



FORNITURA A CONSUMO DI SERVIZI DI SVILUPPO E MANUTENZIONE EVOLUTIVA  
DEL SOFTWARE APPLICATIVO DEI SISTEMI INFORMATIVI DELL'ISTAT

## ALLEGATO 1 AL CAPITOLATO TECNICO

### **Descrizione del contesto applicativo e tecnologico**

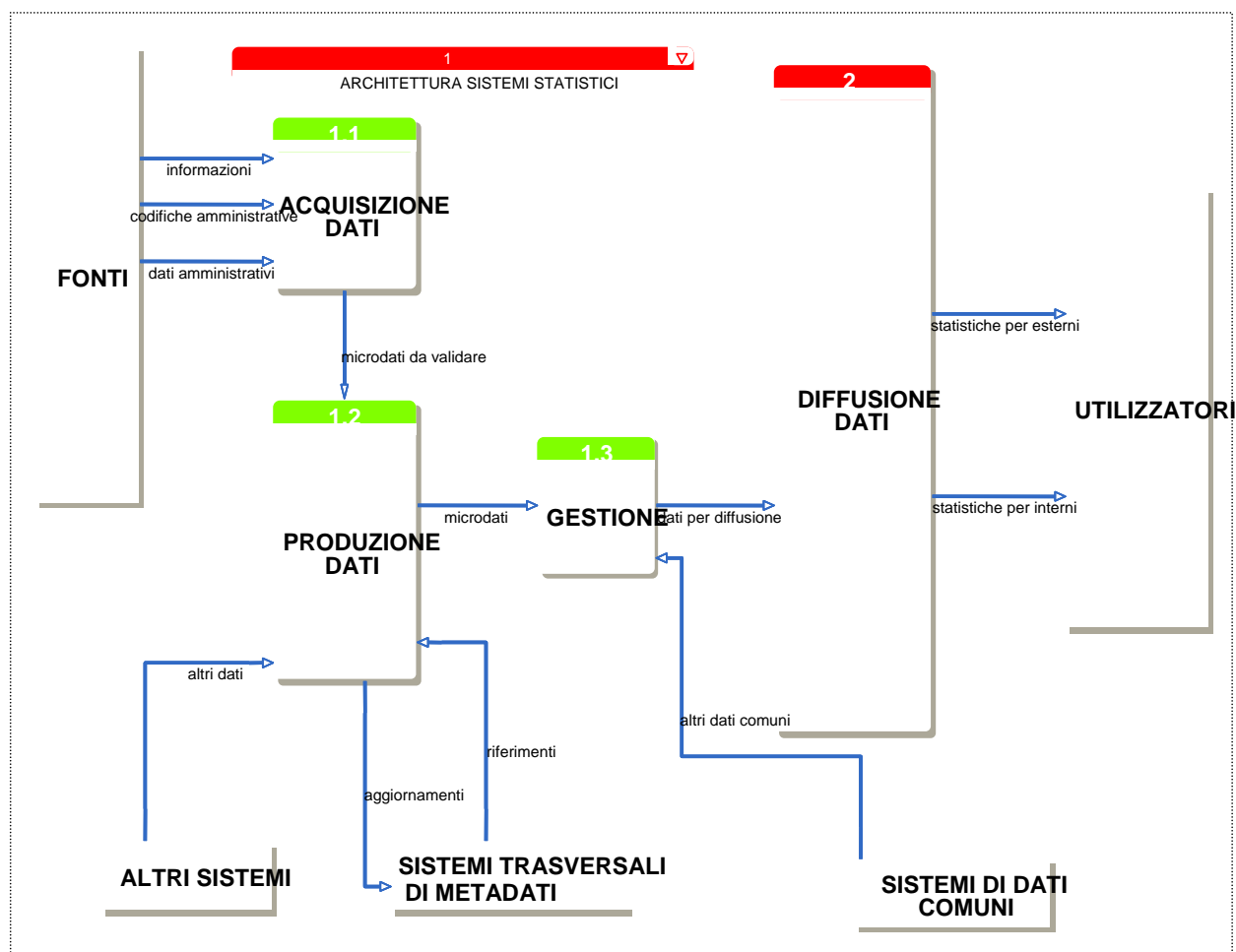
## ***INDICE***

<b>1. ARCHITETTURA DEI SISTEMI ISTAT .....</b>	<b>3</b>
1.1. TIPOLOGIA DEI SISTEMI STATISTICI.....	3
1.2 ARCHITETTURA DEI SISTEMI GESTIONALI.....	6
<b>2. GLI AMBIENTI TECNOLOGICI .....</b>	<b>8</b>
2.1 SISTEMI DI ELABORAZIONE .....	8
2.2 GLI STRUMENTI DI SVILUPPO SOFTWARE .....	8
2.3 LA RETE ISTAT .....	11

# 1. ARCHITETTURA DEI SISTEMI ISTAT

## 1.1. TIPOLOGIA DEI SISTEMI STATISTICI

L'architettura applicativa standard dei sistemi informativi e di elaborazione dati in ambito statistico dell'ISTAT, può essere delineata - ad un primo livello di astrazione - dallo scenario di contesto riportato nella figura seguente:



In tale contesto sono riconoscibili: il complesso dei processi interni al sistema informativo oggetto (area delineata in grassetto nella figura precedente); le interfacce generali di ingresso e uscita; altre interfacce di intercomunicazione, tipicamente rappresentate da altri sistemi informativi collegati, i principali flussi scambiati tra gli elementi del sistema.

Di seguito vengono descritte sinteticamente le componenti caratteristiche del sistema attraverso una decomposizione funzionale utile come riferimento anche per le future articolazioni dei processi:

## **Ambiente di Acquisizione**

E' costituito dalla sfera applicativa avente il compito di:

- acquisire in formati opportuni i dati (anche eterogenei) provenienti dalle varie fonti di input, sia tramite strumenti CASIC (Computer Assisted Survey Information Collection) che via web,
- descrivere i dati e definirne i supporti,
- ma e' anche costituito dagli applicativi inerenti le attività di predisposizione delle unità di rilevazione statistiche quali:
