecipazione integrativa, ha causato un'elevata complessità nello sviluppo del sistema informatico. Nella Figura 2.26 e nella Figura 2.27 sono rappresentati i due flussi di lavorazione previsti dal Sistema a secondo del modello di organizzazione scelto.

Figura 2.26 - Flusso di lavorazione per il modello per Alta partecipazione

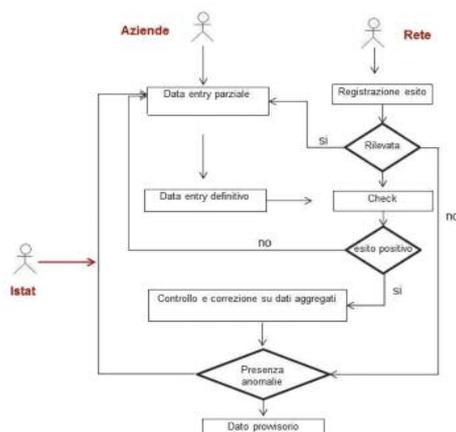
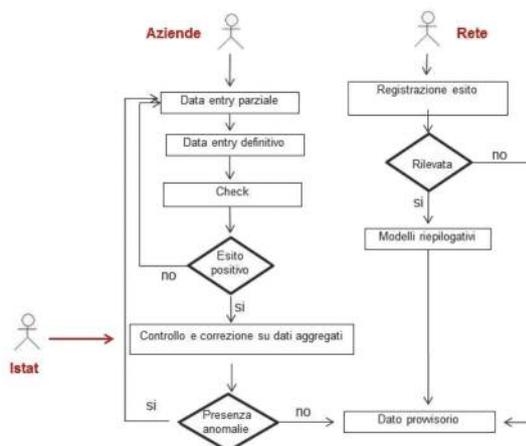


Figura 2.27 - Flusso di lavorazione per il modello per Partecipazione integrativa



I punti di forza di SGR sono stati:

- la disponibilità di un insieme di funzioni in grado di supportare la rete durante tutte le fasi del processo produttivo, mediante un flusso bidirezionale di dati ed informazioni;

- la possibilità di un monitoraggio di massimo dettaglio a disposizione sia della rete sia dell'Istat in tempo reale;
- la disponibilità dei microdati anche prima della chiusura della rilevazione.

Nel complesso, SGR ha compreso oltre 50 funzionalità raggruppate per tipologia in 5 macro-aree:

OPERATORI - consentiva la definizione della rete di rilevazione, dei profili utente e l'assegnazione delle aziende ai vari Rilevatori;

QUESTIONARI - comprendeva tutte le funzioni strettamente legate alla rilevazione (registrazione dell'esito, *data entry*, *check* dei dati,...);

RAPPORTI RIASSUNTIVI - comprendeva un set di rapporti di monitoraggio dell'andamento della rilevazione;

MODELLI RIEPILOGATIVI - comprendeva tutte le funzioni per l'acquisizione e il riepilogo delle variabili primarie necessarie per la pubblicazione dei dati provvisori;

UTILITÀ - comprendeva un insieme di funzioni trasversali all'intero processo di rilevazione di supporto alla rete, come ad esempio: ricerca di un'azienda, la stampa del questionario, la chiusura del censimento, la riapertura in lavorazione dei questionari.

2.3.2 Gestione della rete di rilevazione

Secondo quanto stabilito dal Regolamento di esecuzione⁵ e dal Piano generale di censimento, nei due modelli organizzativi attuati i principali organi erano:

- Ufficio Regionale di Censimento (URC);
- Ufficio Territoriale di Censimento (UTC): costituiti presso le Province, le Camere di commercio, le Comunità montane o gli Enti strumentali della Regione (Area sub regionale UTC);
- Ufficio Comunale di Censimento (UCC): costituiti presso singoli Comuni, presso gli uffici di statistica in forma associata o presso associazioni temporanee di comuni ove definite per ragioni di efficienza organizzativa o di specifiche condizioni territoriali (Area sub regionale UCC).

Nel modello organizzativo a partecipazione integrativa la rete era costituita dai tre suddetti organi. Nel modello organizzativo ad alta partecipazione, invece, la Regione aveva la possibilità di disegnare la rete secondo una delle seguenti modalità:

- solo URC: la Regione in questo caso assume il compito di eseguire la rilevazione sul campo organizzandosi in modo completamente autonomo e indipendente;
- URC e UTC: demandando a questi ultimi la rilevazione sul campo;
- URC e UCC: demandando a questi ultimi la rilevazione sul campo, venendo incontro alla necessità di una maggiore granularità sul territorio;
- URC, UTC, UCC: consentendo una gestione della rilevazione a più livelli di responsabilità territoriale e funzionale, ovvero demandando la rilevazione sul campo agli uffici comunali e assegnando agli uffici territoriali il loro coordinamento e monitoraggio.

Per gli URC, UTC e UCC il Regolamento e il Piano generale di censimento stabilivano le funzioni e i compiti che era possibile assegnare ai responsabili di ufficio di censimento e agli altri membri⁶. I singoli piani regionali di censimento esplicitavano le scelte di ciascuna regione. In SGR si è tenuto conto delle scelte organizzative fatte da ogni

⁵ DPR 23 luglio 2010, n. 154, Regolamento di esecuzione del 6° Censimento generale dell'agricoltura, a norma dell'articolo 17, comma 2, del decreto-legge 25 settembre 2009, n. 135. (10G0168) (GU n. 214 del 13-9-2010).

⁶ Per approfondimenti si rimanda al fascicolo *L'organizzazione della rilevazione e le informazioni censuarie* facente parte della presente pubblicazione.



regione. Inoltre, per consentire una migliore suddivisione delle responsabilità e dei compiti, i Rilevatori sono stati distinti sulla base delle mansioni loro assegnate in:

- Rilevatori ad alta operatività (Ril Alta),
- Rilevatori a media operatività (Ril Media)
- Rilevatori a bassa operatività (Ril Bassa).

A ogni utente del sistema è stato associato un ruolo e un profilo applicativo, in base al quale SGR veniva dinamicamente personalizzato rispetto a:

- visibilità territoriale dei dati;
- insieme di funzionalità accessibili all'interno del sistema.

In particolare, al fine di consentire la gestione distribuita della rilevazione e di garantire al contempo che ogni utente potesse accedere esclusivamente alle informazioni per le quali era autorizzato, ad ogni profilo è stata associata una visibilità territoriale:

- utenti con competenze regionali, provinciali o comunali - rispettivamente URC, UTC, UCC - potevano visualizzare, monitorare e gestire le informazioni relative alle unità di rilevazione del territorio di propria pertinenza;
- utenti con profili più operativi potevano gestire solo le aziende che gli erano state assegnate. Tale granularità a livello di singola azienda ha consentito di gestire anche delle casistiche particolari, quali ad esempio l'individuazione, durante la rilevazione sul campo, e la conseguente lavorazione di aziende con conduttore residente in una regione diversa da quella di pertinenza.

Prospetto 2.1 - Ruoli e profili applicativi in SGR

RUOLO	Profilo	Visibilità territoriale
URC	Responsabile URC Membro URC	Regionale
UTC	Responsabile UTC Membro UTC	Area sub regionale UTC
UCC	Responsabile UCC Membro UCC	Area sub regionale UCC
CiC	RpCiC CiC	Aziende assegnate ai CiC che coordina Aziende assegnate
CoC	CoC	Aziende assegnate ai Ril che coordina
Ril	Ril Alta Ril Media Ril Bassa	Aziende assegnate

Ogni profilo applicativo, inoltre, consentiva di differenziare le operazioni che gli utenti potevano effettuare all'interno del sistema; infatti, sulla base delle responsabilità assegnate nello svolgimento della rilevazione, venivano abilitate le opportune funzioni. Per una gestione completa della rete il sistema offriva già delle funzioni di monitoraggio di dettaglio, oltre alle funzionalità di creazione e di assegnazione. Di seguito sono illustrate le funzionalità comprese nella macro-area **Operatori**:

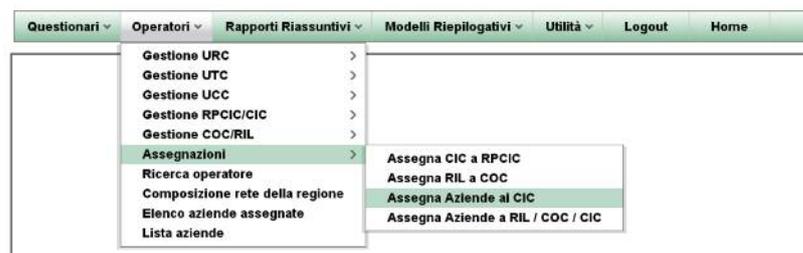
- **gestione operatori:** permetteva di creare, modificare, cancellare operatori con vari profili applicativi, ad eccezione degli operatori nazionali e dei responsabili regionali che sono stati pre-caricati nel sistema prima dell'inizio della rilevazione (Figura 2.28). Le funzioni d'inserimento consentivano di specificare informazioni quali l'anagrafica dell'utente, il profilo associato e, ove necessario, il territorio di pertinenza. In qualsiasi momento, gli utenti abilitati potevano eliminare un operatore dalla rete o modificare i dati di anagrafica degli operatori precedentemente inseriti;

Figura 2.28 - Menù operatori: funzioni di gestione



- **assegnazioni:** garantivano un corretto andamento della rilevazione e un puntuale monitoraggio. Queste funzioni permettevano, infatti, di associare gli operatori ai rispettivi responsabili e le aziende ai Rilevatori. Per offrire una maggiore flessibilità nella gestione della rilevazione le aziende potevano essere assegnate anche ai CIC ed eventualmente ai COC. Tali operazioni dovevano essere effettuate prima dell'inizio della rilevazione sul campo, poiché consentivano la distribuzione del lavoro tra gli operatori della rete (Figura 2.29);

Figura 2.29 - Menù operatori: funzioni di assegnazione



- funzioni di utilità per la **ricerca degli operatori** e per la visualizzazione della **composizione della rete** della regione: la prima funzione consentiva di ricercare un operatore impostando opportuni filtri quali, ad esempio, *userid*, profilo applicativo, area sub regionale, eccetera; la seconda mostrava i profili applicativi previsti per la regione selezionata, l'area sub regionale ove presente, nome e cognome dei responsabili (Figura 2.30). Dalla schermata principale era possibile accedere ad una pagina di dettaglio relativa ai singoli operatori;
- **elenco aziende assegnate:** consentiva il monitoraggio dell'andamento della rilevazione, mostrando per ciascun'azienda lo stato di lavorazione del questionario, l'esito dell'intervista e l'operatore assegnatario. Per i profili applicativi che operavano sul campo era possibile la visualizzazione delle sole aziende di propria competenza. Tale funzione era quindi uno strumento completo: per il rilevatore che aveva un supporto al lavoro sul campo, continuamente aggiornato anche in base alle compilazioni web portate a termine dai rispondenti; per gli operatori di più alto livello che avendo visibilità sul territorio di competenza potevano monitorare l'andamento della rilevazione (Figura 2.31);

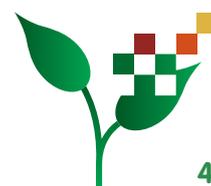


Figura 2.30 - Composizione della rete di rilevazione

Composizione rete di rilevazione per la regione Lazio

Tipo partecipazione: Alta partecipazione
 Area sub regionale UTC: Prevista
 Area sub regionale UCC: Prevista

39 record trovati, visualizzati 1 - 15.
 [Prima/Prec.] 1, 2, 3 [Success./Ultima]

Ruolo	Area sub regionale	Nome cognome / N. componenti	Userid
Responsabile URC		Mario Bianchi	010301
URC		30	
Responsabile URC		Mario Rossi	191803
URC		30	
Responsabile CIC		Mario Verdi	190001
CIC		30	
Responsabile UTC	Provincia di Roma	Mario Gialli	190093
UTC	Provincia di Roma	14	
Responsabile UCC		Mario Neri	190068
UCC		10	
COC		3	
Rilevatore piena operatività		8	
Rilevatore media operatività		5	
Rilevatore bassa operatività		8	

Esporta dati:  

Figura 2.31 - Elenco aziende assegnate

Elenco aziende assegnate 121067 - Mario Rossi

17 record trovati, visualizzati 1 - 10.
 [Prima/Prec.] 1, 2 [Success./Ultima]

Cod. Azienda	Denominazione azienda	Comune residenza	Stato del questionario	Esito Intervista	Provenienza	Nominativo CIC
0110135	AZIENDA AGRICOLA 1	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	Giorgio Bianchi - 120061
0496667	AZIENDA AGRICOLA 2	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	120061
0790986	AZIENDA AGRICOLA 3	Roma	Questionari senza anomalie dopo controlli macro	Azienda in lista non esistente o doppiata: Terreni destinati a soli orti familiari o allevamenti per autoconsumo o aziende esclusivamente forestali	WEB	Giorgio Bianchi - 120061
0849834	AZIENDA AGRICOLA 4	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	120061
0926632	AZIENDA AGRICOLA 5	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	Giorgio Bianchi - 120061
0994385	AZIENDA AGRICOLA 6	Roma	Questionari con check ok dopo controlli macro	Azienda rilevata	WEB	120061
1143080	AZIENDA AGRICOLA 7	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	WEB	Giorgio Bianchi - 120061
1308926	AZIENDA AGRICOLA 8	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	WEB	120061
1320126	AZIENDA AGRICOLA 9	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	Giorgio Bianchi - 120061
1335588	AZIENDA AGRICOLA 10	Roma	Questionari senza anomalie dopo controlli macro	Azienda rilevata	OPERATORE	120061

Esporta dati:  

- **lista aziende:** consentiva di ricercare le aziende presenti o non presenti nella lista pre-censuaria, già assegnate ad un operatore o da assegnare. Era possibile filtrare ulteriormente i risultati della ricerca in base all'esito della rilevazione, allo stato del questionario o ai filtri territoriali. Venivano successivamente mostrate le informazioni relative alle aziende trovate (stato, esito, denominazione, eccetera).

2.3.3 Data entry

Nel sistema SGR era possibile immettere i dati elementari, raccolti tramite questionario cartaceo, utilizzando la stessa applicazione messa a disposizione delle aziende agricole (Figura 2.32).

