

# DOCUMENTI ISTAT

**n. 7/2009**

**Gli standard informatici dell'Istat**

*S. Losco*





---

## **Gli standard informatici dell'Istat**

Silvia Losco

ISTAT, Direzione centrale per le tecnologie e il supporto metodologico (DCMT)

Documenti Istat n. 7 /2009

# Documenti Istat

## Gli standard informatici dell'Istat

Silvia Losco, Istat, Direzione centrale per le tecnologie e il supporto metodologico (DCMT)

**Sommario:** Gli standard informatici sono istruzioni, direttive e raccomandazioni destinate a tutti i settori che si occupano d'informatica in termini di sviluppo di componenti software, di realizzazione di sistemi informativi e in generale di architetture di sistemi e di reti. Essi consistono in modelli, procedure e metodi da applicare in particolare a tutti gli ambiti di sviluppo e di progettazione di applicazioni informatiche all'interno dell'Istituto. Sono inoltre di riferimento all'elaborazione della pianificazione informatica, alla definizione delle strategie IT e alle acquisizioni informatiche dell'Istituto. L'adozione degli standard è da considerarsi fortemente consigliata a tutti i settori di sviluppo dell'Istituto e comunque vincolante per le società esterne di *outsourcing* fornitrici di servizi di sviluppo software.

**Parole chiave:** standard informatici, direttive, raccomandazioni, software, hardware

---

Le collane esistenti presso l'ISTAT - Contributi e Documenti - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT e del Sistan, o da studiosi esterni.

I lavori pubblicati Contributi Istat vengono fatti circolare allo scopo di suscitare la discussione attorno ai risultati preliminare di ricerca in corso.

I Documenti Istat hanno lo scopo di fornire indicazioni circa le linee, i progressi ed i miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.

## Indice

1. Introduzione .....	4
2. Il problema degli standard informatici .....	4
2.1 Il punto di vista dell'Istat .....	5
3. Il Comitato degli standard e le sue attività.....	5
3.1 Organizzazione dei lavori del Comitato.....	6
3.2 Classificazioni e articolazione degli standard .....	7
4. Standard tecnologici per lo sviluppo applicativo .....	7
4.1 Cenni sul Software generalizzato disponibile.....	10
4.2 Architettura complessiva del sistema.....	12
4.3 Le tecnologie standard in uso.....	13
4.3.1 I Data Base .....	13
4.3.2 Tecnologie OLAP .....	15
4.3.3 Tecnologie per l'integrazione .....	15
4.3.4 I linguaggi per il WEB .....	16
4.3.5 Strumenti di sviluppo .....	17
5. L'acquisizione e la codifica dei dati .....	18
5.1 L'architettura per la cattura dati via WEB .....	19
5.2 La rilevazione con tecniche CATI .....	22
5.3 La rilevazione con tecnica CAPI .....	22
5.4 La rilevazione con tecnica CADI.....	23
5.5 Acquisizione secondaria .....	25
6. Controllo e correzione.....	26
7. Produzione di stime e tabulazioni .....	27
8. Produzione di file standard e Data Warehouse .....	28
9. Sistemi informativi trasversali e sistemi di supporto allo sviluppo .....	29
9.1 Sistemi informativi trasversali .....	30
9.2 Strumenti di gestione dei progetti e di supporto allo sviluppo del software.....	31
Riferimenti .....	35

## 1. Introduzione

Gli standard in ambito informatico rappresentano una base di riferimento, un paradigma codificato per la produzione di tecnologie fra loro compatibili e interoperabili. La definizione di standard e il loro rispetto ha notevole importanza in ambito ICT ma in generale in ogni altro ambito anche non strettamente tecnologico.

In ambito ICT il problema assume particolare rilevanza sia per quanto riguarda le componenti hardware, sia per il software, e sia per le infrastrutture di rete.

Il presente documento illustra le attività del Comitato degli standard informatici dell'Istituto nato dall'esigenza di definire gli standard tecnologici di riferimento dell'Istituto; vengono illustrati i risultati conseguiti dal Comitato e una sintesi di quanto riportato sia nella relazione finale del Comitato che nei documenti dei diversi sottogruppi in cui il Comitato si è strutturato in merito alla definizione degli standard per quanto concerne lo sviluppo delle applicazioni in Istituto e per quanto riguarda l'infrastruttura di base di riferimento.

## 2. Il problema degli standard informatici

Gli standard informatici per lo sviluppo del software costituiscono un insieme di indicazioni tecniche dettagliate usate per stabilire criteri di uniformità nello sviluppo del software.

In generale, la standardizzazione è necessaria per permettere in primo luogo la *compatibilità* tra prodotti diversi. Più in generale costituiscono un fondamentale *punto di governo* che guida l'elaborazione della pianificazione della strategia informatica, la definizione dei sistemi informatici, la definizione delle architetture e anche gli acquisti in materia ICT.