- stratificazioni dell'universo,
- estrazione di campioni base ed integrativi,
- ed infine dalle attività di carattere gestionale quali:
  - gestione delle anagrafiche,
  - gestione dei piani di invio, di sollecito e di spoglio dei questionari.

Il prodotto di uscita è rappresentato tipicamente da collezioni di dati grezzi.

Tutti gli ambienti di acquisizione sono progettati e realizzati nel pieno rispetto della normativa sulla privacy (VPN, HTTPS, ...).

## **Ambiente di Produzione**

Sfera applicativa che si occupa delle attività di natura statistico-informatica per il trattamento e la validazione dei dati: si tratta quindi del software, anche di tipo generalizzato, dedicato a:

- controllo quantitativo dei dati, ossia della rispondenza tra invii e ritorni di questionari, della univocità e pulizia delle chiavi identificative, della congruità di eventuali forniture a lotti;
- controllo qualitativo, ossia della compatibilità dei valori raccolti con quelli ammissibili, intra o extra questionario;
- controllo della natura dei dati mancanti;
- correzione interattiva o automatica con imputazione deterministica o probabilistica;
- calcolo di coefficienti di riporto e riporto di dati campionari all'universo.

L'output fornito è costituito da microdati validati.

## **Ambiente di Gestione/Elaborazione**

Sfera applicativa destinata a svolgere le attività di:

- trasformazione dei dati validati in base di dati relazionale, se tale passo non è stato effettuato precedentemente;
- aggiornamento della base di dati;
- verifica e mantenimento della integrità e della coerenza dei dati;
- elaborazioni statistiche per l'aggregazione e la trasformazione dei dati ossia per il passaggio da microdati (o dati elementari) a macrodati (aggregazioni di microdati);
- costruzione, gestione ed aggiornamento dei thesauri informativi e dei metadati ;
- gestione della interconnessione tra sistemi, in particolare con i sistemi dei dati comuni.