Questa funzione, denominata “*data entry*” era presente sotto la voce **Questionari**, ed è stata utilizzata sia per il primo inserimento, sia per la modifica dei dati, sia per la visualizzazione del questionario. La funzione, disponibile solo per gli utenti abilitati, permetteva l’inserimento dei dati solo dopo la verifica di alcune regole di coerenza con il territorio di pertinenza e dello stato del questionario. In particolare si controllava:

- il tipo di modello di partecipazione della regione, permettendo *la registrazione del questionario* solo nel caso dell’alta partecipazione e con registrazione in proprio del questionario;
- la competenza del territorio, permettendo il data entry ad ogni operatore solo per il territorio nel quale poteva operare. Nel caso dei Rilevatori il controllo è stato effettuato a livello di singolo questionario, dando la possibilità di compilare solo i questionari relativi ad aziende loro assegnate;
- lo stato di lavorazione del questionario, ossia l’effettuazione dell’operazione preliminare di registrazione dell’intervista;
- l’esito della rilevazione, ossia che l’unità risulti rilevata.

Le utenze abilitate alla funzione di *data entry* potevano operare in modo diverso sui questionari a seconda del loro profilo applicativo e del loro ruolo:

- gli URC, UTC e UCC su tutte le aziende del loro territorio, loro assegnate sulla base della residenza del conduttore o sede legale dell’azienda, secondo quanto risultava dalla lista pre-censuaria, nonché quelle rilevate ex-novo sul campo;
- i CIC ed i Rilevatori sulle aziende a loro assegnate tramite la funzione di assegnazione;
- i coordinatori sulle aziende assegnate ai loro Rilevatori.

Figura 2.32 - Maschera di accesso al data entry

Questionari - Operatori - Rapporti Riassuntivi - Modelli Riprogettati - Utilità - Logout - Home

Data entry >

Selezionare il questionario:

Numero identificativo Istat

Conferma

Istat - Istituto Nazionale di Statistica
Via Cesare Balbo, 16 00184 - Roma tel. +39 06 46731

© 2010 Istat

Se tutte le condizioni erano verificate, le utenze abilitate potevano eseguire l’immissione dei dati. Il questionario informatico messo a disposizione della rete di rilevazione per il data entry era analogo a quello delle aziende agricole ed aveva quindi il medesimo comportamento illustrato al Paragrafo 2.2. del presente fascicolo. Al termine della compilazione di ogni pagina era richiesto il salvataggio dei dati secondo due modalità: “salva definitivo” e “salva in bozza”. Dopo l’invio definitivo il questionario risultava



chiuso e non poteva essere più modificato se non dopo la fase di *check* e solo in caso di errori sui quali non era stato predisposto un controllo automatico durante la compilazione on-line.

2.3.4 Check

Come anticipato nel Paragrafo 2.3.1, il piano di controllo e correzione dei dati è stato suddiviso in tre momenti diversi, di seguito descritti nel dettaglio.

Nella fase di inserimento dati nel questionario on-line, si attivava un controllo sui dati basato principalmente su regole semplici come il controllo sul tipo di dato digitato dall'utente, i controlli relativi al dominio delle variabili e alla corretta somministrazione delle domande in base alle risposte fornite alle domande filtro, oppure sull'obbligatorietà di risposta per alcuni quesiti. Questo primo tipo di controllo era a disposizione, in maniera automatica, sia dell'unità di rilevazione che dell'operatore incaricato della registrazione dei dati contenuti nei questionari compilati in forma cartacea. In un questionario complesso come quello del censimento dell'agricoltura si è dovuto valutare il corretto *trade off* tra il numero di controlli da inserire nel questionario elettronico e la difficoltà e la complessità di compilazione da parte delle unità di rilevazione. Infatti, tanto più numerosi sono i controlli tanto più complessa è la compilazione telematica e tanto maggiore è il rischio di abbandono della compilazione ovvero di compilazione scientemente errata o incompleta. Inoltre, nella progettazione dei controlli interattivi si è tenuto conto anche delle implicazioni sulla velocità di lavorazione informatica nella registrazione del questionario.

La necessità di garantire una maggiore qualità dei dati raccolti, tuttavia, ha fatto sì che i controlli ulteriori fossero collocati in una fase successiva a quella di raccolta. In questo modo il lavoro di controllo è stato demandato alla rete di rilevazione, ossia, a persone formate e competenti in grado quindi di gestire problematiche più complesse. Questi controlli, nel sistema, sono stati immessi in due fasi diverse denominate "*check micro*" e "*check macro*". La prima tipologia di *check* aveva il compito di eseguire dei controlli di congruenza dei dati immessi nelle varie sezioni del questionario, effettuando anche dei confronti incrociati con dati presi da altre fonti per evidenziare eventuali anomalie. La seconda fase, invece, si proponeva di effettuare dei controlli per evidenziare eventuali *outlier*; questa seconda fase poteva essere eseguita solo dopo aver completato la prima.

Il "*check micro*" è stato eseguito su ogni questionario acquisito via web o registrato dagli operatori della rete. Questo era composto da un insieme di regole esclusivamente di tipo deterministico, anche queste classificate in "Errori" che dovevano essere corretti obbligatoriamente e "Accertamenti", ovvero segnalazioni di possibili anomalie non necessariamente bloccanti il processo di lavorazione. Come abbiamo detto, il "*check micro*" era a disposizione dell'operatore del censimento che attivava una apposita funzione di *SGR*. In caso di dubbi nella risoluzione dei casi erronei evidenziati, l'operatore di censimento doveva ricontrattare l'unità di rilevazione. Il *check micro* poteva essere lanciato su singoli questionari o su interi gruppi, il risultato dell'elaborazione mostrava per ogni questionario tutti gli Errori e gli Accertamenti che si erano presentati. Nel primo caso l'utente era obbligato a rimuovere gli errori per proseguire il flusso di lavorazione, mentre gli accertamenti potevano essere confermati; in questo caso il sistema non li presentava nuovamente qualora il *check* venisse lanciato più volte (Figura 2.33).

Figura 2.33 - Maschera di accesso al check

Le problematiche affrontate, da un punto di vista tecnico, per l'implementazione dei controlli *check* hanno riguardato in primo luogo la scelta architettonica di dove implementare i controlli. Le opzioni prese in considerazione sono state quelle di implementare i controlli all'interno dell'applicazione web (quindi in linguaggio Java), oppure di implementare i controlli in una procedura PL/SQL di Oracle. La scelta è ricaduta sulla seconda, in quanto la perdita di portabilità dell'applicazione era trascurabile se confrontata alle maggiori performance e flessibilità offerte dalla seconda soluzione.

Maggiori dettagli sulla parte di sviluppo delle regole e di gestione del flusso del *check* sono illustrati nella parte relativa alla gestione della base dati (Paragrafo 2.4.2), visto che, come si è detto, è lì che risiedeva il cuore del sistema.

Una volta completato il "*check* micro" era necessario passare alla seconda fase di controllo denominato "*check* macro". Tale fase, era gestita internamente all'Istituto, e i dati sottoposti alla lavorazione non potevano in alcun modo essere lavorati dalla rete fino a quando non venivano rilasciati per ulteriori lavorazioni. Le procedure di controllo e correzione erano basate su metodi probabilistici di individuazione degli *outlier* e avevano quindi bisogno, per risultare efficaci, di un numero consistente di questionari da sottoporre a verifica. Questa fase poteva dunque essere avviata solo se erano pronti i dati di tutte le aziende assegnate ad un singolo CIC (circa duemila) con controlli di "*check* micro" superati. Per questo tipo di controlli SGR forniva blocchi di dati ad un'applicazione interna all'Istituto che veniva eseguita automaticamente e in modo completamente asincrono rispetto all'applicazione web. I risultati delle elaborazioni, ovvero le anomalie rilevate per singolo questionario, venivano poi pubblicati in SGR, per permettere agli organi di rilevazione di confermare l'eventuale anomalia oppure correggere il dato anomalo, dopo le opportune verifiche sul campo. Il protocollo di scambio dati, che permetteva il passaggio dal database di SGR a quello disposto per il controllo e correzione, è illustrato nei paragrafi successivi.

2.3.5 Il monitoraggio: i Rapporti riassuntivi

I **Rapporti riassuntivi** rappresentano lo strumento che ha permesso alla rete un attento e puntuale monitoraggio della rilevazione. Nella macro area a essi dedicata gli utenti hanno avuto a disposizione report di sintesi che permettevano di analizzare diversi aspetti quali: l'andamento e l'esito della rilevazione, il lavoro degli operatori, i risultati della fase di controllo e correzione dei dati, l'ubicazione del centro aziendale. Tutti i



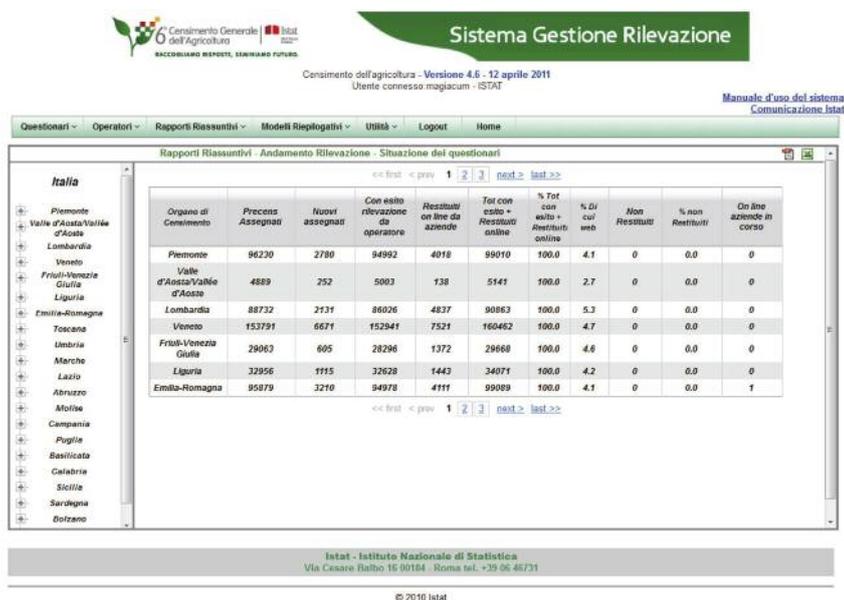
report presenti nel sistema hanno fornito informazioni in tempo reale, in quanto effettuavano un'interrogazione sul database nel momento in cui l'utente richiedeva la visualizzazione dei dati.

Alcuni rapporti riassuntivi hanno assunto anche il ruolo di certificazione della chiusura della attività censuarie per la corresponsione dei contributi spettanti agli organi di censimento. In particolare secondo quanto stabilito nella circolare numero 11 del 21/04/2011 - relativa alla chiusura delle attività censuarie - gli Uffici di censimento dovevano stampare il "Riepilogo territoriale dei principali risultati provvisori", sia per utente connesso sia per ubicazione del centro aziendale, nonché il "Numero di aziende per archivio di provenienza".

I rapporti riassuntivi erano a disposizione degli organi di censimento, dei loro operatori e degli Uffici regionali dell'Istat, relativamente alle unità di rilevazione di propria competenza, e di Istat centrale per l'intero territorio nazionale.

I riepiloghi erano navigabili con tecnica di navigazione a cascata (*drill down*) che permetteva di visualizzare i dati a livelli territoriali di dettaglio sempre maggiore (Figura 2.34). Il territorio non corrispondeva necessariamente ai confini amministrativi, ma alle aggregazioni individuate da ciascuna Regione nel proprio piano di censimento. Tutti i rapporti riassuntivi hanno avuto la medesima struttura con possibilità di esportazione in formato Excel e PDF per ogni livello territoriale visualizzabile.

Figura 2.34 - Rapporto riassuntivo



Andamento della rilevazione

Per monitorare l'andamento della rilevazione, SGR forniva i seguenti tre report di sintesi, ognuno dei quali focalizzava l'attenzione su un aspetto specifico:

- **Situazione questionari:** riepilogando l'andamento complessivo della restituzione dei questionari, mostrava il numero di questionari in lista e le nuove aziende create e riportava le informazioni in valore assoluto e percentuale relative al:
 1. numero di compilazioni on-line del questionario;
 2. numero di questionari cartacei;

3. numero dei non restituiti;
 4. numero delle compilazioni on-line non ancora inviate definitivamente;
- **Stato di lavorazione dei questionari restituiti:** riportava informazioni aggregate sullo stato di lavorazione dei questionari gestiti dalla regione e dalle aree sub-regionali UTC e UCC, se presenti, secondo i diversi stati di lavorazione fino alla fine della prima fase di *check* (controlli micro); in particolare erano visualizzati il numero di questionari nei seguenti stati:
 1. questionari senza *data entry*, per i quali non era iniziato o era ancora in corso l'inserimento dei dati da parte degli operatori ed i questionari compilati on-line con salvataggio parziale;
 2. questionari con *data entry* senza *check*, per i quali era avvenuto l'inserimento delle notizie riportate nel questionario con invio definitivo (attraverso la funzione *data entry*) da parte degli operatori o delle aziende che compilavano on-line e per i quali il *check* non era ancora stato lanciato;
 3. questionari in corso di *check*, per i quali erano rilevate incompatibilità (errori o accertamenti) non ancora risolte;
 4. questionari con *check* ok, per i quali era avvenuto l'inserimento delle notizie riportate sul questionario con invio definitivo (attraverso la funzione *data entry*) e il *check* era andato a buon fine;
 - **Stato di lavorazione dei questionari con *check* micro OK:** mostrava lo stato di lavorazione dei questionari che avevano superato la prima fase di controlli (controlli micro) ed erano stati sottoposti alla fase 2 dei controlli (controlli macro); in particolare forniva informazioni sul numero di questionari nei seguenti stati:
 1. senza anomalie dopo controlli macro, corrispondente al numero di questionari privi di valori da accertare dopo i controlli macro;
 2. con anomalie dopo controlli macro, i questionari per i quali i controlli macro evidenziavano anomalie da verificare ed eventualmente correggere;
 3. in corso di modifica dopo controlli macro, corrispondente ai questionari per i quali era avvenuta la modifica delle notizie riportate nel questionario a seguito delle anomalie riscontrate attraverso i controlli macro effettuati dall'Istat;
 4. in corso di *check* dopo controlli macro, corrispondente ai questionari modificati a fronte delle anomalie segnalate, per i quali il *check* aveva rilevato incompatibilità non ancora risolte;
 5. con *check* ok dopo controlli macro, per i quali era avvenuta la modifica del questionario a fronte delle anomalie ed il *check* era andato a buon fine;
 6. questionari confermati, corrispondente ai questionari con anomalie dopo controlli macro convalidati senza aver effettuato alcuna modifica.

Esito della rilevazione

Un secondo gruppo di rapporti poneva l'attenzione sull'esito della rilevazione fornendo informazioni sulle aziende rilevate, in lista non rilevate, in lista non esistenti o doppioni. Essendo tale aspetto fondamentale per la riuscita delle operazioni censuarie, lo stesso report è stato predisposto rapportando l'esito ad altre informazioni disponibili in Istat. Erano quindi presenti i seguenti rapporti: esito per archivi di provenienza (Figura 2.35), esito suddiviso per liste speciali (Figura 2.36), esito per eleggibilità e per classe di superficie totale. Questo gruppo di rapporti ha supportato la rete nell'individuazione di eventuali problemi di lavorazione dei questionari nelle diverse aree territoriali permettendo all'Istat e agli organi di censimento di intervenire in modo tempestivo.



