

I SISTEMI INFORMATICI PER I CENSIMENTI PERMANENTI





I SISTEMI INFORMATICI PER I CENSIMENTI PERMANENTI

A cura di: Maura Giacummo ed Eleonora Sibilio.

Attività editoriali: Nadia Mignolli (coordinamento), Claudio Bava, Alfredina Della Branca,
Marco Farinacci, Alessandro Franzò e Manuela Marrone.

Responsabile per la grafica: Sofia Barletta.

ISBN 978-88-458-2160-8

© 2024

Istituto Nazionale di Statistica
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti
pubblicati sono soggetti alla licenza
Creative Commons - Attribuzione - versione 4.0.
<https://creativecommons.org/licenses/by/4.0/deed.it>

È dunque possibile riprodurre, distribuire,
trasmettere e adattare liberamente dati e analisi
dell'Istituto nazionale di statistica, anche a scopi
commerciali, a condizione che venga citata la fonte.

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



INDICE

	Pag.
Premessa	7
1. Il questionario elettronico	9
1.1 Configurazione e potenzialità della piattaforma	9
1.1.1 <i>Struttura del questionario</i>	10
1.1.2 <i>Quesiti, flusso e regole di compilazione</i>	12
1.1.3 <i>Gestione dei testi e il supporto per il multilinguismo</i>	14
1.1.4 <i>Integrazione con gli altri sistemi</i>	17
1.2 Gestione dei dati	17
1.2.1 <i>Architettura della base di dati</i>	18
1.2.2 <i>Paradati raccolti durante la compilazione</i>	19
1.3 Sviluppi futuri	20
2. Il Sistema di gestione delle indagini (Sgi)	21
2.1 Digitalizzazione del processo	24
2.2 Progettazione della piattaforma	26
2.3 Aspetti di usabilità	29
2.4 Procedure di configurazione della rete e di caricamento del campione	32
2.5 Creazione, gestione e organizzazione della rete	37
2.6 Rilevazione sul campo	41
2.7 Organizzazione, gestione e monitoraggio della formazione	45
2.8 Base dati	47
2.8.1 <i>Gestione e la manutenzione</i>	49
2.9 Primi risultati e sviluppi futuri	50
3. L'app Rilevo per la gestione dei Censimenti permanenti	53
3.1 Progettazione	55
3.2 Funzionalità	55
3.2.1 <i>Questionario</i>	58
3.3 Sincronizzazione	60
3.4 Sintesi e sviluppi futuri	60

	Pag.
4. La reportistica per le indagini	63
4.1 Piattaforma <i>Microstrategy</i>	63
4.1.1 <i>Rapporti riassuntivi</i>	64
4.1.2 <i>Data integration</i>	67
4.1.3 <i>Configurazione del progetto</i>	70
4.1.4 <i>Tipologie di reportistica</i>	71
4.1.5 <i>Risultati e sviluppi futuri</i>	73
4.2 Piattaforma <i>Apex</i>	74
4.2.1 <i>Rapporti riassuntivi del Censimento permanente delle istituzioni pubbliche</i>	75
4.2.2 <i>Back end: la base dati</i>	75
4.2.3 <i>Front End: web application</i>	77
4.2.4 <i>Risultati e sviluppi futuri</i>	78
5. La formazione a distanza	81
5.1 Piattaforma di formazione integrata per le indagini	82
5.2 Architettura software e integrazione con Sgi e il Sistema utenti	82
5.2.1 <i>Sincronizzazione delle utenze registrate nei Directory Service con Moodle</i>	84
5.2.2 <i>Creazione nel corso Moodle dei gruppi</i>	85
5.2.3 <i>Estrazione della lista di utenti</i>	85
5.2.4 <i>Iscrizione degli utenti al corso di riferimento</i>	86
5.2.5 <i>Inserimento degli utenti e sincronizzazione delle informazioni anagrafiche</i>	86
5.2.6 <i>Sospensione delle utenze già presenti sul sistema Moodle</i>	86
5.2.7 <i>Backup dei corsi ancora attivi sul sistema Moodle</i>	87
5.3 Potenzialità di <i>Moodle</i>	87
5.3.1 <i>Strumenti della piattaforma</i>	88
6. Il Portale della rete	91
6.1 Autenticazione e autorizzazione	91
6.2 Organizzazione del sito	92
7. La gestione dei dati e l'utilizzo dei registri	95
7.1 Registro delle imprese e Censimenti economici	97
7.1.1 <i>Sistema informativo del Registro delle imprese: fonti, database e applicativi a supporto</i>	98
7.1.2 <i>Utilizzo del Registro delle imprese nei Censimenti economici</i>	102
7.2 Registro dei gruppi di impresa e Censimenti economici	103
7.2.1 <i>Architettura del database del Registro statistico dei gruppi di impresa e il processo informatico di costruzione</i>	104
7.2.2 <i>Nuova implementazione del processo informatico di costruzione del Registro statistico dei gruppi di impresa</i>	105
7.3 Industrializzazione dei registri	115
7.4 Indicatori di qualità e di monitoraggio: sperimentazioni sul Registro base degli individui	117
7.5 Supporto per la validazione dei dati	119
7.5.1 <i>Reportistica a supporto</i>	123

	Pag.
8. La piattaforma per la diffusione web dei dati	127
8.1 Requisiti di diffusione web dei dati	127
8.2 Soluzione tecnologica	127
8.3 SDMX Istat <i>Toolkit</i>	129
8.3.1 <i>Meta and Data Manager</i>	129
8.3.2 <i>Data Browser</i>	132
8.4 Disegno architetturale	134
8.5 Evoluzioni programmate	135
8.6 <i>Cubus</i>	135
8.6.1 <i>Cubus per il Censimento generale dell'Agricoltura</i>	136
8.6.2 <i>Cubus per il Censimento permanente delle istituzioni pubbliche</i>	141
9. L'Architettura per l'autenticazione degli utenti	147
9.1 Sistema di <i>single sign-on Shibboleth</i>	147
9.2 Piattaforma <i>Identity provider single sign-on</i> e integrazione con Spid	148
10. La gestione dei tablet	151
10.1 Gestione logistica dei dispositivi	151
10.2 Gestione tecnologica dei dispositivi	152
10.3 <i>Set-up & Configuration Manager</i> dei tablet <i>Knox</i>	153
10.4 Gestione della flotta dei tablet e <i>Private app store</i> mediante il <i>Mobile Device Manager</i>	154
10.5 <i>Accountability</i> e associazione dei diversi profili utente ai <i>device</i>	155
10.6 Gestione delle utenze dei rilevatori sulla piattaforma <i>Teams</i>	156
10.7 Ottimizzazione dei costi di esercizio	158
11. Il service desk	159
11.1 Piattaforma <i>IT Service Management (ITSM)</i> a supporto del <i>service desk</i>	160
11.1.1 <i>Architettura della piattaforma ITSM</i>	160
11.1.2 <i>Interfaccia utente: MyIT Digital Workplace</i>	161
11.1.3 <i>Interfaccia operatore: SmartIT</i>	163
11.1.4 <i>Dashboard di monitoraggio dei processi di Service Management per le indagini: Smart Reporting e Power BI</i>	164
11.2 Monitoraggio dei servizi	166
12. Le architetture middleware	171
12.1 Portale della rete	172
12.2 Piattaforma di formazione	173
12.3 Piattaforma per la gestione delle indagini	174
12.4 Reportistica di raccolta dati	175
12.5 Questionario di raccolta dati	176
12.6 <i>Data warehouse</i> per la diffusione dei dati	177
13. Le architetture delle basi di dati	181
13.1 Architettura <i>Oracle database</i>	181
13.1.1 <i>Ambienti Oracle database</i>	183

	Pag.
13.1.2 <i>Schemi dei dati e accesso utente</i>	184
13.1.3 <i>Accesso applicativo ai dati tramite Wallet</i>	184
13.2 <i>Architettura Microsoft SQL Server</i>	186
13.2.1 <i>Architettura dei database per i tablet</i>	186
13.2.2 <i>Architettura del data warehouse per la diffusione</i>	187
13.2.3 <i>Architettura MySQL per il Portale della rete</i>	188
13.3 <i>Tecnologie correlate alle basi di dati</i>	189
13.3.1 <i>Sistema di replica dei dati in tempo reale</i>	189
13.3.2 <i>Controllo delle attività degli utenti di database</i>	190
13.3.3 <i>Architettura di Oracle Apex</i>	192
14. Le infrastrutture di monitoraggio	195
14.1 <i>Monitoraggio dei server</i>	195
14.2 <i>Monitoraggio delle applicazioni</i>	198
14.3 <i>Monitoraggio delle performance delle basi di dati</i>	200
14.4 <i>Alert, ticket e monitor sintetici</i>	202
15. La gestione della sicurezza informatica	203
15.1 <i>Gestione della sicurezza e i vulnerability assessment</i>	203
15.2 <i>Gestione della sicurezza e i penetration test</i>	205
15.3 <i>Gestione della sicurezza perimetrale</i>	207
15.4 <i>Pubblicazione dei siti web dell'Istat</i>	208
16. Il Sistema di gestione della sicurezza delle informazioni	211
16.1 <i>Sistema di gestione della sicurezza per le indagini censuarie</i>	212
16.1.1 <i>Metodologia</i>	212
16.2 <i>Business Impact Analysis</i>	213
16.3 <i>Risk management</i>	214
16.4 <i>Compliance per gli aspetti di sicurezza e cybersecurity</i>	215
16.5 <i>Sistema di monitoraggio della governance della sicurezza</i>	215
Riferimenti bibliografici	217

PREMESSA¹

Il programma di modernizzazione dell'Istat persegue la finalità di arricchire l'offerta e la qualità delle informazioni prodotte, avvalendosi delle più recenti evoluzioni dell'ICT, e mira a ridurre sia i tempi di diffusione dei dati sia i costi dell'indagine. Coerentemente con questo, dal 2018 in poi l'Istat, grazie al cambiamento metodologico adottato per la strategia censuaria, ha avviato un progetto volto a digitalizzare i Censimenti permanenti. L'obiettivo di tale progetto consiste nell'innovare radicalmente una delle attività di maggiore rilievo e sensibilità per l'Istituto, la rilevazione dei dati per il Censimento della Popolazione e delle abitazioni, mediante la dematerializzazione del questionario e l'incremento della frequenza di rilevazione, da decennale ad annuale, al fine di fornire alle istituzioni e ai cittadini dati aggiornati e di elevata qualità in tempi più rapidi.

La realizzazione di un progetto così rilevante, che impatta direttamente sugli obiettivi dell'Istituto, ha presupposto la realizzazione di specifiche attività volte a reingegnerizzare e innovare molteplici aspetti della catena IT: la governance, l'infrastruttura, le piattaforme, l'interoperabilità e la sicurezza. Per ottenere questo risultato, è stato revisionato completamente il modello di governance e monitoraggio dell'IT favorendo l'accentramento delle iniziative sotto un'unica Direzione centrale informatica che fosse responsabile di tutti i progetti tecnologici e della standardizzazione dei processi nel rispetto delle *best practice*. In tale modo è stato possibile predisporre in tempi brevi l'infrastruttura dedicata alla rilevazione dei dati del nuovo Censimento, digitalizzando il processo di conduzione e garantendo la medesima struttura IT per tutte le indagini censuarie. L'infrastruttura realizzata ha consentito lo sviluppo di una nuova piattaforma di gestione del processo di raccolta dati fondata sull'integrazione di diversi sistemi e software, specializzati in tutte le singole fasi del processo censuario. Infine, le attività svolte per valorizzare l'interoperabilità tra i sistemi e garantire la sicurezza dell'infrastruttura sono state fondamentali per assicurare un'elevata qualità nella realizzazione del progetto.

La digitalizzazione dei censimenti ha comportato una trasformazione importante per l'Istat in quanto ha portato all'eliminazione del cartaceo e ha condotto a un radicale cambiamento nelle modalità di rilevazione dei dati. Difatti, dal 2018 è stato possibile erogare questionari digitali precompilati, somministrati anche mediante l'utilizzo di 14 mila tablet, ed effettuare indagini sul campo che hanno raggiunto circa 1,4 milioni di famiglie, per un totale di 3,5 milioni di persone residenti in 2.852 comuni italiani. Pertanto, l'innovazione apportata ha consentito di effettuare la più grande indagine sul campo svolta in Italia con un risparmio di decine di milioni di euro e una riduzione dell'utilizzo della carta di diverse tonnellate. La conduzione di una indagine su così larga scala è stata possibile anche grazie a una maggiore e più efficace governance del Censimento. La ricezione tempestiva dei dati dai questionari compilati ha permesso all'Istat di monitorare in tempo reale l'andamento della rilevazione, analizzare i risultati e implementare eventuali azioni di comunicazione necessarie a raggiungere il campione rappresentativo della popolazione di riferimento.

Inoltre, la sinergia delle soluzioni infrastrutturali e applicative realizzate ha permesso di ottenere sviluppi fondamentali nel processo censuario tra cui la possibilità di effettuare la rilevazione offline, tale elevato livello di innovazione, raggiunto grazie al lavoro dell'intera

¹ L'ebook è stato curato da Maura Giacommo ed Eleonora Sibilio. La premessa è stata redatta da Massimo Fedeli.



Direzione informatica, ha avuto degli effetti positivi non solo sulle indagini censuarie ma anche sulle altre rilevazioni caratterizzate da un elevato tasso di complessità tecnica.

In conclusione, il progetto raccontato in dettaglio in questa pubblicazione è la prima fase di un processo più ampio di trasformazione digitale dell'Istat, che mira a modernizzare e a innovare le attività censuarie e di indagine svolte dall'Istituto nell'ottica di un miglioramento continuo dei servizi offerti alla collettività.

1. IL QUESTIONARIO ELETTRONICO¹

La possibilità di fare compilare il questionario ai rispondenti in modalità elettronica ha un enorme valore sia in termini di qualità del dato ricevuto sia per la possibilità di usufruire dello stesso in tempo reale. Evita inoltre alcune problematiche come quella della conservazione e della gestione del dato inviato in formato cartaceo.

L'Istat, consapevole dei vantaggi che si potevano raggiungere, già nell'ultimo Censimento decennale della Popolazione e delle abitazioni del 2011 ha adottato tra i diversi canali di restituzione anche la tecnica *Cawi* (*Computer-assisted web interviewing*). A valle dell'esperienza censuaria, ha inoltre avviato delle indagini sperimentali che, dal punto di vista tecnologico, hanno avuto lo scopo di analizzare l'effetto dell'introduzione del questionario elettronico anche per la rete di rilevazione. Questo aspetto, come analizzato anche nel Capitolo 3, ha anche posto il tema di dovere gestire i questionari elettronici in caso di assenza di rete Internet.

Nel corso di dette sperimentazioni del 2016 sono state date agli utenti le seguenti modalità di lavoro:

- il sistema web con la gestione sia della parte online sia offline
- un sistema web per la parte online e un *app* per la gestione dell'offline.

L'utilizzo nella *Capi* (*Computer-assisted personal interviewing*), di un sistema digitalizzato ha semplificato notevolmente il lavoro alla rete di rilevazione che ha potuto limitare l'utilizzo del cartaceo nelle sole situazioni di emergenza.

Nel 2018, con il passaggio ai Censimenti permanenti, le indagini sono state trasformate; oltre i diversi aspetti metodologici, sono state cambiate alcune caratteristiche quali:

- la cadenza, non più decennale ma annuale o al massimo triennale in base al tipo di rilevazione;
- la numerosità, non più indagini esaustive ma campionarie;
- una gestione *paperless*, ossia abbandonando totalmente l'utilizzo della carta.

Questo importante passaggio ha fatto nascere la necessità di avere uno strumento robusto e sicuro e di facile usabilità che potesse gestire le due classi di utenza impattate dal processo: i rispondenti e la rete di rilevazione.

È in questo contesto che si è sviluppato *Panda*, un sistema generalizzato per la progettazione, l'implementazione e la gestione del ciclo di vita dei questionari elettronici.

1.1 Configurazione e potenzialità della piattaforma

Il sistema di generazione dei questionari elettronici, *Panda*, è costituito da un motore che riceve in input un insieme di metadati e genera in output l'applicazione web che implementa il questionario elettronico rappresentato attraverso i metadati, come mostrato nella Figura 1.1. L'approccio descritto ha consentito, da una parte, di concentrare e affrontare in un unico software tutte le problematiche tecnologiche connesse allo sviluppo e alla manu-

¹ Il Capitolo è stato redatto da Luigi Arlotta (paragrafi 1,1.1,1.1.1,1.1.2 e 1.3), Massimo Andreani (paragrafo 1.1.3), Francesco Degni (paragrafi 1.1.4, 1.2, 1.2.1, 1.2.2).

tenzione di un ampio insieme di questionari eterogenei; dall'altra di mantenere gli elevati standard di sicurezza, multimodalità e usabilità richiesti.

Attraverso il sistema, quindi non è più necessario mantenere ed evolvere, per ogni questionario elettronico, una applicazione web complessa, ma è sufficiente gestire e aggiornare i metadati che rappresentano il questionario.

Sarà poi il motore di *Panda*, in fase di preparazione della rilevazione, a generare in automatico l'applicazione web che implementa il questionario o i questionari elettronici necessari sulla base dei metadati forniti in input.

Figura 1.1 - Produzione dei questionari a partire dai metadati forniti in input al motore di *Panda*



Fonte: piattaforma *Panda*

1.1.1 Struttura del questionario

I questionari realizzati attraverso il sistema *Panda* prevedono una struttura standard.

L'accesso al questionario avviene attraverso una *home page* che prende il nome di copertina del questionario (Figura 1.2).

La copertina è una pagina pubblica che ha lo scopo di presentare l'indagine attraverso una breve descrizione e fornire alcune informazioni di base sulla rilevazione.

Dalla copertina è possibile eseguire la procedura di autenticazione che porta il rispondente alla pagina della guida alla compilazione, contenente le istruzioni necessarie per la compilazione e il corretto utilizzo del questionario elettronico.

Da qui il rispondente può iniziare la compilazione del questionario vero e proprio che è generalmente organizzato in sezioni tematiche contenenti i diversi quesiti.

Ogni sezione, se necessario, può essere suddivisa in più pagine.

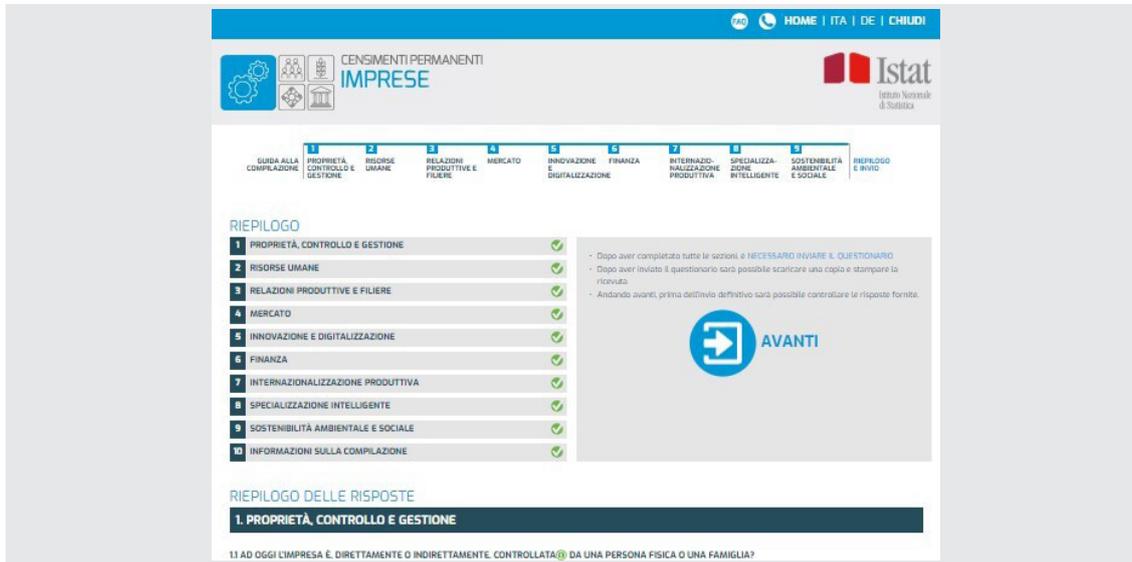
Un indice, sempre presente nella parte alta della pagina, consente di spostarsi tra le diverse sezioni. Completata la compilazione del questionario, il rispondente naviga verso la pagina di riepilogo e invio (Figura 1.2).

In questa pagina è possibile verificare lo stato di compilazione delle singole sezioni di cui il questionario si compone, visionare l'insieme delle risposte fornite e procedere con la chiusura e l'invio definitivo del questionario.

Una volta chiuso il questionario sarà possibile acquisire la ricevuta di avvenuta compilazione in formato elettronico e, se lo si desidera, scaricare una copia del questionario compilato.

1. Il questionario elettronico

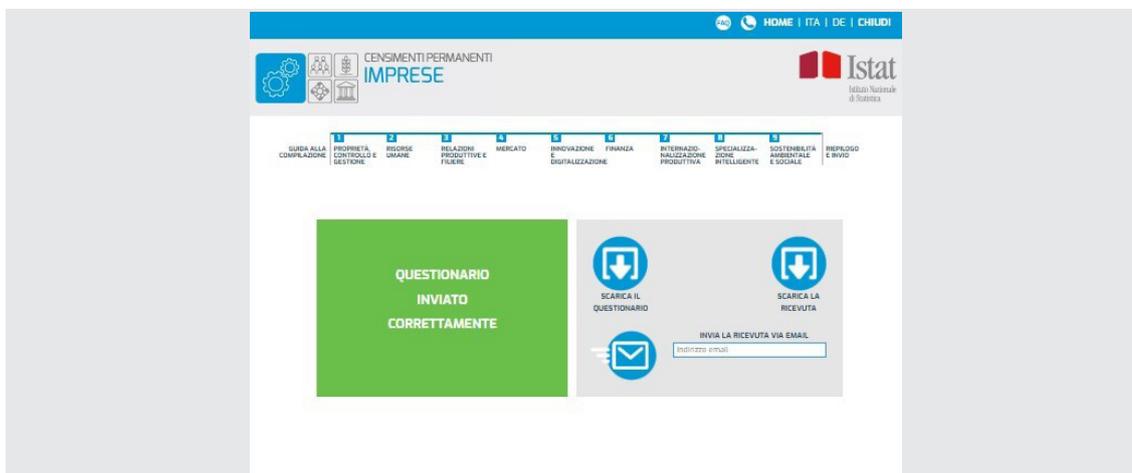
Figura 1.2 - Pagina di riepilogo



Fonte: piattaforma Panda

Tutte le pagine interne al questionario presentano alcuni elementi comuni, sempre a disposizione dell'utente che compila il questionario. Nella barra superiore compaiono i bottoni e i link per la selezione della lingua desiderata, per tornare alla pagina della guida alla compilazione, per abbandonare la compilazione o per contattare l'help desk. La barra descritta ha un formato standard, ma all'esigenza può essere estesa o personalizzata per una particolare indagine aggiungendo o rimuovendo uno o più bottoni. Al di sotto della barra appena descritta è sempre presente un banner contenente il logo del Censimento a cui il questionario si riferisce e il logo dell'Istat. Ultimo elemento comune a tutte le pagine è l'indice del questionario, posto sotto al banner. L'indice consente di spostarsi agevolmente tra le diverse sezioni, di tornare alla pagina della guida o di raggiungere la pagina dedicata alla chiusura e all'invio del questionario. La colorazione dell'indice evidenzia al rispondente lo stato di compilazione delle diverse sezioni: via via che le sezioni vengono compilate il colore della piccola barra posta al di sopra dei nomi delle singole sezioni, dal grigio iniziale, acquisisce il colore specifico del Censimento.

Figura 1.3 - Questionario inviato



Fonte: piattaforma Panda



1.1.2 Quesiti, flusso e regole di compilazione

Il sistema *Panda* prevede la definizione del questionario attraverso un insieme di file di metadati nei quali sono descritti la struttura del questionario, la tipologia dei quesiti e la loro organizzazione in pagine e sezioni tematiche, le regole e i flussi di compilazione. Questi file, che definiscono la struttura e il comportamento del questionario durante la fase di raccolta dati, vengono forniti in input al sistema attraverso una sintassi propria di *Panda*, basata su XML (*eXtensible Markup Language*). Tutti i contenuti testuali del questionario, quali ad esempio i titoli delle sezioni, i testi dei quesiti, i messaggi di errore, sono invece mantenuti in file separati e organizzati come file di *properties* contenenti coppie chiave e valore. La separazione tra struttura e testi semplifica la gestione delle caratteristiche di multilinguismo del sistema e, come vedremo nei paragrafi successivi, semplifica le attività legate ai processi di traduzione. Sarà poi il sistema *Panda*, in fase di elaborazione dell'input, a riconciliare la struttura del questionario con i contenuti testuali. Le sezioni tematiche, di cui il questionario si compone, contengono i quesiti e potranno essere organizzate in una o più pagine a seconda delle caratteristiche degli elementi utilizzati e degli argomenti trattati nel questionario. L'utente che si occupa della costruzione del questionario ha a disposizione un ampio insieme di modelli di quesiti predefiniti che si distinguono per struttura, funzionamento e *layout*. Tra questi abbiamo ad esempio quesiti a risposta singola, quesiti a risposta multipla, quesiti a risposta aperta, varie forme di quesiti tabellari e diversi altri ancora. La Figura 1.4 mostra degli esempi di tipologie.

Figura 1.4 - Diverse tipologie dei quesiti predefiniti

The screenshot displays several question types from the Panda questionnaire system:

- X1.4**: A Likert scale question asking about the centralization of strategic decisions. It has four response options: Molto, Abbastanza, Poco, and Per nulla.
- A.31**: A percentage-based question asking for the percentage of revenue achieved in Italy and abroad. It includes a table for data entry and a 'Totale' row.
- X1.6**: A table-based question asking about strategic objectives for 2021-2022. It has columns for 'Obiettivo Perseguito' (Yes/No) and 'Esito ad oggi' (Fully achieved/Partially achieved/Not achieved).
- 1.2**: A multiple-choice question asking who is responsible for company management, with options like 'Il imprenditore o il socio principale o unico', 'Un membro della famiglia proprietario o controllante', etc.

Fonte: piattaforma *Panda*

La definizione di un quesito avviene attraverso la produzione di un blocco di codice xml e di un insieme di coppie chiave e valore con i testi del quesito. Nel codice xml vengono indicati il tipo del quesito, un identificativo univoco, le variabili ed eventualmente il numero delle possibili modalità di risposta. I testi del quesito sono invece inseriti in un file necessario per la realizzazione del quesito.

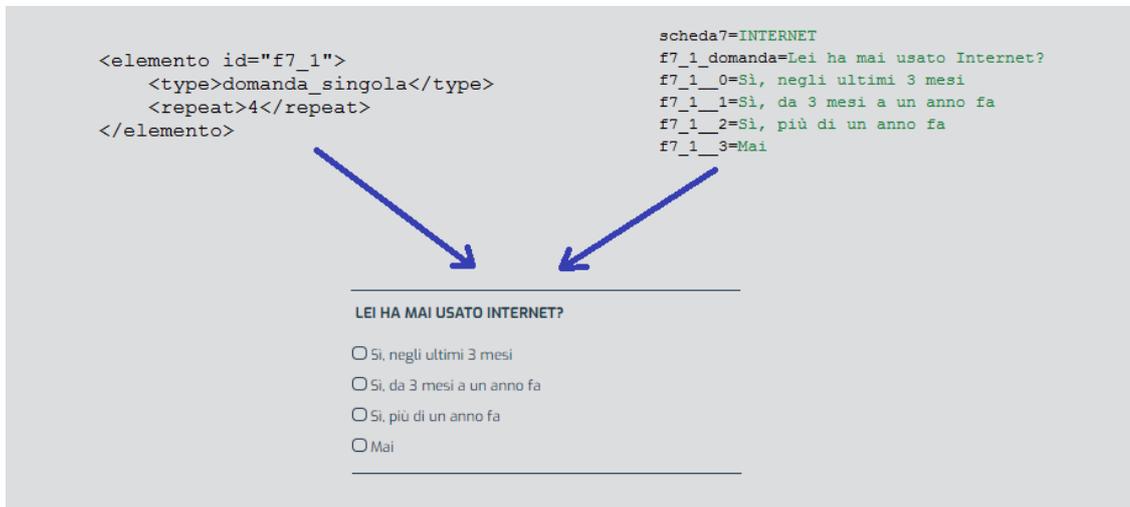
Nella Figura 1.5 sono riportati a titolo di esempio le sezioni dei file di input di *Panda* attraverso le quali il sistema è in grado di produrre un quesito a risposta singola.

Qualora il *set* dei quesiti predefiniti non dovesse soddisfare tutte le esigenze di una specifica indagine è possibile implementare nuovi *template* dedicati a una particolare rilevazione. Il *layout* del questionario, delle pagine e dei singoli quesiti può essere modificato intervenendo sui file di stile predefiniti, file css, tramite opportune estensioni.

Una volta definita la struttura di base e i quesiti da sottoporre ai rispondenti, il questionario può essere arricchito con regole ed eventi che guidano il rispondente durante la compilazione. *Panda* consente l'inserimento di cinque tipologie di regole di base: le regole di flusso, le regole di obbligatorietà, le regole di dominio, le regole di congruenza e le regole di calcolo.

1. Il questionario elettronico

Figura 1.5 - Generazione di un quesito a partire dai metadati



Fonte: piattaforma Panda

Le regole di flusso definiscono quali quesiti debbano essere sottoposti a ogni rispondente in base al suo profilo iniziale o in base alle risposte fornite fino a quel momento. Con riferimento al questionario del Censimento permanente della Popolazione e delle abitazioni, ad esempio i quesiti relativi all'attività lavorativa non vengono visualizzati ai minori di una certa età o a coloro che dichiarano di non avere un lavoro. Questo comportamento del questionario deriva dalle regole di flusso che durante la compilazione analizzano le risposte fornite e definiscono in tempo reale quali quesiti debbano essere sottoposti al rispondente.

Le regole di obbligatorio definiscono l'insieme dei quesiti a cui è necessario fornire una risposta. Sono previsti due diversi livelli di obbligatorio. Il primo livello viene definito bloccante e si applica a quei quesiti la cui mancata risposta deve impedire la prosecuzione della compilazione del questionario. Il secondo livello, detto non bloccante, prevede invece la segnalazione della mancata risposta, ma non inibisce la prosecuzione della compilazione.

Figura 1.6 - Messaggio di errore per la violazione di una regola di obbligatorio



Fonte: piattaforma Panda

Le regole di dominio consentono di imporre, per ogni variabile, l'insieme di valori ammessi. Il sistema può gestire le regole di dominio inibendo l'inserimento di valori non ammessi o segnalando al rispondente la violazione di dominio attraverso la visualizzazione di un messaggio di errore. Le regole di congruenza consentono di verificare che le risposte, fornite fino a quel momento, siano tra loro coerenti. Anche per quanto riguarda le regole di congruenza sono previsti due diversi livelli di errore: il livello bloccante che inibisce la prosecuzione della compilazione fino alla risoluzione dell'anomalia e il livello non bloccante che si limita ad avvertire l'utente di una possibile situazione di incoerenza tra le risposte fornite. Le regole di calcolo infine consentono di calcolare il valore di alcune variabili sulla base delle risposte già fornite dal rispondente.

Figura 1.7 - Messaggio di *warning* per la violazione di una regola di coerenza non bloccante

QUAL È LA SUPERFICIE DELL'ABITAZIONE? ⓘ
 CONSIDERARE LA SUPERFICIE INTERNA E CALPESTABILE ⓘ
 SOFFITTE, GARAGE, BOX AUTO ECC.
 Metri quadrati (arrotondati senza decimali)

⚠ Ha indicato una superficie molto piccola. Se l'informazione inserita è corretta proseguire; in caso contrario correggere.

Fonte: piattaforma *Panda*

Per il trattamento delle regole più complesse *Panda* offre la possibilità di includere nel questionario funzioni scritte in linguaggio Javascript che saranno eseguite sul client, o in linguaggio Java, la cui esecuzione avviene lato server.

Tutte le regole che governano il funzionamento del questionario vengono eseguite continuamente durante la compilazione del questionario e segnalano al rispondente il verificarsi di situazioni anomale per le quali è necessaria un'azione correttiva.

La segnalazione avviene attraverso la visualizzazione di uno specifico messaggio in corrispondenza del quesito in cui è stato rilevato l'errore. La tempestività della segnalazione, così come il contenuto personalizzabile del messaggio di errore, sono di particolare importanza in quanto consentono al rispondente di comprendere l'errore commesso e di eseguire immediatamente le opportune correzioni. Nel complesso, il meccanismo descritto semplifica la compilazione del questionario e influisce positivamente sul tasso di risposta e sulla qualità del dato.

Al termine della compilazione il sistema consente la produzione di una copia elettronica del questionario contenente tutte le risposte fornite.

Allo stesso modo il sistema è in grado di rilasciare una ricevuta elettronica contenente un identificativo univoco che attesta l'avvenuta compilazione. Entrambi i documenti sono prodotti in formato pdf.

1.1.3 Gestione dei testi e il supporto per il multilinguismo

Tutti i testi contenuti nel questionario vengono salvati su file distinti da quelli contenenti i metadati che definiscono la struttura del questionario. Ogni testo viene poi inserito nella posizione corretta attraverso un meccanismo di coppie chiave e valore. Nei file dedicati all'archiviazione dei contenuti testuali del questionario, oltre al semplice testo libero, *Panda* permette l'inserimento di marcatori speciali attraverso i quali è possibile intervenire sulla formattazione e arricchire il questionario con elementi speciali quali *tooltip*, link, immagini, liste, tabelle, valori calcolati, testi variabili e diversi altri elementi. Di seguito vengono presentati alcuni degli elementi usati frequentemente all'interno dei questionari elettronici e viene descritto il meccanismo utilizzato per la loro implementazione.

Un *tooltip* è un elemento grafico il cui scopo è quello di fornire informazioni aggiuntive, senza però appesantire la presentazione del questionario. Si presentano i *tooltip* all'interno del questionario attraverso un'icona verde che consente l'apertura di una piccola finestra informativa formattata utilizzando dimensioni e stili di carattere leggibile. La loro presenza all'interno dei questionari aiuta il rispondente a comprendere meglio i quesiti proposti, riducendo così il tempo richiesto per completare la compilazione. A livello di implementazione, l'elemento di tipo *tooltip* viene inserito direttamente nel testo del quesito, tramite il marcatore #INFO, nel file di *properties* contenente i testi del questionario.

1. Il questionario elettronico

Figura 1.8 - Esempio di utilizzo del marcatore #INFO

```
a5_2_1_domanda=Si tratta di:#BR#CLASS[suggerimento][Sono possibili più risposte]
a5_2_1__0=Box auto#INFO[Locale chiuso, adatto al ricovero di una o più
automobili, ed usato a questo scopo.]
...
```

SI TRATTA DI:
SONO POSSIBILI PIÙ RISPOSTE

Box auto

Posto auto al chiuso (in autorimessa collettiva, garage e simili)

Posto auto all'aperto riservato (condominiali, cortili e simili)

Fonte: piattaforma *Panda*

Spesso i questionari di indagine richiedono l'inserimento del valore di una variabile o di un valore calcolato all'interno del testo di un quesito o di un riquadro informativo. Tale valore in generale sarà funzione di quanto inserito dal rispondente durante la compilazione e dunque non noto in fase di progettazione del questionario.

A questo scopo il sistema *Panda* prevede l'utilizzo del marcatore #VALORE che può essere utilizzato con due diverse tipologie di parametri: se si utilizza il nome di una variabile, il sistema visualizza il valore della variabile richiamata e l'esecuzione avviene in tempo reale. Nel caso in cui nel marcatore sia invece contenuto il riferimento a una funzione, verrà mostrato il risultato della funzione scritta in linguaggio Javascript e attivata al momento dell'esecuzione del questionario. La funzione, che ha accesso a tutti i dati inseriti dal rispondente fino a quel momento, eseguirà una determinata sequenza di istruzioni restituendo in output il risultato che sarà visualizzato all'interno del questionario.

Figura 1.9 - Uso del marcatore #VALORE per l'introduzione di valori calcolati nel testo del quesito

```
a20_domanda=A.20 Rispetto alla Superficie Totale dell'Azienda dichiarata, pari a #VALORE[vsat] ettari,
qual è la superficie per i seguenti titoli di possesso dei terreni?
```

A.20 RISPETTO ALLA SUPERFICIE TOTALE DELL'AZIENDA DICHIARATA, PARI A 201,00 ETTARI, QUAL È LA SUPERFICIE PER I SEGUENTI TITOLI DI POSSESSO DEI TERRENI?

	Ettari	Are
Di proprietà o in usufrutto	150	50
In affitto	40	25
Ad uso gratuito	10	25
SUPERFICIE AGRICOLA TOTALE	201	0

Fonte: piattaforma *Panda*

I questionari in alcuni casi richiedono che uno stesso quesito venga presentato con formulazioni leggermente diverse a seconda del profilo dell'utente o delle risposte fornite in precedenza. Questo meccanismo viene indicato con il nome di testi variabili. *Panda* offre la possibilità di inserire nel questionario testi variabili attraverso il marcatore #SE che consente di modificare la formulazione di un quesito o di parte di esso al verificarsi di condizioni diverse. Anche l'operatore #SE viene valutato durante la compilazione del questionario e ha accesso a tutte le risposte fornite fino a quel momento.

Figura 1.10 - Uso del marcatore #VALORE per l'introduzione di valori calcolati nel testo del quesito

```
wsc3_domanda=Indicare se l'azienda:
wsc3_1=È stata fusa con altre aziende
wsc3_2=È stata smembrata o scissa in più aziende
wsc3_req=È necessario fornire una risposta

wsc4_domanda=#SE[wsc3_0][In quale anno l'azienda è stata fusa con altre
aziende?]<#SE[wsc3_1][In quale anno l'azienda è stata smembrata o scissa in più aziende?]
wsc4_0=Anno
```

INDICARE SE L'AZIENDA:

È stata fusa con altre aziende

È stata smembrata o scissa in più aziende

IN QUALE ANNO L'AZIENDA È STATA FUSA CON ALTRE AZIENDE?

Anno

Fonte: piattaforma *Panda*

L'elemento link, collegamento ipertestuale, ha lo scopo di condurre a ulteriori unità informative per approfondire un argomento o per soddisfare delle esigenze ritenute importanti come supporto a chi compila il questionario. Il sistema di acquisizione *Panda* consente l'inserimento di collegamenti ipertestuali a pagine web o a documenti attraverso il marcatore #LINK. Questo prevede diverse sintassi che consentono di indicare se l'indirizzo della risorsa collegata debba essere visibile o mascherato da un test alternativo, se la risorsa collegata debba essere aperta nella stessa finestra o in una nuova finestra. Nei questionari elettronici i link vengono contraddistinti attraverso una diversa formattazione del testo dell'ancora e l'attivazione avviene attraverso un clic del mouse. Con una sintassi analoga a quella del collegamento ipertestuale è possibile inserire all'interno del questionario collegamenti a indirizzi di email per l'assistenza alla compilazione.

Per la formattazione del testo sono messi a disposizione un insieme di marcatori generici predefiniti, quando questi non sono sufficienti a definire le proprietà grafiche desiderate è possibile ricorrere al marcatore #CLASS che consente di attribuire a un elemento testuale una classe CSS (*Cascading Style Sheet*) che potrà poi essere definita in un apposito file css da collegare al questionario elettronico. In questo modo chi si occupa dello sviluppo del questionario avrà il completo controllo sull'aspetto grafico del testo visualizzato. Tutti i marcatori presentati possono essere combinati tra loro in modo arbitrario. È possibile, ad esempio, inserire un link all'interno di un *tooltip*, o un valore calcolato all'interno di un marcatore #CLASS.

L'esperienza ha mostrato che la formattazione del testo è un aspetto estremamente importante per rendere il questionario elettronico più chiaro e leggibile e per consentire al rispondente di cogliere con immediatezza le informazioni più importanti. La revisione grafica dei questionari elettronici, prima dell'avvio della fase di raccolta dati, è un processo tanto delicato quanto fondamentale allo scopo di ottenere un prodotto finale che risulti facilmente fruibile e graficamente accattivante.

Il sistema generalizzato *Panda* prevede il supporto per il multilinguismo.

Ogni questionario può essere tradotto in un numero arbitrario di lingue semplicemente aggiungendo un nuovo file di testi contenenti la stessa struttura di coppia chiave e valore per ogni nuova lingua. La traduzione del questionario in una nuova lingua avviene semplicemente aggiungendo un nuovo file dei testi in cui i testi originali sono stati sostituiti con la traduzione nella nuova lingua.

1. Il questionario elettronico

Per facilitare la compilazione dei questionari sono state introdotte le lingue francese, tedesco e sloveno riconosciute e tutelate dalla normativa italiana dall'articolo 6 della Costituzione, che riconosce in via di principio la tutela delle lingue minoritarie presenti sul territorio nazionale e la lingua inglese in quanto riconosciuta come la lingua internazionale più comunemente parlata nel mondo.

Questa caratteristica del sistema da una parte semplifica le attività di traduzione del questionario in lingue diverse, dall'altra consente al rispondente di selezionare la lingua desiderata direttamente al momento della compilazione.

Al termine della compilazione il rispondente potrà acquisire copia del questionario compilato o della ricevuta di avvenuta compilazione in formato elettronico e tradotto in una delle lingue messe a disposizione dal questionario.

1.1.4 Integrazione con gli altri sistemi

L'applicazione web che costituisce il questionario sviluppata attraverso il sistema *Panda*, si integra con gli altri sistemi presenti in Istat, come ad esempio il Sistema di gestione delle indagini (Sgi) o il Portale delle imprese. Durante la fase di raccolta dati, infatti, è necessario che il questionario elettronico interagisca con le altre applicazioni scambiando informazioni al verificarsi di determinati eventi. Può essere, ad esempio, richiesta la creazione di un nuovo questionario per una nuova unità da rilevare che non era stata inclusa nel campione iniziale. Questa operazione viene eseguita da un operatore attraverso il sistema di gestione e inviata in tempo reale al sistema che gestisce i questionari. Quest'ultimo crea il nuovo questionario e lo mette a disposizione del rispondente per la compilazione. La comunicazione è biunivoca: il questionario elettronico invia, infatti, una notifica al sistema di gestione ogni volta che un questionario viene consultato, modificato o compilato. In questo modo, il sistema di gestione è sempre aggiornato sull'andamento della rilevazione.

Questa interazione si basa su un insieme di operazioni elementari implementate attraverso microservizi. L'adozione dei microservizi permette di mantenere l'indipendenza dei due sistemi e, allo stesso tempo, garantisce un meccanismo di comunicazione efficiente e per lo più privo di errori. L'architettura a microservizi scelta per la comunicazione è uno standard ampiamente diffuso e dunque predispone i sistemi anche al colloquio con eventuali altre applicazioni interne o esterne.

1.2 Gestione dei dati

Il sistema di acquisizione *Panda* si appoggia su *Oracle*, una base di dati di tipo relazionale. Durante la fase di raccolta, il numero di utenti che, con profili anche diversi, accede ai dati è generalmente piuttosto elevato quindi, per evitare problemi di interferenza, i questionari sviluppati con *Panda* implementano, a livello di database, un meccanismo di *lock* sui singoli questionari. Questo assicura che utenti diversi non accedano contemporaneamente allo stesso questionario evitando così perdite di dati dovuti alla concorrenza degli accessi. Oltre ad acquisire i dati relativi alla compilazione del questionario, ogni questionario elettronico registra diverse informazioni utili sia per monitorare l'andamento della rilevazione sia per valutare la *user experience* dei rispondenti.

1.2.1 Architettura della base di dati

La struttura dei dati prevede un insieme di tabelle condivise da tutte le indagini e altre tabelle che invece sono specifiche per ogni singola indagine. Tra le prime abbiamo ad esempio quelle dedicate all'anagrafica dei rispondenti, alla gestione delle autorizzazioni all'accesso dei questionari e alla memorizzazione dei paradatai raccolti durante la compilazione.

Le tabelle di anagrafica contengono la lista completa dei rispondenti che possono accedere a uno o più questionari elettronici e che dunque sono autorizzati a partecipare a una o più indagini. Ogni rispondente è contraddistinto da uno *username* e da un proprio codice identificativo rappresentato da un numero progressivo univoco. Questa gestione dell'anagrafica, da una parte assicura che ogni rispondente venga inserito una sola volta nella base di dati del sistema a prescindere dal numero di indagini a cui sarà chiamato a partecipare nel corso del tempo; dall'altra consente di ricostruire lo storico delle indagini a lui destinate. L'associazione tra rispondenti, indagini e questionari è rappresentata per mezzo di una tabella dedicata. In questo modo ogni rispondente può essere abilitato all'accesso e alla compilazione di uno o più questionari elettronici che possono fare riferimento alla stessa indagine o a indagini diverse.

Le tabelle dedicate alla memorizzazione dei microdati, ossia delle risposte fornite dai rispondenti ai quesiti proposti, sono invece specifiche per ogni questionario. Come già descritto nei paragrafi precedenti, il sistema *Panda* prevede la definizione del questionario attraverso un insieme di file di metadati. Poiché i file di metadati sono peculiari a una indagine e differiscono da rilevazione a rilevazione, anche le tabelle dei microdati, definite in base al *set* di file di metadati, saranno differenti per numero e struttura da indagine a indagine. Per ogni questionario elettronico in generale è prevista una tabella principale e una o più tabelle collegate. La prima tabella contiene i microdati relativi al flusso principale del questionario, mentre nelle tabelle collegate saranno memorizzati i microdati relativi a eventuali sezioni del questionario che si ripetono. Alcune indagini, infatti, prevedono che il questionario elettronico richieda al rispondente di compilare più volte uno stesso insieme di quesiti, in base ad alcune delle informazioni precedentemente inserite. Ciò accade anche nel questionario del Censimento permanente della Popolazione e delle abitazioni, in cui, ad esempio, è presente una sezione dedicata ai singoli componenti della famiglia, che prende il nome di "scheda individuale". La scheda individuale è composta da un insieme di quesiti ripetuti per ogni componente e dovrà essere compilata un numero di volte pari al numero dei componenti della famiglia. L'integrità referenziale tra la tabella dei microdati principale e le tabelle collegate è garantita dalla presenza di chiavi esterne, ovvero da una colonna, appartenente a tutte le tabelle collegate il cui valore corrisponde alla chiave primaria della tabella principale.

La definizione delle tabelle dei microdati, nel caso di questionari con molti quesiti di diverse tipologie, può diventare piuttosto complessa. Il sistema *Panda* è stato progettato anche per definirne in automatico la struttura delle tabelle dedicate alla raccolta dei microdati e di produrre le istruzioni necessarie alla generazione delle tabelle, degli indici, dei vincoli di integrità e dei *trigger*. Questo è un grosso valore aggiunto che oltre a far risparmiare tempo allo sviluppo evita gli errori causati da una gestione manuale.

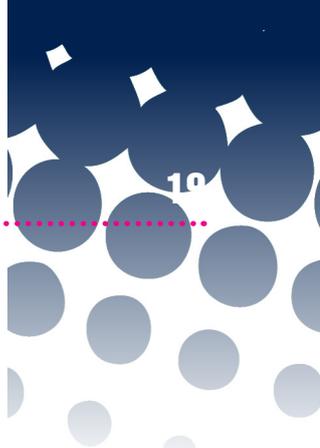
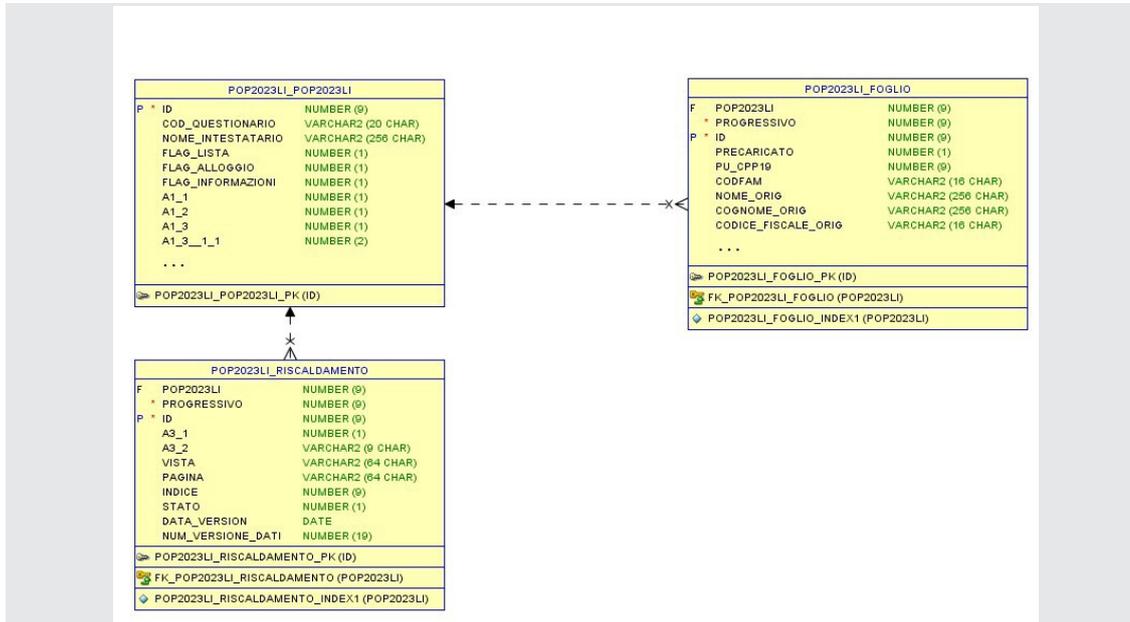


Figura 1.11 - Tabelle per la gestione dei microdati



Fonte: database Oracle

1.2.2 Paradati raccolti durante la compilazione

I questionari elettronici sviluppati attraverso il sistema *Panda* consentono di registrare, durante la fase di raccolta dati, diverse informazioni relative alla interazione tra il rispondente e il questionario elettronico. L'insieme di queste informazioni e la loro analisi durante la rilevazione o a rilevazione conclusa, fornisce un *feedback* estremamente interessante relativo alla *user experience* avuta dagli utenti chiamati a partecipare all'indagine. I dati così raccolti prendono il nome di paradati di compilazione e vengono memorizzati in modo del tutto automatico e trasparente. I paradati sono salvati nella base di dati di *Panda* e organizzati in tabelle diverse per tipologia. Tra le informazioni tracciate abbiamo gli istanti in cui viene effettuato il *login* o il *logout* dal sistema, il numero dei salvataggi intermedi, il tempo impiegato per la compilazione delle diverse pagine di cui si compone il questionario elettronico, l'istante in cui viene effettuata la chiusura del questionario o invio definitivo. Attraverso l'analisi di questo primo set di paradati è possibile calcolare i tempi necessari alla compilazione del questionario o il numero medio di sessioni di lavoro richieste per il completamento dell'intervista. Allo stesso tempo è possibile indentificare le sezioni del questionario che hanno richiesto più tempo per la loro compilazione e che dunque è ragionevole ipotizzare siano risultate più complesse. In presenza di un numero elevato di questionari incompleti è anche possibile avviare delle ricerche sui paradati mirate a individuare la causa o le cause delle interruzioni della compilazione.

Oltre a quelli già descritti, il sistema raccoglie anche altri paradati che riguardano la tipologia di dispositivo utilizzato per la compilazione (pc, smartphone, tablet o altri dispositivi), la lingua selezionata, la tecnica di rilevazione, la documentazione consultata o gli strumenti utilizzati attraverso i collegamenti messi a disposizione del rispondente. Anche l'analisi di questi ulteriori di informazioni risulta estremamente utile al fine di definire quali ottimizzazioni alla struttura o alla presentazione del questionario sia opportuno pianificare nelle edizioni successive di indagine.



Nel complesso l'analisi dei paradata è da considerare a tutti gli effetti una fase rilevante del ciclo di vita dei questionari elettronici in quanto ne determina in parte l'evoluzione sulla base di un effettivo *feedback* ricevuto attraverso l'analisi del comportamento dei rispondenti e dei dispositivi utilizzati per la compilazione.

1.3 Sviluppi futuri

Il sistema *Panda* è attualmente oggetto di un processo di reingegnerizzazione che, da una parte mira ad aumentarne la produttività attraverso la realizzazione di una interfaccia web per la costruzione visuale dei questionari, dall'altra ha l'obiettivo di aggiornarne le tecnologie utilizzate per la produzione del questionario elettronico, migliorandone la fruibilità e le caratteristiche di accessibilità.

Come descritto in precedenza, *Panda* produce l'applicazione web che implementa un questionario elettronico a partire da un input costituito da un insieme di file strutturati in formato xml. La realizzazione dei file di input e la composizione delle espressioni logico matematiche necessarie alla definizione delle regole del questionario richiedono competenze in ambito informatico. Per il trattamento delle regole e dei quesiti più complessi, inoltre, può essere necessario estendere alcuni moduli tramite codice Java o Javascript. Per questi motivi la produzione dei questionari con l'attuale sistema necessita l'intervento di personale specializzato appartenente alla Direzione dell'informatica. L'adozione di una interfaccia web per la costruzione visuale dei diversi componenti del questionario e delle regole che ne determinano il comportamento consentirà l'utilizzo dello strumento anche da parte di utenti privi di conoscenze tecniche avanzate, aumentando la produttività del sistema.

Il questionario prodotto in output dal sistema *Panda* si basa su tecnologie ormai superate e nonostante gli sforzi fatti nel corso degli ultimi anni, alcuni dei quesiti prodotti, quali ad esempio quelli in forma tabellare, non risultano completamente accessibili. L'aggiornamento di queste tecnologie potrà rendere il sistema completamente accessibile e allo stesso tempo migliorerà la *user experience* dei rispondenti che sempre più spesso adottano smartphone o altri dispositivi mobili per la compilazione dei questionari.

La reingegnerizzazione in corso prevede anche l'introduzione di nuove tecniche di crittografia per incrementare il livello di sicurezza del sistema e di riservatezza dei dati acquisiti.

2. IL SISTEMA DI GESTIONE DELLE INDAGINI (SGI)¹

Arrivare in modo capillare sul territorio è una condizione fondamentale per la buona riuscita delle indagini censuarie. Gli organi territoriali hanno infatti da sempre avuto un ruolo attivo nella comunicazione con i cittadini, ma anche hanno supportato la rilevazione tramite, ad esempio, il recupero delle mancate risposte e i centri di assistenza alla compilazione.

Il passaggio dal Censimento decennale al Censimento permanente costituisce un elemento di innovazione metodologica e nello stesso tempo ha reso questo impegno costante nel tempo. Per rendere più agevole questo passaggio e per facilitare il lavoro al territorio, l'Istat da subito ha puntato sulle tecnologie web e sulla digitalizzazione dei processi. Tale innovazione metodologica, però, se non supportata da una forte innovazione tecnologica, avrebbe potuto causare un impegno troppo oneroso per la rete di rilevazione chiamata a svolgere il lavoro annualmente. Per questo motivo l'Istat, fino dal 2018, anno in cui si è svolta la prima *wave* del Censimento permanente della Popolazione e delle abitazioni, ha progettato la piattaforma denominata, Sistema di gestione delle indagini (Sgi) che ha l'obiettivo di digitalizzare non solo la fase di conduzione sul campo, ma anche tutte quelle operazioni a latere che sono di supporto per la buona riuscita della rilevazione. La digitalizzazione ha riguardato le seguenti fasi:

- la creazione e la gestione della rete di rilevazione con la possibilità di assegnare i carichi di lavoro ai diversi utenti. Il responsabile della rete di rilevazione può agire in totale autonomia ed è completamente indipendentemente nell'inserimento delle persone che seguiranno l'indagine; assegna inoltre a ognuno un ruolo che governa la visibilità dei dati e delle funzioni all'interno dell'applicazione. Ogni utente inserito nel Sistema è automaticamente aggiunto nella banca dati dell'Istat ed è accreditato in automatico su tutti i sistemi messi a disposizione per l'indagine. Le persone inserite ricevono comunicazioni con le credenziali di accesso direttamente dal Sistema sulle email personali comunicate. La piattaforma consente anche di organizzare il lavoro della rete di rilevazione assegnando le unità eleggibili agli utenti che devono lavorarle;
- la gestione della formazione con la possibilità di creare eventi, assegnare gli uffici di rilevazione ai corsi a loro designati e permettere al responsabile della rete di inserire le persone che ritiene necessario e monitorare le presenze agli eventi. La digitalizzazione della formazione è iniziata in fase pre Covid-19 con lo scopo di permettere all'Istat di creare gli eventi formativi in presenza e di stampare i fogli firma e, agli uffici di Censimento, di registrare i propri utenti ai diversi eventi. Gli stessi utenti potevano utilizzare anche la piattaforma di formazione online dell'Istat in quanto l'iscrizione era automatica anche su quest'ultima. Nel periodo post Covid-19 questo modulo è stato rivoluzionato sulla base delle nuove esigenze emerse e ha permesso nel giro di poco tempo di lavorare con le aule virtuali integrando l'accreditamento della piattaforma *Microsoft Teams*. Ogni evento formativo, quindi, ha potuto includere personale proveniente da diversi uffici e residente in aree anche molto distanti tra loro, grazie alla possibilità di creare classi online fruibili attraverso altri canali digitali messi a disposizione dall'Istat;

¹ Il Capitolo è stato redatto da Eleonora Sibilio (paragrafi 2.1, 2.6, 2.7 e 2.8), Maura Giacommo (paragrafi 2 e 2.9), Emanuele Bolle (paragrafo 2.2), Serena Pizzicannella (paragrafo 2.3), Giuseppe Giuliano (paragrafi 2.4 e 2.5), Alessandra Pizzella (paragrafo 2.8.1).

- la gestione della rilevazione sul campo ossia di tutte le funzioni connesse alla fase di conduzione vera e propria. Il cuore di questo gruppo di funzioni è costituito dal *Diario*, un cruscotto che racchiude al suo interno le operazioni necessarie per compilare il questionario, per segnalare eventuali lavorazioni fatte sulle unità, per eliminare o aggiungere le stesse o ripristinare i dati inseriti. Ciascuna tipologia di utente ha un proprio cruscotto di lavoro che attiva e disattiva le azioni in base allo stato di lavorazione delle unità. Questo delicato meccanismo, basato su un processo di lavoro guidato, caricato nel Sistema prima dell'inizio dell'indagine, permette al rilevatore di muoversi in autonomia evitando il più possibile errori manuali che potrebbero compromettere la qualità del dato;
- la gestione della comunicazione tra l'Istat e la rete di rilevazione che permette al personale dell'Istituto, se autorizzato, di inserire messaggi personalizzati per classi di utenza e visualizzabili sulla *home page*. Considerando la grande capillarità della rete e l'importanza della tempestività nel dare l'informazione, le *news* online hanno costituito un importante mezzo di comunicazione e di diffusione di aggiornamenti per tutti gli operatori dei vari uffici. Questa funzionalità non sostituisce le comunicazioni ufficiali, ma ha il grande vantaggio di potere veicolare in tempo reale tutte le tipologie di informazioni che si reputano necessarie. Nel corso di ogni indagine sono state inserite note tecniche sulle modalità di lavoro da adottare, tempi e scadenze relativi all'indagine o avvisi su eventi imprevisti.

Accanto a queste funzioni, che permettono la digitalizzazione dell'indagine, sono state sviluppate una serie di funzionalità di configurazione del Sistema che hanno lo scopo di rendere più veloce la creazione delle nuove indagini e di permettere a un utente interno all'Istat, non informatico, di caricare le informazioni che caratterizzano l'edizione di riferimento. Questo processo è in ancora in fase di sviluppo e allo stato attuale permette ad alcuni utenti amministrativi di creare una nuova edizione, configurare la rete di rilevazione e caricare i rispettivi uffici che ne fanno parte, copiare le informazioni di gestione da un'edizione a un'altra, definire il processo di lavorazione e la visibilità dei diversi profili che compongono la rete.

Grazie alle scelte di progettazione e di tecnologia, Sgi permette di gestire reti di rilevazione molto diverse tra loro in termini di visibilità e composizione. La configurazione della rete di rilevazione e dei rispettivi uffici che ne fanno parte determina la visibilità dei dati all'interno del Sistema in modo che ogni ufficio veda solo i dati di propria competenza.

Di seguito sono descritte le reti di rilevazione del Censimento permanente della Popolazione e delle abitazioni, del Censimento generale dell'Agricoltura e del Censimento permanente delle istituzioni pubbliche, che, per le loro caratteristiche, sono rappresentative di gran parte delle reti e quindi possono essere utilizzate anche per configurare altre indagini.

Nello specifico, per la realizzazione del Censimento permanente della Popolazione e delle abitazioni, la rete di rilevazione è sviluppata, secondo quanto stabilito dal Piano generale del Censimento (PGC), sui seguenti livelli di governo territoriali:

1. livello nazionale, in cui operano diversi uffici come la stessa Istat (rete interna) ma anche il Contact Centre (rete esterna);
2. livello regionale, costituito sia da uffici Istat, con gli uffici territoriali di Censimento, sia dall'ufficio della Regione autonoma della Valle d'Aosta/*Vallée d'Aoste*;
3. livello provinciale, in cui hanno operato gli uffici provinciali di Censimento (UPC), uffici di statistica delle prefetture, e le Province autonome di Trento e Bolzano/*Bozen*;
4. livello comunale, in cui sono stati costituiti gli uffici comunali di Censimento (UCC), in forma autonoma o associata.

2. Il Sistema di gestione delle indagini (Sgi)

Ognuno dei livelli descritti, a eccezione di quello nazionale, può includere al suo interno ulteriori aggregazioni necessarie in tutti quei casi dove i confini amministrativi non sono sufficienti: si pensi ad esempio alle unioni dei comuni e ad alcuni uffici di prefettura che non corrispondono con le province.

Un elemento di complessità ulteriore, rispetto a quanto descritto per il Censimento della Popolazione e delle abitazioni, è stato introdotto con il Censimento generale dell'Agricoltura a causa della particolare geografia della rete di rilevazione. Più nel dettaglio, erano previsti i Centri di assistenza agricola (CAA), caratterizzati da una varia diffusione sul territorio nazionale. Accanto a questa peculiarità delle sedi operative era prevista la presenza di uffici di coordinamento, anch'essi distribuiti sul territorio nazionale, e con aree di sovrapposizione. Il Sistema deve quindi prevedere la gestione di una rete non territoriale di due differenti livelli di uffici e una gerarchia di visibilità. Questo ha costituito la maggiore criticità adattativa che è stata risolta da un'associazione puntuale degli uffici operativi dei CAA con le unità di rilevazione, l'unica in grado di rappresentare la reale visibilità degli uffici. Questa impostazione è stata utilizzata anche nel Censimento permanente delle istituzioni non profit.

L'ulteriore difficoltà, caratterizzata dal ruolo di coordinamento degli uffici centrali di Censimento sugli uffici distribuiti sul territorio, è stata superata imponendo nel Sistema la possibilità di avere sia le gerarchie territoriali sia quelle imposte in fase di configurazione dei dati. Questa nuova esigenza ha fatto sì che Sgi si sia evoluto per permettere agli uffici di livello superiore di gestire rete e lavoro e di monitorare in modo attivo non solo i propri uffici, ma anche quelli subordinati.

Anche nel Censimento permanente delle istituzioni pubbliche il Sistema ha dovuto gestire una rete non standard; l'ufficio di Censimento in questo caso è rappresentato dall'unità istituzionale e, per la prima volta, il rispondente è entrato direttamente nel Sistema di gestione delle indagini. Inoltre, le aggregazioni territoriali sono determinate dalle unità locali che vi appartengono.

Indagini complesse, come quelle censuarie, possono quindi possedere diverse tipologie di uffici, coincidenti o meno con il territorio, che governano il Sistema e che possono essere legate tra loro anche da vincoli di gerarchia. L'accesso orizzontale sui dati è ulteriormente affinato da una visibilità sulle funzioni del Sistema, in modo tale da potere gestire diverse figure professionali, con specifici profili, all'interno di uno stesso ufficio; nel Piano Generale di Censimento della Popolazione e delle abitazioni, ad esempio, i ruoli e i compiti tra i diversi attori sono ben descritti e questi devono trovare una corrispondenza all'interno di Sgi. In ogni ufficio del Sistema è sempre possibile inserire un responsabile e un personale di *staff*, mentre per gli uffici che effettuano la rilevazione, oltre a queste due figure, è possibile aggiungere i coordinatori, i rilevatori e il personale di *back office*. Ognuna di queste figure, se presenti, possono nel Sistema prendere visione delle unità di rilevazione del proprio ufficio. Il responsabile dell'ufficio crea la propria rete e monitora il lavoro degli altri, avvalendosi, se necessario del personale di *staff* e del *back office*; quindi, queste figure, se presenti, hanno nel Sistema visibilità di tutte le unità di rilevazione del proprio ufficio. Le figure del coordinatore e del rilevatore invece hanno una visibilità dei dati più limitata: in particolare il rilevatore può monitorare e agire sulle sole unità che gli vengono assegnate, mentre il coordinatore ha il controllo dei dati dei rilevatori assegnatigli. Si crea quindi non solo una gerarchia tra uffici ma anche tra i ruoli. La visibilità ridotta per alcuni ruoli, oltre a essere una esigenza della rete di rilevazione, ha lo scopo di fare vedere a ogni utente le sole unità che deve lavorare, evitando possibili errori e un appesantimento nella visualizzazione dei dati.

La gestione della rete di rilevazione rappresenta quindi una delle maggiori complessità in

un Sistema che vuole garantire la massima flessibilità e autonomia nella conduzione dell'indagine, senza tuttavia fare perdere il controllo e, soprattutto, senza compromettere la qualità del dato e del microdato inviato. L'attuale sistema permette di gestire sia reti territoriali con confini amministrativi più o meno delineati, sia reti costruite come aggregati delle unità di rilevazione. Ogni ufficio inoltre può essere costruito in modo autonomo: ciò significa che possono coesistere uffici territoriali, come i comuni, e uffici formati invece da aggregazioni diverse.

Nei paragrafi che seguono viene analizzato nel dettaglio l'aspetto progettuale e tecnico di Sgi.

2.1 Digitalizzazione del processo

Sgi è un'applicazione web, ad accesso sicuro, utilizzabile da tutti i *browser* più moderni; per accedere è necessario inserire il nome utente e la *password* forniti all'utente. *Shibboleth* è l'*Identity provider* ufficiale dell'Istat che gestisce l'autenticazione degli utenti, sia su *Active Directory* sia su *Lightweight Directory Access Protocol* (LDAP), e garantisce il *single sign-on* (SSO), ossia permette agli utenti già autenticati su Sgi di accedere senza la necessità di reinserire le credenziali anche alle altre applicazioni che confluiscono nell'ecosistema digitale della raccolta dati, quali ad esempio *Panda*, *Microstrategy*, *Apex*, *Portale della rete*, *Moodle* e *Teams*.

È inoltre possibile utilizzare il Sistema pubblico di identità digitale, indicato con Spid o la Carta di identità elettronica, chiamata anche Cie, secondo quanto previsto dalle linee guida dell'Agid. In questo modo tutti i membri della rete di rilevazione territoriale hanno la possibilità di scegliere quale tipo di autenticazione effettuare, mentre il personale interno accede direttamente con le credenziali uniche rilasciate dall'Istituto.

Dopo l'autenticazione, il Sistema permette di scegliere tra le diverse indagini su cui l'utente è abilitato, garantendo un unico punto di ingresso; una volta effettuata la scelta, attraverso le voci di un menu principale, sarà possibile accedere a diverse funzionalità disponibili per la gestione e conduzione della singola indagine che si deve lavorare. Ogni indagine ha infatti le proprie funzionalità e le proprie regole di visibilità sui dati: ognuna si differenzia dall'altra sia per la tipologia di dati sia per modalità di conduzione delle varie fasi. Inoltre ognuna presenta una propria rete di rilevazione formata da diversi profili, ciascuno con precise peculiarità.

L'utente che si autentica su Sgi può accedere alle pagine solo se autorizzato. Tale autorizzazione è determinata da una serie di metadati presenti sul database: essi, infatti, sono progettati per consentire un complesso processo di autorizzazioni legate al profilo assegnato all'utente stesso. Ogni singola pagina è configurata sul server in modo tale che solo le pagine che rispettano determinati prerequisiti vengono aperte.

Altre caratteristiche che vengono utilizzate per la concessione delle autorizzazioni sono quelle legate a ogni singola unità del campione: il server ad esempio controlla, in base all'esito di lavorazione posseduto dalla stessa unità, quali operazioni sono consentite e quali non lo sono e invia al client solo la lista delle azioni permesse.

Inoltre, Sgi è un'applicazione *metadata driven*, ossia la sua configurazione avviene attraverso metadati e questo consente di avere delle specializzazioni sui filtri di ricerca che possono pertanto essere scelti per ogni tipologia di indagine, senza che ciò comporti modifiche sul codice *back end*. Il lavoro viene infatti svolto tutto da un motore *back end* al quale è delegata la creazione e l'esecuzione della richiesta. Anche tale motore è stato realizzato in modo completamente svincolato dall'indagine e dall'output richiesto: esso riceve in input i metadati, li elabora ed è così in grado di generare *query* di varia complessità. L'integrazione con il software di monitoraggio e con il Sistema digitale di raccolta dati (questionario elettronico *Panda*) ga-

2. Il Sistema di gestione delle indagini (Sgi)

rantisce inoltre la possibilità di controllare l'andamento delle operazioni di intervista e di rendicontazione del lavoro svolto da ciascun rilevatore sul campo in tempo reale. Tecnicamente la comunicazione tra i sistemi avviene tramite microservizi esterni che completano l'architettura del Sistema e consentono una integrazione con le altre parti. Esempi di microservizi sono:

- *Utenti*, per la gestione degli inserimenti, le modifiche sulle utenze della rete di rilevazione e per la gestione delle autenticazioni;
- *Territorio*, per l'accesso centralizzato alle informazioni riguardanti il territorio nazionale ed estero, come ad esempio comuni e Stati esteri di nascita, o lo storico dei comuni italiani;
- *Sgi-API e Questionario-API*, per la comunicazione bidirezionale con il questionario elettronico.

A fronte del numero sempre crescente di indagini e della necessità di gestire le conseguenti numerose funzionalità, è stato necessario abbandonare l'architettura *Model-View-Controller* (MVC) basata su *server-side rendering* a favore di una applicazione più agile che non gravasse eccessivamente sul *back end*, in grado di adattarsi automaticamente alle indagini passate e future senza la necessità di riprogettarne ogni volta il codice. Tutto questo si è concretizzato nel rilascio di un software sufficientemente agile da garantire tempi di sviluppo e manutenzione accettabili malgrado l'enorme quantitativo di interfacce da produrre ciclicamente.

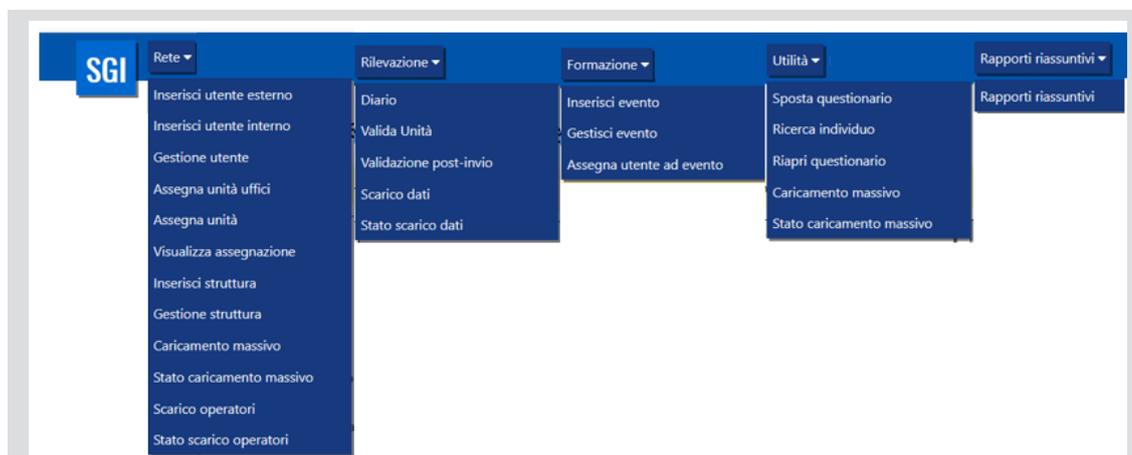
Sgi è una *Single page application* con due motori applicativi indipendenti, uno lato *front end* uno lato *back end*, che permettono di creare interfacce utente dinamiche e flessibili, consentendo la configurazione e la personalizzazione dell'interfaccia senza richiedere modifiche dirette al codice sorgente. La gestione delle interfacce è a carico del *front end* rendendo di fatto Sgi un'applicazione *client-side rendering*. Ciascuno di questi aspetti permette di dedicarsi in maniera distinta e indipendente allo sviluppo e manutenzione di dati, metadati e aspetti grafici, riducendo di conseguenza la generazione di errori e di conflitti tipici dei sistemi centralizzati.

La Figura 2.1 mostra un esempio di possibile menu del Sistema di gestione delle indagini, configurabile in base all'indagine e al ruolo dell'utente che vi accede.

Ogni gruppo di funzionalità, quindi, ha lo scopo di digitalizzare un sotto processo di conduzione e, per ogni indagine, si può decidere quali funzionalità utilizzare in base alle proprie esigenze.

Nei prossimi paragrafi si approfondisce la parte relativa all'architettura software e quella delle funzionalità applicative.

Figura 2.1 - Esempio di menu di Sgi

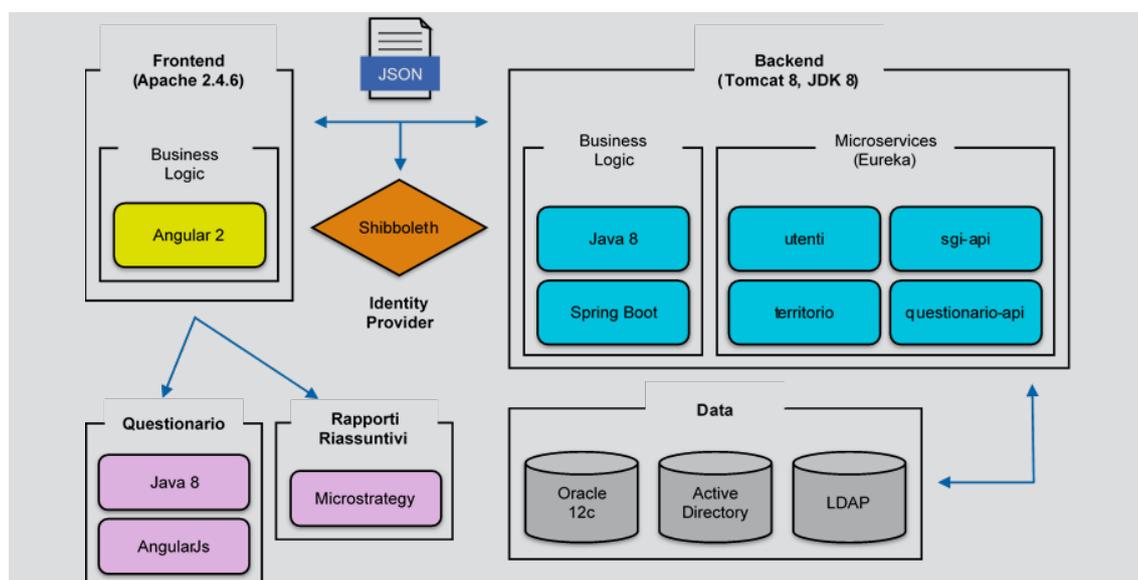


Fonte: Istat, Sistema di gestione delle indagini (Sgi)

2.2 Progettazione della piattaforma

L'applicazione Sgi adotta un approccio innovativo utilizzando json statici definiti con metalinguaggio e interpretati da *Angular* per la creazione dell'interfaccia utente. I dati vengono popolati nell'interfaccia tramite *restfull Application Programming Interface*, denominate API, e microservleti. L'architettura è progettata per garantire scalabilità, modularità e un'autenticazione sicura in Sso. Questa architettura è schematizzata in Figura 2.2.

Figura 2.2 - Architettura di Sgi



Fonte: Istat

Al primo accesso l'utente deve modificare la propria *password*, precedentemente fornita dall'Istat. Può inoltre recuperarla in ogni momento, in caso di smarrimento, o cambiarla. Queste attività sono permesse grazie all'integrazione con il microservizio esterno *Utenti*, che fornisce tutte le API necessarie per la gestione dell'autenticazione della rete di rilevazione. Una volta autenticati, gli utenti autorizzati avranno a disposizione un elenco di indagini da selezionare. Dopo avere scelto l'indagine, il menu principale si popolerà con un *set* di funzionalità disponibili, filtrate in base al profilo posseduto, alle quali attingere per la conduzione e la gestione dell'indagine.

Uno degli aspetti fondamentali di Sgi è l'integrazione con la piattaforma *Panda* con la quale condivide l'utilizzo del microservizio *Territorio* e la cui comunicazione è permessa rispettivamente dai microservizi *Sgi-API* e *Questionario-API*. Questi microservizi esterni completano l'architettura del Sistema, consentendo una interazione stretta tra Sgi e *Panda* durante la fase di raccolta dei dati, permettendo tra l'altro il coordinamento di eventi come la modifica dello stato di un questionario o il suo blocco. L'interazione tra Sgi e *Panda* può avvenire anche in direzione opposta, con il questionario elettronico che comunica a Sgi ogni volta che viene acceduto, modificato o inviato un questionario. Questo meccanismo di comunicazione mantiene Sgi costantemente aggiornato sull'andamento del Censimento. L'implementazione di operazioni basate su microservizi consente a tutti i sistemi, interni o esterni, legati a Sgi di interagire in modo efficace, mantenendo la separazione delle responsabilità e garantendo una comunicazione efficiente e affidabile.

2. Il Sistema di gestione delle indagini (Sgi)

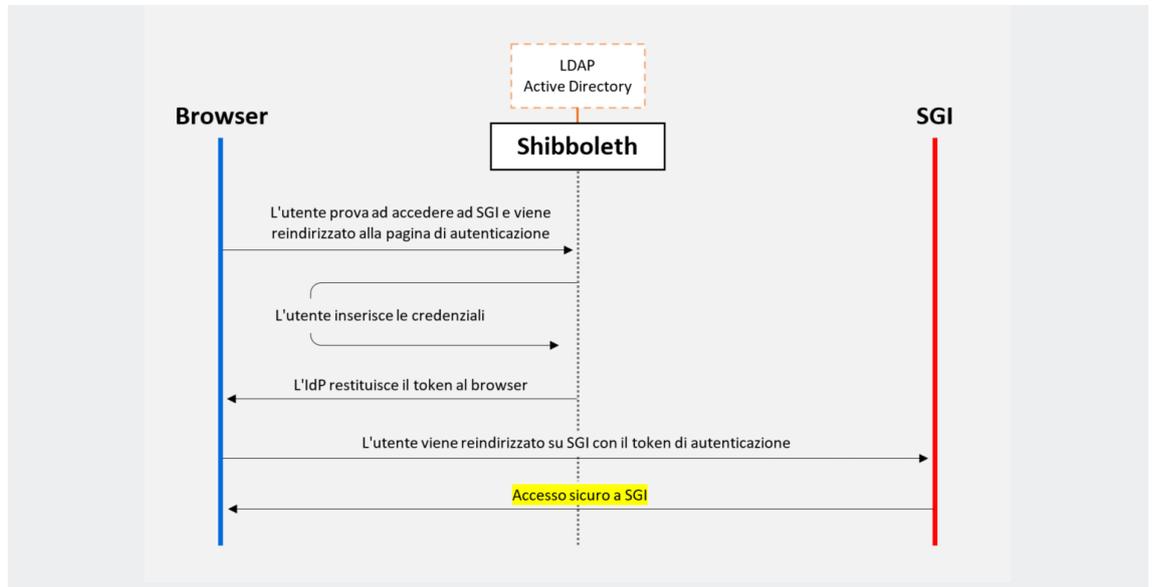
I componenti principali, che fungono da pilastri a questa architettura, sono stati scelti con attenzione per dare al Sistema longevità e robustezza, preferendo a prodotti noti il passaggio a tecnologie di nuova generazione.

La scelta non si è basata solo sulla qualità del prodotto quanto sulla sinergia tra i singoli componenti. Più nel dettaglio, si parla di:

- *Shibboleth*: che verifica in collaborazione con *Active Directory* e LDAP l'autenticazione degli utenti. Questo approccio garantisce un accesso sicuro all'applicazione, consentendo agli utenti di autenticarsi una sola volta per accedere a diverse risorse e servizi;
- *Angular 15*: l'interfaccia utente dell'applicazione è sviluppata utilizzando il potente *framework Angular*. Questo *front end* reattivo utilizza un approccio innovativo, sfruttando json statici definiti con metalinguaggio per delineare la struttura e la composizione dell'interfaccia. Questi json contengono informazioni sulle componenti dell'interfaccia, i campi dei moduli, i *layout* e le validazioni. Il vero punto di forza, che rende particolarmente innovativo l'utilizzo del *framework*, è rappresentato da un motore applicativo personalizzato, sviluppato internamente, che interpreta i json in tempo reale. Questo approccio consente la creazione dinamica dell'interfaccia utente, offrendo una flessibilità senza pari nell'adattamento dell'esperienza utente alle esigenze specifiche;
- *Spring boot*: il cuore operativo dell'applicazione risiede nel *back end* sviluppato utilizzando il *framework Spring boot*. Questo *back end* agisce come il motore che alimenta le operazioni dell'applicazione. Gestisce la comunicazione tra l'interfaccia utente e il Sistema sottostante, fornendo *restfull API* e microservizi. Uno dei punti focali del *back end* è la sua capacità di aggregare dinamicamente dati provenienti da varie fonti. Attraverso un motore applicativo dedicato, il *back end* combina dati provenienti da sorgenti diverse per presentarli in modo coeso all'interfaccia utente;
- *Oracle*: Sgi fa affidamento su *Oracle 12c* come solida base dati per la gestione e la memorizzazione dei dati. Il *back end* interagisce con il database per eseguire operazioni di lettura, scrittura, aggiornamento e cancellazione. *Oracle* fornisce la persistenza necessaria per mantenere dati coerenti e affidabili, consentendo all'applicazione di operare con informazioni aggiornate e consistenti.

Il flusso applicativo di Sgi si può suddividere in tre rami principali (Figura 2.3):

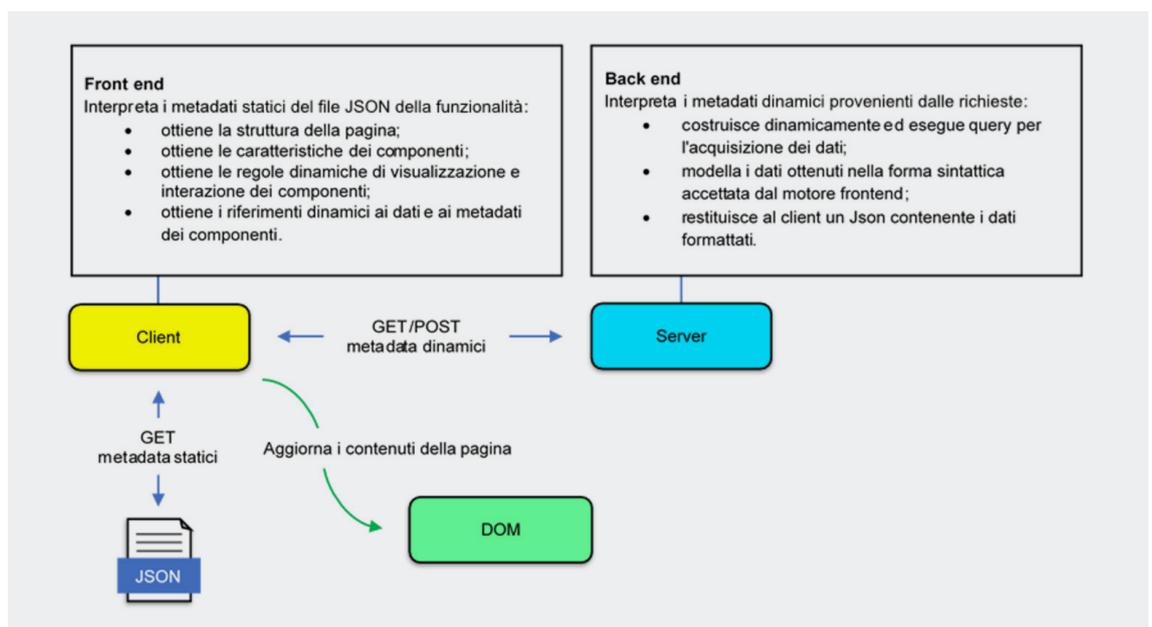
1. **autenticazione Sso**: l'utente richiede l'accesso all'applicazione Sgi, la quale reindirizza a *Shibboleth* per l'autenticazione tramite *Active Directory* o LDAP. Una volta autenticato, l'utente viene reindirizzato con un *token* di accesso, tramite il quale si consente di stabilire sessione utente;
2. **creazione dell'interfaccia utente**: *front end Angular* acquisisce il documento json che fa capo alla funzionalità richiesta; il motore *front end* interpreta il contenuto del documento e inizia la costruzione dell'interfaccia utente. Se specificato, vengono acquisiti e incorporati ulteriori metadati a favore della struttura dell'interfaccia e vengono infine letti e interpretati i dati dinamici dal server per popolare l'interfaccia utente;
3. **acquisizione dei dati**: il *back end Spring boot* riceve le richieste, interroga il database *Oracle* e, quando necessario, aggrega dati da microservizi. Infine, modella i dati in una forma sintattica riconosciuta dal client e invia l'oggetto ottenuto in formato json.

Figura 2.3 - Autenticazione *single sign-on*

Fonte: Istat

La Figura 2.4 mostra come client e server dialoghino tra di loro attraverso uno scambio di dati e metadati in formato json, con il supporto dei due motori applicativi:

- *motore backend*: interpreta le chiamate provenienti dal client e restituisce dinamicamente tutti i dati e i metadati richiesti;
- *motore frontend*: processa i metadati delle interfacce, accorpando i dati relativi alla funzionalità corrente.

Figura 2.4 - Distribuzione delle mansioni tra *front end* e *back end*

Fonte: Istat

2. Il Sistema di gestione delle indagini (Sgi)

I metadati prodotti dal server sono l'esito delle richieste pervenute dal client, costruiti grazie a *query* dinamiche, modellate sulla visibilità specifica dell'utente e adattabili a qualsiasi indagine gestita da Sgi. La loro natura dinamica, oltre a garantire flessibilità al Sistema, ne migliora robustezza e sicurezza.

In conclusione, l'architettura di Sgi si traduce in un ecosistema robusto e sicuro, caratterizzato da scalabilità, modularità, flessibilità e coerenza dei dati. Questi vantaggi si riflettono direttamente sulla esperienza dell'utente finale e sulla gestione ottimizzata delle indagini statistiche:

- **sicurezza avanzata:** l'architettura di Sgi si distingue per la sua solida sicurezza in fase di autenticazione e autorizzazione garantisce che solo gli utenti autorizzati possano accedere alle funzionalità e ai dati dell'applicazione, proteggendo così i dati da accessi non consentiti;
- **scalabilità flessibile:** i componenti dell'applicazione possono essere espansi in modo indipendente, consentendo di bilanciare il carico di lavoro e gestire aumenti del traffico senza compromettere le prestazioni. Questa flessibilità assicura una esperienza utente ottimale anche durante i picchi di utilizzo e consente un adattamento dinamico alle esigenze in crescita degli utenti;
- **modularità e separazione delle responsabilità:** ogni componente dell'applicazione è suddiviso in microservizi indipendenti e tale suddivisione consente una separazione delle responsabilità di manutenzione e sviluppo dei vari moduli;
- **flessibilità dell'interfaccia utente:** l'utilizzo di json statici con metalinguaggio associato al motore applicativo *front end* consente un elevato grado di personalizzazione dell'interfaccia senza richiedere modifiche al codice sorgente.

2.3 Aspetti di usabilità

Secondo lo standard ISO 9241 l'usabilità è definita come “grado in cui un prodotto può essere usato da particolari utenti per raggiungere certi obiettivi con efficacia, efficienza e soddisfazione in uno specifico contesto d'uso”. In pratica definisce il grado di facilità e soddisfazione con cui si compie l'interazione tra l'uomo e uno strumento. Nel caso specifico lo strumento è rappresentato da Sgi, che è stato implementato tenendo conto di questo aspetto fondamentale, mettendo a disposizione dell'utente finale una serie di facilitazioni che consentano di raggiungere l'obiettivo, ossia la configurazione di una nuova indagine, seguendo un percorso il più possibile guidato e intuitivo.