L'output fornito è costituito da microdati validati e macrodati validati.

## **Ambiente di Diffusione**

Sfera elaborativa destinata a svolgere le ulteriori funzioni di trasformazione, a partire dalla base di dati prodotta nell'ambiente di gestione, nonché quelle di preparazione dei prodotti informativi destinati alla diffusione.

Tipicamente questi ultimi possono essere costituiti da:

- banche dati accessibili da utenza esterna, anche via web;
- strumenti di interrogazione facilitata dei dati, preferibilmente con meccanismi di accesso dinamico.
- supporti automatizzati di consultazione archivi (CD-ROM, DVD, ... );
- report e pubblicazioni varie;

Tutti gli ambienti di diffusione elettronica sono realizzati in piena coerenza con le metodologie e le tecniche per il rispetto del segreto statistico.

### **Altri sistemi collegati al sistema statistico**

Un sistema informativo statistico può fare riferimento a *metainformazioni* (definizione e significato dei dati) e a *dati di thesauro* (classificazioni, nomenclature, repertori, etc.) che fanno parte di un patrimonio collettivo e condiviso dell'Istituto e per i quali vengono realizzati appositi sottosistemi informativi.

I *sistemi di dati comuni* sono invece costituiti dall'insieme dei sottosistemi che producono dati di interesse comune come, ad esempio, i dati sulle strutture territoriali per la georeferenziazione dei fenomeni, la produzione di liste per il campionamento, l'archivio dei rilevatori e così via.

## **1.2 ARCHITETTURA DEI SISTEMI GESTIONALI**

Il sistema informativo gestionale e direzionale dell'Istituto, nel suo complesso, è imperniato fondamentalmente su:

- il Sistema Informativo Gestionale e Direzionale (SIGED), che verrà dismesso nel corso del 2011;
- il nuovo Sistema del Personale (URBI), che verrà messo in esercizio nel corso del 2011;
- il Sistema di Ragioneria, Contabilità, Inventario e Patrimonio (LIBRA);
- il Sistema di Programmazione e Controllo (LIBRA-PC);
- il Sistema di Rendicontazione Progetti Finanziati (TIMESHEET);

Il SIGED è un sistema informativo gestionale acceduto da oltre 400 utenti su tutto il territorio nazionale, che mantiene dati ed informazioni relative alle aree Personale e Diffusione e, in misura minore, all'area Provveditorato. Relativamente all'area Personale il sistema gestisce: l'inquadramento giuridico ed economico-fiscale dei dipendenti e dei collaboratori a vario titolo dell'Istituto, le presenze, le assenze, la formazione e le missioni, il calcolo degli stipendi nonché gli adempimenti previdenziali e di quiescenza del personale. In relazione al calcolo della prestazione lavorativa le presenze del personale sono rilevate da lettori di badge a banda magnetica e riversate nel sistema tramite una procedura on line.

È un sistema client-server, il lato server è costituito da una piattaforma Linux che ospita il database Oracle, con procedure e schedatore dei processi sviluppate in linguaggio C. Il lato client è costituito da procedure sviluppate in linguaggio Visual Basic le quali gestiscono l'interazione dell'utente col sistema, eseguono cioè procedure locali o richiamano le procedure in C presenti sul server. Di tale software sono disponibili i sorgenti, sia C che Visual Basic, sui quali viene effettuata una regolare manutenzione correttiva ed evolutiva da parte dell'Istituto (Direzione Generale).

URBI è il sistema che sostituirà nel corso del 2011 tutte le funzioni attualmente svolte dal SIGED. È un sistema interamente web-based, utilizza una base dati Oracle ed è stato acquisito dall'Istituto come software applicativo a pacchetto.

LIBRA, acceduto da oltre 100 utenti su tutto il territorio nazionale, copre l'area funzionale di gestione e controllo di gran parte dei provvedimenti di natura contabile, finanziaria e fiscale dell'Istituto. In particolare permette la gestione della Contabilità Finanziaria, la Contabilità IVA, la Gestione dei Clienti e Fornitori, la Contabilità Generale e Separata, le stampe di Contabilità Finanziaria e Generale, la gestione dell'inventario e del patrimonio dell'Istituto, la gestione dei magazzini modelli, cancelleria e stampati.