Gli standard sono istruzioni, direttive destinate a tutti coloro che sviluppano software siano essi soggetti interni (strutture dell'Istituto delegate allo sviluppo del software) che soggetti esterni intesi come società incaricate di sviluppare software su mandato dell'Istituto stesso.

In generale consistono in modelli, procedure, metodi e tecniche che si applicano allo sviluppo di sistemi informativi e alle applicazioni dell'Istituto.

Lo sforzo posto nella definizione puntuale di standard di riferimento costituisce per l'Istituto il primo passo per uno sviluppo efficiente delle applicazioni a garanzia di compatibilità e interoperabilità.

Un grado elevato di standardizzazione permetterebbe, inoltre, di ridurre la complessità crescente dell'architettura informatica dell'Istituto avviando una conseguente razionalizzazione della spesa e degli sforzi interni in termini di gestione e manutenzione dei sistemi e formazione del personale. Implicitamente infatti vengono favoriti da una parte lo sviluppo di sinergie interne che favoriscono la cooperazione e dall'altra il riuso non solo di componenti software e hardware ma anche di conoscenze.

Generalmente il termine standard è usato in ambiente informatico, mentre il resto dei settori tecnologici usa il termine norma.

Diversi enti a livello internazionale come l'ISO (International Organization for Standardization) e l'IEEE (Institute of Electrical and Electronics Engineers) propongono, concordano e ratificano le norme nei diversi ambiti.

Per tali organizzazioni, uno standard prima di essere considerato tale dalla comunità internazionale, ed essere preso a buon diritto come modello di riferimento, passa attraverso una serie di fasi di analisi e accreditamento che sembra utile in questo contesto riassumere:

L'analisi delle esigenze dell'utenza che conduce chi si occupa di ricerca e sviluppo alla ricerca di soluzioni per i problemi e per le necessità eventualmente riscontrate.

Le specifiche tecniche vengono emesse sotto forma di descrizioni documentate estremamente dettagliate.

Il test e l'utilizzo di tali specifiche da parte della comunità internazionale dei produttori e dei laboratori di ricerca evidenzia le soluzioni migliori.

Le specifiche finali vengono accreditate come standard internazionale da un ente scientifico. Il risultato è un documento che descrive il modello cui le ditte di settore dovranno attenersi, pena l'incompatibilità dei loro prodotti tecnologici.

Di seguito si descrive l'approccio dell'Istat che, prendendo spunto dagli approcci noti in letteratura, ha affrontato il problema della definizione degli standard tecnologici per lo sviluppo del software e per la definizione dell'infrastruttura di base di supporto alla produzione della statistica ufficiale e dei sistemi amministrativi e gestionali dell'Istituto.

## **2.1 Il punto di vista dell'Istat**

In Istat i settori delegati alla produzione di software di supporto alla produzione statistica, hanno spesso orientato le proprie attività nell'ottica di assecondare principalmente il livello di coscienza diffuso nel singolo settore, al fine di ottimizzare l'utilizzo di risorse umane destinate allo sviluppo stesso presenti nella singola struttura di produzione.

Negli anni si è quindi consolidata e diffusa la tendenza a effettuare scelte autonome anche in ottica di originalità.

Da una parte tale differenza di scelte per lo sviluppo del software, effettuata nei diversi settori di produzione dell'Istituto, ha portato a incoraggiare la ricerca di soluzioni innovative per le piattaforme software da adottare, dando grande impulso alla ricerca e arricchendo l'offerta interna di soluzioni.

Dall'altra parte ha invece allontanato negli anni da una politica di utilizzo uniforme delle piattaforme hardware e software con evidente rischio di aumentare la complessità dell'infrastruttura di base dell'Istituto e con inevitabili ricadute sui costi di gestione complessivi in termini sia di spesa viva per l'approvvigionamento di licenze software e piattaforme hardware diverse da aggiornare e mantenere, che di costi indotti calcolabili in termini di necessità di formazione e specializzazione diversificata sulle diverse piattaforme adottate.

In tale contesto è nata la necessità di giungere alla definizione di standard informatici da introdurre e adottare in Istituto in modo condiviso e unitario che tengano ferme le esigenze dei settori tecnici di produzione statistica ma che costituiscano punto di fondamentale convergenza per tutti i nuovi sviluppi e investimenti.

## **3. Il Comitato degli standard e le sue attività**

Il "Comitato delegato all'individuazione degli standard applicativi dello sviluppo del software" è stato creato con deliberazione n. 27/07/DPIS-DCMT; il gruppo, composto da tre coordinatori e venti membri<sup>1</sup>, è stato incaricato di "procedere all'individuazione degli standard tecnologici per lo sviluppo applicativo del software, con la finalità principale di individuare gli strumenti di sviluppo di riferimento uniti agli ambienti e servizi tecnologici di supporto (RDBMS, Web Services) che costituiscono riferimento certo per l'utenza interna e vincolo stringente per i fornitori esterni".