Figura 2.35 - Esiti della rilevazione - Numero Aziende per Archivi di Provenienza

Censimento dell'Agricoltura - Versione 4.6 - 12 aprile 2011
 Utente connesso: magiacum - ISTAT

Manuale d'uso del sistema
 Comunicazione Istat

Questionari - Operatori - Rapporti Riassuntivi - Modelli Riepilogativi - Utilità - Logout - Home

Rapporti Riassuntivi - Esiti - Archivi Provenienza

ESITO DELLA RILEVAZIONE	ARCHIVI								Iste speciali	
	TOTALE	Agea09	Allev09	Reddit09	Cciaa08	Iva07	Catasto09	Asia07		Cens2000
B.1 azienda rilevata	1706330	145122	260872	1137106	842111	649187	1342909	1134276	850483	252078
B.2 azienda in lista non rilevata	74703	45760	9530	35137	33512	22709	48856	41684	25425	5349
a. irreperibilità del conduttore	50430	29541	6338	24252	23892	16198	33515	28989	14233	3682
b. rifiuto	2679	1674	406	1444	1193	913	2013	1513	867	177
c. altra motivazione	21594	14545	2786	9441	8427	5598	13328	11182	10325	1490
B.3 azienda in lista non esistente o doppia	355789	201596	74251	182645	146360	105192	221933	189685	133886	41451
d. Terreni destinati a soli orti familiari o allevamenti per autoconsumo o aziende esclusivamente forestali	50865	23827	15847	26556	13389	9976	36350	19508	15937	3962
e. Soggetto che non ha mai esercitato attività agricola	56763	11783	16518	17602	22333	20985	23603	26076	4550	8159
f. Terreni agricoli definitivamente abbandonati o aziende esclusivamente zootecniche, che hanno totalmente dismesso l'attività senza cessione ad altri	46057	22286	8415	23880	17579	12012	30353	23440	15825	3879
g. Azienda agricola interamente affittata, ceduta, assorbita, fusa o smembrata	161160	123726	20918	95780	76640	52466	108008	102612	85484	19738
h. Unità da ricondurre ad azienda esistente (doppione)	40944	19974	12553	18827	16419	9753	23619	18049	12090	5713
TOTALE	2136822	169857	344653	1354888	1021983	777088	1613698	1365645	1009794	298878

Figura 2.36 - Esiti della Rilevazione - Numero di aziende per lista speciale

Questionari - Operatori - Rapporti Riassuntivi - Modelli Riepilogativi - Utilità - Logout - Home

Rapporti Riassuntivi - Esiti - Numero di aziende per lista speciale

ESITO DELLA RILEVAZIONE	TOTALE	LISTE SPECIALI								
		Istituzioni pubbliche (s13)	Istituzioni non profit	agriturism	prodotti Dop e Igp	vitivinicoli agea (viviv08)	Indagine ortofrutta della Campania (vivaC)	Spa 07	Common land (usi civici)	prodotti vegetali in reddito (quadro RD)
B.1 azienda rilevata	112957	256	143	561	63	6217	2	1796	64	0
B.2 azienda in lista non rilevata	15916	24	53	139	7	801	0	159	12	4
a. irreperibilità del conduttore	12484	8	42	119	6	663	0	110	6	3
b. rifiuto	869	2	2	6	0	38	0	12	0	0
c. altra motivazione	2563	14	9	14	1	100	0	37	6	1
B.3 azienda in lista non esistente o doppia	24509	164	147	179	14	878	1	311	59	2
d. Terreni destinati a soli orti familiari o allevamenti per autoconsumo o aziende esclusivamente forestali	5782	33	29	6	0	221	0	44	8	0
e. Soggetto che non ha mai esercitato attività agricola	4018	106	79	36	3	66	0	36	11	0
f. Terreni agricoli definitivamente abbandonati o aziende esclusivamente zootecniche, che hanno totalmente dismesso l'attività senza cessione ad altri	3772	13	20	11	0	175	0	47	2	0
g. Azienda agricola interamente affittata, ceduta, assorbita, fusa o smembrata	8487	0	10	60	10	328	1	134	6	2
h. Unità da ricondurre ad azienda esistente (doppione)	2450	4	9	66	1	88	0	50	32	0
TOTALE	153382	444	343	879	84	7896	3	2266	135	6

Lavoro Operatori

I rapporti di monitoraggio del lavoro degli operatori della rete sono stati suddivisi per loro tipologia. In particolare per i rilevatori e per i coordinatori intercomunali di censimento erano disponibili in SGR rapporti sulla "situazione dei questionari", sullo "stato

di lavorazione questionari”, sui “questionari con check ok”, tutti aventi la medesima struttura degli omonimi rapporti relativa all’andamento della rilevazione ma con dati riferiti sempre alle aziende assegnate a ciascun rilevatore o coordinatore, indipendentemente dalla persona che aveva eseguito materialmente le varie funzioni.

Ubicazione del centro aziendale

Questo rapporto conteneva dati organizzati per ubicazione del centro aziendale e non per residenza del conduttore, secondo lo schema riportato in Figura 2.37. Era così disponibile il riepilogo dello stato di avanzamento della rilevazione per singolo organo di censimento e per localizzazione del centro aziendale:

- questionari pre-censuari assegnati inizialmente alla Regione o Provincia autonoma;
- questionari totali con *check* ok per i quali era avvenuto l’inserimento delle notizie riportate sul questionario (attraverso la funzione *data entry*) e il *check* era andato a buon fine e relativa superficie totale (in Are);
- questionari lavorati da altri URC con *check* ok e con centro aziendale nella Regione, corrispondente al numero di questionari lavorati dagli altri URC che avevano il centro aziendale nella regione della riga di pertinenza (= questionari acquisiti da altri) e relativa superficie totale (in Are);
- questionari con *check* ok e con centro aziendale in altra Regione, corrispondente al numero di questionari lavorati dall’URC che avevano il centro aziendale in altre regioni (=questionari ceduti ad altri) e relativa superficie totale (in Are);
- questionari totali con centro aziendale nella Regione e relativa superficie totale (in Are).

Figura 2.37 - Ubicazione Centro Aziendale - Esito B1 e check macro completato

The screenshot shows the 'Sistema Gestione Rilevazione' interface. At the top, there is a navigation bar with 'Questionari', 'Operatori', 'Rapporti Riassuntivi', 'Modelli Riepilogativi', 'Utilità', 'Logout', and 'Home'. Below this, a header indicates 'Censimento dell'agricoltura - Versione 4.6 - 12 aprile 2011' and 'Utente connesso: magiacum - ISTAT'. The main content area displays a table titled 'Rapporti Riassuntivi - Ubicazione Centro Aziendale - Esito B1 e check macro completato'. The table has columns for 'Organo di Censimento', 'Questionari Pre-cens. Assegnati', 'Questionari totali', 'Questionari lavorati da altri URC con centro aziendale nella Regione', 'Questionari con centro aziendale in altra Regione', and 'Questionari totali con centro aziendale nella Regione'. Each of these columns is further divided into 'Questionari' and 'Superficie Totale(Are)'. The data rows list regions: Piemonte, Bolzano, Trento, Lombardia, Veneto, and Friuli-Venezia Giulia. Navigation controls like '<< first', '< prev', '1', '2', '3', '4', 'next >', and '>> last' are visible above and below the table. At the bottom, the footer includes 'Istat - Istituto Nazionale di Statistica' and '© 2010 Istat'.

Organo di Censimento	Questionari Pre-cens. Assegnati	Questionari totali		Questionari lavorati da altri URC con centro aziendale nella Regione		Questionari con centro aziendale in altra Regione		Questionari totali con centro aziendale nella Regione	
		(Questionari lavorati dall'organo di censimento)		(Questionari da acquisire)		(Questionari da cedere)			
		Questionari	Superficie Totale(Are)	Questionari	Superficie Totale(Are)	Questionari	Superficie Totale(Are)	Questionari	Superficie Totale(Are)
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)
Piemonte	96230	70647	13096876	194	272253	66	119898	70775	13112116
Bolzano	25661	9	5145	7	771	9	5145	7	771
Trento	22557	12	4515	15	6404	9	2540	18	8379
Lombardia	88732	0	0	437	218286	0	0	437	218286
Veneto	153791	0	0	117	102931	0	0	117	102931
Friuli-Venezia Giulia	29063	23350	28362334	27	5990	24	38954	23353	28329370

2.3.6 Modelli riepilogativi e Scarico microdati

La sezione del menù di SGR dedicata ai **Modelli riepilogativi** è stata largamente utilizzata dagli operatori della rete in quanto ha consentito di effettuare la *data entry* e la



visualizzazione delle variabili primarie⁷ necessarie per la diffusione dei dati provvisori. La funzione di data entry era disponibile agli Uffici comunali di censimento delle sole regioni a Partecipazione integrativa e al Molise, unica regione ad Alta partecipazione che ha scelto la registrazione *in service* dei questionari. Le regioni ad alta partecipazione, invece, potevano solo effettuare la visualizzazione dei modelli riepilogativi, in quanto gli stessi erano compilati automaticamente dal sistema dopo che il questionario era stato compilato ed aveva superato la fase di *check* micro.

L'unica operazione preliminare al data entry dei modelli riepilogativi era la registrazione dell'esito della rilevazione. L'operatore della rete in fase di compilazione doveva confermare residenza e ubicazione dell'azienda e specificare la consistenza delle principali coltivazioni ed allevamenti

Nello stesso sotto-menù, la funzione di Visualizzazione permetteva di visualizzare i dati inseriti, previa specificazione dell'identificativo dell'azienda (Figura 2.38).

Il sottomenù relativo alla generazione dinamica dei Modelli riepilogativi (*Riepilogo Territoriale - Visualizzazione*) permetteva di ottenere un'ampia reportistica per tutte le Regioni e le Province autonome ai vari livelli territoriali: regione, provincia, comune oppure area sub-regionale UTC o area sub-regionale UCC (se presenti) e per residenza del conduttore o ubicazione del centro aziendale (Figura 2.39).

Cliccando sull'apposita funzione il sistema proponeva una maschera che consentiva di selezionare il tipo di prospetto desiderato e di impostare i filtri territoriali per le query di estrazione dei dati.

Figura 2.38 - Visualizzazione variabili primarie inserite dall'operatore

RIEPILOGO QUESTIONARIO: DATA ENTRY



Numero identificativo Istat : 1134343 Denominazione azienda : Azienda Agricola 1

RESIDENZA DEL CONDUTTORE

Regione: Sicilia
 Provincia: Siracusa
 Comune:

UBICAZIONE DEL CENTRO AZIENDALE

Ubicato nello stesso comune del conduttore
 Nell'annata agraria 2009/2010 l'unità agricola è stata: Attiva

AZIENDE ATTIVE INTERVISTATE

con COLTIVAZIONI (Sez. II - Utilizzazione dei terreni)

	Superficie (ettari)	Superficie (are)
Seminativi	0	0
Coltivazioni legnose agrarie	0	86
di cui Vite	0	0
Orti familiari	0	6
Prati permanenti e pascoli	0	0
Prati permanenti e pascoli non più destinati alla produzione	0	0
SAU	0	92
SUPERFICIE TOTALE	0	98

con ALLEVAMENTI (Sez. III - Consistenza degli allevamenti)

IN COMPLESSO	CAPİ (numero)
Bovini	0
Bufalini	0
Equini	0
Ovini	0
Caprini	0
Suini	0
Avicoli	0
Conigli	0
Struzzi	0
Altri allevamenti	Non presenti

7 Cfr. nota 4.

Al termine dell'elaborazione veniva generato un report come quello in Figura 2.39, il cui livello territoriale è quello *comunale*: questi prospetti potevano essere esportati anche nei formati Excel e PDF in modo da poter essere elaborati anche localmente sul proprio personal computer. Dalle risultanze e dai feedback ricevuti, queste funzioni sono state di grande comodità per la rete di rilevazione.

La funzionalità di Scarico microdati, presente nel menù Utilità di SGR, era riservata esclusivamente ai responsabili regionali di Censimento delle regioni ad alta partecipazione che avevano registrato direttamente i questionari cartacei. Essa consentiva di scaricare sul proprio computer i dati elementari relativi alle aziende di propria competenza territoriale, in formato testuale csv, con separatore “;”. In questo modo le regioni ad alta partecipazione hanno avuto accesso ai microdati delle unità di propria competenza, prima della pubblicazione dei dati provvisori.

Una prima maschera permetteva di selezionare non solo l'ambito territoriale, ma anche lo stato dei questionari per i quali si volevano ottenere i dati. Quest'ultima opzione ha permesso di ridurre la dimensione dei *dataset* generati creando *files* specifici, in cui le aziende potevano essere suddivise per “stato” (Figura 2.40). Non era prevista alcuna limitazione al numero degli scarichi che un operatore poteva effettuare.

Figura 2.39 - Riepilogo territoriale dinamico

Riepilogo territoriale dei principali risultati provvisori
 Per Ubicazione Del Centro Aziendale
 Competenza Per Centro Aziendale Dell'utente
 Regione: Emilia-Romagna
 Provincia: Parma
 Comune:

3. AZIENDE RILEVATE NEL COMPLESSO CON CHECK OK	Aziende(num)	Superficie (ettari)	Superficie (are)
3.1 AZIENDE TEMPORANEAMENTE INATTIVE	1		
3.2 AZIENDE ATTIVE	108		
di cui			
3.2.1 con COLTIVAZIONI (Sez. II - Utilizzazione dei terreni)			
Seminativi	87	1528	89
Coltivazioni legnose agrarie	8	11	55
di cui Vite	3	8	23
Orti familiari	4	13	94
Prati permanenti e pascoli	37	407	50
Prati permanenti e pascoli non più destinati alla produzione	8	53	46
SAU	107	2015	34
SUPERFICIE TOTALE	108	2965	14

3.2.2 con ALLEVAMENTI (Sez. III - Consistenza degli allevamenti)	AZIENDE (numero)	CAPİ (numero)
IN COMPLESSO	33	
Bovini	19	1293
Bufalini	0	0
Equini	8	40
Ovini	3	440
Caprini	3	116
Suini	2	828
Avicoli	1	35000
Conigli	0	0
Struzzi	0	0
Altri allevamenti	2	

La procedura di estrazione, scritta in linguaggio PL/SQL, veniva eseguita in background sul database Oracle del Censimento, permettendo all'utente di continuare ad operare col sistema SGR senza dover rimanere sulla schermata in attesa. L'operatore connesso, infatti, aveva la possibilità di vedere, in una schermata apposita, successiva a quella di impostazione dei filtri di estrazione, lo status di tutte le sue richieste pregresse.