È un sistema client-server, utilizza una base dati Oracle ed è stato acquisito dall'Istituto come software applicativo a pacchetto.

LIBRA-PC, acceduto da oltre 200 utenti su tutto il territorio nazionale, copre l'area funzionale di programmazione e controllo della attività dell'Istituto.

È un sistema client-server, utilizza una base dati Oracle ed è stato acquisito dall'Istituto come software applicativo a pacchetto.

TIMESHEET, copre l'area di rendicontazione dei progetti a finanziamento esterno (es. Eurostat, ministeri, ecc.) che impegnano risorse di ruolo o a tempo determinato.

È un sistema interamente web-based, utilizza una base dati SQL Server, è stato sviluppato dall'Istituto con l'ausilio di risorse informatiche esterne. Il linguaggio di sviluppo è VB .NET, sono disponibili i sorgenti e viene effettuata una regolare manutenzione correttiva ed evolutiva da parte dell'Istituto (Direzione Generale).

SW MINORI. L'Istituto ha inoltre sviluppato per proprio conto, con tecnologia Visual Basic, .NET, C#, alcuni software di minore importanza:

- interrogazione dei dati personali del dipendente sulla Intranet;
- elenco telefonico sulla Intranet;
- estrazione dati e reportistica di dati presenti nel sistema del personale;
- software di integrazione fra sistemi (es. Sigid-Timesheet, Sigid-Libra)

Di tali software sono disponibili i sorgenti e viene effettuata una regolare manutenzione correttiva ed evolutiva da parte dell'Istituto (Direzione Generale).

## **2. GLI AMBIENTI TECNOLOGICI**

### **2.1 SISTEMI DI ELABORAZIONE**

L'ISTAT ha sette sedi a Roma, a cui si aggiungono 18 sedi regionali.

L'architettura dei sistemi attualmente implementata in Istat è costituita da una rete di server e di Personal Computer che prevede la specializzazione dei server per funzione. L'Istituto sta procedendo ad una razionalizzazione delle risorse che si svilupperà concentrando le attività elaborative su due Data Center, Balbo e Oceano Pacifico, secondo due linee parallele che vedono il consolidamento dei Server (Server Farm) e il consolidamento delle SAN (Storage Area Network).

In ciascuna delle due sedi Istat esistono quindi più server con funzioni specializzate. In particolare sono state individuate le seguenti tipologie di server:

- Database Server
- Application Server
- Server di gestione

In totale il sistema informatico Istat comprende circa 130 server distribuiti sulle sedi con i seguenti sistemi operativi:

- Windows server 2003 e 2008
- Linux Red Hat Enterprise V. 5.5

Inoltre in ciascuna delle sedi regionali sono installati due server per la gestione locale delle risorse Database Server contengono gli archivi dei dati dell'Istituto; oggi il software che gestisce i DB (DBMS Data Base Management System) è sulla gran parte dei server Oracle, anche se si sta sperimentando un DB open-source, MySQL, oggi utilizzato soprattutto in alcune applicazioni Web. L'accesso ai DB Oracle è effettuato sia attraverso prodotti software della Oracle stessa (SQLplus e simili), sia attraverso strumenti web.

Gli Application Server ospitano gli ambienti di sviluppo, sia in ambiente Web (Java o PHP) che in ambiente tradizionale come SAS ed altri linguaggi di III° generazione.