Il problema dell'usabilità si pone quando il modello del progettista, ossia le idee di questi riguardo al funzionamento del prodotto e trasferito sul *design* del prodotto stesso, non coincide con il modello dell'utente finale, ossia l'idea che l'utente concepisce del prodotto e del suo funzionamento. Il grado di usabilità si innalza proporzionalmente all'avvicinamento dei due modelli. Al fine di raggiungere questo obiettivo, nel nostro caso, il grande sforzo fatto nella progettazione è stato quello di fare aderire quanto più possibile il modello informatico al modello statistico.

Le funzioni rese disponibili in questo contesto sono legate alla funzionalità e usabilità del Sistema. Quest'ultima si determina rispondendo alle domande volte principalmente a stabilire l'obiettivo dell'utente e il contesto in cui opera: perciò, attraverso un'analisi dell'utente e delle sue esigenze, si è giunti nel corso degli anni all'implementazione di questo menu di funzioni.



Quando una nuova indagine deve essere gestita da Sgi, le prime operazioni da effettuare sono quelle per la creazione dei metadati iniziali. A tale scopo esiste nella *home page*, una voce che conduce l'utente a una nuova pagina da cui è possibile accedere alla funzione *Crea edizione*. Questa utilità è visibile solo ad alcuni utenti interni all'Istat.

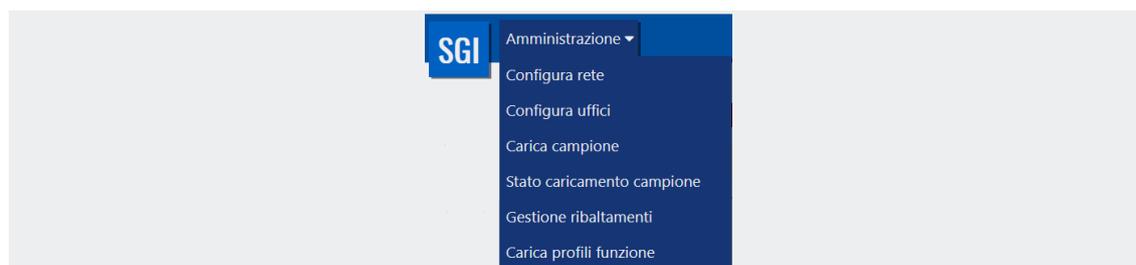
La creazione di una nuova edizione può essere fatta partendo da una già esistente oppure completamente *ex novo*, come si può vedere dalla Figura 2.5.

Figura 2.5 - Pagina *Nuova edizione*

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Dopo avere compilato tutti i dati richiesti, al salvataggio vengono memorizzati sulla base dati tutti i metadati necessari alla fruizione di questa nuova indagine: ad esempio, viene creato l'ufficio Istat nazionale e associato a esso il profilo di responsabile raccolta dati, che sarà l'unico profilo autorizzato ad accedere, in questa fase, alla nuova indagine creata per poterne configurare i vari aspetti. La nuova edizione, una volta così creata, è visualizzata sulla pagina principale e sono messe a disposizione alcune funzionalità che permettono le ulteriori configurazioni presenti nel menu *Amministrazione* (Figura 2.6).

Figura 2.6 - Menu *Amministrazione*



Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Attraverso queste funzioni l'utente, responsabile di un'indagine, può procedere con l'esecuzione di tutti gli altri passaggi preliminari e propedeutici alla gestione dell'indagine dall'applicativo.

Sgi è generalizzata per potere gestire diversi tipi di indagine e attraverso l'applicativo è l'utente finale a decidere quali sono le caratteristiche gestionali e funzionali che il Sistema dovrà possedere. Di seguito analizziamo le voci *Gestione ribaltamenti* e *Carica profili funzione*, le restanti sono approfondite nei paragrafi successivi.

La voce *Carica profili funzione* (Figura 2.7) ha un duplice scopo, in quanto consente sia di definire quali profili entreranno a fare parte della nuova indagine sia di associare a questi le funzioni per le quali saranno abilitati.

2. Il Sistema di gestione delle indagini (Sgi)

Figura 2.7 - Pagina Carica profili funzione

Carica Profili Funzione

Carica Profili Funzione

Edizione di riferimento*

Seleziona

Caricamento file*

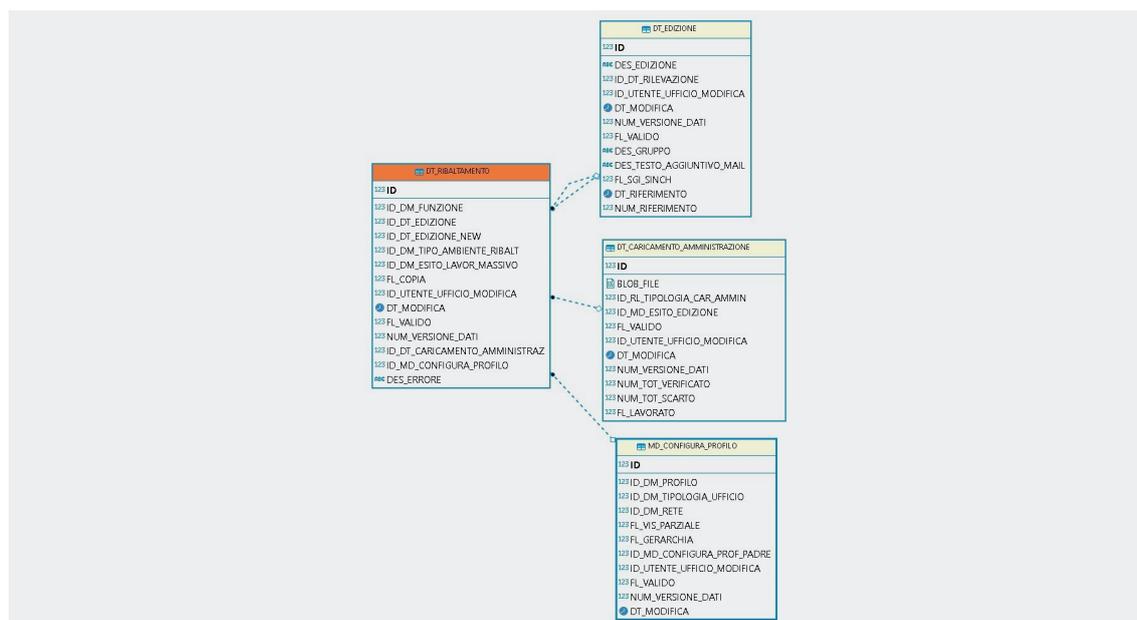
Nessun file selezionato Carica...

Upload Download

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Per fare questo è fornito un *template* in *Excel* da compilare a carico dell'utente sul quale sono presenti due fogli: su uno è riportato l'elenco di tutti i profili disponibili nella base dati di Sgi, e quindi configurabili nel Sistema, sull'altro si trova invece la lista di tutte le voci di menu che è possibile inserire. A ognuna di queste descrizioni è associato un codice univoco da indicare, in un apposito foglio del file, creando l'associazione tra profilo e funzione. Una volta compilato il file, occorre scegliere l'indagine di riferimento da una componente *drop-down* precaricata in pagina ed eseguire l'*upload*. Una procedura PL/SQL programmata lato base dati, tramite un *job*, processa il file, andando a popolare tutte le tabelle che costituiscono il nucleo dei metadati, partendo da quelli esistenti per l'edizione di riferimento indicata e facendone una copia con gli identificativi univoci creati contestualmente e associati alla nuova edizione. Lo schema entità-relazione delle principali tabelle coinvolte in questa fase è riportato nella Figura 2.8. La DT_CARICAMENTO_AMMINISTRAZIONE e la DT_RIBALTAMENTO contengono sia le informazioni fornite dall'utente che configura, come ad esempio il file con le associazioni profilo-funzione, sia tutte le informazioni che la procedura utilizzerà per creare i nuovi metadati per l'indagine in corso di configurazione.

Figura 2.8 - Schema delle tabelle Carica profili funzione



Fonte: Istat

Secondo gli standard della metodologia *DevOps*, adottata per l'ecosistema di raccolta dati nato per il mondo dei Censimenti, gli ambienti su cui si effettuano i *deploy* sono quattro: sviluppo, collaudo, *demo* e produzione. Il risultato del *job* per la funzione *Carica profili funzione* è visibile sull'ambiente sviluppo, mentre quello del *job* associato al *Caricamento funzione*, dalla voce di menu *Gestione ribaltamenti*, provvede a copiare i metadati inseriti su sviluppo su tutti gli altri ambienti. Il risultato dell'esecuzione è la creazione del menu di Sgi, così come è stato configurato su sviluppo, sugli altri ambienti di *deploy*. Dalla pagina dedicata, è possibile inoltre controllare la lista delle operazioni e l'ambiente su cui sono state eseguite: lo stato della lavorazione permettere di capire se l'esecuzione è andata a buon fine o in errore.

Poiché non è sempre desiderabile né possibile spostare le funzioni tutte nello stesso momento, Sgi mette a disposizione la funzione di selezione delle voci da ribaltare. Per poterle scegliere, dopo avere selezionato *Caricamento funzione*, si accede alla pagina sottostante (Figura 2.9) da cui è possibile selezionare, da elementi *drop-down* precaricati in pagina, sia la funzione da ribaltare sia l'ambiente sul quale questa verrà aggiunta.

Figura 2.9 - Pagina *Caricamento funzione*

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Infine, con l'obiettivo di valutare al meglio l'usabilità del Sistema, vengono eseguiti dei test con utenti reali, che sono chiamati a compiere delle operazioni sul Sistema in ambiente di test e con dati di prova. Viene richiesto di simulare il reale andamento delle varie fasi di una rilevazione, al fine di conoscere il loro grado di soddisfazione, per potere intervenire tempestivamente dove necessario a migliorarne il livello.

2.4 Procedure di configurazione della rete e di caricamento del campione

Le operazioni preliminari alla messa in esercizio di una nuova indagine riguardano:

- definizione della rete di uffici che si occuperanno della lavorazione sul campo;
- avvio delle procedure di caricamento automatizzato relativa al popolamento delle tabelle *Oracle* contenente gli uffici e loro relazione gerarchica e i dati del campione da lavorare.

Il menu *Amministrazione* (Figura 2.6) mostra tutte le voci che possono essere utilizzate per la configurazione dell'impianto di una nuova indagine. Questo menu è visibile solo ai profili amministrativi interni all'Istat che sono responsabili della progettazione metodologica e statistica: lo scopo è quello di rendere più veloce la creazione delle nuove indagini e di consentire a un utente di caricare le informazioni che caratterizzano l'edizione di riferimento. In questo paragrafo tratteremo la creazione di rete, uffici e caricamento campione.

La fase iniziale della procedura di creazione della rete richiede la definizione della tipologia degli uffici che si occuperanno di gestire le varie fasi delle attività di Censimento.

2. Il Sistema di gestione delle indagini (Sgi)

Gli uffici vengono classificati mediante l'utilizzo di proprietà specifiche, necessarie a distinguere l'ambito di appartenenza degli stessi. L'utente può effettuare scelte tra tutti uffici gestiti dal Sistema. Gli uffici vengono classificati in base alle seguenti caratteristiche:

- **tipologia:** mostra le opzioni tra cui scegliere, come uffici di rilevazione, uffici provinciali di statistica (Trento e Bolzano/*Bozen*), uffici territoriali Istat, uffici regionali esterni (URC), uffici Istat nazionali, prefetture e così via;
- **rete:** descrive l'identificazione a livello di Paese, Regione, Provincia e Comune o aggregati di diverso tipo;
- **gerarchico:** definisce se l'ufficio ha il ruolo di coordinamento di altri uffici;
- **territoriale:** evidenzia se l'ufficio è legato a un preciso territorio;
- **livello base:** definisce se nella gerarchia è livello "foglia", ossia l'ufficio che si occupa della rilevazione;
- **interno:** specifica se fa parte della rete interna Istat o esterna.

Una volta definita la rete, si deve passare alla lavorazione successiva, ossia a inserire e convalidare gli uffici coinvolti. Sgi mette a disposizione dell'utente un'apposita interfaccia (Figura 2.10), per agevolare la scelta degli uffici sul territorio, e fornisce agli utenti la possibilità di scegliere di caricare gli uffici selezionandoli tramite interfaccia oppure da un foglio *Excel*.

Figura 2.10 - Pagine *Configura uffici*

Configura Uffici

Lista

Cerca nelle righe

Operazioni Tipologia Rete

Operazioni *** Numero verde NAZIONALE

Operazioni *** **Pré Carica uffici** PROVINCIALE

Operazioni *** Pré Carica uffici da file COMUNE

Operazioni *** Prefetture COMUNE

Operazioni *** Prefetture PROVINCIALE

Convalida Uffici

Risultati

Cerca nelle righe

Operazione	Descrizione	Codice Ufficio	Tipologia	Rete
Visualizza Territorio	Anzio	058007	Uffici di rilevazione	COMUNE
Visualizza Territorio	Bari	072006	Uffici di rilevazione	COMUNE
Visualizza Territorio	Bergamo	016004	Uffici di rilevazione	COMUNE
Visualizza Territorio	Bologna	037006	Uffici di rilevazione	COMUNE
Visualizza Territorio	Bolzano/Bozen	021008	Uffici di rilevazione	COMUNE
Visualizza Territorio	Brescia	017009	Uffici di rilevazione	COMUNE
Visualizza Territorio	Cagliari	092009	Uffici di rilevazione	COMUNE
Visualizza Territorio	Caserta	063003	Uffici di rilevazione	COMUNE
Visualizza Territorio	Castellammare di St	063004	Uffici di rilevazione	COMUNE
Visualizza Territorio	Catania	087015	Uffici di rilevazione	COMUNE

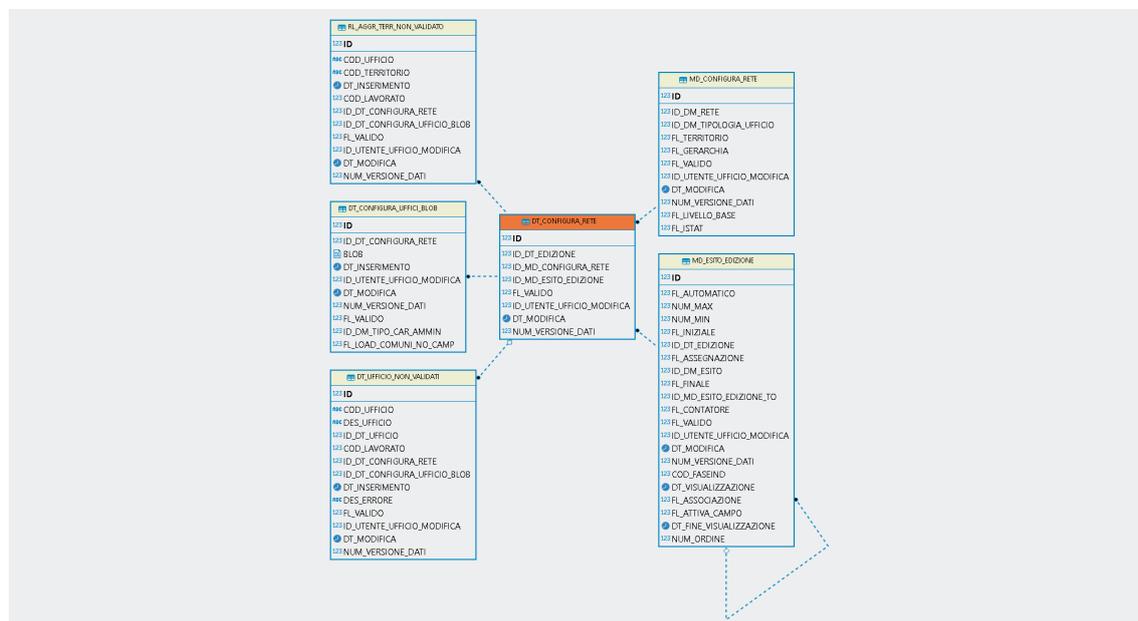
Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Una volta che gli utenti hanno deciso la struttura della Rete, il processo di lavorazione del caricamento e verifica degli uffici è totalmente automatizzato, grazie a una serie di procedure PL/SQL che consentono di eseguire i seguenti compiti basandosi sulla tipologia di rete definita:

- verifica dei dati inseriti dall'utente: è effettuata al fine di eseguire un controllo di congruenza sugli stessi e sul territorio di appartenenza specificato. Il territorio, infatti, potrebbe essere soggetto a modifica rispetto all'ultima indagine effettuata;
- verifica della gerarchia degli uffici: se gli attori interessati appartengono a una gerarchia non territoriale, viene richiesto il caricamento di un ulteriore file in cui si specificano le relazioni padre/figlio;
- ribaltamento dei dati all'interno delle apposite tabelle relative a uffici o strutture gerarchiche dei soggetti.

Per potere gestire la creazione della rete sono state definite delle apposite tabelle di configurazione all'interno della base dati. Nella Figura 2.11 è riportata, tramite diagramma, la relazione tra le tabelle coinvolte:

Figura 2.11 - Schema delle tabelle di configurazione della rete



Fonte: Istat

- **MD_ESITO_EDIZIONE:** tabella trasversale a tutto Sgi che contiene gli esiti per ogni edizione, tra cui quelli che guidano le fasi di caricamento degli uffici e del campione;
- **MD_CONFIGURA_RETE:** raccoglie le varie tipologie di ufficio e di rete con le caratteristiche specifiche relative alla gerarchia, territorio e livello;
- **DT_CONFIGURA_RETE:** ospita la definizione della rete per l'edizione che si sta configurando e serve da guida per le procedure PL/SQL che si occupano del caricamento;
- **DT_CONFIGURA_UFFICI_BLOB:** conserva i dati caricati tramite foglio *Excel* dall'utente con cui è possibile specificare i comuni che parteciperanno alla rilevazione e il loro territorio di appartenenza;
- **DT_UFFICIO_NON_VALIDATI:** tabella di appoggio in cui vengono salvati i dati degli uffici caricati tramite *upload* di un foglio *Excel* contenente le informazioni necessarie al caricamento sulle tabelle *master*;
- **RL_AGGR_TERR_NON_VALIDATO:** al suo interno vengono salvati i dati relativi al territorio degli uffici indicati al punto precedente.

L'esecuzione delle procedure PL/SQL passa attraverso le seguenti fasi di lavorazione:

- verifica della presenza di nuovi *record* sulle tabelle DT_CONFIGURA_RETE, DT_UFFICIO_NON_VALIDATI, RL_AGGR_TERR_NON_VALIDATO, DT_CONFIGURA_UFFICI_BLOB;
- verifica formale e di congruenza dei valori specificati;
- verifica dei dati relativi a rapporti gerarchici specificati dagli utenti per le varie strutture;
- consolidamento dei dati (solo nel caso in cui non siano stati riscontrati errori/incongruenze nei dati) al fine di avviare la fase di lavorazione successiva;
- ribaltamento degli uffici all'interno delle tabelle *master* con conseguente creazione di nuovi uffici in caso di nuove strutture previste per l'edizione in lavorazione;
- ribaltamento delle gerarchie delle strutture all'interno delle tabelle *master*.

2. Il Sistema di gestione delle indagini (Sgi)

Analogamente a quanto descritto per gli uffici, sono state definite anche alcune procedure per il caricamento del campione delle unità di rilevazione. Il caricamento di tali dati viene eseguito utilizzando l'apposita voce di menu *Carica campione*, presente su Sgi nel menu *Amministrazione*. La pagina che viene mostrata all'utente quando accede a tale funzione è riportata in Figura 2.12.

Figura 2.12 - Pagina *Caricamento campione*

Indagine areale 2023

Navigazione: Caricamento Campione

Caricamento Campione

Caricamento Campione

Attenzione: Cliccando sul tasto **Conferma**, si procederà con il caricamento da tabella. Cliccando sul tasto **Upload**, si potrà procedere con il caricamento da file.

Attenzione: Il sistema non accetta file diversi da quello con tracciato record e formati stabilito da Istat, scaricabile tramite il pulsante **DOWNLOAD**.

Caricamento file*

Nessun file selezionato Carica...

Conferma Download Upload

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

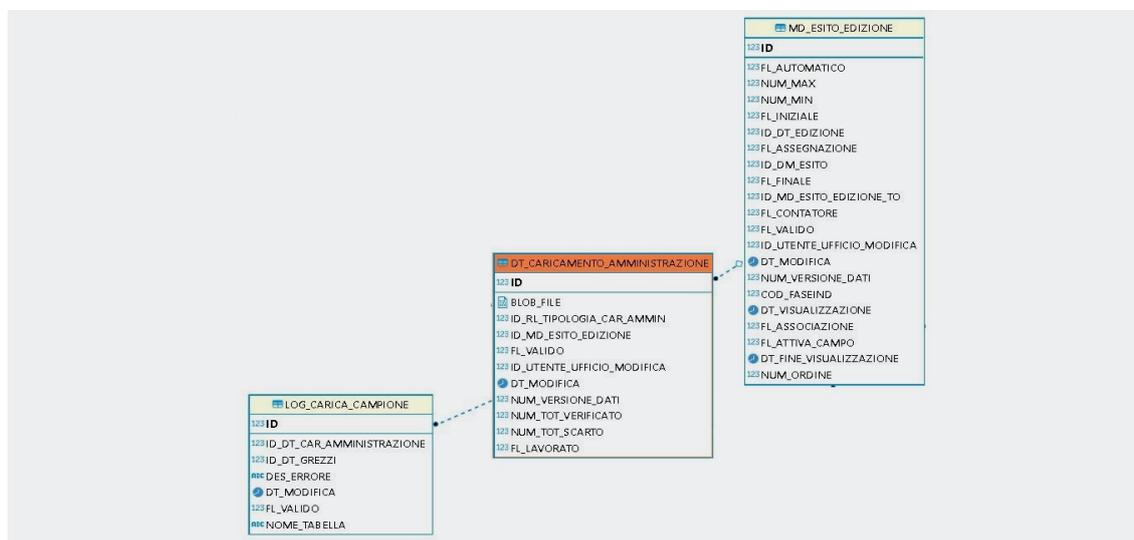
L'utente ha due possibilità in base all'origine dei dati del campione:

- nel caso in cui i dati vengono forniti a Sgi tramite tabelle provenienti da sorgenti messe a disposizione da *Raccolta dati*, sarà sufficiente cliccare su *Conferma* per avviare le procedure di verifica dati e caricamento del campione;
- nel caso in cui i dati non vengano forniti tramite accesso a tabelle *Oracle*, l'utente può utilizzare un apposito *template*, scaricabile da Sgi tramite il pulsante *Download*, compilarlo con i dati del campione ed eseguire l'*upload* del file stesso.

In quest'ultimo caso il file verrà:

- verificato per la corrispondenza dei campi e il numero delle colonne al *template* fornito;
- trasformato in un csv per agevolare la lettura e le procedure di caricamento dei dati;
- salvato in un campo di tipo *blob* nella tabella *DT_CARICAMENTO_AMMINISTRAZIONE*.

Figura 2.13 - Schema delle tabelle del caricamento del campione



Fonte: Istat

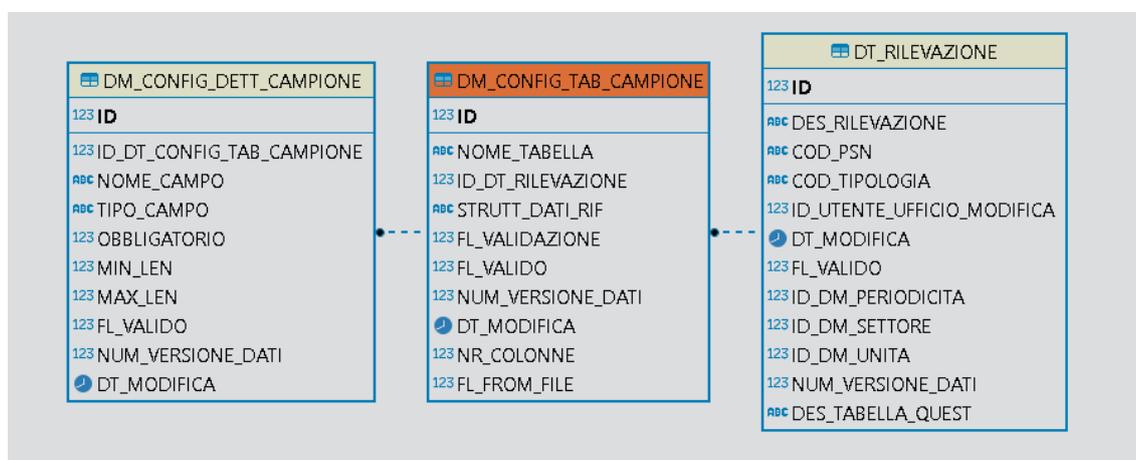
È sempre possibile visualizzare eventuali errori puntuali sui *record* del campione caricato e procedere a un caricamento dell'intera fornitura o di una sua parte. In Figura 2.13 è riportato il diagramma di relazione tra le entità coinvolte.

Uno degli aspetti fondamentali della procedura di caricamento del campione è la necessità che ci sia la massima flessibilità, per potere gestire le stesse operazioni indipendentemente dalla tipologia di campione in lavorazione, differente per ogni Censimento. Per consentire questo tipo di operatività, è stato necessario definire due tabelle di configurazione che raccolgono i dati e i metadati necessari in questa fase.

Le tabelle, il cui schema entità-relazione è illustrato in Figura 2.14, sono state definite come di seguito indicato:

- **DM_CONFIG_TAB_CAMPIONE** contiene i dati relativi alle tabelle di appoggio per ogni rilevazione e le informazioni relative all'origine dati del campione stesso;
- **DM_CONFIG_DETT_CAMPIONE** contiene i metadati delle singole tabelle di appoggio con le informazioni relative a obbligatorietà o meno del campo, tipologia campo (numerico, carattere o data), lunghezza minima e massima del campo e *range* di valori consentiti;
- **DT_RILEVAZIONE** è la tabella trasversale a tutta l'applicazione che definisce la tipologia di Censimento da lavorare.

Figura 2.14 - Schema delle tabelle di caricamento del campione



Fonte: Istat

La procedura effettua verifiche dei dati inseriti dagli utenti, *record per record* e campo per campo. Per eseguire tale operazione vengono messi in relazione i campi specificati nella tabella di configurazione con i singoli valori dei campi inseriti dall'utente. La tecnica utilizzata per tale finalità consiste nel creare una *table function* di tipo *pipelined* che consente di lavorare in modo più veloce: questa soluzione, infatti, permette di ridurre la memoria richiesta per le varie operazioni di interrogazione. Per rendere un'idea, ad esempio, si pensi a un campione composto da circa 1 milione di *record* e con circa 20 attributi da verificare; in questo caso il processo di trasformazione finale, che consiste nel verticalizzare la struttura, restituisce circa 20 milioni di campi totali.

Grazie alla tecnica utilizzata, il processo di verifica è terminato in circa 2 minuti e 30 secondi. Eseguendo lo stesso tipo di controllo, ma utilizzando una tabella fisica di appoggio e una tecnica non *pipelined*, l'attività di controllo ha superato i 40 minuti.

2.5 Creazione, gestione e organizzazione della rete

Le funzionalità di creazione della rete messe a disposizione da Sgi sono le seguenti (in Figura 2.1 sono mostrate le voci di menu):

- **Inserisci utente esterno/interno:** consente di inserire il personale della rete di rilevazione distinguendo tra utenti interni dell'Istat e utenti esterni;
- **Gestione utente:** permette di modificare i dati del personale già inserito;
- **Assegna risorse:** consente di assegnare a ciascun coordinatore uno o più rilevatori;
- **Assegna unità:** consente di assegnare le unità di rilevazione ai rilevatori;
- **Visualizza assegnazioni:** consente di visualizzare le assegnazioni effettuate;
- **Caricamento responsabili:** consente di procedere al caricamento massivo dei responsabili;
- **Stato caricamento responsabili:** consente di visualizzare l'esito del caricamento massivo dei responsabili;
- **Scarico operatori:** consente di procedere allo scarico massivo degli operatori presenti sul Sistema;
- **Stato scarico operatori:** consente di monitorare la richiesta di scarico massivo degli operatori ed effettuare il *download* del file *Excel* una volta terminata la procedura.

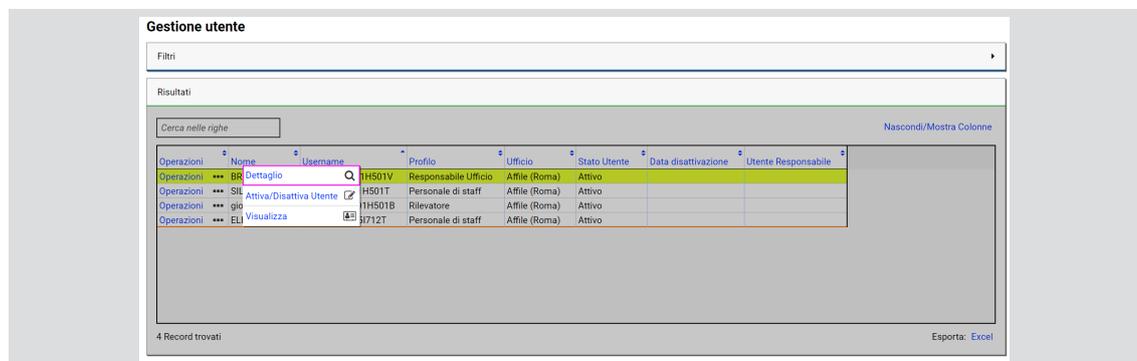
A ogni utente viene associato il profilo scelto dall'operatore in fase di inserimento e viene inserito in uno specifico ufficio.

Lo stesso utente potrà, comunque, essere assegnato ad altri uffici: l'unico vincolo presente è legato al profilo di responsabile di ufficio, perché per ogni edizione all'interno di un ufficio ci può essere un solo responsabile.

L'assegnazione di un profilo vincola la visibilità dell'utente, sia sui dati sia sulle funzioni, e quindi la sua operatività all'interno di Sgi.

Attraverso le operazioni messe a disposizione nella pagina di *Gestione utente* (Figura 2.15) è possibile, inoltre, procedere all'attivazione o disattivazione dell'utente o alla modifica dei dati.

Figura 2.15 - Pagina *Gestione utente: operazioni*



The screenshot shows the 'Gestione utente' interface. At the top, there is a search bar labeled 'Filtri'. Below it, a 'Risultati' section contains a search input 'Cerca nelle righe' and a 'Nascondi/Mostra Colonne' link. The main area is a table with the following columns: Operazioni, Nome, Username, Profilo, Ufficio, Stato Utente, Data disattivazione, and Utente Responsabile. The table contains four rows of data:

Operazioni	Nome	Username	Profilo	Ufficio	Stato Utente	Data disattivazione	Utente Responsabile
Operazioni	BR Dettaglio	IH501V	Responsabile Ufficio	Affile (Roma)	Attivo		
Operazioni	SIL Attiva/Disattiva Utente	IH501T	Personale di staff	Affile (Roma)	Attivo		
Operazioni	gio	IH501B	Rilevatore	Affile (Roma)	Attivo		
Operazioni	EL Visualizza	I712T	Personale di staff	Affile (Roma)	Attivo		

At the bottom left of the table area, it says '4 Record trovati'. At the bottom right, there is an 'Esporta: Excel' link.

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Tramite la voce *Assegna risorse* (Figura 2.16) è possibile assegnare a un utente con profilo coordinatore uno o più rilevatori che svolgeranno l'attività di rilevazione sul campo.

Figura 2.16 - Pagina Assegna risorse: filtri e risultato ricerca

Assegna risorse: selezione risorse

Filtri

Nome: ALDO Cognome: CAVA Profilo: Coordinatore

Username: Cognome:

Cerca Annulla

Risultati

Cerca nelle righe Nascondi/Mostra Colonne

<input type="checkbox"/>	Nome	Cognome	Username	Indirizzo_email	Profilo
<input checked="" type="checkbox"/>	MARIANO	ROSSI	RSSMRN80A01H501H	cuomo@istat.it	Rilevatore
<input checked="" type="checkbox"/>	ROSSANA	VERDI	VRDRSN80M41H501Q	cuomo@istat.it	Rilevatore

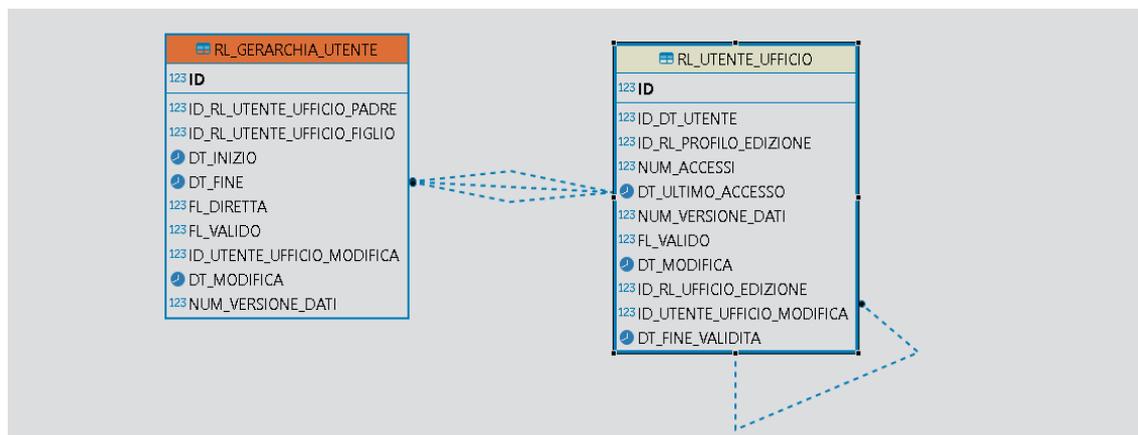
Esporta: Excel

Salva Annulla

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

L'associazione, mostrata nella Figura 2.17, tra coordinatore e rilevatore viene eseguita inserendo sulla tabella RL_GERARCHIA_UTENTE messa in relazione alla tabella RL_UTENTE_UFFICIO, che contiene le informazioni relative agli utenti per l'edizione da lavorare, come l'ufficio di appartenenza.

Figura 2.17 - Schema dell'ufficio di appartenenza e gerarchia dell'utente



Fonte: Istat

Sgi consente di gestire le assegnazioni delle unità di rilevazione ai soggetti interessati alla rilevazione. Attraverso l'utilizzo di pagine appositamente strutturate, infatti, è possibile selezionare l'ufficio e i relativi profili che vi fanno riferimento. Sarà compito di chi effettua le assegnazioni, assegnare le varie unità ai soggetti attraverso le funzionalità apposite, come mostrato nella Figura 2.18.

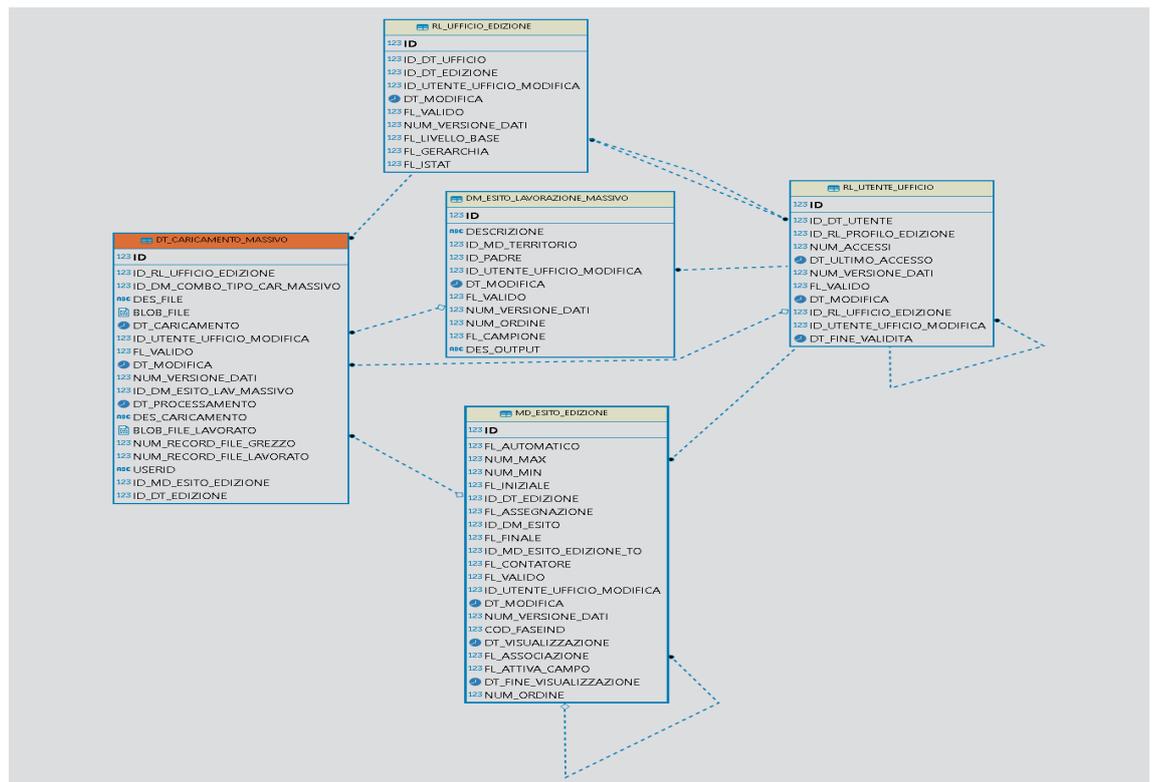
Per evitare assegnazioni non consentite, vengono mostrate all'utente le sole unità eleggibili a questa operazione. È sempre possibile rimuovere le assegnazioni precedentemente effettuate, deselegnando le unità assegnate.

Lo sviluppo applicativo, al fine di definire l'implementazione di quanto sopra esposto, ha richiesto:

- la realizzazione delle funzionalità di interazione tra applicativo e utente;
- lo sviluppo di procedure PL/SQL programmate sul database;
- la definizione di un *batch* che ha anche il ruolo di collaborare con il servizio Utenti.

La scelta di utilizzare procedure PL/SQL e *batch* di comunicazione con servizio Utenti è stata dettata dalla necessità di evitare una diminuzione delle *performance* applicative, con conseguenti tempi di attesa molto lunghi da parte degli utenti. Per effettuare il caricamento dei dati, l'utente utilizza un apposito *template* scaricabile dall'applicativo stesso di cui può effettuare l'*upload* sul Sistema. I dati vengono quindi caricati sul Sistema e vengono eseguite le procedure PL/SQL che eseguono i controlli formali e di congruenza sui campi e sui valori indicati dagli utenti. Infine, la procedura *batch* effettua le operazioni di inserimento specifiche della funzione richiesta. Il seguente schema in Figura 2.20 illustra le tabelle interessate alle attività di caricamento con le rispettive relazioni.

Figura 2.20 - Schema delle tabelle *Caricamento massivo*

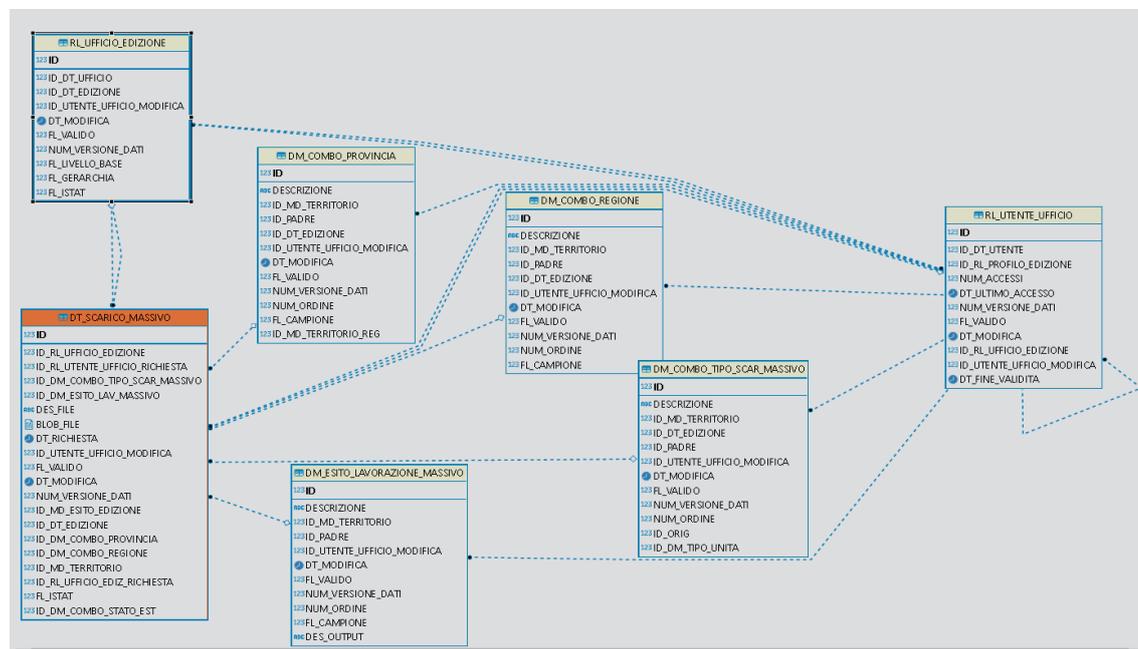


Fonte: Istat

Anche le operazioni di scaricamento massivo degli operatori sono principalmente svolte da procedure PL/SQL, attivate dalla presenza di *record* di richiesta inseriti all'interno della base dati tramite apposita pagina presente in Sgi. La richiesta, così effettuata, genera un *record* all'interno della tabella DT_SCARICO_MASSIVO che viene periodicamente letta dalla procedura. Le tabelle interessate all'attività, con le relative relazioni, sono illustrate in Figura 2.21.

2. Il Sistema di gestione delle indagini (Sgi)

Figura 2.21 - Schema delle tabelle Scarico operatori



Fonte: Istat

La richiesta viene serializzata all'interno del database con i dati relativi all'ufficio per cui è stato richiesto lo scaricamento e l'utenza che ha eseguito la richiesta. La procedura, in presenza di una nuova richiesta, avvia le funzionalità necessarie all'estrazione dei dati e alla creazione di un foglio *Excel* con il risultato ottenuto. Tale foglio è scaricabile tramite la funzionalità presente alla pagina raggiungibile dalla voce di menu *Stato scarico operatori*.

2.6 Rilevazione sul campo

Il "cuore" di Sgi è sotto la voce di menu *Rilevazione*: comprende tutte le funzioni connesse alla fase di conduzione vera e propria. La principale di queste funzioni è costituita dal *Diario*, un cruscotto che racchiude diverse operazioni. Cliccando sulla voce di menu *Diario*, si accede a una pagina in cui è possibile inserire, se lo si desidera, uno o più filtri di ricerca, per affinare la lista dei risultati.

In base al profilo dell'utente che si autentica, i filtri sono impostati in modo da rendere visibili alcune informazioni piuttosto che altre: ad esempio, un profilo con visibilità nazionale vedrà tutto il territorio italiano mentre uno con visibilità più ridotta (provinciale o comunale), vedrà esclusivamente i territori di propria competenza. In altri casi, il profilo dell'utente determina la possibilità di mostrare un filtro o nascondere, dando così l'opportunità di poterlo usare o meno.

Questa scelta è stata fatta nell'ottica di ottimizzare l'usabilità del Sistema, alcune informazioni, infatti, sono utili solo ai responsabili a livello nazionale che devono gestire una mole di dati maggiore rispetto a un profilo con visibilità più ridotta. Un esempio di tali filtri è mostrato in Figura 2.22.

Figura 2.22 - Pagina *Diario*: filtri

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Inoltre, la pagina dei filtri del *Diario* può differire da una indagine all'altra poiché sono impostati in base alle caratteristiche della stessa: ad esempio, in una indagine in cui non è presente la sezione di Censimento, il filtro *Sezione* non comparirà.

La presenza o meno di filtri è determinata in fase di configurazione iniziale, durante la quale avviene il caricamento dei metadati sulla base dati, che è a sua volta letta e interpretata dal motore di inizializzazione delle pagine applicative.

Anche la lista dei dati restituiti, così come i filtri, può differire in base alle esigenze di conduzione e monitoraggio delle varie fasi delle operazioni censuarie, che possono sia variare da una indagine a un'altra sia all'interno di diverse edizioni della medesima indagine. Un *set* di metadati opportunamente configurato istruisce i motori *front end* e *back end*.

Un esempio di risultato dell'elaborazione finale è mostrato in Figura 2.23.

Figura 2.23 - Il cruscotto di Sgi per la conduzione

Operazioni	Esito definitivo	Codice questionario	Sezione	Intestatario	Codice fiscale	Dug	Toponimo
Operazioni	***	7058091001	0580911407006	CARDUCCI GIOSUE	CRDGS174E04H501K	VIA	ANGELICA
Operazioni	***	7058091001	0580911310003	ARGENTO DARIO	RGNDRA66H06H501O	VIA	VIGNALE
Operazioni	***	7058091001	0580911208012	BACH JOHANN SEBASTIAN	BCHJNNS59C02H501J	VIA	LUGNANC
Operazioni	***	7058091001	0580911308008	GUCCINI FRANCESCO	GCCFNC74L14G273F	VIA	DEGLI ESTI
Operazioni	***	7058091001	0580911210011	PRAVO PATTY	PRVPTY89D45F839B	VIA	DI VILLA
Operazioni	***	7058091001	0580911224001	RETTORE DONATELLA	RTDTL87C46F839L	VIA	MARFORI
Operazioni	***	7058091001	0580911433000	ROSSELLINI ISABELLA	RSSLL00L46F839N	VIA	ERMINIO
Operazioni	***	7058091001	0580911433018	SCALA DELIA	SCLDLE99P49F839F	VIA	EPICARM
Operazioni	***	7058091009	0580911209010	SORDI ALBERTO	SRDLRT99P30F205E	VIA	ANTONIO
Operazioni	***	7058091010	0580911210025	VOLTA ALESSANDRO	VLTLSN62R05F205C	VIA	COSTANT

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

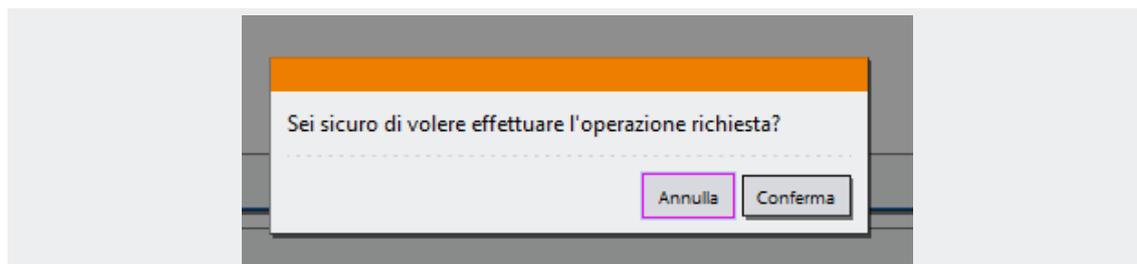
Come è possibile osservare dalla Figura 2.23, all'utente vengono messe a disposizione una serie di operazioni, tramite un elenco interattivo, da cui è possibile scegliere quella desiderata. Tra queste operazioni ci sono anche quelle che consentono di accedere, sia in

2. Il Sistema di gestione delle indagini (Sgi)

compilazione sia in sola visualizzazione, ai dati del questionario. Altre azioni eseguibili in questo contesto riguardano la possibilità di segnalare eventuali lavorazioni fatte sulle unità, per eliminarle o aggiungerle, o ripristinare i dati inseriti.

Cliccando su una di esse, in base alla tipologia, si aprono nuove pagine o *pop up* di conferma delle operazioni richieste, come quando si richiede l'eliminazione di una unità o il ripristino di alcuni dati. Un esempio di *pop up* è mostrato in Figura 2.24.

Figura 2.24 - Pop up di conferma



Fonte: Sistema di gestione delle indagini (Sgi)

Anche la lista delle operazioni non è fissa ma è parametrica e può variare in base al profilo posseduto dall'utente autenticato e in base a proprietà intrinseche della riga identificativa dell'unità di rilevazione, come ad esempio l'esito o lo stato. Il responsabile, inoltre, può intervenire per le situazioni più complesse chiudendo la lavorazione con esiti accessibili esclusivamente a lui.

Come mostrato in Figura 2.25, l'operazione di Dettaglio indirizza verso un riepilogo di alcuni tra i più importanti dati identificativi dell'unità di rilevazione e viene data la possibilità di inserire informazioni fondamentali per conduzione, come esiti o date di appuntamento.

Figura 2.25 - Pagina Dettaglio diario con *drop-down* esiti

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

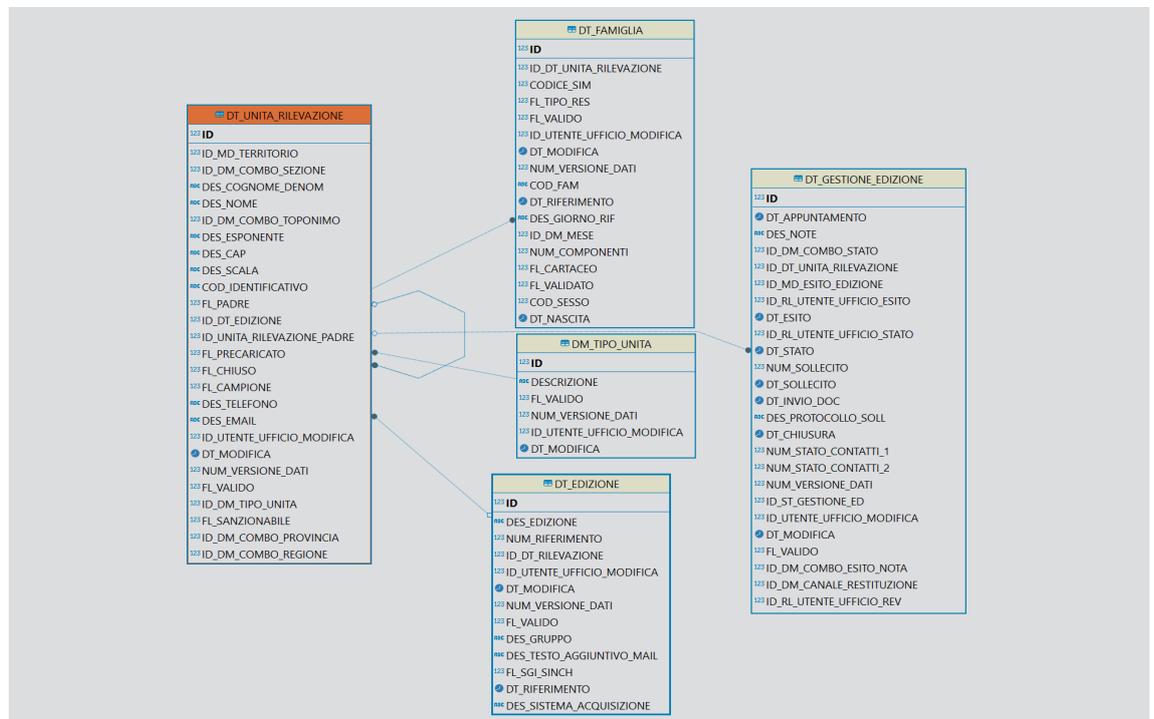
L'aspetto più importante e degno di attenzione all'interno del dettaglio è la configurazione a livello di metadati del *workflow* che regola la visibilità degli esiti e il cambiamento di altre informazioni fondamentali per la conduzione: ogni esito ha un suo predecessore e può essere visualizzato in base a una certa data e a un determinato profilo. Anche questa pagi-

na, come quella dei filtri, può contenere componenti differenti in base al profilo dell'utente connesso, sia tra una indagine e l'altra sia all'interno della medesima indagine: ad esempio, alcuni esiti di indagine possono essere inseriti esclusivamente da un profilo di responsabile e quindi saranno mostrati in modalità scrittura solo agli utenti autenticati con tale profilo. Un altro esempio è il numero di famiglie attese nell'edizione Areale del Censimento della Popolazione e delle abitazioni, che non deve essere visibile al rilevatore, per evitare condizionamenti sulla raccolta delle informazioni e sulle interviste che dovrà effettuare.

Le tabelle coinvolte in questa operazione sono riportate nello schema di relazione di Figura 2.26. La tabella DT_UNITA_RILEVAZIONE contiene tutte le informazioni anagrafiche che permettono di individuare univocamente una unità di rilevazione, così come da campione; a ogni *record* di questa tabella è associato un *record* della DT_GESTIONE_EDIZIONE che contiene invece tutti gli aggiornamenti relativi alla conduzione dell'indagine, specifici per quella unità, come ad esempio stato o esito e i loro autori, ma anche date di lavorazione, informazioni di contatto. Queste due tabelle sono state progettate in modo da essere generalizzate a tal punto da potere essere utilizzate per ogni campione, anche se le rilevazioni in questione presentano un impianto molto differente tra loro.

Esistono poi tabelle, con campi specifici per l'indagine e appartengono alle possibili tipologie di rilevazione e al momento sono presenti in DT_FAMIGLIA, DT_INDIRIZZO e DT_ISTITUZIONE.

Figura 2.26 - Schema delle tabelle Diario



Fonte: Istat

Come già sottolineato, un aspetto molto significativo di una unità di rilevazione è rappresentato dalla coppia esito e stato che possiede o che ha posseduto in un dato momento, sia per conoscere la condizione attuale sia per capire con quale percorso si è giunti a tale condizione: avere quindi a disposizione un riepilogo di tutti gli interventi fatti su tale unità di

2. Il Sistema di gestione delle indagini (Sgi)

rilevazione può risultare utile in particolari momenti della conduzione, ad esempio a ridosso della chiusura per capire se necessario fare solleciti o nella gestione a posteriori dei rifiuti, soprattutto quando si ha a che fare con indagini che entrano nel processo sanzionatorio. L'operazione di *Storico* consente di avere un chiaro monitoraggio di queste necessità, riportando riferimenti dell'utente che ha effettuato i cambi di esito con i conseguenti aggiornamenti di stato, con la data e l'ora.

Figura 2.27 - Pagina Storico diario

Esito	Autore esito	Data esito	Stato	Autore stato	Data stato
unità di rilevazione	ANTONIO	09/06/2022 17:53:51	Da compilare	ANTONIO	09/06/2022 17:53:51
tentativo di contatto	MARIO	15/06/2022 12:25:56	Da compilare	ANTONIO	09/06/2022 17:53:51
appuntamento	LOREDANA	23/06/2022 15:03:20	Da compilare	ANTONIO	09/06/2022 17:53:51
appuntamento	LOREDANA	23/06/2022 15:03:28	Da compilare	ANTONIO	09/06/2022 17:53:51
appuntamento	LOREDANA	23/06/2022 15:03:34	Da compilare	ANTONIO	09/06/2022 17:53:51
intervista	Sibilio Eleonora	09/09/2022 12:53:13	In compilazione	Sibilio Eleonora	09/09/2022 12:53:52
intervista	Sibilio Eleonora	09/09/2022 12:53:13	Da compilare	ANTONIO	09/06/2022 17:53:51
solo persone occasionalmente presenti	Sibilio Eleonora	09/09/2022 12:55:05	Compilato	Sibilio Eleonora	09/09/2022 12:55:05
intervista	ANTONIO	17/01/2023 12:28:00	In compilazione	LOREDANA	25/04/2023 22:28:28
rifiuto	Sibilio Eleonora	08/06/2023 10:33:18	In compilazione	LOREDANA	25/04/2023 22:28:28

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Per alcune indagini è stata data la possibilità al responsabile dell'ufficio, che conduce la rilevazione, di effettuare una convalida dei dati inviati o la modifica, se necessario, degli errori commessi nella fase di raccolta dati. A tale scopo, vengono messe a disposizione sia sotto la voce di menu *Rilevazione* sia sotto quella di *Utilità* delle funzioni specifiche per adempiere alle azioni di validazione. Il vantaggio fornito dal Sistema è quello di potere procedere con le attività di controllo e correzione anche prima del termine della rilevazione censuaria, anticipando sostanzialmente e proficuamente i tempi di valutazione e di aggiornamento degli errori commessi nella fase di raccolta dati.

2.7 Organizzazione, gestione e monitoraggio della formazione

Alcune indagini gestite da Sgi, hanno previsto un periodo dedicato alla formazione della rete di rilevazione; tra le varie modalità di erogazione una di quelle di maggiore successo è la modalità mista, quindi sia con incontri in presenza sia con percorsi ad hoc organizzati su piattaforme online. Per potere gestire e organizzare la formazione dei rilevatori, si è reso necessario uno strumento in grado di garantire un controllo centralizzato delle operazioni di diffusione delle istruzioni a tutta la rete. Durante il periodo della pandemia, tale strumento è stato particolarmente utile a fronteggiare le restrizioni legate al Covid-19 che non consentivano di svolgere lezioni faccia a faccia. Ogni evento formativo, quindi, ha potuto includere personale proveniente da diversi uffici, anche residente in aree anche molto distanti tra loro, grazie alla possibilità di creare eventi online fruibili attraverso canali digitali messi a

disposizione dall'Istat, come la piattaforma *Teams*. Tutte le operazioni necessarie a espletare questa fase sono rese disponibili da Sgi all'interno del menu *Formazione*.

Nella pagina dedicata, l'utente preposto può inserire tutte le informazioni utili alla creazione di un evento formativo; viene data la possibilità di indicare tutti gli uffici il cui personale dovrà aderire alle giornate di formazione. Tali uffici potranno essere selezionati da un elenco che mostrerà solo quelli afferenti alla gerarchia dell'ufficio di competenza. Una volta inserito l'evento, i responsabili della rilevazione possono collocarvi all'interno tutti i discenti dell'evento formativo tramite la funzione *Seleziona partecipanti* della voce di menu *Assegna utente ad evento*: come mostrato in Figura 2.28, il Sistema mostra tutti gli utenti eleggibili che è possibile selezionare.

Figura 2.28 - Pagina *Inserimento partecipanti*

Fonte: Istat, Sistema di gestione delle indagini (Sgi)

Una volta salvata la scelta fatta, gli utenti selezionati sono associati in modo univoco all'evento scelto: ogni utente può essere associato a un solo evento per edizione. In caso di errore, analogamente a quanto implementato per le altre funzioni del Sistema, è possibile variare l'assegnazione deselectando l'utente. Dalla voce di menu *Gestione evento*, come mostrato in Figura 2.29, dopo avere effettuato una ricerca per ufficio, è possibile accedere alla lista di tutti gli eventi già inseriti precedentemente, sia aperti sia chiusi.

Figura 2.29 - *Cruscotto per la formazione*

Operazioni	Evento	Indirizzo	Data classe	Data inizio Fad	Data fine Fad	Comune	Esito
Operazioni ***	Cozze e nuvole	Via della Tellina 123	31/05/2022	20/05/2022	21/05/2022		evento chiuso
Operazioni ***	Chiudi Evento		21/04/2022	21/04/2022	28/04/2022	Latina	evento aperto
Operazioni ***	eve Elimina		25/04/2022	25/04/2022	26/04/2022	Roma	evento aperto
Operazioni ***	eve Modifica		19/05/2022	25/05/2022	27/05/2022	Roma	evento aperto
Operazioni ***	eve Rileva Presenze		19/05/2022	25/05/2022	27/05/2022	Roma	evento aperto
Operazioni ***	evento corona 2	labbkuglpm	10/06/2022	10/06/2022	29/06/2022	Roma	evento aperto
Operazioni ***	Evento dell'organizz via cherso n.110.		20/04/2022	20/04/2022	21/04/2022	Roma	evento aperto
Operazioni ***	evento della piane via delle betulle		14/06/2022	14/06/2022	15/06/2022	Roma	evento aperto
Operazioni ***	Evento della popola via po' n.22		17/05/2022	17/05/2022	19/05/2022	Frosinone	evento aperto
Operazioni ***	Evento della popola via no' n.22		17/05/2022	17/05/2022	19/05/2022	Roma	evento aperto

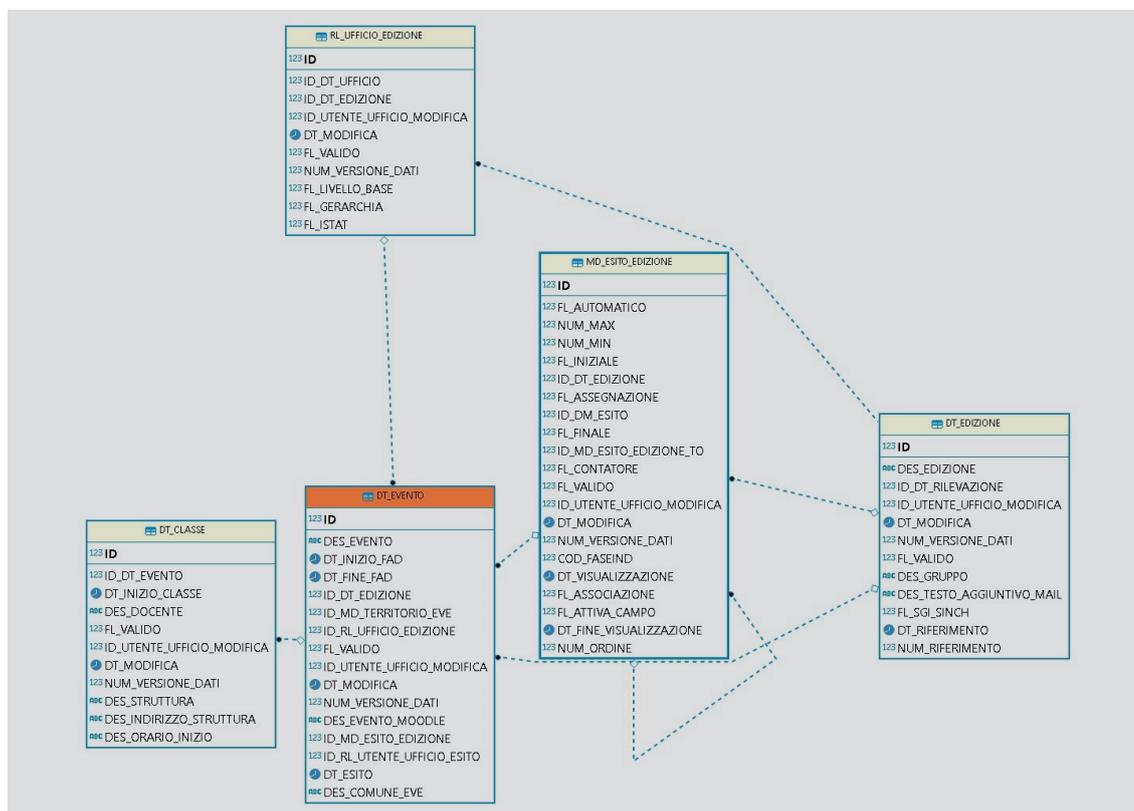
Fonte: Istat, Sistema di gestione delle indagini (Sgi)

2. Il Sistema di gestione delle indagini (Sgi)

Una serie di azioni, messe a disposizione nella sezione *Operazioni*, consente di effettuare attività importanti come la chiusura o la riapertura o l'eliminazione di un evento, la modifica di alcuni dati dello stesso e la rilevazione delle presenze dei partecipanti a un evento, attraverso la quale sarà possibile dichiarare la presenza nella sessione mattutina o pomeridiana o in entrambe o una assenza giustificata.

Il riepilogo della situazione delle presenze viene fornito dalla voce di menu *Visualizza partecipanti*, che consente ai responsabili di monitorare la situazione e avere un quadro complessivo dell'andamento dei percorsi formativi. Le principali tabelle coinvolte, riportate nello schema della Figura 2.30, sono DT_EVENTO e DT_CLASSE, che contengono tutte le informazioni necessarie all'individuazione dell'evento formativo inserito, come l'ufficio di appartenenza che eroga il servizio, gli uffici i cui utenti devono essere iscritti alla formazione, l'indirizzo dell'aula, reale se la classe dovrà vedersi in presenza, virtuale se invece la formazione si svolgerà su *Teams*.

Figura 2.30 - Schema delle tabelle evento



Fonte: Istat

2.8 Base dati

Il modello di Sgi si basa su un database relazionale: nello specifico, la tecnologia utilizzata per la creazione e la gestione del modello è *Oracle*.

Un database relazionale è un tipo di archiviazione che fornisce accesso a *data point* correlati tra loro, basati sul modello relazionale, un modello logico di rappresentazione o strutturazione dei dati implementato su sistemi di gestione relazionali: questi rappresentano un modo intuitivo e diretto di mostrare i dati nelle tabelle, sui cui vengono memorizzati,



attraverso relazioni predefinite. L'assunto fondamentale del modello relazionale è che tutti i dati sono rappresentati come relazioni e manipolati con gli operatori dell'algebra relazionale o del calcolo relazionale, da cui appunto il nome.

Le entità vengono usate per contenere le informazioni sugli oggetti da rappresentare nel database e ogni loro riga è un *record* con un identificativo univoco chiamato "chiave primaria": le righe di tabelle diverse possono essere correlate utilizzando "chiavi esterne". Le colonne contengono invece gli attributi dei dati e ogni *record* di solito ha un valore per ogni attributo, rendendo facile stabilire le relazioni tra i *data point*. È possibile accedere a questi dati in molti modi diversi senza dovere riorganizzare le tabelle di database.

Il modello relazionale consente al progettista di database di creare una rappresentazione consistente e logica dell'informazione. La consistenza è ottenuta inserendo nel progetto del database appropriati vincoli, normalmente chiamati "schema logico".

La teoria comprende anche un processo di normalizzazione in base al quale viene selezionato tra le diverse alternative lo schema maggiormente desiderabile. Alcuni tra gli aspetti fondamentali sono:

1. integrità dei dati: si riferisce alla completezza, all'accuratezza e alla coerenza globali dei dati. I database relazionali usano un insieme di vincoli per garantire tale proprietà. Tali vincoli includono chiavi primarie, chiavi esterne, vincoli *not null*, *unique*, *default*, *check*;
2. transazioni: sono costituite da una o più istruzioni che vengono eseguite come sequenza di operazioni di database per formare una singola unità logica di lavoro. Le transazioni prevedono una asserzione di tipo o tutto o niente, ossia l'intera transazione deve essere completata come singola unità e scritta nel database altrimenti nessuno dei singoli componenti verrà elaborato. Ogni transazione viene trattata in modo coerente e affidabile, indipendentemente dalle altre;
3. conformità ad ACID: per assicurare l'integrità dei dati, tutte le transazioni di database devono supportare le proprietà definite dall'acronimo: *Atomicity* (atomicità), *Consistency* (coerenza), *Isolation* (isolamento), *Durability* (durabilità). L'atomicità implica che la transazione nella sua interezza debba essere eseguita correttamente; se si verifica un errore anche in una sola parte della transazione, dovrà essere interamente invalidata. La coerenza richiede che i dati scritti nel database come parte della transazione rispettino tutte le regole e le restrizioni definite, inclusi vincoli, catene e *trigger*. L'isolamento è fondamentale per ottenere il controllo della concomitanza e garantire che ogni transazione sia indipendente. La durabilità richiede che tutte le modifiche apportate al database siano permanenti dopo il completamento della transazione.

Il database sottostante il Sistema di gestione delle indagini è progettato per dare la massima flessibilità e robustezza all'applicazione. I vincoli tra le varie tabelle sono definiti per garantire la coerenza dei dati scritti dal software, quindi, sono esplicitati su ogni campo ove sia necessario. Le tabelle per la gestione di Sgi sono molto numerose, questo perché si è deciso in fase di progettazione di lasciare sulla base dati molta logica applicativa.

Sono quindi presenti, oltre alle tabelle dei dati, anche tabelle di classificazioni, di metadati, di *logging* e di *backup*. Per fornire quindi una facile lettura e gestione del database è stata data particolare attenzione alla nomenclatura per permetterne l'identificazione e il raggruppamento.

Nei prossimi paragrafi vedremo come la teoria del modello relazionale è stata applicata per progettare e realizzare il database di Sgi.

2. Il Sistema di gestione delle indagini (Sgi)

2.8.1 Gestione e manutenzione

I software utilizzati per la progettazione e la gestione della base dati sono *Data modeler* e *Oracle sql developer*.

Numerosi *package* e procedure, scritti in linguaggio PL/SQL, sono parte integrante dell'applicativo e sono stati sviluppati per consentire una ottimizzazione delle *performance* o una programmazione nel tempo delle operazioni. Ogni processo è realizzato al fine di garantirne una manutenzione sostenibile ed è razionalizzato per essere adeguato in modo dinamico ai requisiti richiesti per ciascuna nuova edizione. Lo sviluppo evolutivo e manutentivo viene adeguato e consolidato a secondo delle complessità previste dalla rilevazione in corso e ne garantisce la congruità dei dati rispettando le tempistiche previste per la conduzione dell'indagine.

Le funzionalità nelle quali sono stati utilizzati gli oggetti *Oracle* messi a disposizione dalla base dati sono quelle che hanno impatto su una mole di dati di grandi dimensioni oppure che hanno la necessità di essere eseguite in momenti specifici. Come descritto nei paragrafi precedenti, dalle voci del menu *Amministrazione* vengono richiamati costrutti procedurali avanzati attraverso cui, per ogni edizione, sono garantiti i caricamenti e i ribalamenti delle strutture e delle unità di rilevazione coinvolte nelle indagini, sono elaborate aggregazioni territoriali e assegnati uffici all'interno della rete di rilevazione. Tutti i caricamenti e gli scarichi dei dati sono gestiti da *package* specifici organizzati per argomento.

Le tabelle della base dati, a livello logico, sono suddivise in base a un prefisso che consente una migliore gestibilità; per tutte le categorie, i prefissi sono seguiti dal nome della tabella che ne indica in modo sintetico il contenuto. Il dettaglio delle categorie presenti è:

- tabelle di classificazione: sono specifiche per ogni edizione e sono identificate dal prefisso DM_. Contengono i dati che vengono gestiti dall'applicativo attraverso le componenti di tipo *drop-down*;
- tabelle associative: identificate dal prefisso RL_, contengono dati legati da relazioni tra due o più entità;
- tabelle di dati: si distinguono per il prefisso DT_ e contengono dati del campione, dati degli utenti, dati degli eventi formativi;
- tabelle di metadati: presentano il prefisso MD_ e contengono le informazioni che alimentano il motore *back end*;
- tabelle per il colloquio: presenti al momento solo per dialogo con la tecnica Cati (*Computer-assisted telephonic interviewing*), si distinguono per il prefisso CATI_;
- tabelle di *logging*: con il prefisso LOG_ vengono distinte tali tabelle e contengono informazioni di esecuzione di tutte le procedure PL/SQL. Sono utili in caso di errore, per individuare tempestivamente la motivazione; in alcuni casi, sono anche visibili dall'applicativo per darne evidenza all'utente;
- tabelle di *backup*: si distinguono per il prefisso BCK_ e vengono utilizzate principalmente in ambito offline per gestire il passaggio dati della sincronizzazione, prendendo tutte le informazioni grezze, prima di qualunque elaborazione.

Inoltre, per permettere un'analisi successiva dei dati e consentire un monitoraggio degli utenti che effettuano ciascuna operazione, ciascuna tabella del database presenta i seguenti campi:

- ID: questo campo è un numerico che rappresenta sempre la chiave prima della tabella.

Questa scelta è stata fatta in fase di progettazione, in modo tale da non dovere gestire chiavi multiple. Le chiavi logiche sono invece gestite attraverso gli indici di unicità;

- ID_UTENTE_UFFICIO_MODIFICA: tiene traccia dell'utente che ha effettuato l'ultima modifica;
- DT_MODIFICA: indica la data dell'ultima modifica effettuata sul *record*;
- NUM_VERSIONE_DATI: *timestamp* necessario nelle fasi gestione dell'offline;
- FL_VALIDO: utilizzato per la visibilità dei *record*.

Al pari della standardizzazione delle tabelle, lavoro analogo è stato svolto sulla nomenclatura dei campi dati: ad esempio, il prefisso DES_ è utilizzato per i campi descrittivi, la dicitura DT_ anticipa sempre un campo data e con il prefisso FL_ si indicano i campi nel formato *flag*, identificati dal dominio 0 e 1.

Questo lavoro di standardizzazione ha riguardato anche tutti gli oggetti *Oracle* che sono stati impiegati nel Sistema. I vincoli di *check* sono identificati dalla dicitura standard CHK_ seguita dal nome della tabella a cui si riferisce. Per le chiavi primarie delle tabelle si utilizza il prefisso PK_ID_ seguito dal nome della tabella a cui si riferisce ed è valorizzato tramite l'utilizzo di una sequenza creata allo scopo di generare automaticamente un valore univoco e sequenziale.

In fase di creazione della sequenza è definito l'incremento di 1, impostati i valori minimo e massimo. Il nome della sequenza è standardizzato all'interno della base dati e definito in coerenza con il nome della tabella alla quale si riferisce, seguito dalla dicitura ID_SEQ. Ogni sequenza è utilizzata dalle procedure di elaborazione dati facendo riferimento ai suoi parametri CURRVAL e NEXTVAL.

La nomenclatura delle chiavi esterne definite per ciascuna tabella rispetta il seguente standard: il prefisso FK_ è seguito sia dal nome della tabella in cui la chiave esterna è definita sia dal nome della tabella cui si riferisce.

Per ottimizzare gli accessi alla base dati e ridurre i tempi di attesa da parte degli utenti sono presenti, oltre a indici generati automaticamente sulle chiavi primarie, anche indici generati sui campi con maggiore frequenza di accesso che mirano a velocizzare le ricerche specifiche. I nomi degli indici rispettano una nomenclatura standardizzata e sono identificati dal prefisso IDX_ seguito dal nome del campo a cui si riferiscono.

Per garantire fluidità nella elaborazione, visto l'utilizzo di un modello normalizzato, e per semplificare la costruzione di *query* più complesse sono utilizzate le viste, elementi di interfaccia utile alla predisposizione degli attributi da estrarre non sempre coincidenti con i campi di una singola tabella. La nomenclatura prevede l'utilizzo del prefisso V_ seguito dal nome che ne identifica sinteticamente il tipo di dati che si rendono disponibili.

Ciascuna fase di gestione e manutenzione della base dati garantisce l'allineamento in ambiente di sviluppo, collaudo, *demo* e produzione a supporto delle omologhe fasi di *deploy* del sistema applicativo.

2.9 Primi risultati e sviluppi futuri

Nel corso degli anni, a partire dal 2018, Sgi ha gestito circa 20 indagini con diversi livelli di complessità in termini di rete di rilevazione. In totale più di 65 mila utenti hanno effettuato l'accesso al Sistema. Molti degli utenti hanno partecipato a diverse indagini,

2. Il Sistema di gestione delle indagini (Sgi)

ottenendo quindi enormi vantaggi nell'avere un unico punto di accesso per la gestione e la conduzione delle stesse. Analizzando i ruoli sulle diverse rilevazioni emerge che: circa millecinquecento utenti appartengono al personale interno Istat, distribuito tra i profili di Responsabile raccolta dati, Monitore, *Staff* raccolta dati. I responsabili di ufficio si attestano sui 120 mila, 50 mila utenti sono presenti come profilo di personale di *staff*, coadiuvati da circa 15 mila operatori di *back office*.

Gli utenti con il ruolo di coordinatore e con quello di rilevatore sono stati rispettivamente circa 2 mila e 67 mila. Infine, circa mille operatori esterni dello *staff Contact center* hanno utilizzato l'applicativo nelle indagini che prevedevano un numero verde al quale il rispondente poteva rivolgersi per ottenere supporto nella compilazione del questionario o per gestire segnalazioni di anomalie del Sistema.

I numeri mostrano che la figura di coordinatore è stata utilizzata solo nelle realtà territoriali di ampie dimensioni; il numero molto ampio di responsabili, invece, è dovuto al fatto che, per ogni tipologia di ufficio, deve essere inserita una figura con tale ruolo.

Le unità di rilevazione lavorate all'interno del Sistema sono state quasi dieci milioni e i questionari compilati, nelle diverse tecniche di rilevazione, poco più della metà. La gestione dei solleciti è stata estremamente onerosa, specie per il Censimento della Popolazione e delle abitazioni dove è stato utilizzato sia il canale postale sia l'invio di messaggi tramite l'*app IO*: in questo caso sono stati inviati nelle diverse *wave* quasi 3 milioni di solleciti postali e un milione tramite l'*app*. Anche i numeri sulla formazione sono estremamente incoraggianti, in quanto sono stati creati più di millecinquecento aule sia virtuali sia reali e tutti gli uffici hanno correttamente iscritto il proprio personale.

Sgi si sta evolvendo sia in termini di funzionalità sia di interoperabilità. Per quanto concerne il primo aspetto, si stanno progettando funzioni per la digitalizzazione della gestione dei pacchi e dei materiali che l'Istat deve sia inviare sia ricevere. Per il secondo aspetto, si sta lavorando all'integrazione con nuove piattaforme di raccolta dati, come *Gino*.

3. L'APP RILEVO PER LA GESTIONE DEI CENSIMENTI PERMANENTI¹

I Censimenti che si avvalgono della collaborazione di una rete di rilevazione prevedono la possibilità di effettuare il recupero delle mancate risposte tramite l'invio dei rilevatori sul campo; questa modalità di lavoro, per risultare efficace, ha la necessità di acquisire le informazioni direttamente sul territorio. Per facilitare queste operazioni è emersa da subito la necessità di consentire le interviste anche in assenza di Internet.

Per l'Indagine Areale, inoltre, la possibilità di lavorare in modalità offline è condizione necessaria per la riuscita della stessa dal momento che l'intera rilevazione è basata sul lavoro dei rilevatori sul campo sia per la fase preliminare di ricognizione dell'area da rilevare con una verifica degli indirizzi, sia per la raccolta dati vera e propria, dal momento che l'unica tecnica di restituzione prevista è la Capi, tramite intervista faccia a faccia.

Per rispondere quindi a questa esigenza nelle prime due *wave* del 2018 e del 2019 del Censimento permanente della Popolazione e delle abitazioni e nel Censimento generale dell'Agricoltura, la piattaforma di gestione delle indagini, Sgi, e di acquisizione dati, *Panda*, sono state progettate per funzionare anche in assenza di Rete, dando la possibilità di compilare il questionario e di inserire gli esiti di caduta e i tentativi di contatto. L'operatore ha potuto quindi utilizzare la stessa applicazione sia in assenza sia in presenza di Rete, scegliendo liberamente la modalità di lavoro.

L'accesso al sistema offline avveniva, così come per l'applicazione online, tramite la digitazione dell'indirizzo URL (*Uniform Resource Locator*) e ciò ha evitato l'installazione di prodotti aggiuntivi sul tablet.

Malgrado questi aspetti positivi, l'analisi dei ticket aperti dagli operatori al *service desk*, ha evidenziato alcune problematiche che hanno reso questa modalità ostica per il rilevatore. La questione più delicata ha riguardato il passaggio involontario dalle due modalità, online e offline, che ha portato, nei casi di lavorazione del medesimo dato, alla perdita di una parte delle informazioni dal momento che veniva salvato il dato più recente.

L'applicazione offline, inoltre, anche se progettata in modo da adattarsi graficamente in maniera automatica a qualsiasi dispositivo, offriva le funzionalità tipiche di un'applicazione web e non di una *app* che sarebbe risultata invece più adatta ai dispositivi in dotazione.

La soluzione adottata dal 2021 è stata quella di separare la parte online da quella offline: è stata quindi rilasciata, per la prima volta all'interno dell'Istat, una *app* Android per l'acquisizione dei dati del Censimento permanente della Popolazione e delle abitazioni.

La progettazione, l'analisi dei requisiti e tutti gli aspetti di organizzazione e di rilascio sono stati seguiti in Istat, con una collaborazione delle diverse direzioni coinvolte, mentre lo sviluppo è stato affidato a una società esterna.

La fase di progettazione si è basata su una tecnica nella quale i bisogni e i desideri dell'utente sono al centro dello sviluppo del software (*user-centered design*).

Per raggiungere questo obiettivo, sono stati creati dei tavoli tecnici con il territorio, in particolare con le associazioni di comuni, e sono stati intervistati alcuni rilevatori che hanno partecipato al Censimento per riuscire ad analizzare in modo completo le loro aspettative.

Per la gestione dell'offline sono state sviluppate le seguenti componenti software:

¹ Il Capitolo è stato redatto da Maura Giacommo (paragrafo 3), Luigi Arlotta (paragrafi 3.1, 3.2.1 e 3.3), Eleonora Sibillio (paragrafi 3.2 e 3.4).

- *mobile app*;
- *API gateway*;
- motore di generazione dei metadati dei questionari;
- microservizio *Synch*.

La *mobile app* denominata *Rilevo*, sviluppata in linguaggio *Dart* e realizzata con *Flutter*, comprende tutte le interfacce per la gestione della rilevazione sul campo e gestisce i contenuti, metadati e dati, ricevuti dal *gateway*.

I dati sono memorizzati in locale utilizzando *SQLite*, una libreria software che implementa un database *Sql* di tipo *ACID* incorporabile all'interno di applicazioni, e vengono criptati con algoritmo 256-bit *Aes* tramite l'estensione *SqlCipher*. Durante la fase di sincronizzazione, i dati transitano nei database interni e rimangono in locale le sole informazioni non complete che devono essere ancora lavorate.

La *API gateway* ha il compito di esporre servizi *restfull* che fanno da ponte verso i microservizi attualmente utilizzati dalle applicazioni *Sgi* e *Panda*: per questo motivo *Rilevo*, *Sgi* e *Panda* presentano il medesimo *stack* applicativo, riportato in Figura 3.1.

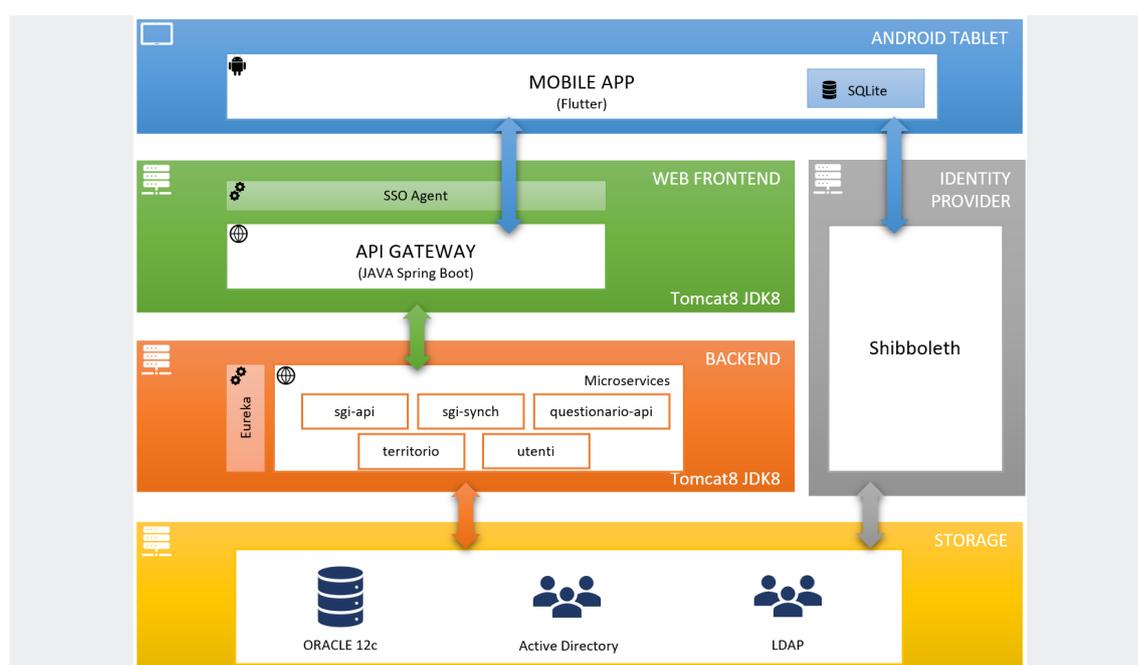
Questa componente ha anche il ruolo di gestire l'autenticazione attraverso *Shibboleth*.

Il motore di generazione dei metadati si occupa di convertire i file di descrizione dei questionari utilizzati dal motore di *Panda*.

Synch fornisce tutti i servizi necessari per trasmettere le informazioni sui dati e metadati in modo bidirezionale, gestisce al suo interno un controllo sul processo, basato su regole impostate in fase di configurazione iniziale che determina l'eventuale scarto dei dati inviati. L'*app* è messa a disposizione dall'Istat sul tablet dei rilevatori; eventuali aggiornamenti possono poi essere scaricati in autonomia dalla rete di rilevazione.

Nel Capitolo 11 è descritta in modo puntuale la gestione del processo di distribuzione.

Figura 3.1 - L'architettura dell'app *Rilevo*



3.1 Progettazione

Nella progettazione dell'app *Rilevo* è stato richiesto di implementare un cruscotto per il monitoraggio delle unità da rilevare e un motore generalizzato per la creazione di questionari elettronici tramite tecnica Capi. Le funzionalità dell'app sono quindi esclusivamente quelle necessarie per le operazioni sul campo e di conseguenza in numero minore rispetto a quelle presenti su Sgi, inoltre, poiché sono sviluppate per adempiere a compiti specifici, hanno in alcuni casi delle situazioni semplificate rispetto a quelle presenti sul sistema online. Ciò è emerso durante la progettazione in cui l'attore principale è stato il rilevatore. Per non destabilizzare l'utente la struttura e la logica del questionario restano le medesime del sistema online. Per garantire tale risultato, tramite un insieme di file strutturati di metadati, sono state messe a disposizione le seguenti parti: la struttura del questionario, la suddivisione in sezioni e pagine, il numero e la tipologia dei quesiti, il *layout*, i testi multilingua e le regole che guidano il rispondente durante la compilazione del questionario. Questi file vengono forniti in input all'app *Rilevo* che è in grado di interpretarli e di costruire il questionario da sottoporre al rispondente. Questo approccio deriva direttamente da *Panda* e presenta notevoli vantaggi in termini di flessibilità e riutilizzo del sistema. Nelle diverse edizioni del Censimento, infatti, al variare dei quesiti proposti e delle regole di compilazione, non è stato necessario modificare il codice dell'app, ma è stato sufficiente agire sui file di configurazione del questionario.

Un lavoro analogo è stato fatto per la parte relativa al cruscotto di gestione del rilevatore: anche in questo caso i metadati riguardanti la modellazione del processo degli esiti e le colonne da mostrare sono inviate tramite dei json gestiti dalla fase di sincronizzazione e derivanti dal database di Sgi.

Alla prima attivazione sul dispositivo *mobile*, *Rilevo* si connette ai server Istat e acquisisce i file di configurazione contenenti i metadati necessari alla composizione del questionario e delle sue regole di compilazione e alla gestione del cruscotto. Allo stesso modo vengono acquisiti tutti i testi delle guide e dei quesiti nelle diverse lingue previste per il Censimento.

Il rilevatore a questo punto può operare anche in assenza di connessione. I dati derivanti dalla compilazione dei questionari vengono memorizzati sul dispositivo in modo cifrato. In fase di sincronizzazione, ossia quando il rilevatore invia le informazioni dal tablet ai server dell'Istat, l'app non solo effettua la sincronizzazione dei dati di rilevazione, ma verifica anche l'eventuale presenza di metadati aggiornati: se sono presenti, l'app li scarica automaticamente e procede ad aggiornare il questionario e il cruscotto di gestione in base ai nuovi metadati. L'utente, quindi, recepisce eventuali modifiche effettuate in modo totalmente trasparente.

3.2 Funzionalità

L'utente, che ha necessità di lavorare offline e ha a disposizione un tablet fornito dall'Istat, trova l'icona associata all'app *Rilevo* sullo schermo e cliccando su di essa accede alle funzionalità per la rilevazione sul campo.

La prima azione è effettuare l'accesso con le stesse credenziali utilizzate per Sgi. Il primo accesso, infatti, deve essere eseguito su Sgi, altrimenti viene restituito un messaggio di errore che ricorda all'utente cosa deve fare. Al fine di verificare la *username* e la *password* inserite, nella schermata di *login* dell'app, il primo accesso di un utente deve avvenire online, in presenza di collegamento di Rete, in modo che la validazione venga effettuata da *Shibboleth*. In caso di esito positivo dell'identificazione, l'app memorizza in locale un *hash* della *password* dell'utente

che verrà utilizzato per verificare le credenziali nei successivi accessi in modalità offline. Quando possibile, l'*app* tenta di validare le credenziali in modalità online ma, nel caso in cui l'*Identity provider* non sia raggiungibile, viene effettuato il confronto con l'*hash* della *password* fornita dall'utente con quello memorizzato in locale. Nella schermata di *login* sarà specificato se la verifica sta avvenendo offline in modo che l'utente sia consapevole di dovere inserire la *password* utilizzata nell'ultimo accesso e che potrebbe non essere quella corrente.

Per motivi di sicurezza all'utente è richiesto di effettuare il *login* nei seguenti casi:

- ogni volta che accede all'applicazione, indipendentemente dallo stato di esecuzione dell'*app*, ossia in caso di apertura dell'*app* non in esecuzione o in caso di rientro in *app* già in esecuzione;
- ogni volta che richiede la sincronizzazione dei dati con il server. provenienti da sorgenti diverse per presentarli in modo coeso all'interfaccia utente.