Il download dei dati, di conseguenza, poteva avvenire in qualsiasi momento successivo al completamento della procedura di estrazione, rimanendo attivi gli scarichi fino al loro “consumo” da parte dell’utente, dopo il quale venivano cancellati anche dal sistema.

Figura 2.40 - Pannello di controllo per lo scarico dei microdati

Scarico dei microdati

Regione*

Provincia

Comune

Stato del questionario

- Salvataggio parziale Web
- Invio via web
- Data entry Effettuato
- Inserimento modello riepilogo questionario
- Questionari con check ok
- Questionari in corso di check
- Questionari con anomalie o in salvataggio parziale dopo controlli macro
- Questionari confermati
- Questionari con invio definitivo dopo controlli macro
- Questionari in corso di check dopo controlli macro
- Questionari con check ok dopo controlli macro
- Questionari pronti Dati Provvisori
- Questionari senza anomalie dopo controlli macro
- Salvataggio parziale Data entry
- Salvataggio parziale Data entry dopo controlli macro
- Modifica Modello Riepilogo Questionario dopo controlli macro
- Questionario abbandonato
- Intervista svolta per questionario abbandonato

Quando lo stato di esecuzione dell’estrazione era nello stato *terminato*, l’operatore poteva salvare sul proprio personal computer il file CSV generato, cliccando sul bottone *Download* (Figura 2.41).

Figura 2.41 - Maschera di riepilogo degli scarichi microdati per utente

Scarico dei microdati - status

Id Utente	Id Richiesta	Data	Descrizione	Stato esecuzione	Data Limite Download	Errore	File
101002	1	01-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Benevento	Terminato	01-Jul-2010 00:00:00	ORA-00923: query failed	Download
101002	11	01-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Napoli Comune : Massa di Somma	Terminato	01-Jul-2010 00:00:00		Download
101002	26	05-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Napoli Comune : Massa di Somma	In corso	05-Jul-2010 00:00:00		Download
101002	29	07-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Benevento	Terminato	07-Jul-2010 00:00:00		Download
101002	30	07-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Benevento	Terminato	07-Jul-2010 00:00:00		Download
101002	31	07-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania Provincia : Napoli Comune : Forio	Terminato	07-Jul-2010 00:00:00		Download
101002	32	07-Jul-2010 00:00:00	Scarico richiesto per Regione : Campania	Terminato	07-Jul-2010 00:00:00		Download

2.4 L'infrastruttura di gestione dei dati

La base dati del 6° Censimento generale dell'agricoltura è di tipo relazionale, ossia è formata da un insieme di tabelle legate tra loro attraverso dei campi denominati "campi chiave". Essa risiede su un Data Base Oracle e include tutte le tabelle per la gestione del questionario on-line e di SGR. Questa soluzione è stata scelta perché ha permesso di trattare nel minor tempo possibile e in modo integrato dati inseriti da sottosistemi diversi. Poiché le applicazioni del questionario on line e quelle di SGR lavoravano utilizzando tabelle diverse, nei paragrafi seguenti i due gruppi di tabelle sono stati descritti in modo separato. Viene, inoltre, illustrato il protocollo di comunicazione che ha permesso il passaggio dei dati dal database di gestione a quello di controllo e correzione, interno all'Istituto. Infine, un paragrafo viene dedicato alla fase di *check* micro, che - come già detto - è stata completamente guidata da procedure che risiedono sul *database* e ha una struttura tabellare che ha permesso di governare tutta la lavorazione.

2.4.1 La base dati del questionario elettronico

Le tabelle del questionario on-line sono state raggruppate in strutture-dati con delle caratteristiche in comune e hanno permesso di trattare i diversi aspetti della rilevazione.

Tabella territoriali

Per la memorizzazione delle informazioni di decodifica del territorio relative - nello specifico - alle regioni, alle province e ai comuni italiani. Tali dati hanno permesso di riconoscere le province di Trento e Bolzano come territori autonomi, alla stregua di regioni, e ne hanno consentito un trattamento diverso rispetto alle altre. Tutte le tabelle sono state pre-caricate utilizzando le codifiche territoriali standard dell'Istituto.

Tabella di metadati

Per la descrizione delle variabili del questionario elettronico e sono state denominate tutte con il prefisso DEC. Tutte queste tabelle hanno una struttura standard che si compone di 4 colonne: 1 colonna contenente i valori che la variabile del questionario può assumere e 3 colonne di descrizione nelle diverse lingue previste per la compilazione on-line: in italiano, tedesco e sloveno. Per esigenze di sviluppo non tutte le variabili del questionario elettronico sono state codificate all'interno della base dati, di seguito sono elencate le variabili del questionario elettronico codificate nella base dati con il riferimento alle Sezioni nelle quali si trovano:

- nella Sezione 1 - "Notizie generali sull'azienda": la forma giuridica, il sistema di conduzione e lo stato di attività dell'azienda;
- nella Sezione 2 - "Informazione per aziende con terreni": la fonte di approvvigionamento dell'acqua irrigua;
- nella Sezione 5 - "Lavoro ed attività connesse": il sesso, la cittadinanza, la condizione professionale, il tempo dedicato, il settore di attività prevalente, la posizione, il contratto e il titolo di studio;
- nella Sezione 6 - "Altre informazioni": la contabilità e l'autoconsumo.

Utilizzare le tabelle di metadati contenenti tutte le classificazioni presenti nel questionario on-line ha consentito di gestire in modo più dinamico i cambi di struttura del questionario.

Questo gruppo di tabelle è stato pre-caricato con i dati presenti nel modello cartaceo.



Tabelle dei microdati

Per la memorizzazione di tutti i microdati del questionario elettronico. I dati venivano inseriti solo dopo che l'applicazione aveva verificato la correttezza del tipo dato e l'obbligatorietà del campo, in questo modo la base dati risultava sempre coerente con quanto previsto dall'applicazione on-line.

Per rendere più leggibile il database, a ciascuna *tabella dei microdati* è stato assegnato un nome con riferimento a una delle sezioni o dei riquadri del questionario; nello specifico:

- la Sezione “Notizie anagrafiche, residenza o sede legale del conduttore” faceva riferimento ai dati dell'azienda e alla sua ubicazione;
- la Sezione 1 - “Notizie generali sull'azienda” faceva riferimento ai corpi aziendali di terreno, agli elementi del paesaggio agrario, all'informatizzazione dell'azienda e al sostegno allo sviluppo rurale;
- la Sezione 2 - “Informazione per aziende con terreni” faceva riferimento all'utilizzazione dei terreni, alle notizie particolari sulla vite e ai metodi di produzione agricola delle colture;
- la Sezione 3 - “Informazioni per aziende con allevamenti” faceva riferimento alla consistenza degli allevamenti e ai metodi di gestione dell'allevamento;
- la Sezione 4 - “Ubicazione dei terreni e degli allevamenti aziendali” conteneva i dati relativi alla consistenza dei terreni e dei capi di bestiame allevati per comune di effettiva ubicazione;
- la Sezione 5 - “Lavoro ed attività connesse” faceva riferimento alla famiglia del conduttore e parenti, all'altra manodopera aziendale in forma continuativa e saltuaria, ai lavoratori non assunti direttamente dall'azienda, alle notizie sul capo azienda, alle attività remunerative connesse all'azienda, al contoterzismo, alla produzione dei mangimi per impiego in azienda e agli impianti per la produzione di energia rinnovabile;
- la Sezione 6 - “Altre informazioni” conteneva i dati su ricavi e commercializzazione dei prodotti aziendali.

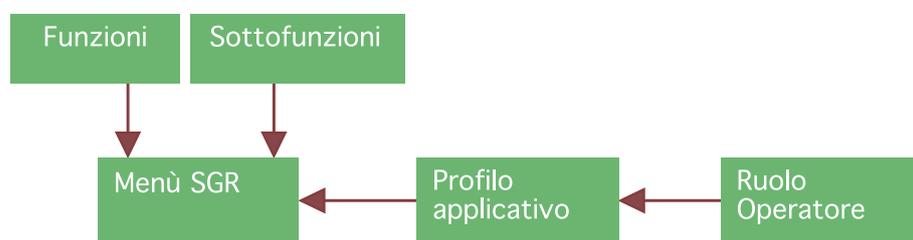
Le tabelle dei microdati venivano popolate esclusivamente con i dati inseriti dall'applicazione web ad eccezione della tabella contenente le notizie anagrafiche, residenza o sede legale del conduttore. Quest'ultima, infatti, è stata popolata da una procedura Oracle con le informazioni provenienti dalla lista pre-censuaria. I dati, visualizzati all'interno dell'applicazione on-line, dovevano essere confermati o modificati dall'azienda al momento della compilazione. La base dati ha conservato entrambe le informazioni, ossia quella del dato presente in lista e quella riferita all'aggiornamento.

Tabelle gestionali

Sono tabelle di servizio create per rendere più dinamica l'applicazione web del questionario.

Di seguito è rappresentato lo schema E/R (Entità/Relazioni) con il dettaglio delle tabelle descritte (Figura 2.42).

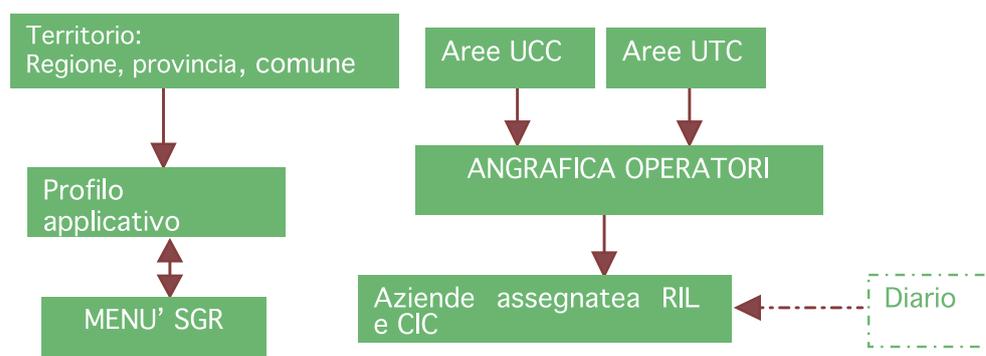
Figura 2.43 - Schema menù dell'applicazione



La rete di rilevazione

La gestione della rete di rilevazione comprendeva le informazioni sul territorio, l'anagrafica degli operatori, le tabelle per la gestione delle assegnazioni. Per tenere conto delle aggregazioni territoriali diverse da quelle amministrative, sono state create delle tabelle contenenti le aree di competenza UCC ed UTC. Inoltre, è stata rappresentata sia la gerarchia interna della rete di rilevazione, sia l'assegnazione delle singole aziende ai Rilevatori e ai Coordinatori intercomunali. Le informazioni relative a territorio, alle aree UCC ed UTC e l'anagrafica degli operatori responsabili sono state comunicate all'Istat dalla rete prima dell'inizio del censimento e pre-caricate nel sistema dai responsabili applicativi della base di dati. Ogni operatore era legato ad un profilo applicativo e attraverso le tabelle di gestione del menù SGR, aveva accesso alle sole funzioni autorizzate (Figura 2.44).

Figura 2.44 - Schema rete di rilevazione



Gestione del Diario

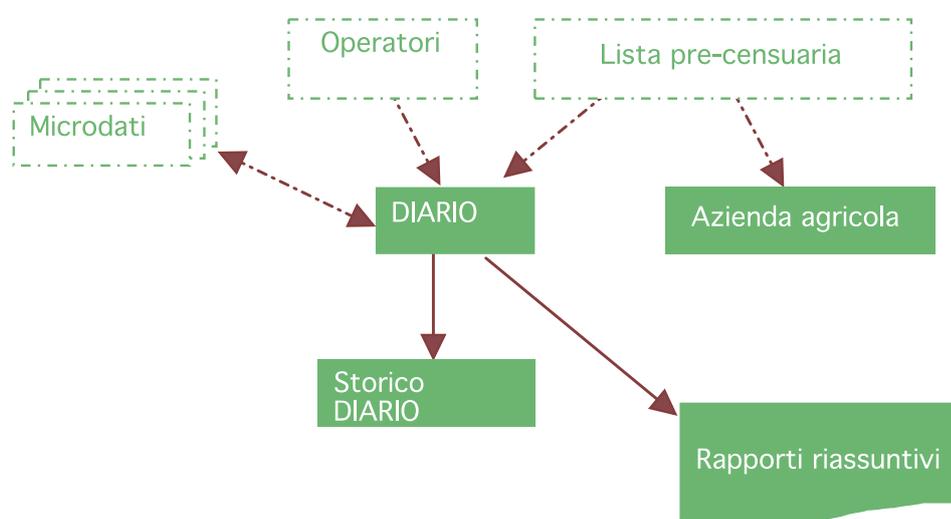
I cambiamenti degli stati dei questionari che tracciavano il flusso di lavorazione di ogni singolo questionario, nonché l'utente che aveva effettuato l'operazione, sono stati registrati in una tabella denominata "Diario". Accanto a tali informazioni, nella tabella sono memorizzate anche le informazioni anagrafiche delle aziende. Tale tabella, che rappresentava il cuore del sistema, è stata pre-caricata con le informazioni anagrafiche derivate dalla lista pre-censuaria e con alcuni dati sull'azienda provenienti da fonti ausiliarie necessarie per attivare alcuni controlli nel momento in cui veniva effettuata la registrazione dell'esito.

Successivamente, l'applicazione web aggiornava il "Diario" mantenendo però sempre memoria della situazione pre-censuaria. Tutte le informazioni salvate sono state fon-

damentali, insieme con quelle degli operatori, per la produzione dei rapporti riassuntivi e per tenere sotto controllo l'andamento della rilevazione.

La tabella "Diario" conteneva un record per ogni questionario, sia della lista pre-censuaria, sia creato dal sistema nel corso della rilevazione. Vista la sua importanza, la tabella veniva storicizzata tramite dei *trigger* che inserivano un nuovo record sulla tabella dello storico ad ogni cambiamento dello stato. Tale storicizzazione era utilizzata dall'applicazione sia per documentare il lavoro fatto su un questionario, sia per ripristinare i dati precedenti in caso di necessità (Figura 2.45).