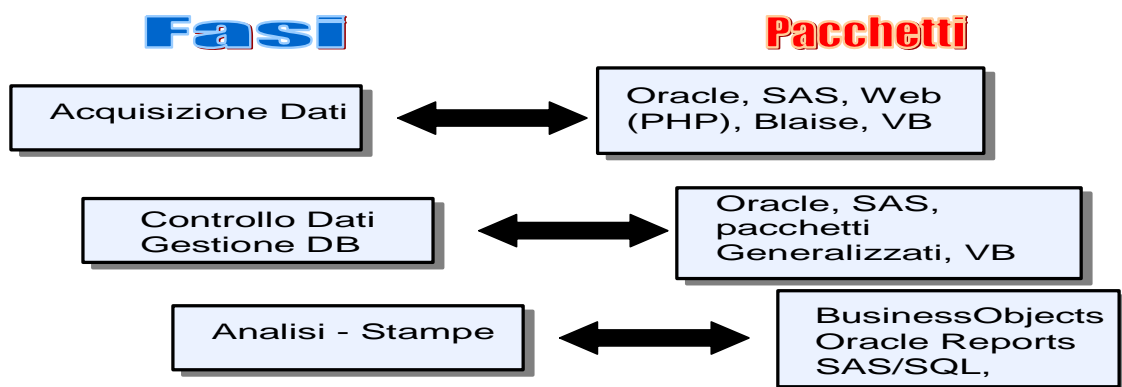
Le postazioni di lavoro sono costituite da PC con sistema operativo Windows XP professional e Windows 7 e pacchetti Windows Office professional nella versione 2003, 2007 e 2010.

### **2.2 GLI STRUMENTI DI SVILUPPO SOFTWARE**

#### **Situazione attuale**

Per le Indagini Statistiche vengono presentati gli ambienti di sviluppo attualmente utilizzati in ISTAT. Si è proceduto inoltre ad una classificazione degli stessi secondo le funzioni di trattamento dati descritte al punto 1.





Per quanto alle altre tipologie di Sistemi Informativi possiamo segnalare i seguenti Strumenti Software :

- SAS : ACCESS, CONNECT, SHARE (per la connettività), Base; AF; SCL; Macro; FSP; GRAPH (per lo sviluppo), SAS/ASSIST, Base (per l'utente), SAS/Warehouse Administrator (per data warehousing);
- Ambiente WEB : Oracle/Web, SAS/IntrNet, - HTML, XML, JAVA, JSP ;
- Ambiente client per architetture client/server: PHP, VisualBasic, HTML, Oracle e relativi ambienti di sviluppo;
- Office Automation : MS-Office, VisualBasic, ecc.
- Modellazione dati e funzioni: prodotti della Computer Associates per la modellazione dei dati logico/fisico e dei processi ErWin e BpWin;
- Gestione della Configurazione: Software Change Manager (Harvest) prodotto della Computer Associates per la gestione del change e versioning del software.

### Prospettive future

In riferimento alla situazione attuale, l'orientamento dell'Istituto sulle tecnologie di sviluppo software è il seguente:

- **SAS:** l'Istituto, considerata l'eccessiva dipendenza dei sistemi informativi di molte indagini e la politica di "licensing" portata avanti dalla SAS, anche in seguito a sollecitazioni in tal senso provenienti dal DigitPA (già CNIPA), ha deciso di ridurre la dipendenza dai pacchetti SAS. Tale riduzione avverrà nel corso di qualche anno e si baserà su una strategia mista:
  - evitare di sviluppare nuove applicazioni con pacchetti SAS: dal vincolo saranno esclusi solo aggiornamenti minori di applicazioni già rilasciate, nel caso che la riscrittura comportasse costi eccessivi.
  - sostituire SAS con altri prodotti specialmente nelle attività correnti di produzione istituzionale delle informazioni statistiche periodicamente diffuse dall'Istituto
  - sostituire SAS nei software generalizzati che ne fanno uso, adottando al suo posto il pacchetto **Open Source "R"**
  - **Software generalizzato:** l'Istituto sta compiendo un grosso sforzo di standardizzazione che ha portato allo sviluppo di una serie di pacchetti generalizzati che coprono diverse fasi della produzione del dato statistico (dall'Acquisizione, al Controllo e Correzione, alla Diffusione). In tutti i progetti si dovrà prima di tutto verificare se le funzioni richieste sono presenti in qualche pacchetto generalizzato, procedendo anche a integrazioni e miglioramenti dei pacchetti stessi, piuttosto che a sviluppi ex-novo.