---

<sup>1</sup>Coordinatori: Giulio Barcaroli, Silvia Losco, Carlo Vaccari;

Membri : A. Guarino, D. Moscara, R. Balestrino, D. Casale, A. Toma, L. Fanfoni, N. Fazio, D. Pagliuca, F. Spagnuolo, R. Torelli, C. Albergamo, A. Laureti Palma, V. Patruno, S. De Francisci, G. Rotondi, E. Calamani, C. Pellegrini, M. Cappella, T. Tumminello, L. Rocchi.

Sin dalla fase di avvio tutti i membri del comitato si sono trovati d'accordo nell'identificare il problema degli standard in Istituto come un problema con radici storiche e comunque serio impedimento allo sviluppo omogeneo e alla manutenibilità e gestibilità dell'infrastruttura complessiva. A partire da questo presupposto, è stato avviato un intenso programma di incontri di analisi tesi a individuare le modalità di approccio più utili ed efficaci per l'Istituto.

In particolare, il quadro che è emerso fin dai primi incontri, è quello di un ambiente in cui la consapevolezza della necessità di standard di riferimento inizia a diffondersi, anche se solo tra gli addetti ai lavori. Il problema rimane invece di poco interesse da parte dei settori a natura più strettamente statistica .

### **3.1 Organizzazione dei lavori del Comitato**

La complessità del tema e le difficoltà riconosciute che l'Istituto ha di adeguarsi a norme e standard generali, in relazione alle peculiarità e complessità dei singoli ambiti di riferimento, ha orientato le attività del comitato su azioni che consentano:

- di accrescere il livello di consapevolezza del problema;
- di esaminare ed evidenziare i processi, le applicazioni e i sistemi informativi in produzione che costituiscono per l'Istituto “buona pratica”;
- di analizzare e considerare prioritarie le tecnologie Open Source come scelte che garantiscano possibile indipendenza dell'Istituto dal mercato, e comunque di affiancare sempre a soluzioni di mercato tecniche Open Source di supporto allo sviluppo

In merito all'ultimo punto, si fa presente che il software open source, più facilmente reperibile, divulgabile e riutilizzabile in realtà differenti, acquista sempre più interesse per l'Istituto. L'ottica di codice aperto permette infatti di sviluppare soluzioni che non sono dipendenti da un unico fornitore o da un'unica tecnologia proprietaria, conducendo a piattaforme e tecnologie di sviluppo standard.

L'organizzazione del Comitato ha fatto leva sull'analisi dell'ambiente di produzione statistica e in particolare del processo di produzione del dato statistico. Si è deciso pertanto di studiare e definire gli standard in relazione ai diversi ambiti di riferimento propri dell'ambiente di produzione dell'Istituto; ha pertanto deciso di affrontare buona parte dei lavori diviso in sottogruppi<sup>2</sup> che affrontassero le tematiche specifiche e ne sviluppassero ampiamente l'analisi.

Tutti i risultati sono confluiti e sono stati discussi all'interno del Comitato stesso che ha approvato i singoli contributi in riunioni plenarie. I sottogruppi hanno affrontato le singole aree e hanno prodotto un'approfondita documentazione tecnica a cui si rimanda per l'approfondimento dei singoli temi.

---

<sup>2</sup> I sottogruppi definiti nell'ambito del Comitato:

**Sottogruppo 1 : Infrastruttura: standard di software di base e reti**

Membri : Guarino, Losco, Moscara

**Sottogruppo 2: Acquisizione dati**

Membri : Balestrino, Casale, Guarino, Losco, Moscara, Toma

**Sottogruppo 3: Controllo e Correzione**

Membri : Barcaroli, Fanfoni, Fazio, Pagliuca, Spagnuolo, Torelli

**Sottogruppo 4 : Stime e tabulazione**

Membri : Albergamo, Barcaroli, Laureti Palma, Pagliuca, Patruno

**Sottogruppo 5: File standard e DW**

Membri : Barcaroli, De Francisci, Laureti Palma, Patruno, Rotondi

**Sottogruppo 6: Sistemi informativi, trasversali e amministrativi**

Membri : Calamani, De Francisci, Pellegrini, Toma, Vaccari

**Sottogruppo 7: Strumenti di progettazione, project management, workflow e monitoraggio**

Membri : Cappella, Casale, Spagnuolo, Pellegrini, Toma, Vaccari



Tutti i documenti sono stati redatti sulla base di un layout predefinito che ha guidato il tavolo di analisi da parte del sottogruppo delegato a esaminare il tema specifico, consentendo di generare report omogenei. Lo schema di riferimento, è stato diviso in sezioni. Di seguito si riporta evidenza delle sezioni costituenti lo schema generale:

#### ***Descrizione dell'aspetto/fase***

- Standard Tecnologici - 1. Linguaggi Web
- Standard Tecnologici - 2. Database e archivi
- Standard Tecnologici - 3. Software Statistico
- Standard Tecnologici - 4. Linguaggi di programmazione
- Standard Tecnologici - 5. Specifiche architetturali
- Standard Tecnologici - 6. Software generalizzato disponibile

### **3.2 Classificazioni e articolazione degli standard**

Nel presentare gli standard per lo sviluppo applicativo viene adottata un'articolazione orientata al contesto della produzione statistica. Molti degli interventi di sviluppo che possono determinarsi nell'ambito statistico si riconducono alle attività del ciclo di vita e di produzione del dato. Pertanto riguardano:

- Acquisizione dati
- Controllo e correzione
- Produzione di stime e tabulazioni
- Produzione di file standard
- Data warehouse.