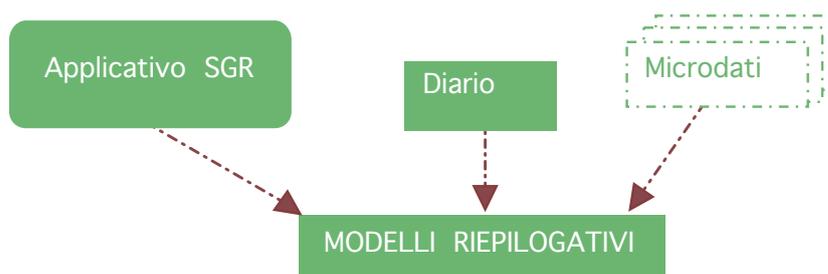
Figura 2.45 - Schema diario



Modelli riepilogativi

Una parte della base di dati SGR è stata dedicata alla registrazione delle variabili primarie, necessarie alla diffusione dei dati provvisori. Queste informazioni venivano registrate nella base dati attraverso due modalità: per le regioni che avevano adottato il modello a partecipazione integrativa e per il Molise con registrazione dei dati *in service* mediante la compilazione manuale di un modello *short-form*; per le altre regioni avveniva in maniera automatica, attraverso una procedura Oracle che le estraeva direttamente dai questionari compilati on-line dalle aziende o dagli operatori (Figura 2.46).

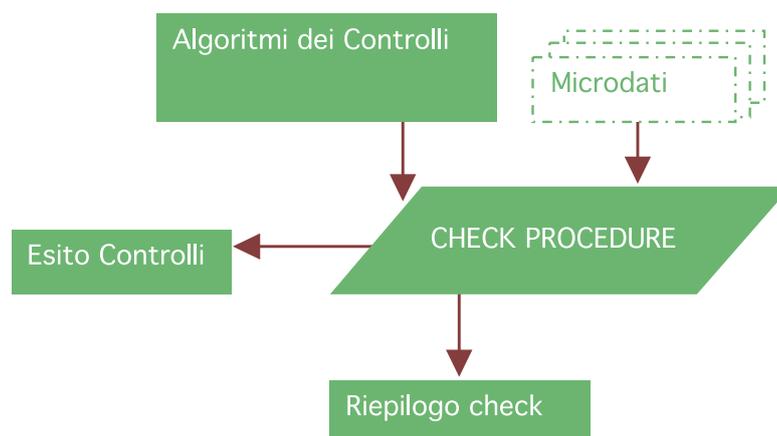
Figura 2.46 - Schema modelli riepilogativi



Check micro

Il processo veniva gestito attraverso una procedura Oracle che leggeva direttamente da tabella gli algoritmi di controllo registrando l'esito del processo in modo da mantenere il legame tra dettaglio degli errori o accertamenti e l'operatore coinvolto (Figura 2.47).

Figura 2.47 - Schema check micro



Database trigger e procedure

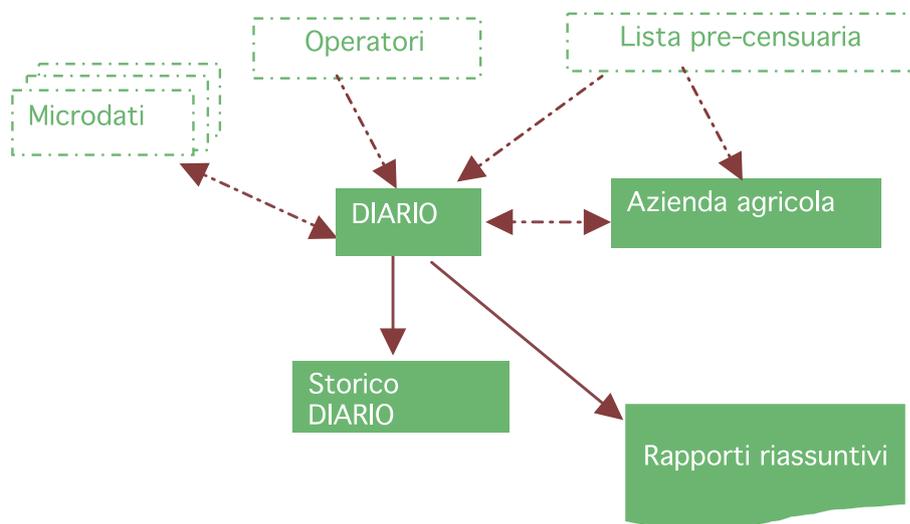
La scelta dell'utilizzo dei trigger di *database* è stata dettata dall'esigenza di centralizzare e automatizzare alcune operazioni chiave del sistema. Mediante il loro impiego in SGR è stato possibile:

- alimentare le tabelle dedicate alla storicizzazione delle operazioni di gestione del diario;
- impostare gli stati di lavorazione dei questionari gestendo il colloquio tra SGR e *data entry* durante le fasi di compilazione e registrazione dell'esito dei questionari da parte dei componenti della rete di rilevazione;
- aggiornare, durante le fasi di assegnazione/de-assegnazione delle aziende agli operatori, alcune tabelle riepilogative di ausilio ai calcoli dei rapporti riassuntivi.

Le *Oracle procedure*, richiamate dall'applicativo, sono state sviluppate con l'obiettivo di ottimizzare la realizzazione delle funzioni di (Figura 2.48):

- scarico dei microdati dei questionari completi della parte anagrafica;
- calcolo dei rapporti riassuntivi per il monitoraggio dello stato di lavorazione dei questionari, dell'attività di Rilevatori e coordinatori intercomunali, degli esiti per classe di superficie totale da lista pre-censuaria, della situazione dei questionari per centro aziendale;
- "check micro" SGR.

Figura 2.48 - Trigger che governano il sistema



2.4.3 Colloquio con i sistemi di controllo e correzione dati

Il protocollo per il colloquio tra SGR e il Sistema di Controllo e Correzione Dati (SCCD), esterno al primo, avveniva tramite un *workflow* basato su tabelle residenti in un'area di transizione. In questa area sono presenti le tabelle dei microdati e una tabella degli stati di lavorazione (STATO_SGR_SCCD), per il passaggio dati riferiti ad un singolo CIC (Coordinatore Intercomunale) o di tutte le aziende di una singola Provincia. La variabile "codice di stato" ha gestito totalmente il protocollo di passaggio dati e ha definito lo stato di lavorazione dei questionari.

Tale codice assumeva i seguenti valori distinti per tipo di operazione:

- avvenuto caricamento, da parte di SGR, di tutti i record relativi ad un CiC, nell'area di transizione;
- avvenuto caricamento, da parte di SCCD, di tutti i record relativi ad un CiC, nella propria area e cancellazione degli stessi, sempre da parte di SCCD, nell'area di transizione;
- avvenuto caricamento, da parte di SCCD, di tutti i report relativi ad un CiC, nell'area di transizione;
- avvenuta acquisizione di tutti i report relativi ad un CiC, da parte di SGR, e cancellazione degli stessi dall'area di transizione, sempre da parte di SGR;
- avvenuto caricamento, da parte di SGR, dei record relativi a un'intera Provincia, nell'area di transizione;
- avvenuto caricamento, da parte di SCCD, dei record relativi a un'intera Provincia, nella propria area e cancellazione degli stessi, sempre da parte di SCCD, nell'area di transizione;
- avvenuto caricamento, da parte di SCCD, dei report relativi a un'intera Provincia nell'area di transizione;
- avvenuta acquisizione, da parte di SGR dei report relativi a un'intera Provincia nell'area di transizione;
- avvenuto caricamento, da parte di SCCD dei record relativi a un'intera Provincia nell'area di transizione;



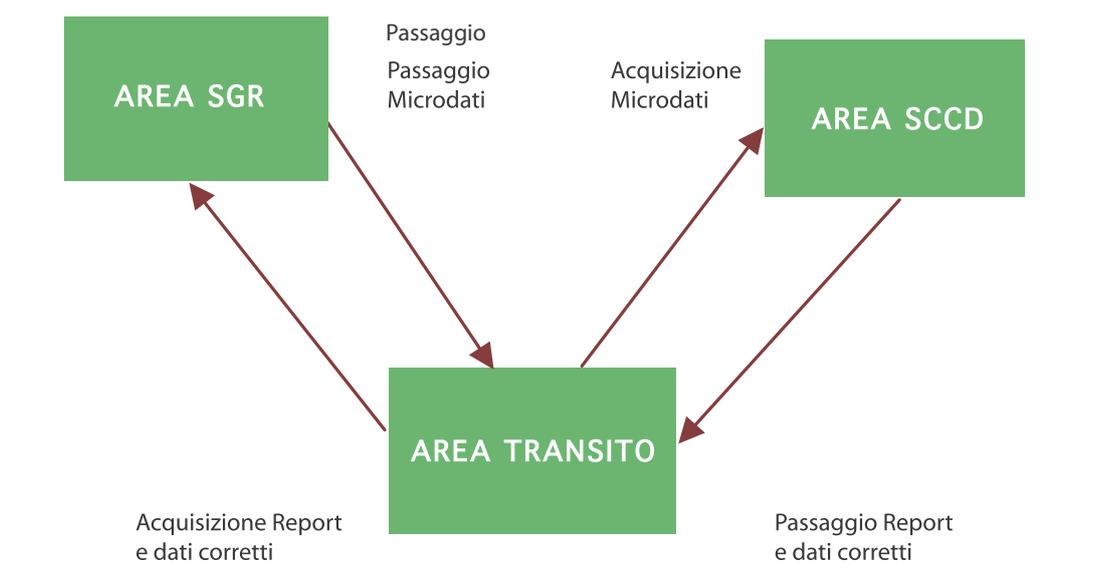
- avvenuta acquisizione da parte di SGR dei record relativi a un'intera Provincia nella propria area e cancellazione degli stessi dall'area di transizione.

Tavola 2.2 - Tracciato record della tabella STATO_SGR_SCCD

NOME_COLONNA	FORMATO	DESCRIZIONE
ENTITA	VARCHAR2(3)	Stringa che identifica se il passaggio è per CIC o Provincia
CODICE	VARCHAR2(9)	Codice CIC o Provincia
STATO	NUMBER	Stato dell'operazione
DATAP	DATE	Data lancio

Di seguito si riporta una rappresentazione grafica del protocollo di passaggio dati tra SGR e SCCD (Figura 2.49).

Figura 2.49 - Protocollo del passaggio dati



2.4.4 La gestione delle regole di controllo

Per gestire le regole di controllo e tenere traccia di tutte le operazioni a esso connesse sono state definite delle strutture tabellari specifiche, integrate nello schema SGR, contenenti le regole sotto forma di procedure informatiche, cioè di *oracle function*. Queste erano corredate con i messaggi di errore previsti in lingua italiana e tedesca.

Ogni volta che la procedura di *check* veniva eseguita tutti i dati coinvolti erano storicizzati in opportune tabelle.

Nel dettaglio, per ogni *check* veniva memorizzato: l'operatore che aveva richiesto l'operazione, la data, l'esito di ogni singolo controllo e un riepilogo del numero di errori o accertamenti evidenziati.

Una *oracle-procedure* gestiva l'intero processo in maniera dinamica, ricevendo in partenza la lista delle aziende da sottoporre a *check*, oltre all'identificativo dell'operatore richiedente. Tale procedura ha rappresentato il cuore del *check* SGR.

Ogni questionario sottoposto a verifica veniva bloccato per impedire l'accesso in *data entry* da parte di un operatore, in modo da non generare incongruenze. Per ogni que-

stionario venivano letti ed eseguiti tutti i controlli previsti secondo il flusso della compilazione del questionario, ad esclusione di quelli soft (accertamenti) eventualmente validati nella fase di *check* precedente; veniva poi registrato ogni esito negativo delle regole. Alla fine della procedura per ogni questionario era calcolato il riepilogo del numero di errori o accertamenti e sbloccato l'accesso al *data entry* solo nel caso di presenza di anomalie da sanare.

La registrazione dell'esito di ogni singola regola permetteva all'operatore la verifica dell'andamento del *check* e la correzione delle anomalie nel questionario o la validazione degli accertamenti a conferma del dato inserito. Una volta validati tutti gli *accertamenti*, in assenza di errori bloccanti, il questionario transitava nello stato di "*check ok*". Viceversa gli errori dovevano essere corretti e il *check* doveva essere nuovamente lanciato fino alla scomparsa di ogni anomalia.

Nel complesso, sono state generate 195 regole, di cui 35 accertamenti e 160 errori, distribuite all'interno delle sezioni del questionario così come riportato in Tavola 2.3. Le sezioni per le quali nelle due fasi di "*check micro*" si sono riscontrati più errori sono quelle relative all'"Ubicazione dei terreni e degli allevamenti aziendali" e alla "Famiglia del conduttore e parenti"; di contro le sezioni in cui è stato rilevato il minor numero di anomalie sono quella sui "Metodi di produzione agricola" e quella relativa alla viticoltura.

Tavola 2.3 - Numero di regole per sezione del questionario

SEZIONE DEL QUESTIONARIO	Regole
A - Notizie anagrafiche, residenza o sede legale del conduttore	11
B - Esito della rilevazione	2
C - Cessioni parziali	1
D - Unità collegate alle aziende in lista	2
E - Ubicazione del centro aziendale	6
Sezione I - Notizie generali sull'azienda	22
Sezione II - Informazioni per aziende con terreni	41
Sezione II - Notizie particolari sulla vite	18
Sezione II - Metodi di produzione agricola	5
Sezione III - Informazioni per aziende con allevamenti	11
Sezione IV - Ubicazione dei terreni e degli allevamenti aziendali	17
Sezione V - Famiglia del conduttore e parenti	31
Sezione V - Altra manodopera aziendale in forma continuativa	10
Sezione V - Altra manodopera aziendale in forma saltuaria e lavoratori non assunti direttamente dall'azienda	9
Sezione V - Notizie sul capoazienda	6
Sezione V - Attività remunerative connesse all'azienda	2
Sezione VI - Altre informazioni	1

La procedura di gestione delle regole di controllo realizzata ha presentato numerosi vantaggi ed innovazioni, come di seguito descritto.

Autonomia: è un concetto completamente indipendente dalle regole di coerenza sui microdati imposte ai questionari, che, esistendo come *oracle function* memorizzate in una struttura tabellare, possono essere modificate, implementate o cancellate, senza la necessità di apportare modifiche alla procedura né al flusso originario del questionario.

Modularità: le regole eseguite per ogni questionario cambiano secondo il flusso seguito nella compilazione, e questo è determinato direttamente dalla definizione delle



regole; i controlli sono classificati in due grandi gruppi, di *validity* e di *constraint*. Per i controlli di *validity* è sufficiente verificare la presenza/assenza delle informazioni in una tabella di riepilogo dei microdati al fine di determinare la correttezza del flusso della compilazione; per i controlli di *constraint* ogni *oracle function* accede puntualmente alla parte di microdati relativa al controllo di congruenza da eseguire.