- **Web:** l'Istituto sta supportando un sempre maggiore utilizzo delle architetture Web, sia nelle applicazioni destinate agli utenti interni, che in quelle destinate agli utenti esterni (altri enti, imprese e famiglie). Sarà dunque favorito lo sviluppo di progetti in architettura Web che consentano di minimizzare l'installazione di software sui client. Per quanto riguarda le diverse architetture Web, sarà favorito lo sviluppo di progetti che facciano uso di architetture "open" (PHP, Java, AJAX) che siano compatibili con le scelte architettureali (Linux, OpenSource) fatte dall'Istituto.
- **Open Source:** in tutte le architetture saranno favoriti gli strumenti di sviluppo "open" e, salvo eccezioni, non saranno accettati progetti che facciano uso di tecnologie proprietarie (VB, .Net) che vincolano l'Istituto a dipendere dalle scelte di fornitori hardware o software. A parziale eccezione saranno accettati progetti sviluppati con la piattaforma applicativa della Oracle, in quanto DB standard dell'Istituto e in quanto aderente ad architetture aperte come J2EE.
- **Database:** l'Istituto sta compiendo uno sforzo di razionalizzazione degli applicativi esistenti che consenta una migliore gestione del patrimonio applicativo esistente. Condizione necessaria per raggiungere questo obiettivo è l'utilizzo, in tutte le applicazioni, della tecnologia Data Base.
- **Software Configuration Management (SCM):** l'Istituto si sta dotando di uno strumento di SCM per migliorare la conoscenza e la gestione del patrimonio applicativo. Tutti i progetti sviluppati (internamente ed esternamente) dovranno integrarsi con il prodotto selezionato, alimentandone opportunamente gli archivi secondo gli standard che saranno stabiliti. Tale attività consentirà tra l'altro di aumentare il riuso delle applicazioni simili sviluppate in diverse indagini.

## **2.3 LA RETE ISTAT**

La Sede di Via Ravà n. 150 a breve sarà trasferita presso la Sede di Viale Oceano Pacifico n. 171.

Sono operative le seguenti reti:

- Rete geografica WAN che comprende 13 uffici regionali dell'ISTAT, collegati con la sede centrale tramite rete IP-MPLS a 4 Mbits/s e altri 4 uffici regionali, collegati con la sede centrale tramite rete IP-MPLS a 20 Mbits/s;
- Rete metropolitana MAN IP-MPLS che comprende tutte le sedi distaccate di Roma collegate alla sede centrale tramite servizi in fibra GBE a 100 Mbits/s e per il CED principale e viale Oceano Pacifico a 300 Mb/s. Queste due ultime sedi sono collegate tra loro anche con un doppio collegamento DWDM a 1 Gbits/s. Fanno parte della rete MAN anche le sedi limitrofe alla sede centrale, che sono collegate a 100Mbit/s tramite Laser.

Attualmente tali collegamenti limitatamente alla parte di trasporto e ad eccezione dei collegamenti a 1+1 Gb/s tra le due sedi principali, sono gestiti dal sistema Pubblico di Connettività (SPC).

Presso ciascun ufficio regionale o sede metropolitana di Roma, compresa la sede centrale, è stata realizzata una rete locale LAN dotata delle seguenti caratteristiche tecniche: cablaggio in categoria 5E/6 a 100 Mbps. In ogni sede romana sono installate reti ethernet di tipo back bone collassato, con dorsali in fibra ottica a 1 Gbps full duplex, che collega tutti i posti di lavoro situati nei piani alla Sala Server di ciascun palazzo situata in luoghi opportunamente protetti e condizionati.

Il CED centrale di Via Balbo e la sede di viale Oceano Pacifico sono cablati in categoria 6A e fibra ottica OM3. Gli apparati attivi switch di piano sono in layer2 e mentre i centri stella sono realizzati con switch layer3.