Figura 3.2 - App Rilevo: home page



Fonte: app Rilevo

Dopo avere effettuato l'accesso, viene mostrata la schermata in cui sono presenti le edizioni di indagine attive per l'utente autenticato e sono mostrate sia funzionalità trasversali all'indagine sia quelle inerenti alla stessa. Per le prime è presente una sezione in cui ci sono le seguenti voci:

- **sincronizza i dati:** la sincronizzazione è contemplata nella doppia direzione, sia per scaricare i dati aggiornati dal server sia per inviare il lavoro svolto;
- **cambia Indagine:** per lavorare su un'altra delle indagini a disposizione;
- **logout:** per uscire in maniera sicura dall'*app*.

Se non si è ancora proceduto alla prima sincronizzazione, il sistema mostra un messaggio che ricorda al rilevatore di effettuare la prima sincronizzazione.

Scegliendo una indagine invece si aprono i menu specifici e vengono mostrate anche alcune informazioni quali la pagina di benvenuto, con alcune informazioni di riepilogo, e su questa pagina, come si può vedere nella Figura 3.2, esiste anche una sezione appositamente dedicata alle *News*, le stesse presenti nella bacheca di Sgi.

Indipendentemente dalle funzionalità scelte, l'*app* contiene le sole unità assegnate all'utente che si è autenticato, a prescindere dal profilo con il quale si sia registrato nel sistema. Ciò significa, ad esempio, che se un responsabile di ufficio effettua il lavoro sul campo, per utilizzare

3. L'app *Rilevo* per la gestione dei Censimenti permanenti

l'app, deve assegnarsi il lavoro e vedrà, non tutte le unità, ma solo quelle che ha preso in carico. Ciò assicura che l'app venga utilizzata esclusivamente per le operazioni da effettuare sul campo.

Tra le funzionalità specifiche sono presenti: la *Verifica del territorio* e il *Diario per l'Area* e la sola voce *Diario per la Lista*. All'interno della *Verifica del territorio* è possibile applicare dei filtri per sezione di Censimento, toponimo ed esito in modo tale da filtrare la lista dei risultati e facilitare le operazioni di raccolta dati sul campo. Inoltre, si può intervenire sulla visualizzazione delle colonne della lista dei risultati, nascondendole o rendendole visibili.

Le operazioni che si possono effettuare sul singolo indirizzo sono:

- **Dettaglio:** nella pagina dedicata, mostrata in Figura 3.3, si può procedere alla validazione degli indirizzi. Si registra l'esito, il dettaglio dell'esito, si segnala il numero degli interni, nel caso di ridenominazione/rinumerazione strada si compila l'apposito campo e, se ci si trova di fronte a un indirizzo problematico, si spunta il campo in basso a destra;

Figura 3.3 - App *Rilevo*: Dettaglio *Verifica del territorio*

The screenshot displays the 'Modifica Indirizzo' interface. At the top, the address 'VIA SANTO STEFANO DI CAMASTRA 5 a' is shown. Below it, there are several input fields: 'Sezione*' with the value '05809114140170', 'Esito attuale' with 'Da confermare', and 'Numero interni'. A dropdown menu for 'Dettaglio Esito' is open, showing options: 'Non presente nel comune', 'SNC', 'Indirizzo fittizio', and 'Altro'. To the right, there is a 'Nuova denominazione ...' field. At the bottom, there is a checkbox 'Spunta se l'indirizzo è problematico' which is checked, and buttons for 'ANNULLA' and 'SALVA'.

Fonte: app *Rilevo*

- **Aggiungi riga:** permette di inserire una riga per ogni alloggio afferente all'indirizzo;
- **Visualizza unità:** tutti i dati del *Dettaglio* sono riportati in sola lettura;
- **Storico:** consente di avere un monitoraggio di tutti i cambiamenti di esito e di stato che l'unità ha subito nel corso della lavorazione, riportando riferimenti dell'utente che ha effettuato i cambi di esito con i conseguenti aggiornamenti di stato, con la data e l'ora.

Le azioni non sono statiche ma si attivano o disattivano in base all'esito posseduto dalla unità di rilevazione, coerentemente a quanto fatto per la corrispondente funzione online su Sgi.

In maniera del tutto analoga al *Verifica del territorio*, la pagina della funzione *Diario* presenta una lista di risultati, con la possibilità di filtrarli e di scegliere le colonne da visualizzare.

Anche questa funzione è centrale e indispensabile al rilevatore per il suo lavoro sul campo. Le operazioni accessibili sulla singola unità sono:

- **Dettaglio:** da questa pagina si può procedere alla correzione di alcuni dati identificativi della famiglia, all'aggiunta di note, all'aggiornamento degli esiti, con relativa data e ora di appuntamento;

- **Storico:** come per il *Verifica del territorio*, consente di monitorare le operazioni e gli utenti che le hanno eseguite sulle singole unità;
- **Elimina:** in caso di inserimento errato, si può eliminare la riga;
- **Inserisci coabitanti:** si potranno inserire tante righe quante solo le famiglie coabitanti in un alloggio, scegliendo una famiglia principale;
- **Visualizza questionario:** consente di leggere tutti i dati di intervista fino inseriti;
- **Compila questionario:** si accede alla compilazione del questionario elettronico;
- **Ripristina:** in caso di inserimento di un esito errato, è possibile tornare al precedente.

L'Indagine da Lista possiede solo la funzione *Diario*, che si presenta in maniera del tutto analoga a quanto già descritto per l'Indagine Areale a eccezione delle azioni **Elimina** e **Inserisci coabitanti**, che non sono possibili in questo caso.

3.2.1 Questionario

Rilevo implementa un motore generalizzato per la creazione dei questionari elettronici da somministrare tramite tecnica Capi. Come avviene per il questionario fruibile in Cawi, anche per l'*app Rilevo*, la struttura del questionario, la suddivisione in sezioni e pagine, il numero e la tipologia dei quesiti, il *layout*, i testi multilingua e le regole che guidano il rispondente durante la compilazione sono definiti tramite un insieme di file strutturati di metadati. I file di metadati utilizzati da *Rilevo* sono ottenuti tramite una rielaborazione automatica di quelli che definiscono il questionario nel sistema *Panda*. Questo aspetto da una parte riduce il lavoro necessario per la preparazione dei questionari dedicati all'acquisizione con tecnica Capi; dall'altra assicura che il questionario elettronico, essendo costruito a partire dagli stessi metadati, si comporti sempre allo stesso modo a prescindere dalla tecnica di rilevazione o dal dispositivo utilizzato. I file di metadati sono memorizzati su una base di dati dei server Istat e vengono inviati all'*app Rilevo* al momento della prima connessione. Una volta acquisiti, *Rilevo* è in grado di costruire *runtime* il questionario da sottoporre al rispondente. Questo approccio, ereditato dal sistema di raccolta dati *Panda*, presenta evidenti vantaggi sia in termini di flessibilità sia di riutilizzo dello strumento. Nelle diverse edizioni dei Censimenti, infatti, al variare dei quesiti proposti e delle regole di compilazione, non è necessario modificare il codice dell'*app* di rilevazione, ma è sufficiente agire sui file di configurazione dei questionari.

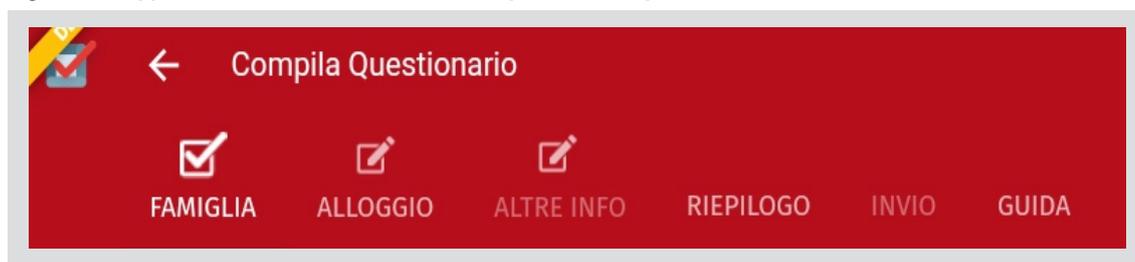
Le schede dedicate alla compilazione del questionario presentano tutte una barra comune nella parte superiore dello schermo. La barra contiene, sul lato sinistro, il menu di navigazione che consente al rilevatore di muoversi tra le diverse sezioni del questionario, visualizzare le schede di riepilogo, l'invio o l'accesso alla guida alla compilazione. Sul lato destro della barra compare invece un selettore che permette di scegliere la lingua del questionario.

La compilazione ha inizio con la lista dei componenti della famiglia che, a seconda della indagine, potrebbe essere già precompilata sulla base dei dati amministrativi. In questa scheda il rilevatore può confermare i dati presenti, apportare modifiche alla composizione della famiglia, indicare il trasferimento o il decesso di uno o più componenti, aggiungere nuovi componenti alla famiglia. Completata la compilazione della lista, sarà necessario confermare la correttezza delle informazioni inserite tramite un apposito selettore che abilita l'accesso alle sezioni successive.

3. L'app *Rilevo* per la gestione dei Censimenti permanenti

Per ogni componente della famiglia, che non risulti traferito o deceduto, sarà richiesta la compilazione di una scheda individuale. Terminate le schede individuali l'intervista prosegue con la compilazione delle sezioni *Alloggio* e *Informazioni finali*. La sezione *Alloggio* contiene quesiti relativi alla dimora e alla zona in cui abita la famiglia; nella sezione *Informazioni finali* sono invece presenti quesiti inerenti alla modalità del questionario e sarà possibile annotare i recapiti che il rispondente deciderà di fornire in fase di compilazione.

Figura 3.4 - App *Rilevo*: evidenza dello stato di compilazione del questionario

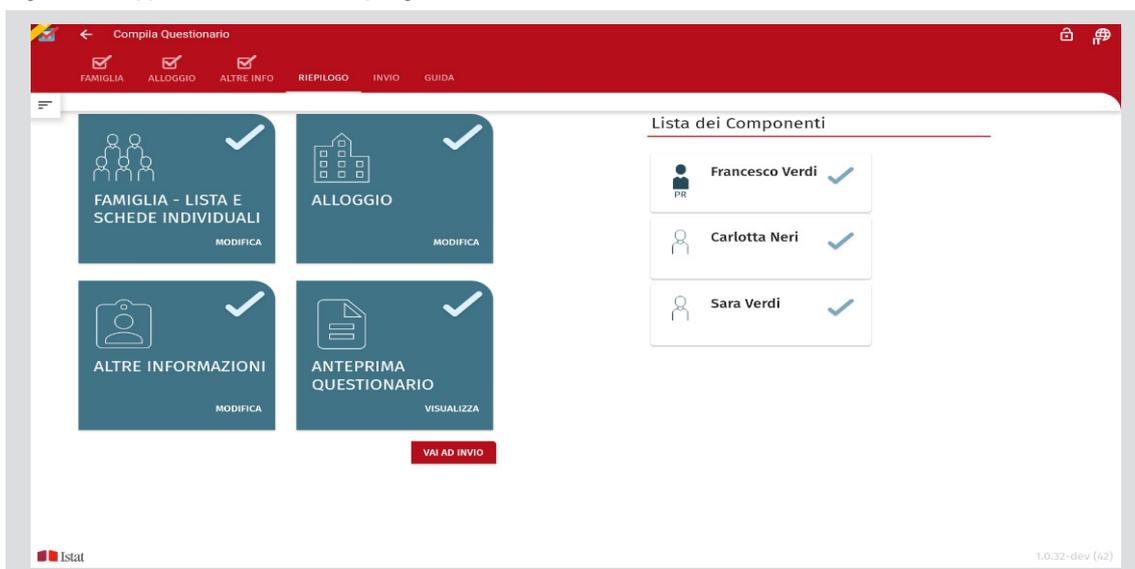


Fonte: app *Rilevo*

Nel corso dell'intervista il sistema indica lo stato di compilazione delle singole sezioni attraverso la visualizzazione di una spunta in corrispondenza delle stesse che risultano corrette e complete (Figura 3.4). In questo modo il rilevatore può sempre verificare lo stato di avanzamento dell'intervista e all'occorrenza tornare sulle sezioni che non risultano ancora completate. Al termine dell'intervista, *Rilevo* mostra una scheda di riepilogo da cui è possibile visionare tutte le risposte fornite attraverso la funzione di anteprima del questionario.

Nella pagina di anteprima, un indice posto sulla sinistra consente di spostarsi rapidamente tra le diverse sezioni agevolando, qualora sia necessaria, la verifica delle risposte fornite (Figura 3.5).

Figura 3.5 - App *Rilevo*: scheda di riepilogo



Fonte: app *Rilevo*

La scheda successiva a quella del riepilogo consente la chiusura del questionario e l'invio definitivo. Dopo l'invio definitivo il questionario non è più modificabile. I microdati relativi ai questionari chiusi e trasmessi all'Istat vengono fisicamente rimossi dall'*app* per questioni di sicurezza e tornano disponibili solo qualora il rilevatore dovesse richiedere la riapertura per effettuare correzioni.

3.3 Sincronizzazione

Non potendo garantire una connessione Internet sufficientemente stabile su tutto il territorio, l'*app* è stata progettata e realizzata per consentire ai rilevatori di portare a termine le interviste in modalità offline, a prescindere dalla disponibilità o meno della Rete. I dati acquisiti tramite le interviste effettuate con l'*app Rilevo* vengono temporaneamente salvati, in modo cifrato, sulla memoria interna del dispositivo e successivamente, quando la connessione di Rete è disponibile, trasmessi ai server Istat tramite un processo che viene indicato con il nome di sincronizzazione. È compito del rilevatore attivare periodicamente e frequentemente il processo di sincronizzazione al fine di mantenere allineati i dati presenti sul dispositivo in suo possesso e i server dell'Istituto. Al momento della sincronizzazione l'*app* si connette ai server dell'Istat tramite un servizio dedicato e trasmette tutti i dati raccolti dal rilevatore nel periodo di tempo intercorso dalla precedente attività di sincronizzazione. Nello stesso momento l'*app* verifica la presenza di eventuali aggiornamenti nei file dei metadati del questionario o di variazioni di stato dei questionari delle famiglie assegnate per le interviste. In presenza di aggiornamenti, *Rilevo* acquisisce automaticamente le nuove informazioni e procede ad allineare il proprio database interno.

Il processo di allineamento dei dati tra l'archivio locale del dispositivo e la base di dati centralizzata dell'Istat ha anche il compito di gestire eventuali incongruenze che possono verificarsi durante questa fase. L'attività del rilevatore su *Rilevo*, infatti, si svolge principalmente in modalità offline e dunque non può tenere conto di eventuali modifiche che sono avvenute online nel periodo di tempo che intercorre tra due attività di sincronizzazione. È possibile, ad esempio, che uno stesso questionario sia stato lavorato offline dal rilevatore e allo stesso tempo online, via web, dal rispondente. Analogamente può capitare che una intervista effettuata da un rilevatore non sia più accettabile in quanto, nel frattempo, l'unità di rilevazione è stata assegnata a un altro rilevatore o è stata definita non più intervistabile. Il servizio che si occupa della sincronizzazione ha il compito di gestire tutte le situazioni che generano incongruenze nei dati, prendere delle decisioni e garantire l'allineamento dei dati tra base di dati centrale e dispositivi mobili in dotazione ai rilevatori. Tutti i dati inviati tramite *app* vengono comunque salvati in apposite tabelle di *backup* in modo da garantire che, a prescindere dalle scelte fatte in automatico dal servizio di sincronizzazione, non si verifichino perdite di dati.

3.4 Sintesi e sviluppi futuri

I dispositivi mobili hanno rappresentato un importante passo avanti nelle indagini sul campo, rendendole potenzialmente più semplici e affidabili. Ciò ha però richiesto uno sforzo notevole sul fronte dello sviluppo del software. Se da un lato creare un'applicazione web sembrerebbe la scelta più economica e ovvia, dall'altro molte delle funzionalità native delle

3. L'app *Rilevo* per la gestione dei Censimenti permanenti

app mobile e necessarie per scopi di rilevamento dovrebbero essere sviluppate da zero. A oggi, anche utilizzando gli strumenti disponibili per creare applicazioni web, puntare a rendere il proprio software reattivo, sicuro, veloce, in grado di funzionare offline e sfruttare il Gps significa spendere tempo e risorse per realizzare ciò che le *app mobile* già hanno e con risultati probabilmente inferiori.

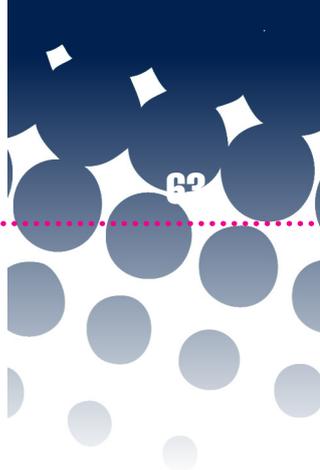
Rilevo, invece, è una *app* nativa per Android in grado di sfruttare appieno le funzionalità che rendono i dispositivi mobili uno strumento particolarmente adatto per le indagini sul campo. Permette ai rilevatori di operare anche in zone montane, insulari o comunque prive di copertura Internet. La geolocalizzazione è veloce e precisa, l'acquisizione e la sincronizzazione dei dati può essere effettuata in momenti separati e, infine, l'interazione con la interfaccia grafica è fluida e affidabile, poiché sfrutta le funzionalità native.

Dato il numero sempre crescente di utenti di dispositivi mobili, il prossimo obiettivo è quello di fornire agli intervistati la possibilità di compilare il questionario di censimento tramite *app mobile*, oltre che tramite versione *desktop*.

Analizzando i risultati nel corso degli anni notiamo un incremento molto importante dei questionari elaborati dopo l'introduzione dell'*app Rilevo*. Infatti, nel 2018, durante il precedente Censimento della Popolazione e delle abitazioni, quando gli addetti al Censimento potevano utilizzare l'applicazione web solo in modalità offline, i numeri erano molto inferiori. Più in dettaglio, vediamo come sono cambiati:

- nel corso del 2018, su un campione composto da oltre 1,6 milioni di famiglie, i questionari elaborati offline sono stati circa 140 mila, con una percentuale pari all'8,66 per cento. Il numero totale dei rilevatori che hanno lavorato nel periodo è stato di oltre 26 mila di cui circa 7 mila hanno utilizzato l'applicazione web in modalità offline;
- nel corso del 2019, con *Rilevo*, su un campione composto da poco meno di 1,6 milioni di famiglie, i questionari elaborati offline sono stati quasi 266 mila, con una percentuale pari al 16,7 per cento. Il numero totale dei rilevatori che hanno operato nel periodo è stato di circa 19 mila, di cui più di 14 mila hanno utilizzato *Rilevo*;
- nel corso del 2021, hanno composto il campione 2,6 milioni di famiglie, mentre i rilevatori sono stati oltre 26 mila di cui quasi 18 mila hanno lavorato con *Rilevo*; il numero di questionari elaborati offline sono stati circa 1,5 milioni con una percentuale pari al 57 per cento;
- nel corso del 2022, hanno composto il campione 1,3 milioni di famiglie, mentre i rilevatori sono stati oltre 14 mila, di cui circa 9 mila hanno lavorato con *Rilevo*; il numero di questionari elaborati offline sono stati circa 646 mila con una percentuale pari al 50 per cento.

La sfida per il futuro sarà rendere l'*app* fruibile anche per altri tipi di indagine e quindi non solo per il Censimento della Popolazione e delle abitazioni, provando ad avvalersi di tutti i vantaggi derivanti dalla facilità di configurazione già sfruttati per gli sviluppi attuali.



4. LA REPORTISTICA PER LE INDAGINI¹

Durante la fase di raccolta dati la disponibilità di una esaustiva reportistica permette di intervenire in modo tempestivo influenzando sia i tassi di risposta sia il lavoro della rete di rilevazione. Pertanto l'Istat ha deciso di dotarsi fino da subito di strumenti specifici a tale scopo. La scelta di rivolgersi al mercato, invece di sviluppare software interno, è motivata dall'ampia offerta di strumenti di *Business Intelligence* che riescono a soddisfare le richieste più esigenti in tempi estremamente rapidi. Nei prossimi paragrafi sono descritte le due piattaforme utilizzate per le indagini con rete di rilevazione e sono analizzati i contesti nei quali sono state utilizzate e le loro potenzialità.

4.1 Piattaforma *Microstrategy*

Microstrategy è una piattaforma *enterprise* di *Business Intelligence & Analytics*, parte integrante dell'ecosistema delle piattaforme digitali dell'Istat a supporto del ciclo di vita del dato. Si colloca come strumento per la realizzazione di analisi e visualizzazioni di dati al fine di sfruttare al meglio il patrimonio informativo delle banche dati disponibili per una migliore e più efficace integrazione dei dati. Le sue funzionalità coprono i differenti aspetti del mondo degli *Analytics*.

Figura 4.1 - *Business Intelligence & Analytics*



Fonte: Istat

La piattaforma consente di collegare i dati provenienti da molteplici sorgenti anche di tipo diverso (database, file *flat*, *big data*, *X*, *web service*, eccetera), integrandole in un modello multidimensionale. Dopo avere raccolto i dati, è possibile pulirli, strutturarli e arricchirli al fine di semplificare i flussi di lavoro di preparazione dei dati.

Il modello dati così definito può essere utilizzato per la predisposizione della tradizionale reportistica con funzionalità OLAP (*Online Analytical Processing*), relazionali, attraverso la definizione di metriche basate su calcoli di dati storici e operazioni di tipo *drill through* per analizzare il dato fino al livello più granulare. Come evoluzione rispetto alla classica *Business Intelligence*, è possibile predisporre grafici, infografiche o cruscotti, utilizzando forme grafiche di nuova concezione come grafici a punti, mappe ad albero o di calore, *word cloud*.

¹ Il Capitolo è stato redatto da: Annunziata Fiore (paragrafi 4.1, 4.1.1 e 4.1.4), Maria Teresa Saccoccio (paragrafi 4.1.2, 4.1.3 e 4.1.5), Domenico Aprile (paragrafi 4.2, 4.2.1, 4.2.2, 4.2.3 e 4.2.4).



Questo permette di effettuare una analisi del dato di tipo:

- descrittiva, facendo uso di metodi numerici quali misure di tendenza centrale (*outlier*, media, moda), misure di posizione (distanza dalla media, dalla mediana), misure di variazione (varianza, deviazione standard), misure di forma (distribuzione di Gauss, distribuzione triangolare, distribuzione lineare);
- predittiva, applicando modelli predittivi per prevedere tendenze (regressioni, *data mining*).

Inoltre, sono di recente introduzione i *tool* di intelligenza artificiale generativa utilizzabili in diverse componenti della piattaforma. Infine, attraverso la predisposizione di interfacce utente sempre più orientate alla semplicità di utilizzo e *browser-based*, la piattaforma fornisce agli utenti uno strumento interattivo di consultazione dei dati, semplificando la navigazione e la condivisione di reportistiche e visualizzazioni.

Tali funzionalità si completano con caratteristiche che rispondono a requisiti di sicurezza legati sia all'autenticazione alla piattaforma, completamente integrata con *Shibboleth*, sia all'accesso differenziato degli utenti in base a differenti profili predefiniti. Tutti questi aspetti hanno portato la piattaforma *Microstrategy* a essere considerata idonea al suo utilizzo in Istat nell'ambito di differenti indagini a supporto delle varie fasi: dalla raccolta, alla validazione fino alla diffusione.

In particolare, la piattaforma è utilizzata per la predisposizione della reportistica e per il monitoraggio delle varie fasi delle operazioni censuarie, condotto dai componenti della rete di rilevazione sulla base del proprio profilo e livello territoriale di competenza.

La reportistica per il monitoraggio delle indagini, nella sua prima versione, è stata sviluppata nel 2018, nell'ambito della prima *wave* del Censimento permanente della Popolazione e delle abitazioni. I *report* sono stati progettati seguendo le stesse caratteristiche di generalizzazione, visibilità differenziata per profili e modi di autenticazione del Sistema di gestione delle indagini (Sgi), in modo da essere completamente integrati con quest'ultimo sistema. Sono stati realizzati sotto forma di visualizzazioni tipiche della *Business Intelligence* tradizionale: uso di servizi OLAP, *drilling*, filtri ed esportazione di dati.

Con il susseguirsi delle varie edizioni del Censimento permanente della Popolazione e delle abitazioni negli anni successivi fino alla più recente del 2023, per la componente della reportistica sono state realizzate delle evolutive di natura:

- adattativa, conseguentemente all'integrazione in Sgi di altre indagini censuarie (Censimento generale dell'Agricoltura 2021, Censimento permanente delle istituzioni non profit) o multiscopo (indagine uso del tempo);
- tecnologica, in affiancamento alla reportistica originaria sono stati aggiunti *dashboard* interattivi con grafici, mappe territoriali, indicatori visivi e una *app mobile* di dati sintetici dell'andamento dell'indagine.

4.1.1 Rapporti riassuntivi

L'insieme della reportistica, chiamata "Rapporti riassuntivi", consente agli utenti impegnati nello svolgimento dell'indagine di tenere sotto controllo lo stato di avanzamento delle attività, verificando le tempistiche delle varie fasi di lavorazione e controllandone le modalità. Inoltre, essa offre informazioni sulla rete di rilevazione, sulle tipologie di attività effettuate dai rilevatori/uffici ed effettua il calcolo dei contributi.

Per la gestione di tale compito sono stati sviluppati dei progetti, oggetti paragonabili a

4. La reportistica per le indagini

delle applicazioni web, sulla piattaforma *Microstrategy* che, a partire dai dati raccolti dalle varie fonti (*Sgi, Moodle, Panda*), forniscono numerose visualizzazioni. L'indice dei *report* di un progetto è richiamabile da una apposita voce di menu di Sgi. Le tavole riepilogative e i cruscotti, una rappresentazione sintetica di più indicatori, mostrano dei valori aggregati, di solito dal livello territoriale fino al dettaglio per singolo utente. L'indice e i dati visualizzati sono opportunamente filtrati per tipologia di rete di appartenenza e profilo utente.

Le tematiche principali affrontate nei *Rapporti riassuntivi* per le diverse indagini sono:

- la formazione degli operatori, che include le informazioni inerenti le caratteristiche e la numerosità degli eventi formativi, le informazioni delle *performance* sul campo e i risultati formativi degli operatori;
- le strutture, che contengono i dati che descrivono le numerosità e le caratteristiche delle strutture coinvolte nella rilevazione;
- la rete di rilevazione, che comprende informazioni sulla numerosità degli utenti della rete di rilevazione per territorio e/o ufficio e le assegnazioni delle unità agli operatori;
- la conduzione dell'indagine, che fornisce dati sullo stato della compilazione dei questionari, sugli esiti della compilazione e sulla reportistica ad hoc a seconda delle indagini;
- cruscotti generali sull'andamento della rilevazione, che permette la visualizzazione e il confronto tra di diversi indicatori;
- contributi da versare all'operatore/ufficio di rilevazione, che descrivono la numerosità delle operazioni effettuate sul campo e del relativo contributo per operatore e ufficio.

Sono stati prodotti anche altri tipi di reportistica, per fasi particolari del processo dell'indagine, come ad esempio i dossier di controllo a livello di regioni per il Censimento generale dell'Agricoltura. Questi consentivano agli utenti di visualizzare delle tavole con dati di *benchmark* e offrivano la possibilità di impostare soglie per verificare la congruenza dei dati raccolti con quelli di indagini precedenti.

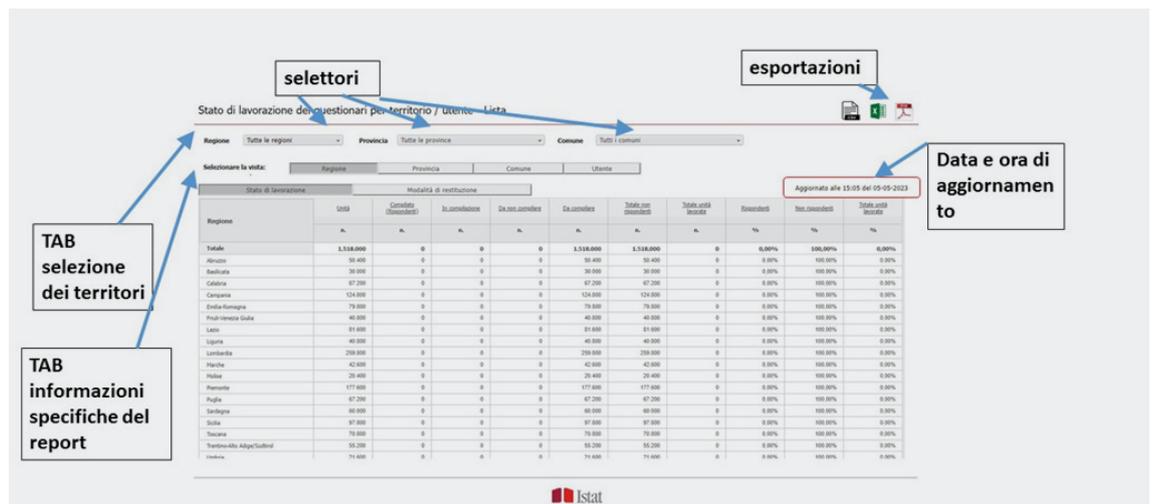
I progetti sono configurati per essere accessibili solo da specifici gruppi di utenti, definiti nei *Microstrategy group* sull'intelligence server, che posseggono ruoli determinati, cioè nell'ambiente *Microstrategy* possono eseguire solo specifiche operazioni. Tali oggetti, sono composti dai gruppi *Active Directory* e *LDAP*, definiti appositamente per la reportistica dell'indagine.

Gli utenti accedono a *Microstrategy* tramite la funzionalità di Sgi denominata *Rapporti riassuntivi*, che richiama il progetto utilizzando un URL. Il passaggio sulla piattaforma di *Business Analytics* avviene in maniera trasparente grazie a *Shibboleth*, utilizzando la sessione utente creata all'accesso in Sgi, senza dovere digitare di nuovo le proprie credenziali.

Nell'URL sono contenuti due parametri necessari per una più dettagliata gestione della visibilità: l'edizione selezionata in Sgi e l'ufficio dell'utente relativo all'edizione scelta. Al momento dell'accesso dell'utente nel progetto, un *report* con dei *prompt*, oggetti simili a filtri dinamici, acquisisce i due parametri ufficio ed edizione della URL e la *User Id* utilizzata nella sessione aperta. Con queste informazioni l'indice iniziale mostra solo l'elenco dei *report* attinenti alla rete e al profilo dell'utente. Inoltre, sempre sulla base dei parametri passati e di sessione, la piattaforma filtra la reportistica successivamente richiesta, visualizzando solo le righe appropriate al ruolo e al territorio o ufficio collegati all'operatore. Per visualizzare un *report* l'utente ne seleziona il link dall'indice e il documento si apre in una nuova pagina del *browser*. Dalle visualizzazioni, di solito in formato tabellare, è possibile effettuare l'esportazione dei dati in formato *Excel* e pdf, imporre dei filtri con dei selettori e navigare tra le schede TAB proposte. In alto a destra è indicata la data e l'ora di aggiornamento dei dati.

La Figura 4.2 illustra le principali possibilità di navigazione di un *report* di tipo *document*, con le selezioni tramite del livello territoriale desiderato. Selezionando il link sulla colonna viene visualizzato un *tooltip* che specifica il contenuto della colonna stessa.

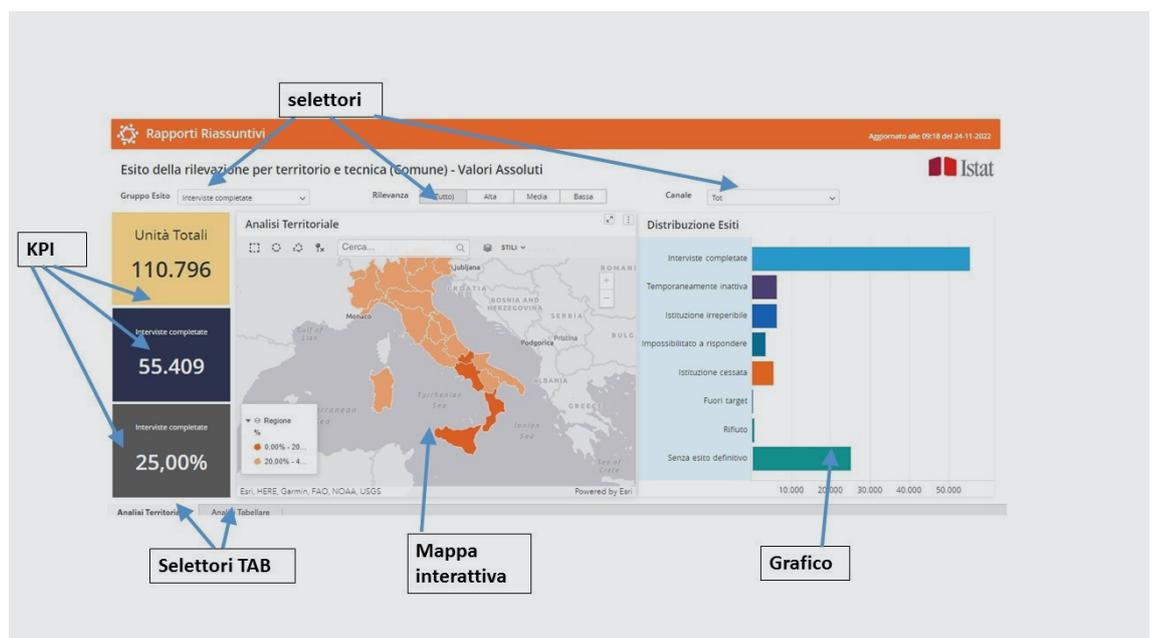
Figura 4.2 - Esempio di *document*



Fonte: Istat, *Rapporti riassuntivi* da piattaforma *Microstrategy*

Nella tipologia di reportistica utilizzata per gli sviluppi più recenti, cioè i dossier, sono presenti delle visualizzazioni migliorate, con aggiunta di maggiori possibilità di interazione. Si è cercato comunque di mantenere la struttura base dei *report* a cui erano abituati gli utenti, aggiungendo nuovi grafici e oggetti di visualizzazione.

Figura 4.3 - Esempio dossier con grafici interattivi



Fonte: Istat, *Rapporti riassuntivi* da piattaforma *Microstrategy*

4.1.2 Data integration

La reportistica delle indagini fornisce diverse tipologie di tavole di dati e visualizzazioni grafiche che consentono agli utenti il controllo dell'andamento della rilevazione. Per fare ciò raccoglie le informazioni necessarie dell'indagine e le elabora aggregandole o conteggiando determinate caratteristiche.

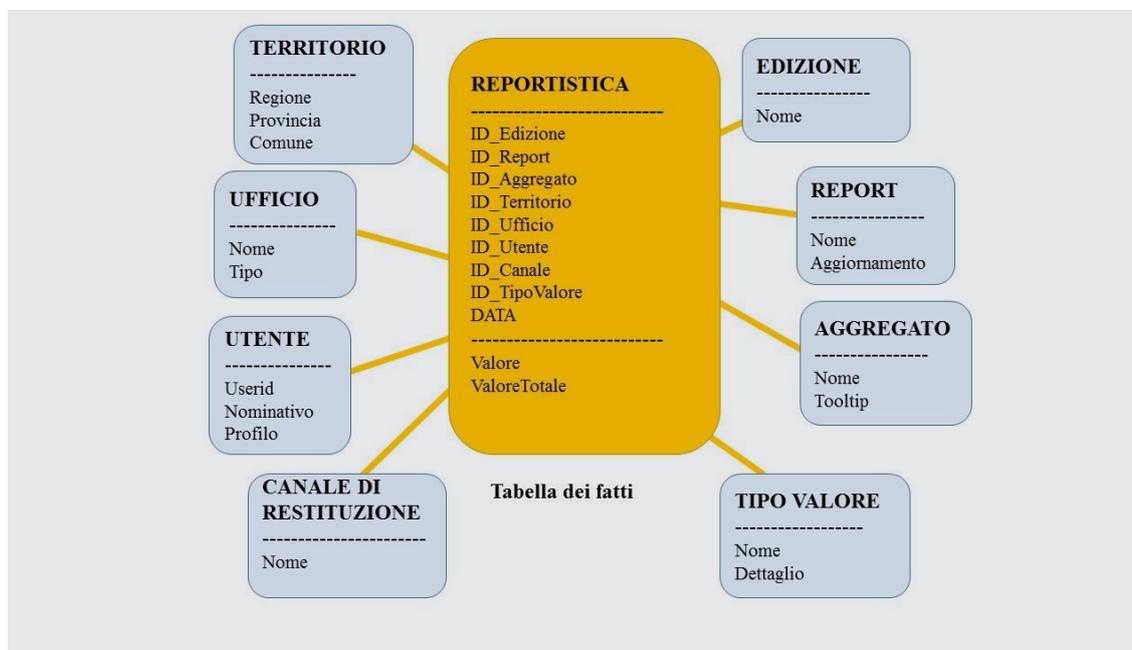
I dati vengono acquisiti direttamente dai diversi sistemi informativi che si occupano della gestione della rilevazione nel suo complesso:

- da Sgi derivano le informazioni quali esiti e stati dei questionari, utenti della rete di rilevazione, assegnazioni delle unità al rilevatore o ufficio;
- da *Moodle* i dati relativi alla formazione degli operatori;
- da *Panda* si recepiscono alcune risposte ai quesiti dei questionari.

Queste informazioni sono memorizzate nei rispettivi database *Oracle* e condivisi con lo schema della reportistica con il supporto di apposite viste o autorizzazioni alla lettura delle tabelle utili. Anche se la piattaforma *Microstrategy* consente di utilizzare ed elaborare direttamente dati provenienti da una o più origini, si è scelto di creare una base dati autonoma su cui mappare i progetti per la reportistica delle indagini.

Questo perché lo schema dati è stato progettato con lo scopo di renderlo quanto più possibile flessibile e utilizzabile da indagini disomogenee.

Figura 4.4 - Schema semplificato di reportistica *data warehouse*



Fonte: Istat

Inoltre, tra le ragioni principali che hanno portato alla creazione di una base dati specifica della reportistica, in cui i dati sono aggregati in precedenza, vi sono quelle di ottenere migliori prestazioni nell'esecuzione dei *report* e quella di rendere il sistema il più possibile indipendente da eventuali modifiche nelle strutture o codifiche dei fornitori dei dati.

Per popolare la base dati sono state sviluppate delle procedure *Extract Transform Load* (ETL), che leggono le informazioni dalle varie fonti, le uniformano, le aggregano per le diverse dimensioni di analisi e memorizzano i valori ottenuti.

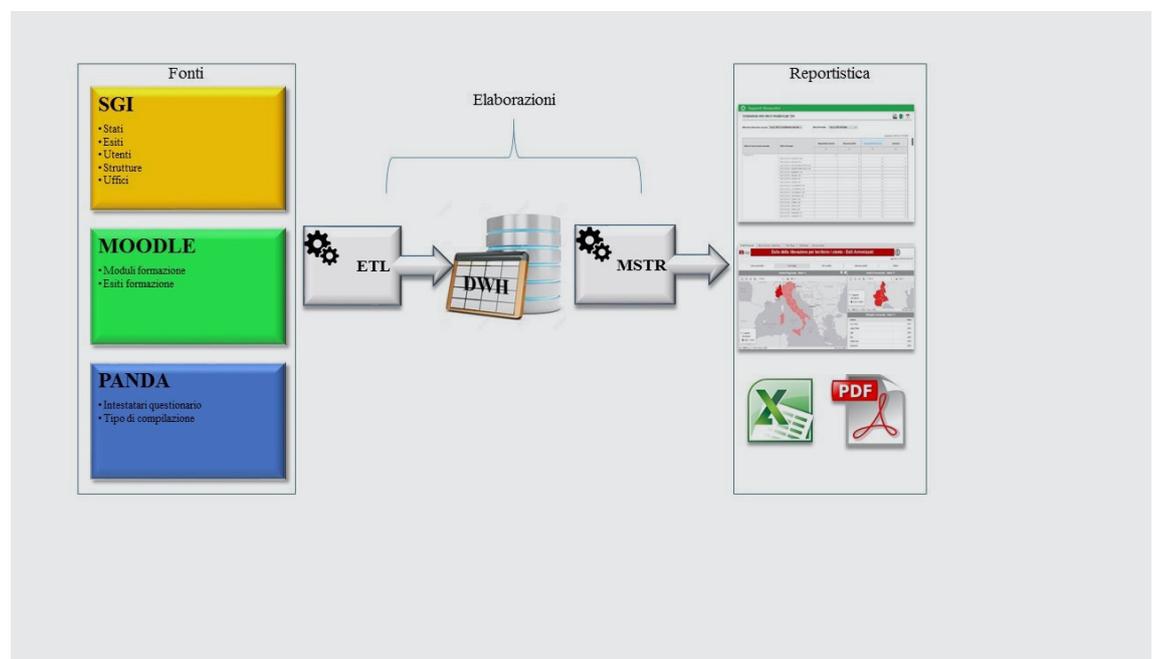
Gli ETL che alimentano la reportistica *Microstrategy* sono procedure e *package* sviluppati nello schema *Oracle* relativo in linguaggio PL/SQL.

Nella prima fase dell'elaborazione le procedure selezionano i dati necessari al *report* dalle varie fonti, quindi li elaborano aggregandoli e codificandoli a seconda dell'argomento di interesse. Inoltre, dove necessario, impongono l'ordine di visualizzazione delle informazioni nella tavola.

Come detto in precedenza, le aggregazioni sono fatte conteggiando o sommando le variabili relative all'argomento da analizzare. Tali variabili sono mappate nel progetto *Microstrategy* come attributi, oggetti che rappresentano i concetti base per analizzare un fenomeno, in questo caso lo svolgimento di una rilevazione. Nei progetti della reportistica per le indagini sono stati definiti, per esempio, gli attributi edizione, ufficio, utente, gli attributi territoriali quali comune, provincia, regione.

L'ultima fase degli ETL è il caricamento delle informazioni elaborate nelle strutture della base dati della reportistica in quelle che vengono chiamate tabelle dei fatti; tali dati aggiornano gli oggetti definiti nel progetto *Microstrategy* come descritto nella Figura 4.5.

Figura 4.5 - Fasi di elaborazione dei dati per la reportistica



Fonte: Istat

Per motivi prestazionali le righe della tabella dei fatti vengono totalmente eliminate e poi ricaricate. La tabella è partizionata sull'edizione in modo da rendere più rapide l'esecuzione di alcune funzioni come ad esempio la cancellazione dei dati di un'edizione.

Il processo è governato da un *job* definito all'interno dello schema *Oracle*, che esegue le procedure degli ETL con pianificazioni orarie. Si passa poi all'aggiornamento dei dati definiti nell'*Intelligent cube*: tale operazione che viene eseguita dall'intelligence server.

4. La reportistica per le indagini

Un *Intelligent cube* è un insieme multidimensionale di dati salvati su un server *Microstrategy* e può essere condiviso con un'unica copia in memoria per essere utilizzato da numerosi *report*, documenti e dossier.

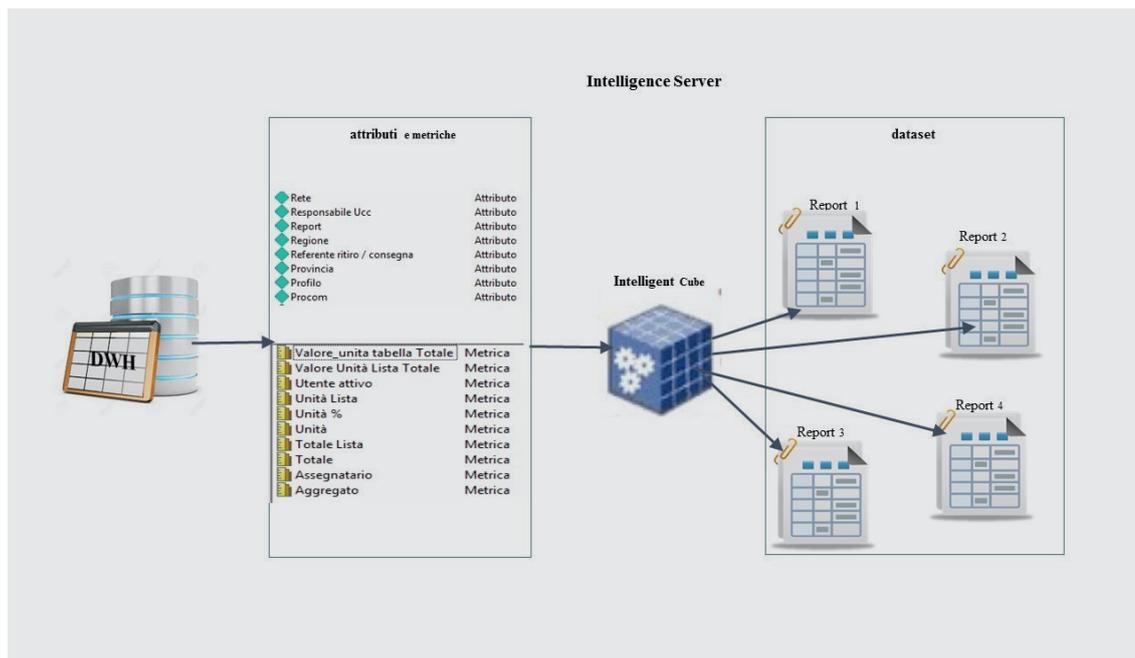
In *Microstrategy* l'origine dei dati di un oggetto di visualizzazione è costituita da uno o più *dataset*, cioè insiemi di attributi e metriche organizzati in una struttura.

I *dataset* possono essere di diverse tipologie, per esempio *report* semplici, *query* dirette sulla base dati, informazioni importate da un'origine dati esterna, *report* derivati da un *Intelligent cube*.

Per la reportistica delle indagini si è optato per quest'ultima tipologia e quindi, quando un utente richiede l'esecuzione di un *report*, *Microstrategy* non esegue una *query* diretta sulla base dati, ma su questa struttura residente in memoria, ottenendo così delle prestazioni migliori soprattutto per *report* complessi.

Gli *Intelligent cube* vengono aggiornati utilizzando delle funzionalità di *Microstrategy* che consentono di definire e schedulare tali attività; le tempistiche sono naturalmente successive alla fine degli ETL che agiscono sulla base dati.

Figura 4.6 - Creazione e utilizzo di un *Intelligent cube*



Fonte: Istat

Nei progetti per la reportistica delle indagini sono stati definiti diversi *Intelligent cube* condivisi da uno o più *report* a seconda degli attributi e metriche rappresentate.

Inoltre, alcuni di essi contengono le informazioni relative agli utenti della rete di rilevazione e provvedono alla funzione di filtro per tutte le visualizzazioni degli utenti territoriali o appartenenti a gerarchie.

La numerosità e l'occupazione di memoria dei cubi totali e di un singolo cubo sono dei parametri che devono essere settati a livello di progetto.

4.1.3 Configurazione del progetto

L'implementazione di un progetto in *Microstrategy* consiste nella progettazione e configurazione degli oggetti che costituiscono il *layer* dei suoi metadati:

- *Schema Objects*: Tabelle, Fatti, Attributi, Gerarchie, Trasformazioni;
- *Analysis Objects*: Metriche, *Prompt*, Filtri, *Custom Group*, *Consolidation*;
- *Report Objects*: Dossier, *Document*, *Report*, *Cube*;
- *Administration Objects*: Gestione Utenti, Ruoli di Sicurezza, *Monitoring*.

La prima operazione effettuata per iniziare lo sviluppo del progetto di reportistica è stata la definizione degli oggetti: schema, tabelle, fatti, attributi, gerarchie.

Dopo avere configurato la connessione al database sul quale sono memorizzati i dati che saranno oggetto della reportistica, si è proceduto a definire il catalogo del *data warehouse*, ovvero l'insieme di tutte le tabelle o viste di riferimento, il cui contenuto è l'output della fase ETL precedentemente descritta. Queste tabelle sono state opportunamente strutturate e organizzate secondo il classico modello schema a stella di un *data warehouse*, costituito da "fatti" e "dimensioni". I fatti producono gli omonimi oggetti in *Microstrategy*, mentre le dimensioni sono associate agli "attributi". Questi ultimi sono organizzati in "gerarchie" nel caso esistano relazioni padre-figlio. Le gerarchie permettono sia la navigazione degli attributi sia le operazioni di *drilling* per le analisi esplorative dei dati.

A partire dai fatti sono definite le "metriche": conteggi, medie, percentuali, indicatori, funzioni calcolate a partire da metriche elementari, ovvero tutti i calcoli eseguiti i cui risultati sono riportati nei *report*.

Per potere generalizzare i *report* sono stati definiti oggetti di tipo *prompt*, legati ad esempio all'edizione dell'indagine, che non sono altro che parametri di input a seconda del quale cambia dinamicamente il contenuto dei dati da visualizzare all'interno di un *report*, *document* o dossier, le cui caratteristiche e differenze saranno descritte nel paragrafo successivo.

Nell'ambito del monitoraggio è stato creato un progetto *Microstrategy* per ciascun tipo delle varie indagini:

- Censimento permanente della Popolazione e delle abitazioni;
- Censimento generale dell'Agricoltura;
- Censimento permanente delle istituzioni non profit;
- Rilevazione *Uso del Tempo* nell'ambito delle Indagini Multiscopo sulle famiglie.

Le numerosità legate alla tipologia di oggetti presenti in un progetto sono:

- 31 tabelle;
- 46 fatti;
- 101 attributi;
- 3 gerarchie;
- 77 metriche;
- 10 *prompt* e filtri;
- 110 *report*;
- 150 documenti;

4. La reportistica per le indagini

- 50 dossier;
- 18 *cube*.

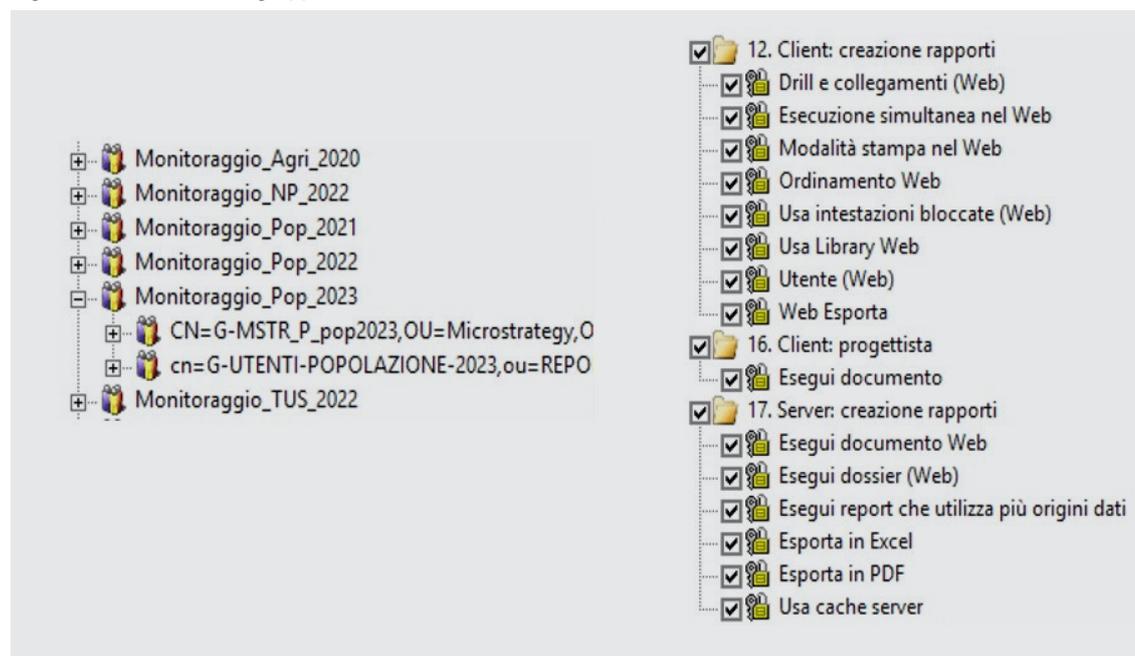
Tutti gli oggetti finora descritti, che compongono i vari progetti, hanno una visibilità per l'utente stabilita da criteri di sicurezza definiti a livello *Administration*.

In particolare, gli oggetti amministrativi permettono di definire gli utenti che accedono alla piattaforma e di organizzarli in gruppi. Questi ultimi possono essere associati soltanto ai progetti di propria competenza e accedervi tramite un ruolo di sicurezza predefinito attraverso il quale sono concessi esclusivamente i privilegi di accesso strettamente necessari. In particolare, per i rapporti riassuntivi sono stati definiti tanti gruppi quanti sono i tipi di indagine per ciascun anno.

A ciascuno di questi gruppi è stato associato direttamente il corrispondente gruppo LDAP definito su *Active Directory* per quella edizione e non i singoli utenti. Attraverso questo meccanismo, *Microstrategy* eredita le autorizzazioni da concedere alla reportistica direttamente dal sistema di autenticazione di Sgi che si interfaccia a LDAP.

La Figura 4.7 mostra l'elenco dei gruppi che sono stati definiti con un esempio di gruppo LDAP associato e il ruolo di sicurezza che consente agli utenti appartenenti a tali gruppi di usufruire di tutti e soli i privilegi elencati.

Figura 4.7 - Gestione dei gruppi e dei ruoli di sicurezza



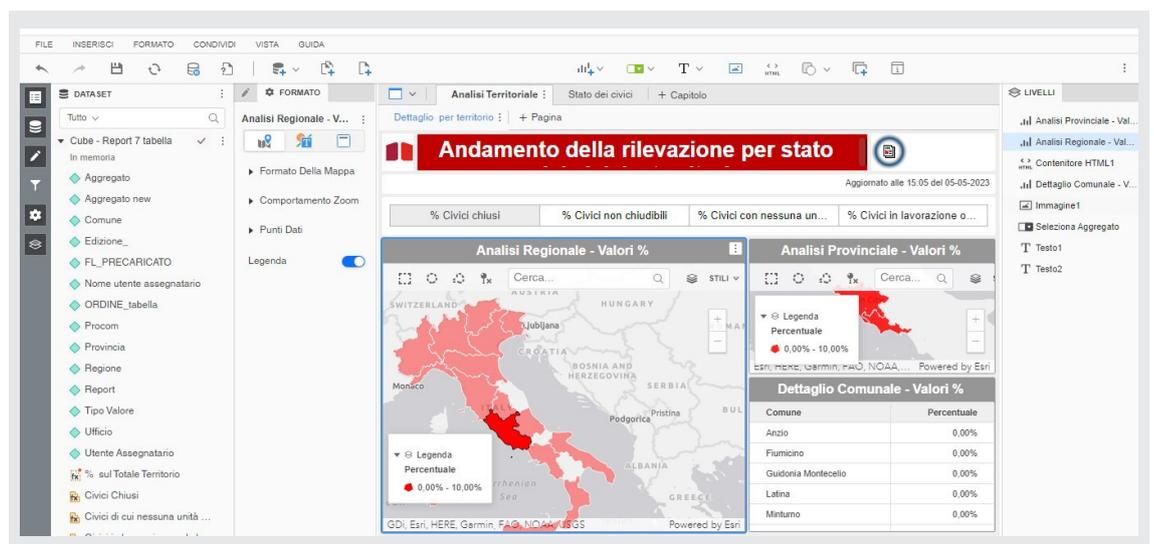
Fonte: Istat

4.1.4 Tipologie di reportistica

A partire dal 2018, anno in cui si è adottato per la prima volta *Microstrategy* per la reportistica dei Censimenti, sono stati implementati principalmente tre distinti tipi di oggetto di visualizzazione, scelti soprattutto in base all'utilizzo, alle caratteristiche dei *report* da produrre e anche al periodo di riferimento dei progetti, considerate l'evoluzione continua della piattaforma e le nuove possibilità di sviluppo offerte nel tempo:

- Document.** offrono la possibilità di creare visualizzazioni graficamente molto accurate e permettono allo sviluppatore di definire precisamente le eventuali interazioni con l'utente. Devono essere lavorati in ambiente web e sono composti da diverse sezioni che consentono di organizzare e progettare in dettaglio il documento. Contengono principalmente tabelle e grafici realizzati inserendo attributi, metriche e altri oggetti provenienti da uno o più *dataset* (*report*, *Intelligent cube*) o essere costituiti da dati importati direttamente da un'origine dati esterna. Un problema riscontrato è che i *document* non sono *web-responsive*, cioè non si adattano automaticamente alle diverse dimensioni dello schermo e dovrebbero essere duplicati e modificati a seconda del dispositivo di utilizzo. Nel caso dei Censimenti, per esempio, la reportistica potrebbe non essere visualizzata correttamente sul tablet del rilevatore. Per mitigare questo limite è stata impostata l'opzione di adattare automaticamente alla pagina offerta dalla piattaforma. Questo tipo di oggetto è utilizzato nella reportistica per l'indagine soprattutto per visualizzazioni di tipo tabellare;
- Dossier:** sono stati introdotti da *Microstrategy* successivamente ai *document*. Come questi ultimi devono essere definiti in ambiente web e hanno la possibilità di mostrare graficamente le informazioni provenienti da uno o più *dataset*. Questi oggetti consentono una implementazione abbastanza rapida da parte dello sviluppatore e inoltre il risultato finale è *responsive*, cioè la visualizzazione si ridimensiona automaticamente a seconda del dispositivo utilizzato. Lo sviluppo di un dossier è piuttosto intuitivo: si procede selezionando le diverse tipologie di visualizzazioni da inserire che possono essere tabelle semplici o composte, grafici e mappe di diverso tipo, gli attributi e le metriche che li devono popolare e poi si definiscono le caratteristiche specifiche di ciascun oggetto. Durante la progettazione è visibile il risultato finale. Il dossier può essere organizzato in pagine e capitoli, può essere salvato dall'utente nella propria *library*, uno spazio dedicato a ciascun utente e che può essere condiviso, e quindi utilizzato anche offline.

Figura 4.8 - Sviluppo di un dossier. Andamento della rilevazione per stato dei numeri civici

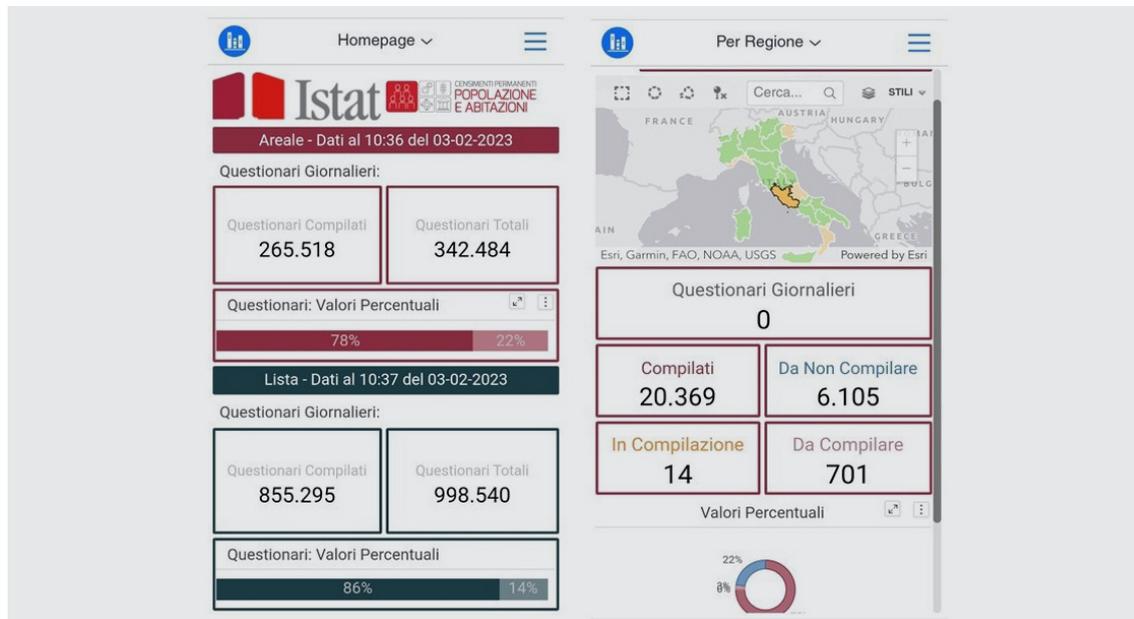


Fonte: Istat, Schermata da piattaforma *Microstrategy*

In questi anni è stata distribuita nella rete Istat, solo per alcuni utenti, anche una *app* consistente in un dossier che mostra in sintesi i dati principali sull'andamento della rilevazione per il Censimento della Popolazione e delle abitazioni, e per quello dell'Agricoltura.

4. La reportistica per le indagini

Figura 4.9 - App Censimento permanente della Popolazione e delle abitazioni. Anno 2022



Fonte: Istat, Schermata da app *Rapporti riassuntivi*

- **Report:** è un oggetto *Microstrategy* che rappresenta un *set* specifico di dati formattati ed è la base della reportistica. Lo strumento più accurato per il suo sviluppo è il *Developer*. Nei rapporti riassuntivi questa tipologia di oggetto acquisisce i suoi dati da un *Intelligent cube* e viene utilizzato come *dataset* e per le esportazioni di file in formato csv;
- **Esportazioni dei dati:** la reportistica delle indagini consente all'utente di esportare file di dati in formato *Excel*, pdf e csv utilizzando le relative icone mostrate nelle visualizzazioni. Anche se le tabelle dei dossier offrono la possibilità di effettuare lo scarico diretto nei primi due formati, si è deciso di creare nuovi oggetti configurati o richiamati tramite URL particolari che ne permettono l'esecuzione nel formato indicato, poiché i *report* prodotti devono contenere ulteriori informazioni rispetto alla tabella visualizzata, come richiesto dagli utenti utilizzatori nel caso dei file *Excel*.

Attualmente i dossier sono utilizzati quasi integralmente in alcune indagini come il Censimento permanente delle istituzioni non profit e in visualizzazioni della reportistica del Censimento permanente della Popolazione e delle abitazioni che includono mappe o grafici particolari.

4.1.5 Risultati e sviluppi futuri

Nel corso di questi anni, dal 2018 a oggi, gli accessi alla piattaforma della reportistica sono stati cospicui, numericamente paragonabili a quelli di Sgi. A partire dal Censimento permanente della Popolazione e delle abitazioni, la funzionalità dei rapporti riassuntivi è stata estesa ad altre indagini.

La reportistica è stata implementata seguendo un principio di generalizzazione delle varie edizioni per la stessa indagine, ma rimane ancora molto specializzata sulla indagine specifica. Tuttavia, andrebbe migliorata progettando, anche insieme agli utenti, un sistema di monitoraggio generalizzato che risponda a esigenze comuni a tutte le indagini.



Con i progressi tecnologici della piattaforma *Microstrategy* nel corso degli anni, anche la reportistica è evoluta, ad esempio con l'introduzione dello strumento del dossier, molto più flessibile e soprattutto *responsive*.

Sebbene attualmente la maggior parte dei *report* siano ancora di tipo *document*, si sta procedendo alla loro revisione per convertirli in dossier, anche con l'obiettivo di fornire agli utenti uno strumento che permetta analisi più approfondite e in maniera autonoma.

Inoltre, per quanto concerne gli ETL, ambito nel quale attualmente sono state sviluppate procedure PL/SQL personalizzate, sarebbe auspicabile l'utilizzo dei prodotti di *Data integration* già presenti in Istat nell'ecosistema delle piattaforme digitali, al fine di standardizzare il processo di caricamento e di aggiornamento dei dati.

Infine, con l'aggiunta di altre indagini, la piattaforma non è sempre stata in grado di supportare il carico a causa principalmente di limitate disponibilità di risorse del server. Per questo motivo *Microstrategy* è stata affiancata ad altre piattaforme per la reportistica, tra cui *Apex* che verrà trattata di seguito.

Parallelamente si sta provvedendo all'ampliamento dei server della piattaforma *Microstrategy* per garantire un miglioramento nelle prestazioni.

4.2 Piattaforma *Apex*

Apex è un *framework* basato sull'infrastruttura dei database *Oracle*, sviluppato dallo stesso produttore per la realizzazione di applicazioni web e data oriented. Presenta un'architettura a tre livelli in cui al *browser* è demandato il compito di inviare richieste tramite un server web. Le richieste vengono gestite da *Oracle Rest Data Services (ORDS)* e inviate a *Oracle database*. La logica di *business* e l'esecuzione dei processi risiedono nel database. Queste caratteristiche presentano vantaggi in termini di integrazione, di sicurezza e di scalabilità, e assumono rilevanza particolare nel contesto dell'Istat in cui *Oracle database* è il più diffuso strumento di gestione dati. Più recentemente è stato affiancato da altri strumenti orientati ai *big data*, come la piattaforma *Cloudera*, con architettura adatta per tipologie, flussi e volumi di dati diversi.

Tuttavia, le peculiari caratteristiche della gran parte dei dati trattati dall'Istat, fortemente strutturati, generalmente omogenei e connessi a esigenze prettamente legate alla sicurezza, all'affidabilità, alla consistenza e alla riservatezza, li rendono più confacenti al paradigma ACID (*Atomicity, Consistency, Isolation, and Durability*) dei database relazionali, piuttosto che a quello BASE (*Basically Available, Soft State, Eventual Consistency*), tipico invece della famiglia di strumenti dei *big data*. Di conseguenza, i database relazionali sono ancora gli strumenti prevalenti presso l'Istat.

Apex nasce con lo scopo di sfruttare le caratteristiche e la diffusione sul mercato dei database *Oracle*, ampliandone le potenzialità *web-oriented*, fruendo dell'architettura e del linguaggio dei database e riducendo al minimo la richiesta di competenze ulteriori. La sua impostazione *low-code*, l'uso di modelli e di procedure guidate, nonché la disponibilità di un ampio catalogo di *plugin*, consente di ridurre i tempi di sviluppo, fornendo inoltre supporto nativo ai grafici e ai dispositivi mobili. Tutto ciò comporta indubbi vantaggi a patto che se ne accettino i compromessi in termini di flessibilità e di personalizzazione delle applicazioni e rende questa tecnologia complementare piuttosto che alternativa ad altri e più specializzati strumenti, quali *Microstrategy*, che ne possono almeno in parte dividerne le funzioni. Vanno quindi ponderati volta per volta i casi che ne rendano l'utilizzo opportuno o meno.

4. La reportistica per le indagini

4.2.1 Rapporti riassuntivi del Censimento permanente delle istituzioni pubbliche

Il Censimento permanente delle istituzioni pubbliche ha rappresentato il primo caso di applicazione *Apex* per la reportistica delle indagini censuarie. La scelta di adottare un diverso strumento, peraltro già disponibile in Istat, nasce dalla necessità di alleggerire il carico sulle risorse afferenti a *Microstrategy*, su cui grava tuttora la gran parte della reportistica delle indagini, sia in termini di hardware sia di licenze software sia di personale impiegato.

Le specifiche richieste sulla reportistica hanno agevolato tale scelta, non essendo necessarie funzioni di presentazione o di *Business Intelligence* particolarmente avanzate. L'obiettivo era di fornire agli utenti interni all'Istat uno strumento per il monitoraggio dell'andamento della rilevazione. Trattandosi di personale con competenze specifiche, l'informazione prevista è stata essenzialmente di tipo quantitativo e basata su formati tabellari, da utilizzare anche per successive analisi ed elaborazioni.

L'accesso all'applicazione avveniva, anche in questo caso, tramite Sgi. Gli utenti erano tutti appartenenti alla rete dell'Istat e non è stato necessario effettuare differenziazioni per profili e per visibilità sui dati. L'unica eccezione era rappresentata dagli utenti di Trento e Bolzano/*Bozen*, esterni alla rete dell'Istat: la visibilità dei dati è stata limitata e, per ridurne i tempi di *release*, questa prima versione è stata fornita in modalità offline. La piattaforma *Apex* è integrata con il sistema di autenticazione *Shibboleth*. La modalità di accesso all'applicazione di reportistica da parte degli utenti di Sgi avveniva in *single sign-on*, tramite un URL.

Il *logout* dall'applicazione *Apex* avveniva attraverso più chiamate *Shibboleth* che effettuavano la stessa operazione anche per l'applicazione Sgi. L'architettura è basata su una componente di *back end* su database *Oracle* per la gestione dei dati e da una componente di *front end*, rappresentata da una applicazione web sviluppata in *Apex*, per la presentazione dei *report* e l'invio delle richieste degli utenti tramite *browser*.

Le linee guida per lo sviluppo dell'applicazione sono state dirette a sfruttare al massimo le prestazioni del database *Oracle*, demandandovi quanto più possibile la logica di *business* e l'esecuzione dei processi. Allo stesso tempo si è cercato di limitare al massimo le personalizzazioni rispetto alle funzioni native di *Apex* deprecando l'uso di codice ad hoc, quale ad esempio Javascript. L'obiettivo di tali linee guida era infatti di semplificare le manutenzioni correttive ed evolutive, di limitare elementi che potessero compromettere la sicurezza e il superamento dei test di vulnerabilità, propedeutici alla messa in esercizio dell'applicazione ed eseguiti per la prima volta in questa occasione su un'applicazione con questa tecnologia.

4.2.2 Back end: la base dati

Per quanto riguarda la base dati, si è scelto di seguire un approccio simile a quello scelto per le applicazioni realizzate con *Microstrategy*, utilizzando uno schema *Oracle* dedicato.

Tale schema è stato implementato sullo stesso server *Oracle* che ospita quello di Sgi, su cui si trovano dati necessari per generare i rapporti riassuntivi. Per il monitoraggio dell'indagine non sono richieste informazioni in tempo reale, ma è sufficiente una pianificazione intragiornaliera, con intervalli di aggiornamento dell'ordine delle ore.

L'alimentazione dei rapporti è molto strutturata ed è stato possibile sintetizzare tutti i dati necessari in un'unica vista con i dati in aggiornati, supportata da un'unica tabella di decodifica statica, da precaricare una volta sola. La soluzione di uno schema dedicato per la reportistica consente maggiore autonomia e flessibilità sulle componenti asservite all'ap-

plicazione *Apex*, anche in prospettiva evolutiva e di limitare l'impatto sullo schema utilizzato da Sgi sia in termini di proliferazione dei componenti sia di carico delle richieste.

La condivisione della stessa istanza di database con Sgi, consente di semplificare le problematiche di integrazione, essendo sufficiente la concessione di *grant* all'utenza applicativa *Apex* e agli oggetti presenti sullo schema dati di Sgi.

Inizialmente era stata presa in considerazione l'opzione di utilizzare direttamente la vista su Sgi come fonte diretta dei *report*.

La sua complessità e i tempi di interrogazione hanno portato a sfruttare l'assenza del requisito di aggiornamento in tempo reale e a propendere quindi per una sua materializzazione nello schema *Apex*, pianificata da parte di un apposito *job*. Per questo caso specifico si è ritenuto sufficiente un numero di quattro aggiornamenti giornalieri, ovviamente suscettibile di adeguamenti.

Ciò ha consentito di ridurre molto la latenza di interrogazione e di presentazione dei *report* da parte del *front end*, di ottimizzare le prestazioni con l'uso di indici, di ridurre gli accessi ai dati di origine sullo schema Sgi e di rendere consistenti tutti i *report* nella singola consultazione.

Alcune tabelle di decodifica statica, non disponibili nella base dati di Sgi, sono state precaricate *una tantum* dalle fonti di origine. Altre tabelle di servizio sono asservite all'applicazione e contengono dati di supporto ai *report* o ne tracciano l'aggiornamento.

Oltre alle tabelle, nello schema sono presenti indici, *job* e un *package* che raccoglie procedure che svolgono vari compiti e supportano le elaborazioni.

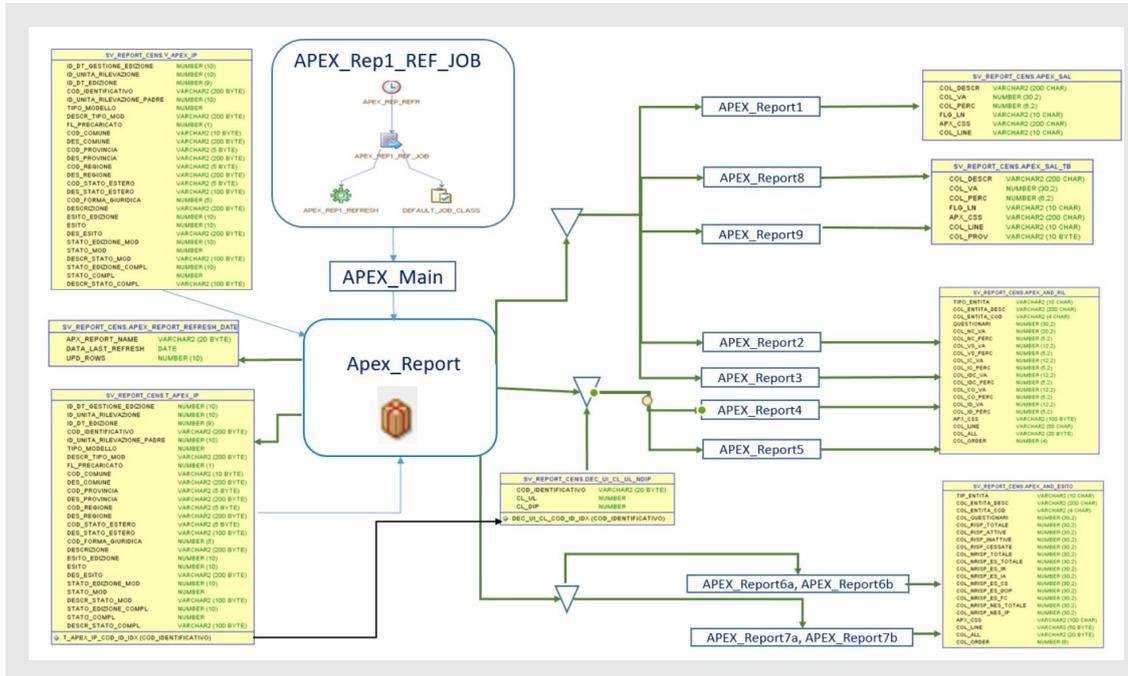
Gli oggetti presenti nella base dati sono i seguenti:

- **dati di origine:** i dati sono esposti tramite una vista presente nella base dati Sgi e materializzati nella base dati dell'applicazione *Apex* nella tabella con nome equivalente. La tabella viene ricreata da un *job* secondo la frequenza indicata nella pianificazione;
- **tabelle di servizio:** tabelle di servizio dell'applicazione, con prefisso APEX_, supportano il tracciamento degli aggiornamenti dei *report* (APEX_REPORT_REFRESH_DATE) e i calcoli dei *report* (APEX_SAL, APEX_SAL_TB, APEX_AND_RIL, APEX_AND_ESITO);
- **tabella di decodifica:** una tabella di decodifica statica con prefisso DEC_, contiene le classificazioni da lista iniziale richieste in alcuni *report* (DEC_UI_CL_UL_NDIP);
- **indici:** indici, con prefisso _IDX, sono presenti sulle tabelle;
- **job:** un *job*, con prefisso _JOB; è finalizzato alla sincronizzazione e produzione dei *report*. Aggiorna la tabella che contiene i dati di dettaglio ed esegue la pulizia delle tabelle *report*, esegue in successione tutte le procedure di calcolo dei *report*;
- **programmi:** il programma APEX_REP1_REFRESH chiama la procedura APEX_REPORT.MAIN;
- **pianificazioni:** la pianificazione APEX_REP_REFR stabilisce la frequenza di esecuzione del *job* con indicazione dei giorni e delle scadenze di esecuzione;
- **package:** un *package* APEX_REPORT raggruppa procedure asservite a varie funzioni e al calcolo dei *report*.

Come già osservato, la logica, le funzioni di calcolo dell'applicazione e l'onere computazionale risiedono prevalentemente a livello di *package*.

4. La reportistica per le indagini

Figura 4.10 - Diagramma del processo



Fonte: Istat

4.2.3 Front end: web application

Il *front end* dell'applicazione è *web-oriented*. Per la realizzazione è stato utilizzato il *framework Apex*, che ha consentito di ridurre i tempi di sviluppo, la scrittura di codice e di avere il supporto ai dispositivi mobili in modo nativo. La versione utilizzata è la 22.2.6.

Per quanto riguarda l'interfaccia, l'impostazione grafica e l'ergonomia richiamano i rapporti riassuntivi realizzati con *Microstrategy*, allo scopo di minimizzare l'impatto sugli utenti, presentando un ambiente e delle funzioni familiari. Essendo gli utilizzatori esclusivamente appartenenti alla rete dell'Istat, non è stato necessario recepire ulteriori requisiti, tipici dei servizi rivolti agli utenti esterni.

L'accesso all'applicazione avviene tramite Sgi. L'utente ha visibilità sulle rilevazioni di sua competenza, quindi può selezionare quella di suo interesse e, successivamente, mediante il menu "Rapporti riassuntivi", accede all'applicazione di reportistica.

Il Censimento permanente delle istituzioni pubbliche ha rappresentato il primo caso di utilizzo della tecnologia *Apex* per un'applicazione accessibile a utenti esterni alla rete dell'Istat. Come da processo standard di rilascio è stato quindi necessario un ciclo di test di vulnerabilità che ha certificato la rispondenza della piattaforma e dell'applicazione ai requisiti di sicurezza.

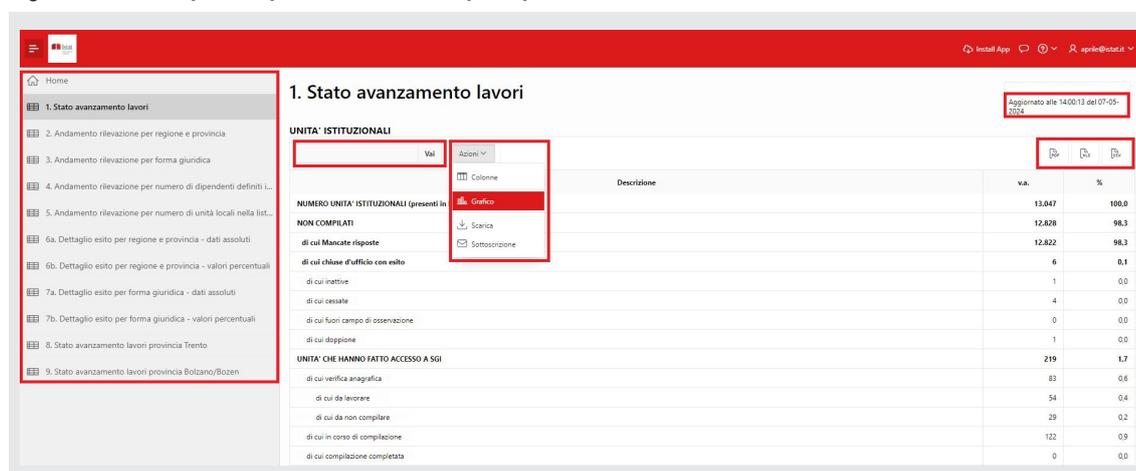
In questa prima *release* non sono stati previsti profili differenziati, pertanto tutti gli utenti avevano lo stesso livello di visibilità sui *report*. La pagina di accesso era quindi la medesima per tutti e conteneva l'elenco dei seguenti undici *report*:

- Stato avanzamento lavori;
- Andamento rilevazione per regione e provincia;
- Andamento rilevazione per forma giuridica;

- Andamento rilevazione per numero di dipendenti definiti in lista;
- Andamento rilevazione per numero di unità locali nella lista di partenza;
- Dettaglio esito per regione e provincia - dati assoluti;
- Dettaglio esito per regione e provincia - valori percentuali;
- Dettaglio esito per forma giuridica - dati assoluti;
- Dettaglio esito per forma giuridica - valori percentuali;
- Stato avanzamento lavori provincia Trento;
- Stato avanzamento lavori provincia Bolzano/*Bozen*.

La navigazione avveniva sempre nell'ambito della stessa finestra. Un'icona collocata nella pagina in alto a sinistra consentiva di attivare o disattivare il menu dei *report* in modo che fosse sempre visibile. Le informazioni relative alla data e all'ora di aggiornamento dei *report* erano sempre visibili così come i pulsanti di esportazione nei formati pdf, *Excel* e csv. Nella zona centrale in alto, si trovava la casella di ricerca testuale. Inoltre, era possibile svolgere ulteriori operazioni quali: la scelta delle colonne da mostrare nel *report* e il loro ordine; la creazione di grafici, selezionando colonne, variabili, funzione e tipo di grafico; la possibilità di inviare i *report* via email. *Apex* offre la possibilità di rilasciare automaticamente ulteriori funzioni agli utenti, ma si è scelto di limitarle per questo primo rilascio in modo da evitare un'eccessiva complessità dell'interfaccia.

Figura 4.11 - Esempio di *report* con le funzioni principali



Fonte: Istat, Schermata *Rapporti riassuntivi* da piattaforma *Apex*

4.2.4 Risultati e sviluppi futuri

La reportistica del Censimento permanente delle istituzioni pubbliche ha rappresentato, per alcuni aspetti, un caso pilota per l'utilizzo della tecnologia *Apex*. In questa occasione è stata sancita l'integrazione con l'infrastruttura dell'Istat, il suo sistema di autenticazione e la rispondenza ai requisiti di sicurezza.

Questa esperienza consente di prospettare possibili usi futuri per i casi in cui i requisiti non richiedano specificamente l'uso di *framework* più specializzati, quali *Microstrategy*.

La relativa semplicità di uso dell'ambiente di sviluppo, l'integrazione con il database

4. La reportistica per le indagini

Oracle e con il suo linguaggio PL/SQL, consente inoltre di sfruttare le infrastrutture e le competenze già presenti in Istat.

Riguardo all'applicazione esistente, ulteriori evoluzioni possono riguardare principalmente l'introduzione della profilazione e l'aggiunta di funzioni quali mappe o *report* interattivi, che consentano maggiore flessibilità e autonomia agli utenti.

Infine, un considerevole potenziamento sarà consentito dall'integrazione con la piattaforma di virtualizzazione *Denodo*, attualmente in corso, che renderà molto più flessibile l'accesso alle fonti dei dati e renderà disponibile un ulteriore strato per la distribuzione delle funzioni.

5. LA FORMAZIONE A DISTANZA¹

Le prime esperienze in Istat di ricorso a piattaforme di *e-learning* risalgono al 2007, all'interno di un progetto di formazione condotto dalla sede territoriale Istat della Sicilia in collaborazione con il servizio statistica della Regione Sicilia, rivolto ai referenti statistici regionali e comunali. Successivamente, nel 2010/2011, si avvalgono della formazione online anche i Censimenti agricoltura, popolazione e industria e le statistiche sull'ambiente, il Verde urbano e l'*Eco-management*.

Con la realizzazione del portale <https://formazione.istat.it/> a inizio 2016, e del portale <https://formazionereti.istat.it/> nel 2018, l'Istat si è dotata di uno strumento per la formazione a distanza ben strutturato e sicuro, che ha permesso la diffusione di contenuti formativi idonei al raggiungimento dei propri obiettivi istituzionali in tema di formazione di soggetti interni ed esterni: personale Istat, enti del Sistan e reti di rilevazione.

Sono stati chiari, fin dalle prime esperienze, i vantaggi che una piattaforma *e-learning* avrebbe potuto offrire e che di fatto ha reso possibili:

- **abbattimento delle distanze e dei costi di spostamento:** seguire un corso di formazione, interagire con l'insegnante o con gli altri partecipanti è semplice in qualsiasi luogo in cui ci si trovi, purché si abbia a disposizione un *device* e una connessione Internet;
- **materiali didattici sempre disponibili e consultabili da qualsiasi *device*:** il corsista che ha un dubbio, o non ricorda bene come affrontare una situazione particolare, può consultare il materiale formativo in ogni momento o chiedere supporto al docente o ai colleghi di corso;
- **contenuti flessibili e sempre aggiornati:** la formazione è continua e inserire o modificare i contenuti è un'azione che avviene in tempo reale, anche in caso di cambiamenti in itinere, modifica di una direttiva, precisazioni sui processi, aggiornamenti di linee guida, eccetera;
- **conciliazione dei tempi di vita:** i corsi sono modulari e possono essere seguiti in qualsiasi momento, anche assecondando i propri ritmi quotidiani, è il corsista a decidere quando e dove seguire il corso. È possibile interrompere un modulo, riprendendo la sessione in un momento successivo;
- **monitoraggio costante, valutazione e autovalutazione:** docenti e studenti possono analizzare l'andamento dei corsi, sia il proprio sia quello dell'intera classe; ognuno, in base al proprio ruolo, può conoscere il livello di apprendimento;
- **appeal ed engagement:** le lezioni possono essere integrate da audio, animazioni, video, *web conference*, *wiki*, *feedback*, eccetera, rendendo il corso più interessante e interattivo;
- **capacità di gestione dell'intero processo formativo:** il sistema permette di soddisfare le esigenze dei diversi soggetti interessati: dall'erogazione di un corso alla sua fruizione, passando per l'interazione fino alla fase di valutazione.

¹ Il Capitolo è stato redatto da Antonio Pitrone (paragrafi 5.1, 5.2, 5.2.1, 5.2.2, 5.2.3, 5.2.4, 5.2.5, 5.2.6 e 5.2.7), Emiliano Maniccia (paragrafi 5.3 e 5.3.1).

5.1 Piattaforma di formazione integrata per le indagini

Con i Censimenti permanenti l'infrastruttura, già utilizzata per il portale della formazione, è stata potenziata permettendo di soddisfare la formazione online di una rete di rilevazione composta da diverse migliaia di unità. Il portale *Formazionereti* a oggi conta 77 mila utenti con un accesso annuale tra i 15 e i 25 mila utenti attivi su almeno uno dei corsi offerti.

Il portale destinato alle reti di rilevazione è stato integrato sia con Sgi sia con il Sistema utenti, dedicato alla gestione centralizzata dell'autenticazione e della definizione di profili di autorizzazione degli utenti esterni che utilizzano i sistemi informatici dell'Istat.

L'integrazione del portale con il sistema Sgi è ottenuta tramite alcuni *plugin* originari della piattaforma, adeguatamente configurati, e grazie ad altre procedure automatizzate sviluppate ad hoc. I risultati ottenuti sono:

- la registrazione sulla piattaforma *Moodle* delle utenze registrate su Sgi;
- la sincronizzazione con l'anagrafica presente su Sgi: nome, cognome, email dell'intervistatore e territorio su cui opera;
- la registrazione delle utenze al corso appropriato;
- la creazione di nuovi gruppi di utenze e la cancellazione dei gruppi vuoti;
- l'iscrizione a gruppi di utenze, sulla base dell'appartenenza territoriale e della partecipazione a eventi formativi in aula definiti su Sgi.

L'integrazione con il Servizio utenti e il meccanismo di Sso migliorano l'utilizzo dell'applicativo in quanto evitano all'utente già collegato ad altri sistemi dell'Istat di dovere inserire nuovamente le proprie credenziali. Inoltre è possibile predisporre il sistema al supporto per Spid e Cie, permettendo l'utilizzo di credenziali uniche per l'accesso ai sistemi informatici di tutta la Pubblica amministrazione. Vengono infine offerte interfacce semplificate per il cambio e recupero *password*.

5.2 Architettura software e integrazione con Sgi e il Sistema utenti

Il cuore del sistema è *Moodle*, della cui offerta in termini di contenuti parleremo più avanti. È installato su una infrastruttura software che comprende:

- *Linux*;
- *Apache*;
- *Oracle*;
- *PHP: Hypertext Preprocessor* (PHP);
- *Shibboleth*;
- *OpenLDAP*;
- *Active Directory*.

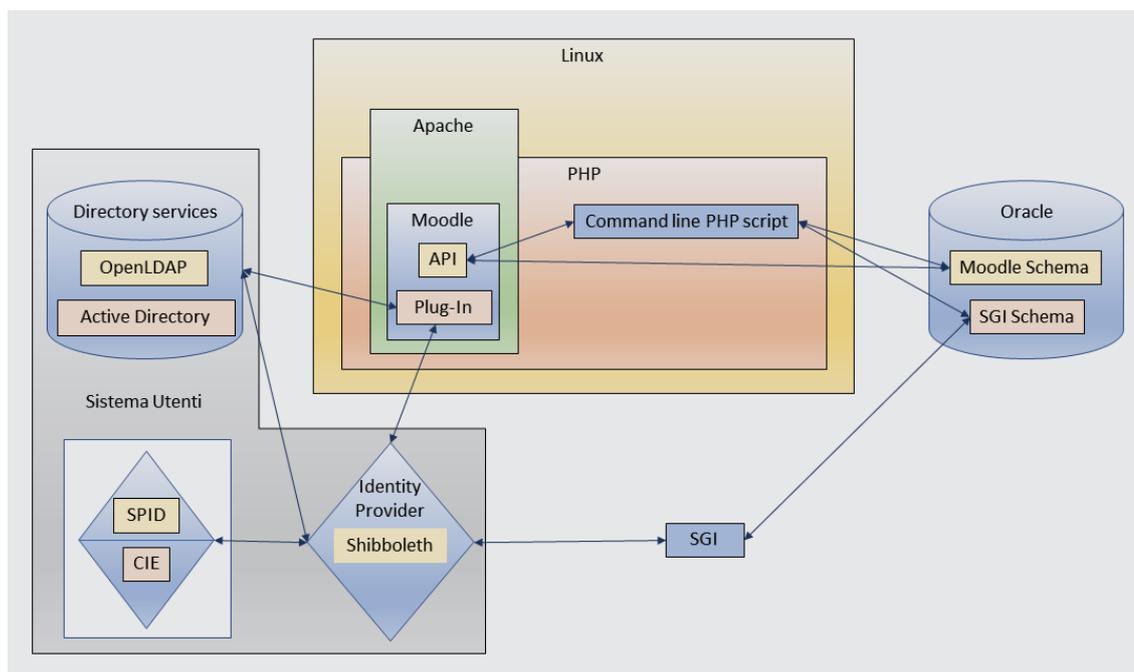
La piattaforma è stata resa più solida e più affidabile rispetto a quella originaria grazie a un database *Oracle* con due nodi in *cluster* e il potenziamento delle macchine di *front end* dedicate all'interfaccia web lato utenti.