Riuso del software: questa progettazione strutturata modularmente ha determinato un importante miglioramento delle performance in termini di giornate/uomo necessarie allo sviluppo del software, in quanto - con bassi costi - potrà essere riusato in altri sistemi. Solo gli algoritmi delle regole dovranno essere riscritti di volta in volta, secondo le esigenze dei sistemi di acquisizione.

Monitoraggio: attraverso la storicizzazione dei *check*, eseguiti per ogni azienda, si riesce a preservare le scelte dell'operatore relativamente alle validazioni degli *accertamenti* e a monitorare e verificare il lavoro svolto dai vari operatori.

Capitolo 3

La diffusione dei risultati⁸

3.1 Architettura e componenti generali del sistema di diffusione dei dati

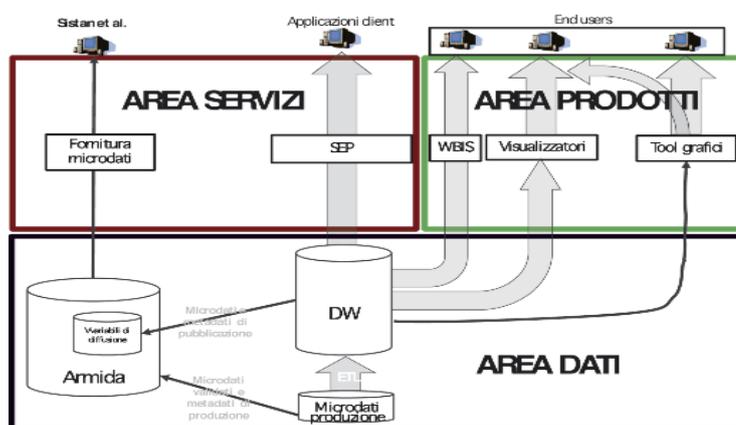
Il perno centrale della diffusione dei risultati definitivi del 6° Censimento generale dell'agricoltura è rappresentato dal *Web data warehouse CensimentoAgricoltura2010*, pagina tematica del data warehouse I.Stat, per la prima volta utilizzato per la diffusione di dati censuari, al quale sono agganciati altri prodotti e servizi informativi innovativi dedicati a utenti più o meno specializzati (*Istat.eXplorer*, *Censimento in breve*, *Single Exit Point*). Completano l'offerta editoriale, le pubblicazioni tradizionali dedicate a un'utenza più generale: un volume cartaceo nazionale che riassume i principali risultati strutturali delle aziende agricole, alcuni volumi tematici, in formato telematico, che approfondiscono aspetti specifici di analisi e gli Atti del 6° Censimento generale dell'agricoltura che documentano le attività svolte per la progettazione e la realizzazione del Censimento stesso.

Il sistema di diffusione dei dati del Censimento dell'agricoltura è stato concepito ponendo alla sua base un quadro architettonico "a geometria variabile" rivolto a diversi livelli e tipi di utenti e costituito da due tipologie di output: *prodotti* e *servizi* informativi (Figura 3.1). Della prima tipologia fanno parte output *pre-confezionati*, rivolti a utenti finali prevalentemente non specialistici, mentre la seconda si riferisce a soluzioni rivolte ad utenti specialistici e all'interscambio e condivisione di dati (*data sharing*) tra applicazioni a livello client, in ottica *machine-to-machine* (prevalentemente in ambito regionale e Sistema statistico nazionale (Sistan)).

In termini generali, l'architettura del sistema di diffusione è quindi caratterizzata da tre componenti:

1. **dati** - microdati di produzione, ARMIDA e componente dati del data warehouse (DW);
2. **servizi informativi di output** - fornitura microdati e Single Exit Point;
3. **prodotti informativi di output su web** - sistema di navigazione di dati aggregati, applicazioni per la visualizzazione di tavole, tool grafici interattivi e dinamici (comprese funzioni di *storytelling*).

Figura 3.1 - Architettura del sistema di diffusione

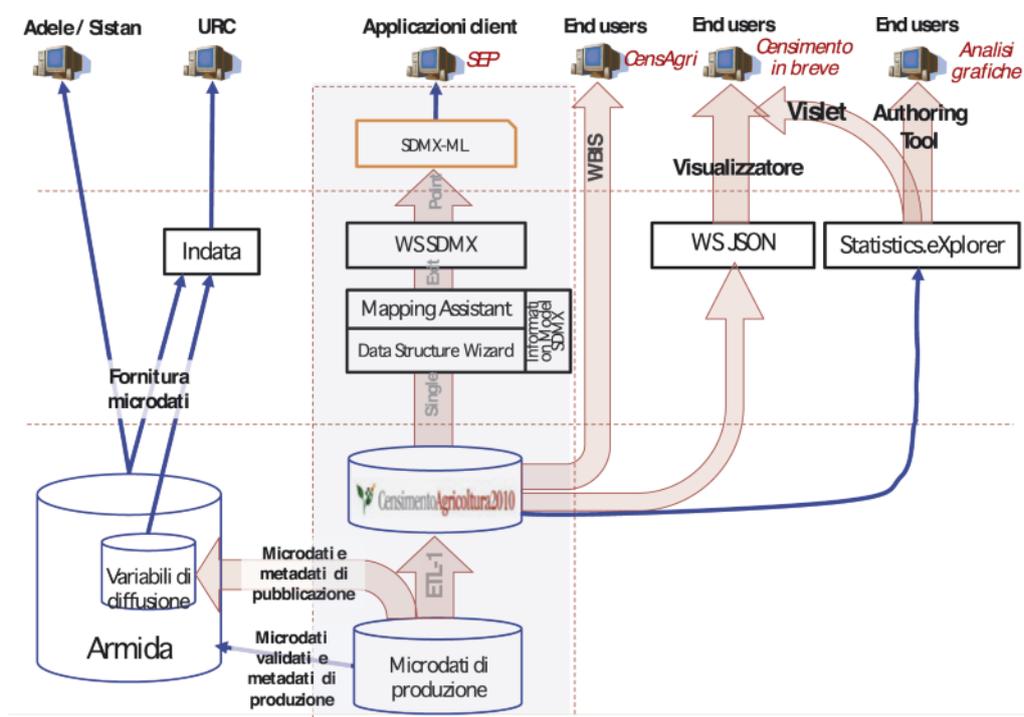


⁸ Autori: M. Greco, F. Lorenzini, S. De Francisci § 3.1; M. Cammarota § 3.2; A. Cardacino § 3.3.3; S. De Francisci §§ 3.3.1, 3.3.2, 3.4.2; S. De Francisci, A. Pitrone § 3.4.1.



Per sostenere questa organizzazione informativa, l'architettura è incentrata su una dorsale composta da tre principali nodi: i microdati (di produzione e/o validati), il DW di diffusione e il *Single Exit Point*. Su queste componenti si innestano tutti i principali prodotti e servizi informativi offerti agli utilizzatori, illustrati in Figura 3.2 e descritti brevemente di seguito:

Figura 3.2 - Prodotti e servizi informativi offerti agli utilizzatori



- **End user** e **Applicazioni client** sono lo strato di front-end del sistema;
- **WBIS** (Web Browser I.Stat) è l'applicazione di front-end su Web di I.Stat;
- **SDMX-ML** rappresenta il formato XML dell'interrogazione via Web Service SDMX⁹ realizzata attraverso il *Single Exit Point* (SEP);
- **SDMX-RI** è la *Reference Infrastructure* di SDMX realizzata da Eurostat e adottata dall'Istat per il SEP. Risulta composta da:
 - **Mapping Assistant**: componente per la mappatura dei dati in formato SDMX;
 - **Data Structure Wizard**: componente per il disegno delle Structure Data Definition (DSD) degli ipercubi provenienti dal DW;
 - **IM SDMX**: Information Model di SDMX;
- **CensimentoAgricoltura2010**, DW del Censimento dell'agricoltura, basato sulla tecnologia DotStat;

⁹ SDMX sta per Statistical Data and Metadata eXchange e rappresenta uno standard basato sul linguaggio XML per lo scambio di dati e metadati statistici. Nato da una iniziativa promossa e incoraggiata da alcune istituzioni internazionali tra le quali la Bank for International Settlements, la Banca centrale europea, l'Ufficio Statistico della Comunità Europea (Eurostat), il Fondo Monetario Internazionale, l'OCSE, le Nazioni Unite (Statistics Division) e la Banca Mondiale, si è ormai affermato come uno standard mondiale per la condivisione di dati statistici aggregati.

- **ETL¹⁰⁻¹**, componente applicativa per la generazione dei CSV dei dati aggregati definiti sulla base dei piani di spoglio del censimento;
- **Microdati di produzione**, strato di back-end che ha costituito la fonte primaria dei dati individuali da aggregare negli ipercubi (dataset di dati aggregati). In questa prima dorsale ci si riferisce a dati di produzione validati.

3.2 La fornitura dei microdati

I file di microdati prodotti dalle rilevazioni effettuate dall'Istat vengono archiviati centralmente nel sistema ARMIDA (**AR**chivio di **MI**cro**DA**ti) per due principali finalità:

- conservare i microdati validati e i metadati delle rilevazioni condotte dall'Istat;
- favorire il riutilizzo dei microdati validati per finalità statistiche da parte di utenti esterni.

In particolare, i microdati archiviati in ARMIDA vengono usati per rispondere alle richieste degli enti del Sistan e sono accessibili presso il Laboratorio per l'Analisi dei Dati ELEMENTARI (Laboratorio ADELE) da parte di ricercatori e studiosi di università, istituti o enti di ricerca.

Per ottenere i file di microdati i richiedenti possono presentare richiesta attraverso il *Cont@ct Centre* dell'Istat, oppure mediante il sistema web per l'acquisizione e il trattamento on-line delle richieste di informazioni statistiche e dei servizi di diffusione (<http://contact.istat.it/>).

Il laboratorio ADELE è un *Research Data Center*, situato presso la sede centrale e gli Uffici Regionali dell'Istat, ovvero un luogo sicuro dove si possono recare ricercatori e studiosi di università, istituti o enti di ricerca per effettuare direttamente analisi statistiche sui microdati. Il servizio è destinato a un'utenza specializzata in grado di individuare la rilevazione statistica di interesse, utilizzare gli strumenti hardware e software messi a disposizione nel laboratorio e interpretare i dati e le elaborazioni realizzate. Gli utenti prelevano i risultati delle elaborazioni realizzate solo in seguito all'applicazione, da parte dei responsabili del laboratorio, dei previsti controlli finalizzati a escludere rischi di violazione della riservatezza.

Come tutti i microdati delle rilevazioni condotte dall'Istat, anche quelli del 6° Censimento generale dell'agricoltura e quelli armonizzati relativi ai Censimenti agricoli dal 1982 al 2010 sono stati archiviati in ARMIDA.

La fornitura dei file di microdati attraverso ARMIDA può essere realizzata in diversi formati: formato fisso (.dat) o delimitato (.txt), Access, Excel 2003, Excel 2007, Xml. Inoltre, la fornitura avviene attraverso il "confezionamento" di un *package* contenente, oltre ai microdati, anche i metadati utili a una corretta interpretazione dei dati: tracciato record, classificazioni, altri allegati. Nel caso del Censimento agricoltura gli allegati sono rappresentati dal questionario di rilevazione, dalla nota metodologica e dal glossario.

Particolarmente rilevante è l'innovazione apportata in occasione dell'archiviazione dei microdati relativi al Censimento dell'agricoltura. In ARMIDA sono state archiviate, oltre ai microdati rilevati tramite il questionario, anche le *variabili derivate* con l'obiettivo di consentire all'utente specializzato di utilizzarle nelle proprie elaborazioni, in modo standardizzato.

Tali variabili sono state individuate analizzando i Piani di spoglio (Pds) degli ipercubi

10 ETL sta per *Extract, Transform, Load* e si riferisce a una particolare tipologia di applicazioni software specializzate per gestire il processo di estrazione, trasformazione e caricamento dei dati da sistemi sorgenti eterogenei, di tipo transazionale, file di testo comuni o altri sistemi informatici in un sistema di data warehousing.



diffusi con I.Stat e generate attraverso apposite procedure SAS. L'archiviazione in ARMIDA ha riguardato le variabili derivate relative a: Caratteristiche delle aziende, Altre caratteristiche, Utilizzazione del terreno delle aziende, Consistenza degli allevamenti, Capo azienda, Manodopera familiare, Altra manodopera aziendale, Contoterzismo delle aziende, Attività remunerative connesse alle aziende.

Le tipologie di variabili derivate individuate e archiviate riguardano:

- variabili originate da funzioni di somma;
- variabili definitorie;
- variabili classificatorie.

Le *variabili originate da funzioni di somma* sono variabili ottenute dalla somma di variabili presenti nel questionario. Ad esempio, per la consistenza degli allevamenti, mentre nel questionario relativo al 6° Censimento generale dell'agricoltura si rileva il numero di capi bovini per maschi e femmine di età inferiore ad 1 anno, tra le variabili derivate è stato considerato anche il numero totale di bovini di età inferiore ad 1 anno (maschi + femmine).

Le *variabili definitorie* sono variabili derivate dalla combinazione di più variabili. Ad esempio, per l'ipercubo di I.Stat relativo al Capo azienda sono state calcolate nuove variabili che indicano la connotazione o meno per la singola azienda delle seguenti caratteristiche:

- azienda con allevamenti e coltivazioni;
- azienda con coltivazioni;
- azienda con allevamenti;
- azienda solo con coltivazioni;
- azienda solo con allevamenti.

La nuova variabile presenta modalità 1=presenza e 0=assenza ed è associata alla singola azienda.

Le *variabili classificatorie* riorganizzano variabili quantitative in classi. Ad esempio, queste si riferiscono alla classe di SAU, alla classe di età del capoazienda, alla classe di giornate di lavoro e contengono le modalità presentate nelle tavole diffuse in I.Stat.

Le circa 230 variabili derivate sono state archiviate in ARMIDA per singola azienda e, nel caso di alcune variabili derivate relative al lavoro, per singolo lavoratore.

3.3. La diffusione dei dati aggregati

3.3.1 Il data warehouse su Web: CensimentoAgricoltura2010

Per la diffusione su Web dei dati aggregati del censimento dell'agricoltura è stata adottata la stessa tecnologia del corporate data warehouse dell'Istat, I.Stat¹¹, implementando un impianto tematico dedicato esclusivamente ai dati censuari.

CensimentoAgricoltura2010 è basato su alcuni concetti fondamentali:

- insiemi di dati statistici aggregati, organizzati in ipercubi multidimensionali (dataset) e disponibili a livello di singolo valore;
- meccanismi di ricerca basati sulle funzionalità dei data warehouse;

11 Il sistema è basato sulla tecnologia del sistema di diffusione dell'OECD (.STAT), piattaforma di software e servizi sviluppata per il rilascio su Web di dati statistici multidimensionali secondo un approccio di data warehouse. Le attività di personalizzazione del sistema sono state condotte in collaborazione con OECD, secondo quanto concordato in un Memorandum of Understanding ("MOU") stipulato tra i due Istituti.