La rete interna e le sottoreti dedicate ai server sono protette da un sistema firewall ubicato presso le due sedi principali dell'Istituto, che assicura anche i servizi di terminazione VPN e ispezione del traffico dati.

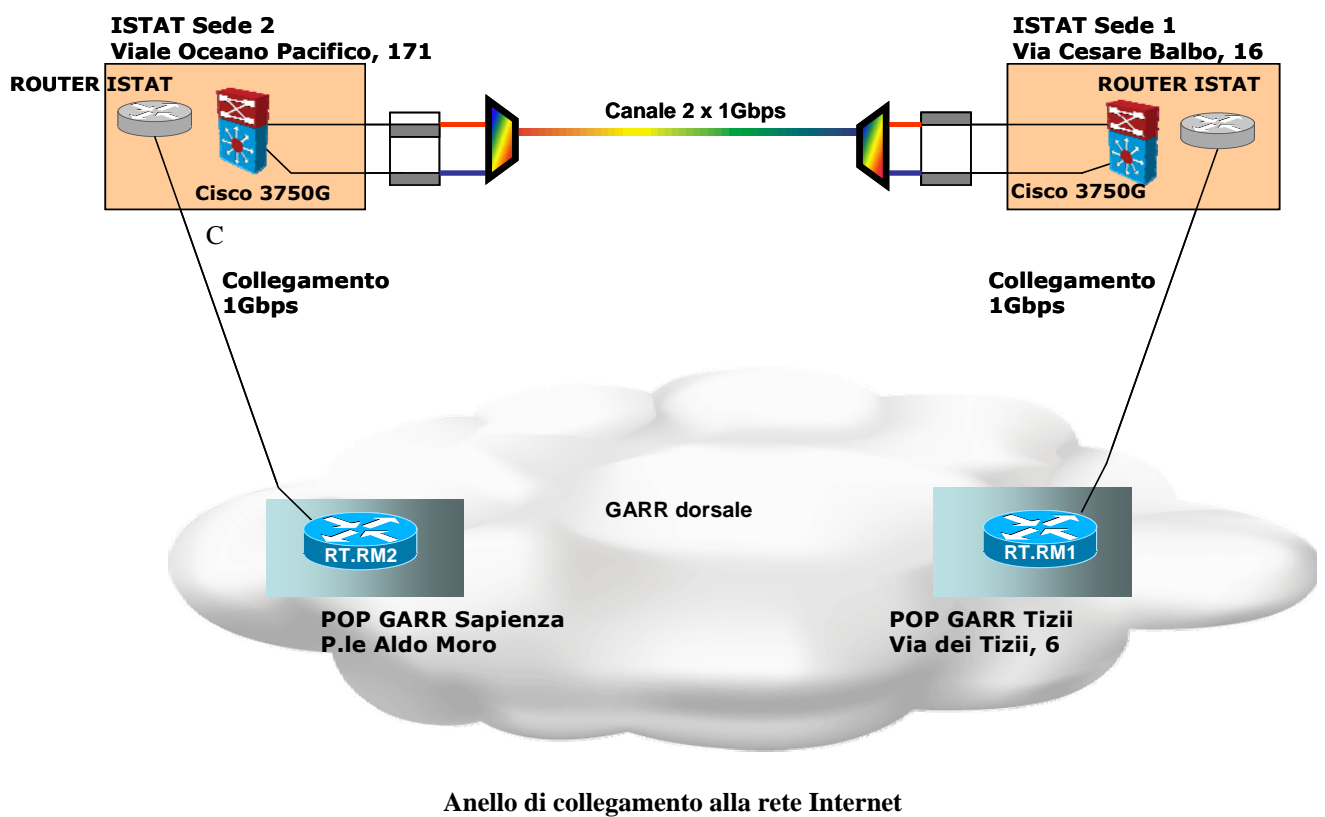
L'accesso ad Internet è assicurato tramite la rete GARR su rete di trasporto in fibra ottica TELECOM a 1 Gb/s sia dalla sede di Via Balbo che dalla sede di Viale Oceano Pacifico. La ridondanza è assicurata tramite il protocollo BGP.