Si è sentita inoltre la necessità di indicare standard per lo sviluppo di sistemi di carattere trasversale, indipendenti dal processo di produzione delle singole indagini. A tale classificazione fanno parte :

- Sistemi informativi e amministrativi
- Strumenti di supporto allo sviluppo applicativo.

Si ritiene utile fornire infine una descrizione generale dell'architettura della rete e dei sistemi, delle modalità della comunicazione verso il mondo esterno, dei vincoli e misure di sicurezza che tutte le applicazioni, per qualunque finalità sviluppate, dovranno rispettare.

## **4. Standard tecnologici per lo sviluppo applicativo**

Le tecnologie ritenute standard secondo i criteri evidenziati in precedenza vengono in questa sezione elencati sinteticamente secondo una strutturazione guidata dagli ambiti e dalle fasi del processo di produzione statistica.

Nella scelta dei prodotti definiti come standard, si è scelto di inserire sempre almeno un prodotto open source per ogni categoria, in modo da rispondere positivamente alle necessità espresse dalla direttiva "Stanca" (Direttiva 19 Dicembre 2003, Sviluppo ed utilizzazione dei programmi informatici da parte delle pubbliche amministrazioni, GU n. 31 del 7-2-2004).

Per quanto riguarda i criteri di scelta tra diversi i prodotti elencati, si è raccomandato di adottare, a parità di funzionalità offerte, quelli operanti in architetture Web, quelli open source e quelli supportati da DBMS relazionali.

Relativamente all'ambiente tecnologico di riferimento, si è considerato nello schema di sintesi la separazione tra architettura esposta e interna, nel senso di infrastrutture rivolte agli utenti e ai sistemi esterni l'Istituto, o infrastrutture per utenti e sistemi interni. All'interno di quest'ultimo ambito si sono distinte le architetture distribuite per sistemi client-server tradizionali, e le architetture web, per applicazioni basate su pagine dinamiche html, con thin client (browser) e dati centralizzati su dbms.

Per la descrizione del contesto tecnico e delle architetture di sistema attualmente operative in Istituto, si rimanda al documento prodotto dal sottogruppo 1<sup>3</sup> e alla sintesi esposta nel paragrafo sull'architettura. In particolare, si ricorda, per quanto riguarda le architetture web esposte, che lo standard di Istituto, per quanto riguarda il web server, è stato da tempo individuato in Apache.

E' utile anche in questo contesto precisare che lo sviluppo applicativo di procedure di produzione deve in linea generale far riferimento a Linux come sistema operativo. Da tempo l'Istituto ha infatti indicato i server Linux come l'ambiente di produzione per eccellenza. Detto questo, possono esistere procedure che in tutto o in parte richiedono Windows come ambiente elaborativo: ciò dipende dai requisiti di sistema di alcuni software generalizzati (Teleform, Blaise, ACTR, CANCEIS) i quali possono operare solo in tale ambiente. Per quanto riguarda i sistemi generalizzati prodotti all'interno dell'Istituto (CONCORD, MAUSS, GENESEES), questi sono stati realizzati nelle due versioni, Linux e Windows based.

Per le descrizioni di dettaglio delle tecnologie relative a ambienti, linguaggi e software generalizzati, si rimanda ai documenti prodotti dai sottogruppi tematici<sup>4</sup>. In questi documenti, oltre alle tecnologie qui elencate, sono descritte e contestualizzate le cosiddette "isole tecnologiche applicative", quegli ambienti cioè per cui sono adottate tecnologie non ritenute standard. Viene escluso dallo schema sintetico l'ambito relativo agli strumenti di progettazione, project management, workflow e monitoraggio. Per le scelte e le descrizioni di dettaglio di tali strumenti di supporto, si rimanda allo specifico paragrafo di riferimento e ai documenti prodotti dal sottogruppo n.7<sup>5</sup>.