5. La formazione a distanza

Moodle è un software modulare e la *community* che lo supporta, negli anni, ha sviluppato funzionalità aggiuntive offrendo numerosi *plugin*, richiamabili anche da riga di comando, che possono essere gestiti con un sistema centralizzato e flessibile. Inoltre, offre API con un'ampia offerta di funzionalità che permettono di sviluppare software e procedure integrati alla sua logica applicativa al fine di agevolare l'accesso ai contenuti e l'interconnessione con altri sistemi.

L'architettura software e le interconnessioni tra i sistemi *Moodle*, Sgi, Sistema utenti e *Identity* è rappresentato in Figura 5.1.

Figura 5.1 - Architettura software e interconnessioni



Fonte: Istat

Il processo realizzato per la gestione dei corsi si sviluppa in otto fasi che sfruttano i *plugin* offerti da *Moodle* e dalle sue API, richiamate da un insieme di *script* in PHP realizzati internamente. Tali *script* sono lanciati da linea di comando e, quindi, non integrati direttamente nel software *Moodle*; questa scelta è stata fatta per tenerli isolati dall'applicazione, assicurando che il livello di sicurezza offerto da *Moodle* non sia compromesso. L'esecuzione di *plugin* e *script* è stata pianificata per essere poi avviata periodicamente in maniera automatica dal sistema *Linux*, in particolare dal suo pianificatore *Cron*.

Le otto fasi in cui si sviluppa il processo sono:

1. sincronizzazione delle utenze registrate nella *Active Directory* con il sistema *Moodle*;
2. sincronizzazione delle utenze registrate nell'*OpenLDAP* con il sistema *Moodle*;
3. creazione nel corso *Moodle*, dedicato a una data indagine, dei gruppi utili a ripartire le utenze per ambito territoriale e sulla base della partecipazione agli eventi formativi in aula;
4. estrazione della lista degli utenti da iscrivere sia al corso sia ai gruppi creati nella fase precedente;
5. iscrizione degli utenti, estratti nella fase 4, al corso di riferimento;
6. inserimento degli utenti iscritti al corso nei gruppi creati nella fase 3 e sincronizzazione

- delle informazioni anagrafiche, estratte nella fase 4 dai profili utente di Sgi;
7. sospensione delle utenze già presenti sul sistema *Moodle*, ma non più censite all'interno delle *directory service* prima citate;
 8. *backup* dei corsi ancora attivi nel sistema *Moodle*.

In una prospettiva di gestione oculata delle risorse informatiche, finalizzata a non sovraccaricare il sistema e garantire un servizio ottimale per gli utenti, in considerazione anche della diversa priorità di esecuzione delle singole fasi rispetto alla loro frequenza, le otto fasi sono state raggruppate in quattro maxi fasi secondo il seguente schema:

- maxi fase 1: comprende le fasi 2, 4, 5, 6 con esecuzione ogni due ore tra le ore 6:00 e le ore 12:00 e alle ore 16:00 di ogni giorno;
- maxi fase 2: comprende le fasi 2, 3, 4, 5, 6, 7 con esecuzione alle ore 6:00 e alle ore 14:00 di ogni giorno;
- maxi fase 3: comprende le fasi 1, 2, 3, 4, 5, 6, 7 con esecuzione giornaliera alle ore 18:00;
- maxi fase 4: comprende le fasi 1, 2, 3, 4, 5, 6, 7, 8 con esecuzione a giorni alterni.

Inoltre, poiché l'esecuzione di ciascuna fase è propedeutica alla successiva, ciascuna maxi fase in esecuzione è bloccante per l'avvio di un'altra.

Ciò è stato possibile utilizzando un meccanismo di semafori che ricorre alla creazione di file di *lock* per ciascuna maxi fase. Infine, in caso di conflitti di esecuzione o di esecuzioni prolungate delle maxi fasi, è stato previsto l'invio di un *alert* agli amministratori della piattaforma *Moodle*.

Nei paragrafi seguenti sono descritte nel dettaglio le 8 fasi del processo.

5.2.1 Sincronizzazione delle utenze registrate nei Directory Service con Moodle

Il presupposto per l'iscrizione delle utenze registrate su Sgi nel portale *Formazionereti* per l'integrazione con il Sistema utenti è l'attestazione a un gruppo LDAP o *Active Directory* degli utenti abilitati all'accesso al portale *e-learning*; questa attività è svolta da Sgi quando l'utente è inserito in una indagine.

Successivamente il processo sfrutta due *plugin* offerti da *Moodle*:

- autenticazione su server LDAP: un *plugin* consente la configurazione dell'autenticazione tramite un server LDAP esterno e può leggere gli attributi dell'utente da LDAP e mappare i campi in *Moodle*;
- autenticazione su *Shibboleth*: questo metodo permette l'autenticazione agli utenti.

Il primo *plugin* scansiona il server LDAP e *Active Directory*, filtrando solo gli utenti attestati sul gruppo FORMAZIONERETI, e li registra alla piattaforma riportando *username*, nome, cognome ed email recuperati dai corrispondenti attributi LDAP. È stato opportunamente modificato in modo da procedere regolarmente con la registrazione delle informazioni di cui sopra, ma impostando per ciascuna utenza il metodo di autenticazione *Shibboleth*.

Il secondo *plugin* permette a *Moodle* di tenere traccia sia del processo di autenticazione che dovrà essere soddisfatto dall'utente con attestazione per l'utilizzo dell'autenticazione *Shibboleth*, sia dei parametri necessari affinché l'autenticazione sia riconosciuta come valida.

5.2.2 Creazione nel corso Moodle dei gruppi

Durante la fase 3 viene eseguito il primo degli *script* PHP sviluppati internamente; il suo compito è quello di preparare il corso dedicato a una data indagine per il raggruppamento delle utenze per ambito territoriale e sulla base della partecipazione agli eventi formativi in aula.

I raggruppamenti così definiti permettono una condivisione semplificata delle informazioni con la rete territoriale dell'Istat e propone un sistema più efficace di tracciamento, valutazione e monitoraggio delle attività didattiche.

Lo *script* accede a una vista, condivisa con lo schema *Oracle* di Sgi, relativa alle informazioni dei profili degli utenti registrati per la data indagine. Inoltre individua tutti i possibili gruppi necessari a soddisfare l'obiettivo e procede alla creazione dei gruppi, utilizzando l'appropriato metodo offerto dalle API di *Moodle*.

Infine, si procede all'attestazione, nei gruppi creati, delle utenze del personale Istat precedentemente definiti con il ruolo di *tutor* o docente. Tali ruoli permettono ai colleghi di potere accedere alle valutazioni di tutte le utenze della rete di rilevazione. L'attestazione di queste utenze ai vari gruppi consente di filtrare le valutazioni ottenute dagli studenti sia sulla base della distribuzione territoriale sia della partecipazione agli eventi formativi tenuti in aula.

5.2.3 Estrazione della lista di utenti

Anche nella fase 4 è eseguito uno degli *script* PHP sviluppati internamente. Il suo scopo è quello di recuperare, per una data indagine, le informazioni dei profili utente presenti su Sgi e, sulla base di queste ultime, creare due file di testo necessari alle fasi 5 e 6. Le informazioni raccolte accedendo allo schema *Oracle* di Sgi sono:

- *username*;
- nome;
- cognome;
- email;
- territorio di appartenenza (Provincia o Regione);
- ufficio di rilevazione;
- ufficio di coordinamento a cui appartiene l'ufficio di rilevazione;
- ruolo ricoperto nell'ambito della rete di rilevazione;
- evento formativo in aula a cui è stato iscritto.

Con le informazioni recuperate sono generati due file: il primo, in formato .csv, viene generato secondo le specifiche del *plugin* di *Moodle* e ha lo scopo di iscrivere gli utenti al corso. Il secondo, invece, è generato in base alle specifiche per l'inserimento massivo di utenti sulla piattaforma *Moodle* e consente sia l'iscrizione ai gruppi creati con la fase 3 sia la sincronizzazione tra *Moodle* e Sgi delle informazioni anagrafiche dei profili utenti.

Il formato e la struttura scelta per il secondo file consentono di avere a disposizione un'alternativa valida in caso di necessità: infatti, il file può essere utilizzato per ripristinare in poco tempo sia le iscrizioni a un dato corso sia l'articolazione in gruppi delle utenze, utilizzando la funzionalità nativa di *Moodle* per l'importazione massiva delle utenze. È stato inoltre previsto che questo file sia sottoposto a *backup* secondo una pianificazione prefissata.

5.2.4 Iscrizione degli utenti al corso di riferimento

Per soddisfare la fase 5 si è ricorso al *plugin* nativo di *Moodle* chiamato *Flat file enrollment*, che controlla la presenza di un file di testo nel percorso impostato: se il file viene trovato, si procede alla elaborazione e alla iscrizione degli utenti al corso.

I presupposti per l'utilizzo di *Flat file enrollment* sono i seguenti:

- le utenze devono risultare già registrate su *Moodle* per accedere alla piattaforma e tale condizione è soddisfatta con le fasi 1 e 2;
- il file csv deve essere creato secondo un dato formato, sia in termini di struttura sia di codifica caratteri e tale condizione è soddisfatta con la fase 4.

5.2.5 Inserimento degli utenti e sincronizzazione delle informazioni anagrafiche

Per inserire le utenze nei gruppi creati durante l'esecuzione della fase 3 e per l'aggiornamento dell'anagrafica utente, sincronizzata con le informazioni provenienti da Sgi, è stato implementato un ulteriore *script* in PHP eseguito durante la fase 6.

Per soddisfare le esigenze informative, i profili utente standard di *Moodle* sono stati integrati con dei campi aggiuntivi in cui sono registrate ulteriori informazioni provenienti da Sgi. Le maggiori informazioni di dettaglio riportate, per ciascun utente, sono:

- territorio di appartenenza (Provincia o Regione);
- ufficio di rilevazione;
- ufficio di coordinamento a cui appartiene l'ufficio di rilevazione;
- ruolo ricoperto nell'ambito della rete di rilevazione (supervisore od operatore).

Lo *script*, accedendo alle informazioni del secondo file, generato durante la fase 4, e interrogando *Moodle*, con l'utilizzo dei metodi offerti dalle sue API, verifica per ciascun utente iscritto al corso:

- se sia stato già inserito in ciascuno dei gruppi di sua pertinenza, creati durante la fase 3;
- se l'anagrafica del profilo risulti sincronizzata con le ulteriori informazioni di Sgi presenti nel file creato in fase 4.

Nel caso in cui l'utente non sia presente in uno dei gruppi di sua pertinenza, lo *script* procede a inserirlo nel gruppo, richiamando l'opportuno metodo offerto dalle API di *Moodle* per l'inserimento in un gruppo di un dato corso.

Nel caso in cui una delle informazioni dell'anagrafica del profilo utente non corrisponda a quella presente su Sgi, lo *script* procede ad aggiornarla richiamando l'opportuno metodo offerto dalle API di *Moodle* per la modifica dei campi del profilo.

5.2.6 Sospensione delle utenze già presenti sul sistema Moodle

Le utenze registrate sul Sistema utenti possono, nel tempo, risultare non più attive per i più svariati motivi quali ad esempio pensionamento del personale, inattività per lungo periodo, termine incarico per una data indagine. Durante la fase 7 le utenze che non hanno più possibilità di accesso al sistema *Moodle* vengono modificate e configurate nello stato

"sospeso", rendendole candidabili alla cancellazione, che avviene durante le regolari fasi di manutenzione della piattaforma. La cancellazione comporta la perdita di ogni informazione relativa all'utenza e registrata sul sistema *Moodle*.

5.2.7 Backup dei corsi ancora attivi sul sistema Moodle

L'ultima fase del processo qui descritto permette di effettuare il *backup* dei corsi presenti nel sistema. A tale scopo si è preferito non utilizzare il sistema di gestione dei *backup* nativo di *Moodle*, bensì un ulteriore *layer* applicativo denominato *Moosh*. È uno strumento a riga di comando che consente di eseguire le attività più comuni di *Moodle* utilizzando i metodi offerti dalle sue ampie librerie.

La scelta è ricaduta su *Moosh* per la grande offerta di impostazioni possibili nella configurazione dei *backup* e alla flessibilità del suo utilizzo, oltre naturalmente alla possibilità di averlo isolato dall'applicazione, analogamente a quanto fatto per gli *script* in PHP di cui abbiamo parlato precedentemente. In questo modo si garantisce che il livello di sicurezza offerto da *Moodle* non venga compromesso.

L'utilizzo di *Moosh* ha permesso di impostare a giorni alterni i *backup* solo per i corsi ancora attivi, mentre i dati dei *backup*, riguardanti i corsi precedenti alle ultime due settimane, vengono eliminati contestualmente all'ultima esecuzione di *Moosh* che ha dato esito positivo. Ciò non solo permette di potere ripristinare facilmente qualsiasi corso attivo nella condizione esistente nell'arco degli ultimi quattordici giorni, ma anche garantire che il sistema non sia sovraccaricato durante la sua esecuzione, nell'ottica di una gestione oculata delle risorse informatiche.

5.3 Potenzialità di Moodle

Moodle è un software *CMS open source* rilasciato con licenza GPL (*General Public License*), gratuito e modificabile da qualsiasi programmatore. Questa tipologia di licenza ha favorito la formazione di una vasta *community* internazionale, che nel tempo ha garantito un costante aggiornamento del software e un supporto tecnico sempre disponibile.

La scelta della piattaforma, da parte dell'Istat, è ricaduta sul software *Moodle* non solo per la sua notevole diffusione, con oltre 100 mila siti in 227 paesi del mondo, ma anche per la sua versatilità di applicazione in differenti contesti organizzativi: università, scuole superiori e primarie, dipartimenti governativi, organizzazioni sanitarie, organizzazioni militari, aeroporti o attività di liberi professionisti.

Gli elementi determinanti nell'orientare l'Istat alla scelta della piattaforma *Moodle* sono stati:

- sicurezza, potenza e offerta da una piattaforma ben strutturata;
- *layout* personalizzabile;
- struttura modulare dei contenuti;
- flessibilità di percorsi formativi online a supporto e integrazione dei tradizionali corsi in aula e in relazione alle differenti tipologie di utenza;
- offerta di supporto tecnico.

5.3.1 Strumenti della piattaforma

Il nucleo centrale della struttura di *Moodle* è il corso. L'iscrizione a un corso può avvenire con diverse modalità:

- autoiscrizione;
- iscrizione manuale da parte del docente;
- *script* automatici che importano le informazioni da altre sorgenti di dati.

I corsi possono essere organizzati in categorie predefinite e gli utenti possono ricoprire differenti ruoli per i singoli corsi o categorie di corso; chi accede a *Moodle* non ha privilegi speciali, finché non gli venga assegnato un ruolo per un corso specifico, una categoria di corsi o per l'intera piattaforma. I ruoli possono essere predefiniti (docenti, studenti, *tutor*, eccetera) o definiti all'occorrenza dall'amministratore della piattaforma.

Il materiale didattico può essere realizzato con tre differenti modalità:

- attività: uno strumento per favorire l'interazione tra studenti (*forum*, collaborazione in una *wiki*) o tra uno studente e un docente (domande/risposte di un quiz);
- risorsa: un elemento a supporto dell'apprendimento, ad esempio un file, un video o un collegamento a un sito web;
- blocco: a margine del corso, per fornire informazioni supplementari o link per facilitare l'apprendimento, attraverso una sorta di *widget* che mostra i risultati di un quiz, un calendario, i collegamenti a *blog*, un glossario o un file.

Alcune attività, risorse e blocchi sono standard ma ne possono essere implementate altre extra, tramite *download* o direttamente sviluppate ad hoc.

In riferimento ai due principali ruoli di docente e studente e agli strumenti standard forniti con la piattaforma, un docente può:

- attivare lezioni e attività didattiche collaborative sotto forma di dispense, attività ed esercitazioni utilizzando le risorse libro e pagina o inserendo bozze, presentazioni attraverso file o link a materiale esterno;
- organizzare le valutazioni per mezzo delle attività attraverso quiz, scegliendo tra domande a scelta multipla, risposta breve, risposta numerica, o compito, elaborato presentato da uno studente o da un gruppo. La valutazione, inoltre, può avvenire impostando degli obiettivi, associabili al completamento di una o più attività;
- gestire la classe;
- programmare le attività didattiche grazie al calendario.

La gestione della classe comprende:

- organizzazione in sottogruppi per differenziare attività e risorse secondo distinte tipologie di utenze;
- distribuzione certificati virtuali, non legalmente validi, di avvenuta partecipazione al corso al completamento di quest'ultimo;
- somministrazione di sondaggi tematici o per l'analisi dei *feedback*.

5. La formazione a distanza

Ogni componente didattica è inserita all'interno degli argomenti che sono le parti costitutive dei corsi e garantiscono la modularità dei contenuti in base alle esigenze del docente. Oltre ai moduli già citati, il docente potrà attivare ulteriori elementi, quali *chat*, *forum*, *wiki*, *blog*, glossario o database.

La piattaforma indirizza gli studenti verso un approccio attivo all'apprendimento, offrendo possibilità di:

- rappresentare le proprie conoscenze attraverso *blog*, *wiki*, glossario, test, database;
- comunicare, collaborare, condividere attraverso *chat*, *forum*, *blog* e *wiki*.

Infine, l'accesso alla piattaforma può avvenire sia tramite *browser* sia tramite l'*app mobile* fornita dagli *store* Apple e Android.

6. IL PORTALE DELLA RETE¹

Il Portale della rete è stato predisposto dall'Istat per supportare l'organizzazione delle rilevazioni censuarie ed è accessibile a tutti i soggetti che fanno parte della rete territoriale di rilevazione del Censimento permanente.

Il sito è stato sviluppato su piattaforma CMS (*Content Management System*) denominata *Typo3*, un sistema di gestione dei contenuti sviluppato in PHP che utilizza un database *MySQL*. Per gestire diverse esigenze architetture emerse e per integrare l'applicativo con le altre piattaforme, è stata effettuata l'analisi, la progettazione e lo sviluppo di moduli specifici come ad esempio l'interfacciamento a *Shibboleth*, l'interrogazione di database *Oracle* per la fase di autorizzazione, la gestione di ricerca delle mappe.

6.1 Autenticazione e autorizzazione

L'accesso al sito è consentito solo dopo la procedura di autenticazione che avviene tramite *Shibboleth*. Anche per questo sito, analogamente a quanto fatto per tutti gli altri, è possibile utilizzare sia le credenziali inviate da Istat sia Spid e Cie. Dopo la fase di autenticazione, segue la procedura di autorizzazione per certificare che l'utenza possa accedere alle informazioni richieste, attraverso il controllo sull'anno di riferimento e sull'area censuaria di competenza.

Figura 6.1 – Suddivisione in anni e tipologie di indagini per il Censimento permanente della Popolazione e delle abitazioni

Il sito della rete è il portale che l'Istat ha predisposto per supportare l'organizzazione delle rilevazioni censuarie Areale e da Lista ed è accessibile a tutti i soggetti che fanno parte della rete territoriale di rilevazione del Censimento permanente.

Dopo l'autenticazione è necessario selezionare la rilevazione di interesse (Areale o da Lista).

Il sito contiene tutta la documentazione necessaria allo svolgimento delle rilevazioni e si configura come un **grande repository** per tutta la documentazione ufficiale tra l'Istat, gli organi intermedi di rilevazione e le famiglie. Scegliendo l'indagine di interesse si accede alla documentazione dedicata (lettera per il sindaco, circolari, comunicazioni, PGC, lettere informative per le famiglie, lettere di sollecito, ecc.) e a tutto il materiale di rilevazione (locandina, tesserino del rilevatore, ricevuta di avvenuta compilazione, biglietto da visita, questionario e guida anche in altre lingue, mappe delle sezioni di censimento, ecc.).

Il costante aggiornamento del Sito lo delinea come il principale luogo di raccolta di tutta la documentazione disponibile.

Fonte: Istat

¹ Il Capitolo è stato redatto da Andrea Marsico.



Per profilare l'utenza nella maniera opportuna, vengono effettuate interrogazioni alla base dati Sgi e a quella del portale della formazione. Nel caso in cui l'utenza sia abilitata a più tipologie di rilevazione, questa può accedere selezionando quella di interesse. La Figura 6.1 mostra la scelta per l'accesso all'indagine areale e alla lista del Censimento permanente della Popolazione e delle abitazioni suddivisa per i diversi anni.

6.2 Organizzazione del sito

Il sito contiene tutta la documentazione necessaria allo svolgimento delle rilevazioni e si configura come un unico *repository* per tutti i documenti ufficiali necessari alle interazioni tra l'Istat e gli organi intermedi di rilevazione. Il suo costante aggiornamento lo delinea come il principale luogo di raccolta di tutto il materiale disponibile. Il sistema è tradotto anche in lingua tedesca per agevolare le province di Trento e Bolzano/*Bozen*, che partecipano in modo attivo alle rilevazioni censuarie.

Figura 6.2 – Tipologia di materiale per le indagini

The screenshot displays the Istat website interface for permanent censuses. At the top, there are navigation links: 'Mappa del sito', 'Scelta Rilevazione', 'Cambio password', 'Logout', and language options 'ITA' and 'DE'. The main header features the Istat logo and the title 'CENSIMENTI PERMANENTI POPOLAZIONE E ABITAZIONI'. Below the header is a navigation bar with 'HOME', 'DOCUMENTI', 'MATERIALE DI RILEVAZIONE', 'STRUMENTI', 'CONTATTI', and 'FAQ'. The 'DOCUMENTI' section is active, showing a breadcrumb trail: 'Home > Scelta Rilevazione > Areale 2022 > Documenti'. A 'MENU' box lists several document types: 'Lettera informativa', 'Proroga della rilevazione', 'Circolari', 'Comunicazioni', and 'Piano Generale di Censimento (PGC)'. The main content area explains that this section contains official documents for the 'Areale' survey, including the 'Piano Generale di Censimento' which details objectives, characteristics, and methodology. It also mentions that circulars and communications are available, along with a letter sent to participating families. The text notes that documents are available in multiple languages (Italian, German, Slovenian, French, English, Spanish, Albanian, Romanian, Arabic, Chinese, Russian) to facilitate contact with foreign residents.

Fonte: Istat

Dopo avere scelto l'indagine di interesse, si accede alla consultazione suddivisa per sezioni in base alla tipologia di materiale caricato.

Nella sezione **Documenti** sono presenti, ad esempio, lettera per il sindaco, circolari, comunicazioni, Piano generale di Censimento, lettere informative per le famiglie, lettere di sollecito e altri documenti ufficiali di riferimento per la rilevazione.

Il Piano generale di Censimento è uno dei principali documenti: in esso sono contenuti infatti gli obiettivi, le caratteristiche, le regole metodologiche, le modalità operative, il campo di osservazione e l'organizzazione generale della rilevazione.

Nella stessa sezione si trovano anche le circolari e le comunicazioni che regolamentano specifiche fasi o singole attività della rilevazione. Inoltre è disponibile la lettera informativa spedita per posta alle famiglie coinvolte nella rilevazione. Questi atti ufficiali sono scritti anche nella versione bilingue in italiano-tedesco, italiano-sloveno e italiano-francese; sono tradotti anche in altre lingue, come inglese, spagnolo, albanese, rumeno, arabo, cinese e russo, in modo tale da agevolare il contatto del rilevatore con le famiglie straniere residenti in Italia.

6. Il Portale della rete

La sezione **Materiali di rilevazione** contiene documenti di ausilio per la rilevazione sul campo, come la locandina, il tesserino del rilevatore, la ricevuta di avvenuta compilazione, il biglietto da visita, il questionario e la guida. Anche per questi materiali è previsto il multilinguismo, sempre con lo scopo di facilitare il lavoro del rilevatore durante la compilazione del questionario da parte delle famiglie straniere.

Nella sezione **Strumenti** sono presenti utilità specifiche a ogni indagine: ad esempio, per la rilevazione areale è presente la voce *Mappe*, in cui sono contenute le mappe delle sezioni di Censimento campionate, come strumento di ausilio per la verifica preliminare del campione di indirizzi e delle sezioni e la ricognizione del territorio. Dopo avere selezionato la regione di interesse, si accede per ciascuna provincia all'elenco dei comuni campione. Per ogni comune è disponibile il file kml; se la porzione del territorio in cui condurre la rilevazione è l'intera sezione, si dispone delle carte in pdf per ciascuna delle singole sezioni di Censimento campionate.

Il sito contiene anche un'area dedicata alle **Faq** (*Frequently asked questions*), organizzate per area tematica. Esiste una funzione specifica che permette di individuare Faq di interesse e, attraverso il *download*, scaricare l'elenco completo in formato *Excel*.

Nell'area **Contatti** è possibile reperire i riferimenti per richiedere assistenza e chiarimenti durante tutto il periodo delle attività censuarie. Selezionando la regione di interesse, è possibile raggiungere l'indirizzo email dedicato da ogni ufficio regionale di Censimento.

Il sistema permette in qualsiasi momento di raggiungere sia Sgi sia la piattaforma per la formazione statistica e il sito istituzionale Istat. Completano l'offerta del sito il navigatore Ateco e il navigatore delle professioni, presenti sul sito istituzionale dell'Istat, che possono essere consultati durante la compilazione del questionario elettronico per permettere al rispondente di identificare rispettivamente l'attività economica e la professione. All'interno del portale è presente un'area riservata esclusivamente agli utenti Istat.

7. LA GESTIONE DEI DATI E L'UTILIZZO DEI REGISTRI¹

L'Istat negli ultimi anni sta lavorando a un progetto ambizioso e strategico: la costruzione del Sistema integrato dei registri statistici (Sir). A partire dall'esperienza maturata sui registri per le statistiche economiche, si stanno realizzando registri relativi a individui e famiglie, luoghi, lavoro, redditi, istruzione e formazione. L'obiettivo è la costruzione di un'architettura dati di qualità, popolata a partire dall'acquisizione, armonizzazione e integrazione degli archivi amministrativi e supportata dai risultati di indagine, che consenta analisi annuali e longitudinali nel tempo, *intra-domain* e *inter-domain*, garantendo la tutela della riservatezza, utilizzando appieno il potenziale informativo dei dati amministrativi e riducendo l'impiego delle indagini.

Il cambiamento in corso è profondo: il passaggio da un approccio *survey-based register assisted* a quello *register-based survey assisted* pone i registri alimentati dalle fonti amministrative al centro della produzione statistica. Questo cambiamento di approccio ha consentito il passaggio dai Censimenti generali decennali ai Censimenti permanenti. I registri diventano così sia punto di partenza per le indagini, con la fornitura dei dati per il campionamento e per il supporto sul campo, sia punto di arrivo, attraverso il riporto all'universo dei dati raccolti, la validazione e la diffusione. La costruzione dei registri è un tema molto complesso e sfidante, sia dal punto di vista statistico-metodologico sia dal punto di vista informatico e di governance, in quanto gli aspetti da affrontare sono molteplici:

- la governance degli accessi ai dati e l'aderenza alle indicazioni del GDPR e alle *policy* di sicurezza dell'Istat;
- la qualità dei processi e delle variabili prodotte;
- l'utilizzo armonizzato delle fonti in modo da snellire e condividere gli sforzi per le analisi e le elaborazioni comuni a più registri;
- la definizione delle regole e delle metodologie di deduplicazione, codifica, integrazione, stima, controllo, correzione e validazione dei dati;
- la gestione della semantica dei dati attraverso l'introduzione di ontologie e opportuna metadatazione;
- la produzione di output su singolo registro e integrati da più registri, annuali e longitudinali nel tempo;
- l'automatizzazione e il monitoraggio dei processi;
- la documentazione dei processi e dei prodotti e la loro accessibilità per gli utenti.

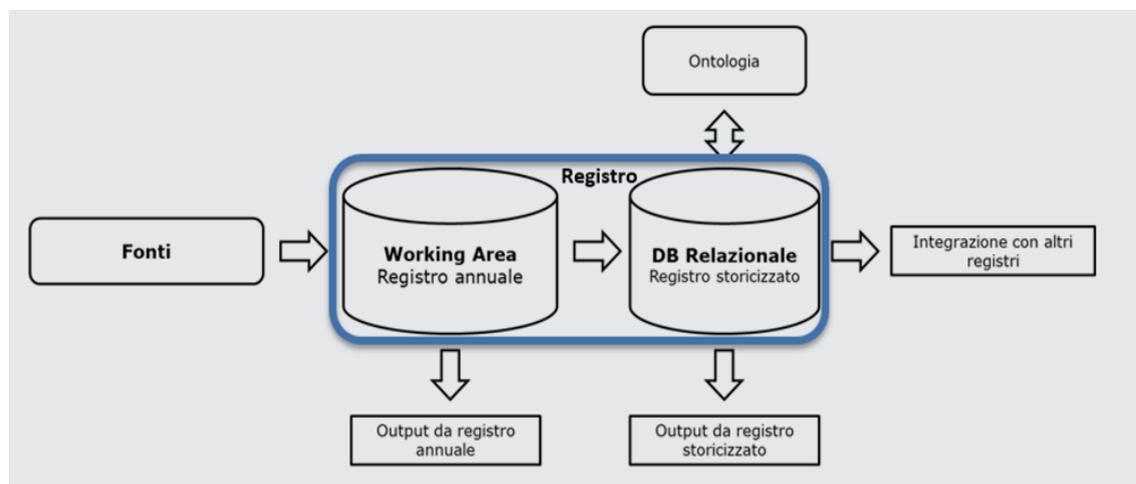
La realizzazione dei singoli registri e delle loro integrazioni è pertanto complessa e necessita di svariate competenze, per cui viene svolta attraverso la stretta collaborazione di diverse strutture dell'Istat: dai settori tematici, responsabili della produzione e validazione dei dati di uno specifico dominio, a quelli che offrono servizi trasversali, quali la raccolta dei dati, la predisposizione di metodologie statistiche e architetture per la definizione dei processi, la realizzazione e manutenzione di database e di sistemi informativi a supporto.

¹ Il Capitolo è a cura di Giulia Vaste e Daniela Casale, ed è stato redatto da: Giulia Vaste (paragrafo 7); Manuela Marrone (paragrafi 7.1, 7.1.1, 7.1.2); Franco Tulli (paragrafi 7.2, 7.2.1, 7.2.2, 7.2.3); Maria Pia di Maio (paragrafo 7.3); Roberta Benedetti (paragrafo 7.4); Francesco Bosio (paragrafo 7.5); Annunziata Fiore (paragrafo 7.5.1).

Si procede attraverso un approccio incrementale e iterativo, sia per la progettazione e la realizzazione sia per la manutenzione: da una parte i registri molto complessi sono stati suddivisi in moduli costruiti progressivamente, dall'altra anche una semplice modifica introdotta in una fonte amministrativa va analizzata, ne va valutato l'impatto sui processi esistenti e va gestita apportando le modifiche necessarie all'interno dei processi stessi. Inoltre l'impiego di una nuova fonte o un differente e più approfondito impiego di una fonte già in uso può arricchire e migliorare il potenziale informativo di un registro, per cui la sua manutenzione evolutiva è un'operazione estremamente frequente e delicata.

Lo schema generale, che si applica ad alcuni registri, definito dai metodologi responsabili dell'architettura apposita, è illustrato sinteticamente in Figura 7.1. I registri sono archivi *Oracle*, anche di milioni di *record*, popolati attraverso procedure PL/SQL, *Sas*, *R* e sistemi *batch* esterni. Le fonti amministrative sono acquisite dall'Istat tramite la Direzione centrale per la raccolta dati e vengono poi messe a disposizione, prive di dati identificativi, dei registri che le utilizzano, che provvedono a deduplicarle, integrarle, elaborarle per ottenere delle tabelle annuali, contenute nella *working area*. Da tali tabelle si producono degli output e si popola lo schema relazionale del registro, contenente la storicizzazione delle variabili di interesse, che consente analisi longitudinali nel tempo e su cui è mappata l'ontologia del dominio di interesse. Il popolamento delle tabelle annuali viene lanciato generalmente un paio di volte l'anno per singolo registro, per produrre la versione provvisoria, elaborata con un sottoinsieme di fonti amministrative, e la versione consolidata, elaborata invece con tutte le fonti disponibili per l'anno di interesse.

Figura 7.1 - Schema generale di costruzione di un registro



Fonte: Istat

Alcuni registri sono molto articolati e suddivisi in più moduli. Il livello di realizzazione complessivo dei singoli registri e dei relativi moduli è molto vario: alcuni sono in fase di analisi e progettazione; altri vengono raffinati con versioni successive che sfruttano più ampiamente alcune fonti o ne integrano di nuove; altri ancora sono già utilizzati per i processi di produzione dell'Istat e subiscono manutenzioni evolutive a seguito, ad esempio, di modifiche nelle fonti di input o inserimento di nuove variabili. I registri economici sono nati già da diversi anni e sono correntemente utilizzati in produzione, supportati da interfacce grafiche dedicate, descritti nei paragrafi 7.1 e 7.2.

Per i registri nati più recentemente, nel paragrafo 7.3 si illustrano gli aspetti informatici che ne supportano la produzione e l'utilizzo, anche ai fini censuari e si rimanda a pubblicazioni tematiche relative all'impiego del Registro base degli individui e del Registro statistico base dei luoghi per il Censimento permanente della Popolazione e delle abitazioni e del Registro tematico del lavoro per il Censimento permanente delle istituzioni pubbliche. Il paragrafo 7.4 illustra una sperimentazione, condotta sul Registro base degli individui, sugli indicatori di qualità e di monitoraggio dei processi dei registri, che si stanno studiando presso l'Istat. Infine, il paragrafo 7.5 è dedicato al supporto informatico alla validazione dei dati censuari.

7.1 Registro delle imprese e Censimenti economici

Il Registro delle imprese si inserisce nell'ambito del più complesso sistema dei registri economici, realizzati tramite una stretta collaborazione tra i settori tematici delle statistiche economiche e il settore informatico, aventi ciascuno un diverso grado di sviluppo e di industrializzazione.

I principali registri che compongono il sistema dei registri economici sono:

- il Registro delle imprese, nel quale l'unità statistica di riferimento è l'unità giuridica, che nella maggioranza dei casi corrisponde all'impresa;
- il Registro delle unità locali, nel quale l'unità statistica di riferimento è l'unità locale appartenente all'unità giuridica, che nella maggioranza dei casi ha più unità locali associate;
- il Registro dei gruppi di impresa, nel quale l'unità statistica di riferimento è il gruppo di impresa, che rappresenta una struttura complessa che può includere diverse unità giuridiche e diverse imprese, collegate da legami funzionali o proprietari;
- il Registro delle istituzioni pubbliche, nel quale l'unità statistica di riferimento è l'unità istituzionale del settore pubblico;
- il Registro delle istituzioni non profit, nel quale l'unità statistica di riferimento è l'unità istituzionale del settore privato che opera senza fini economici con finalità solidaristiche.

Il Registro delle imprese ha la più lunga tradizione e la più consolidata struttura produttiva in Istat. Nasce nel 1996 in base al Regolamento (CEE) n. 2186/1993 del Consiglio dell'Unione europea, relativo al coordinamento comunitario dello sviluppo dei registri di imprese utilizzati a fini statistici, poi sostituito dal Regolamento (CE) n. 177/2008, a sua volta abrogato dal Regolamento (UE) 2019/2152 del Parlamento europeo e del Consiglio.

L'impresa si definisce come la più piccola combinazione di unità giuridiche che costituisce un'unità organizzata per la produzione di beni e servizi che fruisce di una certa autonomia decisionale in particolare per quanto attiene alla destinazione delle sue risorse correnti. Un'impresa esercita una o più attività in uno o più luoghi. Il Registro delle imprese è costituito dalle unità economiche che esercitano arti e professioni nelle attività industriali, commerciali e dei servizi alle imprese e alle famiglie e fornisce informazioni:

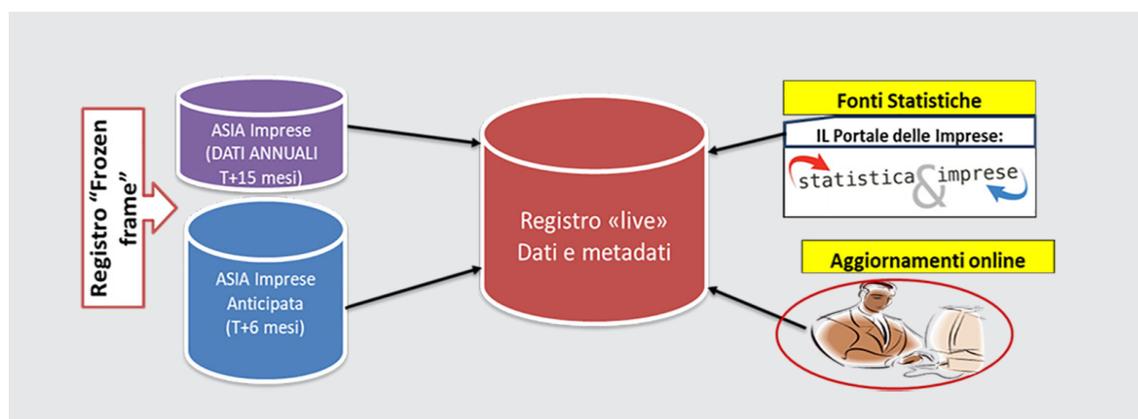
- di struttura, per quanto concerne attività economica, addetti dipendenti e indipendenti, forma giuridica, data di inizio e fine attività, volume di affari, stato di attività, informazioni sulla residenza sul territorio nazionale;
- identificative, come codice identificativo dell'impresa, indirizzo e codici amministrativi della localizzazione della sede legale.

Sono escluse dal campo di osservazione del Registro delle imprese, indipendentemente dalla loro attività economica, le unità classificate come istituzioni pubbliche e come istituzioni private non profit, che confluiscono nei rispettivi registri economici di competenza.

7.1.1 Sistema informativo del Registro delle imprese: fonti, database e applicativi a supporto

Il sistema informativo del Registro delle imprese, anche definito “registro *live*” delle imprese, è il risultato di una complessa integrazione di fonti amministrative, fonti statistiche sulle imprese e processi di aggiornamento online ed è volto a contenere le informazioni più complete e aggiornate per ogni componente del sistema (Figura 7.2).

Figura 7.2 - Schema della composizione del Registro delle imprese



Fonte: Istat

Le principali fonti amministrative, gestite da enti pubblici o da società private, oggetto della fase di integrazione nel registro delle imprese sono:

- gli archivi gestiti dall'Agenda delle entrate per il Ministero dell'economia e delle finanze, quali l'Anagrafe tributaria delle persone giuridiche e delle persone fisiche con partita Iva, le dichiarazioni annuali delle imposte indirette, le dichiarazioni dell'imposta regionale sulle attività produttive, gli studi di settore, i dati del modello Unico;
- i Registri delle imprese delle Camere di commercio, industria, artigianato e agricoltura e gli archivi collegati dei soci delle società di capitale e delle persone con cariche sociali;
- gli archivi dell'Istituto nazionale della previdenza sociale;
- le denunce retributive mensili EMens per gli occupati dipendenti e le dichiarazioni trimestrali della manodopera agricola;
- la cassa integrazione a pagamento diretto;
- le posizioni contributive degli imprenditori artigiani e commercianti;
- la gestione separata parasubordinati;
- l'archivio delle denunce contributive lavoratori dello sport e dello spettacolo (*ex* Enpals);
- l'archivio dei bilanci consolidati e di esercizio delle Camere di commercio.

7. La gestione dei dati e l'utilizzo dei registri

Se le fonti amministrative costituiscono la base informativa primaria del Registro delle imprese, le fonti statistiche sono invece coinvolte nella fase di aggiornamento e sono il risultato di processi differenti, che vedono coinvolti i revisori del registro, responsabili di tale attività durante tutto l'anno.

Si possono individuare fonti statistiche interne al Registro, come quelle provenienti dall'aggiornamento del Registro dei gruppi di impresa e dall'attività di *profiling* svolta dai revisori.

Sono invece considerate fonti statistiche esterne al registro, le informazioni acquisite mediante lo scambio di dati con le principali indagini statistiche strutturali e congiunturali dell'Istat. Ciò avviene tramite la condivisione delle informazioni raccolte da *Statistica & Imprese* (<https://imprese.istat.it/>), il Portale delle imprese gestito dall'Istat. Si tratta di un sistema di servizi statistici dedicato alle imprese e agli operatori economici per la raccolta e la restituzione di informazioni nell'ambito di tutte le rilevazioni condotte dall'Istat.

La complessa mole di dati che compone il sistema informativo del Registro delle imprese è memorizzata e organizzata all'interno di un database relazionale, che poggia fisicamente su una istanza di database *Oracle Exadata*. Il database del Registro delle imprese è individuato dalle seguenti entità principali:

Componente

Le informazioni contenute nel database si riferiscono a un soggetto specifico definito per l'appunto "componente", che ha anche una specifica codifica della tipologia. Alcune componenti sono le unità legali, le imprese, i gruppi di imprese, le unità locali, le istituzioni pubbliche, le unità locali pubbliche.

Carattere

Il componente ha caratteristiche che variano nel corso del tempo e sono rappresentate dal "carattere". Ogni specifico carattere è di proprietà di uno specifico componente e ha anch'esso una specifica codifica della tipologia.

I caratteri dei componenti sono distinti per tipologia nei seguenti gruppi:

- caratteri di classificazione, che forniscono informazioni espresse dal valore di variabili qualitative, come ad esempio lo stato legale, l'attività economica principale, secondaria o ausiliaria;
- caratteri dimensionali, che forniscono informazioni espresse dal valore di variabili quantitative quali ad esempio il numero dei dipendenti o il fatturato;
- caratteri di localizzazione, quali l'indirizzo dell'impresa, corredato del codice Istat della provincia, del codice Istat del comune e del codice di avviamento postale (CAP);
- caratteri di tipologia, che identificano l'appartenenza di alcuni soggetti a particolari sottoinsiemi, quali: l'azienda stagionale, i gruppi di imprese, le istituzioni;
- caratteri di trasformazione, che identificano particolari eventi, ad esempio la variazione della partita Iva, la fusione e il cambiamento di forma giuridica;
- caratteri di identificazione, quali il codice fiscale e la ragione sociale;
- caratteri di relazione con altri componenti, quali la relazione tra l'impresa e l'unità giuridica;
- caratteri di dimensione, che individuano la distribuzione di una variabile quantitativa secondo i valori di una variabile di classificazione, come l'impiegato per sesso.

Ogni carattere di un componente ha i seguenti attributi principali: la specifica provenienza dei dati, se da Agenzia delle entrate e da Censimento o da altra indagine dell'Istat; la modalità di attribuzione del valore, che può essere ad esempio osservato o stimato; il riferimento al processo che lo ha identificato; l'integrazione della fonte amministrativa o della rilevazione statistica o dell'acquisizione diretta; il riferimento temporale.

Tempo di riferimento

Il componente può modificare le sue caratteristiche nel corso del tempo. Tutti i valori che i caratteri assumono nel tempo vengono memorizzati nel database del Registro delle imprese. Ogni cambiamento di un carattere è memorizzato sul database con:

- data di riferimento, che nel database è denominata DATA_RIF_PROCESSO;
- inizio della data di validità, che nel database è denominata DATA_INIZIO_CARATTERE;
- data di fine validità, che nel database è denominata DATA_FINE_CARATTERE.

Il valore corrente di ogni carattere ha la DATA_FINE_CARATTERE non valorizzata. Dal punto di vista fisico il Registro delle imprese presenta le seguenti tipologie di tabelle:

Tablelle dei metadati

Sono tabelle che descrivono le informazioni del Registro, quali ad esempio la classificazione della forma giuridica o la classificazione dell'attività economica, e le relazioni tra i dati; contengono i percorsi fisici da seguire per la gestione dei dati, mediante tabelle associative.

Una tabella dei metadati fondamentale è DBASIA_TIP_COMP_CARATTERE, che contiene:

- le informazioni sul componente proprietario del carattere;
- la tabella dei dati in cui sono memorizzate le informazioni;
- il percorso per recuperare le informazioni a partire da un componente non proprietario.

Di seguito, una visualizzazione dal database del registro del contenuto della tabella DBASIA_TIP_COMP_CARATTERE, come esempio di relazione tra un carattere e un componente.

Figura 7.3 - Estrazione dei dati della tabella DBASIA_TIP_COMP_CARATTERE

CAR	TIP	TAB
01001	001	DBASIA_GRUPPI
01101	001	DBASIA_GRUPPI
03015	001	DBASIA_CARQUANT
05007	001	DBASIA_CARCLASS
05010	001	DBASIA_CARCLASS
06002	001	DBASIA_CARREL
06005	001	DBASIA_CARREL
07001	001	DBASIA_CARIDENT
07002	001	DBASIA_CARIDENT
07003	001	DBASIA_CARIDENT
07013	001	DBASIA_PAROLE
07013	001	DBASIA_PAROLE_GRUPPI
08005	001	DBASIA_CARTIPOL
08006	001	DBASIA_CARTIPOL

Carattere Componente «proprietario» Nome Tabella dei Dati

Tablelle dei dati

Queste contengono le informazioni specifiche del registro, in particolare i valori dei caratteri: il codice di attività economica, il numero dei dipendenti, la forma giuridica, lo stato di attività, la ragione sociale. Un esempio di questo tipo di tabelle è quella dei dati relativa ai caratteri di localizzazione: DBASIA_CARLOCAL. Un esempio di dati relativi alla localizzazione dell'impresa (cod_carattere 09001) è riportato nella Figura 7.4.

Figura 7.4 - Estrazione dei dati della tabella DBASIA_CARLOCAL

CAR	TIP	ID	PR	MOD	PRO	RIF	DA	A	PRO	COM	CAP	INDIRIZZO
09001	002	1	00000	009	001	09-OTT-06	01-GEN-96	01-GEN-02	092	009	09100	VIA S. GIOVANNI 28
09001	002	1	00000	009	001	24-LUG-08	01-GEN-02		092	009	09127	VIA LOGUDORO

Informazioni sui dati (punta a PR, MOD, PRO)
 data di aggiornamento (punta a RIF, DA, A)
 cod provincia e cod comune (punta a PRO, COM)
 cap (punta a CAP)
 indirizzo (punta a INDIRIZZO)

Fonte: Istat, Database del Registro

Nell'esempio in Figura 7.4 sono stati selezionati i valori assunti nel tempo dal carattere di localizzazione dell'impresa con identificativo ID uguale a "1". Nella prima riga sono riportati tutti i valori validi dal 1 gennaio 1996 al 1 gennaio 2002 e sono: l'indirizzo "VIA S. GIOVANNI 28", il codice Istat della provincia "092", il codice Istat del comune "009", il CAP "09100".

I dati del secondo *record*, quindi l'indirizzo "VIA LOGUDORO", il codice Istat della provincia "092", il codice Istat del comune "009" e il CAP "09127" rappresentano i valori validi dal 1 gennaio 2002 alla data corrente.

La struttura fisica del database del Registro delle imprese consente la sua indipendenza dai cambiamenti delle informazioni, per cui alla definizione di un nuovo carattere o di una relazione diversa corrisponde una sola attività di aggiornamento dei valori contenuti nelle tabelle del database senza modificarne la struttura. Per consentire agli utenti dell'Istat una navigazione interattiva su tale database è stata creata un'applicazione web ad hoc denominata *Asi@web*, che è su un *server* web ed è richiamabile dal *browser* di qualsiasi PC della rete dell'Istat.

Le principali funzioni implementate in *Asi@web* sono:

- visualizzazione e modifica dei dati relativi alle imprese e alle unità giuridiche;
- creazione di nuove imprese o di unità giuridiche;
- collegamento di unità giuridiche con le imprese esistenti;
- gestione degli eventi demografici, quindi visualizzazione, inserimento, modifica, eliminazione;
- visualizzazione della presenza di un'impresa sul Portale delle imprese;
- visualizzazione e gestione delle notizie, come inserimento, modifica, eliminazione.

Un altro componente importante del sistema informativo del Registro delle imprese è il Sistema di gestione ASIA (Sga), un'applicazione web che rappresenta un insieme eterogeneo di strumenti di *information technology* che fornisce supporto:

- al processo anticipato e definitivo di produzione del Registro delle imprese attive al tempo t , anche conosciuto con il nome di registro *frozen frame* delle imprese attive (cfr. Figura 7.2);

- al processo di produzione del Registro dei gruppi di imprese al tempo t ;
 - al processo di produzione del Registro delle unità locali al tempo t ;
 - al processo di produzione della lista S13 annuale;
 - al processo di produzione della lista per il Censimento permanente delle istituzioni pubbliche.
- I moduli software in cui Sga è suddiviso sono:
- modulo **Amministrazione**, che è composto a sua volta da tutti i moduli che gestiscono gli utenti del software;
 - modulo **Controllo dei processi dei registri economici**, attraverso il quale tutti i processi di produzione dei registri economici, quali il registro *frozen frame* delle imprese attive, la lista S13 annuale, la lista per il Censimento permanente delle istituzioni pubbliche possono essere eseguiti, controllati e programmati in autonomia dall'utente;
 - modulo **Imprese**, che è composto dall'insieme di specifici strumenti a supporto del processo di produzione della lista definitiva delle unità attive al tempo t , quali: la gestione delle regole distinte per ciascun carattere dell'impresa, la visualizzazione dei *report* di applicazione delle regole, eccetera;
 - modulo **Unità locali**, che è composto dall'insieme di specifici strumenti a supporto del processo di produzione del registro delle unità locali al tempo t , quali: la gestione delle regole distinte per ciascun carattere dell'unità locale, la visualizzazione dei *report* di applicazione delle regole;
 - modulo **Gruppi di imprese**, che fornisce tutte le funzionalità per gestire i processi di generazione dei gruppi di imprese estere e nazionali;
 - modulo **ASIWEB_ONLINE**, che consente di collegarsi da Sga all'applicazione *ASIA@Web*.

7.1.2 Utilizzo del Registro delle imprese nei Censimenti economici

Il Registro delle imprese ha un ruolo centrale nella produzione delle statistiche economiche. Tradizionalmente la sua funzione è quella di fornire la popolazione delle unità statistiche da cui potere estrarre i *frame* e i campioni per le indagini, così come gli strumenti per supportare la fase di monitoraggio o per la riduzione dell'onere statistico. In aggiunta, oggi il Registro è utilizzato direttamente per la produzione di statistiche sulla struttura della popolazione delle imprese per settore di attività economica, dimensione delle imprese e territorio.

È la base per le analisi sull'evoluzione della struttura delle imprese e sulla loro demografia, ed è anche base per la realizzazione di nuovi prodotti statistici ottenuti come risultato della sua integrazione con altri registri statistici o archivi amministrativi.

Uno dei prodotti fondamentali del Registro delle imprese, che scaturisce da un complesso processo di produzione annuale, è sicuramente il registro *frozen frame* delle imprese attive per l'anno t , meglio noto come Registro statistico delle imprese attive (ASIA-Imprese). Tale registro rappresenta la popolazione delle unità statistiche valida per un determinato periodo di riferimento, indicato con anno t , con i principali caratteri. È la base di riferimento per il riporto all'universo dei dati campionari ed è anche utilizzato per le stime di Contabilità nazionale.

A partire dal 2011 il Registro delle imprese rappresenta inoltre, il contenuto informativo strutturale del Censimento dell'industria e dei servizi (Cis). Infatti, i dati pubblicati relativi alla numerosità delle unità, alla loro forma giuridica, attività economica, localizzazione e occupazione, sono derivati direttamente dal Registro delle imprese e non da una rilevazione di

campo esaustiva, contribuendo in modo rilevante alla riduzione sia dei costi dell'operazione censuaria sia del fastidio statistico per le imprese.

Per potere utilizzare le informazioni del Registro delle imprese come dato censuario sono tuttavia necessarie specifiche rielaborazioni dei dati del Registro, poiché le due popolazioni presentano specifiche differenze, nel campo di osservazione e nel riferimento temporale dei dati. Analogamente a quanto avviene per il Cis, dal 2021 per il Censimento permanente delle istituzioni pubbliche è prodotta una lista precensuaria a partire dai dati del Registro delle istituzioni pubbliche, diffuso dall'Istat per la prima volta nel 2020.

Il Registro delle imprese e, più in generale il sistema dei registri economici, rappresenta una solida infrastruttura per assicurare coerenza nella produzione di tutte le statistiche economiche dell'Istat e per lo sviluppo di processi di produzione più efficienti.

7.2 Registro dei gruppi di impresa e Censimenti economici

Il Registro statistico dei gruppi di imprese è costruito in ottemperanza alle disposizioni del Regolamento (UE) 2019/2152, che ha abrogato il Regolamento (CE) n. 177/2008, e del Regolamento (CEE) n. 696/1993, secondo una metodologia armonizzata approvata da Eurostat.

Il Regolamento (CEE) n. 696/1993 definisce il gruppo di impresa come “più imprese con legami giuridico-finanziari. Il gruppo di imprese può avere diversi centri decisionali, in particolare per quel che concerne la politica della produzione, della vendita, degli utili; esso può unificare certi aspetti della gestione finanziaria e della fiscalità. Esso costituisce un'entità economica che può effettuare scelte con particolare riguardo alle unità alleate che lo compongono”. Tale definizione, sufficientemente generale per potere cogliere le molteplici configurazioni dei gruppi di impresa, non fornisce criteri immediati per una sua implementazione negli archivi statistici, dove devono essere registrati i legami di controllo tra le imprese e alcune caratteristiche salienti del gruppo di appartenenza.

A questa esigenza operativa ha risposto la metodologia armonizzata messa a punto da Eurostat in collaborazione con gli istituti statistici dei paesi membri nell'ambito del *Working group on enterprise groups* e pubblicata nel Capitolo 21 del *Business register. Recommendations manual* (Eurostat 2010). La realizzazione del Registro per l'Istat è in carico ai tematici delle strutture che si occupano delle statistiche economiche.

Le informazioni utilizzate per la costruzione del Registro statistico dei gruppi di impresa provengono in prevalenza da fonti amministrative che forniscono tutti i legami di partecipazione dove una unità partecipante detiene una quota di partecipazione attraverso cui esercita la propria influenza su una unità partecipata.

L'utilizzo di una base informativa così ampia mira a garantire la copertura totale del fenomeno sia dei gruppi a controllo privato sia dei gruppi a controllo pubblico. La metodologia di integrazione delle diverse fonti amministrative, partendo dai dati elementari sulla struttura delle partecipazioni dirette di tutte le società, individua i legami di controllo, esercitati sia direttamente sia indirettamente, cui è soggetta ogni società.

Per ciascuna società controllata viene individuata la sua controllante prossima, definita come il primo soggetto fisico o giuridico che in linea gerarchica esercita per prima su di essa un controllo diretto o indiretto. La struttura del gruppo è infine ricostruita attraverso la sequenza continua dei legami tra le controllanti prossime, fino alla attribuzione del vertice ultimo all'intero gruppo.

7.2.1 Architettura del database del Registro statistico dei gruppi di impresa e il processo informatico di costruzione

La prima versione del Registro statistico dei gruppi di impresa in Istat risale al 2005 ed è stato presentato per la prima volta a Wiesbaden nel 2007 in occasione dell'*International Roundtable on Business Survey Frames*. Già dalla prima implementazione è stato deciso di sviluppare tutta la procedura in ambiente *Oracle*, con l'utilizzo del linguaggio PL/SQL per i programmi e *Forms 6i* per l'interfaccia grafica, che ne gestiva solo una minima parte. Nasce quindi un processo ciclico annuale al cui fondamento, partendo dalle fonti di input, viene costruito un Registro dei legami di base caratterizzato da una struttura molto semplice, dove era presente il legame di partecipazione o di controllo tra due imprese, persone fisiche o giuridiche, italiane o estere, individuate univocamente dal codice ASIA e dalla relativa tipologia. Le successive elaborazioni sull'archivio stesso e sugli archivi di supporto, a completamento del processo, portavano alla costruzione finale di un unico Registro statistico dei gruppi di impresa e alla relativa pubblicazione.

Inizialmente sono state utilizzate solo tre fonti di input: bilanci (consolidati e di esercizio), Consob e soci. Successivamente sono state introdotte altre fonti. L'Archivio bilanci è fornito dalla *Cerved*, società informatica delle Camere di Commercio, e comprende i bilanci riclassificati di tutte le società di capitale. I bilanci, consolidati e di esercizio, contengono informazioni sui legami diretti e indiretti tra società italiane, e tra società italiane e società estere.

La Consob è il soggetto pubblico preposto alle attività di vigilanza sulle società quotate e dei gruppi quotati in Borsa; gestisce l'Archivio sulla trasparenza proprietaria a cui hanno l'obbligo di dichiarazione tutti i soggetti, fisici e giuridici, che detengono partecipazioni superiori al 2 per cento in società quotate, nonché tutte le società quotate che possiedono partecipazioni superiori al 10 per cento in società non quotate. La Banca dati soci, gestita dalla società *Infocamere* per conto della Camere di commercio, registra l'elenco dei soci, persone fisiche e giuridiche, che hanno partecipazioni in società di capitale e iscritte nel Registro delle imprese, residenti non quotate in borsa e aventi forma giuridica di società per azioni (s.p.a.), società in accomandita per azioni (s.a.p.a.), società a responsabilità limitata (s.r.l.), cooperative, consorzi.

La disponibilità di tutte le fonti non è garantita per tutti gli anni e, in mancanza, si fa ricorso allo storico. Schematicamente il primo impianto del processo si suddivide in due macrofasi: la prima costruisce il Registro dei legami, la seconda il Registro dei gruppi di impresa. La prima macrofase comprende:

1. acquisizione delle fonti di input;
2. riconoscimento delle unità giuridiche e assegnazione del Codice ASIA;
3. applicazione dei link di sostituzione;
4. caricamento dei legami di base;
5. generazione dei legami di controllo indiretto;
6. selezione dei legami di controllo;
7. accorpamento delle fonti.

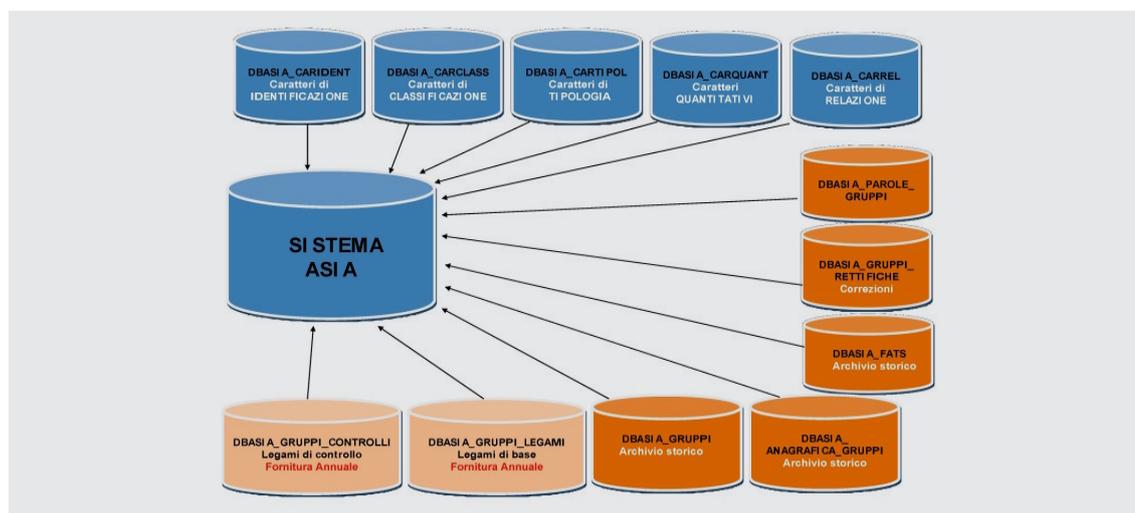
La seconda macrofase, per successive elaborazioni cicliche, individua i gruppi e i livelli gerarchici all'interno della struttura piramidale di controllo, integrando gli interventi manuali dei revisori realizzati tramite l'inserimento di nuovi link e rettifiche, fino alla completa risoluzione delle eventuali incongruenze o divergenze dalla situazione a un determinato tempo.

La seconda macrofase comprende:

1. applicazione dei link di sostituzione;
2. costruzione dei gruppi provvisori;
3. risoluzione delle incongruenze in base alle priorità;
4. assegnazione dei livelli per gruppo;
5. esclusione dei gruppi impropri.

Il processo è completamente gestito da personale informatico; ai revisori statistici è consentito esclusivamente l'introduzione di nuovi link o rettifiche dopo la verifica dei risultati di costruzione e la stampa delle incongruenze. La Figura 7.5 illustra lo schema del processo: in colore blu le tabelle esistenti e utilizzate dal Registro delle imprese ASIA; di differente colore le tabelle create ad hoc per il processo del Registro dei gruppi di impresa.

Figura 7.5 - Tabelle principali utilizzate dal processo



Fonte: Istat

7.2.2 Nuova implementazione del processo informatico di costruzione del Registro statistico dei gruppi di impresa

Lo sviluppo in Eurostat di EGR 2.0, la nuova versione dell'*EuroGroups Register*, ha conseguentemente richiesto altri requisiti alle statistiche nazionali per accompagnare i cambiamenti. Ciò ha reso necessario una ristrutturazione e una riprogettazione del Registro dei gruppi di impresa che accogliesse tutte le nuove esigenze. Il nuovo progetto è iniziato nel 2016 con l'obiettivo finale di realizzare un Registro dei gruppi di impresa con le seguenti caratteristiche:

Flessibile

Le fonti disponibili, disomogenee tra loro, variano di anno in anno, così come spesso anche le loro caratteristiche. È stato quindi necessario che le procedure non dipendessero né dal numero delle fonti disponibili, né dalla loro disponibilità temporale e, solo in misura minima, dalla loro struttura. Il rilascio annuale del Registro dei gruppi di impresa ha tempistiche e caratteristiche secondo le diverse necessità, come ad esempio per Eurostat i gruppi nazionali o i gruppi pubblici.

Pertanto, sono necessari vari scenari di sviluppo del Registro, anche contemporanei.

Per disporre di un Registro dei gruppi di impresa il più aggiornato e corretto possibile, è necessario avere la possibilità di intervenire con delle correzioni che abbiano un margine temporale di applicazione e avere, appena disponibile, il segnale di cessazione dell'unità giuridica. Poiché era già in corso l'implementazione del nuovo concetto di impresa, il Registro ha dovuto già implementare una struttura semplice in grado di adattarsi a nuovi concetti e nuove unità.

Veloce

Il Registro dei gruppi di impresa non è un registro di natura semplice, dove ogni elemento è indipendente dall'altro, come può essere quello delle unità giuridiche, ma è un registro complesso e dinamico dove ogni singolo cambiamento incide su una parte della struttura: ogni mutamento prevede una completa rielaborazione. Per permettere, quindi, un utilizzo in tempo reale al singolo utente, è necessario che i tempi di processo di rilascio del risultato siano sufficientemente rapidi. Pertanto l'elemento di base deve essere il più semplice possibile, facilitando anche un eventuale intervento sul dato.

Integrato

Per avere una minore ridondanza di dati e un utilizzo del software esistente, è stato necessario cercare la massima integrazione con il Registro delle imprese attive, utilizzando tutte le informazioni legate all'unità giuridica e all'impresa registrate, e aggiornate costantemente, rinunciando quindi a una propria anagrafica. Viene dato un pieno supporto a EGR 2.0, con la possibilità di integrare i dati provenienti da EGR e nello stesso tempo di generare i file di dati necessari alle varie fasi di scambio con Eurostat.

Sicuro

Essendo il nuovo Registro dei gruppi di impresa dinamico e aggiornato in tempo reale, è necessario che le informazioni di partenza rimangano inalterate e siano solo un rimescolamento delle informazioni disponibili al momento della generazione del gruppo, che può avvenire più volte e con diverse modalità. Pertanto le tabelle contenenti le informazioni sui legami provenienti dalle forniture non devono mai essere modificate: ciò garantisce sempre l'integrità del dato di partenza e la possibilità al revisore di effettuare continue modifiche e correzioni di eventuali errori senza temere il mancato recupero del dato iniziale. Un nuovo algoritmo di costruzione dei gruppi è messo a disposizione del revisore in modo tale che possa lavorare un singolo gruppo. Tale algoritmo prende in considerazione aspetti precedentemente non elaborati nel processo globale, realizzando quindi due modalità con molte differenze logiche, ma con un unico risultato possibile. Il confronto dei due risultati ha garantito l'assenza di errori nel processo di elaborazione del dato finale.

Storico

Viene messo a disposizione anche un Registro storico dei gruppi, collegato nei vari anni dalla presenza dello stesso vertice e quindi assegnando il medesimo codice in fase di consolidamento del gruppo. Con questo Registro è possibile utilizzare le informazioni storiche per coprire l'eventuale mancata presenza per l'anno corrente di una o più fonti.

Grafico

L'interfaccia grafica gestisce vari processi:

- modifica, inserimento e cancellazione di alcune tipologie di dati, per permettere l'intervento manuale dei revisori sulla struttura del singolo gruppo;
- gestione completa dei processi di generazione del singolo gruppo per i revisori e dei processi di generazione dei gruppi globali per i *super user*. Ciò permette di rendere pienamente indipendente il singolo revisore dall'assistenza del personale informatico;
- gestione dei gruppi bloccati o profilati, ossia i gruppi di maggiore peso lavorati puntualmente dai revisori statistici;
- gestione dei processi di stampa ed estrazione per l'analisi e la diffusione dei dati.

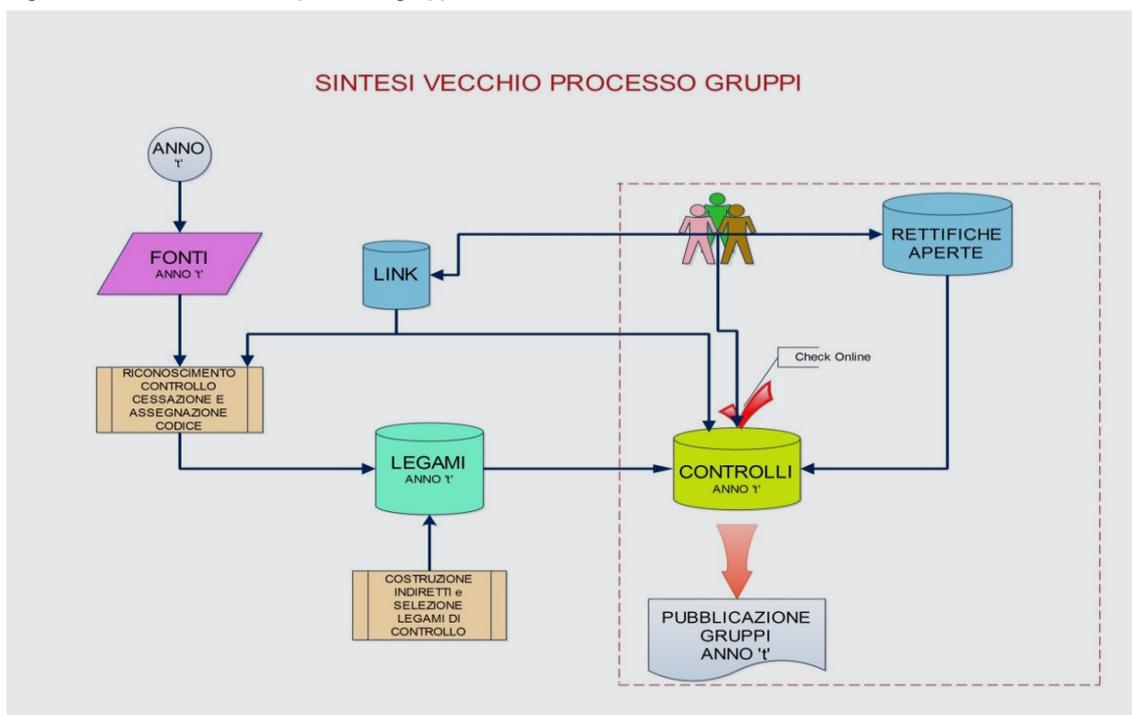
Registro dei gruppi di impresa v.2.0: l'implementazione

Nella nuova implementazione, come ambiente di base è stato confermato *Oracle*, ma per quanto riguarda l'interfaccia grafica, compatibilmente con le direttive informatiche e le nuove tecnologie impiegate dall'Istat, si è deciso di utilizzare Java, invece di *Forms*, e integrarla come nuova voce di menu in Sga, dove già era presente l'applicativo di controllo di processo per la gestione del Registro delle imprese attive (ASIA).

Rispetto alla prima versione, è la seconda macrofase che subisce dei radicali cambiamenti, mentre la prima macrofase rimane inalterata nella logica e in gran parte dei processi. Di fatto si sono costruiti solo dei moduli di passaggio per l'integrazione con la nuova struttura che è andata a sostituire completamente la seconda macrofase.

Sinteticamente, dallo schema in Figura 7.6 si è passati allo schema in Figura 7.7.

Figura 7.6 - Sintesi del nuovo processo gruppi



Fonte: Istat

7. La gestione dei dati e l'utilizzo dei registri

Dalla fonte vengono estratte tutte le informazioni relative ai legami, sia quelle di partecipazione sia di controllo, e caricate insieme a poche altre informazioni: la forma giuridica della società partecipante, la quota di controllo e il *flag* qualitativo, che indica se la partecipazione è reale o figurativa. Contestualmente, interrogando la tabella delle cessazioni, costruita appositamente e alimentata giornalmente, viene caricata l'eventuale data di cessazione della partecipata e della partecipante. Quindi al termine di questo *step* si ottiene un Registro dei legami costituito da due tipologie principali:

1. legame di partecipazione con quota controllo minore o uguale al 50 per cento;
2. legame di controllo con quota controllo maggiore del 50 per cento.

Se la partecipazione è di tipo reale sarà presente "Q" nel *flag* qualitativo sempre insieme a un valore di quota, altrimenti vi sarà una lettera compresa tra "A" ed "E" o un numero compreso tra 1 e 5 se il controllo è figurativo. Raramente, a seconda delle fonti, può essere presente una combinazione dei due casi. Dopo il caricamento di questi legami, avvenuto fonte per fonte, a seconda della disponibilità, viene applicato un algoritmo per il calcolo degli indiretti, sempre fonte per fonte, che è un legame di controllo calcolato, costruito tramite un programma e inserito con un *flag* qualitativo contrassegnato dalla lettera "I". Terminato questo *step*, si può ritenere conclusa la costruzione del Registro dei legami.

Dal passo successivo inizia il nuovo processo di costruzione dei gruppi.

Dal Registro dei legami si selezionano tutti i legami di controllo e vengono caricati sulla tabella dei controlli. Uno specifico programma individua, all'interno della stessa fonte, tutti i casi opposti e cerca di risolverli mediante una serie di controlli. Gli opposti non sono altro che due controlli in cui controllato e controllante sono invertiti: "A" controlla "B" e "B" controlla "A". Risolti in automatico i casi già elaborati negli anni passati, i restanti sono corretti con un'analisi puntuale.

Eseguita questa fase, la fonte viene caricata nella tabella delle forniture, dove non subisce più alcuna modifica fino a una nuova fornitura della stessa fonte. Campi fondamentali sono: l'anno, in quanto forniture diverse potrebbero avere anno diverso; la data di validità della relazione che determina l'utilizzo o meno del dato; la forma giuridica della società partecipante per potere esperire una serie di controlli, ad esempio una persona fisica non può essere partecipata; il campo fonti è un campo posizionale e segnala tutte le forniture che hanno indicato la relazione presente nel *record*.

Il processo di costruzione del singolo gruppo, così come quello globale, può utilizzare quindi, a seconda delle opzioni selezionate, le seguenti tabelle:

- forniture: sono presenti i legami di controlli provenienti dalle varie fonti;
- cessazioni: sono presenti le date di cessazione delle unità giuridiche;
- link di sostituzione: sono presenti i codici delle unità giuridiche estere che sostituiscono altre;
- rettifiche: sono presenti cancellazioni e inserimenti provenienti dai revisori;
- profilati/bloccati: sono presenti i legami di controllo di un gruppo considerati validi e bloccati dal revisore nell'anno.

Tramite l'utilizzo di queste tabelle il processo lanciato, unendo le varie informazioni provenienti dalle varie fonti e risolvendo le incongruenze sulla base di priorità, si costruisce il gruppo o l'intero registro, riportando il risultato in una tabella partizionata per revisore, con un tracciato molto simile alla tabella delle forniture, ma con dei campi necessari all'identificazione

della struttura dei gruppi, per poi potere essere utilizzata dai successivi programmi di stampa o di scarico dal medesimo utente. Le differenze rispetto alla tabella delle forniture sono: l'utente, con il quale la tabella è partizionata; il codice gruppo provvisorio, che viene assegnato durante il processo di elaborazione, così come, in assenza di incongruenze, viene assegnato il livello all'interno del gruppo; la provenienza del dato che indica da quale tabella viene prelevato il dato.

L'assegnazione del codice ASIA

L'assegnazione del codice ASIA rappresenta il primo importante step per ogni nuova fonte che deve essere caricata. A differenza delle unità residenti, l'assegnazione del codice ASIA per le unità estere è un compito tipico del processo di costruzione dei gruppi di impresa e avviene tramite l'applicazione di un algoritmo basato unicamente sulla nazionalità e denominazione. Un primo passo riguarda la standardizzazione della denominazione, che avviene applicando le seguenti procedure e che rende la denominazione legale pronta per il riconoscimento:

- trasformazione della denominazione legale in maiuscolo;
- normalizzazione della denominazione legale con ricerca di stringhe e loro sostituzione con stringhe standard diverse per ciascun paese;
- applicazione di altre normalizzazioni standard di stringhe valide per tutti i paesi;
- conversione di caratteri speciali in altri come impostazione predefinita;
- sostituzione dei caratteri alfabetici caratteristici di altri paesi con i caratteri corrispondenti inglesi;
- soppressione di spazi multipli tra le stringhe di un nome.

Una prima ricerca utilizza direttamente la denominazione interrogando la tabella dei caratteri identificativi del Registro delle imprese, selezionando unicamente quelle con il medesimo paese di appartenenza. In caso di risultato negativo, vengono individuate le stringhe di parole che compongono la denominazione ed eliminata la forma giuridica si effettua una ricerca sulla tabella delle stringhe costruita sulla base delle unità giuridiche precedentemente registrate.

Se la ricerca ottiene un solo codice, lo si assegna all'unità, altrimenti ne viene generato uno nuovo progressivo tramite un'apposita *sequence*. Questa nuova unità giuridica estera viene caricata nelle relative tabelle della base dati ASIA nel seguente modo:

- il codice ASIA e la denominazione sono inseriti nella tabella dei caratteri identificativi;
- il codice ASIA e il codice paese sono inseriti nella tabella dei caratteri di classificazione;
- il codice ASIA e ogni stringa di parole vengono inserite nella tabella relativa.

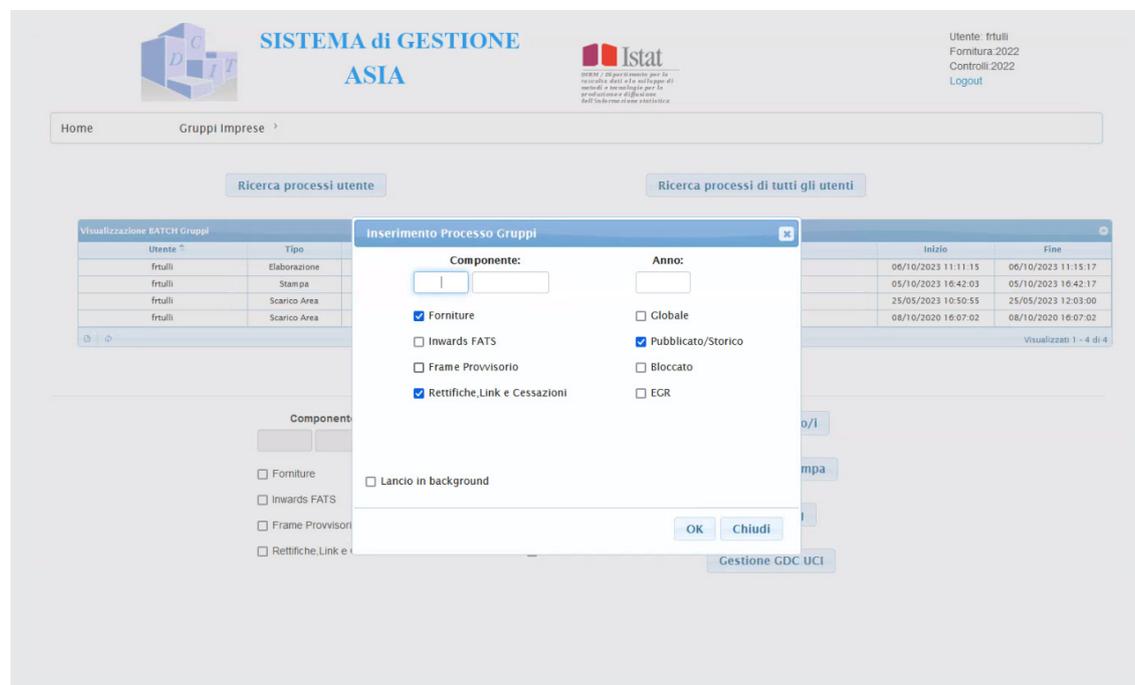
Periodicamente e dopo ogni introduzione di numerose unità giuridiche estere sulla base dati ASIA, un processo *batch* globale analizza la situazione e cerca di risolvere la presenza di unità giuridiche duplicate per denominazione simile e per stesso codice paese, che vengono registrate in una tabella apposita dove un codice unità diventa il sostituto di altri codici esteri.

La costruzione dei gruppi di impresa

La costruzione di un singolo gruppo o dell'intero Registro dei gruppi di impresa può avvenire in ogni momento dietro una semplice richiesta effettuata tramite un pannello di controllo dell'interfaccia grafica (Figura 7.8) e in base alla scelta di alcuni parametri.

7. La gestione dei dati e l'utilizzo dei registri

Figura 7.8 - Interfaccia del sottomenu per la generazione dei gruppi



Fonte: Istat, Sistema di Gestione ASIA

In base ai parametri selezionati viene lanciata una sequenza di procedure “P” guidate dalla tabella degli scenari “S” e “N” presenti nel campo scenario di Figura 7.9.

Attualmente si va da un minimo di un singolo programma avviato, partendo da “P1”, fino a un massimo di 14 presente nel campo “P14”.

Figura 7.9 - Alcuni scenari selezionabili

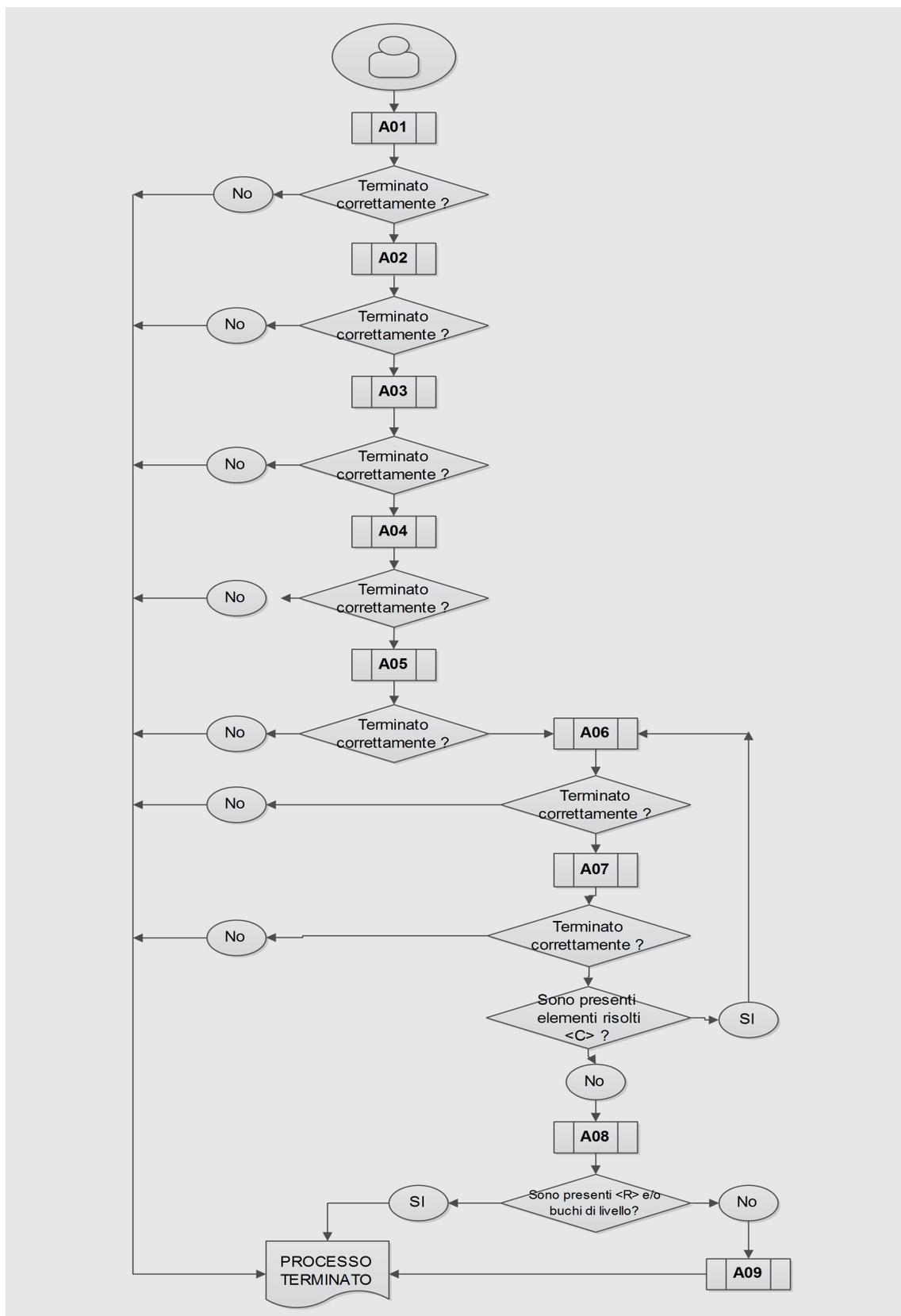
SCENARIO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14
GSNNNNSSNN	G01	A02	A03	G04	G05	G04	A03	G06	G07	G08	G09	P02	(null)	(null)
SSNNNNNSNN	A01	A02	A03	A04	P06	A05	A04	A03	A06	A07	A08	A09	(null)	(null)
SNSNNNNNNN	C01	(null)												
SSSNNNSNN	A01	C02	A11	A02	A03	A04	A05	A04	A03	A06	A07	A08	A09	(null)
GSNNNNNSNN	G01	A02	A03	G04	G05	G04	A03	G06	G07	G08	G09	G10	(null)	(null)
GSSNNNSNN	G01	C12	A02	A03	G04	G05	G04	A03	G06	G07	G08	G09	G10	(null)
SNNNNSSNN	P01	P04	P06	A12	A02	A03	A04	A05	A04	A03	A06	A07	A08	A09
SNNNNSSNN	P01	P04	A04	P06	A05	A04	A03	A06	A07	A08	A09	(null)	(null)	(null)
SNSNNNSNN	C04	A04	P06	A05	A04	A03	A06	A07	A08	A09	(null)	(null)	(null)	(null)
SNNNNNSNN	P01	P05	(null)											
GNSNNNNNNN	C03	(null)												

Fonte: Istat, Sistema di Gestione ASIA

Una volta terminato, ogni passo viene registrato in un apposito campo *clob*, visualizzabile dall'interfaccia: se è terminato correttamente, vengono avviati i passi successivi, applicando, a seconda dello scenario, un classico algoritmo (Figura 7.10).



Figura 7.10 - Algoritmo utilizzato dallo scenario per la costruzione di un gruppo singolo



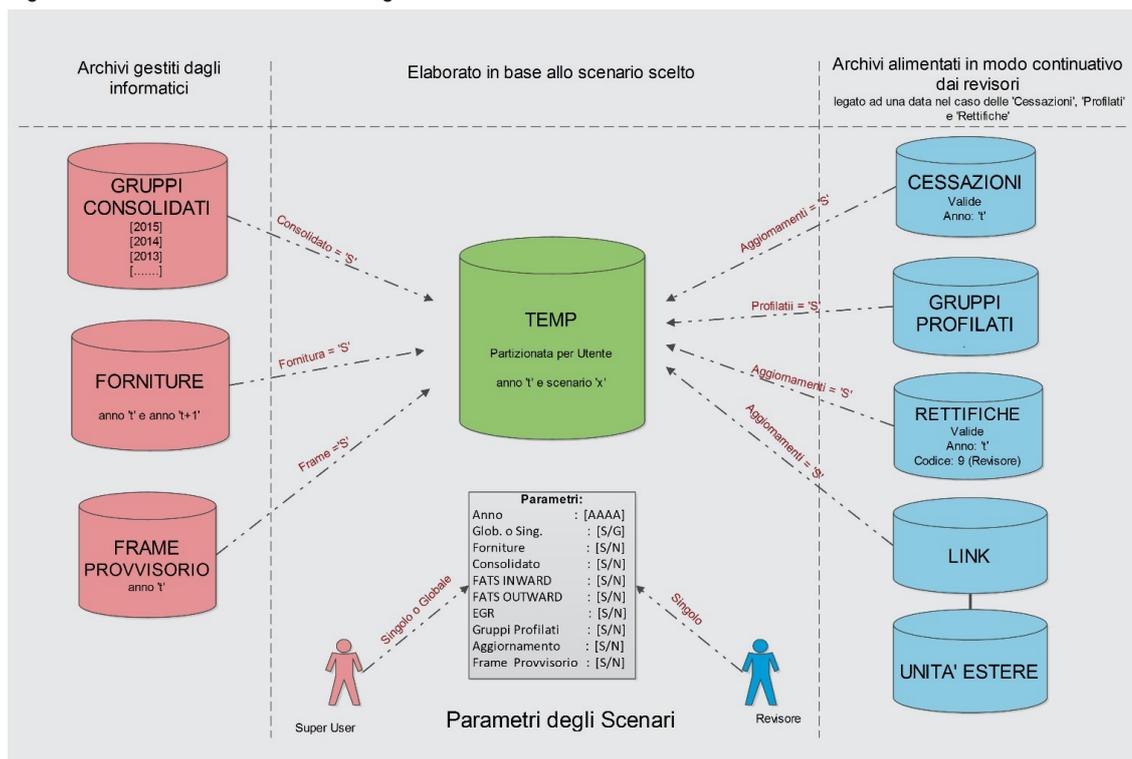
7. La gestione dei dati e l'utilizzo dei registri

Tutte le procedure avviate dall'applicativo sono decodificate tramite una tabella in cui è presente per ogni elemento la descrizione della procedura, il nome per esteso, la tipologia (procedura o *package*) e i parametri necessari al lancio. Il primo carattere alfabetico del codice, poi seguito sempre da un numero a due cifre, indica una specifica funzione:

- “S” indica le procedure o i *package* per realizzare stampe di vario tipo;
- “G” indica le procedure o i *package* per la costruzione dei gruppi globali;
- “A” indica le procedure o i *package* per la costruzione dei gruppi singoli;
- “P” indica le procedure o i *package* di partenza o di integrazione.

Sulla base dello scenario scelto, sono coinvolte una o più tabelle dalle quali sono caricati i dati ed elaborati tramite un processo ciclico. Come illustrato nella Figura 7.11, alcune tabelle sono gestite esclusivamente dal personale informatico (indicate in colore rosa), mentre altre tabelle sono gestite dai revisori statistici tramite alcuni pannelli dell'interfaccia grafica Sga o *ASIA@Web* (indicate in colore azzurro).