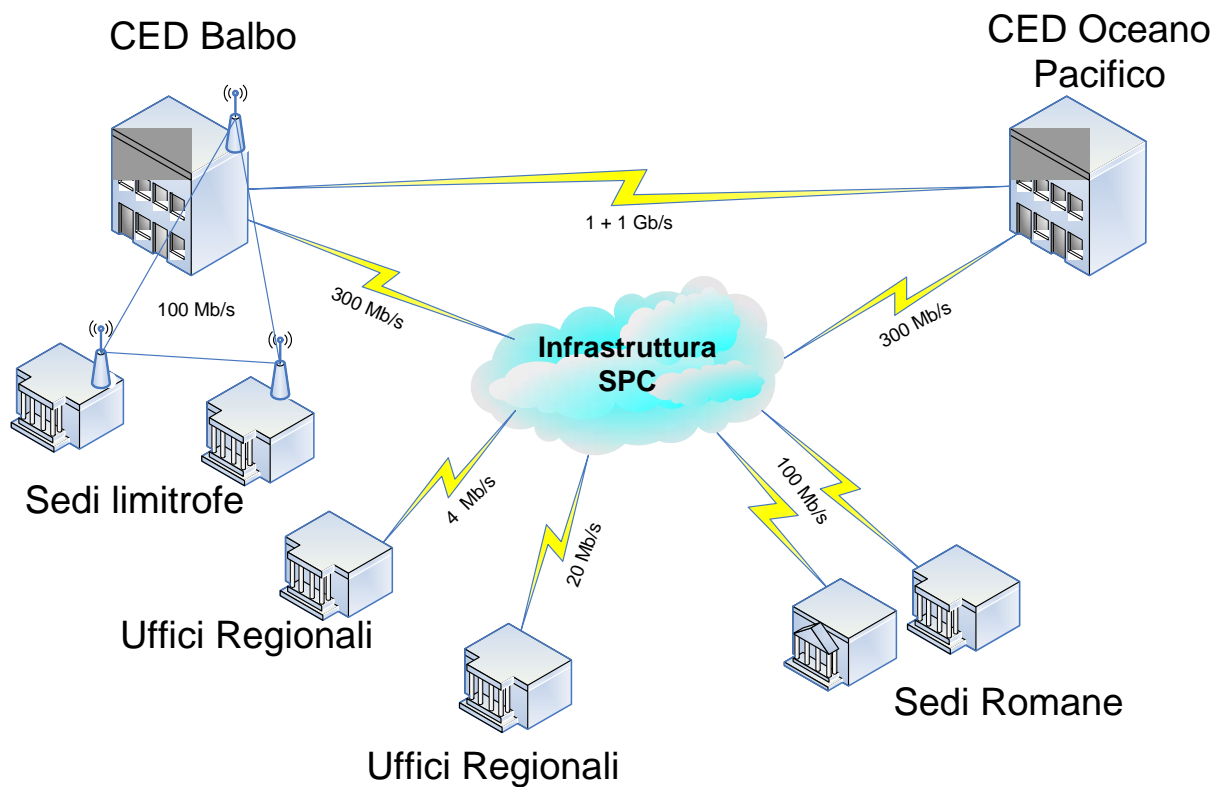
Per garantire la ridondanza dei collegamenti per i server del data center sono stati previsti due apparati di centro stella ossia apparati modulari ad elevate prestazioni in grado di processare un elevato throughput di traffico.

Infine, anche per i tratti di rete in fibra ottica, ogni server o apparato di storage ha a disposizione due collegamenti attestabili su due apparati distinti.

A livello di rete i centri stella svolgono anche funzioni di routing ed anche in questo caso è stata prevista la ridondanza dei gateway della rete, gestita tramite il protocollo VRRP.

A livello di routing l'uso del protocollo OSPF consente una gestione dinamica dei collegamenti della rete fornendo in modo automatico il riadattamento del routing in caso di caduta di un link.





**Struttura della rete ISTAT**