Relativamente ai software DBMS, per tutti gli ambiti e contesti applicativi, i due database Oracle e MySQL sono da assumersi standard. Lo stesso dicasi per il "linguaggio" XML, compresi i linguaggi di

---

<sup>3</sup> A. Guarino, S. Losco, D. Moscara, *Infrastruttura: standard di software di base e reti*, Documento interno <http://wiki.doku.php/standard:documenti:architettura>

<sup>4</sup> R. Balestrino, G. Barcaroli, S. De Francisci, L. Fanfoni, S. Losco, C. Vaccari, *Relazione finale del Comitato degli Standard*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:relazione\\_finale](http://wiki.istat.it/doku.php/standard:documenti:relazione_finale)

R. Balestrino, D. Casale, A. Guarino, S. Losco, D. Moscara, A. Toma, *Standard per l'acquisizione dati*, Documento interno <http://wiki.doku.php/start?idx=standard>

G. Barcaroli, L. Fanfoni, L. Fazio, D. Pagliuca, F. Spagnuolo, R. Torelli, *Controllo e correzione dei dati*, , Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_3\\_correzione](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_3_correzione)

File: standard\_documenti\_template\_correzione\_sottogruppo3\_rev7.doc

C. Albergamo, G. Barcaroli, A. Laureti Palma, D. Pagliuca, V. Patruno, *Stime e tabulazioni*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_4\\_stime](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_4_stime)

G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno

[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)

File : template\_dwh\_sottogruppo5\_def3.doc

E. Calamani, S. De Francisci, C. Pellegrini, A. Toma, C. Vaccari, *Sistemi informativi, trasversali e amministrativi*, Documento interno

[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_6\\_sistemi\\_informativi\\_trasversali](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_6_sistemi_informativi_trasversali)

<sup>5</sup> M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno [http://wiki.doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki.doku.php/standard:documenti:sottogruppo_7_strumenti)

gestione, elaborazione e trasformazione basati su XML, dato il diffondersi di queste tecnologie specialmente nei sistemi di integrazione e nelle applicazioni che rendono disponibili “web services”.

I linguaggi di programmazione elencati nello schema sono organizzati per famiglie:

- Suite di sviluppo Oracle (10AS), che comprende i linguaggi PL/SQL e Forms, e gli strumenti di sviluppo di applicazioni (client-server, web).
- Java, che comprende anche le JSP e il linguaggio JavaScript
- PHP, per applicazioni web-based
- R, inteso come linguaggio per applicazioni statistiche
- C, inteso come linguaggio “General Purpose” per applicazioni in cui la performance è prioritaria.

I software e i sistemi generalizzati disponibili per i singoli ambito/fase sono citati nel presente documento e descritti e contestualizzati in dettaglio nei documenti prodotti dai sottogruppi.

La tavola seguente riassume gli standard per lo sviluppo applicativo definiti dal comitato. La tavola mette in evidenza per ogni fase di produzione del dato statistico l’ambiente di riferimento , inteso o come architettura esposta (Web) o come architettura interna (client/server, Web), i linguaggi organizzati per famiglie individuati come standard, e i software generalizzati in uso in Istat e ritenuti ormai standard.

**Tavola 1** : *Schema di sintesi degli standard per lo sviluppo applicativo*<sup>6</sup>

	<b>Ambiente di Riferimento</b>	<b>Linguaggi</b>	<b>SW Generalizzati</b>
<b>Acquisizione Dati</b>	Architetture esposte web (indata), Architetture interne distribuite	Suite Oracle, Java, PHP, C	Teleform, DATISTAT, exFolia, BLAISE, CsPro, LimeSurvey, ACTR
<b>Controllo e Correzione</b>	Architetture interne client/server e web	Suite Oracle, Java, PHP, C, R	CONCORD, CANCEIS, DIESIS
<b>Stime e Tabulazioni</b>	Architetture interne client/server e web	Java, PHP, C, R	SMOL, IMPAGINA, R (SURVEY, EVER), GENESEES-R
<b>File Standard e DW</b>	Architetture esposte web,interne client/server e web	Suite Oracle, Java, PHP, C, R	MU-ARGUS, TAU-ARGUS, SMOL, Foxtrot.META, Foxtrot.PD, Web.PD, Foxtrot.MD, Web.MD, Sdmx.ISTAT
<b>Sistemi Informativi</b>	Architetture esposte web,interne client/server e web	Suite Oracle, Java, PHP, C	

<sup>6</sup> Fonte : R. Balestrino, G. Barcaroli, S. De Francisci, L. Fanfoni, S. Losco, C. Vaccari, *Relazione finale del Comitato degli Standard*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:relazione\\_finale](http://wiki.istat.it/doku.php/standard:documenti:relazione_finale)

## 4.1 Cenni sul Software generalizzato disponibile

Di seguito si riporta breve descrizione dei software generalizzati citati nella tavola 1 e ritenuti standard nel loro utilizzo per gli ambiti specificati nello schema di sintesi. Per l'approfondimento dei sw generalizzati si rimanda comunque all'Osservatorio Tecnologico Software generalizzati per la produzione statistica (OTS), organizzato secondo il classico ciclo di produzione delle informazioni statistiche.<sup>7</sup>

BLAISE è la tecnologia standard di supporto alle indagini CATI sviluppato da Statistics Netherlands e già largamente diffuso presso gli altri Uffici Nazionali di Statistica. Attualmente ne è disponibile la versione 4.7. L'ambiente di produzione del sistema è composto dall'integrazione di tre moduli:

- il questionario elettronico,
- il sistema di schedulazione delle chiamate
- il sistema di indicatori di monitoraggio delle interviste.