- insiemi di metadati di tipo referenziale¹² disponibili a tutti i livelli informativi, dai dataset fino alle singole celle delle tavole;
- cataloghi di informazioni su aree tematiche, insiemi di dati, dimensioni, modalità delle dimensioni e serie di riferimento per la gestione del sistema.

Il sistema fornisce un ambiente integrato di analisi e diffusione on-line, tramite il quale gli utilizzatori possono accedere ai dati statistici prodotti dal Censimento. Attraverso il browser del sistema è possibile visualizzare i dati statistici aggregati, costruire tavole statistiche personalizzate ed estrarre i dati di proprio interesse per successive elaborazioni.

Dal punto di vista della navigazione il sistema consente di selezionare le statistiche per soggetto, ordinate per voci e organizzate in un albero tematico gerarchico, costruito sulla base delle aree informative presenti nel questionario del censimento, da quelle più generali alle più specifiche. I dati prodotti possono essere visualizzati come tabelle raggruppate in cartelle di lavoro e, attraverso un pannello di comando, se ne può modificare la struttura di tabulazione agendo su: variabili statistiche visualizzate, periodi di riferimento, disposizione degli elementi scelti in ascissa e ordinata e altri parametri di layout, tra cui il mascheramento di linee e colonne prive di dati o la scelta del numero di decimali cui limitare la formulazione dei dati. Il popolamento del sistema avviene attraverso una componente interna di amministrazione e caricamento (*Data Provider Interface*).

Ogni insieme di dati comprende metadati dettagliati per agevolare la comprensione dei contenuti trattati.

All'interno di CensimentoAgricoltura2010 i dati statistici sono organizzati in forma di dataset, strutture di dati aggregati (denominate ipercubi), tramite le quali è possibile rappresentare in un opportuno formato tavole statistiche multidimensionali. Un dataset è costituito da un insieme di valori numerici e dalle rispettive informazioni testuali associate, e comprende tutti i valori che condividono un insieme comune di dimensioni. Ogni dataset è riferito a uno specifico tema o sottotema dell'albero tematico.

Le dimensioni di un dataset corrispondono agli assi lungo i quali sono descritti i dati. Le dimensioni possono essere rappresentate sia come classificazioni semplici sia come classificazioni gerarchiche. Ogni dimensione contiene un elenco predefinito di voci, chiamate modalità della dimensione. Il data warehouse prevede la definizione di un sottoinsieme di dimensioni comuni (quali ad es. territorio, tempo, classi di età, ecc.) che offrono la possibilità di collegare tra loro più dataset. Nella dimensione territorio le modalità sono, ad esempio, l'elenco delle regioni, province, eccetera.

I metadati sono dati qualitativi associati ai dati quantitativi. I metadati possono essere riferiti ai vari livelli del data warehouse, a partire da una descrizione sintetica a livello di intero dataset, oppure si può trattare di note applicabili a singole serie di dati entro un dataset, fino al livello di singola cella di una tavola dati.

CensimentoAgricoltura2010 dispone di un navigatore su Web (*browser*) che rende disponibile agli utilizzatori l'ambiente on-line per accedere alle statistiche del censimento. Il browser è stato realizzato per permettere a tutti gli utenti, sia quelli esperti sia i nuovi utilizzatori, di individuare e recuperare in modo rapido i dati di proprio interesse con i rispettivi metadati.

12 I metadati referenziali descrivono il contenuto e la qualità dei dati statistici [si vada: <http://stats.oecd.org/glossary/detail.asp?ID=7076>]



Le principali caratteristiche del browser di CensimentoAgricoltura2010 includono:

- COLLEGAMENTI ALLE TAVOLE DI DATI PIÙ FREQUENTEMENTE RICHIESTE - accesso alle versioni più recenti delle serie richieste più spesso all'Istat;
- VISTA DI TAVOLA - visualizzazione di dati, flag e metadati di una determinata tavola attraverso l'utilizzo di una serie di funzionalità per la composizione personalizzata dei layout grafici;
- PANIERE DI DATI - possibilità di salvare le interrogazioni (query) di dati per una visualizzazione successiva e di condividerli con altri utilizzatori;
- RICHIESTE COMBinate (MERGED QUERY) - possibilità di unire le interrogazioni per permettere confronti fra dataset che condividono una o più dimensioni comuni;
- METADATI - accesso a tutti i metadati, dal livello del set di dati fino alle note a piè di pagina del dataset;
- RICERCA FULL TEXT - possibilità di individuare in modo rapido serie di dati attraverso dataset, dimensioni e metadati.

3.3.2 Il data warehouse: aspetti tecnologici

L'architettura per la diffusione su Internet dei dati censuari è caratterizzata da una completa scalabilità orizzontale e verticale secondo il modello descritto nella Figura 3.3 e Figura 3.4.

La componente per la navigazione in internet è composta di tre nodi fisici distinti, mentre quella in ambiente intranet è composta da due ambienti, dedicati rispettivamente alle funzionalità di navigazione e a quelle di amministrazione (DPI).

Figura 3.3 - Architettura per la diffusione su internet dei dati censuari

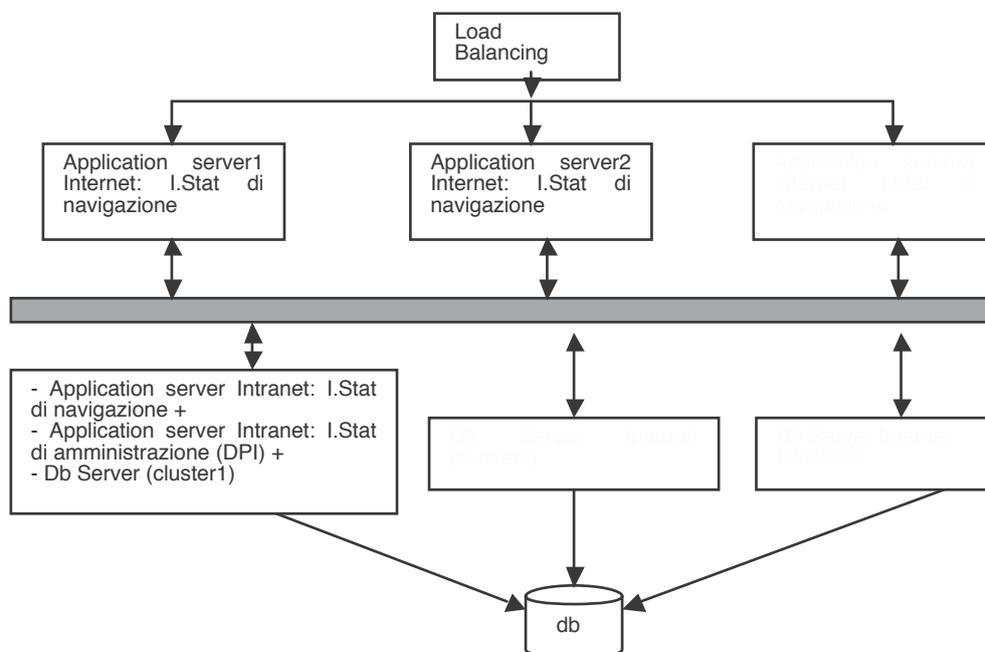
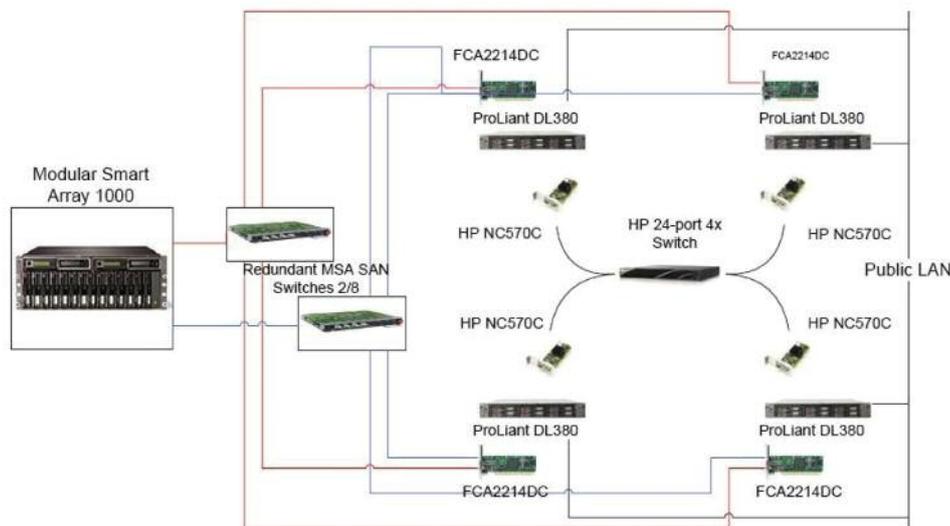


Figura 3.4 - Architettura fisica di rete



3.3.3 Il Single Exit Point e la SDMX Reference Infrastructure

Il SINGLE EXIT POINT del Censimento agricoltura rende possibile la diffusione via web service (*machine-to-machine*) e in formato SDMX dei dati contenuti all'interno del data warehouse del Censimento stesso.

Il sistema si basa sulla medesima architettura applicativa del Single Exit Point per i dati del Corporate data warehouse di Istituto I.Stat: tale architettura è a sua volta basata sulla piattaforma *SDMX Reference Infrastructure* prodotta da Eurostat. La caratteristica principale di tale piattaforma è che, al fine di rendere disponibili i dati in formato SDMX via web service, non si opera alcuna trasformazione preventiva dei dati stessi (che restano nelle basi dati locali), ma essi sono "mappati" sui costrutti SDMX in modo tale che la trasformazione avvenga a *runtime* sulla base di query SDMX sottoposte al web service da parte delle applicazioni client.

L'infrastruttura si basa su un database di registro (usualmente chiamato *Mapping Store*) che contiene:

- metadati SDMX relativi ai dataset disponibili;
- parametri di connessione ai database di diffusione e di mappatura dei relativi dataset.

Per predisporre il rilascio dei dati tramite il Single Exit Point sono state effettuate le seguenti operazioni:

- modellazione degli ipercubi del Censimento secondo i costrutti SDMX per la generazione delle Data Structure Definition (DSD), ovvero delle strutture dati SDMX su cui sono stati poi mappati i dataset locali. Esse sono costituite dalle variabili dimensionali (dimension), le cui combinazioni identificano il dato, dagli attributi (attributes) descrittivi delle caratteristiche dei dati (es. stato delle osservazioni, decimali, unità di misura) e dalle misure. Tale attività è stata effettuata utilizzando lo strumento Data Structure Wizard;
- popolamento del Mapping Store con le DSD precedentemente modellate e gli artefatti (costrutti) SDMX ad esse correlati come le codelist (o classificazioni) ed i *concept scheme* (ovvero gli insiemi di concetti a cui si riferiscono dimensioni ed attributi delle DSD); effettuazione della mappatura dei dati locali su dimensioni, attributi e misure delle DSD, nonché della transcodifica dei codici locali su quelli presenti nelle codelist associate alle varie dimensioni.



L'*end-point* del web service è il seguente:

- http://sdmx.istat.it/WS_CENSAGR/NsiStdV20Service.asmx,

mentre il Web Services Description Language (WSDL) del web service è:

- http://sdmx.istat.it/WS_CENSAGR/NsiStdV20Service.asmx?WSDL.

Gli utenti possono avere accesso al servizio previa comunicazione degli indirizzi IP delle macchine che si devono connettere al web service.

I dati sono interrogabili tramite query SDMX e restituiti in formato SDMX-ML.

Il web service permette l'interrogazione dei dati in formato SDMX tramite i seguenti metodi:

- GetCompactData (che restituisce i dati in formato SDMX-ML Compact);
- GetGenericData (che restituisce i dati in formato SDMX-ML Generic);
- GetCrossSectionalData (che restituisce i dati in formato SDMX-ML Cross-Sectional).

Il metodo che permette l'interrogazione dei metadati (Strutture dati SDMX o Data Structure Definitions, Classificazioni o CodeList, insiemi dei concetti collegati a dimensioni o attributi della Data Structure Definition ovvero Concept Scheme) è il QueryStructure.

3.4 Visualizzazione e rappresentazione grafica dei dati

3.4.1 Il visualizzatore delle tavole on-line

Il **Censimento dell'agricoltura in breve** presenta le principali informazioni ottenute con il Censimento del 2010. I dati, tutti a livello regionale, sono organizzati in cinque temi - Struttura delle aziende agricole, Coltivazioni, Allevamenti, Manodopera, Altre attività -, cui si aggiunge una sezione dedicata alle Serie storiche.

L'obiettivo è di offrire una consultazione agile dei risultati, anche grazie all'ausilio di visualizzazioni grafiche interattive, che possono essere salvate nei formati jpg e png.

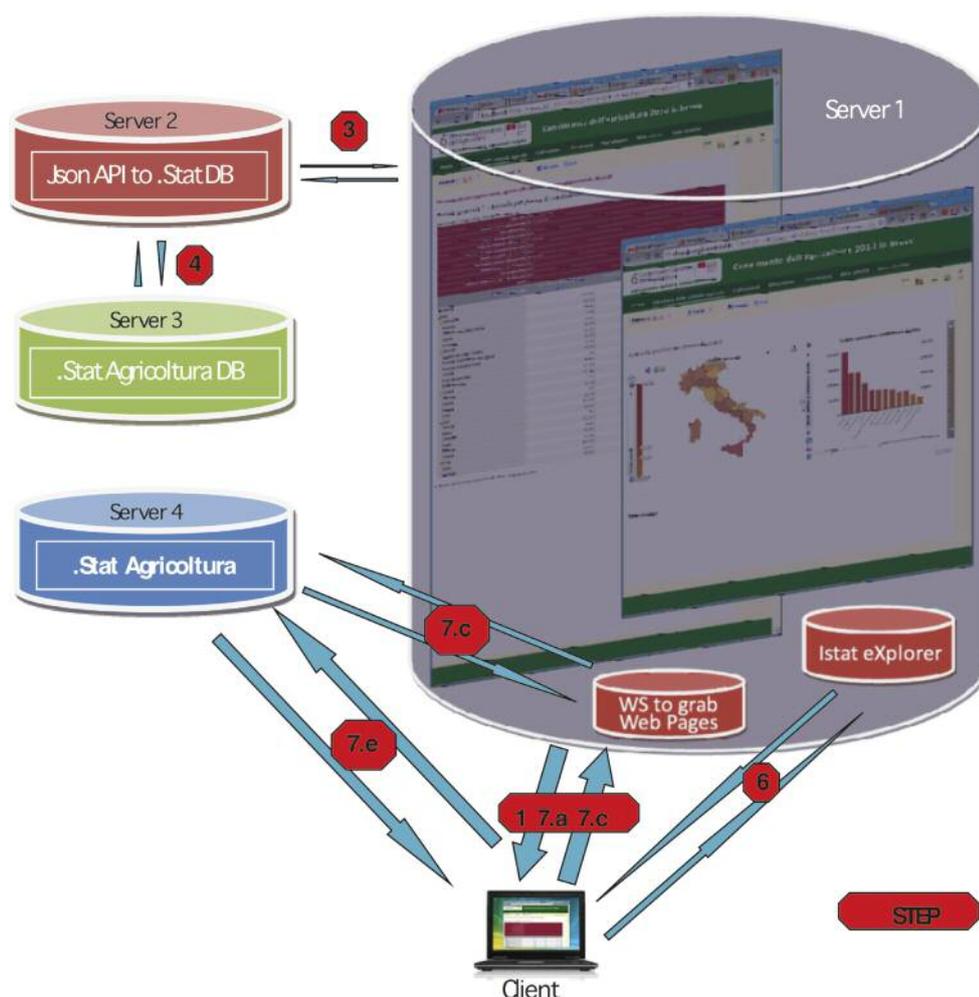
Ogni tavola, esportabile in formato Excel, csv e sdmx, consente inoltre di ottenere maggiori dettagli informativi attraverso appositi link al *data warehouse* generale del Censimento dell'agricoltura.