Figura 7.11 - Alimentazione dei dati negli archivi



Fonte: Istat

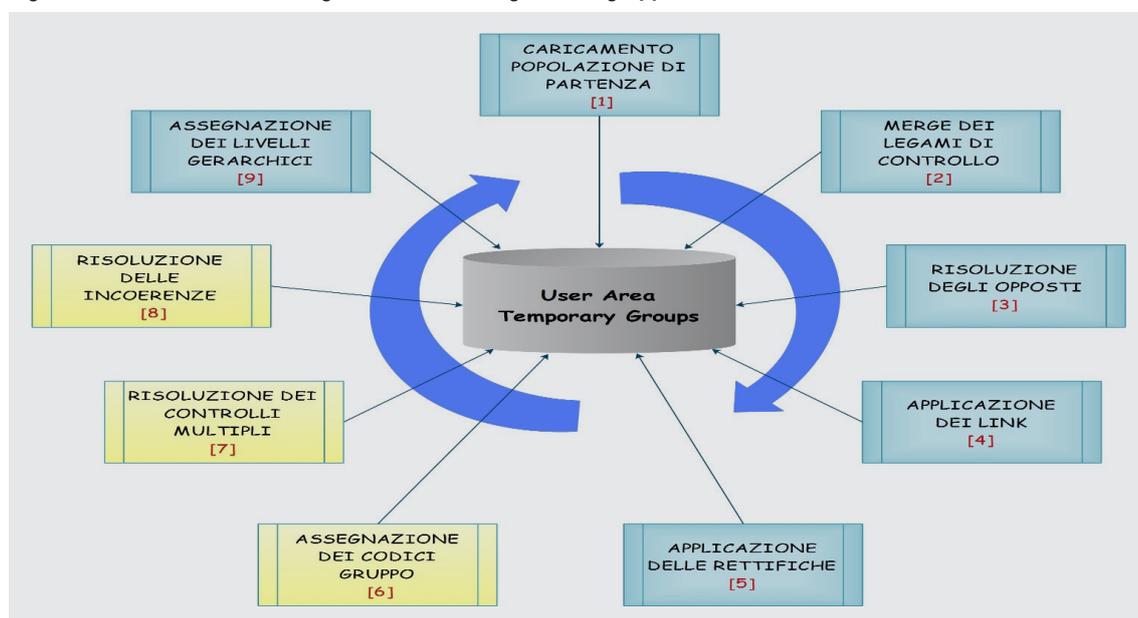
Il ciclo del processo, sia di un singolo gruppo sia globale, può essere sintetizzato con lo schema in Figura 7.12. Alcuni *step* possono girare più volte a ogni inserimento di nuovi dati o per la risoluzione di incongruenze; al termine del processo di costruzione, se esente da errori, nel caso sia stato scelto di generare il gruppo globale, ci sarà un decimo *step* che sostituirà una parte dei gruppi elaborati con quelli controllati, lavorati e certificati dai revisori, e presenti nella tabella dei gruppi profilati e bloccati. Al termine scaricherà le strutture con una serie di informazioni aggiuntive anche in tabelle che potranno poi essere analizzate da altri processi.

Tramite le varie voci presenti nel menu di Sga relativo ai gruppi sarà dunque possibile eseguire una serie di operazioni: stampe di vario tipo, scarichi in più formati, bloccaggi e sbloccaggi.

Ogni utente ha un'area di lavoro separata ed è quindi possibile lavorare e lanciare in modo simultaneo. Per tutti gli utenti le tabelle di partenza sono sempre le stesse, non sono modificabili dal processo di creazione e messe a disposizione solo tramite le relative applicazioni.

Ciò garantisce la totale integrità dei dati e la possibilità di ripartire in ogni situazione fino alla diffusione in vari periodi dell'anno del Registro dei gruppi di impresa per le diverse esigenze, come l'integrazione con EGR, la costruzione del Registro delle ENT, inteso come il nuovo concetto di impresa, svariate statistiche economiche o per i Censimenti economici.

Figura 7.12 - Sintesi del ciclo di generazione del Registro dei gruppi



Fonte: Istat

7.2.3 L'utilizzo del Registro dei gruppi di impresa nei Censimenti economici

Il processo di aggiornamento dei gruppi di impresa si svolge parallelamente a quello delle altre unità statistiche. Una volta completata l'individuazione delle unità giuridiche della base dell'anno t , il processo si conclude con la realizzazione di una lista annuale di oltre 90 mila gruppi, diffusa autonomamente e in tempi diversi rispetto alle altre unità. Il Registro dei gruppi di impresa ha un minore utilizzo nella diffusione dei dati censuari economici rispetto alle altre unità. Si è fatto uso, infatti, dei soli attributi appartenenza a un gruppo e tipo gruppo per tutte le imprese nel campo di osservazione del Censimento e sono state utilizzate sia come variabile di classificazione dei dati diffusi sulle imprese, sia nella fase di creazione della lista per la rilevazione sulle unità complesse e per la successiva analisi delle informazioni rilevate con tale indagine. Più in generale, le informazioni sui gruppi sono utili per l'analisi dei risultati e delle informazioni qualitative acquisite attraverso il questionario della rilevazione multiscopo. L'obiettivo è di profilare in maniera più completa le imprese nel quadro economico e di potere analizzare e interpretare in maniera più compiuta i comportamenti delle imprese stesse.

7.3 Industrializzazione dei registri

Lo sviluppo dei registri più recenti è iniziato con la realizzazione di prototipi in via sperimentale, via via sempre più consolidati. Per i registri e i moduli più maturi si sta passando, in modo incrementale, da una esecuzione manuale e frammentata dei processi a una gestione industrializzata, ossia controllata, monitorata, documentata, supportata da strumenti e interfacce, e quanto più possibile automatizzata. Si sta procedendo a una ingegnerizzazione, a una strutturazione organica e una ottimizzazione del codice esistente, includendo i controlli e le verifiche effettuati manualmente nelle varie fasi di caricamento e produzione degli output. L'attenzione è stata posta sulla razionalizzazione, ristrutturazione e ottimizzazione del codice e dello schema delle basi di dati. Questo comporta che il codice viene organizzato in procedure e *package* ed è reso più monitorabile e robusto inserendo controlli e *log*. Inoltre, si utilizzano sinonimi per riferirsi a tabelle o fonti il cui nome può variare in base all'anno e alla versione del registro che si sta costruendo. Si introducono partizionamenti per ottimizzare le prestazioni e si sta procedendo sia all'analisi dei processi, per individuare e descriverne le fasi, produrre *log* e gli indicatori opportuni per il relativo monitoraggio e valutazione della qualità, sia all'introduzione di interfacce per la reportistica e per il disegno, l'esecuzione automatica e il monitoraggio dei processi.

Su questo fronte, per individuare gli strumenti adeguati, sono state condotte delle sperimentazioni su un caso reale, ossia un modulo del Registro tematico del lavoro, e delle analisi comparative che hanno coinvolto anche dei tirocinanti universitari. Le principali funzionalità richieste agli strumenti sono:

- orchestrare l'esecuzione delle procedure esistenti in modo semplice per l'utente, attraverso l'utilizzo di interfacce web che consentano di visualizzare e monitorare in modo intuitivo e immediato l'andamento dei processi;
- automatizzare i processi attraverso la gestione del flusso, ossia prevedendo attività in sequenza e in parallelo, punti di decisione manuali o automatici attraverso regole di controllo predefinite, gestione degli errori o inserimento di parametri da interfaccia;
- visualizzare reportistica di output, di processo, di validazione;
- gestire la profilazione degli utenti e la relativa autorizzazione ai processi e funzionalità di competenza;
- inviare email per le notifiche in modo automatico.

A seguito delle sperimentazioni effettuate, gli strumenti che si è deciso di adottare sono *Apex* e *Flows for Apex*. *Apex* consente sviluppi rapidi attraverso la configurazione di funzionalità standard ed è impiegato in ambito registri per la realizzazione di applicativi di gestione. In particolare, consente in modo semplice e veloce la predisposizione di menu, la gestione utenti, la creazione e visualizzazione di reportistica, l'inserimento dati, la produzione di alcune tipologie di grafici, eccetera. Ad esempio, per il Registro tematico del lavoro è ampiamente utilizzato per la reportistica e per la gestione dei metadati, ossia per consentire agli esperti tematici di gestire da interfaccia i *mapping* tra le codifiche di fonte e le codifiche da impiegare nel registro, effettuando anche verifiche e confronti rispetto all'anno precedente. Il disegno, l'esecuzione e il monitoraggio dei processi vengono realizzati con *Flows for Apex*, che si integra nell'applicazione realizzata in *Apex* in modo trasparente per l'utente, come voce di menu dedicata. *Flows for Apex* si basa su *Business Process Model and Notation* (BPMN), che rappresenta uno standard *de facto* per la modellazione dei processi.

La potenza del BPMN risiede nell'essere una notazione grafica semplice e intuitiva, che consente di indicare la sequenza delle attività del processo, specificando ad esempio condizioni o rami paralleli.

Ciò rende il BPMN di semplice comprensione anche per utenti non tecnici e, al contempo, si presta bene a una traduzione dei *task* in software funzionante: di fatto si propone come strumento di interfaccia utile per la comunicazione tra utenti e tecnici. Si riportano in Figura 7.13 i simboli BPMN attualmente utilizzati per la descrizione dei processi di caricamento dei Registri.

Figura 7.13 - Simboli BPMN utilizzati per la descrizione dei processi di caricamento dei Registri

	Simbolo	Nome	Descrizione
Eventi		Inizio del processo	Indica l'inizio del processo.
		Fine del processo	Indica la fine del processo
Tasks Indica una singola attività / operazione del processo		User Task	Sono attività che richiedono l'interazione con l'utente tramite il supporto del software
		Service Task	Sono attività implementate tramite software (es. lancio di procedure PL/SQL o invio di mail automatizzate)
Flow		Linea di flusso o di percorso	Indica un collegamento fra le varie fasi / attività del processo. Il verso della freccia indica il verso di percorrenza
Gateway Punto di decisione		Exclusive Gateway	Il gateway gestisce rami mutuamente esclusivi: dirige il flusso in base alla condizione imposta
		Parallel Gateway	Il gateway gestisce rami paralleli
Artefatto Informazioni rilevanti per il modello		Annotazioni	Consentono al modellatore di descrivere parti di flusso aggiuntive del modello o della notazione

Fonte: Istat

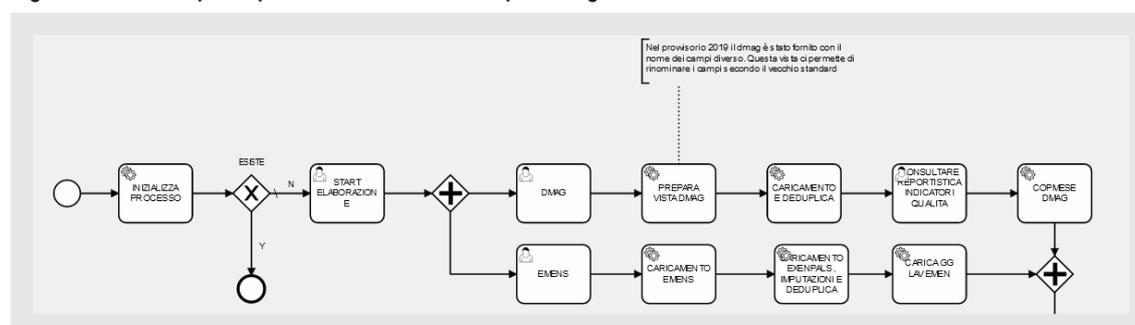
Flows for Apex consente di disegnare i diagrammi BPMN, associarvi agevolmente le procedure PL/SQL che implementano le attività, lanciare il processo e monitorarne l'avanzamento in modo intuitivo attraverso un'interfaccia semplice che modifica i colori dei *task* via via che vengono eseguiti. In particolare, il colore verde indica i *task* in esecuzione, il colore grigio i *task* già eseguiti, il colore rosso eventuali *task* andati in errore.

7. La gestione dei dati e l'utilizzo dei registri

La Figura 7.14 mostra un esempio di una porzione del flusso di caricamento del Registro tematico del lavoro modellata in BPMN attraverso *Flows for Apex*. Dal punto di vista dell'impiego degli strumenti si stanno effettuando sui vari registri analisi per individuare:

- i *task*, distinti in manuali, che sono quelli individuati dall'icona che rappresenta la figura stilizzata di un uomo, e automatici, individuati invece dall'icona che rappresenta degli ingranaggi;
- le procedure, con i relativi input e output, che implementano i *task* automatici;
- i controlli, che attualmente vengono effettuati manualmente e che possono essere automatizzati od organizzati in reportistica dedicata, a disposizione dell'utente tematico che monitora il processo, corrispondenti a *task* manuali;
- il flusso secondo cui i *task* devono essere eseguiti.

Figura 7.14 – Esempio di porzione di flusso BPMN per il Registro tematico del lavoro



Fonte: Istat

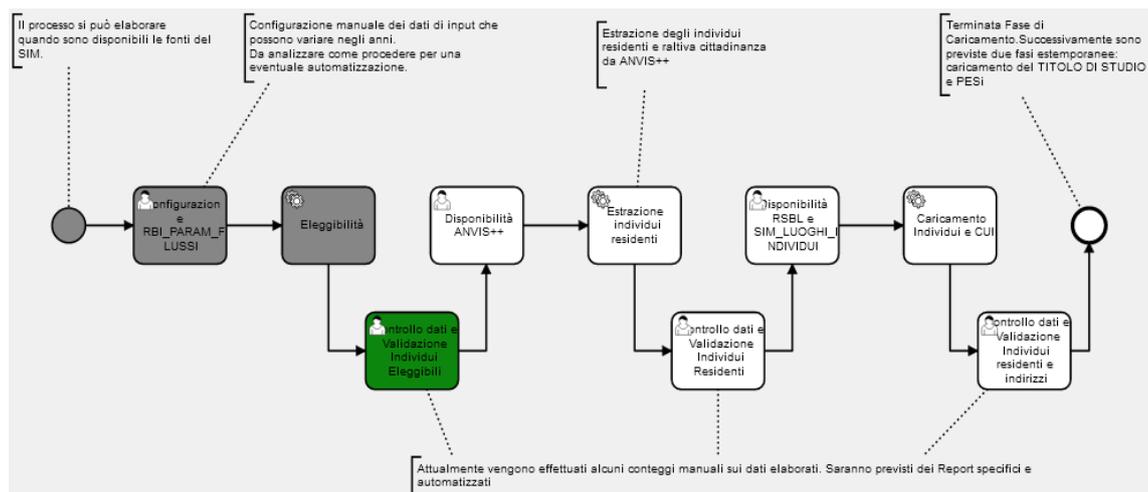
7.4 Indicatori di qualità e di monitoraggio: sperimentazioni sul Registro base degli individui

Oltre agli aspetti di automazione, negli ultimi anni ci si sta concentrando sugli aspetti di qualità e di monitoraggio dei processi di caricamento dei registri, attraverso gruppi di lavoro multidisciplinari dedicati, che includono professionalità differenti. È stato definito un *framework* teorico che consente di documentare le fasi dei processi di costruzione dei registri, basate sugli standard internazionali come il *Generic Statistical Business Process Model* (GSBPM) e il *Generic Statistical Information Model*. Il *framework* propone una serie di indicatori di qualità per il monitoraggio e la valutazione di ogni fase. Sono state condotte alcune sperimentazioni ed è in corso un gruppo di lavoro per l'implementazione del *framework* sul Registro base degli individui e sul Registro tematico del lavoro.

In particolare il Registro base degli individui è stato oggetto di sperimentazione in un precedente gruppo di lavoro multidisciplinare per la definizione della documentazione e degli indicatori di qualità relativi ai processi di caricamento dei registri. Un documento tecnico interno era già stato prodotto con lo scopo di illustrare una proposta di impostazione dei flussi di caricamento dell'area di *working* e del relazionale, disegnati in BPMN attraverso *Flows for Apex*. In Figura 7.15 è illustrato un esempio di una porzione del flusso relativo al caricamento della *working area*. I *task* con l'icona raffigurante degli ingranaggi sono automatici e corrispondono a specifiche procedure del processo di caricamento, i *task* con l'icona raffigurante un uomo stilizzato sono manuali, ad esempio il *task* "Controllo dati e validazione individui eleggibili" prevede il calcolo della numerosità degli individui per fonte di provenienza, sesso, totali, confrontati con i rispettivi valori dell'anno precedente.

Il flusso in Figura 7.15 è stato oggetto della sperimentazione in quanto si è osservato che il task “Eleggibilità” include più sottoprocessi secondo lo standard del *Generic Statistical Business Process Model*. La reportistica prevista per il task del controllo dati e validazione individui eleggibili andrà quindi arricchita in modo da sottoporre all’utente gli indicatori di tali sottoprocessi.

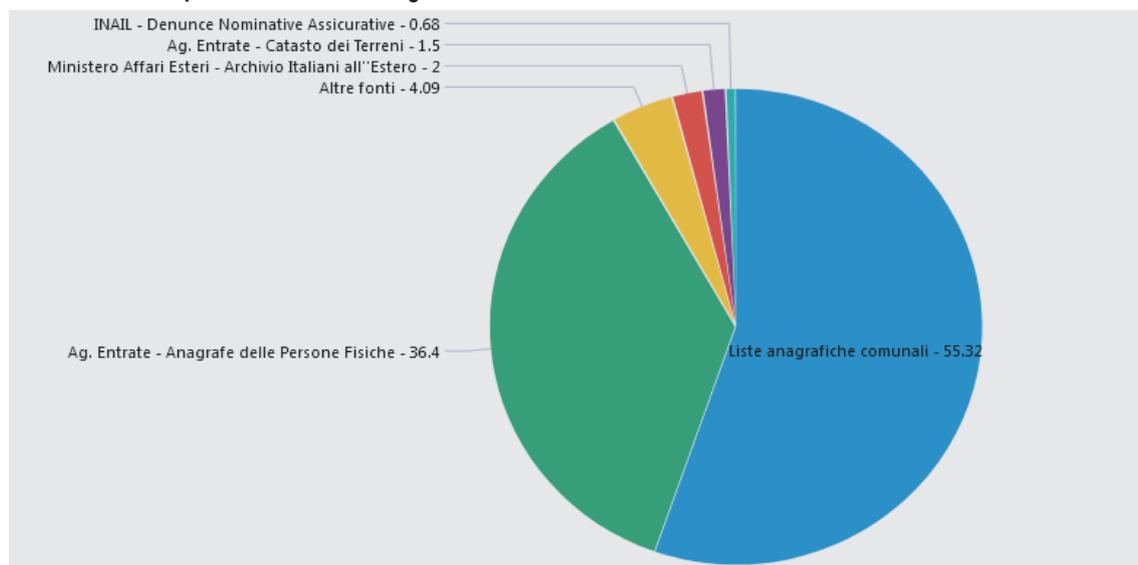
Figura 7.15 - Esempio di una parte del flusso relativo al caricamento della *working area* del Registro base degli Individui



Fonte: Istat

Nel corso della sperimentazione sono stati calcolati alcuni indicatori relativi all’integrazione al controllo e correzione della variabile sesso ed è stato proposto un esempio di interfaccia per la visualizzazione degli indicatori associati ai relativi sottoprocessi. Ogni indicatore può poi essere mostrato in forma tabellare o tramite vari tipi di grafici: in Figura 7.16 è, ad esempio, riportata la visualizzazione del grafico a torta di un indicatore relativo al task “Eleggibilità” della scheda del sottoprocesso GSBPM “Integrazione”

Figura 7.16 - Visualizzazione tramite grafico a torta di un indicatore relativo al task “Eleggibilità” della scheda del sottoprocesso GSBPM “Integrazione”



Fonte: Istat

La sperimentazione è stata significativa nell'ambito dei lavori del gruppo poiché ha consentito di testare il *framework* teorico proposto sul processo effettivo di caricamento del registro. Il nuovo gruppo di lavoro prevede di selezionare gli indicatori significativi per il processo, di calcolarli e di implementare la relativa reportistica.

Per i registri in progettazione o in fasi iniziali di sviluppo, si intende utilizzare un approccio di industrializzazione *by design* provvedendo fin dalle prime implementazioni a strutturare opportunamente il codice, suddividere il processo in fasi e sottofasi, inserire gli indicatori necessari e progettare la reportistica opportuna.

7.5 Supporto per la validazione dei dati

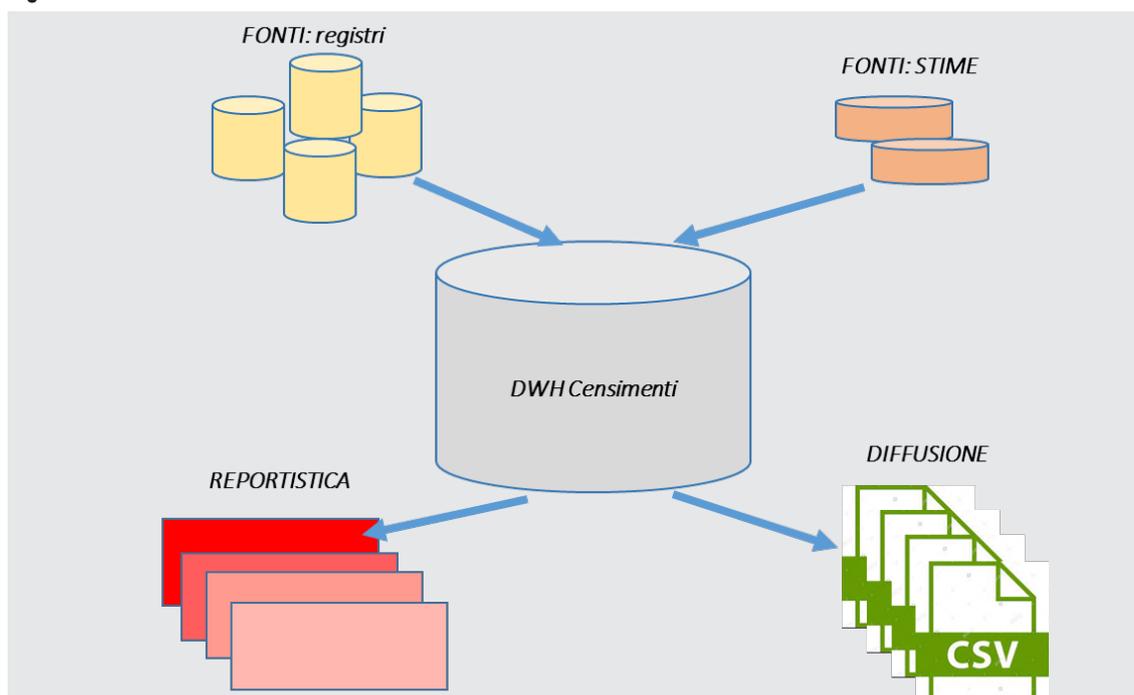
Il *data warehouse* di seguito descritto viene alimentato dalle fonti dati del Censimento tramite:

- registri, in particolare Registro base degli individui, eventualmente integrato con i dati di sottocopertura e di sovracopertura;
- le cosiddette stime: dati riferiti a una determinata variabile in una determinata popolazione di riferimento. Per ogni singolo elemento della popolazione viene fornita la distribuzione di probabilità per ognuna delle modalità che può assumere la variabile. Ad esempio, se stiamo parlando della popolazione residente e la variabile è "stato civile", per ogni data persona la stima ci dirà quale è la probabilità che sia celibe o nubile, quale è la probabilità che sia coniugato e così via.

Il *data warehouse* viene utilizzato:

- in fase di validazione dei dati, in quanto alimenta i vari *report* che vengono utilizzati;
- per la generazione dei file con cui vengono alimentati i sistemi di diffusione dati.

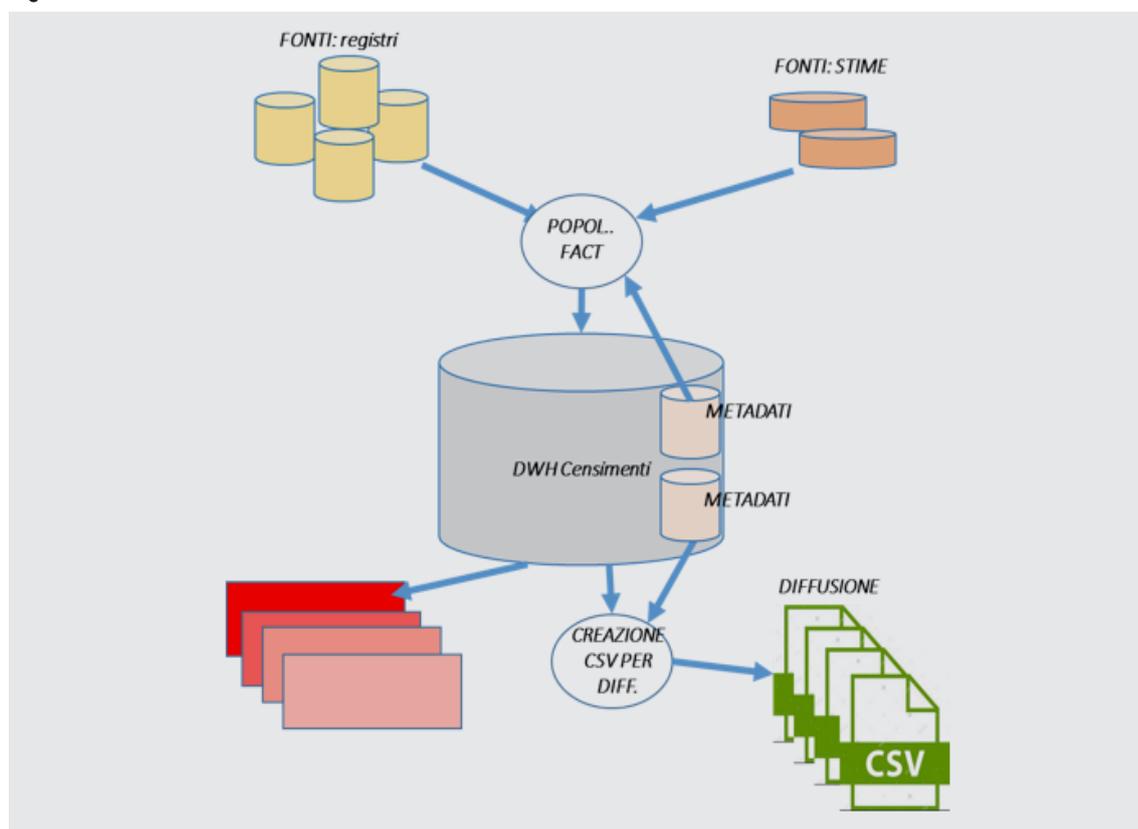
Figura 7.17 - Schema di alimentazione e utilizzo del *data warehouse*



Nella progettazione si è partiti da un tradizionale schema di *data warehouse* a stella, con relative tabelle dei fatti, dati aggregati secondo varie variabili, e delle dimensioni, descrizioni di variabili e relative modalità. Si è dovuto peraltro tenere conto di due principali considerazioni:

1. rischio di eccessiva numerosità dei dati nelle tabelle dei fatti: il processo di produzione dei dati, nella misura in cui in parte si basa su stime e non su dati dei registri, porta alla creazione di tabelle con una numerosità molto elevata. Infatti, in tale caso il numero di righe è dato solo dal numero delle dimensioni coinvolte e del numero di modalità di ciascuna di esse, non limitato dalla numerosità della popolazione base. Consideriamo, per fare un esempio, una tabella di fatti, dove abbiamo come base la popolazione italiana (circa 60 milioni). Supponiamo di avere quattro variabili: comune di residenza (8 mila modalità), età (110 modalità), cittadinanza (200 modalità) e settore di attività economica (40 modalità). Se tale tabella dei fatti viene generata a partire da dati presenti in registro, la sua numerosità non potrà mai eccedere le 60 milioni di unità. Viceversa, quando il processo di produzione parte da stime di probabilità per ogni variabile, la numerosità della tabella corrispondente risulterebbe essere: $8.000 \times 110 \times 200 \times 40 = 7.040.000.000$, anche se la popolazione di riferimento fosse di solo circa 60 milioni;
2. tabelle a geografia variabile: il *data warehouse* deve essere in grado di alimentare *report* che confrontino, a pari riferimento territoriale (comune, provincia, regione), dati riferiti a un determinato anno con quelli passati (uno, due, dieci anni precedenti). Bisognerà quindi potere assicurare, qualora vi siano state variazioni territoriali, per esempio accorpamento di comuni, tra un anno e l'altro, che i confronti avvengano sempre tra dati territorialmente omogenei.

Figura 7.18 - Focus sui metadati nello schema di alimentazione e utilizzo del *data warehouse*



Dal primo punto consegue che non è possibile avere, per ogni popolazione di riferimento (individui, famiglie, abitazioni, eccetera) una singola tabella dei fatti contenente tutte le dimensioni relative, e tante altre tabelle di *roll up* ottenute per aggregazione dalla principale.

Diventa, invece, necessario avere un numero relativamente elevato di tabelle dei fatti contenenti un numero più limitato di dimensioni.

Bisogna progettarle in modo da essere sufficiente, come tipologia di incroci tra le variabili in esse contenute, ad alimentare i *report* necessari al processo di validazione e i file da produrre per i sistemi di diffusione dei dati.

Per evitare di dovere creare e mantenere un numero elevato di tabelle con i relativi *script* di alimentazione a partire dalle fonti, come registri e dati di stima, è stato necessario predisporre un sistema di metadati che permettesse di parametrizzare tali operazioni.

Un altro sistema di metadati viene impiegato dai processi che producono i vari file in formato csv utilizzati per alimentare i sistemi di diffusione e che vengono consegnati a Eurostat.

Tenuto conto di quanto già descritto, il *data warehouse* è stato progettato come segue.

Tipologia di oggetti del *data warehouse*

- Tabella dei fatti chiamate *Fact*, che contengono dati aggregati secondo una o più dimensioni. Queste dimensioni includono sempre una dimensione territoriale. Le tabelle *Fact* si distinguono in: *Fact* contenenti dati di produzione, ossia tabelle a partire dalle quali saranno creati i file csv, e *Fact* di *benchmark* che contengono dati di confronto, utilizzati nel processo di validazione;
- Tabella *Dim* (dimensioni): contengono le dimensioni delle grandezze utilizzate nelle tabelle *Fact*, contenenti anche classificazioni a più livelli (ad esempio, età per singolo anno, per classi quinquennali, decennali, eccetera);
- Tabella contenenti metadati. Questi si distinguono in metadati utilizzati dalle procedure di popolamento delle *Fact* e quelli utilizzati nella creazione dei file per la diffusione;
- Procedure e funzioni facenti parte del database, che, guidate dai metadati, compiano le seguenti operazioni: popolamento delle *Fact* a partire dalle fonti (registri e stime), adattamento dei dati delle *Fact*, così che rappresentino la “dinamicità” del territorio, e creazione dei file che alimentano i sistemi di diffusione.

Struttura delle tabelle *Fact*

Le tabelle *Fact* si compongono dei seguenti campi:

- ANNO: anno a cui si riferiscono i dati;
- ANNO_TERR_INIZIO e ANNO_TERR_FINE: anni a cui si riferisce il riferimento territoriale ID_COMUNE;
- colonne che specificano dimensioni (ad esempio, ID_COMUNE, ID_SESSO, ID_ETA): hanno sempre una corrispondenza nelle tabelle *Dim* (ad esempio, ID_ETA viene decodificata nella tabella DIM_ETA_IT);
- colonne che specificano valori: per una *Fact* di solito vi è una sola colonna di questo tipo, ma ce ne possono anche essere due. Di solito si tratta del conteggio delle occorrenze.

Struttura dei metadati

I metadati guidano la creazione degli oggetti del *data warehouse* e dei file csv e descrivono:

- la struttura delle tabelle *Fact* e il processo della loro creazione a partire dalle fonti;
- La struttura dei file csv.

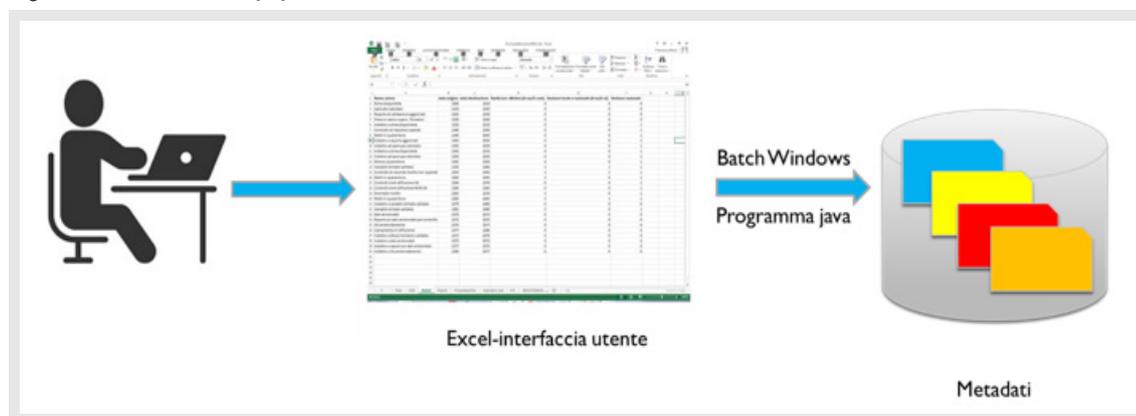
Le tabelle di *staging* per il processo di creazione dei metadati (prefisso IMPORT_) vengono utilizzate nella gestione dei metadati, pur non contenendo metadati in senso stretto.

Soluzione adottata come interfaccia per il popolamento e la gestione dei metadati

Come rudimentale interfaccia per l'utente, mediante la quale è possibile inserire i metadati, che descrivono sia le tabelle *Fact* sia i file csv, è stato ideato un sistema di fogli *Excel* a struttura fissa e definita. A partire da tali file *Excel*, vi è un processo che popola le tabelle dei metadati nel *data warehouse*. Esistono inoltre processi diversi per il popolamento dei metadati per le *Fact* e per i file csv, ma in entrambi i casi essi consistono di:

- un *batch bat Windows*;
- un programma Java, attivato dal *batch* che legge i fogli *Excel* e popola le tabelle di *staging*; in ognuna di queste ultime sono riversati i dati contenuti in un foglio dati di uno dei file *Excel*. Infine, il programma lancia una procedura PL/SQL che, a partire dalle tabelle di *staging*, popola le tabelle dei metadati strutturati.

Figura 7.19 - Schema del popolamento delle tabelle di metadati



Fonte: Istat

File *Excel* per il popolamento dei metadati per la creazione delle *Fact*

Ogni file *Excel* contiene i dati che descrivono una determinata tabella *Fact*, caratterizzata da un determinato incrocio, ad esempio FACT_POPOLAZIONE_LEGALE, basata dall'incrocio delle variabili COMUNE DI RESIDENZA, SESSO, ETA', CITTADINANZA a due modalità italiano/straniero, in un determinato anno. Ognuno di tali file *Excel* ha una struttura fissa e si suddivide in varie sezioni. Le più importanti sono:

1. informazioni di base: nome della *Fact* da popolare, anno di riferimento, fonti dati;
2. informazioni sulle dimensioni che fanno parte di una *Fact*;
3. informazioni sul processo di aggregazione dei dati di stima.

File *Excel* per il popolamento dei metadati per la creazione dei file *csv*

La struttura dei file in questione replica i cosiddetti piani di spoglio, documenti interni Istat in cui vengono definite le strutture e il contenuto dei file *csv* da produrre.

Processi di popolamento dei dati delle *Fact*

I processi di popolamento, guidati dagli appositi metadati, leggono i dati delle fonti, come registri e stime, e popolano le *Fact*. Per motivi di velocità di esecuzione, il popolamento avviene attraverso varie fasi e vengono utilizzate tabelle temporanee che vengono create e popolate attraverso comandi SQL generati dinamicamente a partire dai metadati. Si tratta per lo più di operazioni di aggregazione; l'esecuzione di tali procedure è più complesso e lungo, quante più sono le variabili che fanno utilizzo di stime e quante più sono le modalità associate a tali variabili.

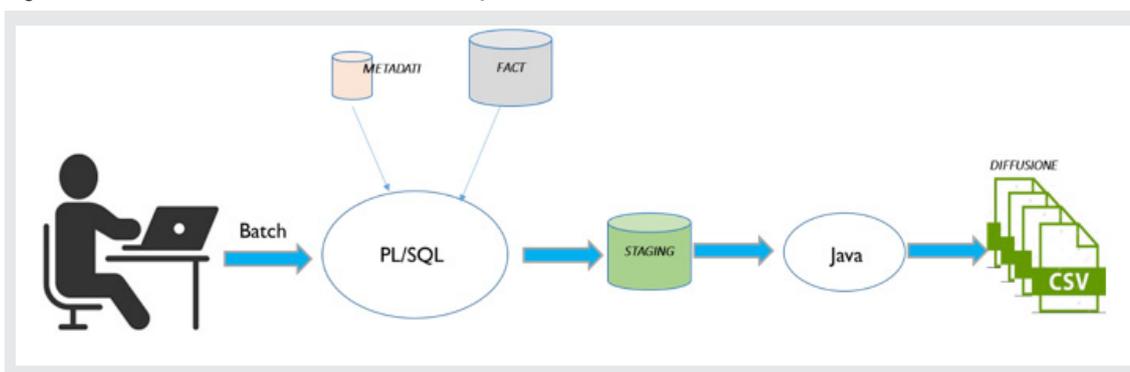
Creazione dei file *csv* per la diffusione

La creazione dei file *csv* per la diffusione avviene in modo simile rispetto ai processi di popolamento dei metadati, anche se in senso inverso. Infatti un *batch bat Windows* attiva:

- una procedura PL/SQL, attingendo ai dati delle *Fact* e ai metadati, calcola i dati da inserire nei file *csv*, attraverso vari passaggi che utilizzano tabelle temporanee di *staging*;
- un programma Java che legge dalle tabelle di *staging* e genera i vari file *csv*.

La Figura 7.20 illustra tale flusso.

Figura 7.20 - Schema della creazione dei file *csv* per la diffusione



Fonte: Istat

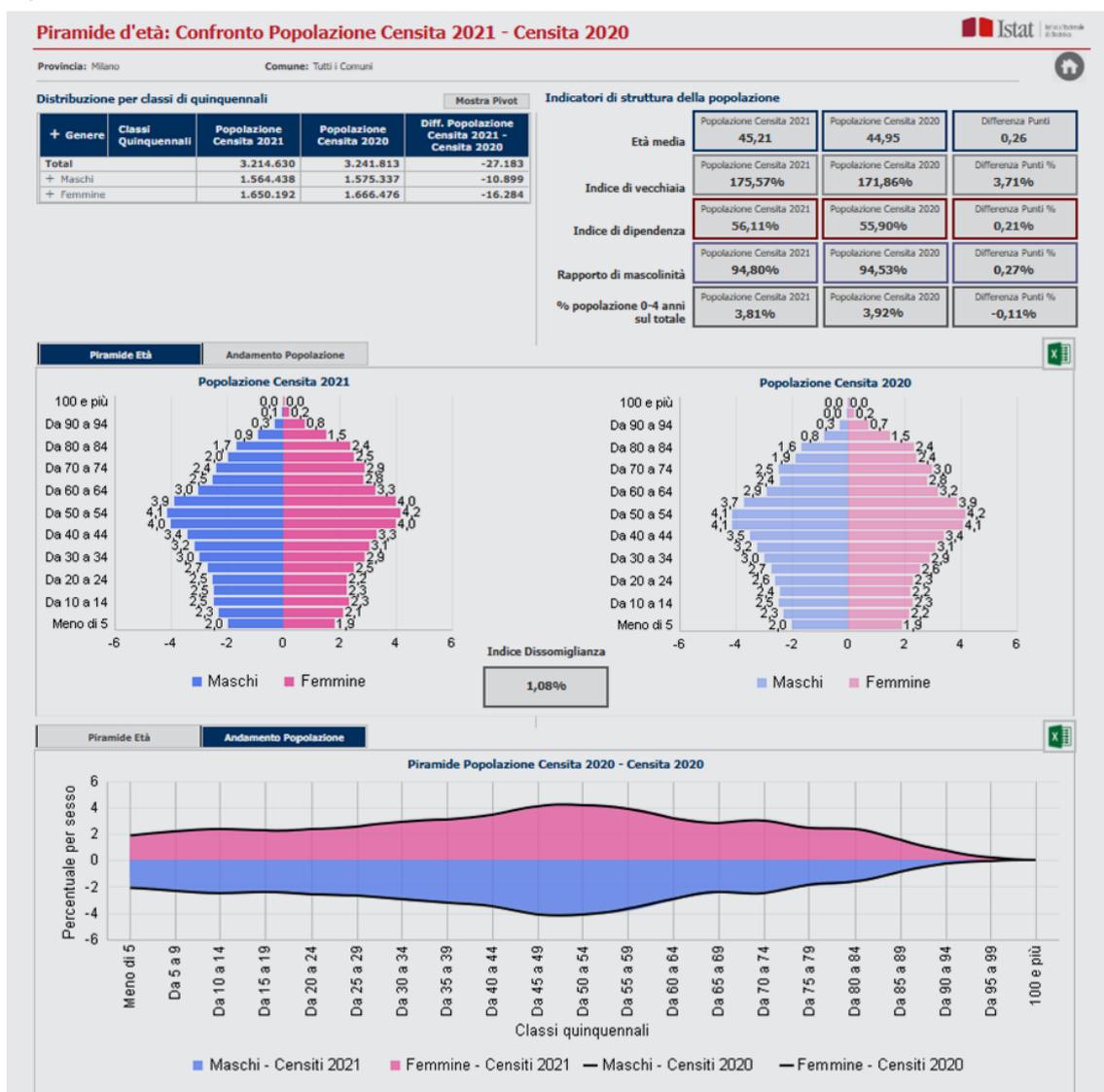
7.5.1 Reportistica a supporto

A supporto della fase di validazione dei dati prima e della successiva fase di predisposizione della diffusione, sia italiana sia europea, sono stati prodotti *report* riassuntivi sulle varie tematiche, come popolazione, famiglie, nuclei, alloggi, utilizzando la piattaforma *Microstrategy* di *Business Intelligence & Analytics*, interfacciandosi direttamente con gli oggetti del *data warehouse*. Tali *report* sono stati resi fruibili agli utilizzatori via web in maniera condivisa, nel rispetto dei vincoli di autorizzazione e visibilità di utenti specifici, completamente integrati con i sistemi di collaborazione e autenticazione in uso in Istat.

Durante la prima fase di validazione, tali *report* sono stati utilizzati dai referenti tematici e revisori per verificare eventuali anomalie nei dati, attraverso un confronto degli stessi dati con diverse fonti di *benchmark*, come ad esempio i dati del Censimento di anni precedenti o dati raccolti da altre indagini dell'Istat. I *report* sono stati predisposti come un insieme organizzato di indicatori, tavole e grafici atti a evidenziare scostamenti di valori anche in relazione a soglie prefissate. L'obiettivo è quello di effettuare una prima analisi dei valori che possano essere considerati fuori soglia per evidenziare successivamente ai revisori le situazioni che realmente hanno necessità di una ulteriore revisione più accurata e dettagliata, tramite analisi esplorative più approfondite sui dati stessi. Una tipica applicazione di questo processo è il caso del revisore che, su indicazioni del tematico di riferimento, vada a verificare i dati di una provincia o di un comune specifico che si ritiene debba essere approfondita.

La Figura 7.21 riporta un caso del genere applicato alla provincia di Milano, riportato come esempio tipico di una rappresentazione dei dati predisposta per la validazione del conteggio della popolazione.

Figura 7.21 - Esempio di *dashboard*



La *dashboard* mostra la popolazione censita del 2021 confrontandola con quella del 2020, usata come fonte di *benchmark*, rappresentandole tramite visualizzazioni differenti.

È visualizzata una classica tavola di distribuzione per età, una griglia di indicatori di struttura e due grafici: le piramidi delle età e l'andamento della popolazione.

Sono stati altresì predisposti ulteriori *report* a matrice aventi una struttura fissa, rispecchiando principalmente la stessa visualizzazione delle tavole successivamente diffuse sulla piattaforma Istat dedicata. Questa tipologia di *report* è presentata in forma di dossier, ovvero di una raccolta di *report* organizzata in pagine o capitoli. Ciascuno di questi *report*, come anche quelli di validazione descritti in precedenza, sono stati realizzati con modalità tale che agli utilizzatori sia sempre consentito potere effettuare operazioni di *drill through*, per visualizzare dati di maggiore dettaglio. Ad esempio, si è data la possibilità di potere navigare i dati all'interno dell'intero territorio italiano, partendo dal livello "Italia" di massima aggregazione, attraverso quello ripartizionale, regionale e provinciale fino a livello comunale. La navigazione può avvenire sia utilizzando la modalità "raggruppamento" sia attraverso specifici filtri di selezione.

A titolo di esempio, la Figura 7.22 mostra il dossier del conteggio della popolazione straniera residente suddivisa per genere, area e paese di cittadinanza (la cui gerarchia è navigabile in modalità "raggruppamento") e territorio italiano navigabile attraverso l'uso dei filtri posti nella parte superiore. Il dossier rispecchia lo stesso *layout* e contenuto della diffusione su *I.stat* successiva alla fase di validazione.

Figura 7.22 - Esempio di dossier orientato alla diffusione

SIMILDIFFUSIONE PAESE DI CITTADINANZA					
		Istat		Istituto Nazionale di Statistica	
Tutte le Regioni		Tutte le Province		Tutti i Comuni	
+ Continente	+ Area Geografica	Genere Paese Cittadinanza	Maschi CP 2021	Femmine CP 2021	Total CP 2021
Total			2.468.202	2.562.514	5.030.716
- Europa			990.294	1.407.543	2.397.837
	+ Unione europea		571.757	817.574	1.389.331
	+ Europa centro orientale		400.423	569.533	969.956
	+ Altri paesi europei		18.114	20.436	38.550
+ Africa			703.505	432.251	1.135.756
+ Asia			627.954	498.628	1.126.582
- America			145.338	222.723	368.061
	- America settentrionale		7.135	9.348	16.483
		Canada	794	1.193	1.987
		Stati Uniti d'America	6.341	8.155	14.496
	+ America centro meridionale		138.203	213.375	351.578
+ Oceania			783	1.076	1.859
+ Apolide			328	293	621

Fonte: Istat, Sistema di Reportistica

In aggiunta e a completamento della reportistica prodotta e predisposta sopra descritta, è stato progettato e realizzato un ambiente di analisi a uso degli utenti di produzione, i quali hanno la possibilità di costruire rapidamente e in autonomia nuovi *report*, a partire dai dati collezionati utilizzando strumenti di semplice utilizzo.

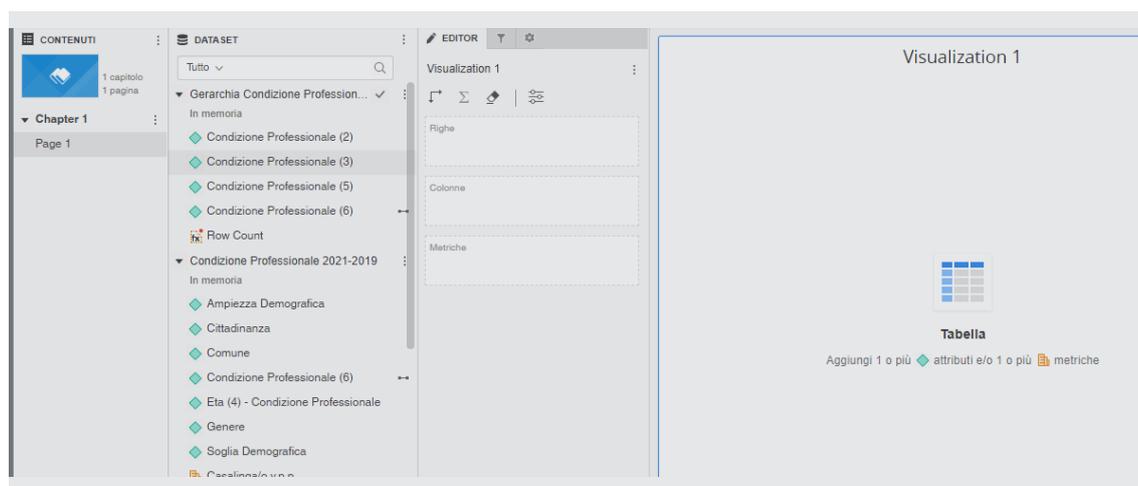
L'ambiente così realizzato risponde all'obiettivo di cercare di svincolare l'analista statistico dal personale informatico e consentire la gestione autonoma dell'analisi dei dati attraverso la predisposizione di estrazioni dati e di *report*, offrendo anche la possibilità di

fare personalizzazioni del modello dati.

L'ambiente di analisi è stato predisposto riutilizzando gli stessi oggetti *Microstrategy* e di conseguenza gli stessi dati che compongono la reportistica precedentemente descritta, riorganizzati però, in una modalità di accesso più semplificata e configurata opportunamente in base a una profilazione utenti predefinita. Il fatto di utilizzare gli stessi oggetti, senza crearne di nuovi, fornisce agli utenti la garanzia di visualizzare lo stesso identico dato indipendentemente se lo si sta analizzando su un *report* di validazione, orientati alla diffusione o nell'ambiente di analisi.

I *dataset* preaggregati, sono stati realizzati sulla piattaforma nella forma di cubi e memorizzati in un'unica area centralizzata non visibile direttamente agli utenti finali, ma sono resi disponibili come semplici collegamenti ai cubi organizzati in cartelle tematiche in un'area condivisa. L'utente, dopo l'accesso alla cartella tematica di interesse, può procedere autonomamente alla creazione di un dossier, tramite semplici passaggi sull'interfaccia a disposizione sulla piattaforma: un esempio è presentato nella Figura 7.23. Viene anche data la possibilità di personalizzazione delle strutture dati, ad esempio con la creazione di nuove classificazioni a partire da quelle date. Infine, dopo aver completato il dossier, si potrà condividere con altre persone.

Figura 7.23 - Interfaccia per la creazione di un dossier



Fonte: Istat, Sistema di Reportistica

8. LA PIATTAFORMA PER LA DIFFUSIONE WEB DEI DATI¹

Il Censimento permanente della Popolazione e delle abitazioni del 2020 ha rappresentato la prima occasione di utilizzo della nuova piattaforma generalizzata di riferimento per la diffusione web dei dati. Sono seguite altre implementazioni e, in particolare, la pubblicazione, a ottobre 2022, del nuovo *corporate data warehouse*. Ciò ha consentito di rispondere in modo tempestivo ed efficace alle diverse esigenze dell'utenza, legate a una maggiore disponibilità di strumenti informatici web per la visualizzazione, l'analisi e il trattamento dei dati rispetto al passato. I punti cardine di questo processo sono stati: l'adozione di moderni strumenti di diffusione, l'orientamento verso una sempre più elevata industrializzazione dei processi di diffusione e la standardizzazione dei contenuti diffusi tramite l'utilizzo degli standard statistici internazionali, in particolare lo *Statistical Data and Metadata Exchange (SDMX)*.

8.1 Requisiti di diffusione web dei dati

Le direttrici strategiche per la definizione di una nuova architettura di diffusione dati si sono basate sul soddisfacimento dei seguenti requisiti:

- facilitare l'accesso ai dati da parte degli utenti tramite vari possibili dispositivi (PC, smartphone, tablet, eccetera) e sulla base di percorsi di navigazione lineari e non ridondanti, tramite interfacce intuitive e lo sfruttamento dei motori di ricerca;
- sviluppare l'interoperabilità semantica nell'ambito dell'interscambio di dati e di metadati tra l'Istat e le altre organizzazioni a livello nazionale e internazionale, puntando su tecnologie *service-oriented* e sull'utilizzo di standard statistici come SDMX;
- effettuare una razionalizzazione delle soluzioni applicative (servizi, applicativi web, eccetera) tendendo a consolidare ed evolvere il più possibile soluzioni generalizzate, *open source*, riutilizzabili e basate su standard statistici anche sviluppate da altre organizzazioni.

I risultati attesi a seguito dell'implementazione di una nuova architettura di diffusione basata sui punti sopra descritti sono i seguenti:

- riduzione dei costi e degli *effort* di sviluppo nel medio-lungo periodo;
- soluzioni software durevoli nel tempo;
- interoperabilità semantica.

8.2 Soluzione tecnologica

Sono state analizzate in Istat differenti opzioni per l'implementazione della nuova architettura di diffusione:

- soluzioni commerciali;

¹ Il Capitolo è stato redatto da: Alessio Cardacino (paragrafi 8, 8.1, 8.2, 8.3, 8.3.1, 8.3.2, 8.4, 8.5), Luca Ramadori (paragrafo 8.6), Paola Giorgetti (paragrafo 8.6.1), Francesco dell'Orco (paragrafo 8.6.2).

- *tool* disponibili nell'ambito della comunità statistica a partire da quelli sviluppati nell'ambito della *Statistical Information System Collaboration Community* (SIS-CC) coordinata dall'*Organization for Economic Co-operation and Development* (OECD);
- evoluzione di un prototipo sviluppato internamente specificamente per il progetto di *hub* della statistica pubblica. In quest'ultimo caso era stato sviluppato un *set* di *tool* basato sullo standard SDMX per la diffusione centralizzata dei dati statistici tramite un portale web, che rendesse disponibile in tempo reale i dati da parte dei vari enti in modo delocalizzato.

Dopo un'attenta analisi dei costi/benefici e di aderenza ai criteri sopra indicati, la soluzione scelta è stata di evolvere il prototipo sviluppato internamente e integrarlo con *tool open source* basati su standard statistici sviluppati in altri contesti da altre organizzazioni a livello internazionale, in particolare quelli sviluppati da Eurostat.

L'orientamento verso l'utilizzo dello standard SDMX è stato dettato dai seguenti motivi:

- è uno standard per la condivisione e l'interscambio di dati statistici sviluppato e mantenuto da sette organizzazioni internazionali;
- possiede un modello informativo per la strutturazione dei dati e dei metadati;
- supporta i più diffusi formati *open data* (xml, json, csv, *RDF Data Cube*);
- guida l'armonizzazione dei dati favorendo l'interoperabilità semantica;
- sono disponibili un elevato numero di *tool* e librerie *open source* sviluppati da grandi organizzazioni internazionali (come la *SDMX Reference Infrastructure* e le librerie SDMX API di Eurostat).

I principi di progettazione della nuova architettura hanno tenuto conto dei requisiti relativi alla gestione dei processi di diffusione e di quelli legati alle esigenze dell'Istat di rendere disponibile un patrimonio informativo di notevole entità e qualità. Nello specifico:

- disporre di un sistema efficiente di visualizzazione e navigazione dei dati che utilizzi un sistema di *cache* adeguato e consenta di visualizzare tavole di grandi dimensioni;
- non incrementare il *burden* nei confronti dei settori di produzione permettendo il riutilizzo delle procedure già esistenti per il popolamento e l'aggiornamento dei dati e dei metadati;
- disporre di un'infrastruttura di gestione e di navigazione dei dati in tempi brevi;
- parallelizzare il canale di navigazione web interattiva dei dati rispetto al canale degli utenti *machine-to-machine* al fine di mantenere prestazioni adeguate;
- permettere agli utenti e alle organizzazioni di estrarre i dati *machine-to-machine* con una conoscenza basata su standard riconosciuti e di facile acquisizione;
- avere un sistema di navigazione e di ricerca delle informazioni efficace, affiancando una navigazione trasversale a quella tematica verticale. Il motore di ricerca testuale agisce in simbiosi con la navigazione tematica in modo da consentire il filtro dei risultati.

Da un punto di vista di progettazione architettonica e tecnologica, si è dovuto tenere conto di varie tipologie di utenza, da quella generalista a quella altamente specialistica; è stato adottato quindi un approccio *business-driven*, iniziato con una fase pilota e una implementazione prototipale dei *tool*, che ha permesso di raccogliere il *feedback* degli utenti. Inoltre, la scelta di soluzioni di tipo *plug-and-play* e *ready-to-use*, che non richiedono competenze tecniche particolarmente avanzate, ha facilitato l'installazione e la configurazione dei *tool*.

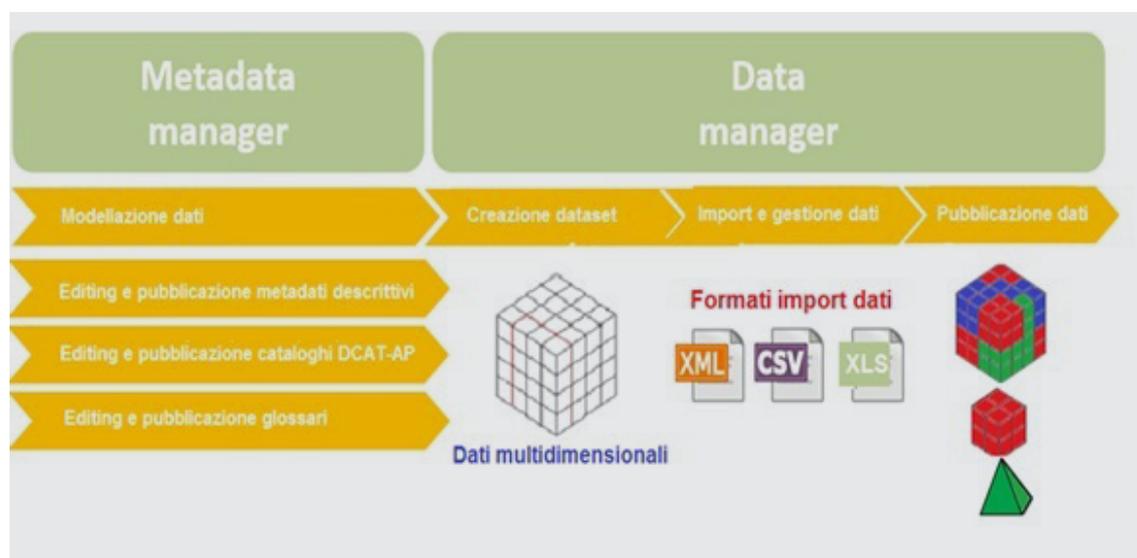
8.3 SDMX Istat Toolkit

L'*SDMX Istat Toolkit* sviluppato dall'Istat è una piattaforma *open source* basata su standard internazionali che facilita la standardizzazione e l'industrializzazione dei processi di *business* relativi alla diffusione e allo scambio dei dati. Il *Toolkit* permette la realizzazione di una architettura a servizi che garantisce l'estensibilità quando è necessario implementare nuove funzionalità. Il *Toolkit* è basato sui due moduli principali, *Meta e Data Manager* e *Data Browser*, ciascuno dei quali si compone di un *front end*, che permette l'interazione attraverso un'interfaccia grafica, e un *back end* che gestisce le interazioni dell'utente e i servizi che implementano le varie funzionalità.

8.3.1 Meta and Data Manager

Meta and Data Manager è un'applicazione web che consente di progettare e gestire *dataset* multidimensionali e i relativi metadati strutturali e di riferimento secondo il Modello Informativo SDMX. L'obiettivo è di pubblicarli in modalità *machine-to-machine* tramite l'*NSI Web Service* dell'infrastruttura di riferimento SDMX di Eurostat, e l'API dei metadati di riferimento.

Figura 8.1 - Schema delle funzionalità del *Meta and Data Manager*



Fonte: Istat

Le principali funzionalità riguardano:

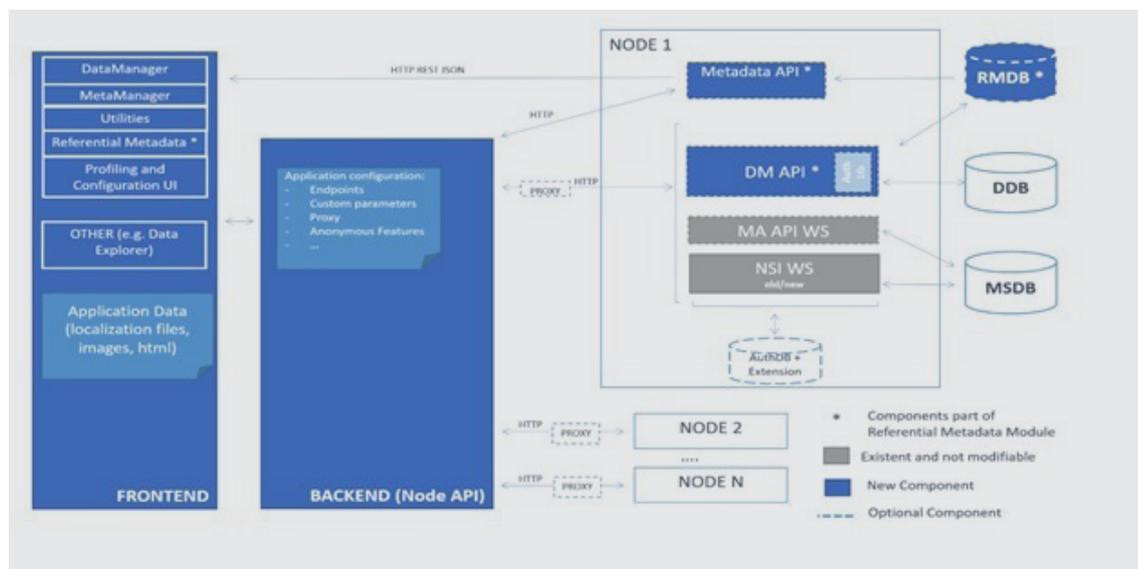
- modellazione di tavole statistiche attraverso la definizione dei relativi metadati che specificano il ruolo che ciascuna variabile assume nella struttura delle tavole, oltre che alla loro rappresentazione attraverso opportune liste di codici o di classificazioni;
- creazione e pubblicazione di metadati descrittivi (concettuali, metodologici, di qualità);
- creazione e pubblicazione di cataloghi dei *dataset* che si intende diffondere. Tali cataloghi vengono pubblicati utilizzando come *DCAT vocabulary* secondo lo standard RDF per i *Linked Open Data*. Di *default* viene fornito tutto il necessario per implementare

l'*application profile* DCAT-AP_IT realizzato dall'Agenzia per l'Italia Digitale (AgID);

- creazione, gestione e pubblicazione di glossari tematici;
- creazione e gestione di database statistici organizzati come cubi multidimensionali in cui inserire e aggiornare i dati a partire da file csv, *Excel* e xml;
- esposizione di tabelle statistiche multidimensionali attraverso un'apposita API secondo lo standard SDMX.

L'architettura è orientata ai servizi che sono attivati, sulla base delle azioni compiute dagli utenti, da un *front end* che interagisce con il servizio di *back end*, *Node API*, il quale organizza l'azione degli altri servizi di gestione e di interrogazione dei dati e dei metadati. Inoltre, alcuni servizi, inclusi nell'architettura, evidenziati in colore grigio nella Figura 8.2, sono stati sviluppati da Eurostat come parte della *SDMX Reference Infrastructure*. I servizi sono: l'*NSI Web Service*, per la gestione e l'interrogazione dei metadati strutturali SDMX e dei dati; il *Mapping Assistant Web Service*, per l'implementazione delle funzionalità di pubblicazione dei dati ottenuti tramite il *web service* precedente. Il servizio DM_API_WS permette la gestione dei dati, la creazione ipercubi, importazione, modifica, cancellazione, ai fini della loro pubblicazione. Infine il METADATA_API rende disponibili, in modalità *machine-to-machine*, i metadati di riferimento, in formato SDMX, e i cataloghi DCAT-AP dei *dataset* presenti nel sistema tramite opportune funzionalità pubbliche conformi a tale standard. Tale servizio può essere navigato da una interfaccia *client-side* interattiva che è contenuta all'interno del modulo di *front end* del *Meta and Data Manager*, ma che può essere anche utilizzata in modo autonomo e separato da tale applicazione.

Figura 8.2 - Architettura applicativa del *Meta and Data Manager*



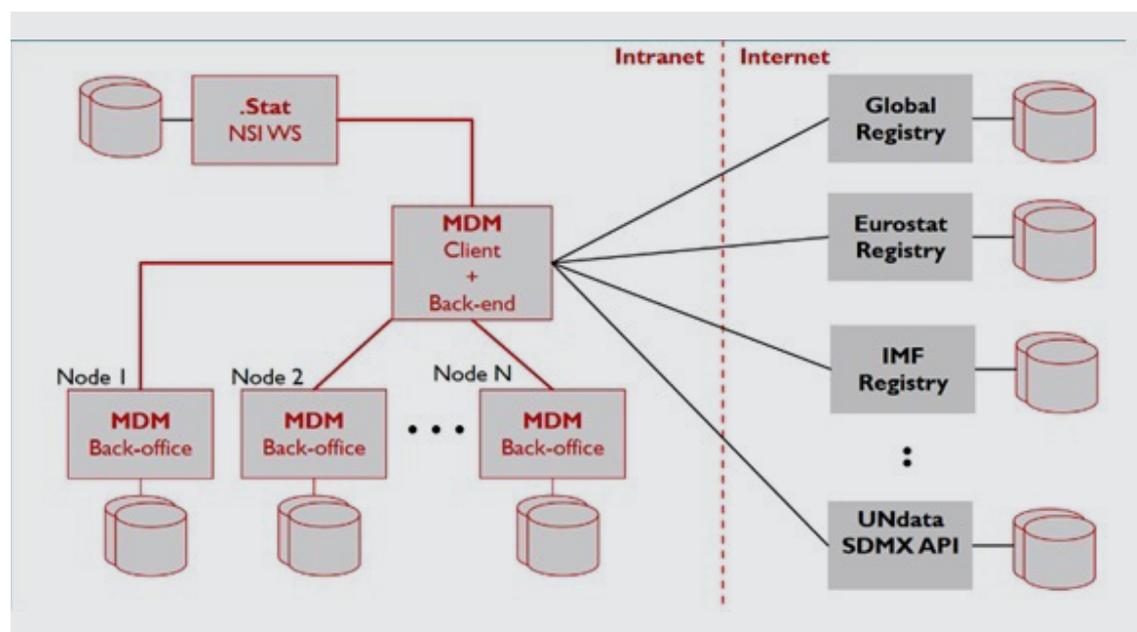
Fonte: Istat

L'applicazione è dotata anche di un meccanismo molto flessibile e granulare di gestione dei profili di accesso degli utenti, che permette di definire tali profili a partire dalle singole funzionalità e selezionando i domini o i *set* di dati a cui gli utenti con profilo operativo possono accedere. Tramite una singola istanza dell'applicazione si può accedere a più di un ambiente di lavoro per la gestione di una sorgente di dati e di metadati, previa selezione

8. La piattaforma per la diffusione web dei dati

dalla *home page* di quello su cui operare. Inoltre è possibile navigare i registry SDMX di metadati strutturali resi pubblicamente disponibili da altre organizzazioni, tramite la connessione *machine-to-machine* al relativo *end-point*.

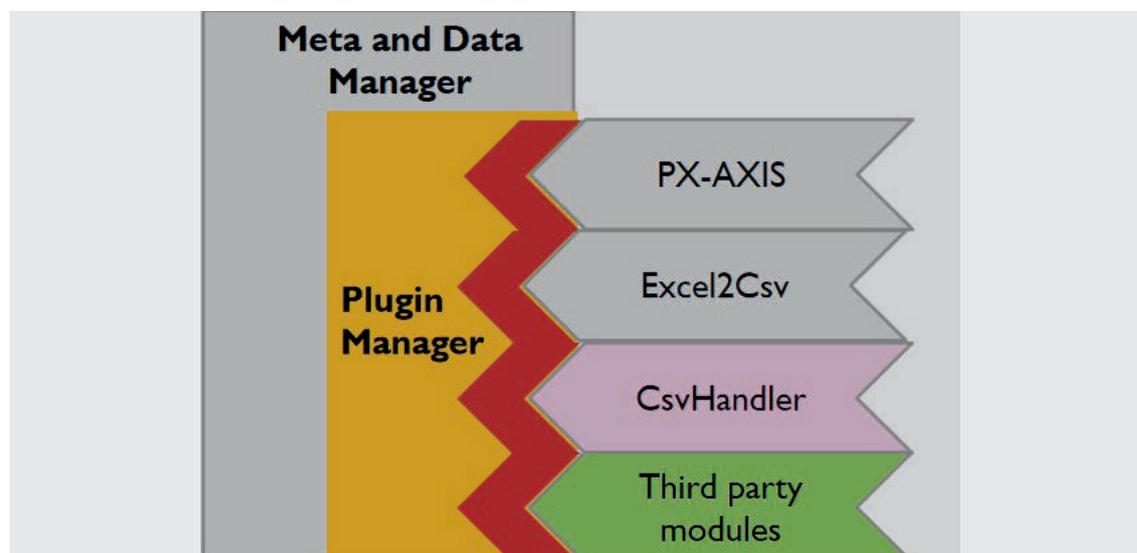
Figura 8.3 - Architettura di dispiegamento del *Meta and Data Manager*



Fonte: Istat

Attualmente il sistema è in fase di evoluzione, in particolare tramite l'aggiunta di una funzione che permette di inserire alcune procedure di trasformazione di strutture e formati dei dati, per effettuare caricamenti come funzioni a se stanti all'interno del sistema (Figura 8.4).

Figura 8.4 - Architettura funzionale del *Plugin Manager* di *Meta and Data Manager*



Fonte: Istat

8.3.2 Data Browser

Il *Data Browser* è una applicazione web che interagisce con API SDMX consentendo agli utenti di navigare, ricercare, interrogare, visualizzare tavole statistiche e *dataset*, ed effettuare il *download* in vari formati xml, csv, json. La rappresentazione dei *dataset* può avvenire sotto forma di tabelle multidimensionali, grafici e mappe tematiche.

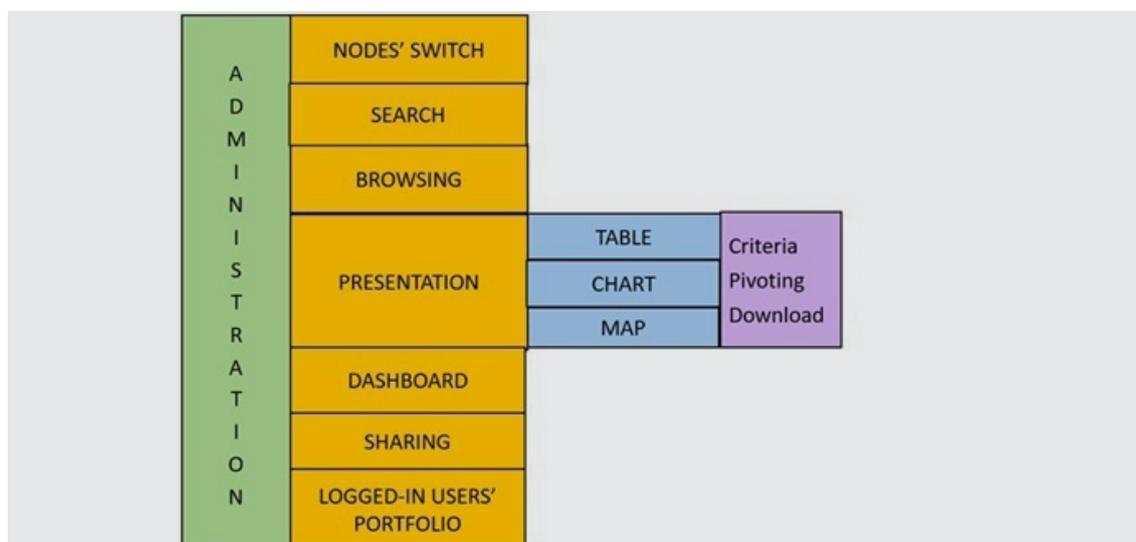
Tale applicazione può essere utilizzata all'interno di una singola organizzazione al fine di diffondere *dataset* archiviati in uno o più database, oppure nell'ambito di una architettura distribuita su server web raggiungibili attraverso il protocollo HTTP nella rete Internet.

Le principali funzionalità di tale applicazione sono le seguenti:

- la possibilità in un'unica istanza dell'applicativo di navigare, visualizzare e scaricare dati provenienti da più nodi, ovvero sorgenti delocalizzate. Ciò può avvenire in modalità *dispatcher*, ovvero navigando un nodo alla volta, o in modalità *hub*, ovvero accedendo a un nodo che permetta di interrogare in tempo reale i *dataset* provenienti da differenti sorgenti, come se tali dati fossero memorizzati nella stessa base dati;
- navigare e ricercare *dataset* all'interno di ogni singola sorgente dati digitando parole chiave o selezionando il *dataset* di interesse attraverso appositi alberi dei temi gerarchici;
- visualizzare i *dataset* in formato di tavole multidimensionali, di grafici o di mappe tematiche, potendo personalizzare il *layout* di tutte le forme di visualizzazione;
- effettuare il *download* dei dati in vari formati, anche *open data* come xml, SDMX-ml, SDMX-csv, SDMX-json, csv, *Excel*, jpeg e pdf;
- creare *dashboard* per la visualizzazione, all'interno di una stessa pagina web, di tabelle, grafici, mappe e testo i cui dati possono provenire da differenti *dataset* e sorgenti dati;
- la possibilità, per gli utenti registrati, di potere salvare le proprie visualizzazioni dei dati all'interno di un portafoglio consultabile una volta effettuata il *login* nel sistema.

Il sistema è inoltre dotato di un modulo di amministrazione che permette di configurare sia l'applicativo sia la connessione alle varie sorgenti delocalizzate tramite l'interfaccia web.

Figura 8.5 - Schema delle principali funzionalità del *Data Browser*

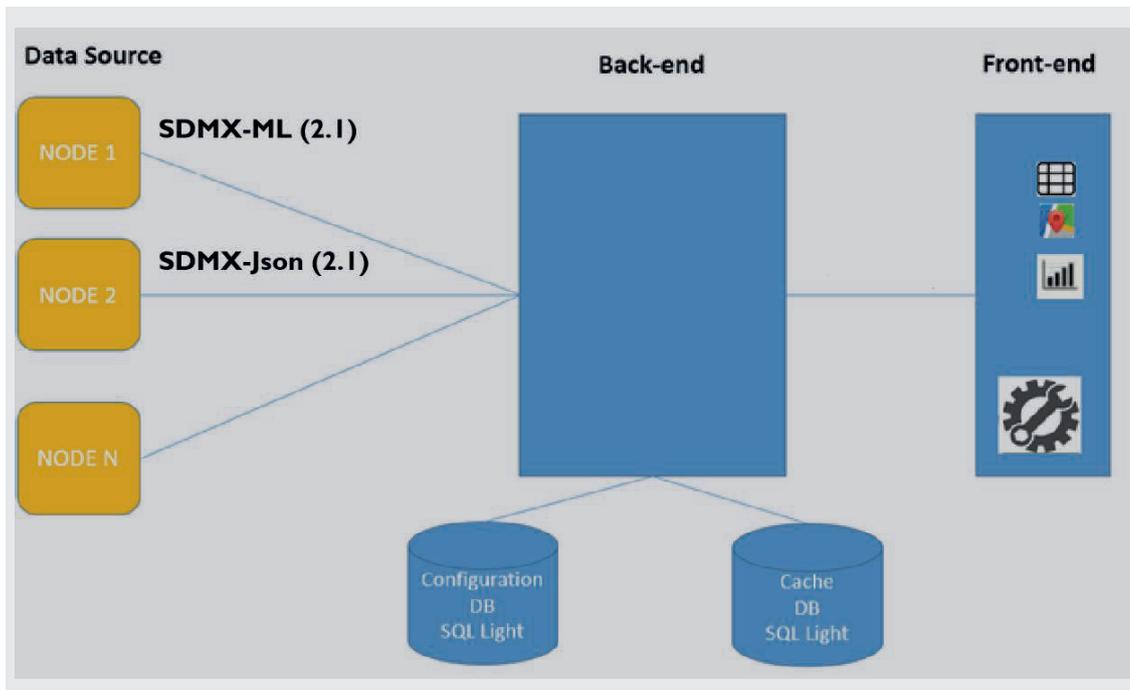


8. La piattaforma per la diffusione web dei dati

L'architettura di tale applicativo è anch'essa completamente orientata ai servizi, con un *front end*, lato client, che interroga un servizio di *back end*, il quale a sua volta gestisce tutti gli aspetti di navigazione e rappresentazione dei dati e interagisce con le sorgenti dati delocalizzate connettendosi ai rispettivi *web service* SDMX.

Il sistema è dotato di un sofisticato sistema di *cache* che permette di memorizzare i risultati di tutte le navigazioni effettuate dagli utenti sui vari *dataset* e di riutilizzarle per le successive interrogazioni, in conformità con i criteri di aggiornamento dei *dataset* medesimi che invalideranno le *cache* precedentemente create. Ciò consente di rendere altamente performanti le navigazioni, le interrogazioni e le visualizzazioni di dati, anche in presenza di *dataset* particolarmente corposi: può infatti effettuare la rappresentazione di un milione di celle dati memorizzate all'interno di *dataset*. L'architettura dell'applicativo *Data Browser* nel suo complesso è rappresentata in Figura 8.6.

Figura 8.6 - Architettura applicativa *service-oriented* del *Data Browser*



Fonte: Istat

Il sistema è inoltre dotato di tutta una serie di ulteriori funzionalità che permettono di personalizzare e rendere agevole la navigazione e la visualizzazione dei dati come:

- impostare filtri per restringere il campo di ricerca;
- scegliere la modalità di presentazione dei dati disponendo le dimensioni di analisi negli assi *x*, *y* e *z*, attraverso la funzione di *pivoting*;
- personalizzazione grafica dei *layout* di grafici e mappe, colori, *label*, titoli, posizionamento degli oggetti di visualizzazione;
- la possibilità di tenere aperte, nella medesima sessione di lavoro, differenti pagine di visualizzazione dei dati per potersi muovere da una pagina all'altra.

8.4 Disegno architetturale

L'*SDMX Istat Toolkit* è la nuova piattaforma applicativa di riferimento per l'implementazione dei sistemi di diffusione in Istat, a partire dal portale dati del Censimento permanente della Popolazione e delle abitazioni, che è stata la prima implementazione. La reingegnerizzazione dello stesso *Corporate Data warehouse* e dei sistemi tematici di diffusione esistenti, così come l'implementazione di nuovi sistemi di diffusione devono essere effettuati, pertanto, considerando i moduli dell'*SDMX Istat Toolkit* come la soluzione applicativa a cui riferirsi, a meno di motivate eccezioni.

L'architettura di dispiegamento basata su tale architettura prevede l'utilizzo del *Meta and Data Manager* come strumento di *back office* che permette di:

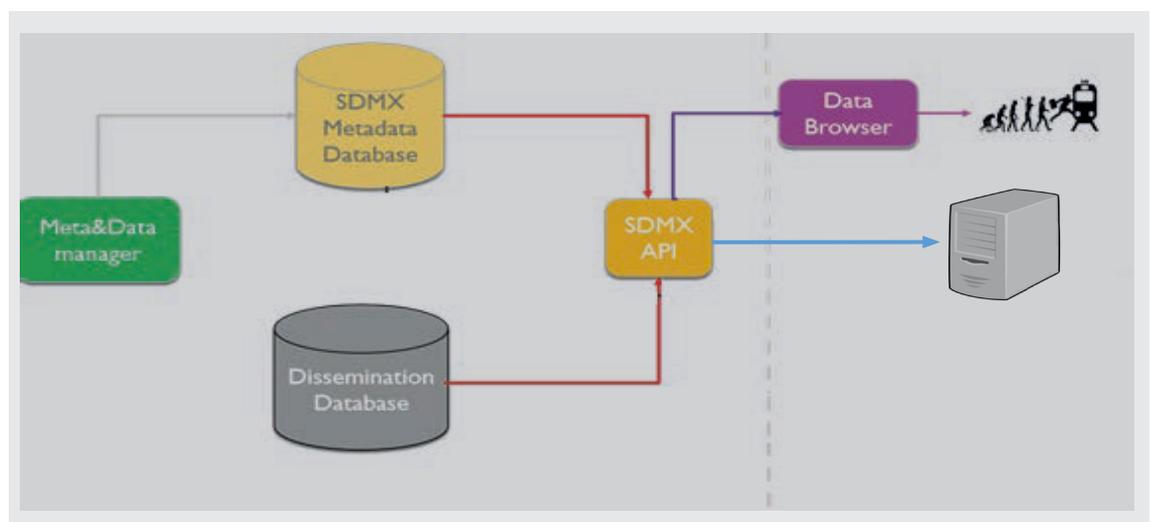
- modellare gli ipercubi multidimensionali che conterranno i dati da diffondere;
- gestire la memorizzazione dei dati all'interno dell'architettura;
- pubblicare in modalità *machine-to-machine* i *dataset*, ciascuno basato sulla selezione dei dati memorizzati all'interno di uno degli ipercubi multidimensionali, esponendoli tramite il *web service* SDMX;

Il *Data Browser* rappresenta lo strumento di *front end* dell'architettura che, connettendosi ai *web service* SDMX che pubblicano i dati e i relativi metadati di struttura dei *dataset*, permette di potere navigare e visualizzare i contenuti messi a disposizione tramite un'apposita interfaccia per gli utenti finali.

L'accoppiamento tra lo strumento di *back office* e lo strumento di *front end* è pertanto basato solo sull'utilizzo del *web service* SDMX; quest'ultimo pubblica i dati e i metadati del *back office* e può essere oggetto di connettività da parte dell'applicazione *Data Browser*.

Inoltre, in tale tipo di architettura, a partire dagli stessi *repository* di dati e di metadati è possibile effettuare la diffusione verso gli utenti finali sia in modalità interattiva, ossia tramite *data browser*, sia in modalità *machine-to-machine*.

Figura 8.7 - Architettura di dispiegamento complessiva del *Corporate Data warehouse* Istat



8.5 Evoluzioni programmate

Le funzionalità evolutive, che nel breve e medio periodo saranno applicate alla piattaforma in modo generale all'*SDMX Toolkit*, *Meta and Data Manager* e *Data Browser*, e di cui potrà beneficiare anche l'infrastruttura applicativa relativa al portale dati del Censimento permanente della Popolazione e delle abitazioni, riguardano i seguenti aspetti principali:

- ulteriori ottimizzazioni nell'interrogazione di *dataset* di grandi dimensioni, anche di decine di milioni di *record*, tramite alcune modifiche relative alla tipologia di *query* SDMX utilizzate, al fine di sfruttare al meglio, anche con il supporto della *cache* di dati, l'utilizzo dei filtri per la contestualizzazione delle *query*;
- la possibilità di condividere visualizzazioni di dati relative ai *dataset*, alle tavole, ai grafici e alle mappe, in modo che possano essere inserite all'interno di interfacce web esterne al *Data Browser*;
- la modifica del *workflow* di pubblicazione dei dati con la possibilità di sincronizzare applicativamente gli inserimenti e le modifiche di dati e metadati, effettuate in ambiente interno di *staging*, con l'ambiente di produzione esposto senza dovere ripetere manualmente le operazioni.

8.6 Cubus

Il software *Cubus* è uno strumento per i settori di produzione finalizzato all'elaborazione massiva degli ipercubi di dati aggregati che vanno ad alimentare i sistemi di *data warehouse* interni all'Istat, *I.Stat* e il più recente *Data Browser*, partendo dai microdati e dalle classificazioni, sia su file locali, *Excel*, *csv*, *txt*, sia su database server *Oracle*, *Microsoft SQL Server*, *MySQL*, *PostgreSQL*.

Tali microdati e classificazioni vengono presentati come *dataset* sui quali è possibile effettuare una serie di operazioni propedeutiche alla composizione dei piani di spoglio di pubblicazione, necessari per il computo dei dati aggregati.

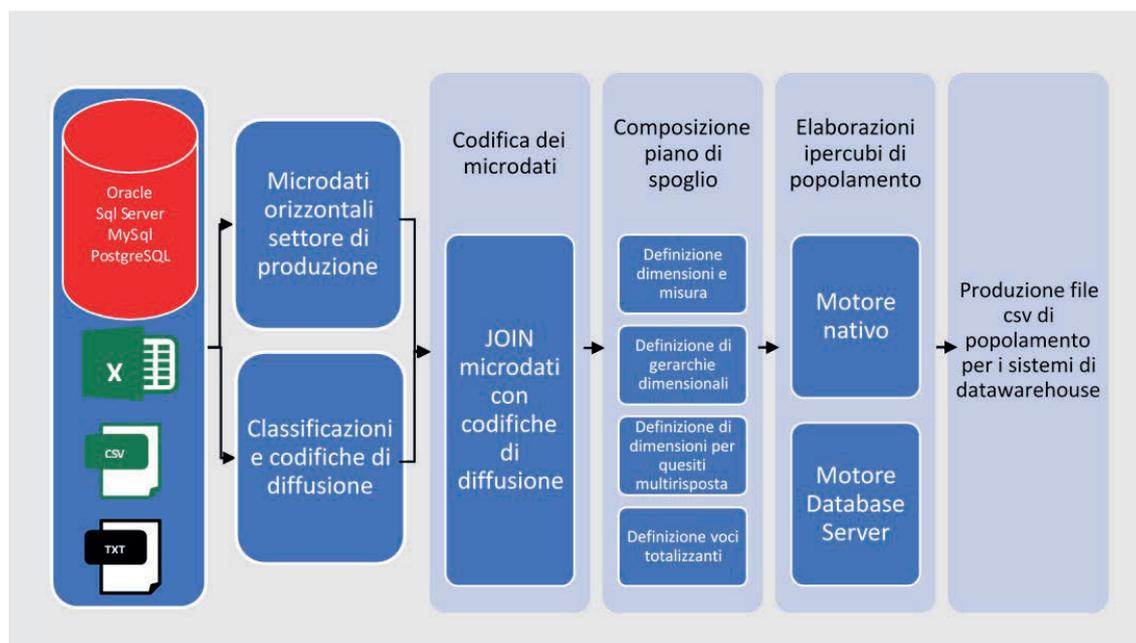
L'attività di composizione di un piano si esplica essenzialmente attraverso dei semplici trascinalamenti in una griglia, *drag and drop*, dei campi dei *dataset* che costituiscono le dimensioni dell'ipercubo e attraverso la definizione della misura di analisi, funzione o espressione di raggruppamento, che si intende valorizzare.

Per ogni dimensione è necessario definire la cosiddetta voce totalizzante, ossia il valore che esprime l'insieme di tutte le modalità assunte dalla dimensione stessa, o specificare se si intende non valorizzare tale voce.

L'elaborazione di un ipercubo da uno o più piani di spoglio è un processo massivo di calcolo combinatorio applicato a istruzioni SQL di aggregazione di microdati; gli ipercubi possono poi essere esportati per popolare le piattaforme di *data warehouse* interne.

Il software è progettato per lavorare su volumi di dati di milioni di *record* ed è realizzato in C#, architettura *Microsoft Windows Forms .Net* a 64 bit, *.NET Framework 4.7*, installato di *default* su tutti i sistemi operativi Windows.

Il flusso di lavoro del software *Cubus* è mostrato in Figura 8.8.

Figura 8.8 - Flusso di lavoro di *Cubus*

Fonte: Istat

8.6.1 *Cubus* per il Censimento generale dell'Agricoltura

La diffusione dei dati pervenuti dalle rilevazioni delle indagini e dei censimenti avvengono in forma aggregata, e le indicazioni su come aggregare i dati e su quali informazioni pubblicare sono definite da uno o più piani di spoglio (PDS).

I PDS vengono definiti dai referenti e responsabili del dato censuario, in collaborazione con il responsabile della diffusione che ha il compito ulteriore di verificarne la congruità, la fattibilità o e la compatibilità con le classificazioni pubbliche Istat.

Il risultato della elaborazione del PDS sui microdati verrà poi caricato nel sistema di diffusione *Data Browser* come ipercubo di informazioni, in formato csv con un tracciato fisso definito a priori.

Per la diffusione dei dati del Censimento generale dell'Agricoltura del 2020 si è scelto di utilizzare il software applicativo *Cubus*, che permette di produrre con facilità gli ipercubi che devono essere forniti al *Data Browser* rispettando tutte le caratteristiche richieste.

La base di lavorazione sono i PDS definiti e approvati, che sono trasmessi come file *Excel* contenente più fogli; nel primo foglio è descritto ciò che si vuole diffondere e sono elencate in colonna:

- i tipi dato, vale a dire il fenomeno che si vuole studiare, un esempio è costituito dal numero di aziende o la superficie aziendale;
- le dimensioni di analisi per le quali si vuole studiare il fenomeno, viste come classificazioni, come ad esempio la forma giuridica o il tipo di coltivazione;
- i livelli territoriali per i quali si intende fare diffusione: Italia, regione, provincia;
- il tempo che, nel nostro caso, è l'anno di rilevazione.

Nelle righe sono rappresentati i diversi incroci tra tipi dato e dimensioni di analisi che si vogliono realizzare e che definiscono ogni singolo PDS; le caselle contrassegnate da 'X'

8. La piattaforma per la diffusione web dei dati

all'intersezione di righe e colonne, indicano quali incroci, tra i tanti tipi dato e dimensioni di analisi presenti, devono essere effettivamente sviluppati per il PDS.

Nel Prospetto 8.1, ad esempio, per il *progr 1* si vogliono conoscere il numero di persone capo azienda che hanno frequentato corsi, suddivise per forma giuridica, sesso ed età a livello territoriale Italia e regione; per il *progr 3* si vogliono conoscere invece 2 fenomeni distinti: la Sau (superficie agricola utilizzata) e il numero di aziende con superficie agricola utilizzata, entrambi suddivisi per tipo di coltivazione a livello territoriale Italia e Regione.