Il sistema Blaise genera automaticamente un set limitato di indicatori, inerenti essenzialmente agli ultimi esiti cumulati e per intervistatore, nonché la distribuzione degli appuntamenti per fascia oraria e priorità.

CONCORD <sup>8</sup>(CONtrollo e CORrezione dei Dati) si compone di tre moduli:

SCIA (*Sistema di Controllo e Imputazione Automatica*): “esegue il controllo e la correzione di variabili qualitative secondo l'approccio probabilistico (solo per errori casuali). Per ciascun record errato il sistema prima individua il minimo numero di variabili da imputare e poi effettua l'imputazione garantendo azioni di imputazione coerenti con un predefinito insieme di regole di controllo”;

GRANADA (*Gestione delle Regole e ANALisi dei DATi*): “esegue l'imputazione di variabili qualitative e quantitative secondo l'approccio deterministico, ossia mediante l'applicazione di regole del tipo SE [condizione di errore] ALLORA [azione di correzione]. Mediante questo modulo è possibile eseguire anche il solo controllo dei dati (separazione in esatti ed errati) secondo regole di incompatibilità che ammettono operatori logici e aritmetici (e quindi valide per variabili qualitative e quantitative)”;

RIDA - (*Ricostruzione dell'Informazione con Donazione Automatica*): “esegue l'imputazione di variabile qualitative e quantitative mediante donatore di minima distanza. Operazioni propedeutiche sono:

la classificazione delle unità in esatte ed errate;

la loro registrazione in due file distinti;

la identificazione dei valori da imputare mediante un predefinito carattere (di errore)”.

BANFF è un software generalizzato per il controllo e la correzione dei dati (imputazione) per le variabili numeriche, continue e non-negative, sviluppato da Statistics Canada.

Sono implementate le seguenti funzionalità:

- definizione delle regole di consistenza;
- verifica della coerenza delle regole di consistenza;

---

<sup>7</sup> <http://www.istat.it/strumenti/metodi/software/osservatorio.html>

D. Pagliuca, R. Cianchetta, M. Broccoli, T. Buglielli, R. di Giuseppe, D. Zardetto, *L'Osservatorio Tecnologico per i Software Generalizzati (OTS) nel 2008*, Collana Documenti Istat, n°4/2009

<sup>8</sup> Tratto da G. Barcaroli, L. Fanfoni, L. Fazio, D. Pagliuca, F. Spagnuolo, R. Torelli, *Controllo e correzione dei dati*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_3\\_correzione](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_3_correzione)

- localizzazione degli errori;
- identificazione dei valori anomali;
- imputazione.

CANCEIS è un software generalizzato per il controllo e la correzione di variabili qualitative e quantitative sviluppato da Statistics Canada per il controllo e l'imputazione dei dati censuari, in primo luogo quelli relativi ai componenti delle famiglie.

DIESIS è stato sviluppato per il Censimento della Popolazione del 2001, in particolare per il controllo e la correzione delle variabili anagrafiche e di stato civile (dati con struttura gerarchica).

Il Sistema consente il trattamento congiunto di variabili quantitative e qualitative e il trattamento di dati anche con struttura gerarchica.

GENESEES<sup>9</sup> (GENERALISED software for Sampling Estimates and Errors in Surveys) “è un software che calcola e stampa in automatico le stime aggregate a un livello corrispondente al dominio di stima previsto, allo scopo principale di mostrare i relativi errori campionari. Per l'ottenimento di tutte le stime di interesse è possibile utilizzare i pesi finali, calcolati tramite la funzione di Riponderazione, che è applicabile in tutti i casi in cui esistono informazioni ausiliarie, espresse in termini di totali noti rispetto a variabili ausiliarie legate alle variabili di interesse.

Le modalità d'uso prevedono una interazione tra sistema ed utente (il responsabile dell'indagine) che, utilizzando le form dell'interfaccia AF, provvede a definire tutti i parametri richiesti”.

IMPAGINA “è un software per l'impaginazione di tavole statistiche, secondo gli standard tipografici dell'Istat prendendo come input la tavola prodotta da un'elaborazione precedente

Il sistema, consente, inoltre, la gestione di titoli, note, numerazione di pagina, la stampa di più tavole nella stessa pagina, la suddivisione della tavola in più pagine con ripetizione del titolo e i margini di stampa”.