I grafici e le tavole possono essere condivisi attraverso una serie di *social buttons* messi a disposizione in ogni pagina.

Dal punto di vista tecnico il processo di visualizzazione (Figura 3.5) si articola nei seguenti passi:

1. l'utente richiede una determinata tabella o grafico in modalità interattiva;
2. un *Web Server* richiama l'applicazione web;
3. l'applicazione Web richiama una API JSON che effettua un'interrogazione al data warehouse;
4. l'API JSON riceve dal data warehouse i dati e tutte le informazioni correlate (tema e dataset) in risposta alla richiesta effettuata da un utente;
5. viene eseguita una query e generato un codice HTML che viene inviato al Client, il quale a sua volta carica il codice html e tutti gli altri componenti di riferimento (.css, .cs, .jpg, .png, file swf).

Figura 3.5 - Processo di visualizzazione



3.4.2 Gli strumenti per l'analisi grafica interattiva

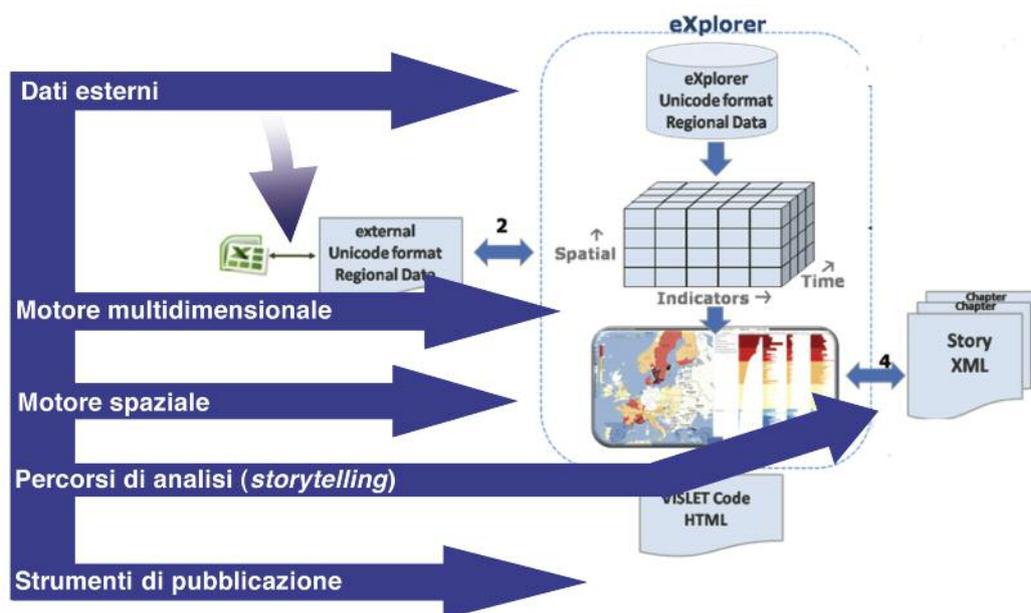
Oltre al prodotto **Censimento dell'agricoltura in breve**, il sistema di diffusione dei dati del censimento dell'agricoltura si è avvalso di uno strumento avanzato di analisi grafica. È stata utilizzata per questo la suite Istat.eXplorer, un sistema di visualizzazione grafica interattiva e dinamica di dati statistici spazio-temporali. L'applicazione è basata sulla personalizzazione del tool **Statistics eXplorer**, sviluppato dal centro di ricerche svedese NComVA¹³.

L'architettura generale dell'applicazione è descritta nella Figura 3.6.

¹³ Statistics eXplorer è stato sviluppato dal centro ricerche NComVA (Norrköping Communicative Visual Analytics), una *spin-off company* dell'università svedese di Linköping con la quale Istat ha attivato una collaborazione tecnica.

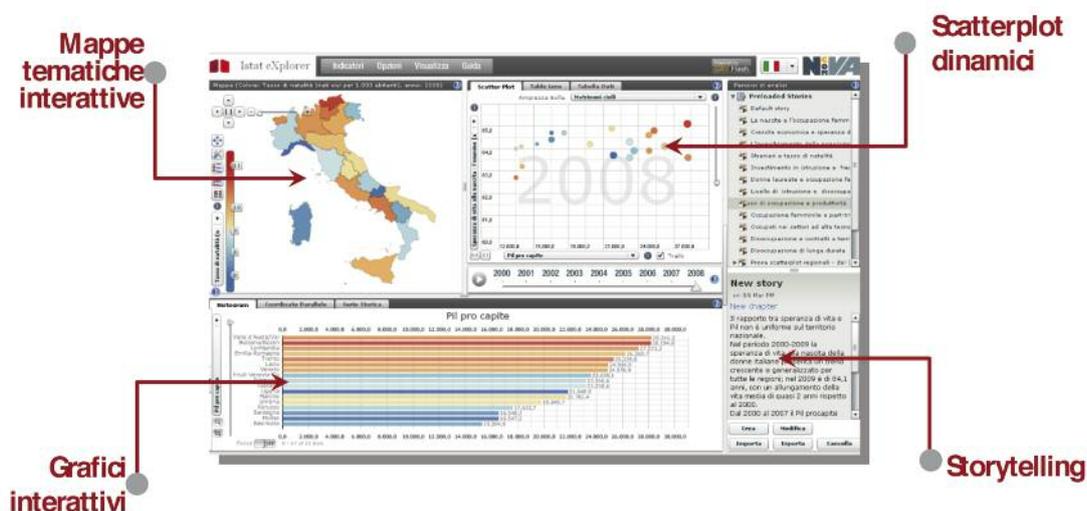


Figura 3.6 - Architettura Istat.eXplorer



Il sistema offre la possibilità di rappresentare i dati attraverso mappe tematiche e strumenti grafici avanzati (diagrammi a bolla, *scatterplot*, *table lens*, grafici a coordinate parallele, grafici in serie storica, eccetera) (Figura 3.7). È uno strumento **interattivo** che consente agli utilizzatori di caricare propri dati, di combinare on-line gli indicatori presenti, di personalizzare formati e contenuti, di costruire e condividere “storie” e percorsi di analisi associati ai dati. Lo strumento consente anche analisi dinamiche offrendo la possibilità di seguire l’andamento spazio-temporale dei fenomeni, visualizzandone il movimento grafico sullo schermo.

Figura 3.7 - Rappresentazioni grafiche



Il sistema adotta un potente e innovativo metodo di visualizzazione dei dati (*Vislet*), basato su tecnologie *widget*, che offre una serie di oggetti dinamici incorporati nella struttura testuale di base della pubblicazione e permette agli utenti di gestire analisi di serie temporali, confronti a livello europeo, nonché la selezione rapida di aree territoriali alle quali si riferiscono gli indicatori.

Attraverso la tecnologia *Vislet* e le funzioni di *Storytelling*, *Statistics eXplorer* consente anche di realizzare rappresentazioni visuali dinamiche *preconfezionate*, composte da grafici e testo, le quali possono essere “catturate” e incorporate in siti e altre applicazioni Web ed essere messe in tal modo a disposizione degli utenti, nonché di condividere e inter-scambiare le elaborazioni effettuate con altri utilizzatori.

Tutte le componenti grafiche sono integrate tra loro e interattive, offrendo agli utilizzatori un “laboratorio” visuale di analisi e sintesi grafica, che consente anche la pubblicazione e la diffusione su Web dei risultati ottenuti e dei prodotti realizzati.



Conclusioni

I sistemi informatici a supporto delle fasi di acquisizione, monitoraggio, controllo, correzione e diffusione dei dati del 6° Censimento generale dell'agricoltura 2010 sono stati realizzati con tecniche di progettazione, modellazione e sviluppo software in un'ottica di standardizzazione complessiva mediante:

- l'adozione di metodi e metodologie condivise;
- l'utilizzo di tool standard per la gestione di tutte le fasi che compongono il ciclo di vita del software;
- l'utilizzo di strumenti di sviluppo basati su framework;
- l'adozione di sistemi e di procedure per il controllo della qualità del codice;
- l'adozione di sistemi e di procedure per misurare la vulnerabilità del codice rispetto a possibili attacchi esterni;
- l'adozione di sistemi di misura del carico atteso e di test per la definizione delle configurazioni ottimali dei sistemi di esercizio;
- l'adozione di metodologie e strumenti di versioning e di testing;
- la predisposizione di documentazione di progetto standard;
- l'adozione di strumenti di project management per la gestione della pianificazione e del monitoraggio del progetto in termini di risorse, attività, tempi;
- l'utilizzo di strumenti *open source* adottati in Istat, con conseguente contenimento dei costi.

L'adozione di metodologie e modelli standard per lo sviluppo delle applicazioni informatiche 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, di produrre la documentazione appropriata hanno consentito:

- la creazione di un Knowledge Management System;
- il riuso di gran parte delle componenti software;
- la realizzazione di sistemi ad alto grado di generalizzazione per la gestione su piattaforma web di processi simili a quelli censuari basati su una rete di rilevazione, la gestione della multi-canalità nella fase di acquisizione e la predisposizione di questionari elettronici.

I sistemi realizzati, pur non essendo "generalizzati", sono stati progettati in modo da consentire un elevato grado di riuso 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, tanto da essere stati di fatto completamente riutilizzati, migliorandoli, per la gestione del 15° Censimento della Popolazione e delle Abitazioni e del 9° Censimento dell'Industria e dei Servizi.

Vanno comunque evidenziati alcuni elementi di criticità. La complessità della rete e dell'organizzazione del lavoro ha richiesto un sistema 'complesso'. La fase di start up nell'utilizzo del sistema è stata più lunga e laboriosa del previsto ed è stato necessario un maggiore affiancamento tra Uffici regionali Istat e rete di rilevazione con un impegno superiore al previsto degli Uffici regionali di censimento. Inoltre, all'avvio del sistema di compilazione on-line da parte delle aziende agricole, il consistente numero di accessi contemporanei da parte delle aziende agricole ha messo in evidenza un comportamento anomalo dell'applicazione, non evidenziato in fase di test. È stato necessario, a poche ore dalla partenza, sospendere il servizio e sottoporre l'applicazione a test più approfonditi, sia sulla sicurezza sia sulle funzionalità del sistema nei momenti di picco al fine di rimuovere gli errori. Inoltre la particolare complessità del

questionario di rilevazione ha richiesto ulteriori verifiche. Questo ha consentito di riaprire il Sistema di gestione della rilevazione dopo 48 ore, mentre il sistema di acquisizione on-line è stato riaperto alle aziende dopo due settimane. Lo stop del sistema dopo l'avvio del censimento ha di fatto determinato un abbassamento delle risposte via web da parte delle aziende agricole rispetto alle aspettative iniziali.

In generale il sistema ha permesso la costruzione di una rete articolata composta da più di 20.000 operatori; a fronte delle oltre 2.000.000 di unità incluse nella lista pre-censuaria iniziale. I questionari relativi alle oltre 2 milioni di unità sono stati tutti registrati on line, circa 1.352.000 questionari sono stati inseriti via web dagli operatori di censimento delle regioni ad alta partecipazione (66,1 per cento); circa 636.000 notizie di base riferite alle variabili primarie¹⁴ sono state inserite dagli operatori degli UCC delle regioni a partecipazione integrativa attraverso l'apposita funzionalità a disposizione nel sistema di gestione delle rilevazione (31,0 per cento) e circa 61.000 questionari sono stati compilati via web direttamente dalle aziende agricole (2,9 per cento).

¹⁴ Registrazione del "Modello riepilogativo per singolo questionario". Nota 2



Riferimenti bibliografici

“Piano generale del 6° Censimento generale dell’agricoltura”, Istat, 2010.

E. Berntsen, “Istruzioni per la rilevazione, 6° Censimento generale dell’agricoltura”. Istat, 2010.

G. Bellini, C. Conti, M. Giacommo, G. Mottura, M. Venturi, “Web Use For The 6TH General Agriculture Census In Italy”, Journées de Méthodologie Statistique, 2012, Insee - Paris, Janvier 2012.

C. Cavaness: Programming Jakarta Struts, 2nd Edition, O’Reilly Media, 2004.

C. Begin, B. Goodin, L. Meadors: iBATIS in Action, Manning Publications, 2007.

W. Crawford, J. Kaplan, “J2EE Design Patterns”, Ed. O’Reilly, 2003.

B. Kurniawan, “Struts 2 Design and Programming”, Ed. BrainySoftware, 2008.

L. Vetti Tagliati, “UML”, Ed. Hoepli, 2003.

Il fascicolo è dedicato al sistema informatico integrato, che ha rappresentato un importante strumento operativo di supporto alla rete di rilevazione, consentendo di gestire l'acquisizione e la memorizzazione dei dati, il monitoraggio della rilevazione e la prima fase del piano di controllo dei dati. Gli aspetti descritti in modo dettagliato riguardano l'architettura generale, l'infrastruttura tecnologica, le funzionalità realizzate e, infine, le soluzioni adottate per la diffusione dei dati. I manuali d'uso prodotti e le regole di controllo applicate sono contenuti, in formato digitale, nel cd-rom collocato all'interno del primo fascicolo dell'opera, *Il quadro normativo*.