Prospetto 8.1 - Estratto del primo foglio del file Excel

Progr	VC1		VC2 tipo dato				VC3	VC4	VC7	VC16	VC17	VC35
	Livello territoriale	Superficie a boschi	Persone capo azienda che hanno frequentato corsi	Aziende agricole	Sau	Azienda con superficie agricola utilizzata	Forma giuridica	Titolo di possesso dei terreni	Tipo di coltivazione	Sesso	Classe di età	Tempo
1	Paese - regione		X				X			X	X	2020
2	Paese - regione			X						X	X	2020
3	Paese - regione				X	X			X			2020
4	Paese - regione	X			X	X						2020

Fonte: Istat

Negli altri fogli del file *Excel* si trova il dettaglio delle modalità delle classificazioni previste per le variabili di analisi di solo interesse per la pubblicazione, come ad esempio la forma giuridica (Prospetto 8.2).

Prospetto 8.2 - Classificazione della forma giuridica

CODICE	DESCRIZIONE
11_X11	Imprenditore o azienda individuale o familiare
12	Società di persone
13	Società di capitali
14	Società cooperativa
15	Consorzio di diritto privato e altre forme di cooperazione tra imprese
OTH_PADM_CA2020	Altri enti privati
X216	Amministrazione o ente pubblico
X299	Ente (comunanze, università, regole, eccetera) o Comune che gestisce le proprietà collettive
TOT	Totale

Fonte: Istat

Ciò che si deve ottenere è definito come misura, un numero dato dagli incroci delle dimensioni di analisi per il tipo dato desiderato, ottenuto utilizzando funzioni di raggruppamento quali conteggio, somma eccetera. Sulle singole unità della rilevazione, possono essere richieste anche elaborazioni più complesse che necessitano di uno sviluppo di funzioni ad hoc.

Una fase propedeutica alla produzione degli ipercubi è stata quella, di analizzarli per verificare ulteriormente la corrispondenza tra i dati raccolti nella rilevazione e le pubblica-

zioni previste, ad esempio si è prestata particolare attenzione nel controllare che per tutte le codifiche di classificazioni delle dimensioni di analisi, non si verificassero casi di modalità presenti in pubblicazione e non nei dati e viceversa.

Si è poi proceduto a una analisi di contenuto per definire un nuovo database di lavorazione, dove i microdati consolidati, sono stati elaborati al fine di renderli conformi alla pubblicazione in termini di coerenza con i tipo dato e le dimensioni di analisi previste.

Per consentire la pubblicazione è quindi necessario effettuare alcune modifiche quali:

- riclassificazione di variabili per le quali sono previste modalità o voci di classificazione differenti tra dato raccolto e classificazione pubblica dell'Istat. Un esempio è la dimensione sesso che potrebbe essere raccolto in rilevazione con i valori M e F, ma che nella classificazione pubblica prevede i valori 1 per Maschio e 2 per Femmina;
- creazione di nuove variabili; un esempio è dato dalla dimensione di analisi 'classi di età del conduttore dell'azienda' che prevede valore 1 se l'età è compresa tra 15 e 19 anni, valore 2 se l'età è compresa tra 20 e 49 anni, valore 3 se l'età è 50 e oltre; poiché nella rilevazione è presente solo la data di nascita, devono essere definite due nuove variabili, una per l'età calcolata e un'altra che rappresenta il valore del raggruppamento della classe di età, determinata sulla base della nuova variabile età calcolata, che assumerà il valore 1, 2 o 3;
- creazione di nuove variabili ricalcolate da operazioni su altre; un esempio è il calcolo delle superfici in ettari con decimale, quando nella rilevazione il dato è stato raccolto in due variabili distinte, ettari e are;
- verifica che le variabili ricalcolate non generino problemi di arrotondamenti;
- riclassificazione del territorio secondo la codifica NUTS.

Contestualmente e a completamento di questa fase di analisi e di ricalcolo del microdato, si è studiato anche come procedere per la realizzazione degli ipercubi attraverso *Cubus* cioè quali strutture fosse meglio creare nel database per meglio sfruttare le potenzialità dello strumento.

La scelta operativa per il Censimento generale dell'Agricoltura del 2020 è stata quella di creare una tabella nel nuovo database che rappresentasse il dato esattamente come doveva essere pubblicato.

La tabella contiene tutte le dimensioni di analisi, opportunamente ricreate e/o riclassificate e il territorio, riportato alla codifica NUTS.

Tale impostazione ha semplificato il lavoro su *Cubus* così da essere svolto non necessariamente da personale informatico.

Le lavorazioni dei piani di spoglio su *Cubus* hanno previsto i seguenti passaggi:

- si definisce inizialmente il *dataset* di lavorazione che punta in remoto alla tabella definita nel database *Oracle* sul quale viene definita la connessione;
- in due sezioni della schermata di *Cubus* (Figura 8.9) il contenuto del *dataset* viene rappresentato sia in termini di tracciato, cioè di colonne esistenti, sia in termini di contenuto, attraverso la visualizzazione di una anteprima dei dati;
- si iniziano a definire i singoli PDS che sono uno per ogni tipo dato di ogni riga del file *Excel* ove presente la biffatura 'X';
- alla definizione di un nuovo PDS viene generato un prospetto con colonne e righe inizialmente vuote che deve essere costruito sulla base delle variabili di analisi, tipo dato e livello territoriale, trascinando nelle intestazioni delle colonne le variabili del

dataset che rappresentano ciascuna dimensione e creando una colonna fissa per il tipo dato;

- lo standard dell'output dei csv da fornire per la pubblicazione prevede che nella prima colonna sia presente il territorio, nella seconda il tipo dato e poi a seguire le dimensioni di analisi che possono essere tutte quelle previste nel foglio *Excel* contenente i PDS, oppure solo quelle del PDS in lavorazione e, infine, la misura che si vuole ottenere. Nel caso del Censimento generale dell'Agricoltura del 2020, per le colonne delle dimensioni, la scelta è stata di riportare sempre tutte le dimensioni di analisi, anche quelle che non rientrano come incrocio nello specifico PDS, in modo che il loro modello di importazione del dato nel *Data Browser* fosse sempre il medesimo;
- per il territorio viene definita una sola colonna con i vari livelli di gerarchia territoriale previsti; su questa colonna si indica anche al sistema come gestire la gerarchia;
- per le altre colonne delle dimensioni si può operare in due modi dipendenti dallo sviluppo o meno della dimensione. Nel primo caso viene creata una colonna vuota e poi si trascina al suo interno il campo del *dataset* che rappresenta la dimensione. Nel secondo caso viene creata una colonna con valore fisso, impostando il valore della voce che indica la totalità dei valori della dimensione della classificazione corrispondente alla dimensione;
- per tutte le colonne non a valore fisso, deve essere indicata la voce totalizzante, ovvero la sua assenza. Se viene indicata la voce totalizzante, *Cubus* provvede, in maniera automatica, alla generazione delle righe di totale, compresa quella del territorio;
- l'ultima colonna da definire è quella che calcola la misura di analisi; per questa si possono utilizzare funzioni di raggruppamento classiche come conteggio, somma, media eccetera su un singolo campo, oppure scrivere espressioni SQL complesse di raggruppamento;
- una volta definite tutte le colonne, il PDS deve essere generato e il risultato viene esportato in formato csv, fornendo al sistema indicazioni come nome e *directory* del file di esportazione, se creare o meno un file compresso e il carattere separatore delle colonne;
- poiché *Cubus* offre anche la possibilità di esportare più piani di spoglio in un'unica operazione, uno standard utilizzato è quello di assegnare subito un nome per il file di esportazione a ciascun PDS. Ciò ha semplificato la fase di esportazione, sia se si doveva esportare un unico PDS perché si manteneva sempre il nome, sia se si esportavano i PDS in massa, perché in questo modo ognuno aveva il suo nome e non si creava confusione.

Dopo avere generato ed esportato il file, si è provveduto al controllo del risultato prodotto, sia verificando con il personale del Servizio Statistiche e rilevazioni sull'agricoltura che avevano già pubblicato dati per Eurostat, sia ricreando delle *query* sulla base dati originale. La scelta di fare il controllo sulla base dati originale e non sulla nuova tabella definita per essere utilizzata da *Cubus*, è stata fatta allo scopo di verificare che il risultato del csv fosse corretto e, ancora prima, per assicurarsi che non si fossero generati errori nella definizione della nuova tabella.

Nella Figura 8.9 è rappresentata la schermata di *Cubus* con le varie parti che la compongono; la parte centrale è il prospetto di definizione del PDS dove si costruiscono le colonne che comporranno il file csv.

Figura 8.9 - Schermata di lavorazione di *Cubus*

Anteprima del dataset 'SV_CENSAGR_MICRO_SV_ISTAT_QUESTIONARIO_NO_CLFILTRATA'

UTONOMA	PROVINCIA_AUTONOMA_ORIG	BOSCHI_FILTRO	WOODTR_FILTRO	MUSHRO_FILTRO	SANU2_FILTRO	CORSI_CAPO
[NULL]	[NULL]	0	2	2	2	2
[NULL]	[NULL]	0	2	2	2	2
[NULL]	[NULL]	0	2	2	2	2
[NULL]	[NULL]	0	2	2	2	2
[NULL]	[NULL]	0	2	2	2	2
[NULL]	[NULL]	1	2	2	2	2

Trascina i campi del dataset sulle intestazioni di colonna. Definisci la misura cliccando con il tasto destro sull'intestazione dell'ultima colonna. Definisci un valore fisso per un campo cliccando con il tasto destro sull'intestazione della colonna. Premi CANC per cancellare righe o colonne selezionate.

5 - Prospetto di definizione del PDS

TERRITORIO	TIPODATO=PE_CA_AGFO	CAPO_SESSO	ALTRECOL=TOT	CAPO_ETA_GRI	ANNO=2020	MISURA
PROVINCIA_AUTONOMA NUTS_RIP_5 NUTS_REGIONE NUTS_PRO						COUNT(CASE WHEN CORSI_CAPO = 1 THEN 1 ELSE NULL END)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Fonte: Istat, software *Cubus*

Nella Figura 8.10 è evidenziata la presentazione dell'anteprima dell'ipercubo generato sulla quale è possibile fare operazioni di ordinamento, scorrimento per singola riga o pagina, e conoscere il totale delle righe generate; quest'ultima è un'importante caratteristica poiché è possibile eseguire già delle prime verifiche sul risultato; nel caso ipercubi molto complessi, con molte righe e colonne da verificare, si rivela senz'altro più comodo eseguire l'esportazione.

Figura 8.10 - Schermata di *Cubus* con anteprima ipercubo

Anteprima ipercubo 1 del piano PDS_CAPI_CORSI del dataset SV_CENSAGR_MICRO_SV_ISTAT_QUESTIONARIO_NO_CLFILTRATA

TERRITORIO	TIPODATO=PE_CA_AGFO	CAPO_SESSO	ALTRECOL=TOT	CAPO_ETA_GRI	ANNO=2020	MISURA
IT	PE_CA_AGFO 9	TOT	TOTAL	2020	327755	
IT	PE_CA_AGFO 9	TOT	Y30-44	2020	57373	
IT	PE_CA_AGFO 9	TOT	V45-59	2020	115488	
IT	PE_CA_AGFO 9	TOT	V60-74	2020	103142	
IT	PE_CA_AGFO 9	TOT	V_0E75	2020	40041	
IT	PE_CA_AGFO 9	TOT	V_LIN29	2020	11761	
IT	PE_CA_AGFO 1	TOT	TOTAL	2020	260599	
IT	PE_CA_AGFO 1	TOT	Y30-44	2020	48696	
IT	PE_CA_AGFO 1	TOT	V45-59	2020	91738	
IT	PE_CA_AGFO 1	TOT	V60-74	2020	87217	
IT	PE_CA_AGFO 1	TOT	V_0E75	2020	35635	
IT	PE_CA_AGFO 1	TOT	V_LIN29	2020	9233	
IT	PE_CA_AGFO 2	TOT	TOTAL	2020	59246	
IT	PE_CA_AGFO 2	TOT	Y30-44	2020	12687	
IT	PE_CA_AGFO 2	TOT	V45-59	2020	23760	
IT	PE_CA_AGFO 2	TOT	V60-74	2020	19205	
IT	PE_CA_AGFO 2	TOT	V_0E75	2020	4406	
IT	PE_CA_AGFO 2	TOT	V_LIN29	2020	2528	

Trascina i campi del dataset sulle intestazioni di colonna. Definisci la misura cliccando con il tasto destro sull'intestazione dell'ultima colonna. Clicca con il tasto destro sull'intestazione di una colonna per definire un valore fisso, aggiungere o rimuovere colonne, rimuovere campi da colonne multiple. Premi CANC per cancellare righe o colonne selezionate.

Anteprima ipercubo

5 - Prospetto di definizione del PDS

TERRITORIO	TIPODATO=PE_CA_AGFO	CAPO_SESSO	ALTRECOL=TOT	CAPO_ETA_GRI	ANNO=2020	MISURA
PROVINCIA_AUTONOMA NUTS_RIP_5 NUTS_REGIONE NUTS_PRO						COUNT(CASE WHEN CORSI_CAPO = 1 THEN 1 ELSE NULL END)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Fonte: Istat, software *Cubus*

Nello sviluppo dei numerosi PDS previsti per la pubblicazione dei risultati del Censimento generale dell'Agricoltura del 2020, *Cubus* si è rivelato uno strumento prezioso poiché ha permesso una produzione in minore tempo e con minori risorse rispetto a quella fatta con *query* SQL dedicate da sviluppare per ogni PDS.

Poiché è prevista la generazione delle righe corrispondenti alle voci totalizzanti in maniera automatica, è stato possibile fare le verifiche sul totale Italia partendo dal presupposto che se il totale Italia è corretto, lo è anche quello degli altri livelli territoriali. Ciò ha permesso di velocizzare significativamente le operazioni di controllo; infatti, solo se questi controlli non andavano a buon fine, si entrava maggiormente nel dettaglio per capire l'origine dell'errore e procedere con i corretti aggiustamenti che potevano essere fatti o sul database o direttamente su *Cubus*, qualora si fossero impostate in modo errato delle condizioni.

8.6.2 *Cubus* per il Censimento permanente delle istituzioni pubbliche

La fase di diffusione del Censimento permanente delle istituzioni pubbliche sul sito istituzionale dell'Istat, come per gli altri Censimenti, comporta un lavoro di produzione di output di reportistica dei conteggi per ogni raggruppamento di unità considerato.

La definizione di questi conteggi e delle modalità con cui si effettuano i raggruppamenti di unità è, come è stato detto, il piano di spoglio, le cui specifiche sono veicolate dal personale dell'Istat della Direzione centrale per la comunicazione, informazione e servizi ai cittadini, in particolare del servizio che si occupa della diffusione dei dati.

I dati dell'Istat, fatta eccezione per casi particolari, vengono essenzialmente diffusi in forma aggregata, quindi il piano di spoglio definisce anche come si passa dai dati grezzi ai dati aggregati, privati così delle connotazioni che rendano le unità identificabili.

Il file con il piano di spoglio è fornito sotto forma di file *Excel* multifoglio; nel primo foglio vengono riportati i campi che costituiscono l'intestazione del file csv da riprodurre in output e che viene successivamente fornito al personale che si occupa della diffusione, comprensivo della variabile che visualizza i conteggi delle unità associate.

Gli altri fogli contengono i dettagli delle modalità da riportare per ogni grandezza da misurare, corrispondenti alle voci da espandere sui menu visualizzabili sul sito.

Nella Figura 8.11 è riportato il dettaglio delle voci del menu del Censimento delle istituzioni pubbliche, a ognuna delle quali corrisponde un file del piano di spoglio.

Figura 8.11 - Menu sul *frame* a sinistra di I.Stat

The screenshot shows the 'CENSIMENTI PERMANENTI data warehouse' interface. On the left, there is a navigation menu with the following items:

- ISTITUZIONI NON PROFIT
- CENSIMENTO DELLE ISTITUZIONI PUBBLICHE
 - Istituzioni, risorse umane e genere
 - Istituzioni, risorse umane e genere - dati comunali
 - Unità locali, risorse umane e genere
 - Unità locali, risorse umane e genere - dati comunali
 - Attività formativa del personale
 - Occupazione per tipologia contrattuale e regime orario
- CENSIMENTO DELLE IMPRESE
 - Censimento industria, istituzioni pubbliche e non profit 2011

The main content area displays the following text:

Benvenuto in CensimentiPermanenti.Stat

L'Istat rende disponibili i primi risultati dei censimenti permanenti, realizzati attraverso rilevazioni continue a cadenza annuale triennale. Con la tornata censuaria del 2010-2011 si è quindi chiusa una lunga fase della storia della statistica pubblica caratterizzata dai censimenti generali decennali.

I censimenti permanenti non coinvolgono tutti i cittadini, le imprese e le istituzioni, ma di volta in volta dei campioni rappresentativi restituiti al Paese dei dati ottenuti è di tipo censuario, quindi riferibile all'intero campo d'osservazione.

Grazie all'integrazione tra fonti amministrative e rilevazioni campionarie i nuovi censimenti garantiscono l'eshaustività, l'aumentata qualità dell'offerta informativa, il contenimento del fastidio statistico sui rispondenti, la riduzione dei costi complessivi.

La strategia dei censimenti permanenti è estesa a tutte le aree tematiche: popolazione e abitazioni, imprese, istituzioni non pubbliche. A partire dal 2021 anche il censimento dell'agricoltura diventerà permanente.

Per saperne di più visita il sito <https://www.istat.it/it/censimenti-permanenti>

Per la preparazione dei file in output con estensione csv si è utilizzato il software *Cubus*, sviluppato interamente in Istat.

Prima che *Cubus* possa produrre l'output è necessario compiere le seguenti attività, avendo come punto di partenza le specifiche riportate sui piani di spoglio:

- codificare le risposte al questionario in modo che i dati grezzi riportati sul database vengano elaborati così da produrre le modalità testuali che andranno a costituire parte dell'intestazione sul file di output. A tale proposito sono costruite delle viste di dati che combinano le variabili di questionario tramite opportune formule;
- gestire le variabili che comportano la suddivisione dell'universo dei rispondenti in modo da avere modalità non mutuamente esclusive. Ad esempio, le unità ricadenti nella classe dei codici ATECO a 2 cifre comprendono come sottoinsieme i codici a 4 cifre secondo la codifica NACE e quindi i codici non sono mutuamente esclusivi.

Nella Figura 8.12 è riportato un foglio *Excel* con il prospetto generale relativo al piano di spoglio: il lavoro per la produzione dei file finali di output prende l'avvio dalla lettura di tale foglio.

Figura 8.12 - Estratto dal prospetto generale del piano di spoglio del Censimento permanente delle istituzioni pubbliche 2020

INCROCI	VC1	VC2	VC3					VC4		VC5
Variabili di Classificazione (VC)										
Codice DS (solo se ci sono più DS)	Territorio (N) (Territory)	Tipo dato (Data type)	Ateco 2007 (Nace2007)					Forma giuridica (Legal form)		Classe di unità locali (Class of local units)
			Tot.	N1	N2	N3	N4	N5	agg.	
		x(ip)	x							x
		x(ip)	x	x	x	x	x	x		x
		x(ip)	x	x	x	x	x	x		x
		x(ip)	x	x	x	x	x	x		x
		x(ip)	x	x	x	x	x	x		x
		x(ip con pers eff)	x	x	x	x				x
		x(ip con dip)	x	x	x	x				x
		x(pers eff)	x	x	x	x	x			x
		x(pers eff)	x	x	x	x	x	x		x
		x(dip)	x	x	x	x	x			x
		x(dip)	x	x	x	x	x	x		x
		x(vol)	x	x	x	x	x	x		x

Fonte: Istat

Le variabili da riportare sull'intestazione dell'output sono rappresentate dai gruppi di colonne indicati con la dicitura VC sulla riga 2, con un numerico successivo che può assumere i valori interi da 1 a 9. Ogni *query* ha una modalità o un complesso di modalità da esplicitare, a partire da quella con il contrassegno VC5.

A loro volta, per ognuna delle variabili di classificazione, viene riportato un foglio *Excel* di dettaglio, ad esempio per la variabile VC1 corrisponde il foglio denominato C1. Le "X" indicate nelle celle del foglio di prospetto generale corrispondono alle modalità da esplicitare nell'output relativamente alla *query* indicata dall'ID.

8. La piattaforma per la diffusione web dei dati

Le variabili di classificazione possono essere sia strutturali, ossia estratte a partire da una tabella dati di anagrafica, sia ricavate dalle risposte al questionario.

Di seguito un esempio di foglio di dettaglio nel quale viene esplicitata la variabile VC5, la classe di unità locali. Per come è strutturato il piano di spoglio quest'ultima andrà espansa soltanto in corrispondenza della *query 1* in tutte le sue voci, data la presenza di una "X", mentre negli altri casi si riporta soltanto la voce di totale.

Prospetto 8.3 - Foglio di dettaglio di una variabile di classificazione

Classificazione pubblica di riferimento				
Nome	Codice			
Numerosità	NUMEROSITA			
Modalità della classificazione				
Codice	Descrizione in italiano	Descrizione in inglese	Ordine di visualizzazione	Modalità utilizzate
N1	1	1		x
N2	2	2		x
N3-5	3-5	3-5		x
N6-10	6-10	6-10		x
N11-49	11-49	11-49		x
N50-99	50-99	50-99		x
N100-499	100-499	100-499		x
N500-999	500-999	500-999		x
N1000_GE	1.000 e più	1,000 and over		x
N1GE	1 e più	1 and over		x (voce di totale)

Fonte: Istat

Ciò significa che in corrispondenza della *query 1* verranno riportate le voci N1, N2, N3-5, N6-10, N11-49, N50-99, N100-499, N500-999, N1000_GE, N1_GE, corrispondenti alle classi di numerosità di unità locali riportate nella colonna contrassegnata dalla dicitura "Descrizione in italiano". In corrispondenza delle altre *query*, per le quali la variabile della classe di unità locali non riporta la "X", sull'output si dovrà indicare semplicemente la voce di totale, ossia la N1_GE.

Si passa quindi all'implementazione su *Cubus*: dopo avere avviato il programma, vengono scaricati gli ultimi aggiornamenti e viene configurata la connessione alla base dati *Oracle*. La vista contenente i dati classificati secondo le variabili di interesse viene quindi selezionata e collegata a *Cubus*.

A questo punto, sulla schermata di destra viene visualizzata una griglia su un sottoinsieme di dati mentre su quella di sinistra compaiono tutti i pulsanti con cui è possibile implementare un piano di spoglio per ogni *query* sul foglio *Excel*.

Con riferimento alla *query 1* nella Figura 8.12, le colonne di nostro interesse vengono trascinate e riportate sulla parte in basso del *frame* destro di *Cubus*.

Quando su una colonna viene trascinato più di un campo, viene data la possibilità di selezionare una tra le seguenti modalità: separa, unisci e unisci non nulli.

In alternativa, può essere riportato un valore fisso su una colonna.

Sull'ultima colonna va quindi inserita la funzione di raggruppamento dei campi per ottenere il valore riportato aggregato sull'output finale.

In corrispondenza delle colonne estratte dalla base dati che sono associate a una voce totalizzante del piano di spoglio, si apre una specifica interfaccia, qui di seguito

riportata, che consente di indicare quali sono le modalità da prendere in considerazione per il calcolo del totale.

Figura 8.13 - Definizione delle voci di totale

Nome campo	Voce totalizzante	Non elaborare la voce totalizzante*
COD_PRO_I_STAT	IT	<input type="checkbox"/>
COD_REG_I_STAT	IT	<input type="checkbox"/>
COD_RIP_I_STAT	IT	<input type="checkbox"/>
N3	0010	<input type="checkbox"/>
N2	0010	<input type="checkbox"/>
N1	0010	<input type="checkbox"/>
N5	0010	<input type="checkbox"/>
N4	0010	<input type="checkbox"/>
FG_AGGR_COM		<input checked="" type="checkbox"/>
CLASSE_UL	N1_GE	<input type="checkbox"/>

Fonte: Istat, Software *Cubus*

Si popola così la parte in basso a destra come in Figura 8.14.

Figura 8.14 - Definizione dell'ipercubo per l'output

separa unisci unisci non nulli		separa unisci unisci non nulli					
TERRITORIO		ATECO_2007					
COD_RIP_I_STAT	TIPO_DATO=APIN	N1	FG_AGGR_COM	CLASSE_UL	TIPO_RISORSA=99	CLASSE_TOT_A7=TOT	CONTA
COD_REG_I_STAT		N2					
COD_PRO_I_STAT		N3					
		N4					
		N5					
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	count(*)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	

Fonte: Istat, Software *Cubus*

Successivamente si può visionare un'anteprima dell'output, come mostrato in Figura 8.15, pronta per essere salvata sul file csv finale.

8. La piattaforma per la diffusione web dei dati

Figura 8.15 - Anteprima output finale

Anteprima ipercubo '1 del piano Piano 02/11/2023 14:23:07:92 del dataset PD_CENSIP_CHECK_UI_V_ANAGRAFICA_2020_QIP20_CB'

TERRITORIO	TIPO_DATO	ATECO_2007	FG_AGGR_COM	CLASSE_UL	TIPO_RISORSA	CLASSE_TOT_A7	CONTA
IT108	APIN	G	X1510_1520_16_17_18_23	N3-5	99	TOT	1
IT108	APIN	O	2420	N3-5	99	TOT	1
IT108	APIN	O	2430	N1	99	TOT	9
IT108	APIN	O	2430	N2	99	TOT	13
IT108	APIN	O	2430	N3-5	99	TOT	20
IT108	APIN	O	2430	N6-10	99	TOT	12
IT108	APIN	O	2430	N11-49	99	TOT	1
IT108	APIN	O	X2440_2450	N3-5	99	TOT	2
IT108	APIN	O	X1510_1520_16_17_18_23	N1	99	TOT	1
IT108	APIN	O	X1510_1520_16_17_18_23	N6-10	99	TOT	1
IT108	APIN	P	X1510_1520_16_17_18_23	N1	99	TOT	1
IT108	APIN	Q	2500	N11-49	99	TOT	3
IT108	APIN	Q	X1510_1520_16_17_18_23	N1	99	TOT	1
IT108	APIN	Q	X1510_1520_16_17_18_23	N3-5	99	TOT	1
IT108	APIN	Q	X1510_1520_16_17_18_23	N6-10	99	TOT	1
IT108	APIN	R	27	N1	99	TOT	5
IT108	APIN	S	27	N1	99	TOT	7

Fonte: Istat, Software *Cubus*

9. L'ARCHITETTURA PER L'AUTENTICAZIONE DEGLI UTENTI¹

Per facilitare l'uso dei servizi informatici dell'Istat da parte degli utenti esterni, l'Istituto ha adottato una modalità di autenticazione a servizi web basata sull'identificazione personale dei soggetti coinvolti. Nella maggior parte dei casi, le utenze vengono quindi assegnate a singole persone, alle quali vengono attribuite anche le relative autorizzazioni alle risorse informatiche.

Tale modalità di autenticazione presenta diversi vantaggi sia per l'Istat sia per gli utenti. I vantaggi per l'Istat sono di identificare e di fornire le opportune autorizzazioni agli individui che accedono ai servizi informatici; il vantaggio per gli utenti è di accedere con un unico identificativo a tutti i servizi dell'Istituto. Inoltre, poiché l'Istat utilizza un sistema di *single sign-on* per le applicazioni web, l'utente non deve digitare più volte la *password* se si autentica su più applicazioni dell'Istituto nell'ambito della stessa sessione di lavoro.

Per accedere alle risorse informatiche dell'Istat, l'utente viene identificato mediante il proprio codice fiscale. Numerosi sono gli aspetti positivi di questa modalità: ogni utente è già in possesso del codice fiscale, che è univoco ed è rilasciato da un'autorità sicura; l'utente probabilmente già lo utilizza per identificarsi in altri contesti e quindi non ha difficoltà a ricordarlo. Inoltre il codice fiscale è uno degli attributi obbligatori utilizzati nel sistema di autenticazione Spid, grazie al quale è stato possibile fare convivere il sistema di autenticazione interno con Spid, dando la possibilità all'utente di scegliere in fase di *login* le credenziali rilasciate dall'Istat oppure la propria identità digitale Spid.

9.1 Sistema di *single sign-on* Shibboleth

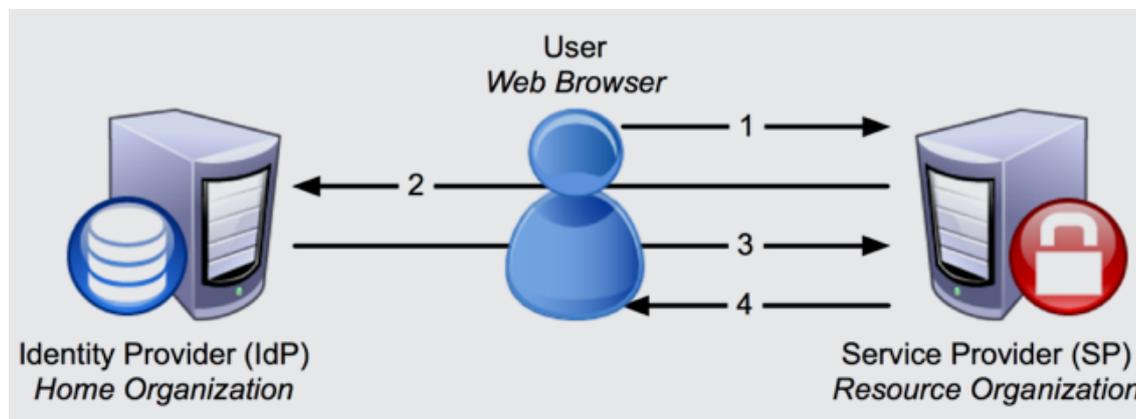
Il sistema di autenticazione adottato dall'Istat è basato su *Shibboleth*, un progetto del gruppo Internet 2 che ha sviluppato una soluzione *open source* per l'accesso a risorse e servizi web condivisi tramite credenziali di autorizzazione. Il sistema *Shibboleth*, basato sul protocollo SAML (*Security Assertion Markup Language*), è comprensivo di almeno due componenti che separano la gestione delle identità degli utenti dalla gestione delle risorse autenticate:

- *Identity provider* (IdP): un servizio che si interfaccia a uno o più *repository* delle identità (LDAP, *Active Directory* o altro) e si occupa della gestione delle sessioni di autenticazione, della richiesta delle credenziali oppure della verifica che l'utente abbia già una sessione valida in *single sign-on* e del rilascio degli attributi degli utenti autenticati;
- *Service provider* (SP): il servizio web accessibile agli utenti che richiede l'autenticazione indirizzando l'utente al suo IdP. Oltre all'esito dell'autenticazione, il *Service provider* può richiedere all'IdP gli attributi utente necessari per gestire la fase di autorizzazione.

La Figura 9.1 mostra le fasi dell'autenticazione: quando l'utente richiede l'accesso a una risorsa protetta da *Shibboleth* (*Service provider*) viene reindirizzato alla pagina di *login* dell'IdP, dove immette le credenziali di autenticazione. Se l'autenticazione va a buon fine, l'IdP invia al *Service provider* informazioni attendibili sull'utente sotto forma di attributi. Con queste informazioni il *Service provider* può prendere decisioni di controllo, di accesso e di autorizzazione.

¹ Il Capitolo è stato redatto da Andrea Credendino.

Figura 9.1 - Fasi di autenticazione



Fonte: Istat

Nel caso dell'Istat, l'IdP si interfaccia con due *repository* di identità per gestire, nell'ambito dello stesso sistema di autenticazione, sia gli utenti interni sia quelli esterni:

- il server *Active Directory* per gli utenti interni;
- il server *OpenLDAP* degli utenti esterni.

9.2 Piattaforma *Identity provider single sign-on* e integrazione con Spid

Un *Identity provider* (IdP) è un gestore delle identità digitali che offre l'autenticazione dell'utente, detto *principal*, come servizio alle applicazioni a esso federate, *relying party*. Le applicazioni che ne usufruiscono esternalizzano quindi la fase di autenticazione dell'utente demandandola completamente all'IdP di fiducia, il quale ha l'ulteriore compito di gestire il *single sign-on* per consentire l'accesso dell'utente ad altre risorse federate senza dovere reinserire le credenziali.

Spesso un *Identity provider* assume il ruolo di *Attribute provider* cioè fornisce gli attributi utente alle applicazioni che ne fanno richiesta. Alle applicazioni è demandata invece la gestione delle autorizzazioni all'accesso che avviene attraverso la valutazione degli attributi rilasciati. La comunicazione tra *Identity provider* e *relying party* avviene, per le applicazioni web, tramite il protocollo SAML. Quest'ultimo è basato sulla produzione di asserzioni di autenticazione o di attributi da parte dell'IdP, che quindi deve svolgere le tre seguenti fasi essenziali del protocollo: ricevere e validare una richiesta di autenticazione SAML da un *relying party* tramite un browser web, autenticare l'utente e infine rispondere al *relying party* con un'asserzione di autenticazione SAML per il *principal* e un'asserzione SAML per gli attributi del *principal*. In questo contesto, un *relying party* in grado di ricevere ed elaborare un'asserzione SAML di autenticazione o di attributi si definisce *Service provider* SAML.

Le funzionalità di *Identity provider*, *Attribute provider* e *Service provider* SAML in Istat sono state realizzate attraverso la piattaforma basata su software *Shibboleth*. Il sistema di autenticazione è stato progettato per integrare l'identità digitale Spid, la Carta di identità elettronica (Cie), l'autenticazione degli utenti interni e l'autenticazione con le credenziali del sistema utenti. Quest'ultimo rappresenta il sistema centralizzato di gestione delle identità digitali dell'Istat che permette l'accesso ai soggetti esterni che utilizzano i servizi informatici dell'Istat.

Di seguito viene riportata la descrizione dettagliata del flusso di autenticazione sul sistema *identity* per le applicazioni del Censimento:

1. l'utente, tramite browser, richiede l'accesso a una delle risorse web del Censimento;
2. il *Service provider* SAML, che protegge la risorsa web, intercetta la richiesta di accesso e invia un messaggio SAML, *AuthRequest*, all'*Identity provider* SAML su cui è abilitato, *identity*;
3. se *AuthRequest* viene validato, l'IdP esamina il tipo di autenticazione richiesta che può essere locale o Spid/Cie;
4. nel caso di autenticazione locale, *identity* assume il ruolo di *Identity provider* operando l'autenticazione utente attraverso i connettori LDAP sui *repository* di autenticazione per gli utenti interni ed esterni, e assume anche il ruolo di *Attribute provider* rilasciando tutti gli attributi ricavati mediante i connettori LDAP;
5. nel caso di autenticazione Spid o Cie la richiesta viene inoltrata all'*Identity provider* Spid su cui l'utente ha attivato la sua identità digitale, Poste/TIM/Sielte, eccetera, oppure all'*Identity provider* Cie;
6. l'IdP remoto autentica l'utente e invia un'asserzione SAML di autenticazione e di attributi al sistema *identity*;
7. l'IdP dell'Istat esegue i meccanismi di riconciliazione dell'utente appositamente configurati per arricchire gli attributi rilasciati dall'IdP remoto con quelli eventualmente presenti nei *repository* LDAP, qualora l'utenza sia già presente su tali sistemi. In tale modo l'IdP dell'Istat esercita la funzione di *Attribute provider* anche in caso di autenticazione Spid o Cie;
8. l'IdP dell'Istat risponde al *Service provider* applicativo attraverso un messaggio SAML di *AuthResponse*, sia se si tratta di autenticazione locale sia nel caso di Spid o Cie;
9. Il *Service provider* applicativo, valida la *response* e, in caso sia conforme alle regole di sicurezza che implementa, autorizza l'utente ad accedere alla risorsa.

Il sistema, una volta verificata la correttezza delle credenziali di autenticazione, può rilasciare al *Service provider* un *set* di attributi utente, che differiscono a seconda del tipo di autenticazione (credenziali dell'Istat oppure con Spid o Cie) e altri attributi predefiniti o personalizzati. Nel caso di autenticazione con le credenziali dell'Istat, gli attributi rilasciati sono i seguenti:

- *uid*: identificativo utente;
- *cn*: nome e cognome;
- *sn*: cognome;
- *givenName*: nome;
- *mail*: indirizzo di posta elettronica;
- *memberOf*: elenco dei gruppi di appartenenza;
- *pwdChangedTime*: data dell'ultima modifica della *password* (solo per utenti esterni);
- *whenChanged*: data dell'ultima modifica della *password* (solo per utenti interni);
- *employeeID*: matricola (solo per utenti interni);
- *department*: dipartimento (solo per utenti interni);
- *division*: direzione (solo per utenti interni);

- *physicalDeliveryOfficeName*: servizio (solo per utenti interni);
- *title*: livello (solo per utenti interni).

L'attributo *memberOf* viene normalmente utilizzato per la profilazione degli utenti. Nel caso di autenticazione con Spid, il sistema può rilasciare gli attributi che ha ricevuto dagli *Identity provider* Spid, in particolare i seguenti:

- *name*: nome;
- *familyName*: cognome;
- *spidCode*: codice identificativo univoco assegnato dal gestore di identità Spid;
- *fiscalNumber*: codice fiscale in formato europeo;
- *email*: indirizzo di posta elettronica.

Nel caso in cui l'autenticazione venga effettuata da parte di un utente che utilizza Spid ed è anche presente sul sistema utenti, vengono valorizzati sia gli attributi Spid sia gli attributi del sistema utenti.

Gli attributi *mail* ed *email* possono avere valori diversi: il primo è l'indirizzo di posta elettronica registrato sul sistema utenti e rappresenta l'indirizzo che l'Istat utilizza per le comunicazioni con l'utente; il secondo è l'indirizzo che l'utente ha scelto di indicare al momento dell'attivazione dell'identità Spid.

Nel caso di autenticazione con Cie, il sistema può rilasciare gli attributi che ha ricevuto dall'*Identity provider* Cie, e in particolare i seguenti:

- *name*: nome;
- *familyName*: cognome;
- *fiscalNumber*: codice fiscale in formato europeo;
- *dateOfBirth*: data di nascita.

Gli obiettivi raggiunti attraverso l'implementazione del sistema *identity* sono:

- realizzazione di un unico sistema di autenticazione per le applicazioni web;
- gestione centralizzata delle informazioni utente;
- accesso utente alle risorse web consentito anche se la risorsa non è gestita dall'organizzazione;
- controllo sul rilascio delle informazioni utente;
- le credenziali utente sono uniche per tutte le risorse federate;
- gli aggiornamenti delle informazioni utente si propagano a tutte le applicazioni federate;
- le applicazioni web non devono realizzare meccanismi di autenticazione, ma solo di controllo delle autorizzazioni di accesso dell'utente alla risorsa.

10. LA GESTIONE DEI TABLET¹

A partire dall'anno 2018, nel quale hanno avuto avvio il Censimento permanente della Popolazione e delle abitazioni, l'Istat, in coerenza con gli obiettivi di dematerializzazione e digitalizzazione delle indagini censuarie, si è dotato di una flotta di tablet da assegnare ai rilevatori. In particolare, nell'anno dell'avvio del Censimento, ne ha acquisiti in noleggio 14.470, modello Samsung Galaxy Tab A10.1² con sistema operativo Android, corredati da 13.773 SIM assegnate ai rilevatori comunali.

Successivamente, in relazione allo svolgimento delle varie edizioni di indagine e alle variate esigenze della rilevazione, il numero dei tablet ha subito diverse variazioni fino a raggiungere circa 20 mila unità nel 2021. I tablet, inoltre, sono stati soggetti a un rinnovo tecnologico che ha portato alla sostituzione dell'intero parco macchine con modelli più recenti e con versioni aggiornate del sistema operativo.

Allo stato attuale l'Istat dispone di circa 13.500 tablet dotati di relative SIM. Un numero così elevato di dispositivi distribuiti sull'intero territorio nazionale ha richiesto una logistica rigorosa e la ricerca di soluzioni originali che facessero uso di tecnologie innovative al fine di assicurarne un utilizzo corretto e in piena sicurezza.

10.1 Gestione logistica dei dispositivi

Dal punto di vista logistico la movimentazione annuale di un numero consistente di tablet, legato alla variazione dei comuni che partecipano alla rilevazione e all'avvicinarsi delle diverse edizioni, richiede una accurata programmazione al fine di rispettare le tempistiche di indagine e un attento controllo delle assegnazioni dei dispositivi stessi per evitarne la dispersione.

In particolare, i tablet sono distribuiti ai comuni che effettuano la rilevazione in base alle numeriche definite dalla Direzione della raccolta dati. Per i Censimenti permanenti della Popolazione e delle abitazioni, i tablet sono assegnati in parte ai comuni Auto Rappresentativi (AR), in media circa il 43,5 per cento, e in parte ai comuni Non Auto Rappresentativi (NAR), che costituiscono in media rispettivamente il 43,5 per cento e il 56,5 per cento del totale dei comuni che partecipano alla rilevazione. La lista dei NAR varia annualmente e pertanto ciascun anno è necessario ritirare i tablet a essi assegnati l'anno precedente per consegnarli a quelli dell'anno in corso.

Tale operazione viene effettuata nella finestra temporale di inattività che si verifica nei primi mesi dell'anno, prima dell'avvio delle operazioni legate al Censimento. Per garantire maggiore sicurezza e limitare gli eventuali furti e smarrimenti degli apparati, le operazioni di ritiro e di consegna vengono effettuate richiedendo l'autenticazione del personale selezionato dai comuni a prendere in consegna tali dispositivi.

Inoltre, si rileva che, per garantire la sicurezza dei dati, al termine di ogni indagine e

¹ Il Capitolo è stato redatto da: Alessandro Ghizzoni (paragrafo 10.4), Erika Lucarelli (paragrafo 10.5), Giuliano Gialli (paragrafo 10.6), Paolo Rondinone (paragrafo 10.3) e Rosa Elia (paragrafi 10, 10.1, 10.2 e 10.7).

² Una parte dei tablet è stata utilizzata come scorta per ottimizzare i tempi di sostituzione dei dispositivi in caso di guasto.

prima della movimentazione dei tablet viene inviato un comando di cancellazione di tutti i dati in essi residenti.

Inoltre, in fase di movimentazione, i tablet sono configurati in maniera da essere inutilizzabili da parte di un eventuale malintenzionato che se ne dovesse appropriare. Prima della movimentazione dei dispositivi viene inoltre effettuata l'operazione di *staging* che consiste:

- nell'assemblaggio del kit composto di tablet e SIM in caso di nuova fornitura dei dispositivi;
- in una revisione dei kit dei tablet e dei loro accessori in caso di movimentazione dei dispositivi da un comune all'altro.

Quest'ultima operazione contribuisce a migliorare notevolmente la qualità dei servizi offerti agli utenti finali in quanto, dopo l'utilizzo prolungato, i tablet spesso vanno incontro a danneggiamenti o alla perdita di accessori, come per esempio la perdita degli alimentatori non sempre segnalata dai comuni in fase di restituzione del dispositivo.

In mancanza di una ricognizione approfondita, i tablet potrebbero essere inviati al nuovo comune malfunzionanti o mancanti dei necessari accessori. In fase di avvio di indagine, ciò può comportare rallentamenti a causa del mancato arruolamento del tablet sui sistemi centralizzati, in attesa della sostituzione delle parti mancanti o malfunzionanti. Inoltre, le segnalazioni di malfunzionamenti sono un ulteriore carico di lavoro per il *service desk* nella fase iniziale della rilevazione, già critica a causa dell'assistenza agli utenti.

10.2 Gestione tecnologica dei dispositivi

Dal punto di vista della gestione informatica dei tablet è stato necessario creare un impianto che garantisca: rapidità e semplicità di configurazione iniziale; definizione uniforme e propagazione in tempo reale della configurazione e dell'aggiornamento dei dispositivi; flessibilità nella gestione e massimizzazione della sicurezza informatica, anche in ottemperanza di quanto previsto dalle indicazioni contenute nel GDPR - Regolamento (UE) 2016/679.

Pertanto, l'approccio utilizzato per la gestione dei tablet è stato di avvalersi di un'infrastruttura centralizzata con il compito di gestire i dispositivi sia dal punto di vista dell'hardware e delle configurazioni del sistema operativo, sia dal punto di vista delle applicazioni utilizzate. La scelta è ricaduta sulla piattaforma *Samsung Knox* per gestire i primi aspetti e sulla piattaforma *Citrix Xen Mobile* per controllare il secondo.

Entrambe le piattaforme contribuiscono a realizzare un ecosistema all'interno del quale tutti gli aspetti della gestione del dispositivo sono affrontati e governati in modo strutturato e dinamico secondo le esigenze delle varie indagini, sempre mantenendo una configurazione che protegga l'Istat da possibili violazioni di sicurezza sui dati dei rispondenti.

Inoltre, particolare attenzione è stata posta ai necessari aggiornamenti del sistema operativo che sono realizzati, sempre garantendo un controllo centralizzato, attraverso la piattaforma *Samsung E-Fota*. Nei paragrafi successivi sono descritte in modo dettagliato le tecnologie impiegate.

10.3 Set-up and Configuration Manager dei tablet Knox

La piattaforma *On Cloud Knox* di Samsung fornisce tutti i requisiti necessari vista la perfetta integrazione con la flotta di tablet di marca Samsung utilizzata.

Knox permette di utilizzare soluzioni di sicurezza, gestione e produttività ad alto livello grazie all'integrazione software e hardware. Poiché i tablet contengono i dati personali dei rispondenti, l'Istat adotta le necessarie *policy* di sicurezza sia durante la fase di rilevazione sia in caso di smarrimento o di furto. Nello stesso tempo, per garantire un processo agile, i tablet devono essere configurati in modo che siano di facile utilizzo e assicurino un alto livello sicurezza.

Dato l'elevato numero di tablet utilizzati, per le diverse fasi della gestione dei dispositivi è necessario adottare configurazioni standard massive da effettuare con pochi passaggi tramite un'interfaccia web e che al contempo assicurino il controllo costante dei dispositivi. Altrettanto importante è avere la possibilità di operare anche sul singolo tablet in maniera indipendente per interventi specifici.

La gestione tramite piattaforma *Knox* permette anche di demandare, predisponendo le opportune configurazioni, la gestione dell'ambiente di lavoro con soluzioni che non siano *Knox Samsung*. La configurazione e la gestione dei dispositivi, la predisposizione dell'ambiente di lavoro e l'aggiornamento degli stessi è effettuata tramite tre differenti soluzioni presenti nella piattaforma *Knox* e integrate tra loro:

- *Knox Configure*: permette la configurazione e la gestione da remoto;
- *KME*: si occupa della gestione dei parametri dell'ambiente di lavoro;
- *E-Fota*: consente l'aggiornamento sistema operativo e delle *patch* di sicurezza.

Il *set-up* della flotta dei tablet dell'Istat avviene con l'inserimento dei dispositivi nella *console* di gestione, con la collaborazione del rivenditore autorizzato.

Si basa sulla *IMEI*, *International Mobile Equipment Identity*, che rappresenta un numero univoco di identificazione del dispositivo e permette all'Istat di poterli gestire da remoto. In questa fase i tablet sono inutilizzabili per i possessori, i quali devono attendere la configurazione definitiva per lo svolgimento delle operazioni censuarie. Il secondo passaggio prevede la configurazione standard, che è la medesima per tutti i dispositivi.

In tale fase, si configura il dispositivo in modo che, in caso di smarrimento o di furto, ne venga bloccato l'uso e vengano cancellati tutti i dati presenti secondo le *policy* adottate dall'Istat e dalle leggi vigenti. La prima configurazione e quella smarrimento o furto sono gestite tramite *Knox Configure*.

Lo svolgimento del Censimento della Popolazione e delle abitazioni del 2023 con l'utilizzo dei dispositivi mobili ha reso necessario l'adeguamento delle soluzioni tecnologiche e di sicurezza del sistema operativo dei tablet. Questa necessità è stata soddisfatta con la soluzione *E-Fota* integrata nella piattaforma *Knox*, che permette l'aggiornamento del sistema operativo e relativi aspetti di sicurezza. Tale percorso è obbligato, documentato e reso disponibile agli utilizzatori finali del dispositivo.

L'integrazione con *Knox Configure* avviene con una configurazione espressamente creata a tale scopo per facilitare la gestione di eventuali problematiche che potrebbero verificarsi nella fase di aggiornamento tramite apposito *help desk*.

I dispositivi sono configurati per gestire le operazioni censuarie in maniera semplice, sicura e ottimizzata, secondo le seguenti modalità:

- il blocco degli *store Google Play* e *Samsung Store* dal momento che le *app* saranno unicamente distribuite tramite uno *store* privato dell'Istat;
- la navigazione web, tramite il *browser Samsung Internet*, consentita solo verso i siti autorizzati: i siti del dominio *istat.it* e il sito ufficiale del Ministero dell'Interno;
- l'*app Secure Hub* per la Gestione dell'ambiente di lavoro;
- restrizioni hardware e software sul dispositivo secondo *policy*: trasferimento dei dati verso l'esterno tramite USB, *Bluetooth*, *cloud* di terze parti *One Drive*, *Google Drive*, utilizzo client di posta, *account* personali, *screenshot*.

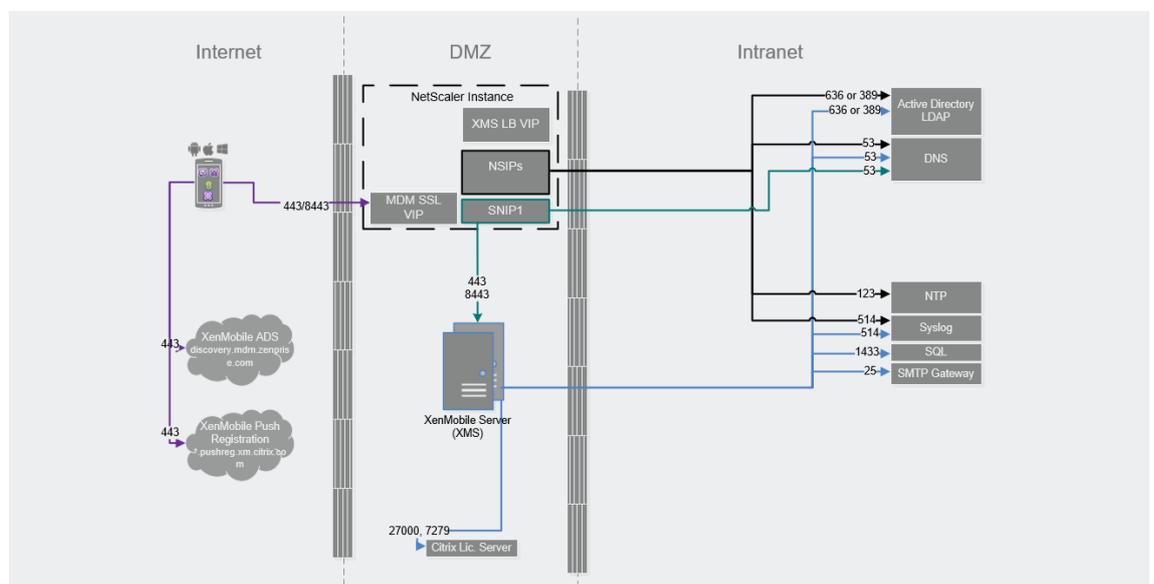
L'ultimo passaggio consiste nella predisposizione dell'ambiente di lavoro e delle sole *app* consentite nelle operazioni di rilevazione. Con la soluzione KME si trasferiscono i parametri e le relative credenziali di accesso in maniera trasparente e facile per gli utilizzatori e si avvia la fase successiva al sistema *Citrix Mobile Management*.

10.4 Gestione della flotta dei tablet e *Private app store* mediante il *Mobile Device Manager*

La sicurezza dei tablet inviati ai comuni per la rilevazione viene garantita dall'utilizzo della piattaforma *XenMobile*. Come già descritto, i tablet sono distribuiti su tutto il territorio nazionale e sono necessari per svolgere le attività di lavoro sul campo: occorre quindi adottare una configurazione tale da garantire che siano utilizzati esclusivamente per i compiti istituzionali.

L'ambiente *XenMobile* è *on-premises* ed è configurato per essere utilizzato sui due Centri elaborazione dati (Ced) dell'Istat, situati nella sede centrale dell'Istat e dell'Istituto Nazionale Assicurazione contro gli Infortuni sul Lavoro (Inail). Sono stati installati due ambienti del tutto analoghi nelle configurazioni: uno di collaudo per le fasi di test e uno di produzione per la messa in esercizio. L'architettura dell'infrastruttura realizzata per la gestione dei dispositivi è la medesima per entrambi i siti Istat e Inail (Figura 10.1).

Figura 10.1 - Architettura per la gestione dei dispositivi



Le *virtual machine* necessarie al corretto funzionamento della soluzione sono state create sull'attuale infrastruttura di virtualizzazione avente l'hardware *Nutanix* con base *VmWare*. Nella predisposizione di questa infrastruttura è fondamentale che:

- i dispositivi siano tenuti in sicurezza il più possibile; l'utente, infatti, non deve potere effettuare nessuna operazione non consentita. Come ulteriore garanzia di protezione, viene utilizzata la funzionalità *Kiosk* disponibile sui dispositivi *Samsung*;
- i dispositivi siano gestiti centralmente e, in caso di necessità, è possibile installare ulteriore software o modificare le *policy* inizialmente previste.

In totale sono gestiti 13.500 tablet con sistema operativo Android, attualmente aggiornato alla versione 11. Dalla versione 10 di Android si è passati da una gestione *legacy* a quella *enterprise*, nella quale *Google* ha svolto un ruolo molto più attivo; in concomitanza con questo passaggio, infatti, si sono dovuti integrare i *cluster* del *Mobile Device Manager* (MDM) con lo *store Google Play*. Tale operazione è stata necessaria sia per l'ambiente di collaudo sia per l'ambiente di produzione; in questa nuova modalità, infatti, *Google* permette all'Istat di creare uno *store* privato sul quale caricare tutte le *app* che devono essere installate sui tablet dei rilevatori.

L'obiettivo di garantire la gestibilità è stato raggiunto riunendo gli utenti in specifici gruppi, in base alla tipologia di rilevazione; questi gruppi poi vengono utilizzati per definire lato piattaforma *MDM Citrix* un *delivery group*, ossia un raggruppamento di *policy* da applicare ai tablet gestiti dall'infrastruttura e un *set* di applicazioni da installare sui tablet in base alla rilevazione. Tutte le operazioni avvengono in modo centralizzato e ogni modifica viene poi distribuita a tutti gli utenti appartenenti allo stesso *delivery group*. Questa pratica viene adottata anche quando si installano nuove applicazioni necessarie al rilevatore.

La più importante tra le varie misure applicate dal *delivery group* per garantire la sicurezza è quella relativa alle *restriction*, una *policy* che impedisce all'utente di utilizzare impostazioni non consentite. L'introduzione di *Android enterprise* ha permesso di impostare delle politiche di restrizione molto più granulari e di effettuare eventuali modifiche centralmente per poi propagarle a tutti gli utenti appartenenti a un dato *delivery group*.

Un altro aspetto da evidenziare sempre nell'ambito della sicurezza è che, con l'introduzione di *Android enterprise*, i tablet hanno una modalità *Kiosk* nativa: in questo modo i rilevatori hanno a disposizione sul tablet solo le applicazioni necessarie alla loro rilevazione.

In caso di smarrimento o di furto del tablet dalla *console* di gestione è necessario bloccare il tablet ed eseguire il comando per riportare il dispositivo alle impostazioni di fabbrica. Inoltre, occorre eseguire una cancellazione di tutti i dati relativi alle applicazioni.

10.5 *Accountability* e associazione dei diversi profili utente ai *device*

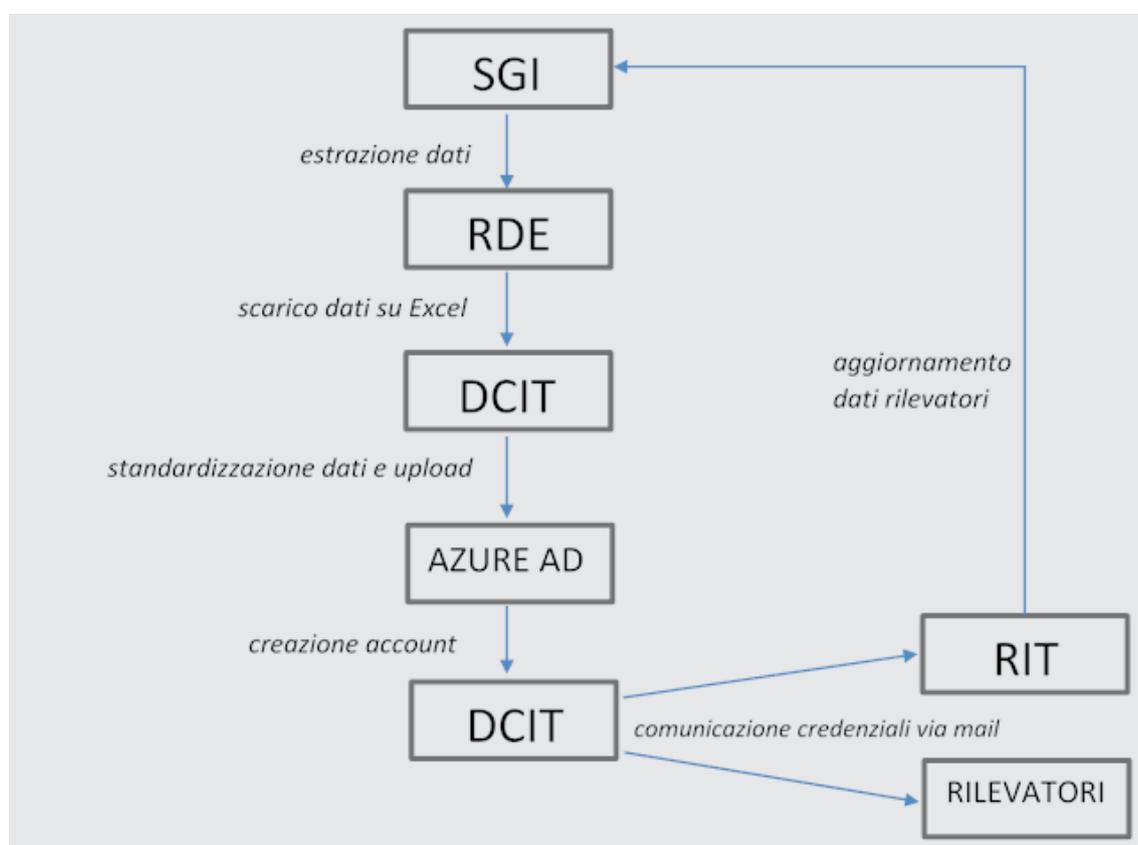
In ottemperanza a quanto previsto dalle indicazioni contenute nel GDPR - Regolamento (UE) 2016/679 in materia di sicurezza informatica si è reso necessario implementare una struttura in grado di rispettare i principi di *accountability*. Agli oltre 13 mila tablet assegnati ai rilevatori sono state associate altrettante utenze di dominio che consentono l'accesso alle applicazioni necessarie alla rilevazione. Attraverso tali utenze, i rilevatori vengono abilitati a livello centralizzato a un *set* specifico di servizi e sui tablet possono essere impostate le *policy* del *delivery group* di appartenenza.

10.6 Gestione delle utenze dei rilevatori sulla piattaforma *Teams*

L'emergenza causata dal Covid-19 di inizio 2020 ha reso necessario individuare, in un lasso temporale molto breve, uno strumento di videoconferenza adatto alla formazione a distanza dei rilevatori. La scelta è ricaduta su *Microsoft Teams*: questo strumento, infatti, ha il vantaggio di essere già stato adottato dall'Istat come strumento di comunicazione interno tra i dipendenti in *smart working*. Sono state quindi acquistate dall'Istat circa 30 mila licenze.

Inizialmente, la gestione di tali *account* è stata impostata grazie a una stretta collaborazione tra il servizio di raccolta dei dati e i servizi interni alla Direzione informatica. Il processo, poiché avviato in fase emergenziale, prevedeva sia operazioni manuali sia automatiche. Nella Figura 10.2 sono evidenziate le fasi del processo e gli attori coinvolti.

Figura 10.2 - Processo di gestione degli *account Teams*



Fonte: Istat

Questa procedura ha da subito mostrato le seguenti debolezze:

- non prevede una cadenza stabilita di lavoro, coinvolge diversi gruppi di lavoro e necessita di operazioni manuali, specialmente nella parte iniziale del processo;
- presenta un elevato rischio di errori nelle operazioni di creazione, modifica e cancellazione degli *account*;
- sono necessarie diverse credenziali per accedere a *Microsoft Teams* e a Sgi;
- presenta problemi legati alla sicurezza in quanto gli utenti esterni entrano nel dominio Istat come membri interni.

10.7 Ottimizzazione dei costi di esercizio

L'impianto tecnologico e organizzativo appena descritto, oltre che sfidante dal punto di vista logistico e gestionale, necessita anche di notevoli investimenti economici. Ai fini dell'ottimizzazione e della massima efficienza possibile, sono state messe in atto una serie di strategie:

- i tablet vengono utilizzati dall'Istat per la conduzione non solo delle indagini censuarie, ma anche di altre che si svolgono in tempi diversi. Il numero dei dispositivi è ottimizzato al fine di garantirne il massimo riutilizzo, movimentandoli opportunamente nel corso dell'anno. In particolare i tablet sono stati utilizzati per: il Censimento della Popolazione e delle abitazioni (anni 2018, 2019, 2021, 2022, 2023); l'Indagine Aspetti della vita quotidiana (AVQ) dal 2019 al 2023; il Censimento generale dell'Agricoltura (anno 2021); il Censimento permanente delle istituzioni non profit (anno 2022); i Censimenti sperimentali della Popolazione (anno 2021); le Rilevazioni sui prezzi dall'anno 2019 in poi;
- negli anni passati si è proceduto al riscatto della prima fornitura di tablet noleggiati. L'acquisto dei dispositivi ha consentito un notevole risparmio economico legato al mancato pagamento del canone di noleggio per il periodo di utilizzo, per una durata superiore a un anno dopo il riscatto, a fronte di un esborso una tantum per il riscatto equivalente a una mensilità per ciascun dispositivo;
- durante il periodo in cui i tablet non sono utilizzati, vengono disabilitate temporaneamente le schede SIM; le stesse vengono successivamente riattivate prima dell'avvio delle rilevazioni;
- la numerosità del personale esterno del *service desk* è definita in base alle esigenze delle indagini: in particolare viene fortemente rafforzato al momento dell'avvio del Censimento della Popolazione e delle abitazioni, che rappresenta l'indagine che ha maggiormente impatto dal punto di vista delle richieste di assistenza, mentre viene ridotto a un presidio minimo durante il resto dell'anno, adeguandone comunque il numero alle esigenze delle rilevazioni.

11. IL SERVICE DESK¹

Contestualmente all'avvio delle indagini censuarie è stato anche predisposto il *service desk* incaricato di fornire assistenza tecnica ai rilevatori per la configurazione, l'utilizzo e la risoluzione dei malfunzionamenti sui tablet loro assegnati. Il *service desk* è contattabile dai rilevatori attraverso un numero verde gratuito e un indirizzo email dedicato. In origine incaricato dell'assistenza tecnica per le problematiche hardware, il servizio si è in breve evoluto per acquisire un ruolo strategico nella conduzione delle rilevazioni, estendendo le sue competenze al supporto agli applicativi di indagine utilizzati dai rilevatori.

Il *service desk* è strutturato in più livelli la cui funzionalità viene riportata di seguito:

- *I livello*. È il punto unico di contatto per i rilevatori dei Censimenti permanenti e delle Indagini continue, e offre assistenza per i problemi tecnici collegati ai tablet, quali malfunzionamenti hardware, software e delle SIM, problemi su applicativi di indagine e informazioni di carattere generale sulla conduzione delle rilevazioni. Raccoglie e traccia tutte le segnalazioni di anomalia o malfunzionamento (*incident*) e provvede ad aprire i corrispondenti ticket sulla piattaforma, dall'apertura alla sua chiusura, di *trouble ticketing*. È responsabile del ciclo di vita del ticket e di assicurare la sua risoluzione in tempi congrui, monitorandone l'avanzamento. Nel caso non sia nella facoltà di risolvere autonomamente l'anomalia, ripristinando la normale operatività dei dispositivi, attiva i livelli di servizio superiori, *service desk* di II livello, inoltrando i ticket in base agli ambiti di competenza. In caso di malfunzionamento hardware provvede ad attivare la società esterna e alla sostituzione del dispositivo guasto attingendo alla "scorta calda"² per minimizzare i tempi di sostituzione. Provvede inoltre a curare i contatti con i comuni per concordare gli appuntamenti sia nella fase di consegna e ritiro degli apparati, sia in caso di sostituzione, sia di movimentazione dei tablet. Il servizio è attivo dal lunedì al venerdì dalle ore 8:30 alle ore 20:30, il sabato dalle 8:30 alle 16:00;
- *Il livello tecnico*. È suddiviso in due gruppi: *service desk* di II livello *Knox* e *service desk* di II livello MDM. Il primo si occupa di intervenire sulla piattaforma *Knox* per effettuare *reset* dei dispositivi o assegnazioni di profili configurativi. Il secondo esegue le operazioni ordinarie di configurazione in coordinamento con il II livello *Knox*. Il servizio è attivo dal lunedì al venerdì in orario di ufficio;
- *Il livello applicativo*. Fornisce assistenza per la risoluzione di anomalie legate agli applicativi di indagine. È a sua volta diviso in vari gruppi di assistenza che si occupano delle problematiche relative a ciascun applicativo. Il servizio è attivo dal lunedì al venerdì in orario di ufficio.

La strutturazione del *service desk* è in continua evoluzione e punta al miglioramento continuo dei servizi erogati alla rete di rilevazione. Fin dalla sua costituzione i servizi sono stati progettati facendo riferimento ai principi dello standard ITIL, *Information Technology Infrastructure Library*. Il flusso delle interazioni tra i vari attori coinvolti è stato strutturato e sono state definite le procedure operative per le varie attività come descritto nei paragrafi successivi.

¹ Il Capitolo è stato redatto da: Rosa Elia (paragrafo 11.2), Rosa Izzo e Valeria Prigiobbe (paragrafi 11, 11.1, 11.1.1, 11.1.2, 11.1.3, 11.1.4).

² La "scorta calda" è un numero di tablet acquistati in eccedenza che rimangono a disposizione del magazzino per rendere più efficienti le sostituzioni dei dispositivi in caso di guasto.

Tali procedure sono state migliorate e consolidate nel corso del tempo, in particolare quelle relative agli incidenti di sicurezza che potrebbero implicare una violazione dei dati personali dei rispondenti. Per migliorare il servizio agli utenti, inoltre, il personale impiegato nel *service desk* frequenta, in corrispondenza all'avvio di ciascuna indagine, una formazione mirata alla conoscenza necessaria a garantire l'assistenza nei vari ambiti. Viene altresì fornita al servizio la documentazione di indagine nella quale è compresa anche una lista di risposte a domande note, che vengono aggiornate nel corso del tempo.

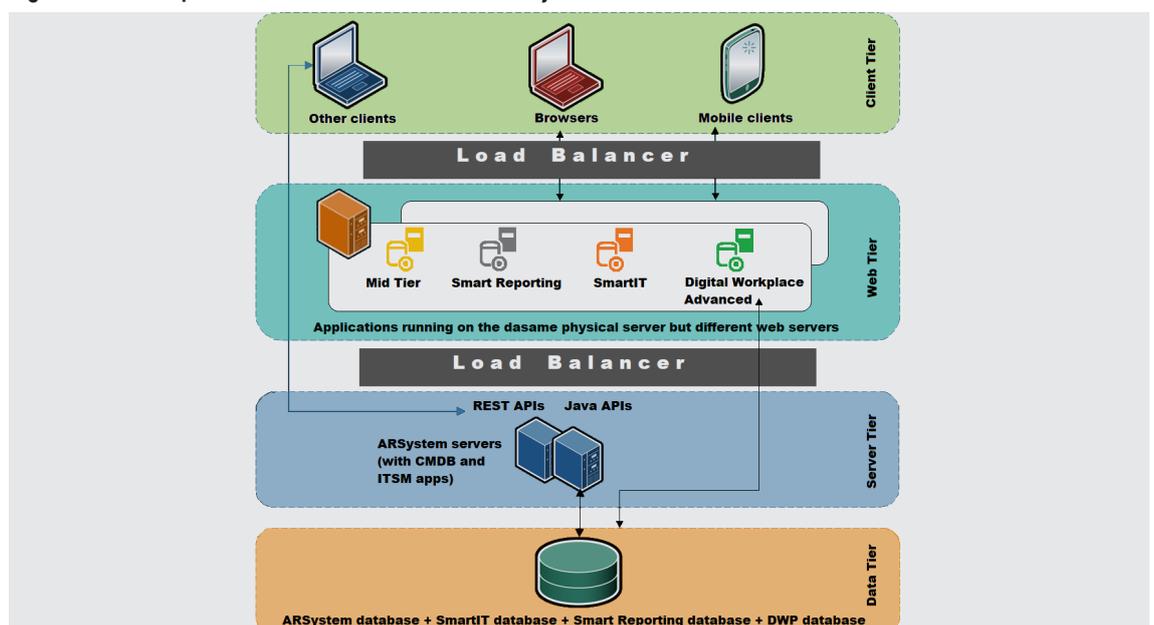
11.1 Piattaforma *IT Service Management (ITSM)* a supporto del *service desk*

La piattaforma *IT Service Management (ITSM)* utilizzata dal *service desk* del Censimento per tracciare e gestire le richieste di supporto da parte della rete di rilevazione è *BMC Remedy Service Management Suite*. Il *service desk* del Censimento è stato organizzato come *Single Point of Contact (SPOC)* e raccoglie le segnalazioni attraverso due canali: un numero verde e un indirizzo email dedicati. Le richieste di supporto vengono successivamente smistate verso i gruppi di supporto specialistici creati ad hoc per ogni singola tematica. La piattaforma ITSM utilizza il flusso del processo di *incident*, opportunamente configurato secondo le *best practice* ITIL per la gestione dei malfunzionamenti segnalati e la loro risoluzione.

11.1.1 Architettura della piattaforma ITSM

Lo strumento per la gestione delle richieste informatiche a supporto delle indagini si compone della parte di *front end*, *MyIT Digital Workplace*, e *back end*, *SmartIT*, a cui è possibile accedere con apposita autorizzazione e opportuna configurazione dell'utente. La Figura 11.1 mostra la relazione tra i componenti all'interno di ciascuno degli ambienti funzionali dell'architettura di distribuzione di *BMC Remedy*, i livelli dell'architettura e il *layout* delle applicazioni in tali livelli.

Figura 11.1 - Componenti dell'architettura *BMC Remedy*



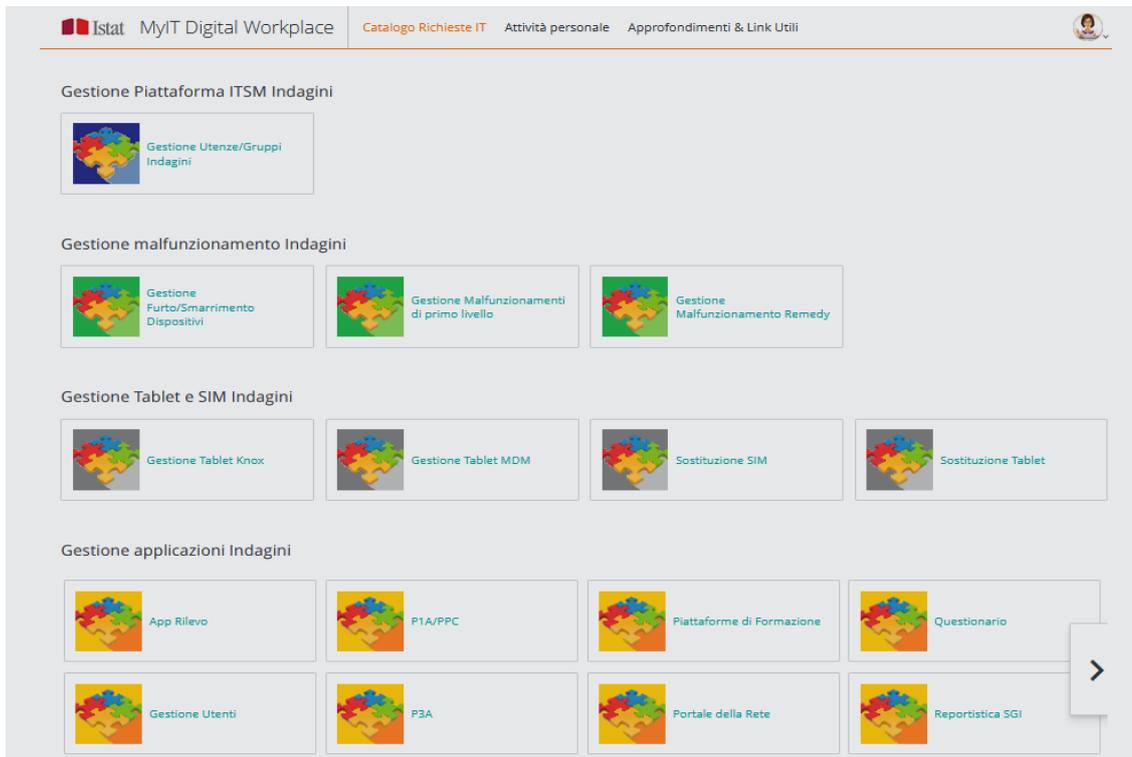
L'architettura *BMC Remedy Service Management Suite* è un'architettura client/server multilivello costituita dal livello Client, dal livello Web, dal livello Server e dal livello Dati.

Il livello Client è costituito da strumenti che permettono l'accesso a un servizio reso disponibile da un server. Gli strumenti di questo livello sono responsabili della presentazione dei servizi e della visualizzazione dei dati agli utenti attraverso varie interfacce: *browser*, PC, *mobile*. Il livello Web è costituito da applicazioni che consentono di accedere al livello Server da un *browser* e rendere accessibili i servizi web. Il livello Server è costituito da applicazioni e flussi di lavoro che applicano la logica dei processi aziendali. Il livello Dati contiene i dati che le applicazioni possono creare, accedere e gestire.

11.1.2 Interfaccia utente: MyIT Digital Workplace

MyIT Digital Workplace è l'interfaccia web di supporto all'operatività degli utenti finali per l'utilizzo del sistema di gestione dei Servizi IT, implementato con la piattaforma *BMC Remedy* a cui si accede tramite la finestra di *login* con la loro utenza di dominio. Il *service desk* dedicato ai Censimenti è abilitato ad accedere alla piattaforma, a utilizzare le richieste dedicate alla gestione del supporto relative all'indagine del Censimento organizzate in un cruscotto.

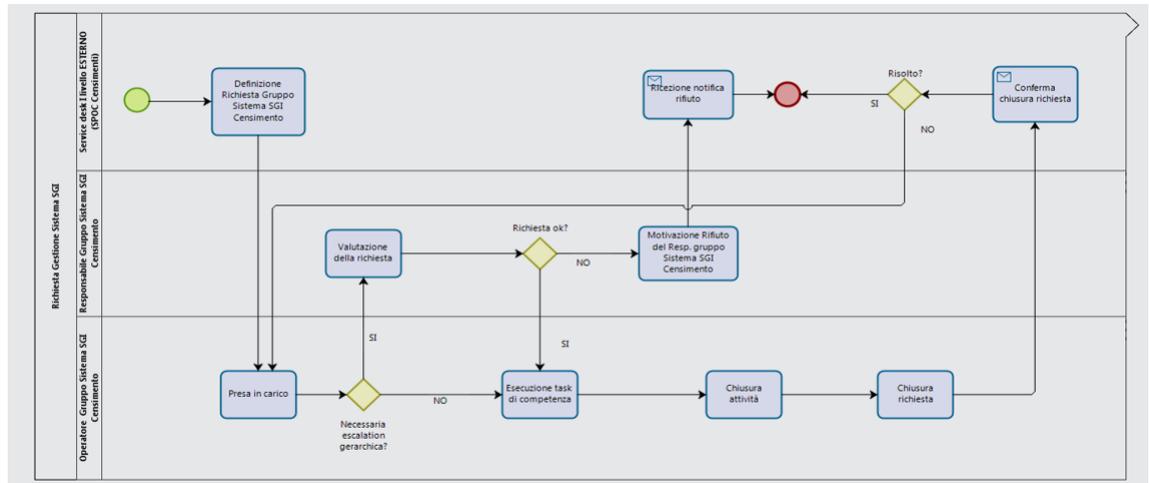
Figura 11.2 - Cruscotto della piattaforma *MyIT Digital Workplace*



Fonte: Piattaforma *MyIT Digital Workplace*

Le richieste implementate sono state definite nell'ambito del Processo di *Incident Management* e, per ognuna di esse, è stata redatta una procedura gestionale in cui sono descritte le attività e i ruoli, gli attori e le responsabilità. Nella Figura 11.3 è mostrato un esempio di flusso con notazione *Business Process Model and Notation* (BPMN) presente in una procedura gestionale.

Figura 11.3 - Esempio di flusso di una procedura gestionale con notazione BPMN



Fonte: Istat

Attraverso queste richieste, opportunamente categorizzate, vengono raccolte le esigenze della rete di rilevazione, ingaggiati i gruppi di supporto competenti e tracciate tutte le attività di supporto svolte per la risoluzione dei malfunzionamenti.

Il cruscotto *MyIT* attuale è frutto di un'evoluzione, avvenuta in questi ultimi anni, con l'obiettivo di migliorare l'esperienza dell'utente e per rispondere all'esigenza di scalabilità e flessibilità dell'applicazione in base alle nuove indagini da gestire.

Questi cambiamenti sono stati di natura funzionale e grafica: l'interfaccia, che con il tempo si è arricchita di nuove richieste per la gestione di diverse tematiche per le indagini, è stata ridisegnata per renderla graficamente più semplice e intuitiva. Dal punto di vista applicativo, è stata sviluppata una nuova funzionalità che permette di configurare facilmente le richieste generiche quando vengono aggiunte nuove tematiche o nuove indagini.

Figura 11.4 - Maschera per la segnalazione dei malfunzionamenti per indagine

Fonte: Piattaforma *MyIT Digital Workplace*

Ciò ha contribuito anche a una maggiore semplificazione per gli utenti sulla piattaforma *MyIT*: tutte le segnalazioni di un malfunzionamento, infatti, hanno un unico punto di accesso rappresentato dalla funzionalità *Gestione malfunzionamento indagini* (Figura 11.4), dalla quale l'utente può selezionare l'ambito impattato dal malfunzionamento.

11.1.3 Interfaccia dell'operatore: *SmartIT*

SmartIT è l'interfaccia web operativa della piattaforma ITSM che permette di gestire il ciclo di vita di una richiesta da parte dei gruppi di supporto IT specialistici, di I e di II livello. Gli operatori abilitati possono accedere all'applicazione tramite una pagina di *login* con la loro utenza di dominio. Tutte le richieste vengono identificate con codici univoci e smistate secondo la tipologia a un gruppo di supporto specialistico.

In particolare a ogni richiesta di supporto del *front end*, identificabile con un nome come REQ0000XXX, viene creato un codice univoco di *back end*, indicato con INC0000xxxx, che contiene tutte le informazioni inserite dall'utente e tutte le informazioni di risoluzione dell'operatore.

I gruppi di supporto possono interfacciarsi trasferendo la richiesta a un altro gruppo di supporto competente.

I ticket seguono vari passaggi di stato fino alla loro completa risoluzione:

- assegnato: il ticket è stato inoltrato al gruppo di supporto preposto per risolverlo;
- in corso: l'operatore ha preso in carico la segnalazione e provvede a eseguire le attività di propria competenza;
- pendente: l'operatore decide di mettere il ticket in sospeso se necessita di informazioni aggiuntive da parte dell'utente;
- risolto: l'operatore, espletate tutte le attività e chiude;
- chiuso: il sistema chiude il ticket automaticamente dopo 15 giorni, impostando lo stato in risolto.

Ogni operatore può visualizzare i ticket personalmente presi in carico e quelli di appartenenti al proprio gruppo di supporto sulla *console* dedicata; inoltre, per rendere più efficiente l'attività di risoluzione, il gruppo di supporto di riferimento riceve una notifica per ogni richiesta aperta.

Gli operatori dei gruppi di supporto possono comunicare con gli utenti richiedenti attraverso la piattaforma, tenendo traccia della conversazione; quando il ticket viene risolto il *service desk* riceve la notifica di completamento delle attività e, a sua volta, può informare il rilevatore che ha segnalato il problema.

Prima di rilasciare la piattaforma, per tutti gli operatori coinvolti è stato organizzato un percorso formativo, articolato in due sessioni: la prima, svolta in aula, ha avuto lo scopo di illustrare la piattaforma; la seconda invece, in remoto, con l'obiettivo di chiarire eventuali dubbi. Ciò ha permesso di implementare il servizio in base alle nuove esigenze così emerse.

Infine, è stato reso disponibile un modulo di *Knowledge Base* per la condivisione e consultazione di documenti e *Faq* utili per potere risolvere in autonomia le segnalazioni pervenute al numero verde o all'indirizzo di posta dedicati.

11.1.4 Dashboard di monitoraggio dei processi di Service Management per le indagini: Smart Reporting e Power BI

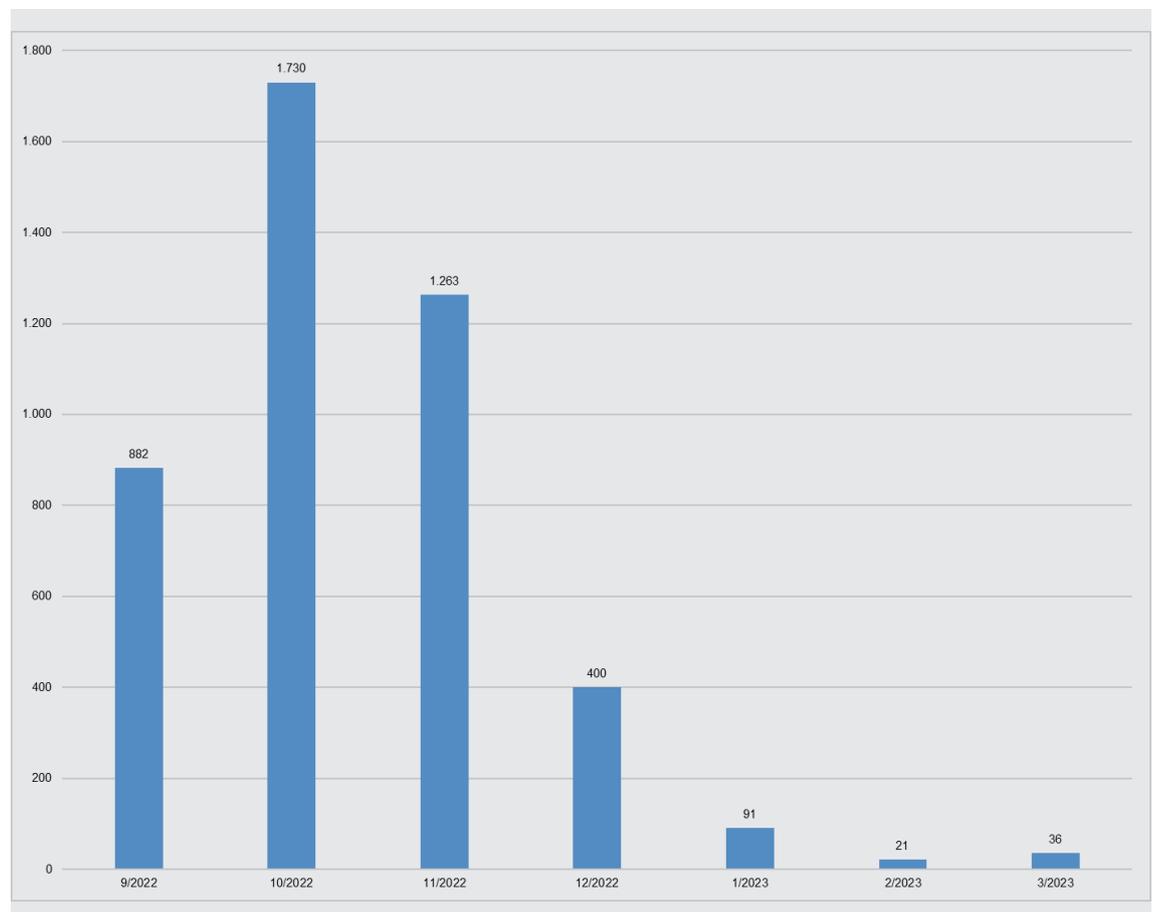
A supporto delle indagini è stato definito e implementato un modello di monitoraggio e controllo dei processi di *Service Management* mediante un sistema integrato di indicatori di processo (*Key Performance Indicator - KPI*), misurati e raccolti sulla piattaforma ITSM. Per ogni KPI sono stati costruiti dei *report* e delle *dashboard* sugli strumenti di *Business Intelligence* dell'Istat: *Smart Reporting* e *Power BI*.

Smart Reporting è lo strumento disponibile nella suite *BMC Remedy* che permette la creazione di *dashboard* e *report* utilizzando i dati presenti nel *CMDB*, *Configuration Management database*. A supporto del monitoraggio delle attività di *service desk* per le indagini sono stati creati dei *report* specifici che attraverso i filtri disponibili (tempo, gruppi di supporto, categoria del ticket) permettono di analizzare diversi aspetti di queste attività, come ad esempio:

- richieste pervenute;
- richieste per indagine;
- distribuzione dei ticket ai gruppi di assistenza;
- *trend* delle richieste.

Nella Figura 11.5 è riportato un esempio di *report*.

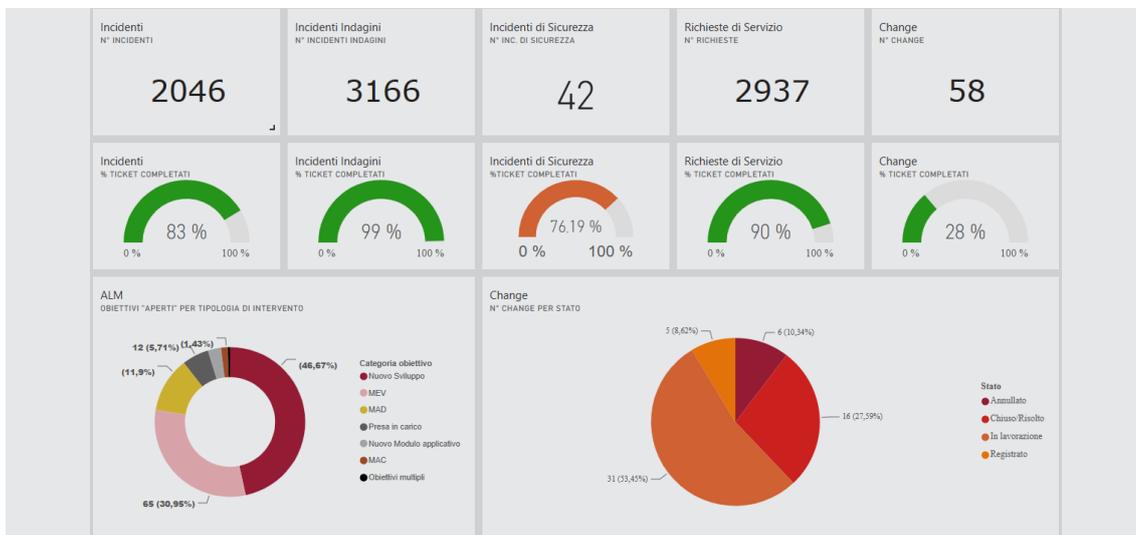
Figura 11.5 - *Trend* mensile dei ticket di malfunzionamento risolti. Settembre 2022 - marzo 2023



Fonte: Piattaforma *Smart Reporting*

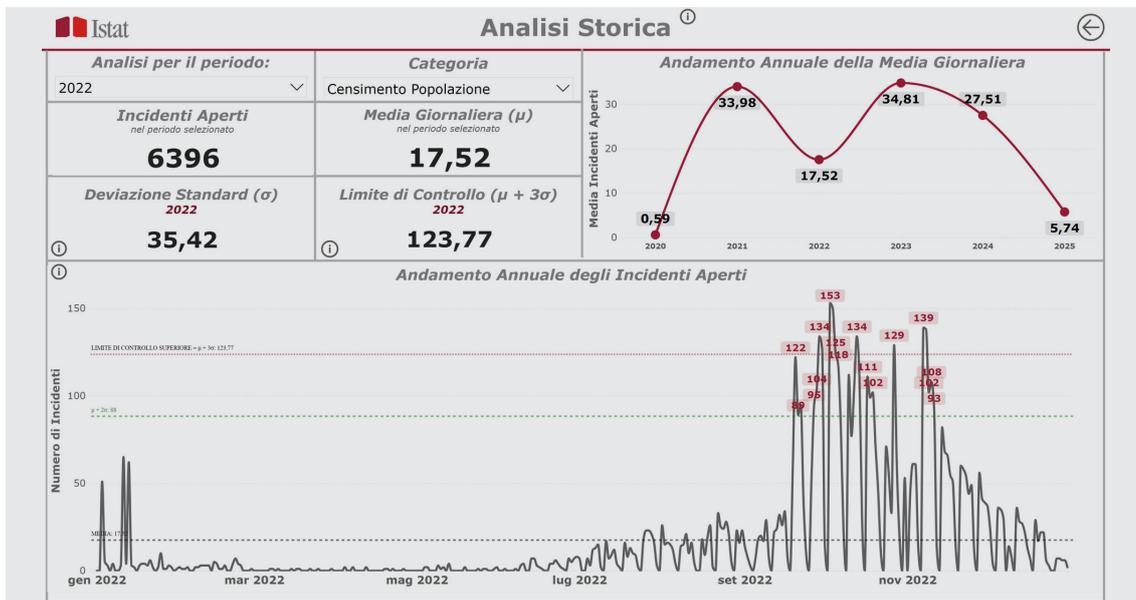
La piattaforma *Power BI*, applicazione appartenente alla *suite Office 365* di *Microsoft*, ospita il *Cruscotto Indicatori DCIT* (Direzione centrale per le tecnologie informatiche) dell'Istat al fine di assicurare una visione olistica dell'andamento delle attività della Direzione e di monitorarne gli indicatori secondo diverse prospettive di analisi, utilizzando i monitoraggi attualmente implementati e concentrando in un unico punto la raccolta e la consultazione dei dati. La *home page* di tale cruscotto rappresenta il punto di accesso principale allo strumento di monitoraggio e offre una panoramica completa dei principali ambiti monitorati e degli indicatori più rappresentativi di ciascuna area. Grazie alla struttura modulare del cruscotto, gli utenti possono accedere alle diverse *dashboard* specifiche (Figura 11.17).