EX-FOLIA “permette di trattare in modalità automatica fogli Excel utilizzati come questionari per l'acquisizione dei dati e trasmessi dai rispondenti; la funzione primaria è l'estrazione dei microdati presenti in un foglio excel e il loro inserimento in un database. L'estrazione dati avviene in ambiente web/Apache; l'applicazione è sviluppata interamente in PHP e si avvale di apposite classi originariamente reperite in ambito open source e adattate alle esigenze interne. I dati estratti dai singoli fogli di lavoro sono inseriti contestualmente in una tabella Oracle o Mysql che conterrà un record per questionario e a sua volta, qualora l'utente finale ne abbia esigenza, potrà essere successivamente esportata in un file sequenziale o foglio Excel”.

MU-ARGUS<sup>10</sup> “è un software che, dato un file di dati di indagine, permette: (a) di valutare il rischio di violazione della riservatezza, cioè la possibilità di collegare i dati ad unità nella popolazione; (b) di diminuire tale rischio effettuando opportune manipolazioni dei dati nelle variabili che potenzialmente permettono il collegamento (tipicamente, cancellazione o perturbazione di valori o aggregazione di modalità).

<sup>9</sup> Tratto da C. Albergamo, G. Barcaroli, A. Laureti Palma, D. Pagliuca, V. Patruno, *Stime e tabulazioni*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_4\\_stime](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_4_stime)

<sup>10</sup> Tratto da G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)

La modalità di funzionamento è la seguente: il responsabile dell'indagine definisce le variabili di collegamento nel file dei dati, e stima il rischio di violazione connesso alle diverse unità (famiglie, individui). Se questo è troppo elevato, procede con le manipolazioni dei dati, fino a far rientrare il rischio al disotto di una soglia definita accettabile”.

TAU-ARGUS “è un software che assume in input due diversi file: (1) il file dei microdati, con le associate descrizioni, oppure (2) il file dei diversi aggregati che si intende diffondere, con le descrizioni delle tavole. Nel primo caso, è previsto anche un passo di specifica delle tavole e di tabulazione.

Una volta definite, o lette, le tavole, queste sono analizzate per identificare le celle a rischio di violazione. Una volta identificate, queste sono sottoposte ad uno dei seguenti trattamenti: (a) soppressione con quattro possibili differenti metodi o (b) arrotondamento controllato. Segue un passo di ridisegno interattivo delle tavole modificate, ed un ulteriore controllo sulle tavole così ottenute. Alla fine, viene generato il file dei dati aggregati “sicuro” dal punto di vista della violazione della riservatezza”

SMOL (Statistical Multidimensional OLAP Light) “è una versione “leggera” degli strumenti OLAP di aggregazione ed analisi on line di dati in formato multidimensionale. Sviluppata inizialmente in ambiente SAS, è ora disponibile anche in tecnologia PHP. Permette di acquisire un modello dimensionale predefinito a partire da fonte generica e generare tavole statistiche di dati aggregati. Il software opera a partire dai file di dati e metadati di ARMIDA”.

FOXTROT.META, “partendo dall'archivio dei microdati validati, offre all'utente statistico la possibilità di definire tramite passaggi guidati una struttura logico-fisica dei dati, organizzati. Il software opera a partire dai file di dati e metadati di ARMIDA”.

SDMX.ISTAT consente di navigare trasversalmente sulle diverse banche dati esposte, caratterizzate nella loro versione originaria da una forte eterogeneità tecnologica, applicativa e di front-end, come un unico data warehouse corporate di tipo virtuale.

## **4.2 Architettura complessiva del sistema**

Per quanto riguarda l'infrastruttura di base, l'Istituto ha deciso di adottare due sistemi operativi:

- Linux RedHat Enterprise Linux per i server
- Microsoft Windows per le postazioni client (PC)

Esistono isole tecnologiche che usano altri sistemi operativi (si citano a titolo di esempio IBM AIX, Microsoft Windows 2003 Server e MacOS) usati in ambienti specifici per vincoli applicativi rilevanti o economie di scala. I server sono specializzati per funzione e suddivisi in Application server (su cui vengono installati web server come Apache e Tomcat), Database server (su cui si installano Oracle e/o MySQL) o file server. Normalmente i sistemi di storage sono esterni e collegati in fibra ottica alle Storage Area Network (SAN). Per garantire un'alta disponibilità del servizio tutti i componenti dei server ed il collegamento alla SAN sono presenti in numero doppio, in modo da sopperire ad un'eventuale interruzione di funzionamento di uno di essi (ridondanza). Per le applicazioni ritenute critiche viene adottata una configurazione in cluster dei server stessi. Al momento attuale non esiste uno schema di autenticazione unico per gli utenti della rete e dei servizi della rete, anche se è allo studio una soluzione di Single-Sign-on con utilizzo del protocollo LDAP. Nella zona esposta della rete dell'Istituto sono compresi tutti i Web server e i servizi accessibili sia da Internet sia da altri enti

attraverso il Sistema Pubblico di Connettività (SPC), o attraverso collegamenti diretti. A protezione di tutte le risorse esposte, in questa area è installato il sistema firewall che suddivide l'insieme dei sistemi in zone a protezione differenziata, su cui vengono applicate le politiche di sicurezza dell'Istituto. L'architettura della zona esposta è basata sulla presenza di server di Front-end nella zona pubblica, su cui vengono concluse tutte le comunicazioni provenienti dall'esterno, e server di Back-end (Application server e Database server) nella zona più protetta. Per quanto riguarda l'accesso alle applicazioni dell'Istituto da parte di personale "mobile" (FOL, Prezzi, Telelavoro) si utilizza una Rete Privata Virtuale (VPN) con autenticazione forte e protocollo sicuro che garantisce la cifratura di tutti i flussi trasmessi e ricevuti.