Figura 11.7 - Cruscotto Indicatori DCIT. Monitoraggio dei processi ITSM



Fonte: Piattaforma *Power BI*

Figura 11.8 - Numero di incidenti aperti per il Censimento della Popolazione e delle abitazioni. Anno 2022



Fonte: Piattaforma *Power BI*



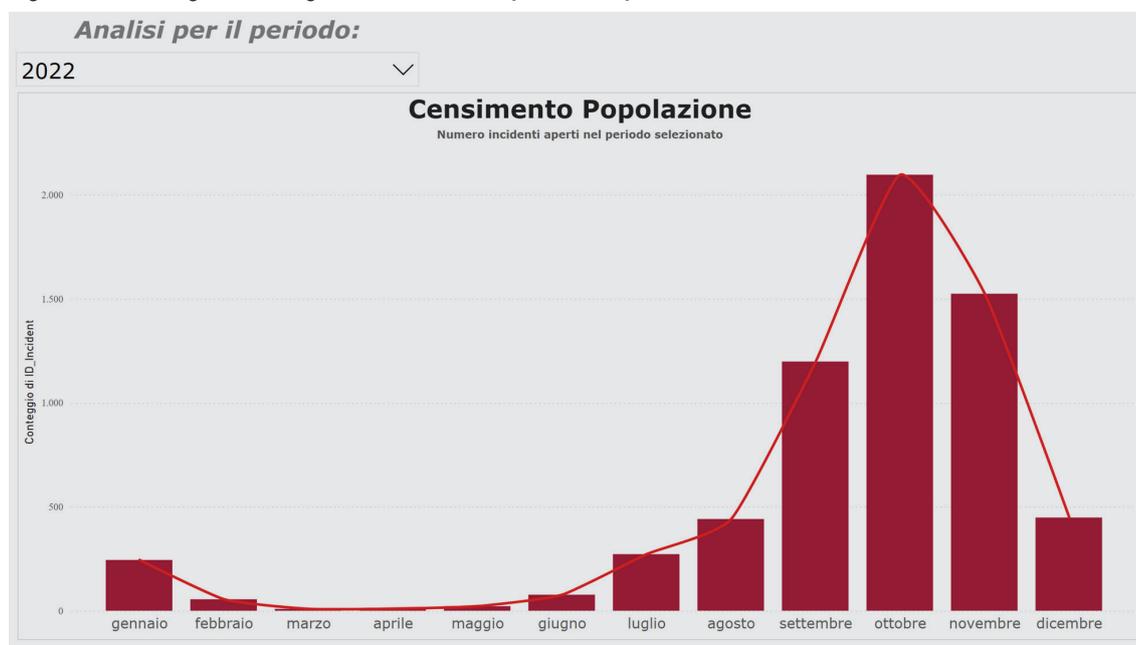
Sul cruscotto è possibile visualizzare, oltre agli indicatori di direzione, gli indicatori riferiti alle richieste di malfunzionamento per le indagini come mostrato nella Figura 11.8.

La Figura 11.8 mostra la *dashboard* del volume di segnalazioni di malfunzionamento, nel caso del Censimento della Popolazione e delle abitazioni del 2022. In particolare viene indicato il tempo di risoluzione medio per ogni richiesta di supporto e il tempo medio di riassegnazione del ticket tra i gruppi di supporto.

Lo strumento *Power BI* permette di espandere, in una specifica visualizzazione, i dati aggregati per mezzo della funzione di *drill down*. Nella stessa visualizzazione, viene presentata un'esplosione del dato, mostrando le informazioni relative ai mesi dell'anno su cui si è cliccato.

La Figura 11.9 mostra il dettaglio, per l'anno 2022, del volume delle richieste pervenute, suddivise per mese e per stato del ticket relativi al Censimento della Popolazione e delle abitazioni.

Figura 11.9 - Dettaglio delle segnalazioni suddivise per mese e per stato. Anno 2022



Fonte: Piattaforma *Power BI*

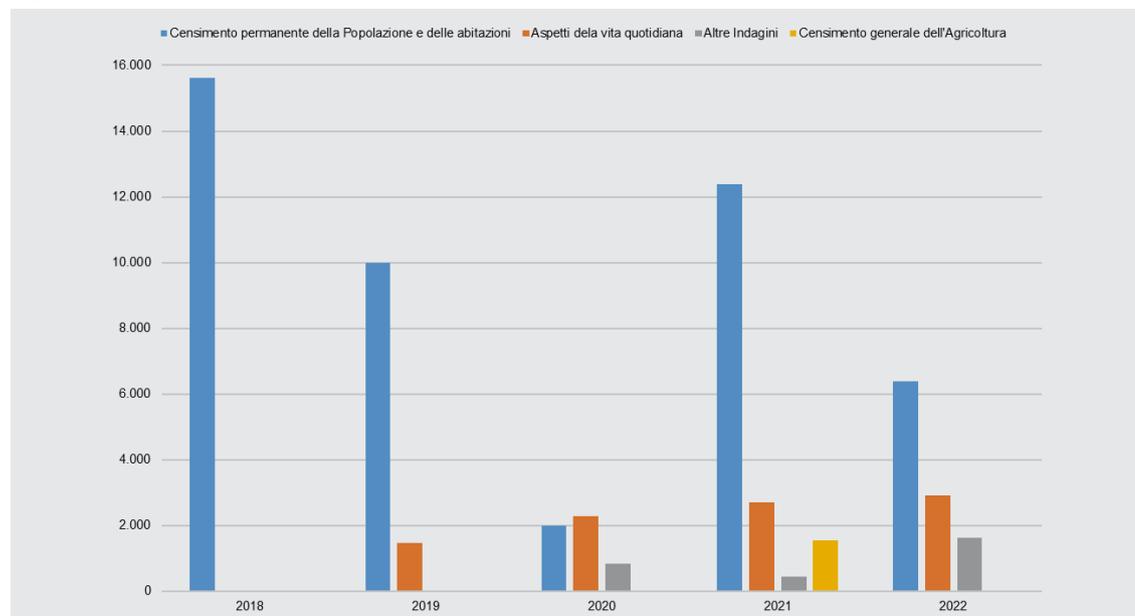
Il modello di monitoraggio delle *performance* dei servizi e dei processi segue il ciclo di Deming, *Plan-Do-Check-Act*, per garantirne il miglioramento continuo. I KPI selezionati delle diverse prospettive vengono costantemente monitorati per verificare l'effettivo raggiungimento del target previsto. La Figura 11.10 riporta il numero complessivo di ticket per anno, evidenziando in colore diverso i ticket relativi a ciascuna indagine.

11.2 Monitoraggio dei servizi

Il servizio del *service desk* ha subito, a partire dal suo avvio, diverse azioni mirate al miglioramento dei livelli di servizio offerti agli utenti. In tale ottica di particolare importanza è l'attività di monitoraggio che viene svolta periodicamente facendo uso degli strumenti di reportistica descritti nel paragrafo precedente. Per comprendere meglio l'andamento delle richieste nel corso degli anni si riportano di seguito alcuni grafici relativi alle segnalazioni, pervenute tramite

ticket, a partire dall'anno 2018 fino all'anno 2022. La Figura 11.10 riporta il numero complessivo di ticket per anno, evidenziando in colore diverso i ticket relativi a ciascuna indagine.

Figura 11.10 - Andamento dei ticket per anno e per indagine. Anni 2018-2022

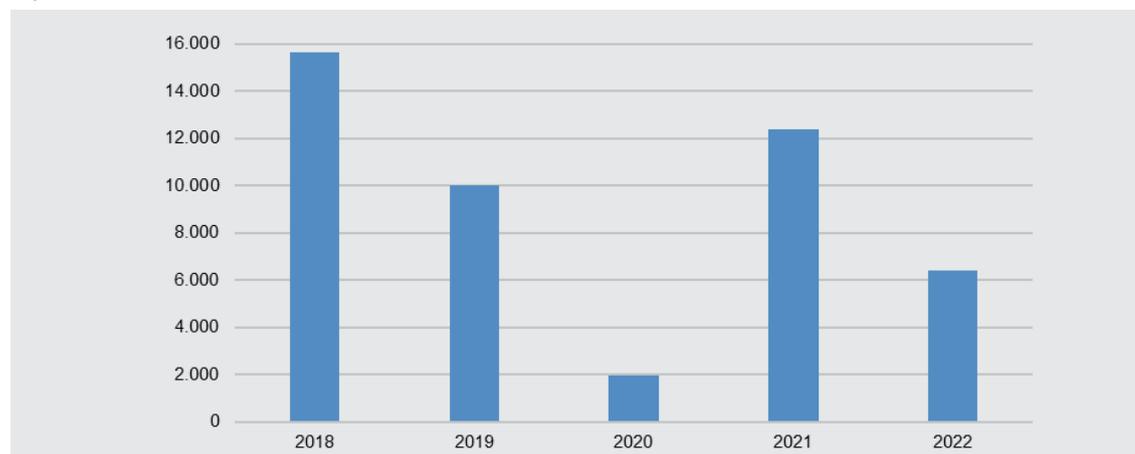


Fonte: Piattaforma Smart Reporting

Esaminando il grafico si evince che il numero maggiore di ticket viene generato dal Censimento permanente della Popolazione e delle abitazioni, anche in virtù del numero di dispositivi utilizzati, e che il numero di ticket è in costante decrescita nel corso degli anni.

Caso a sé rappresenta l'anno 2021 nel quale tale Censimento ha utilizzato un numero di rilevatori, e quindi di dispositivi, nettamente superiore alla media degli anni precedenti e successivi, 16.661 dispositivi a fronte di 13.700 nel 2108, e di un numero inferiore negli anni successivi (Figura 11.11). Tale variazione è giustificata dal fatto che, nell'anno 2020 l'indagine censuaria non è stata svolta a causa della pandemia e pertanto nell'anno 2021 sono stati coinvolti i comuni NAR sia del 2020 sia del 2021.

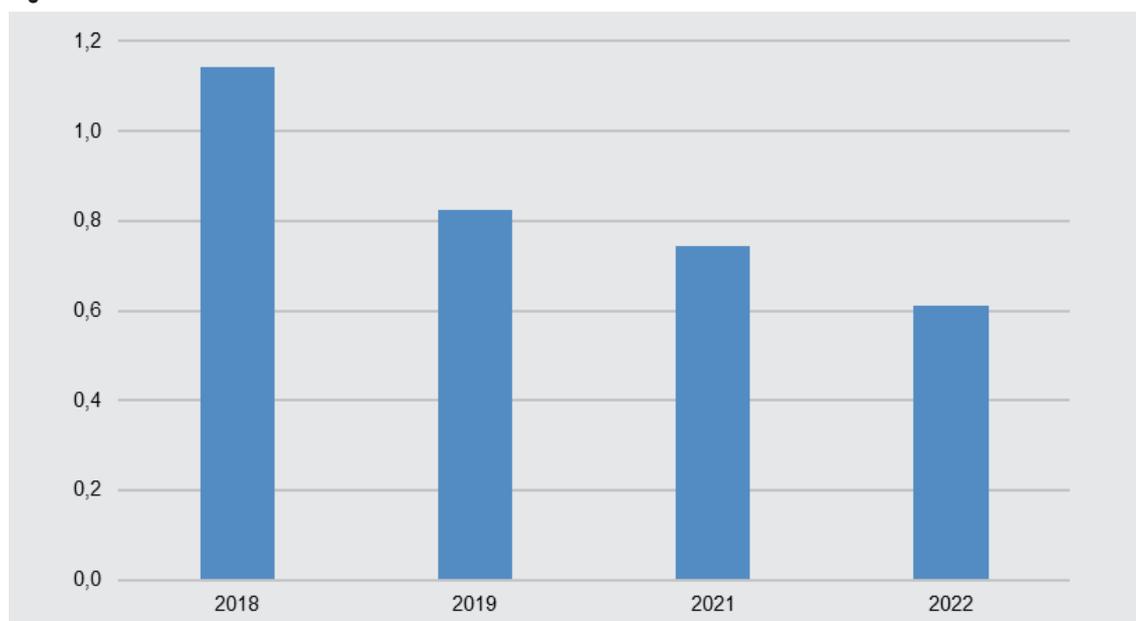
Figura 11.11 - Numero di ticket per il Censimento della Popolazione e delle abitazioni. Anni 2018-2022



Fonte: Piattaforma Smart Reporting

Per evidenziare ancora meglio il miglioramento in termini di efficienza del servizio in esame, la Figura 11.12 mostra, per ciascun anno, il numero medio di ticket in rapporto al numero di tablet utilizzati.

Figura 11.12 - Media dei ticket in relazione ai tablet nei diversi anni. Anni 2018-2022



Fonte: Piattaforma *Smart Reporting*

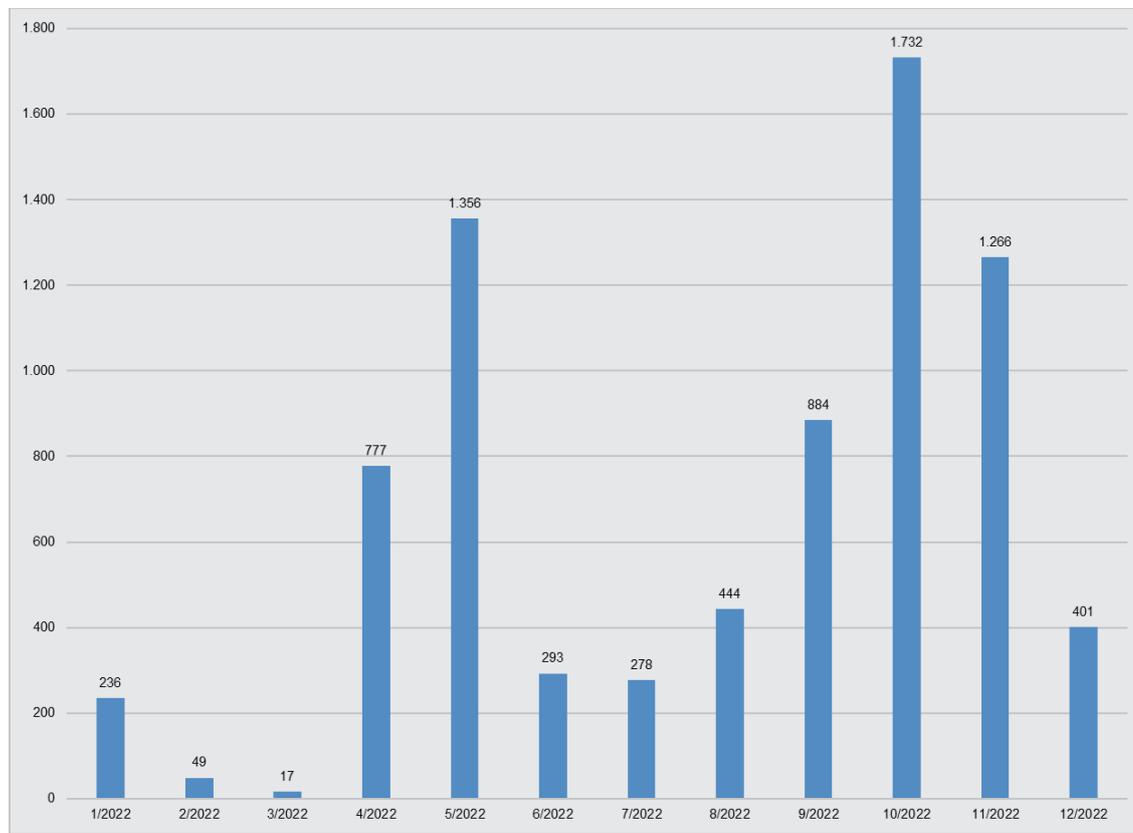
L'incremento dell'efficienza del servizio è in effetti il frutto di diversi fattori che hanno contribuito a diminuire il numero di segnalazioni di *incident* da parte degli utenti finali:

- miglioramento generale della gestione dell'indagine e introduzione di correttivi e miglioramenti funzionali apportati agli applicativi di indagine;
- per i comuni autorappresentativi (AR) nel corso degli anni si è consolidata una maggiore dimestichezza nell'utilizzo degli strumenti informatici messi a disposizione dall'Istat;
- maggiore attenzione nella compilazione e diffusione dei manuali per la configurazione e per l'utilizzo dei tablet, e creazione di *Faq* destinate agli utenti finali.

La principale criticità del servizio risiede in ogni caso nei picchi del numero di segnalazioni che si verificano in fase di avvio delle indagini, in particolar modo del Censimento della Popolazione e delle abitazioni, che ha un numero significativo di rilevatori. Per mitigare questo effetto, oltre all'incremento degli operatori impiegati, nel corso degli anni si è cercato di anticipare l'avvio delle operazioni di arruolamento dei dispositivi sui sistemi centrali da parte dei comuni. Tale operazione, infatti, genera il maggior carico dei ticket in fase di avvio di indagine e risulta particolarmente impegnativa per il *service desk*, poiché spesso fornisce assistenza agli utenti per l'intera fase di esecuzione delle procedure di configurazione del tablet. A titolo di esempio riportiamo la distribuzione su base mensile delle richieste di assistenza ricevute dal *service desk* nel corso dell'anno 2022 (Figura 11.13).

In tale grafico risultano particolarmente evidenti i picchi in corrispondenza dell'avvio dell'indagine AVQ e del Censimento della Popolazione e delle abitazioni. Per quest'ultimo si può osservare che un aumento significativo dei ticket è già presente nel mese di settembre, in virtù delle operazioni propedeutiche condotte prima dell'avvio del Censimento.

Figura 11.13 - Trend incidenti delle indagini su base mensile. Anno 2022



Fonte: Piattaforma Smart Reporting

12. LE ARCHITETTURE *MIDDLEWARE*¹

Per sostenere le piattaforme applicative descritte nei Capitoli precedenti si è reso necessario progettare delle architetture e infrastrutture che potessero essere standard, interoperabili, resilienti, versatili e altamente scalabili, in modo tale da potere costituire un ecosistema di riferimento per tutte le indagini censuarie. Tale approccio ha permesso di gestire efficientemente l'intera filiera applicativa e di processo, che una sfida così complessa impone, orientando al meglio tutte le diverse esigenze.

Il passaggio dal Censimento decennale al Censimento permanente ha spostato l'asticella del *time-to-market* di un fattore moltiplicativo, stressando certamente gli aspetti metodologici, ma riportando il focus verso gli ambiti di carattere tecnologico e palesandone la loro centralità e criticità. Tale sfida ha portato l'Istat a investire in modo massiccio in tecnologie innovative, proiettando l'intera Direzione informatica verso una vera e propria trasformazione digitale in termini di governabilità dei processi e tecnologie e trasformando il processo fino ad arrivare all'azzeramento dell'uso della carta.

Il processo è estremamente articolato e si compone di diverse fasi correlate e di elevata complessità tecnologica, dalla raccolta del dato grezzo alla pubblicazione del dato lavorato, aggregato e opportunamente organizzato, secondo appropriati aspetti metodologici e di diffusione. Le diverse fasi rivestono un'importanza cruciale per la buona riuscita dell'intero processo censuario, partendo dalla gestione della fase informativa, formativa e comunicativa, in termini di gestione della rete di rilevazione; il Portale della rete e la piattaforma per la formazione svolgono in questo contesto un ruolo fondamentale.

Il Sistema di gestione delle indagini, inoltre, consente di gestire l'intera filiera di raccolta dati, integrandosi sinergicamente con il sistema di reportistica.

Per quanto concerne la fase di raccolta dati, l'innovativo portale del questionario elettronico, *Panda*, riesce a rendere estremamente versatile ed efficace la fase di raccolta dei microdati.

Infine, la fase di controllo e correzione dei dati è fondamentale per consolidare i dati nella forma più appropriata; mediante l'uso del *data warehouse* di diffusione, si raggiunge la fase finale del processo con un ritorno per il cittadino sotto forma di diffusione dei dati. Si può dunque scorgere come il grado di difficoltà sia estremamente complesso, visto che l'intera infrastruttura ha, non solo la necessità di erogare un servizio di grande stabilità, resilienza e scalabilità, ma deve integrarsi tra le diverse architetture in maniera completa e le molteplici e differenti tecnologie.

Le varie infrastrutture *middleware* messe in campo dall'Istat devono garantire la *business continuity* sui due *site* e rendere le architetture resilienti alle infrastrutture di base, dispiegando il servizio su due differenti tipologie.

Possiamo dunque affermare che non si tratta di una semplice applicazione web, ma di un vero e proprio ecosistema di architetture tecnologiche in relazione, finalizzate a sostenere l'obiettivo censuario nel suo complesso.

Per garantire una fluida sinergia tra i servizi e le applicazioni, e rendere modulare le diverse componenti software, si sono messi in campo diversi meccanismi di comunicazione tra i sistemi, puntando particolarmente sull'uso dei microservizi, come descritto nel Capitolo 2.

¹ Il Capitolo è stato redatto da: Gianluca D'Amato (paragrafo 12.6), Gianluca Ferri (paragrafi 12.3 e 12.4), Manuela Lo Bosco (paragrafo 12.5), Mario Magarò (paragrafo 12), Francesca Masotti (paragrafo 12.2), Sergio Sortino (paragrafo 12.1).

Questo approccio innovativo, basato su tecnologia *Eureka Service Registry* e *Spring Boot*, ha permesso di ridurre i tempi di sviluppo e manutenzione, disaccoppiando l'implementazione delle tante interfacce, rispetto alle componenti di logica, scindendo così il livello *front end* da quello *back end*, in una strategia *multilayer*.

Eureka Server e *Spring Boot* costituiscono un'architettura che governa le informazioni di tutte le applicazioni del servizio client. Ogni microservizio verrà registrato nel server *Eureka* che conosce tutte le applicazioni client in esecuzione su ciascun *socket*. *Eureka Server* è anche comunemente noto come *Discovery Server*.

Infine, possiamo affermare che non basta innovare inserendo tecnologie all'avanguardia, se non gestite opportunamente da personale qualificato e definendo gli opportuni processi di manutenzione, monitoraggio e un proficuo modello di *deployment* efficace ed efficiente, seguendo quanto stabilito dalla metodologia *DevOps*. Tale approccio metodologico assicura che le nuove modifiche apportate al codice vengano regolarmente compilate, testate e unite in un *repository* condiviso, risolvendo così il problema dei conflitti tra le numerose diramazioni di un'applicazione in fase di sviluppo. *DevOps* si pone l'obiettivo di garantire una distribuzione continua, aumentando la qualità del software e la sicurezza, rendendo sempre più robusto il software e pertanto il sistema in generale.

Ciò premesso, risulta evidente che la tecnologia, volta a garantire un valore aggiunto, ha la necessità di essere affiancata da opportune logiche di gestione e di controllo. Alla base della qualità, stabilità e resilienza delle architetture *middleware*, vi è una sfida sempre più difficile, che è quella di innovare, mantenendo nel tempo il medesimo livello di stabilità e di sicurezza.

12.1 Portale della rete

Il Portale della rete è una piattaforma web basata su TYPO3, scritto in PHP e supportato da un database *MySQL*; il sito viene eseguito su web server Apache che utilizza il protocollo HTTPS (*Hypertext Transfer Protocol Secure*). Quest'ultimo è noto per essere veloce, affidabile e sicuro, oltre che altamente personalizzabile, permettendo di soddisfare le esigenze di molti ambienti diversi grazie all'utilizzo di estensioni e di moduli.

L'autenticazione da parte degli utenti avviene attraverso la tecnologia *Shibboleth*: grazie al *single sign-on*, l'utente effettua il *login* una sola volta e ottiene l'accesso a diverse applicazioni senza la necessità di immettere nuovamente le proprie credenziali.

La protezione del sito da attacchi dannosi e dal traffico Internet indesiderato è affidata al *Web Application Firewall* (WAF), un dispositivo specializzato nell'analisi del traffico a livello applicativo per i servizi web; in particolare, il *firewall* analizza il traffico di rete e, attraverso apposite regole, decide se bloccarlo o meno.

Per garantire agli utenti prestazioni elevate e alta affidabilità sono state adottate soluzioni tecnologiche di ottimizzazione del carico di lavoro e di dislocazione fisica dei server: il *load balancing*, permette a un gruppo di server, *cluster*, di fare fronte ai picchi di traffico, ma anche di prevedere una soluzione di emergenza in caso di guasto; i server, in particolare quelli di produzione, sono stati attestati su due Ced differenti in *business continuity*, situati nella sede centrale dell'Istat e dell'Inail.

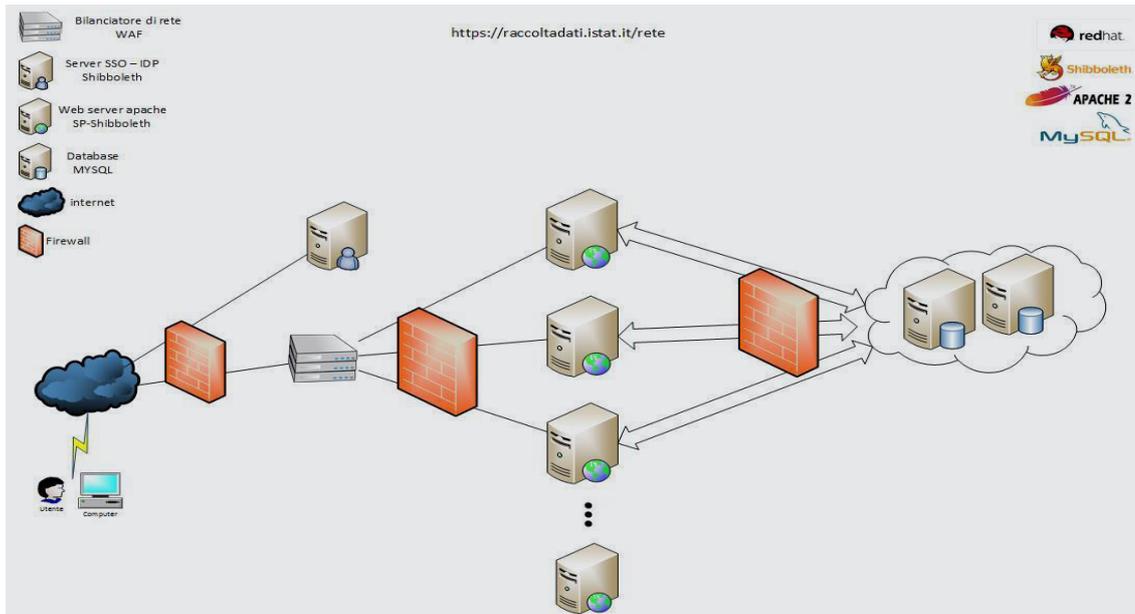
L'architettura è composta da server virtuali gestiti dal software *VMware* e il sistema operativo utilizzato è *Red Hat Enterprise Linux*, che garantisce un'alta flessibilità e affidabilità grazie a costanti e continui aggiornamenti eseguiti tramite il componente aggiuntivo *Red*

Hat Satellite, permettendo così di soddisfare i più recenti requisiti di sicurezza e conformità.

Nella Figura 12.1 viene illustrata l'architettura tecnica del sito della rete.

Secondo gli standard ITIL, l'infrastruttura prevede anche i relativi ambienti di collaudo e sviluppo.

Figura 12.1 - Architettura del sito della rete



Fonte: Istat

12.2 Piattaforma di formazione

Per i Censimenti è stata predisposta una piattaforma di *e-learning*, denominata *Formazione Reti*, che consente la formazione a distanza.

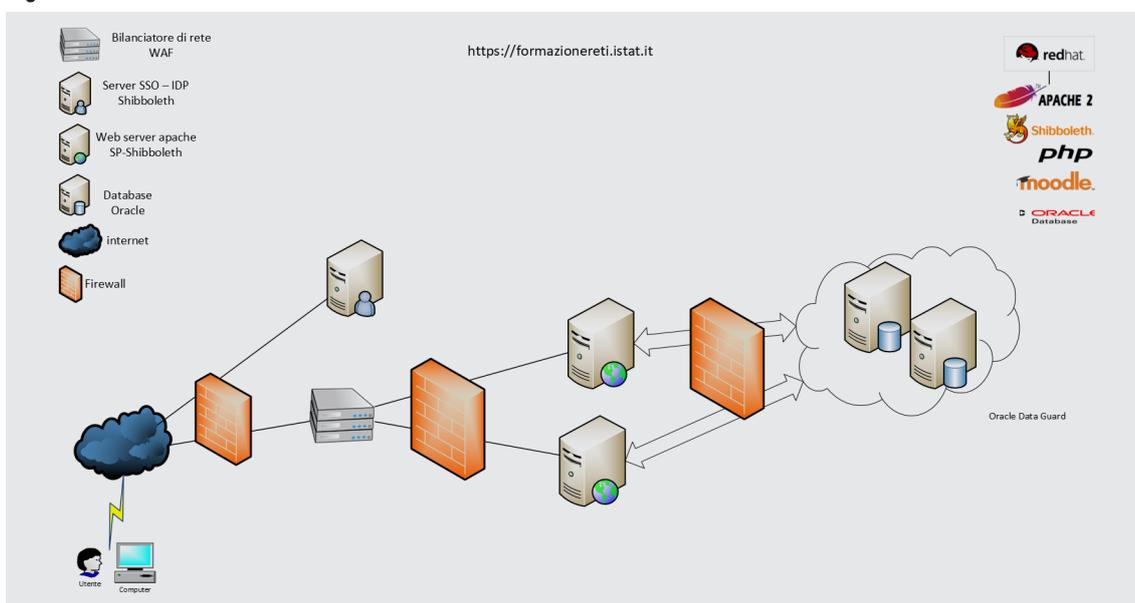
Per soddisfare i bisogni della rete di rilevazione, è necessario dotarsi di un ambiente dinamico per l'apprendimento che, oltre a garantire i requisiti funzionali richiesti, abbia caratteristiche di fruibilità e versatilità,

Le esigenze descritte sono soddisfatte grazie all'utilizzo della piattaforma *Moodle*, un'applicazione web completamente *open source* e scritta interamente in linguaggio PHP, come descritto nel Capitolo a essa dedicato, che supporta la tradizionale formazione in aula e permette di pubblicare e rendere accessibile alla rete di rilevazione il materiale didattico e video. L'infrastruttura pensata per il sistema è composta da un'architettura software *multi-tier* con due server virtuali dislocati sul *Vcenter VMware* ad alta affidabilità su infrastruttura iperconvergente *Nutanix*. Ogni server è dotato di 8 CPU e 16 GB di RAM, con sistema operativo *Red Hat Enterprise Linux* e *file system* condiviso GSF2. Il *file system* ha lo scopo di fornire un *set* di funzionalità il più vicino possibile a un *file system* locale, garantendo allo stesso tempo la piena coerenza del *cluster* tra i nodi: avendo accesso allo *storage* condiviso, si migliorano le prestazioni mediante la memorizzazione nella *cache* locale, dei dati utilizzati più frequentemente. Il sito viene gestito da Apache, il web server *open source*, ampiamente utilizzato, perché offre alte prestazioni, stabilità e sicurezza. Infine supporta una vasta gamma di protocolli, HTTP, HTTPS e FTP, ed è in grado di gestire più richieste simultanee, ideale per siti web complessi.

La protezione di tutta l'infrastruttura è gestita dal WAF, che permette di monitorare il traffico in entrata e in uscita. Il WAF mette a disposizione un puntamento VIP (*Virtual IP*) che espone la URL all'utente finale in modalità sicura. L'utente, una volta collegato attraverso l'autenticazione *Shibboleth*, accede alla piattaforma *e-learning Moodle* la quale rende accessibile il materiale didattico per la formazione. A sua volta la piattaforma, attraverso l'utilizzo di un'utenza applicativa, resa *proxy user* dai DBA *Oracle* e con adeguate *grant* di esercizio fornite dal responsabile, si collega in totale sicurezza al database *Oracle*.

In accordo con lo standard ITIL, l'infrastruttura prevede anche un ambiente di sviluppo e di collaudo. Di seguito un esempio semplificato di architettura tecnica dell'ambiente di produzione.

Figura 12.2 - Architettura del sito della formazione



Fonte: Istat

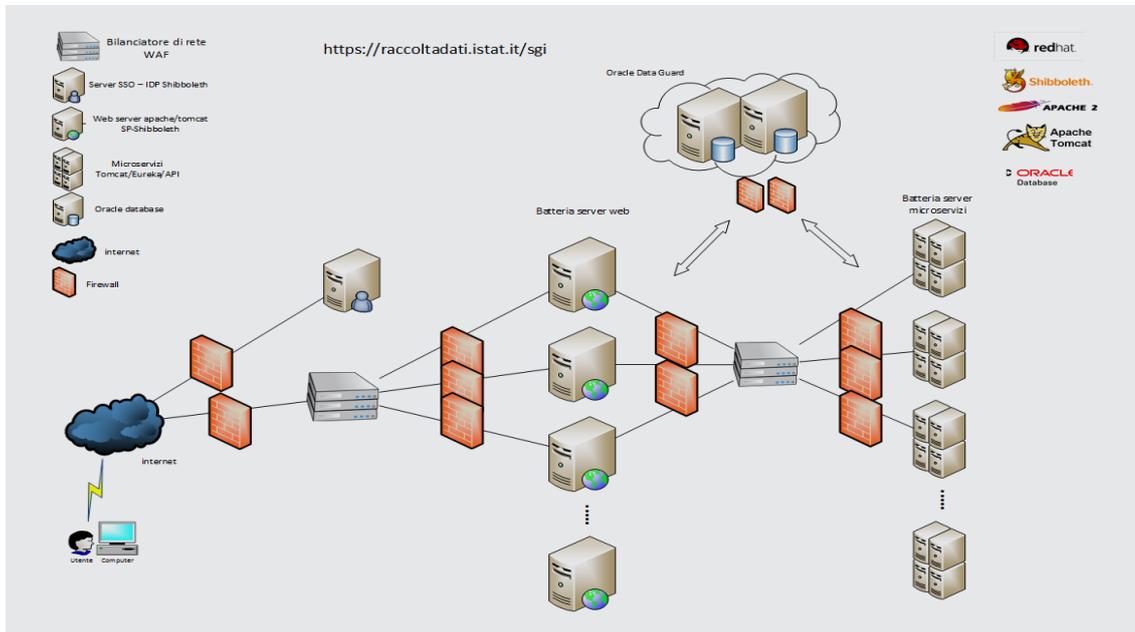
12.3 Piattaforma per la gestione delle indagini

Il Sistema di gestione delle indagini (Sgi) è un'applicazione web distribuita su più livelli: il *front end* (o *client-side*), il *back end* (o *server-side*), il database e i microservizi.

Nello schema logico viene raffigurata l'infrastruttura composta da diverse componenti quali le macchine virtuali, il bilanciatore di rete, il *firewall* e il sistema di autenticazione, che possono essere virtuali o fisiche. Le prime sono allocate su nodi fisici *ESXi VMware (Elastic Sky X Integrated)*, un sistema software che permette di eseguire più sistemi operativi su un singolo server fisico. La pila software dove si appoggia e lavora l'applicativo Sgi è composta dal sistema operativo *Linux*, distribuzione *Red Hat*, dal web server *Apache*, da *Tomcat* utilizzato come contenitore di applicazioni/servizi scritti in *Java*, dal database *Oracle* per i dati, e dal sistema *Shibboleth* per il *single sign-on* che comprende la parte *Service provider* e *Identity provider*.

La Figura 12.3 mostra anche il livello di affidabilità in termini di carico e di sicurezza dell'ambiente, dato rispettivamente dalle batterie dei server web e dei microservizi bilanciati dal VIP e protette dal WAF e *firewall* di rete.

Figura 12.3 - Schema logico infrastruttura di Sgi



Fonte: Istat

La configurazione a regime prevede tutti i server operativi sempre attivi; altri server sono disabilitati ma comunque mantenuti e aggiornati in maniera tale che, per sopperire a una richiesta applicativa improvvisa o per fare fronte a un malfunzionamento di uno dei server attivi, si possa facilmente introdurre nella batteria un'ulteriore macchina semplicemente abilitandola nel VIP. Specularmente a quanto fatto per l'ambiente di produzione, è stato configurato e allestito anche l'ambiente *demo*, utile e necessario alla rete di rilevazione sia per momenti di formazione sia per attività di *training on the job* degli operatori abilitati. Secondo gli standard ITIL, l'infrastruttura prevede anche i relativi ambienti di collaudo e di sviluppo.

12.4 Reportistica di raccolta dati

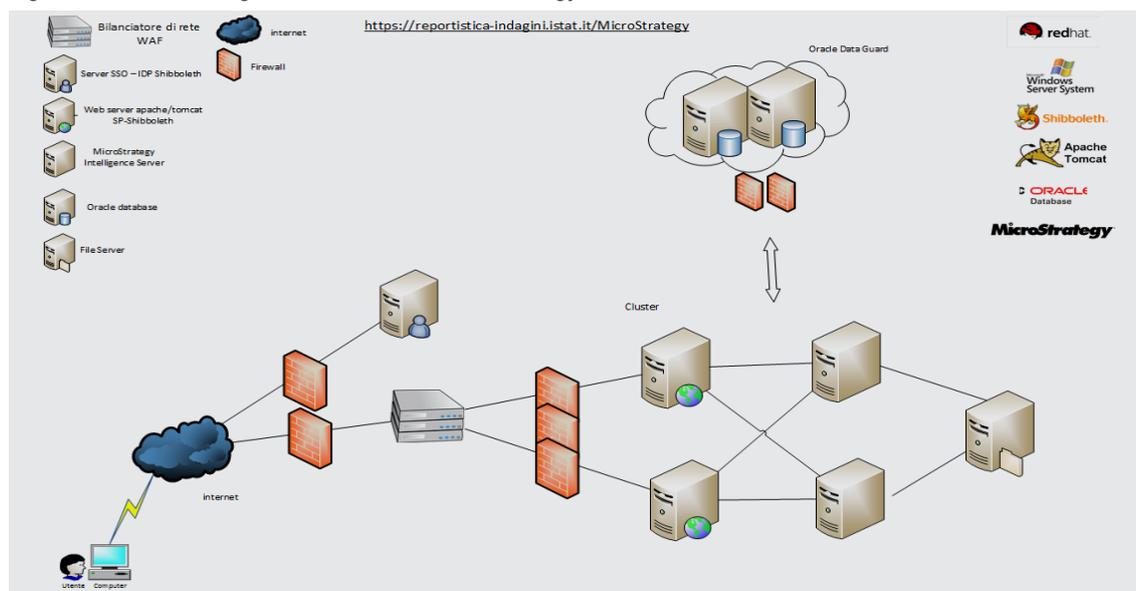
La reportistica di raccolta dati viene prodotta utilizzando il software *Microstrategy*, piattaforma di *Business Intelligence*, che fornisce strumenti avanzati per l'analisi dei dati, il *reporting* e la creazione di *dashboard*.

La Figura 12.4 mostra l'infrastruttura composta da diverse componenti, che possono essere virtuali o fisiche: macchine virtuali, bilanciatore di rete, *firewall* e sistema di autenticazione. Le componenti virtuali sono allocate su nodi fisici *ESXi VMware (Elastic Sky X Integrated)*, che permette di eseguire più sistemi operativi su un singolo server fisico. Il *cluster Microstrategy* è costituito da due web server e due intelligence server e da un file server.

La pila software dove si appoggia e lavora l'applicativo è composta dal sistema operativo *Linux*, distribuzione *Red Hat*, da *Tomcat*, utilizzato come contenitore di applicazioni o servizi scritti in Java sui due web server, da *Windows Server*, dall'applicativo *Microstrategy*, sopra gli intelligence server, dal database *Oracle* per i metadati e i dati, e dal sistema *Shibboleth* per il *single sign-on*, che comprende la parte del *Service provider* e *Identity provider*.

La Figura 12.4 mostra anche il livello di affidabilità in termini di carico e di sicurezza dell'ambiente, dato rispettivamente dal *cluster* di server bilanciato dal VIP, e protetto dal WAF e *firewall* di rete. Secondo gli standard ITIL, l'infrastruttura prevede anche i relativi ambienti di collaudo e sviluppo.

Figura 12.4 - Schema logico infrastruttura di *Microstrategy*



Fonte: Istat

12.5 Questionario di raccolta dati

Come descritto nel Capitolo 1, *Panda* è un sistema generalizzato per la produzione di questionari elettronici per ogni tipologia di censimento che l'Istat propone.

L'architettura infrastrutturale è composta da 8 server virtuali basati su tecnologia *Red Hat Enterprise Linux*, ognuno con 4 CPU e 16 GB di RAM, dislocati su due *Vcenter VMware*, per garantire l'alta affidabilità del servizio e prestazioni elevate di risposta all'utente, sfruttando le potenzialità del *load balancing*. All'interno di ogni *Vcenter*, inoltre, i server sono stati allocati su nodi fisici *ESXi* differenti e attestati su *Ced* dislocati fisicamente su due poli differenti, la sede centrale dell'Istat e dell'Inail, in *business continuity*.

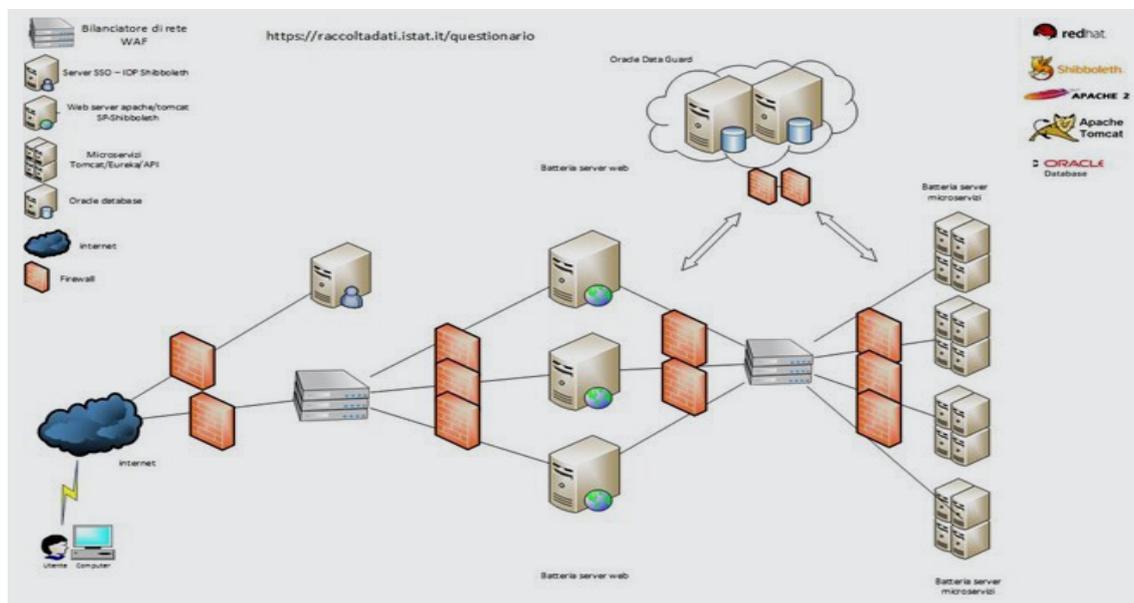
Tutta l'architettura infrastrutturale è protetta dal WAF, che mette a disposizione un puntamento VIP che espone la URL per l'accesso all'utente finale in modalità sicura. Fornisce anche un primo bilanciamento di rete relativamente alle chiamate ai vari server di *back end*, onde evitare colli di bottiglia dell'I/O (Input/Output) o il potenziale sovraccarico di un solo server.

Come descritto per le altre applicazioni, la configurazione a regime prevede alcuni server sempre attivi e altri server disabilitati ma comunque mantenuti e aggiornati.

La Figura 12.5 mostra un esempio semplificato di architettura dell'ambiente di produzione.

Per quanto riguarda lo *stack* tecnologico, il questionario sfrutta le potenzialità del servizio web Tomcat supportato da una Java Virtual Machine. Per la *login* centralizzata sfrutta il *single sign-on* del software *Shibboleth*. Per la connessione al database *Oracle*, sfrutta le potenzialità e la continuità di servizio dell'*application pool*.

Sono stati configurati e allestiti anche i relativi ambienti *demo*, collaudo e sviluppo, secondo gli standard ITIL.

Figura 12.5 - Schema logico infrastruttura di *Panda*

Fonte: Istat

12.6 Data warehouse per la diffusione dei dati

Il sistema informativo *Esploradati* implementa un *data warehouse* distribuito basato sullo standard SDMX (ISO IS-17369) consultabile dagli utenti esterni tramite interfaccia web e ospita al suo interno la banca dati dell'Istituto Nazionale di Statistica.

L'architettura di rete (Figura 12.6) è costituita da:

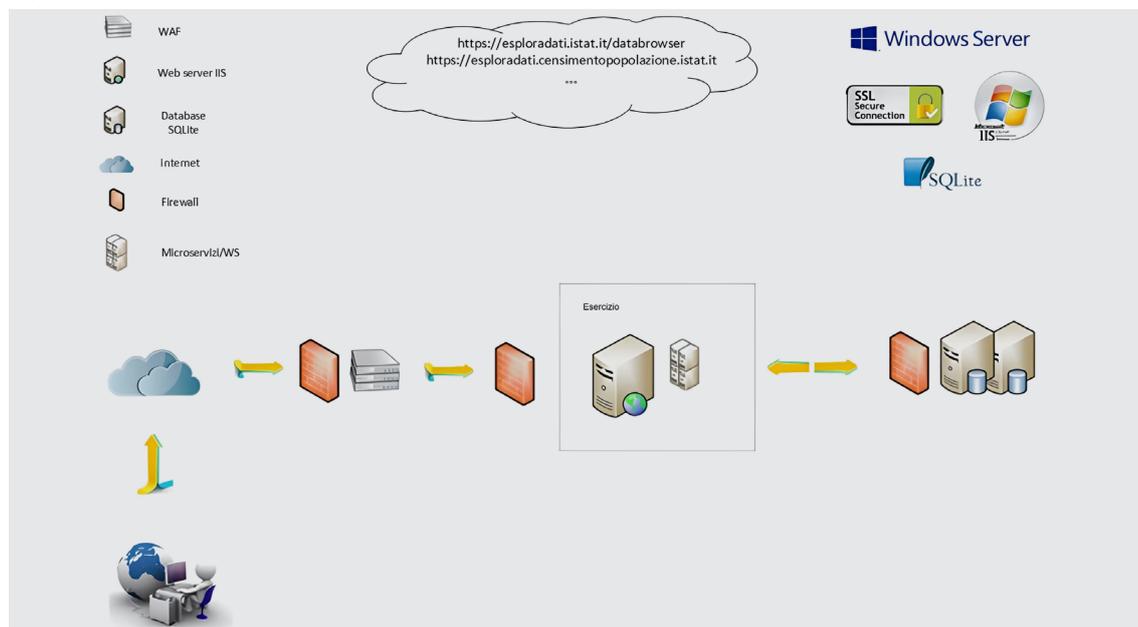
- tecnologia *Windows Server*, la piattaforma per la creazione di un'infrastruttura di applicazioni, reti e servizi web;
- *Internet Information Services (IIS)* di *Microsoft*, web server con un'architettura completamente modulare che offre vantaggi importanti quali l'estendibilità e l'integrazione ASP.NET;
- *.NET Core Runtime* e il modulo *ASP.NET Core*, *bundle* che consente l'esecuzione delle app ASP.NET Core con IIS;
- *Web Application Firewall* per la protezione delle applicazioni web da attacchi dannosi e da traffico Internet indesiderato, e da attacchi *SQL injection* e *denial of service*;
- *firewall hardware* collocato tra la rete interna e la rete Internet in modo da innalzare la barriera che impedisce ai malintenzionati di penetrare all'interno del perimetro di sicurezza per installare *malware*, estrarre dati e svolgere qualsiasi attività che possa costituire una minaccia alla sicurezza dei dati e all'infrastruttura;
- *SQLite* che implementa un DBMS SQL di tipo ACID, cioè atomico, coerente, isolato e durevole;
- un microservizio *core*, che incapsula tutte le principali funzionalità dell'applicazione, e un microservizio dedicato alla funzionalità *news*, che ha un proprio servizio indipendente, con dominio disaccoppiato dal *core*;
- *Secure Sockets Layer*, tecnologia standard che garantisce la sicurezza di una connessione a Internet mediante crittografia dei dati inviati tra un sito web e un *browser*

(o tra due server) per impedire la visualizzazione o il furto delle informazioni trasferite, inclusi dati personali o finanziari;

- modulo *Meta & Data Manager* (MDM) per l'importazione dei dati. L'architettura del *Meta & Data Manager* prevede uno o più nodi di *back end* con cui un *Node API* è in grado di dialogare. Ciascun nodo ha la medesima struttura ed è costituito da un insieme di API. L'importazione dei dati avviene sempre attraverso le API del *Meta & Data Manager* e pertanto le strutture dati create da questo modulo saranno le stesse prodotte dall'MDM;
- *Single Exit Point* per la diffusione via *web service* in formato SDMX dei dati del Censimento della Popolazione e delle abitazioni, memorizzati all'interno del *corporate data warehouse* dei Censimenti;
- un database di *test area staging* per il passaggio dei dati in un'area di *staging* affinché questi vengano ulteriormente lavorati (aggregazioni, correlazione, arricchimento, decodifica, eccetera).

L'infrastruttura ha anche i corrispondenti ambienti di collaudo e sviluppo. Per quanto riguarda il sistema di autenticazione, la banca dati è consultabile liberamente, ma tramite un servizio di registrazione, ogni nuovo utente può autenticarsi e sfruttare il vantaggio di creare proprie viste, visualizzarle e cancellarle se necessario.

Figura 12.6 - Architettura di rete del sito di diffusione



Fonte: Istat

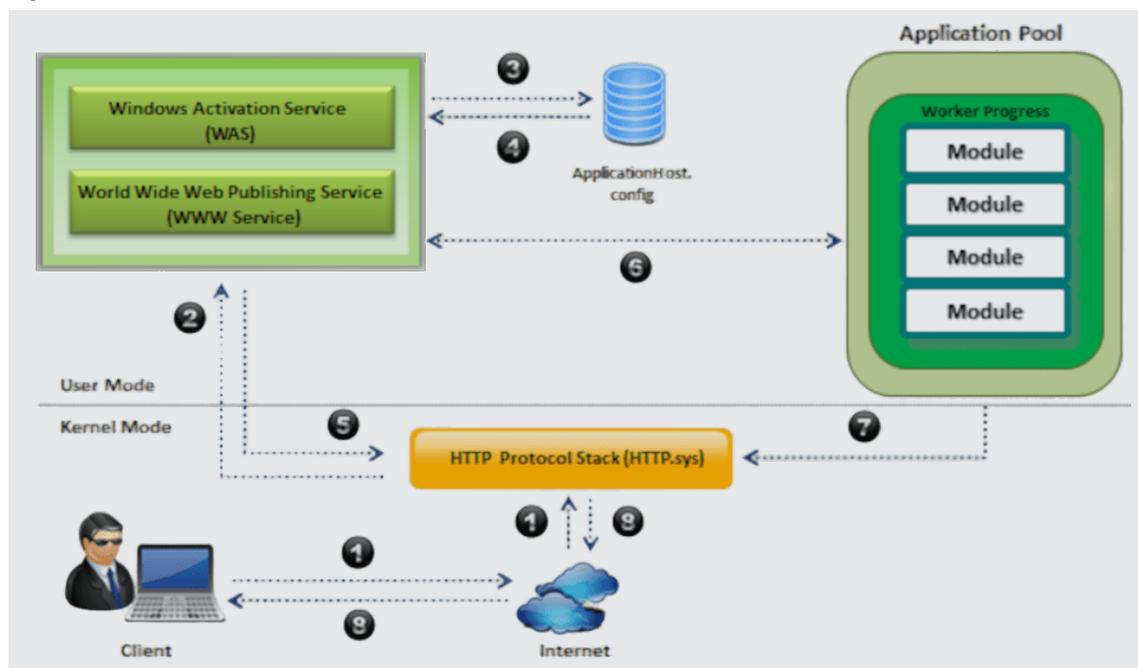
Di seguito una descrizione dell'elaborazione di richieste HTTP in IIS. L'elenco descrive il flusso di elaborazione delle richieste illustrato in Figura 12.7:

1. il *browser* avvia una richiesta HTTP per una risorsa nel server web, HTTP.sys intercetta la richiesta;
2. HTTP.sys contatta *Windows Activation Service* (WAS) per ottenere informazioni dall'archivio di configurazione;
3. WAS richiede informazioni di configurazione dall'archivio di configurazione `applicationHost.config`;

12. Le architetture *middleware*

4. il servizio *World Wide Web* (WWW) riceve informazioni, ad esempio il *pool* di applicazioni e la configurazione del sito;
5. il servizio WWW utilizza le informazioni di configurazione per configurare HTTP.sys;
6. WAS avvia un processo di lavoro per il *pool* di applicazioni a cui è stata effettuata la richiesta;
7. la richiesta viene elaborata e restituisce una risposta a HTTP.sys e il client riceve una risposta.

Figura 12.7 - Flusso di elaborazione HTTP in IIS



Fonte: Istat

13. LE ARCHITETTURE DELLE BASI DI DATI¹

I dati rappresentano l'elemento fondante dell'Istat e, data la loro importanza, tutte le informazioni provenienti dalle indagini statistiche sono archiviate e gestite all'interno di infrastrutture di database dedicate. A seconda del tipo di dato, della modalità di acquisizione e dal tipo di trattamento necessario, viene scelta la tecnologia di database più adatta alle finalità.

I *Relational Database Management System* (RDBMS) *Oracle* sono l'infrastruttura principale per la quasi totalità dei sistemi di produzione statistica in Istat; inoltre nei sistemi informatici di diffusione e di supporto alla rilevazione, nell'ambito delle indagini censuarie, sono utilizzati *MySQL* e *SQL Server* che completano il panorama delle infrastrutture di database.

13.1 Architettura *Oracle* database

L'*Exadata X7-2 database Machine*, di cui l'Istat è dotato, garantisce, mediante potenti server a elevate *performance*, le prestazioni, la transazionalità, la scalabilità, la sicurezza e l'alta affidabilità dei servizi di accesso e trattamento del dato.

L'attuale versione *Oracle* database è la 12c ma è in corso la migrazione verso la versione 19c *Multitenant* che consentirà a un database *Oracle* di funzionare come contenitore, CDB, di più database *pluggablePDB*. Sarà quindi possibile consolidare diversi database in un'unica istanza condividendo le stesse risorse di sistema e mantenendo, però, una separazione logica e amministrativa tra i database. Ai sistemi informatici dei Censimenti verrà dedicato un CDB comune all'interno del quale verranno realizzati più database tematici in cui risiederanno tutte le utenze necessarie per ogni Censimento e fase di esso.

Attualmente i database server sono fisicamente collocati presso i due *Data Center* dell'Istat di Roma: il Centro elaborazione dati (Ced) all'interno della sede centrale dell'Istat e il Ced Inail presso la sede Inail.

L'infrastruttura *Exadata Database Machine*, essendo complessa e di grandi dimensioni, è stata logicamente partizionata in 3 *rack*:

- *rack* 1 denominato DM03 presso il Ced della sede centrale dell'Istat;
- *rack* 2 denominato DM04 presso il Ced della sede dell'Inail;
- *rack* 3 denominato DM05 presso il Ced della sede centrale dell'Istat.

Nei database *Machine* DM03 e DM04 sono destinati i dati raccolti e mantenuti dai sistemi di produzione statistica interna ed esterna tra cui i dati provenienti da tutte le fasi di indagine statistica dei Censimenti; il database DM05 è utilizzato per lo sviluppo e il collaudo di applicazioni e di procedure e in esso sono presenti dati di test.

Ogni *Exadata Database Machine* consta di due *compute node*; ciascun *compute node* è stato configurato con due macchine virtuali dedicate ai database, una interna e una esterna.

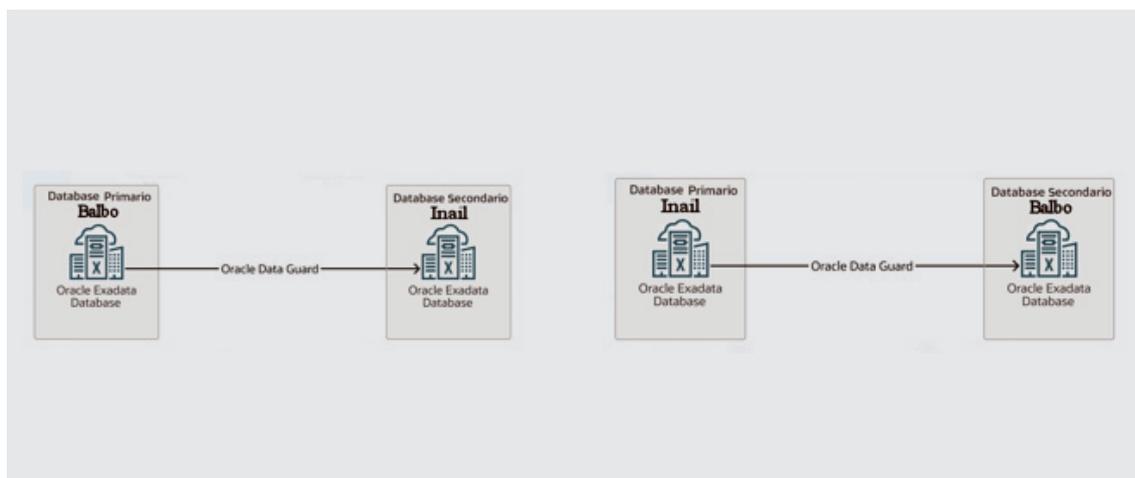
Parte integrante di *Exadata* è il RAC, *Oracle Real Application Clusters*, che consente il *clustering* delle istanze di database per la distribuzione dei carichi di lavoro su più nodi di sistema.

¹ Il Capitolo è stato redatto da: Linda Corradi (paragrafi 13, 13.1 e 13.3.1), Claudia Gioia (paragrafi 13.1.1 e 13.1.2), Fulvio Giannetti (paragrafi 13, 13.1.3, 13.3.1 e 13.3.3), Emiliano Montefiori (paragrafi 13.2.1, 13.2.2 e 13.2.3), Giuseppe Ruggiero (paragrafi 13.1.2, 13.2.3 e 13.3.2).

Ciò offre alle elaborazioni dei dati, come ad esempio quelle di caricamento, di analisi, di correzione e di aggregazione dei dati censuari, di essere processate parallelamente e un utilizzo migliore delle risorse, ottenendo così prestazioni elevate e tempi di risposta più rapidi.

L'utilizzo di RAC con una configurazione attiva permette ai dati archiviati di essere replicati in modo continuativo tra le istanze di produzione, *primary* e *standby* dislocate geograficamente nei due Ced dell'Istat. Il componente *Oracle*, che cura questa attività di sincronizzazione in tempo reale tra le istanze dei database *Machine* e che garantisce coerenza e integrità dei dati, si chiama *Data Guard* (Figura 13.1). Il suo obiettivo è di rendere il servizio di accesso ai dati sempre disponibile e garantire così la *business continuity*. In caso di guasto oppure in caso di manutenzione o di aggiornamento dei sistemi è possibile isolare un'istanza di database e accedere all'altra in modo del tutto trasparente agli utenti interni o esterni, quali i cittadini o gli operatori delle reti di rilevazione.

Figura 13.1 - Replica tra *primary database* e *standby database* mediante *Data Guard*



Fonte: Istat

Altro elemento chiave volto a garantire la *business continuity* e la resilienza dell'architettura, oltre alle procedure di *export* bisettimanali di tutte le utenze *Oracle* di produzione, è rappresentato dal sistema avanzato di *backup* e ripristino fornito da RMAN, il *Recovery Manager di Oracle Exadata*.

Grazie al RMAN è possibile proteggere i dati e tutte le procedure che risiedono nei database da accadimenti non desiderati come cancellazioni causate da errori umani o attacchi informatici, corruzione dei dati o guasti.

In ottica di risparmio temporale RMAN può eseguire il *backup* in parallelo sfruttando il carico di lavoro bilanciato su *Exadata*; inoltre, per ridurre il consumo delle risorse *storage* i *backup* vengono generati sia in modalità *full* sia incrementale. In quest'ultimo caso viene prodotta una copia delle sole modifiche apportate ai dati rispetto all'ultimo *backup* precedentemente eseguito. Si ottengono così, su disco, copie compresse di dati che in caso di necessità possono essere ripristinate rapidamente e con sicurezza, con tempi minimi di fermo, riportando i database a uno stato precedente congruente e funzionante.

Tutte le copie dei microdati censuari acquisiti e archiviati nel database *Machine* esposto, prodotte da RMAN, vengono conservate per un periodo di tre mesi.

13.1.1 Ambienti Oracle database

I dati censuari sono archiviati negli RDBMS *Oracle* messi a disposizione dalla piattaforma *Exadata X7-2*, progettata per fornire prestazioni elevate e affidabilità per carichi di lavoro di database critici. La progettazione prevede l'utilizzo di tre differenti ambienti ognuno con le proprie funzionalità e peculiarità:

- **ambiente di sviluppo**, presente sui due *compute node* del database. L'ambiente di sviluppo è utilizzato per progettare, sviluppare e testare nuove applicazioni o per effettuare modifiche al database. Gli sviluppatori hanno un accesso più ampio per consentire la creazione e la modifica degli oggetti di database e per scopi di *debug*. Le macchine di sviluppo possiedono una configurazione meno onerosa a livello di risorse ma comunque performante dal momento che la quantità di dati da processare è sicuramente inferiore a quella di produzione in quanto campionaria;
- **ambiente di collaudo**, presente anch'esso sugli stessi due *compute node* dell'ambiente di sviluppo ed è utilizzato per testare le modifiche apportate all'applicazione o al database prima che vengano implementate nell'ambiente di produzione. La configurazione dell'*Exadata X7-2* riflette quella dell'ambiente di produzione per garantire test accurati e realistici;
- **ambiente di produzione**, collocato fisicamente sulle due *virtual machine* esposte è destinato all'esecuzione delle applicazioni in produzione e contiene tutti i dati relativi alle indagini censuarie. Dato il patrimonio di informazioni in esso contenute è di primaria importanza implementare tutte le regole per garantire la sua sicurezza e la disponibilità dei dati in esso archiviati. La configurazione è tale da garantire alte prestazioni, elevata affidabilità e scalabilità.

Tutti gli schemi che appartengono ai Censimenti possiedono il medesimo profilo in modo da avere delle configurazioni standard nei parametri quali: impostazione della massima complessità della *password* delle utenze, uguale numero di sessioni concorrenti (pari a 300) e impostazione del parametro del tempo di connessione a *unlimited*.

La protezione dei dati in un ambiente di produzione *Oracle* è di fondamentale importanza per garantire la sicurezza, la *privacy* e l'integrità delle informazioni sensibili, di seguito sono illustrati tutti gli strumenti utilizzati per proteggere i dati censuari in ambito *Oracle* su tutti e tre gli ambienti sopra descritti:

- accesso limitato alle utenze mediante l'utilizzo della figura di responsabile dello schema che richiede la concessione del privilegio di *proxy user* a utenze *Oracle* personali di colleghi che vi collaborano per la coamministrazione;
- creazione di una *password* casuale che rispetti i criteri di complessità e che garantisca l'accesso alle utenze esclusivamente in modo tracciato così come imposto dall'ultima normativa del GDPR;
- utilizzo del sistema di crittografia della *password* mediante il *Secure External Password Store* per evitare di inserire in chiaro le credenziali dello schema nel software;
- abilitazione dell'*auditing* mediante il prodotto *Audit Vault* per tenere traccia dell'attività degli utenti e per registrare gli accessi, le modifiche o la cancellazione di dati, consentendo di rilevare i comportamenti sospetti o non autorizzati;
- applicazione di puntuali politiche di *backup full* e incrementale, affiancate a un *export* completo delle utenze bisettimanale, per facilitare il processo di ripristino e garantire la disponibilità dei dati in caso di perdita o di danneggiamento.

13.1.2 Schemi dei dati e accesso utente

Tutte le utenze, suddivise tra i vari ambienti di sviluppo, di collaudo e di produzione rispettano una nomenclatura standard utilizzata per facilitarne l'identificazione. Ogni indagine censuaria prevede l'utilizzo dei seguenti prefissi secondo le diverse tipologie di utenza:

- SV_ (per le utenze di sviluppo);
- TS_ (per le utenze di collaudo);
- DE_ (per le utenze di *demo*);
- PD_ (per le utenze di produzione);
- AP_ (per le utenze applicative);
- UP_ (per le utenze personali).

Le prime quattro tipologie di utenza possiedono uno spazio fisico più o meno grande dove memorizzare i dati in oggetti di database come ad esempio tabelle, procedure, *package*. Le utenze applicative vengono utilizzate esclusivamente per accedere alle applicazioni web; possiedono una *password* casuale e complessa memorizzata attraverso un sistema di crittografia, realizzato tramite *Oracle Secure External Password Store*. Quelle personali, invece, sono utilizzate per accedere alle utenze di progetto nelle modalità stabilite dai rispettivi responsabili.

Gli schemi sono tutti identificati da una *password* casuale e complessa che non viene memorizzata nemmeno dagli amministratori delle basi di dati. Le utenze *Oracle* per accedere agli schemi sono solo di tipo personale, create tramite il sistema di *Identity Access Management* che consente la gestione centralizzata delle credenziali di autenticazione legata al dominio dell'Istat.

Ogni schema afferisce a un responsabile che si preoccupa della sua gestione, manutenzione e sicurezza. Ogni responsabile, che conosce quali sono le utenze personali autorizzate ad accedere e a trattare i dati censuari di riferimento, ha il compito di richiedere in modo esclusivo il privilegio di *proxy user* per le utenze che necessitano di accedere allo schema che permette a un'utenza personale di connettersi a uno schema senza conoscerne la *password* garantendo una maggiore sicurezza. Questa modalità è stata introdotta a seguito della normativa del GDPR circa il tracciamento degli accessi a cui l'Istat si è adeguato in modo tempestivo. Le utenze autorizzate ad accedere in modalità *proxy user* sono a tutti gli effetti dei coamministratori dello schema potendo effettuare ogni tipo di operazione su di esso.

Qualora un'utenza personale necessiti di avere una visione parziale dei dati o di potere effettuare esclusivamente determinate operazioni, sarà cura del responsabile concedere solo le *grant* opportune sui singoli oggetti dello schema.

13.1.3 Accesso applicativo ai dati tramite Wallet

Del complesso apparato del sistema informatico dedicato a tutte le rilevazioni dei Censimenti, fanno parte numerose applicazioni sviluppate in linguaggio Java.

In seguito a un'analisi di sicurezza effettuata durante le fasi di sviluppo e collaudo si è ritenuto opportuno utilizzare una *feature* di *Oracle Secure External Password Store*, il *Wallet*.

Questa *feature* evita la scrittura in chiaro, nei file di configurazione degli applicativi, delle informazioni relative alla connessione al database e contrasta quindi un eventuale accesso fraudolento ai microdati del Censimento.

Il *Wallet* è un contenitore di dati criptati che consente di eseguire il collegamento a un database *Oracle* utilizzando un *alias*, denominato credenziale, in esso contenuto, al quale vengono associate le informazioni di connessione. Deve risiedere sul server da cui l'applicazione viene distribuita. Essendo i Censimenti serviti da una batteria di server applicativi si è proceduto alla creazione dei *Wallet* e della configurazione delle credenziali necessarie per ogni server della batteria.

Il primo passo consiste nel creare una *directory* destinata a contenere il *Wallet* e i file a esso necessari per il suo funzionamento. Questi file sono:

- il *SQLNet.ora*, che contiene informazioni riguardanti la locazione del *Wallet*;
- il *TNSNames.ora*, che contiene le informazioni di instradamento verso i database come l'*Host*, la porta e il *service-name*.

Al momento della creazione del *Wallet* viene richiesta una *password* che risulta necessaria per la futura gestione dello stesso e che rimane in possesso dei soli componenti del gruppo dedicato alla gestione delle basi di dati.

Subito dopo il *Wallet* sono creati, nella *directory* dedicata, due ulteriori file:

- *cwallet.sso*, che consente al *Wallet* di essere sempre aperto e fruibile;
- *ewallet.p12*, che contiene le credenziali per l'accesso ai database.

Come ulteriore misura di sicurezza è stata creata sulla *directory* una *Access Control List*, ACL Linux (Figura 13.2) per permettere l'accesso in lettura ed esecuzione agli utenti del sistema operativo abilitati e proprietari dei processi applicativi in esecuzione.

Figura 13.2 - Access Control List

```
[ENV=CLIENT21] # getfacl *.*
# file: cwallet.sso
# owner: oracle
# group: dba
user::rwx
user:apache:r-x
group:---
mask::r-x
other:---

# file: ewallet.p12
# owner: oracle
# group: dba
user::rwx
user:apache:r-x
group:---
mask::r-x
other:---

# file: sqlnet.ora
# owner: oracle
# group: dba
user::rwx
user:apache:r-x
group:---
mask::r-x
other:---

# file: tnsnames.ora
# owner: oracle
# group: dba
user::rwx
user:apache:r-x
group:---
mask::r-x
other:---
```

Fonte: Istat

Il passo seguente consiste nella creazione delle credenziali per accedere alle utenze di database che successivamente saranno utilizzate nel software applicativo.

Una credenziale è la combinazione di tre informazioni:

1. l'*alias* per la connessione;
2. il nome dell'utenza per la connessione al database;
3. la *password* dell'utenza per la connessione al database.

Per la generazione delle credenziali viene creata una utenza *Oracle* applicativa con *password* randomica e complessa dotata dei soli privilegi necessari verso le utenze *Oracle* di progetto che archiviano i microdati. Per completare la configurazione e rendere così possibile il suo utilizzo, al file *TNSNames.ora*, residente nella *directory* del *Wallet*, viene aggiunto l'*alias* del *Wallet* con le informazioni utili al raggiungimento del database.

La Figura 13.3 mostra le differenze tra le stringhe di connessione delle applicazioni Java rispettivamente senza e con l'utilizzo della *Secure External Password Store*.

Figura 13.3 - Utilizzo del *Wallet* nella stringa di connessione

```
spring:
  datasource:
    type: com.zaxxer.hikari.HikariDataSource
    url: jdbc:oracle:thin:@(DESCRIPTION=(ADDRESS_LIST=(LOAD_BALANCE=on)(ADDRESS = (PROTOCOL = TCP)(HOST = dm05-scan.istat.it)(PORT = 1521)))(CONNECT_DATA=(SERVICE_NAME=XXXXXXXXX)))
    username: <nome_utenza>
    password: <password_utenza>
    driver-class-name: oracle.jdbc.OracleDriver

spring:
  istatDataSource:
    url: jdbc:oracle:thin://<alias_wallet>
    username:
    password:
```

Fonte: Istat

La connessione applicativa del Censimento mediante *Secure External Password Store* è stata implementata con successo nel 2019 e, in seguito a una fase di sperimentazione, il suo utilizzo è stato adottato come *policy standard* di sicurezza per tutte le applicazioni compatibili sviluppate in Istat.

13.2 Architettura *Microsoft SQL Server*

13.2.1 Architettura dei data base per i tablet

L'Istat utilizza una piattaforma in *Microsoft SQL Server* con *Always On*, che garantisce disponibilità e il ripristino in caso di emergenza, per la configurazione dei tablet utilizzati per i Censimenti permanenti e le indagini continue. Tale piattaforma è costituita da un *setup manager* e da un *Mobile Device Manager*, i quali garantiscono uniformità delle configurazioni e delle *policy* di sicurezza su tutti i tablet della flotta, non modificabili dall'utente finale.

Il *setup manager* gestisce la configurazione del dispositivo inibendo l'uso delle funzioni di base non ritenute necessarie. In particolare:

- esegue l'installazione del sistema operativo e del software necessario per le indagini;
- disabilita le funzioni di base non necessarie, come la possibilità di installare applicazioni da fonti esterne o di modificare le impostazioni del dispositivo;

- configura le *policy* di sicurezza, come le regole di autenticazione e di accesso ai dati.

Il *Mobile Device Manager* effettua il controllo e la gestione centralizzata delle applicazioni distribuite sul tablet tramite uno *store* privato dell'Istat. Questo contiene le sole applicazioni funzionali alle indagini da svolgere.

Una piattaforma in *Microsoft SQL Server* con *Always On* offre i seguenti vantaggi per la configurazione dei tablet utilizzati per i Censimenti permanenti e le indagini continue:

- disponibilità elevata, *Always On* garantisce la disponibilità dei dati e delle applicazioni anche in caso di guasto di un server;
- scalabilità, *Always On* è progettato per scalare in modo lineare per soddisfare le esigenze di crescita dei dati e degli utenti;
- sicurezza, *Always On* utilizza una varietà di funzionalità di sicurezza per proteggere i dati da accessi non autorizzati.

In questo modo si garantisce una configurazione uniforme e sicura dei tablet utilizzati per i Censimenti permanenti e le indagini continue. La piattaforma è costituita da due componenti principali:

1. il *cluster* di server *Microsoft SQL Server*, che è distribuito su due siti geografici, uno locale e uno remoto. Ciò garantisce una maggiore disponibilità del servizio, in caso di problemi di guasto o di manutenzione su uno dei due siti;
2. il sistema di replica dei dati per garantire che questi siano sempre disponibili, anche in caso di guasto di un server. La replica può essere sincrona o asincrona. La prima garantisce che i dati siano sempre aggiornati in tempo reale su entrambi i siti, mentre la replica asincrona consente di ridurre il carico sulle risorse del sito locale.

La piattaforma in *Microsoft SQL Server* con *Always On* utilizza una varietà di funzionalità di sicurezza per proteggere i dati da accessi non autorizzati. Queste funzionalità includono l'autenticazione, l'autorizzazione e la crittografia:

- l'autenticazione è utilizzata per verificare l'identità degli utenti che accedono ai dati;
- l'autorizzazione è indispensabile per determinare quali dati gli utenti possono visualizzare o modificare;
- la crittografia è necessaria per proteggere i dati da accessi non autorizzati.

13.2.2 Architettura del data warehouse per la diffusione

I risultati del Censimento permanente della Popolazione e delle abitazioni vengono diffusi dall'Istat, a partire dal 15 dicembre 2020, su diverse piattaforme che offrono funzionalità e contenuti differenziati pur basandosi sullo stesso tipo di architettura:

- *data warehouse* tematico dei Censimenti permanenti;
- banca dati dedicata;
- dati per sezione di Censimento;
- dati per griglia regolare;
- mappe Gis.

Anche l'architettura interna del sistema di diffusione dei dati del Censimento è basata su *Microsoft SQL Server* in *cluster Always On*, con tutti i vantaggi già descritti nel paragrafo precedente.

Il trasferimento dei dati tra i due siti è sincrono sul motore primario di ogni sito e asincrono sui secondari; il trasferimento sincrono garantisce che i dati siano aggiornati in tempo reale su entrambi i siti, mentre il trasferimento asincrono consente di ridurre il carico sulle risorse del sito locale. Esiste un *backup* continuo delle informazioni, per far sì che i dati siano protetti in caso di perdita o modifica accidentale.

Tutti gli accessi al sistema sono regolamentati da *policy* interne, monitorati e salvati, in questo modo gli accessi al sistema sono autorizzati e tracciabili.

L'architettura interna del sistema di diffusione dei dati del Censimento è progettata per garantire i noti principi di affidabilità, efficienza, scalabilità e sicurezza.

I dati sono archiviati in un ambiente sicuro e protetto da accessi non autorizzati.

13.2.3 Architettura MySQL per il Portale della rete

Il portale è stato sviluppato per offrire il supporto tecnico e organizzativo alla rete di rilevazione del Censimento della Popolazione e delle abitazioni, ed è accessibile al personale interno dell'Istat, alle province autonome di Trento e Bolzano/*Bozen* e a tutta la rete di rilevazione comunale che svolge le operazioni censuarie.

Il portale è, inoltre, sviluppato su una piattaforma *MySQL* ridondata: i dati sono memorizzati su più database in modo da garantire la loro disponibilità anche in caso di guasti hardware o software.

Il database contiene i dati principali del portale, i contenuti del *repository* e le informazioni operative; è sottoposto a regolari *backup* per l'archiviazione di una copia di tutti i dati secondo le *policy* interne dell'Istat.

La piattaforma *MySQL* ridondata è gestita da un sistema di monitoraggio che rileva eventuali anomalie e attiva le procedure di ripristino necessarie, ed è così caratterizzata:

- i database sono distribuiti su due server in diverse sedi geografiche;
- i dati vengono sincronizzati tra i database in modo automatico;
- il sistema di monitoraggio è basato su un software *open source* chiamato *Nagios*.

La piattaforma è stata progettata per soddisfare i requisiti di affidabilità e sicurezza del portale secondo le misure di sicurezza adottate in Istat, essenziali per il corretto svolgimento delle operazioni censuarie.

Il database *MySQL* è stato scelto per i seguenti motivi:

- affidabilità, si tratta di un database relazionale ampiamente utilizzato e supportato, noto per la sua affidabilità e disponibilità;
- scalabilità, può essere facilmente adattato alle esigenze di un portale web di grandi dimensioni;
- sicurezza, offre una serie di funzionalità di sicurezza per proteggere i dati.

Oltre a garantire la disponibilità dei dati in caso di guasti, la replica può essere utilizzata per migliorare le prestazioni di un sistema *MySQL* distribuendo le *query* di lettura su più server e per implementare un sistema di *disaster recovery* in modo da ripristinare i dati in caso di disastro naturale o altro evento imprevisto. *MySQL* è uno dei database già gestiti e disponibili in Istat e ciò ha consentito di sviluppare e mantenere il portale in modo efficiente

e a costi contenuti. L'Istat ha condotto un'analisi comparativa di diversi database relazionali prima di scegliere *MySQL*. L'analisi ha dimostrato che *MySQL* era la soluzione più adatta per soddisfare i requisiti del portale.

13.3 Tecnologie correlate alle basi di dati

Oracle, oltre al prodotto di punta per la gestione degli RDBMS, offre svariate soluzioni di piattaforme integrate per lo sviluppo software, per la distribuzione e la gestione di applicazioni informatiche.

Alcune di esse sono di supporto trasversale a tutte le elaborazioni, le procedure e le applicazioni informatiche dell'Istat; altre sono specifiche di alcuni sistemi di indagine, tra cui quelli dedicati ai Censimenti. Tra i prodotti trasversali più significativi vi sono:

- *Oracle Audit Vault* per le funzionalità avanzate di *auditing* e la sicurezza dei database;
- *Oracle Application Express (Apex)* per lo sviluppo rapido di applicazioni web;
- *Oracle GoldenGate* per la replica e la sincronizzazione dei dati tra database eterogenei.

Apex viene utilizzato per la realizzazione della reportistica dei Censimenti delle Istituzioni pubbliche mentre *Oracle GoldenGate* è impiegato a supporto del Sistema di gestione delle indagini.

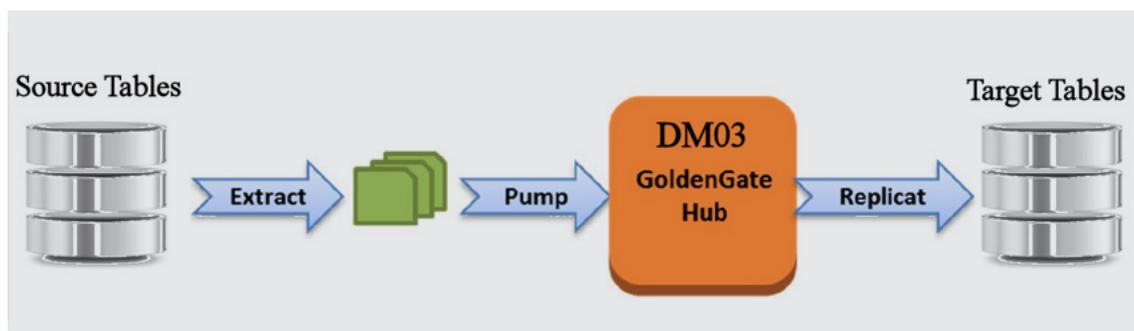
13.3.1 Sistema di replica dei dati in tempo reale

Nel corso del 2023, per rendere disponibili in tempo reale nel Sistema di gestione delle indagini i dati lavorati su istanze di database *remote* di produzione interna, è stato acquisito il prodotto *Oracle GoldenGate Microservices v.21.9* in architettura *hub*.

Il traffico continuo dei dati è unidirezionale dall'istanza di database interna (sorgente) verso quella esposta (target). Si tratta di una soluzione implementata per la replica distribuita dei dati con garanzia di alta affidabilità e continuità del servizio dei database coinvolti nella replica. Nella configurazione *hub* i processi di *GoldenGate* non sono nel database server, ma in una macchina separata situata vicino alla rete del database di destinazione e la connettività viene gestita da *Oracle Net Service*.

Nel contesto dell'Istat il ruolo dell'*hub* viene invece svolto da uno dei due nodi del *cluster* interno dell'istanza *Exadata* (Figura 13.4).

Figura 13.4 - Sistema *hub* primario



Fonte: Istat

Al fine di monitorare il corretto funzionamento dell'infrastruttura, sia dal punto di vista delle risorse necessarie sia da quello dei dati replicati, sono stati predisposti degli *script* e, in caso di criticità, è stato pianificato l'invio automatico di messaggi di allarme tramite email.

Per garantire l'alta affidabilità, in caso di indisponibilità del nodo, l'istanza di *GoldenGate* è ricollocata sul nodo attivo. Se, invece, è l'intero *cluster* a non erogare il servizio, l'istanza viene riattivata su uno dei nodi del *cluster* interno dell'*Exadata* DM04. Se entrambe le istanze *primary* del database sorgente non sono nelle condizioni di erogare il servizio per una manutenzione programmata o per una criticità imprevista, la continuità operativa è assicurata dalla ripartenza di *GoldenGate* su una delle istanze target del database *standby*.

Una volta installato il prodotto, si è proceduto alla configurazione e all'attivazione dei flussi di replica, nel caso specifico dei Censimenti, tra i database sorgente interni e il database target esposto.

In particolare, il flusso si occupa di replicare un sottoinsieme di tabelle di alcuni schemi dei database sorgenti verso il database target.

I componenti fondamentali di *Oracle GoldenGate* sono i processi di *extract* e *replicat*. Essi sono configurati tramite file di parametri specifici che definiscono le tabelle di origine e di destinazione, le regole di filtraggio e l'eventuale trasformazione dei dati. Il processo di *extract* cattura le modifiche dal database sorgente e le memorizza nel file di *trail*; il processo di replica legge le transazioni dal file di *trail* e le applica sul database target, consentendo la replica dei dati in tempo reale.

13.3.2 Controllo delle attività degli utenti di database

I database dell'Istat contengono dati riconducibili a persone fisiche o giuridiche il cui accesso, per mezzo delle utenze *Oracle* personali o applicative, deve essere registrato e monitorato in ottemperanza alle *policy* di sicurezza vigenti in Istat.

Nel 2017 la Direzione informatica ha acquisito la soluzione di sicurezza delle applicazioni e dei dati denominata *SecureSphere*. La sua implementazione è stata finalizzata in particolare alla protezione dei dati nei database rispetto a minacce esterne e interne utilizzando tecnologie di rilevamento, monitoraggio e analisi delle attività volte a individuare anomalie e tentativi di accesso non autorizzati. Le informazioni di dettaglio raccolte circa gli accessi degli utenti e le operazioni eseguite hanno permesso di generare *report* e dimostrare la conformità degli standard di sicurezza applicati.

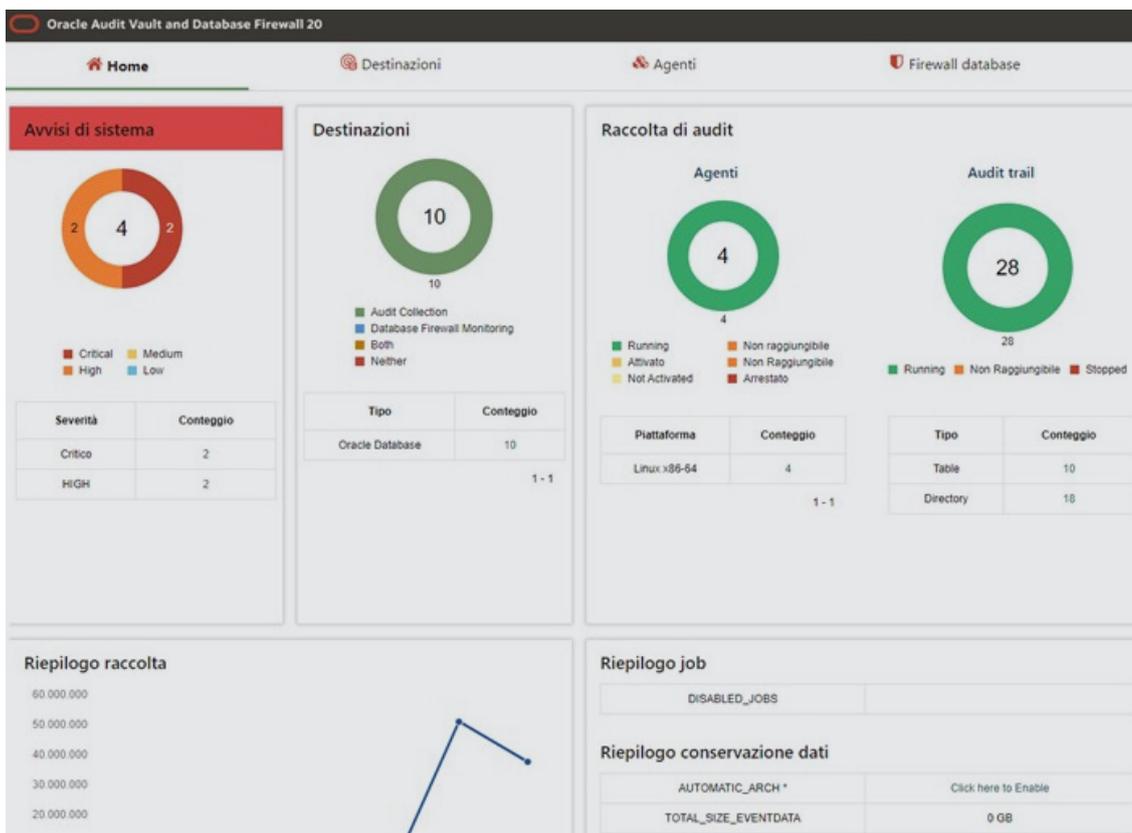
Dal 2020, dopo un periodo di affiancamento, la soluzione *SecureSphere* è stata sostituita con il prodotto *Audit Vault* della *Oracle* la cui applicazione ha riguardato tutti i database, da quelli destinati alla produzione ove risiedono i dati raccolti dalle indagini censuarie a quelli per lo sviluppo e il collaudo di applicazioni e di procedure.

Il prodotto *Audit Vault* è legato al concetto di *Database activity monitoring*: è una tecnologia utilizzata per il monitoraggio e le analisi riguardanti le attività sui database e applicata per identificare operazioni illegali, non connesse a quelle tipiche del sistema.

La soluzione *Oracle Audit Vault* ha messo a disposizione funzionalità gestionali, mediante apposita strumentazione (Figura 13.5) per le regole di raccolta, di acquisizione e di archiviazione cifrata dei dati di *audit* provenienti da tutti i database monitorati, l'analisi centralizzata dei file di *log* e il controllo continuo in tempo reale.

Il risultato produce una visione completa di tutta l'attività eseguita sui database, utile a rilevare minacce e a verificare i requisiti di conformità a garanzia del patrimonio dati che l'Istat acquisisce, gestisce, diffonde e archivia nel tempo.

Figura 13.5 - Console Audit Vault



Fonte: Piattaforma Oracle Audit Vault

Nel dettaglio, mediante l'implementazione e l'abilitazione dell'*auditing* è stato possibile generare reportistica *audit trail* che fornisce informazioni specifiche sulle operazioni eseguite, l'utente responsabile del processo, l'*host* di provenienza, quali oggetti dello schema siano coinvolti, l'orario dell'esecuzione e il comando SQL eseguito. Tutte le azioni eseguite sugli schemi del database indipendentemente dall'origine locale o remota delle stesse, possono essere catturare in file.

L'acquisizione del traffico SQL basato sulla rete e la raccolta dati di *audit* sono utili a gestire il livello di sicurezza di un database. L'implementazione e la gestione delle politiche di sicurezza riguardano l'*auditing*:

- sull'accesso ai dati personali, sensibili o comunque riconducibili a una persona fisica o giuridica;
- sulle attività di utenti privilegiati;
- sulle attività critiche dal punto di vista della sicurezza.

Gli utenti che accedono al database con funzioni quali *Database Administrator* o qualsiasi utente dotato di credenziali di sistema sono considerati utenti privilegiati e sono sottoposti ad *auditing* al fine di identificare tempestivamente eventuali comportamenti difformi dallo standard e quindi potenzialmente pericolosi.



Per mezzo dell'analisi di alcune operazioni ritenute critiche, come la concessione anomala di privilegi, la creazione o modifica non autorizzata di utenti e/o schemi, l'esecuzione di istruzioni SQL di lettura o di modifica dei dati personali o sensibili, si possono individuare pericoli che minacciano la sicurezza delle basi di dati.

I *report* memorizzano, oltre alle attività generali di accesso e le istruzioni eseguite, anche le operazioni su tutti gli oggetti remoti tramite *Database Link* e i *login/logoff* degli utenti per rilevare tempi di attività non congrui.

I file di *auditing* vengono archiviati con meccanismi di crittografia e controllo degli accessi volti a garantire la sicurezza, l'integrità delle informazioni raccolte e a non permettere modifiche o cancellazioni. Essi vengono mantenuti 12 mesi online e 12 mesi in archivio.

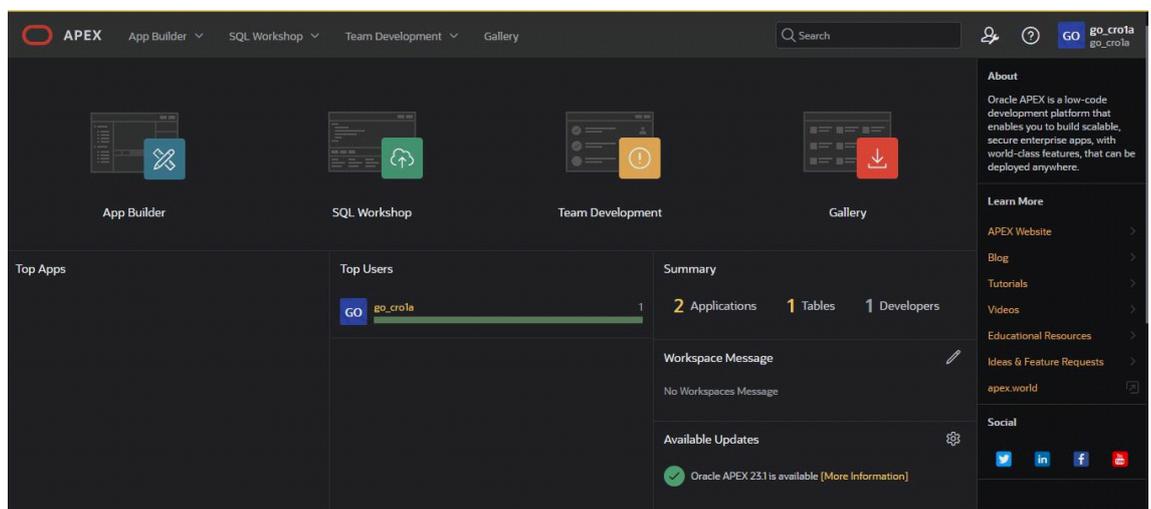
13.3.3 Architettura di Oracle Apex

Oracle Application Express (Apex) è una piattaforma di sviluppo di applicazioni web che consente agli sviluppatori di creare soluzioni moderne e funzionali utilizzando il database *Oracle* come *back end*. La versione attualmente presente in Istat è la 22.2.6, in modalità *multilanguage*, l'infrastruttura è stata installata e configurata anche sui nuovi database *multitenant* versione 19c, in vista della prossima migrazione dalla versione 12c. *Apex* è oggetto di continua manutenzione e aggiornamento da parte del gruppo di gestione dei database per offrire agli utenti l'uso delle più recenti funzionalità del prodotto.

Dal punto di vista sistemistico, consiste in uno schema di database contenente tutti gli oggetti come tabelle, procedure, *function*, *package*, eccetera, che consentono agli sviluppatori di realizzare avanzate interfacce grafiche per accedere ai dati del database. L'ambiente di lavoro è raggiungibile, via web tramite l'utilizzo di un modulo di *Oracle* database, *Oracle REST Data Service (ORDS)*, in grado di creare ed esporre servizi *RESTful* per l'accesso ai dati del database.

In Istat, dopo una prima fase in cui ORDS è stato posizionato direttamente sui server *Exadata*, si è deciso di adottare un'architettura a due livelli, installandolo su application server equipaggiati con *Tomcat*, per gestire con più facilmente la configurazione, la manutenzione e gli aggiornamenti del prodotto. Ogni sviluppatore viene fornito di una *workspace*, una vera e propria area di lavoro, all'interno della quale è in grado di scrivere i propri programmi (Figura 13.6).

Figura 13.6 - *Dashboard* di sviluppo *Apex*



Fonte: Istat, Piattaforma *Apex*

Negli ambienti di sviluppo e test è stata installata la versione *full* di *Apex* per consentire agli sviluppatori di apportare modifiche alle applicazioni. Per motivi di sicurezza, nell'ambiente di produzione del Censimento, si è proceduto a installare solo la parte *runtime*. Ne consegue che il *deploy* delle applicazioni in questo ambiente, dopo l'autorizzazione del gruppo della sicurezza, viene effettuato dal gruppo gestione database.

I dati censuari risiedono sui database esposti. Per potere utilizzare *Apex* su questi dati e garantire un alto grado di sicurezza, si è deciso di implementare la validazione *Shibboleth* a monte del suo utilizzo rendendo quindi necessaria una preventiva validazione tramite la propria utenza di dominio.

14. LE INFRASTRUTTURE DI MONITORAGGIO¹

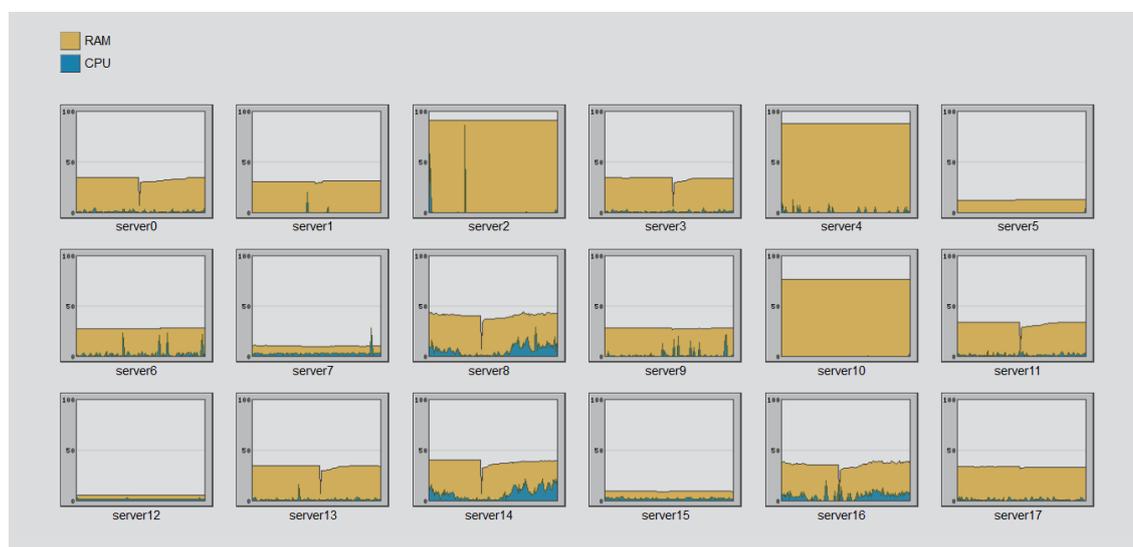
14.1 Monitoraggio dei server

Il monitoraggio delle prestazioni dei server è stato realizzato, durante le fasi di indagine dei Censimenti, con il supporto di software specifico. In particolare, sono state usate le piattaforme *Splunk Enterprise* e *Dynatrace OneAgent*.

Splunk Enterprise è uno strumento specializzato nella raccolta e nell'analisi di grandi quantità di dati. Viene usato efficacemente nella consultazione e nell'analisi dei *log*, ma è adattabile al trattamento di qualsiasi tipo di informazione. I client di *Splunk*, installati di *default* su tutti i server Linux, ricevono i file di dati in diversi formati e con diverse provenienze: *log* di accessi, *log* di sistema, output di procedure create appositamente, eccetera. In particolare, *Splunk* consente di tenere sotto controllo l'utilizzo e la distribuzione temporale di alcune variabili come l'utilizzo della memoria, dei processori e dello spazio su disco. Questo prodotto è stato spesso utilizzato in combinazione con procedure e software sviluppati appositamente, sfruttando la capacità specifica di *Splunk* nella raccolta e nella distribuzione dei dati.

Nella Figura 14.1 sono mostrate le miniature dei grafici di utilizzo delle risorse RAM e CPU per i server coinvolti nel Censimento permanente della Popolazione e delle abitazioni.

Figura 14.1 - Miniature dei grafici di utilizzo della RAM e CPU dei server Linux nell'intervallo di 24 ore

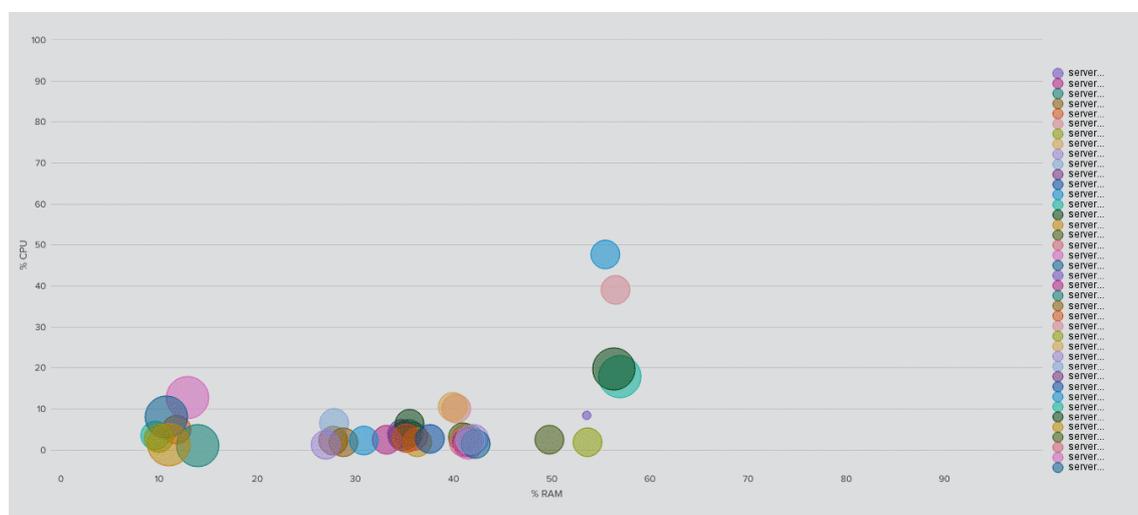


Fonte: Piattaforma *Splunk Enterprise*

In questo caso si utilizza l'*agent* di *Splunk* per la distribuzione dei dati di sistema dei server; mentre viene utilizzato il linguaggio PHP insieme alle librerie grafiche JpGraph per la realizzazione delle miniature. Con questa visione di insieme si possono identificare in modo rapido situazioni critiche nell'utilizzo delle risorse. Selezionando una delle miniature, si ottiene il grafico di dettaglio del server con la possibilità di modificare l'intervallo temporale (24 ore, settimana, mese).

¹ Il Capitolo è stato redatto da Fausto Panicali.

Figura 14.2 - Grafico a bolle dell'utilizzo delle RAM e CPU dei server

Fonte: Piattaforma *Splunk Enterprise*Figura 14.3 - Tabella di utilizzo dei *file system* dei server

host	mount	Type	Tot Spazio	Spazio Residuo	PercentUsedSpace
server5	/web	xfs	60G	4.5G	93 %
server3	/web	xfs	60G	4.5G	93 %
server4	/web	xfs	60G	4.5G	93 %
server1	/web	xfs	60G	4.5G	93 %
server5	/usr	xfs	10G	2.1G	80 %
server2	/usr	xfs	10G	2.1G	80 %
server4	/usr	xfs	10G	2.2G	79 %
server3	/usr	xfs	10G	2.2G	79 %
server1	/usr	xfs	10G	2.2G	79 %
server7	/usr	xfs	10G	2.2G	79 %

Fonte: Piattaforma *Splunk Enterprise*

Per il monitoraggio dei server si è rivelata molto utile anche la piattaforma *Dynatrace OneAgent*. Con questo prodotto un singolo agente si occupa di raccogliere i dati relativi al monitoraggio di qualunque componente dell'infrastruttura. I punti di forza sono la forte interazione tra l'*agent* e l'*host*, che garantisce un'osservabilità avanzata. Si tratta infatti di una tecnologia in grado di tracciare e rendere visibile ogni singola transazione *end to end* di esercizio. Anche *Dynatrace OneAgent* consente quindi di controllare il funzionamento dei sistemi verificando, in tempo reale parametri come l'utilizzo della CPU, della memoria, dello spazio su disco e dell'utilizzo della rete. Consente inoltre di tracciare accuratamente i problemi delle prestazioni usando soglie personalizzate, con possibilità di indagare le cause del problema a livello di codice. L'interfaccia web di *Dynatrace* fornisce, ad esempio, un'utile tabella di riepilogo sulle condizioni dei server (Figura 14.4).

14. Le infrastrutture di monitoraggio

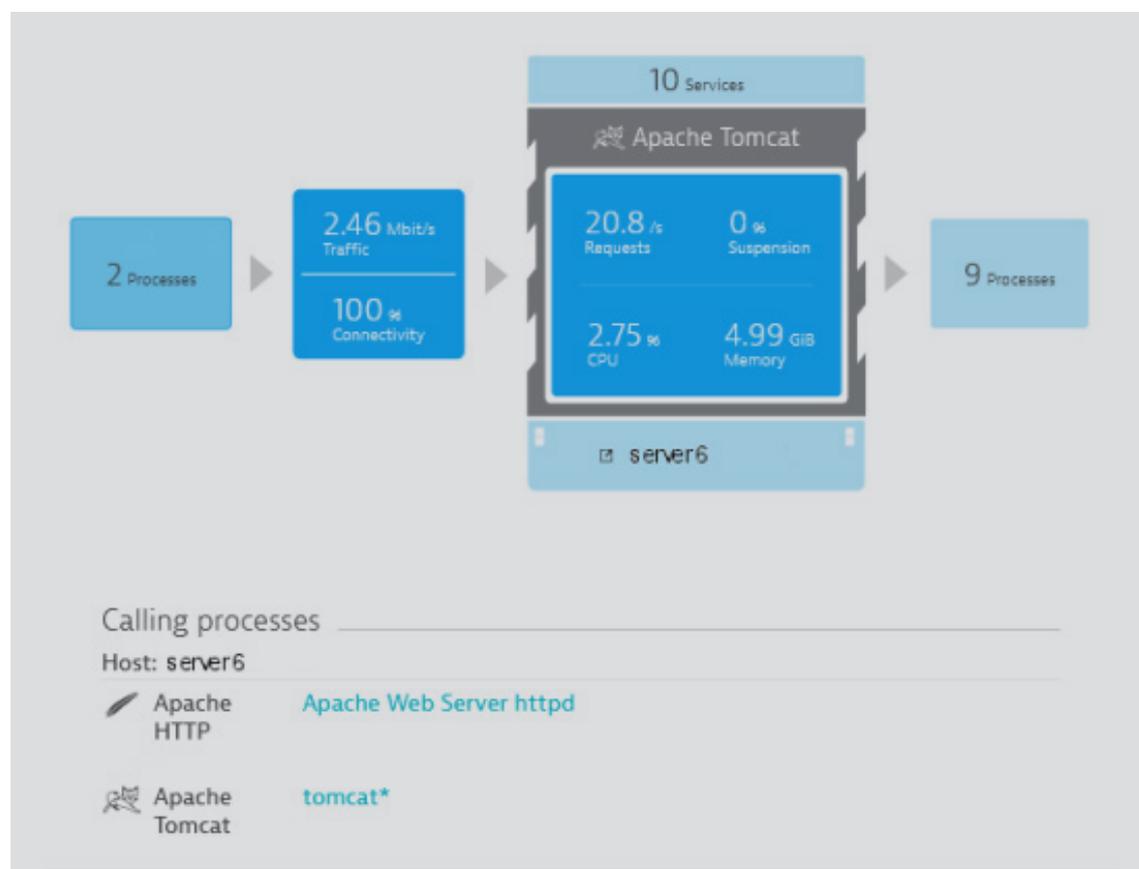
Figura 14.4 - Riepilogo delle condizioni dei server

Name	Operating system	Virtualization	CPU usage	Memory usage	Disk latency	Network traffic
server1	Linux	VMware	2.82 %	43 % of 15.5 GiB	2.04 ms	1.57 Mbit/s
server2	Linux	VMware	3.9 %	42 % of 15.5 GiB	2.04 ms	1.45 Mbit/s
server3	Linux	VMware	13 %	50 % of 15.5 GiB	3.79 ms	22 Mbit/s
server4	Linux	VMware	2.47 %	49 % of 15.5 GiB	2.61 ms	296 kbit/s
server5	Linux	VMware	1.53 %	49 % of 15.5 GiB	8.95 ms	237 kbit/s
server6	Linux	VMware	4.17 %	42 % of 15.5 GiB	3.19 ms	2.09 Mbit/s
server7	Linux	VMware	2.65 %	42 % of 15.5 GiB	3.53 ms	412 kbit/s
server8	Linux	VMware	12 %	47 % of 15.5 GiB	2.81 ms	23.2 Mbit/s
server9	Linux	VMware	1.51 %	48 % of 15.5 GiB	7.76 ms	227 kbit/s
server10	Linux	VMware	14 %	48 % of 15.5 GiB	131 ms	251 kbit/s

Fonte: Piattaforma *Dynatrace OneAgent*

La tabella è ordinabile in base ai campi, permettendo di identificare eventuali valori critici. Selezionando un server si accede a una pagina di dettaglio, con dati di sistema quali RAM, CPU, rete, dischi e processi. Un esempio di visualizzazione è riportato nella Figura 14.5, in cui vengono sintetizzate le risorse coinvolte dai due processi *Apache* e *Tomcat*. Ogni componente dello schema è selezionabile per ottenere ulteriori dettagli.

Figura 14.5 - Schema dei processi *Apache* e *Tomcat*



Fonte: Piattaforma *Dynatrace OneAgent*

14.2 Monitoraggio delle applicazioni

Per il monitoraggio delle applicazioni, analogamente a quello dei server, è stato utilizzato con successo il prodotto *Dynatrace OneAgent*. Grazie al codice inserito ed eseguito lato *browser*, infatti, si possono ottenere anche informazioni sulla *user experience*.

Nella Figura 14.6, ad esempio, vengono mostrate le azioni più frequenti nell'utilizzo del sistema web *Panda*. Le azioni sono ordinate per numero di richieste e mostrano la componente rete e server nel tempo di caricamento delle pagine.

Figura 14.6 - Tabella delle azioni su *Panda*

Action Name	Count	AVG Duration	AVG serverTime	AVG networkTime
loading of page /questionario/secure/pop2023li/cawi/page	23,825	3,991 ms	1,849 ms	171 ms
loading of page /questionario/	1,413	2,708 ms	90 ms	174 ms
loading of page /questionario/secure/pop2023li/cawi/home	1,019	2,328 ms	89 ms	216 ms
loading of page /questionario/secure/pop2023li/cawi/invio	716	2,508 ms	119 ms	183 ms
loading of page /questionario/secure/pop2023li/cawi/confirm	648	2,692 ms	108 ms	178 ms
loading of page /it/censimenti/popolazione-e-abitazioni/compilazione-online	599	2,358 ms	99.6 ms	220 ms
loading of page /questionario/secure/pop2023li/cawi/close	580	2,363 ms	100 ms	154 ms
loading of page /questionario/secure/pop2023li/capi/gtp_0_0	531	2,399 ms	91.2 ms	152 ms
loading of page /	366	2,495 ms	98.1 ms	182 ms
loading of page /questionario/secure/pop2023li/capi/gtp_1_0	345	2,605 ms	107 ms	269 ms

Fonte: Piattaforma *Dynatrace OneAgent*

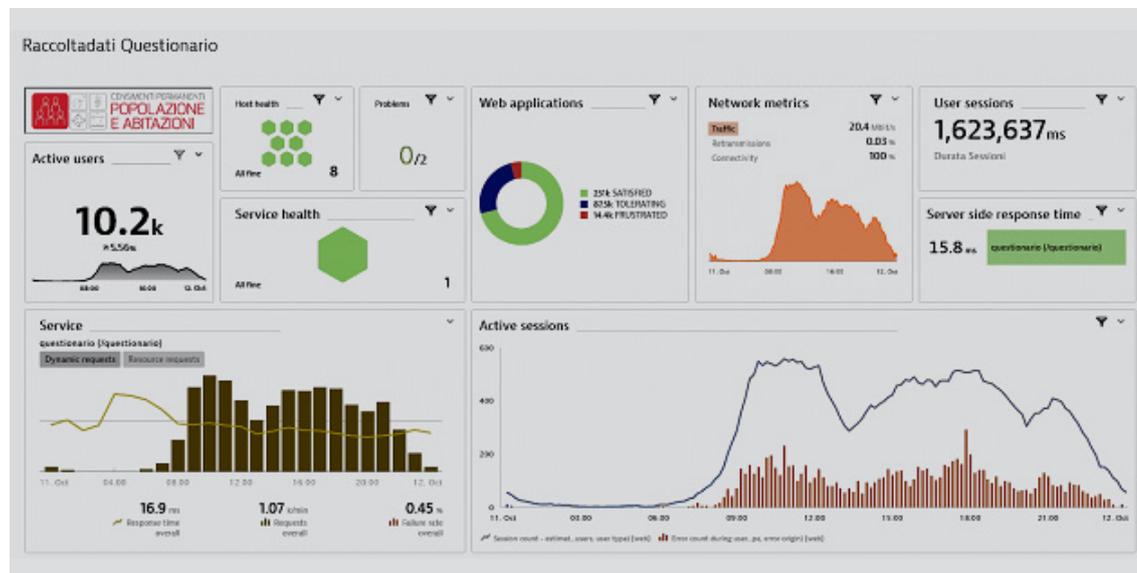
Il prodotto *Dynatrace* permette anche di avere informazioni sui tempi di risposta delle applicazioni, anche secondo il *browser* utilizzato.

Sull'interfaccia web di *Dynatrace OneAgent* sono state create diverse *dashboard* specifiche per alcune delle applicazioni che compongono il Censimento: *Panda*, *Sgi*, *Microservizi*. Questi pannelli sono utili per sintetizzare e per evidenziare in modo immediato lo stato delle applicazioni e dei diversi componenti. Forniscono, inoltre, una sorta di indice per accedere al dettaglio delle diverse visualizzazioni. Sfruttando la particolarità di *Dynatrace*, di seguire il processo fino al livello utente (monitoraggio *end to end*), vengono mostrati alcuni indici di efficienza percepita. In Figura 14.7 è rappresentato l'andamento giornaliero dell'applicazione per il questionario elettronico. Ogni *report* mostrato nella Figura ha un proprio significato;

- il *report* misto (*single value timechart*) quantifica gli utenti attivi;
- il grafico ad anello mostra il grado di soddisfazione dell'utente, che è stimato in tre livelli (soddisfatto, tollerabile, non adeguato). Questi indicatori sono individuati utilizzando come parametri i tempi di risposta e il numero di errori, e possono essere modificati e adattati al tipo di applicazione;
- il grafico a nido di ape evidenzia lo stato globale dei server di *back end* coinvolti, che è indicato con il colore verde se lo stato è ottimale, mentre se vi è un'anomalia è evidenziato in colore rosso;
- il grafico misto (*line in bar*) mostra le sessioni attive insieme al conteggio degli errori;
- il *report* numerico rappresenta la durata media in millisecondi delle sessioni;
- il *report* a rotazione mostra lo stato della rete.

14. Le infrastrutture di monitoraggio

Figura 14.7 - Dashboard dedicata alla piattaforma Panda

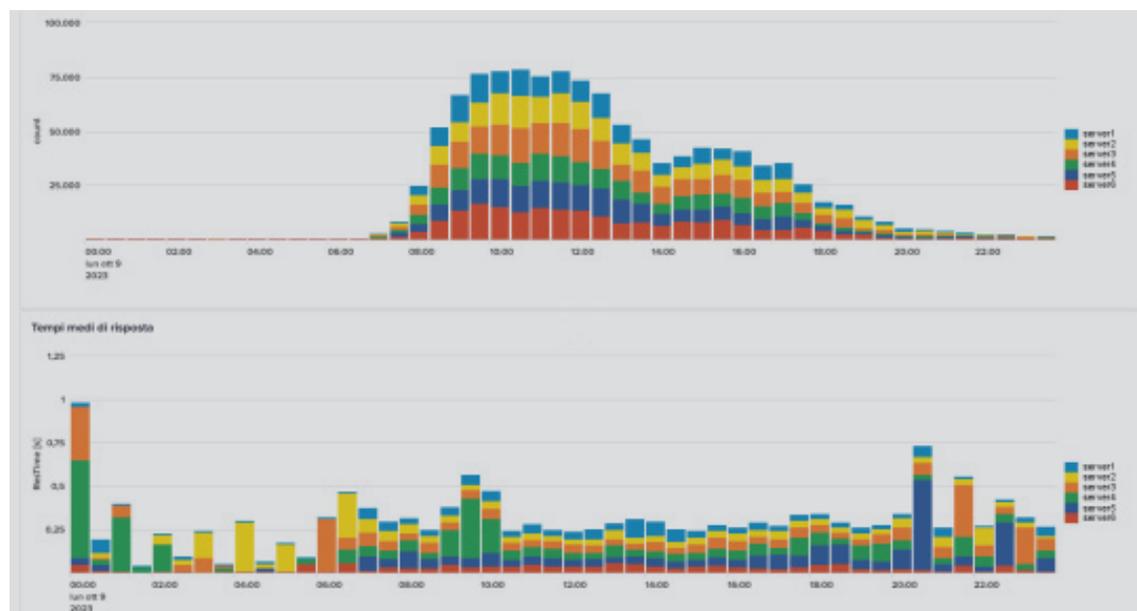


Fonte: Piattaforma Dynatrace OneAgent

Una *dashboard* dedicata al Censimento permanente della Popolazione e delle abitazioni è stata preparata anche sull'interfaccia web di *Splunk*. Con una selezione a tendina è possibile scegliere l'applicazione, ma si può anche scegliere di includerle tutte.

La Figura 14.8 mostra un dettaglio del pannello relativo all'applicazione Sgi; si tratta dei grafici temporali a barre in pila che rappresentano, nella parte superiore, il numero di richieste e, nella parte inferiore, i tempi di risposta in secondi durante un'intera giornata di funzionamento del sito. Una rappresentazione di questo tipo è molto utile per verificare, in un *cluster* di server, il reale utilizzo e il comportamento di tutti i server coinvolti. Si possono così individuare eventuali problemi nel bilanciamento dei server o nella capacità degli stessi server di gestire le richieste.

Figura 14.8 - Numero di richieste per server e tempi di risposta per l'applicazione Sgi



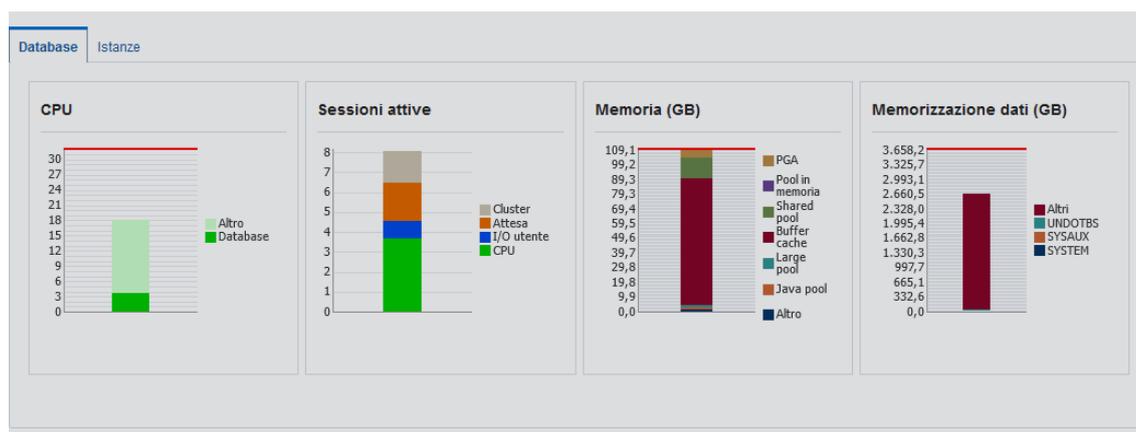
Fonte: Piattaforma Dynatrace OneAgent

In questi grafici, come nei grafici temporali del pannello *Dynatrace*, appare evidente la caratteristica distribuzione a doppia gobba degli accessi alle applicazioni (mattinata e pomeriggio).

14.3 Monitoraggio delle *performance* delle basi di dati

Le *performance* dei database durante tutte le fasi di indagine dei Censimenti sono strettamente monitorate con modalità eterogenee e con il supporto di strumentazione ad hoc, come la *suite Oracle Enterprise Manager* combinata con *Oracle Cloud Management*. Quest'ultima *suite*, in fase di sperimentazione, offre il servizio di monitoraggio di *Oracle Cloud Infrastructure* attraverso cui è possibile individuare anomalie, navigando sulle *dashboard* per visualizzare il rapporto tra le richieste SQL e il consumo delle risorse oppure mediante la predisposizione di avvisi e notifiche in tempo reale riguardanti l'utilizzo della CPU, la memoria, il traffico di rete e lo *storage* utilizzato.

Figura 14.9 - Grafico di monitoraggio di un database esposto di produzione



Fonte: *Oracle Cloud Infrastructure*

L'impatto sulle risorse dell'istanza di database a servizio della produzione statistica esposta, prodotto dalle funzioni applicative di acquisizione dei questionari e dal sistema Sgi, sono controllate anche attraverso procedure *Oracle* di reportistica sviluppate ad hoc in PL/SQL e dedicate alle sole utenze di database dei Censimenti.

Grazie a tali procedure è possibile:

- verificare l'origine e i tempi dei processi in corso nonché le attese onerose;
- individuare *lock* tra sessioni concorrenti che producono uno stallo a causa dell'accesso conflittuale alla medesima risorsa;
- monitorare il numero delle sessioni esistenti aggregate per stato, utente di database e application server;
- intervenire puntualmente sulle sessioni inattive con durata anomala mediante *trigger* di database al fine di contenerne il numero secondo i parametri prefissati e di garantire i servizi sempre disponibili all'utenza.

Tra le attività dedicate al controllo delle sessioni attive le *performance* delle singole istruzioni SQL rappresentano la più importante.

Mediante il confronto dei piani di esecuzione in serie storiche si ottengono segnalazioni di degrado improvviso o anomalo, secondo determinate soglie, a cui fanno seguito delle osserva-

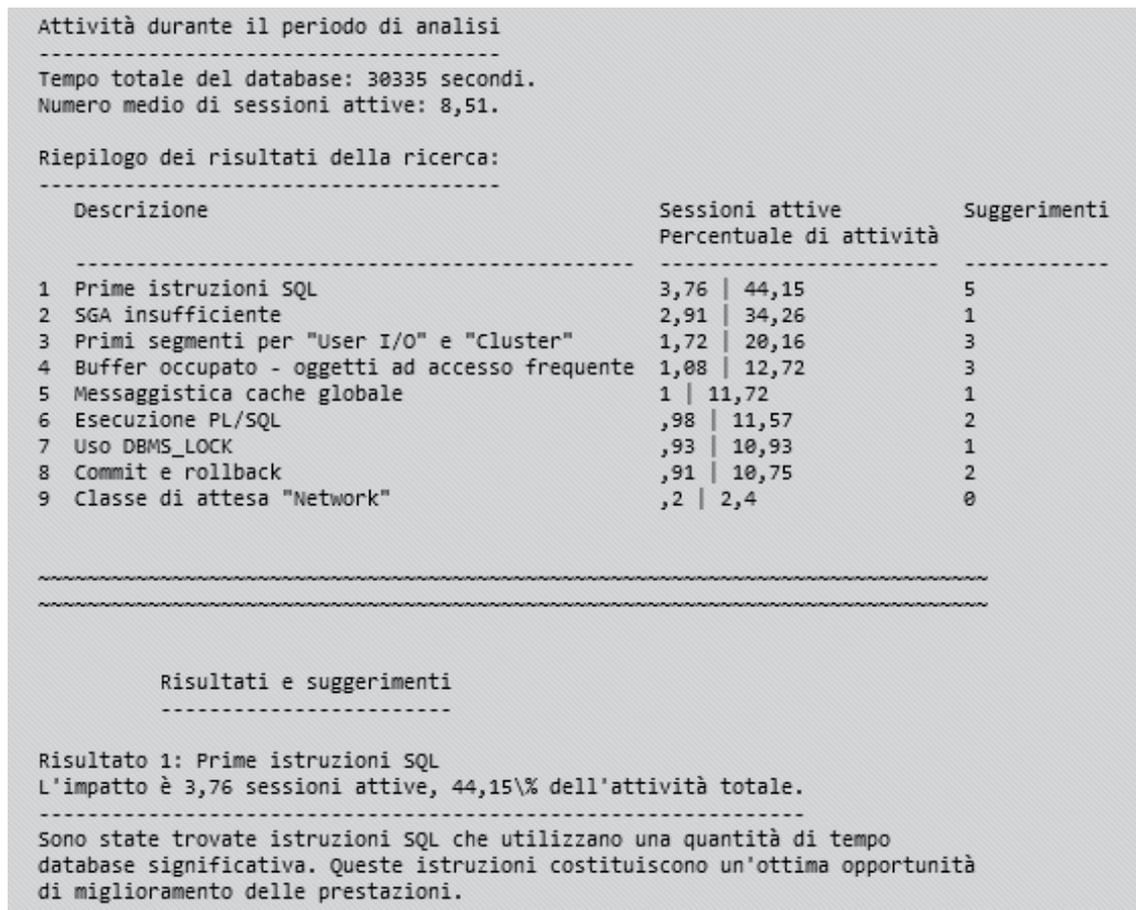
zioni manuali mirate o prodotte con l'ausilio strumentale della *suite Oracle Enterprise Manager*.

In particolare, durante le fasi di collaudo e di avvio della produzione censuaria, attraverso la funzione *SQL Tuning Advisor*, le *query* applicative vengono analizzate al fine di individuare quelle che possono beneficiare di significativi miglioramenti prestazionali.

Sono oggetto di controllo la scrittura dei comandi SQL, le statistiche sulle tabelle e sui campi coinvolti, la presenza e l'uso di indici per l'accesso rapido alle informazioni. A questa attività di analisi fa seguito l'ottenimento di raccomandazioni operative volte a superare le criticità, da implementare in modalità automatica o manuale – in collaborazione con il settore di sviluppo informatico – a livello di codice software o di oggetti di database, oppure attraverso la modifica di parametri di configurazione del database o delle applicazioni informatiche.

Durante le fasi di picco segnalate dalle *dashboard* di *Oracle Enterprise Manager* e *Oracle Cloud Management*, un prezioso strumento di raccolta delle informazioni utili a rappresentare il livello di prestazione dei sistemi è il *report Automatic Workload Repository (AWR)* generato da *Enterprise Manager* mediante l'analisi dei dati raccolti.

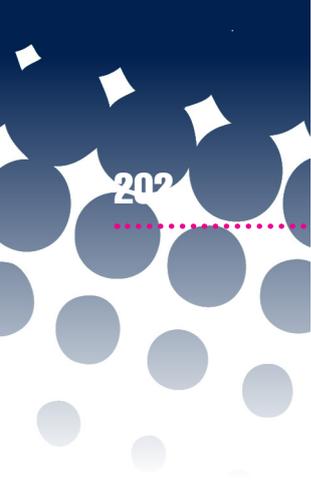
Figura 14.10 - Sezione del *Report AWR* con indicazioni generali e raccomandazioni



Fonte: *Oracle Automatic Workload Repository (AWR)*

Il *report* fornisce indicazioni sia generali sia dettagliate circa l'efficienza del database quali i tempi medi di utilizzo delle risorse, i tempi medi di risposta e gli eventi di attesa più significativi, l'utilizzo della CPU, il carico I/O e il *networking*, fino a segnalare, in sezioni specifiche, le *query* più onerose in termini di tempi di esecuzione e di consumo di risorse, complete di numero di esecuzione, di piani di esecuzione e di statistiche correlate.





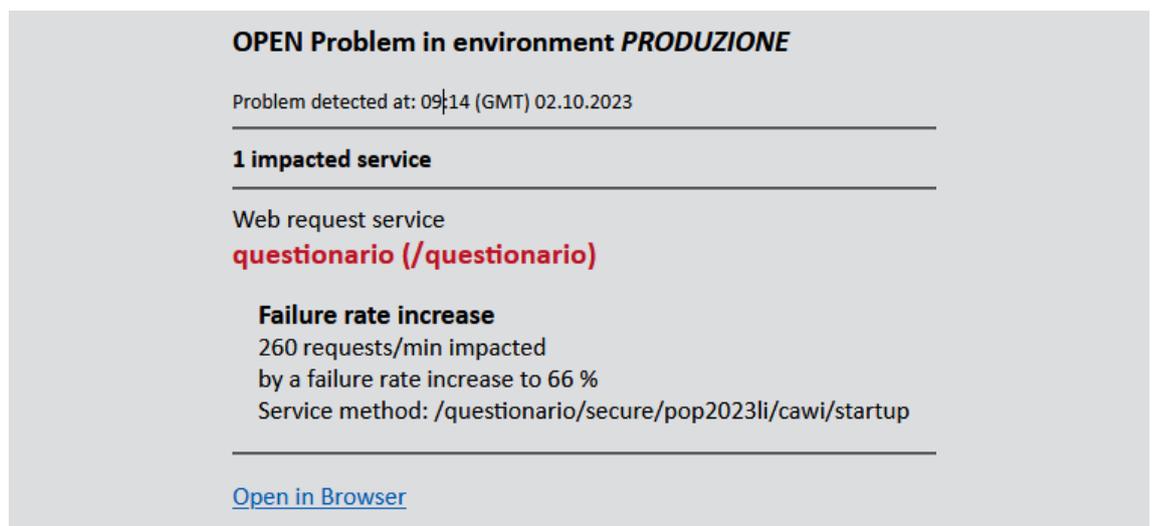
Si tratta, quindi, di un documento in grado di offrire una visione molto accurata su più livelli, utile sia a coloro che svolgono il ruolo di amministratori di sistema sia agli appartenenti dei gruppi informatici responsabili dello sviluppo software applicativo.

14.4 *Alert*, ticket e monitor sintetici

La piattaforma *Dynatrace OneAgent* fornisce un ricco servizio di *alerting*, ampiamente configurabile con soglie, frequenze, finestre di controllo e di esclusione, manutenzione, festività, eccetera.

Nella Figura 14.11 sono mostrate le due fasi, *open* e *resolved*, di una tipica email di *alert* relativa all'applicazione per il questionario; nell'*alert* si evidenziano una breve descrizione del problema e un link per entrare nel dettaglio sull'interfaccia web di *Dynatrace*.

Figura 14.11 - Esempio di email di apertura e chiusura nella segnalazione di problemi



Fonte: Piattaforma *Dynatrace OneAgent*

Un altro metodo per ottenere una verifica del corretto funzionamento di un sito e dei servizi associati è quello dei monitor sintetici. Questa tecnica consiste in un'applicazione client, posizionata in diverse postazioni geografiche, che trasmette una serie di transazioni automatizzate all'applicazione da testare. In genere i controlli vengono eseguiti ogni 15 minuti, ma è possibile configurarli per frequenze diverse. Nella pagina di configurazione viene indicato, oltre alla provenienza geografica, anche il tipo di rete, il *range* degli indirizzi IP, il tipo e la versione del *browser* simulato. I tipi di *alert* che vengono generati da esecuzioni del *browser* sintetico sono simili a quelli mostrati nella Figura 14.11.

Con un collegamento ottenuto grazie a un'interfaccia di integrazione tra *Dynatrace* e il software *Remedy*, l'apertura di un ticket è associata ad alcuni tipi di *alert*. Il ticket viene automaticamente inviato a uno o più *team*, in base al tipo di problema riscontrato. Dal momento in cui, a seguito di nuovi controlli, il problema non viene più riscontrato, il ticket viene chiuso in modo automatico. Sebbene richieda un notevole lavoro preliminare per la definizione dei criteri di attuazione e per la scelta delle destinazioni dei ticket, questo automatismo costituisce, una volta adeguatamente configurato, un valore aggiunto della piattaforma *Dynatrace*.

15. LA GESTIONE DELLA SICUREZZA INFORMATICA¹

Negli ultimi anni, le applicazioni web sono diventate gli strumenti fondamentali per la gestione delle indagini; la fase di acquisizione dei dati, infatti, si sta sempre più avvalendo delle tecniche Cawi o Capi. La dismissione dei moduli cartacei ha, quindi, lasciato il passo alle tecnologie informatiche, che permettono al rispondente di compilare il questionario o in autonomia sui siti dell'Istat, o con l'ausilio di *app* presenti sui tablet forniti ai rilevatori.

Pertanto, è stato necessario adottare provvedimenti di sicurezza informatica che garantiscano, oltre al rispetto del segreto statistico, secondo quanto stabilito dall'articolo 9 del d.lgs n. 322/1989, anche la protezione dei dati, come descritto dai requisiti normativi espressi nella vigente legislatura italiana ed europea, Regolamento (UE) 2016/679 e d.lgs n. 196 del 30 giugno 2003, e dagli standard tecnici attuali, ISO/IEC 27001 e ISO/IEC 27032:2012. In particolare, l'articolo 32 del Regolamento (UE) 2016/679, noto come *General Data Protection Regulation* (GDPR), indica di effettuare una valutazione del rischio sul trattamento dei dati, attuando delle procedure che lo riducano o lo evitino. Nei paragrafi seguenti sono descritte le principali azioni messe in atto dall'Istat per garantire la sicurezza informatica dell'intera infrastruttura censuaria.

15.1 Gestione della sicurezza e i *vulnerability assessment*

La valutazione delle vulnerabilità dei sistemi informatici, *vulnerability assessment*, è un processo che ha lo scopo di verificare la presenza di punti di debolezza nel software o nelle configurazioni di un sistema, che potrebbero essere sfruttati per un attacco informatico. Mediante questo processo le vulnerabilità vengono classificate in base alla tipologia e ne viene fornita una misura della gravità. Sulla base di queste informazioni, unite ad altri fattori quali per esempio la criticità del sistema e la sua esposizione, viene assegnato un livello di rischio a ogni singola vulnerabilità.

Nell'ambito della sicurezza informatica, il *vulnerability assessment* è una delle misure di sicurezza più importanti per la prevenzione degli incidenti di sicurezza informatica che potrebbero essere condotti sia sull'infrastruttura sia sugli applicativi ospitati. La mitigazione delle vulnerabilità individuate costituisce un'azione importante per la protezione dei dati gestiti da un sistema informatico o per garantire la disponibilità del sistema stesso, il cui funzionamento potrebbe essere manomesso temporaneamente o in modo definitivo attraverso lo sfruttamento di vulnerabilità presenti sui sistemi. La mitigazione delle vulnerabilità note si ottiene nella maggior parte dei casi mediante l'aggiornamento del software, in particolare installando le *patch* di sicurezza che i produttori di software mettono rapidamente a disposizione ogni volta che viene individuata una nuova vulnerabilità nei loro prodotti.

Il *vulnerability assessment* viene effettuato mediante piattaforme automatizzate, che effettuano scansioni, appositamente configurate, consentendo di verificare l'installazione del software, la presenza di eventuali *patch* o aggiornamenti di sistema, le configurazioni effettuate.

¹ Il Capitolo è a cura di Cecilia Catalano ed è stato redatto da: Roberto Arena (paragrafo 15.4), Marco Catena (paragrafo 15.3), Sabrina Filippini (paragrafi 15 e 15.2), Aureliano Secchiero (paragrafo 15.1).

Per gestire il processo del *vulnerability assessment*, l'Istat si avvale di una piattaforma commerciale, *Tenable Security Centre*, composta da una serie di sonde, che effettuano periodicamente le scansioni per individuare le vulnerabilità, e di una console centrale che consente di avere una visione di insieme dello stato dell'intera infrastruttura dell'Istat e una gestione controllata delle vulnerabilità riscontrate sui sistemi oggetto di monitoraggio.

In particolare per quanto riguarda l'ambito dei Censimenti, i sistemi che sono sottoposti in modo continuativo a questo tipo di monitoraggio sono i seguenti:

- gli apparati di rete (Wi-Fi, DNS, eccetera);
- gli apparati di sicurezza perimetrale (*firewall*, *Web Application Firewall*);
- le applicazioni web relative al mondo della raccolta dati (Sgi, *Panda*, Microservizi);
- le configurazioni dei sistemi;
- la batteria dei server per ogni singolo applicativo monitorato;
- i database utilizzati dalle applicazioni web di raccolta dati.

Le scansioni sono effettuate singolarmente per ogni apparato o applicativo. Per esempio, per l'applicazione di Sgi, vengono scansionati i seguenti ambiti:

- l'applicazione;
- le configurazioni sul sistema;
- la batteria di server che ospita;
- il database.

Tutti gli apparati di rete e di sicurezza vengono sottoposti a verifica così come le applicazioni web. Queste ultime vengono eseguite in concomitanza ai *penetration test*.

Eventuali segnalazioni relative alle configurazioni e allo stato dei software presenti sui server vengono segnalate ai gruppi di competenza, attraverso il processo di *Change management* previsto dall'Istat, affinché vengano sanate o mitigate. Le scansioni su tutti gli apparati coinvolti vengono eseguite con cadenza mensile.

Al termine di ogni scansione viene redatto un *report* contenente le informazioni sull'analisi effettuata. Nel *report* sono indicati i server o gli apparati di riferimento con il numero e la tipologia di vulnerabilità riscontrate, la classificazione del rischio, la mitigazione o il rimedio da applicare.

A questo scopo sulla piattaforma sono stati predisposti degli avvisi automatici, *alert*, che inviano segnalazioni via email al gruppo di sicurezza informatica.

Se vengono pubblicamente divulgate vulnerabilità a cui vengono assegnati identificativi *Common Vulnerabilities and Exposures (CVE)*², con un possibile impatto classificato come critico o alto, viene eseguita tempestivamente una scansione ad hoc su tutti i sistemi che possono essere affetti da tale vulnerabilità. Qualora ne venga individuata la presenza, viene subito attivato il protocollo relativo per la mitigazione della vulnerabilità.

Ogni notte il database della piattaforma viene aggiornato con le informazioni dei vari produttori di software e hardware attraverso una serie di *web service* che effettuano delle interrogazioni sui siti del produttore.

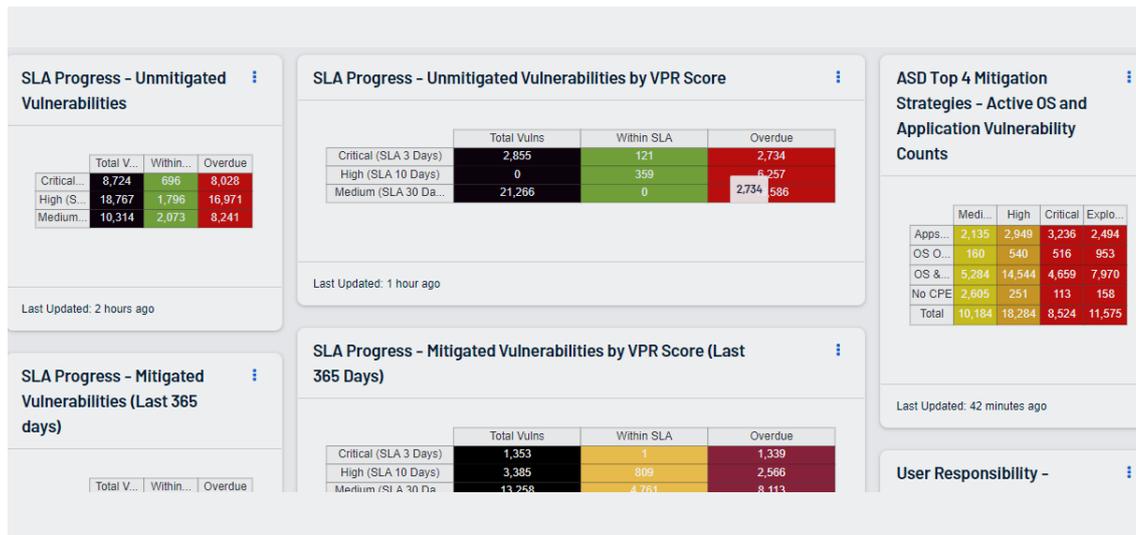
² *Common Vulnerabilities and Exposures (CVE)* è un database pubblico gestito dalla *Mitre Corporation* e contenente le vulnerabilità conosciute e identificate tramite un codice CVE.

15. La gestione della sicurezza informatica

Questo continuo processo di prevenzione, tracciatura, aggiornamento e mitigazione consente di rimuovere tutte le eventuali vulnerabilità presenti sui sistemi critici dell'Istat.

All'interno della piattaforma è possibile creare una serie di *dashboard* personalizzate, profilate adeguatamente, che consentono di avere sotto controllo la situazione dell'intero ecosistema.

Figura 15.1 - *Dashboard* di monitoraggio di sicurezza



Fonte: Piattaforma *Tenable Security Centre*

Sulle *dashboard* sono presenti una serie di indicatori che consentono di verificare i *Service level agreement* (SLA), basati sui tempi di risoluzione, metriche e rischio. In particolare le metriche sulle vulnerabilità possono essere: totali, presenti, mitigate, suddivise per arco di tempo (settimanale o mensile) o per gravità (critica, alta, media, bassa, info).

Nel caso di vulnerabilità conosciute e che, nel particolare contesto in cui si trovano, non costituiscono una minaccia, è possibile accettare il rischio di mantenere la vulnerabilità o rimodularlo assegnando un valore differente.

15.2 Gestione della sicurezza e i *penetration test*

I test di vulnerabilità hanno l'obiettivo di simulare un attacco *cyber* in modo da mettere in luce le problematiche di sicurezza che potrebbero intaccare la cosiddetta triade CIA: *Confidentiality, Integrity and Availability*. Costituiscono, quindi, uno strumento fondamentale per la stima dei potenziali rischi ai quali un applicativo potrebbe essere esposto. Per questo motivo l'Istat effettua un'esecuzione periodica degli stessi su tutti i siti necessari per lo svolgimento dei Censimenti, dalla fase della loro raccolta alla fase della loro diffusione.

Allo scopo di garantire la disponibilità, l'integrità e la riservatezza dei dati trasmessi dai cittadini rispondenti tramite le tecnologie informatiche, l'Istat ha pianificato cicli di verifiche di sicurezza su tutti gli applicativi che trattano i dati censuari e sull'applicazione Android installata sui tablet a disposizione dei rilevatori. I test garantiscono che i dati possano essere letti e modificati soltanto dai diretti interessati e che non possano essere letti, manipolati o distrutti da agenti esterni malevoli o a causa di avvenimenti accidentali.

Gli applicativi sono stati sottoposti a due diverse modalità di verifiche di sicurezza:

- modalità *black box*, in cui si simula un attacco compiuto da un agente di minaccia privo di credenziali di accesso e/o informazioni sul sistema;
- modalità *white box*, in cui si simula un attacco eseguito avvalendosi di credenziali valide per l'accesso e altre informazioni, che un potenziale agente malevolo avrebbe potuto ottenere mediante precedenti attacchi di *phishing* o *bruteforce* o altro attacchi andati a buon fine.

Le verifiche di sicurezza sono state effettuate tramite:

- test automatici, impiegando *tool* che basano le loro scansioni e attività sulle *best practice* e le liste di vulnerabilità conosciute e utilizzabili;
- test manuali che, implementati in maniera empirica e in base alle conoscenze, sono finalizzati a scartare eventuali falsi positivi segnalati dai test automatici, a verificare la presenza di ulteriori vulnerabilità e a identificare eventuali problematiche sconosciute.

Nella loro iterazione i test, soprattutto quelli manuali, vanno adattati agli specifici scenari, partendo da una specifica definizione del target che sarà sottoposto alle verifiche.

Tutti i test, sia automatici sia manuali, sono generalmente effettuati seguendo le metriche suggerite dall'*Open Web Application Security Project* (OWASP) e dalla documentazione ufficiale del *National Institute of Standards and Technology* (NIST), concentrandosi sulle vulnerabilità riconosciute come più pericolose nella classifica OWASP.

La metrica di valutazione delle vulnerabilità, scelta dall'Istat, determina gli indici di rischio in modo qualitativo, associando le vulnerabilità a delle categorie (lieve, moderata, alta e critica) in base al tipo di danno potenziale, alla probabilità che la vulnerabilità possa essere sfruttata e i tempi di recupero delle funzionalità dopo un eventuale attacco.

L'approccio scelto dall'Istat tende a non trascurare nessuna delle vulnerabilità riscontrate perché lo sfruttamento di una vulnerabilità categorizzata come lieve o moderata, in accordo con altre vulnerabilità del sistema, potrebbe avere conseguenze dalla portata sconosciuta.

Pertanto, sono state adottate anche altre buone pratiche, come la rimozione di tutte le risorse e librerie non strettamente necessarie all'esecuzione della corretta funzionalità della applicazione web.

Nel limite del possibile, i test di vulnerabilità sono stati pianificati con il necessario anticipo rispetto all'avvio del Censimento, a causa dell'imprevedibilità del numero delle vulnerabilità riscontrate, del tempo necessario per la loro mitigazione e del numero di test ai quali sottoporre l'applicativo prima che il suo livello di sicurezza sia considerato adeguato.

La validità degli esiti di un test di vulnerabilità è temporanea, non permanente: i test, infatti, sono il risultato di una fotografia di un sistema in un determinato momento. Nel tempo possono presentarsi varie situazioni che richiedono la ripetizione ciclica dei test, come ad esempio la scoperta di nuove falle di sicurezza oppure la necessità dell'aggiornamento o sostituzione di alcuni prodotti non più supportati dai relativi produttori.

Qualsiasi cambiamento all'applicativo o al contesto nel quale viene inserito compromette, inoltre, la validità dei test di vulnerabilità che devono perciò essere ripetuti. La ripetizione, necessaria sia per normativa sia per motivi tecnici, rappresenta una *best practice* supportata anche dalle documentazioni del *project management* nell'ottica del miglioramento continuo suggerito dal ciclo di Deming.

15.3 Gestione della sicurezza perimetrale

Secondo le statistiche sugli attacchi informatici, nel mondo, ogni minuto, una grande azienda o un'infrastruttura nazionale subisce un evento potenzialmente malevolo proveniente da Internet e diretto alla propria rete interna. Gli apparati informatici che fungono da primo scudo di protezione sono perciò di fondamentale importanza.

Il più importante strumento utilizzato per la protezione perimetrale di una rete è il *firewall*, questa tipologia di apparati viene utilizzata con l'obiettivo di creare una separazione tra zone contenenti sistemi informatici con requisiti di sicurezza diversi. Nei casi più semplici, le zone su cui operano i *firewall* sono: Internet, la rete privata, la cosiddetta *Demilitarized zone* (DMZ) contenente i server che rendono disponibili i servizi sulla rete Internet. La separazione tra le zone viene realizzata mediante il controllo dei loro flussi di traffico. Nel caso semplice, il *firewall* è configurato per impedire qualsiasi accesso da Internet alla rete interna e permette alcune connessioni verso la DMZ, limitatamente ai servizi che devono essere messi a disposizione verso l'utenza esterna all'organizzazione.

Per adeguarsi all'evoluzione delle minacce informatiche, gli apparati di tipo *firewall* sono stati dotati di altri servizi di sicurezza, oltre alla funzione di filtraggio del traffico. Tipicamente un *firewall* moderno è in grado di fare l'analisi del contenuto delle connessioni per rilevare la presenza di virus e di *malware*, per applicare politiche di *content filtering* e per rilevare la presenza di anomalie riconducibili a tentativi di intrusione.

Il *firewall* viene spesso affiancato dal *Web Application Firewall* (WAF), che non si limita a verificare la provenienza e la destinazione di una comunicazione per applicare le regole configurate per un determinato flusso, ma entra nel dettaglio degli scambi a livello applicativo per individuare la liceità o meno dello scambio stesso. Pertanto, il fattore che rende un apparato di sicurezza perimetrale efficace, oltre all'accurata implementazione dell'architettura della rete, è la sua corretta configurazione dal punto di vista delle regole di filtraggio e dell'analisi dei contenuti. La scelta delle *policy* e le configurazioni da implementare sugli apparati rappresenta una fase delicata e di estrema importanza, che può preservare da grandi disastri durante i tentativi di attacco.

Gli apparati di sicurezza perimetrale devono quindi garantire la disponibilità dei dati tramite l'esposizione di servizi su Internet e, contemporaneamente, proteggere l'infrastruttura, evitando ogni accesso non autorizzato e ogni modifica non autorizzata ai dati e ai sistemi. L'Istat si avvale di diversi apparati utilizzati per implementare queste funzioni, tra cui i principali sono *firewall* e WAF.

Le misure di protezione realizzate mediante questi apparati sono applicate all'intera infrastruttura dell'Istat, tra cui anche i sistemi dedicati a tutte le fasi di trattamento dei dati censuari.

L'accesso da Internet verso i siti del Censimento, degli applicativi e relativi accessi ai dati vengono filtrati dal *firewall* tramite *policy* di sicurezza e, in particolare, i flussi vengono mitigati dai seguenti moduli: *Antivirus*, *Intrusion Prevention System*, *Intrusion Detection System*, *Application Control*, *Web Filter*, *File Filter*.

Inoltre i *log* degli apparati di sicurezza perimetrale vengono inviati in tempo reale a un sistema centralizzato di raccolta, sul quale vengono effettuate delle analisi automatiche per individuare comportamenti anomali riconducibili a tentativi di attività illecite sui sistemi.

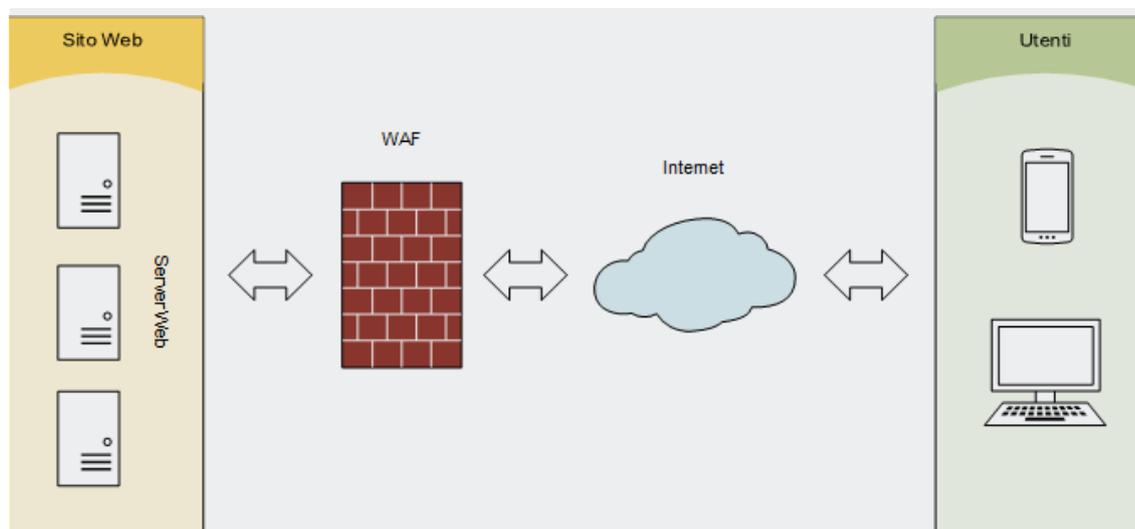
15.4 Pubblicazione dei siti web dell'Istat

La pubblicazione dei siti web dell'Istat, tra cui quelli realizzati per gestire la raccolta dati, il monitoraggio della rilevazione e tutti i servizi accessori per le attività censuarie, avviene mediante la configurazione dei WAF. L'uso dei WAF costituisce una delle misure di sicurezza per proteggere i dati da accessi non autorizzati e permette anche di garantire un'alta affidabilità e la distribuzione del carico su una batteria di server.

Il WAF crea un confine, a livello applicativo, tra l'infrastruttura interna dell'Istat e la rete Internet, bloccando gli accessi che presentano caratteristiche anomale riconducibili a tentativi di attacco informatico e permettendo quelli ritenuti leciti. In particolare è progettato per consentire la distribuzione sicura di applicazioni basate sul protocollo *Hypertext Transfer Protocol* (HTTP).

Il WAF ispeziona il traffico di rete a livello di applicazione, ovvero a un livello più granulare rispetto al tradizionale *firewall* di rete, implementando funzioni di distribuzione delle applicazioni (*Application Delivery Control* - ADC), di bilanciamento del carico del server (*Server Load Balancing* - SLB), offrendo protezione da attacchi che sfruttano vulnerabilità come, per esempio, *SQL injection*, *cross-site scripting*, filtrando e bloccando le comunicazioni verso i server interni.

Figura 15.2 - Schema di connessione a un sito web



Fonte: Istat

I siti web che offrono il servizio di raccolta dei dati sono disponibili agli utenti attraverso degli indirizzi IP virtuali (*Virtual Internet Protocol* - VIP) definiti sul WAF, correlati a diversi server interni che ospitano i siti web.

Per permettere un alto livello di affidabilità e per la continuità operativa del servizio offerto, il nodo del WAF è configurato in modo da bilanciare il carico uniformemente su tutti i server web con l'erogazione del servizio anche in caso di malfunzionamenti di uno o più server. Il bilanciamento si basa anche sulla persistenza della connessione sugli indirizzi IP in modo da rendere più affidabile sia l'autenticazione dell'utente, sia la sessione di interazione con il servizio web.

Il livello di sicurezza applicativo avviene utilizzando esclusivamente il protocollo di comunicazione crittografico *HyperText Transfer Protocol over Secure Socket Layer* (HTTPS), che è il connubio tra il protocollo di comunicazione HTTP e del *Transport Layer Security*.

15. La gestione della sicurezza informatica

Grazie alle relative configurazioni *cipher suite*, ossia un insieme di algoritmi di cifratura, sono rispettate le proprietà di integrità e di riservatezza dei dati in transito. Questa tipologia di protezione è completata da una procedura operativa di verifica di sicurezza proattiva, implementando sui server web e sui siti web ospitati, le attività sistematiche e ripetute di *vulnerability assessment* e di *penetration test*.

16. IL SISTEMA DI GESTIONE DELLA SICUREZZA DELLE INFORMAZIONI¹

L'Istat effettua le indagini censuarie nell'ambito delle sue attività istituzionali, in linea con il Piano statistico nazionale e con le regole deontologiche per trattamenti a fini statistici o di ricerca scientifica del Sistema statistico nazionale. Tale attività ha riguardato le seguenti principali aree tematiche:

- Sistema Integrato Censimento e Indagini Sociali (SICIS);
- Censimenti permanenti delle unità economiche;
- Censimento generale dell'Agricoltura.

D'altra parte, in questo contesto l'Istat si trova a dovere affrontare e gestire le minacce di sicurezza informatica e di cybersicurezza, e quindi gli attacchi informatici che sono tra i principali rischi per le infrastrutture digitali della Pubblica amministrazione, con sempre nuovi ed emergenti scenari di esposizione e di rischio. L'attuale contesto geopolitico, inoltre, ha introdotto strumenti cyber-offensivi altamente più sofisticati rispetto al passato. Come evidenziato dal *Rapporto Clusit 2023*, nel periodo compreso tra il 2018 e il 2022, gli attacchi globali hanno registrato una crescita del 60 per cento e, nello stesso periodo, la media mensile di attacchi gravi a livello globale è passata da 130 a 207. Infine, l'indice di gravità dei danni indotti da tali attacchi è drasticamente aumentato.

Il *Rapporto Clusit 2023* indica che nell'immediato futuro si assisterà a un ulteriore incremento sia della frequenza degli attacchi informatici sia della gravità dei danni, con gravi conseguenze anche in diversi ambiti della società, della politica, dell'economia e della geopolitica. Tale previsione si basa sui risultati dell'analisi del 2022 dove gli attacchi gravi, pari al 45 per cento del totale, si sono verificati nei primi sei mesi dell'anno, mentre quelli con impatto critico hanno rappresentato un terzo del totale. Nel complesso, gli attacchi con impatto critico e alto sono stati il 78 per cento degli attacchi riusciti. In questo ambito l'Italia, sempre secondo il *Rapporto Clusit 2023*, risulta inevitabilmente coinvolta, come dimostra il significativo incremento di attacchi andati a segno nel 2022 e ricevendo ormai il 7,6 per cento degli attacchi globali rispetto al 3,4 per cento del 2021.

Per fare fronte a questo scenario è opportuno agire secondo un approccio che includa l'adozione di misure di prevenzione e mitigazione del rischio volte a innalzare la resilienza delle infrastrutture digitali, per evitare fermi dei sistemi informativi con conseguenti danni reputazionali.

Da tale contesto sono scaturite, quindi, le considerazioni che hanno portato alla definizione della strategia nazionale di Cybersicurezza 2022-2026 e alla istituzione nel 2004 dell'*European Union Agency for Cybersecurity*, la quale contribuisce oggi attivamente alla politica dell'Unione europea in materia di sicurezza nel settore informatico.

Occorre evidenziare che dal 2019 l'Istat ha intrapreso un percorso di certificazione ISO/IEC 27001 (*Information Security Management Systems - Requirements*), partendo dal perimetro *Micro-Data Exchange* (MDE) fino a estenderlo alla progettazione e allo sviluppo di applicazioni, supportando così una governance sempre più capillare e una gestione efficiente ed efficace di tutti gli aspetti di sicurezza delle informazioni.

¹ Il Capitolo è stato redatto da: Giuseppe De Marco e Daniela Segreto (paragrafo 16). Giuseppe De Marco (paragrafi 16.1, 16.1.1, 16.2, 16.3 e 16.4), Giuseppe De Marco, Daniela Segreto e Marco Cellucci (paragrafo 16.5).

16.1 Sistema di gestione della sicurezza per le indagini censuarie

A supporto della strategia sulla sicurezza delle informazioni è stato predisposto un Sistema di gestione della sicurezza delle informazioni (SGSI) aderente allo standard ISO/IEC 27001, con l'obiettivo di mettere in atto un insieme di iniziative e attività di sicurezza per la protezione del patrimonio informativo dell'Istat mediante la definizione e l'implementazione di una serie di misure organizzative, tecniche e di verifica fondate sul rispetto dei principi di riservatezza, di integrità e di disponibilità delle informazioni trattate. In particolare l'SGSI garantisce la protezione delle informazioni da un'ampia gamma di minacce; deve assicurare, inoltre, la continuità della missione dell'Istat, minimizzando i danni in caso di incidenti e massimizzando l'efficienza negli interventi di sicurezza.

L'adozione di un SGSI conforme allo standard ISO/IEC 27001 supporta, in linea con il suo miglioramento continuo, il raggiungimento degli obiettivi di sicurezza definiti sulla base di un approccio sistematico basato sull'analisi e sul trattamento dei rischi. Il processo di analisi e di gestione dei rischi permette l'individuazione delle minacce relative agli *asset* informativi ritenuti rilevanti per il raggiungimento degli obiettivi e della missione dell'Istat, e consente l'individuazione dei controlli e delle contromisure atte a gestire tali rischi. Le azioni da intraprendere per la riduzione del rischio sono commisurate al valore degli *asset* informativi da proteggere e in linea con la strategia di sicurezza definita.

16.1.1 Metodologia

In linea con quanto previsto dal SGSI sono state analizzate le indagini censuarie al fine di indirizzare:

- gli obiettivi di sicurezza delle informazioni e le regole per il loro raggiungimento, coerentemente con la strategia dell'Istat e conformemente alle leggi, alle normative e alle regolamentazioni in ambito di sicurezza, di *cybersecurity* e di *privacy*;
- la valutazione dei rischi informatici a fronte dell'esecuzione di una periodica attività di analisi dei rischi di sicurezza, che permetta un confronto nel tempo dei risultati e una revisione dell'efficacia delle misure di sicurezza adottate;
- un'adeguata protezione degli *asset* informatici in modo commisurato al loro valore, alla classificazione delle informazioni e coerentemente con gli esiti dell'analisi dei rischi.

La metodologia adottata ha individuato una serie di fasi che hanno permesso di tenere sotto controllo, in maniera strutturata e continuativa, le seguenti tematiche:

- *Business Impact Analysis* per gli aspetti IT;
- *Risk Management*;
- *Compliance Management*.

16.2 Business Impact Analysis

La *Business Impact Analysis* (BIA) e la gestione dei rischi sono delle attività finalizzate a garantire la continuità operativa e la resilienza dell'Istat. La BIA permette di pianificare e prendere decisioni per proteggere le attività operative attraverso la comprensione delle attività critiche; la gestione dei rischi permette di mitigare le potenziali minacce al fine di ridurre gli impatti di eventi imprevisti sulle operazioni.

La fase di *Business Impact Analysis* ha permesso di individuare i processi critici definendo i requisiti necessari per il ripristino dei processi aziendali secondo i tempi stabiliti; permette, dunque, di valutare la criticità dei processi e degli *asset* associati, attraverso l'analisi degli impatti a fronte della perdita dei requisiti di riservatezza, di integrità e di disponibilità.

Il perimetro della BIA è stato individuato nelle rilevazioni statistiche di interesse pubblico affidate all'Istat, nell'ambito del Sistema statistico nazionale, dal Programma statistico nazionale. Tali attività devono considerarsi come quelle dirette all'ottenimento del prodotto finale di interesse degli *stakeholder*.

L'attività di BIA svolta dall'Istat si è basata su un approccio integrato e coordinato alle questioni di sicurezza delle informazioni e ha inteso, inoltre, raccogliere le valutazioni sugli impatti da utilizzare nell'ambito dell'analisi dei rischi relativi alla sicurezza in modo da assicurare piena sinergia con le attività di continuità operativa. L'obiettivo prioritario, quindi, è stato di assicurare la conformità alla normativa ISO 22301:2019 e utilizzare i risultati della BIA quale input per la determinazione degli impatti legati alla sicurezza delle informazioni e per la protezione dei dati personali.

In tale contesto sono state, quindi, analizzate le indagini censuarie e i relativi impatti nell'eventualità in cui l'Istat fosse coinvolto in uno scenario di emergenza che possa mettere a repentaglio la corretta erogazione dei servizi.

In linea con il Programma statistico nazionale è riportato di seguito l'elenco delle indagini censuarie analizzate con le relative aree tematiche:

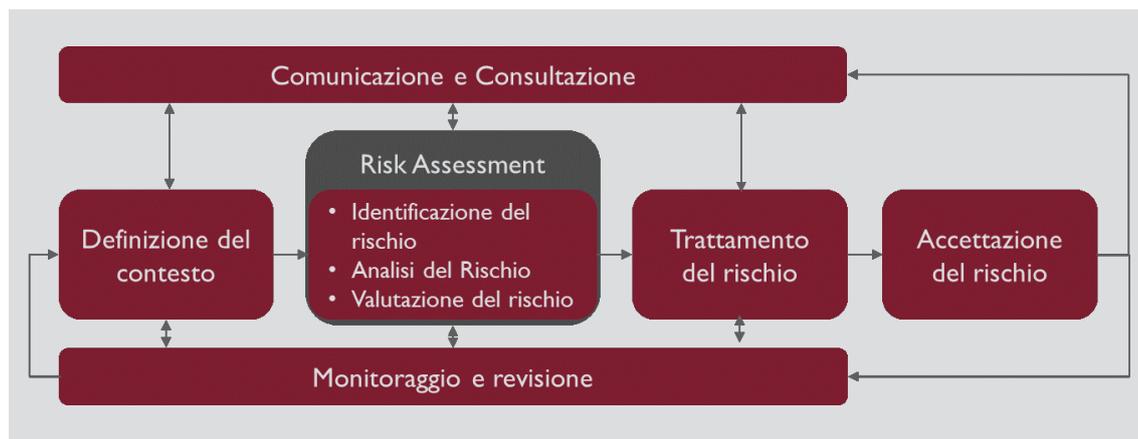
- IST-02493 - Sistema Integrato Censimento e Indagini Sociali, componente areale;
- IST-02494 - Sistema Integrato Censimento e Indagini Sociali, componente da lista;
- IST-02578 - Censimenti permanenti delle unità economiche - Rilevazione campionaria sulle istituzioni non profit;
- IST-02575 - Censimenti permanenti delle unità economiche - Rilevazione censuaria delle istituzioni pubbliche;
- IST-02623 - Censimenti permanenti delle unità economiche - Rilevazione multiscopo qualitativa sulle imprese;
- IST-02792 - Censimento generale dell'Agricoltura 2020.

L'attività di BIA ha evidenziato che gli impatti più significativi per le Indagini in oggetto sono legati a perdite di integrità e di riservatezza, mentre l'aspetto della disponibilità non costituisce una variabile critica. Per quanto riguarda l'impatto degli aspetti economici le indagini censuarie di maggiore attenzione per l'Istat sono il Censimento permanente della Popolazione e delle abitazioni, il Censimento permanente delle unità economiche e quello delle istituzioni pubbliche; mentre per quanto riguarda l'impatto degli aspetti, oltre che economici, anche normativi è il Censimento generale dell'Agricoltura.

16.3 Risk Management

L'approccio di analisi e gestione del rischio adottato ha come riferimento la norma ISO 31000:2018; per quanto riguarda la sicurezza delle informazioni, invece, il riferimento specifico è la norma ISO/IEC 27005:2018 le cui fasi sono riportate di seguito.

Figura 16.1 - Metodologia di gestione del rischio



Fonte: Istat

La metodologia di gestione del rischio ha previsto quindi:

- definizione del contesto, ovvero definizione del perimetro e individuazione delle informazioni e dei criteri di valutazione e accettazione del rischio;
- identificazione del rischio, ovvero determinazione degli eventi che possono causare una potenziale perdita di riservatezza, integrità e disponibilità delle informazioni;
- analisi dei rischi, ovvero stima qualitativa degli impatti, delle probabilità di accadimento delle minacce e dell'efficacia mitigatoria apportata dai controlli nell'ambito del trattamento del rischio stesso. Il piano di trattamento dei rischi è definito dal responsabile della *Security Governance* dell'Istat.

In particolare la fase di analisi del rischio è stata condotta considerando le minacce che possono avere impatto sugli *asset* che insistono sui Censimenti secondo la classificazione dell'Agenzia per l'Italia Digitale (AgID), indicata di seguito:

- abuso dei sistemi informatici;
- accesso non autorizzato a credenziali valide;
- attacchi all'infrastruttura fisica dell'organizzazione;
- compromissione dei sistemi informatici di fornitori o partner dell'organizzazione target;
- conduzione di un attacco DoS (*Denial of Service*);
- effetti indesiderati di un cambiamento;
- errori accidentali o per negligenza da parte di utenti;
- furto o perdita di sistemi informativi;
- introduzione di *malware* nei sistemi informatici;
- introduzione di codice non autorizzato nelle applicazioni o nei software;
- mancata soddisfazione dei requisiti contrattuali da parte di terzi;
- manipolazione interpersonale;
- monitoraggio o modifiche non autorizzate delle comunicazioni;
- *phishing*;
- *probing* o *scanning* non autorizzati sulla rete;

16. Il Sistema di gestione della sicurezza delle informazioni

215

- raccolta di informazioni sull'organizzazione pubblicamente disponibili;
- sfruttamento di vulnerabilità nei sistemi informatici sull'organizzazione;
- sfruttamento di architetture di rete progettate in maniera non efficace;
- sfruttamento di errori di configurazione nei sistemi informatici dell'organizzazione;
- sfruttamento dello smaltimento non sicuro dei sistemi informatici dell'organizzazione;
- sfruttamento delle vulnerabilità nei meccanismi di autenticazione;
- *session hijacking*.

La valutazione delle probabilità associata a ogni minaccia, in linea con la metodologia di riferimento, è stata effettuata facendo riferimento alle esperienze precedenti, alle statistiche degli eventi sui singoli *asset* e ai *trend* per ogni singola minaccia.

16.4 Compliance per gli aspetti di sicurezza e cybersecurity

In linea con la strategia nazionale di *cybersecurity*, è stata effettuata un'analisi di *compliance* e maturità, con particolare attenzione per le attività legate alle indagini censuarie che l'Istat svolge.

L'analisi è stata effettuata in termini di *compliance*, verso le applicabili normative per la Pubblica amministrazione e, in particolare, verso:

- il perimetro di sicurezza nazionale cibernetica (decreto legge del 21 settembre 2019, n. 105, convertito con modificazioni dalla legge del 18 novembre 2019, n. 133, Decreti attuativi);
 - la direttiva NIS2 (Direttiva UE 2022/2555 del Parlamento Europeo e del Consiglio del 14 dicembre 2022, pubblicata sulla Gazzetta ufficiale dell'Unione europea del 27 dicembre 2022) in vista della sua attuazione per gli Stati membri dell'Unione europea (17 ottobre 2024) come abrogazione della direttiva NIS;
 - il regolamento sui servizi *cloud* della Pubblica amministrazione (Regolamento di cui all'articolo 33-*septies*, comma 4, del decreto legge 179/2012).
- Per ognuna delle precedenti normative, tale analisi ha permesso quindi di:
- verificare la maturità delle misure tecnico/organizzative di gestione dei rischi di *cybersecurity* al momento in essere e che le normative stesse richiedono;
 - individuare le azioni e gli investimenti in ambito tecnico/organizzativo che si devono mettere in campo per indirizzare le eventuali non *compliance*;
 - individuare i fattori che concorrono a una gestione efficiente ed efficace della sicurezza dell'Istat, il cui ambito di analisi, comprende politiche, processi, strutture organizzative e misure di protezione.

16.5 Sistema di monitoraggio della governance della sicurezza

Il monitoraggio della governance della sicurezza assume un ruolo fondamentale nelle attività di indagini censuarie che l'Istat svolge. La capacità, infatti, di definire e visualizzare puntualmente gli indicatori di monitoraggio dell'efficienza ed efficacia del sistema di gestione stesso è vitale non solo per l'adeguamento al *framework* ISO/IEC 27001, ma rappresenta anche di fatto una *best practice* in ambito sicurezza delle informazioni.

Un sistema di monitoraggio ben progettato offre, quindi, una visione continua e aggiornata dello stato del SGSI, consentendo l'identificazione tempestiva di potenziali problemi o anomalie. Tale strumento, perciò, può agire per prevenire o mitigare rischi e minacce alla sicurezza delle informazioni.

La Direzione informatica, pertanto, sta implementando una piattaforma di analisi dei dati personalizzabile che consenta di verificare, monitorare e analizzare gli indicatori di mo-



monitoraggio del SGSI di Istat. Tale strumento è basato su una *dashboard* di visualizzazione dei dati di monitoraggio che facilitano la comprensione dello stato del SGSI e supportano le decisioni riguardo a eventuali azioni correttive da intraprendere. Un quadro visivo chiaro e immediato degli indicatori consente, quindi, un'interpretazione rapida e precisa, permettendo una risposta tempestiva a potenziali problemi di sicurezza.

Attraverso l'implementazione di questa piattaforma sarà possibile raggiungere risultati efficaci sia nella gestione proattiva dei rischi relativi alla sicurezza delle informazioni sia nell'abilitazione e automazione delle attività di monitoraggio basate sui dati stessi di governance del SGSI.

RIFERIMENTI BIBLIOGRAFICI

- Agenzia per l'Italia Digitale - AgID. 2019. "Linee guida per il rilascio dell'identità digitale per uso professionale". *Linee guida disponibili sull'area web dedicata*. Roma, Italia: AgID. https://www.agid.gov.it/sites/default/files/repository_files/linee_guida_identita_digitale_per_uso_professionale_v.1.0.pdf.
- Arkills, B. 2003. *LDAP Directories Explained: An Introduction and Analysis*. Boston, MA, U.S.: Addison-Wesley Professional.
- Ascari, G., G. Brancato, A. Bruni, S. Daddi, G. Di Bella, S. Giavante, F. Rocci, R. Sanzo, A.M. Sgamba, G. Simeoni, e S. Spirito. 2023. "Il monitoraggio e la valutazione della qualità del Sistema Integrato dei Registri". *Istat working papers*, N.7/2023. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/il-monitoraggio-e-la-valutazione-della-qualita-del-sistema-integrato-dei-registri/>.
- Bernardini, A., D. Bonardo, A. Dentini, M. Giacummo, M. Mazziotta, and A. Preti. 2020. "The Italian Continuous Censuses: Managing the Complexity between Registers, Statistics and Territories". *Rivista Italiana di Economia Demografia e Statistica*, Volume 74, N. 1: 85-96. http://www.sieds.it/listing/RePEc/journal/2020741P08_028_sieds_2019.pdf.
- Borel, B. 2023. *The Chicago Guide to Fact-Checking. Second Edition*. Chicago, IL, U.S.: University of Chicago Press.
- Boselli, C. 2023. "Le piattaforme di diffusione dei dati aggregati". In Cammarrota M. (*a cura di*). *Sistema integrato di diffusione delle statistiche dell'Istat: percorso e stato attuale*: 35-49. *Lecture Statistiche - Metodi*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/sistema-integrato-di-diffusione-delle-statistiche-dellistat-percorso-e-stato-attuale/>.
- Bowman, J., S.L. Emerson, and M. Darnovsky. 1996. *The Practical SQL Handbook: Using Structured Query Language. Third Edition*. Boston, MA, U.S.: Addison-Wesley.
- Cantoni, F., e G. Mangia. 2005. *Lo sviluppo dei sistemi informativi nelle organizzazioni. Teoria e casi*. Milano, Italia: Franco Angeli.
- Casale, D., M.P. Di Maio, e G. Vaste. 2022. "Tirocini informatici online in Istat: esperienze di learning-by-doing di coding e di utilizzo di strumenti low-code". In Marafioti, R.S., P. Ciancarini, P. Ravotto, e M. Gentile (*a cura di*). *Atti DIDAMATICA 2022. La trasformazione digitale nella Scuola, negli ITS, nell'Università e nella formazione professionale*: 314-321. Milano, Italia: Associazione Italiana per l'Informatica ed il Calcolo - AICA. <https://www.aicanet.it/documents/10776/4555506/ATTI+Didamatica+2022/469726b1-58d5-4f48-8a8f-0fa1d2fb0f34>.
- Cecconi, N. (*a cura di*). 2021. *La conduzione della raccolta dei dati del Censimento permanente della popolazione e delle abitazioni 2019*. *Lecture Statistiche - Metodi*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/la-conduzione-della-raccolta-dei-dati-del-censimento-permanente-della-popolazione-e-delle-abitazioni-2019/>.
- Codd, E.F. 1970. "A relational model of data for large shared data banks". *Communications of the ACM*, Volume 13, N. 6: 377-387. <https://doi.org/10.1145/362384.362685>.
- Di Zio, M., e A. Ferruzza (*a cura di*). 2024. *L'informazione statistica microterritoriale: l'esperienza del Registro statistico di base dei luoghi*. *Lecture Statistiche - Territori*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/informazione-statistica-microterritoriale-lesperienza-del-registro-statistico-dei-luoghi/>.
- Eurostat, European Commission. 2010. "Business registers. Recommendations manual". *Eurostat Methodologies and Working papers*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/ks-32-10-216>.
- Fedeli, M. (*a cura di*). 2021. *Il ruolo e l'evoluzione dell'IT nella statistica ufficiale: l'esperienza dell'Istat*. *Lecture Statistiche - Metodi*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/il-ruolo-e-levoluzione-dellit-nella-statistica-ufficiale-lesperienza-dellistat/>.

- Feuerstein, S., and B. Pribyl. 2002. *Oracle PL/SQL Programming. Third Edition*. Sebastopol, CA, U.S.: O'Reilly & Associate.
- International Organization for Standardization - ISO. 2022a. *ISO/IEC 27001:2022. Information security, cybersecurity and privacy protection - Information security management systems - Requirements*. Vernier: ISO.
- International Organization for Standardization - ISO. 2022b. *ISO/IEC 27002:2022. Information security, cybersecurity and privacy protection - Information security controls*. Vernier: ISO.
- International Organization for Standardization - ISO. 1997. *ISO 9241-1:1997. Ergonomic requirements for office work with visual display terminals (VDTs)*. Vernier: ISO.
- Istituto Nazionale di Statistica - Istat. 2022a. *Conti Economici delle Imprese e dei Gruppi di Impresa. Anno 2020*. Statistiche Report. Roma, Italia: Istat. <https://www.istat.it/it/archivio/276829>.
- Istituto Nazionale di Statistica - Istat. 2022b. *Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2021*. Roma, Italia: Istat. <https://www.istat.it/it/files//2022/12/Nota-metodologica-censipop-2021.pdf>.
- Istituto Nazionale di Statistica - Istat. 2022c. *Popolazione residente e dinamica demografica. Anno 2021*. Roma, Italia: Istat. <https://www.istat.it/it/files/2022/12/CENSIMENTO-E-DINAMICA-DEMOGRAFICA-2021.pdf>.
- Istituto Nazionale di Statistica - Istat. 2018. *Piano generale del censimento permanente della popolazione e delle abitazioni*. Roma, Italia: Istat. <https://www4.istat.it/it/files/2016/03/PGC-CENSIMENTO-POPOLAZIONE-E-ABITAZIONI.pdf>.
- Istituto Nazionale di Statistica - Istat. 2015. "Il Censimento delle imprese". *Atti del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit*. Fascicolo 2. Roma, Italia: Istat. https://www.istat.it/it/files//2016/02/Atti_CIS_2.pdf.
- Macchia, S., E. Berntsen, M.C. Congia, e D. De Siena (a cura di). 2023. *Il Censimento Permanente delle Istituzioni Pubbliche: registri e rilevazione censuaria multiscopo*. Letture Statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/ii-censimento-permanente-delle-istituzioni-pubbliche-registri-e-rilevazione-censuaria-multiscopo/>.
- Project Management Institute - PMI. 2021. *PMBOK® Guide. A Guide to the Project Management Body of Knowledge. Seventh Edition*. Newtown Square, PA, U.S.: PMI.
- Radini, R., M. Di Zio, e G. Vaste. 2021. *Il Sistema Integrato dei Registri dell'Istat: un sistema di dati integrato a servizio del Paese*. Presentazione alla XIV Conferenza Nazionale di Statistica. Roma, 30 novembre - 1° dicembre 2021. <https://www.slideshare.net/slideshow/14a-conferenza-nazionale-di-statistica-250769570/250769570>.
- Regolamento (CEE) n. 696/93 del Consiglio, del 15 marzo 1993, relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità. *Gazzetta ufficiale dell'Unione europea*, L 76, 30 marzo 1993: 1-11. <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:31993R0696>.
- Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati. *Gazzetta ufficiale dell'Unione europea*, L 119, 4 maggio 2016: 1-88. <https://op.europa.eu/s/zJvJ>.
- Saad, E., and R. Mitchell. 2020. *Web Security Testing Guide (WSTG)*. Wilmington, DE, U.S.: Open Worldwide Application Security Project - OWASP. <https://github.com/OWASP/wstg/releases/download/v4.2/wstg-v4.2.pdf>.
- Simeoni, G., F. Rocci, F. Brait, G. Brancato, M. Di Zio, R. Filippini, M. Riccio, R. Sanzo, and G. Vaste. 2022. "A Model for Documenting and Monitoring Quality of Statistical Registers According to GSBPM and GSIM." In *ModernStats World Workshop*. Belgrade, 27-29 June. https://unece.org/sites/default/files/2022-07/MWW2022_Presentation_Italy_Simeoni.pdf.
- Vaste, G., M.P. Di Maio, P. Giorgetti, M. Passacantilli, e F. Dell'Orco. 2021. "Il sistema integrato dei registri (SIR): sperimentazione su strumenti di controllo di processo". Presentazione alla XIV Conferenza Nazionale di Statistica. Roma, 30 novembre - 1° dicembre 2021.