### 4.3 Le tecnologie standard in uso

La tavola 2 riassume le tecnologie ritenute standard per l'Istituto in termini di sistema operativo di base operante sulle piattaforme dei server e dei client, i Data Base in uso e i Web Server. Per una descrizione di dettaglio si rimanda alla documentazione prodotta dal sottogruppo che ha analizzato la tematica nell'ambito del Comitato<sup>11</sup>. In ultimo vengono citati i linguaggi di programmazione in uso.

**Tavola 2:** *Tecnologie Standard per l'Istituto*

<b>Sistema Operativo Server</b>	Linux
<b>Sistema Operativo Client</b>	Windows, Linux
<b>DBMS</b>	Oracle, MySql
<b>WebServer</b>	Apache, TomCat
<b>Linguaggi</b>	Suite Oracle, Java, Php, R, C

Di seguito si fornisce breve descrizione delle tecnologie adottate suddivise per tipologia, rimandando alla documentazione tecnica nota in letteratura, fornita comunque in bibliografia, per eventuali approfondimenti.

#### 4.3.1 I Data Base

##### **Oracle 10g**<sup>12</sup>

“È il DB commerciale ufficiale dell'Istituto, supporta tutte le componenti necessarie all'organizzazione dei dati, alla loro navigazione e accesso, e a tutte le attività di DWH.

Una delle componenti essenziali per rendere efficiente il processo di estrazione delle informazioni è l'algoritmo di indicizzazione, in Oracle sono supportati sia gli indici BTree che i Bitmap. Gli indici Bitmap sono preferiti negli ambienti di DWH in quanto occupano meno spazio degli indici Btree e consentendo quindi di realizzare, senza un utilizzo eccessivo di spazio disco, tabelle completamente indicizzate su singole colonne.

<sup>11</sup> A. Guarino, S. Losco, D. Moscara, *Infrastruttura: standard di software di base e reti*, Documento interno, <http://wiki/doku.php/standard:documenti:architettura>

<sup>12</sup> Tratto da G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)

Oltre all'ambiente relazionale ottimizzato, in Oracle è presente l'Analytic Workspace (AW) che opera una trasformazione fisica dei dati per una gestione ottima dell'informazione con specifici motori, sia per schemi ROLAP che per schemi MOLAP, quest'ultimo è supportato dal modulo Oracle OLAP che si avvale di un modello di dati multidimensionale a supporto di query analitiche complesse direttamente all'interno di Oracle Database 10 g.

L'AW consente di definire gli oggetti dimensionali in termini di misure, dimensioni, livelli, gerarchie e attributi al fine di mapparli in data source proprietari. L'operazione di trasformazione è applicata al livello logico attraverso l'ambiente analitico, e le funzioni di mappatura sono realizzate all'interno del Analytic Workspace Manager aggiornando l'Active Catalog (i metadati del workspace).

Gli strumenti di interfaccia proprietari Discoverer Plus OLAP e Spreadsheet Add-In utilizzano l'Active Catalog per effettuare le query nel workspace per le operazioni OLAP. Le regole di aggregazione identificano gli aggregati memorizzati per ogni misura, quando una applicazione interroga l'analytic workspace il sistema è in grado di recuperare una l'interrogazione se già calcolata o può calcolarla on the fly da un sottoinsieme di aggregati memorizzati. Quest'ultima funzionalità è estremamente performante per calcoli multidimensionali.

Oracle inoltre fornisce API Java con le componenti OLAP con cui è possibile definire direttamente in Java il modello dimensionale ed istanziarlo direttamente nell'*analytic workspace*. Le API implementano un modello ad oggetti multidimensionale e forniscono gli strumenti di istanziazione consentendo di definire le relazioni tra gli oggetti nel modello (le OracleBI Beans non sono incluse con la OLAP option).

C'è da segnalare infine che Oracle fornisce un semplice tool (Oracle Warehouse Builder) completo per la gestione dell'intero ciclo di vita (ETL) di dati e metadati".

## MySQL

“E' un DBMS relazionale, il codice di MySQL è di proprietà della omonima società, viene distribuito con la licenza TGU GPL oltre che con una licenza commerciale. MySql supporta sia la versione client/server che la configurazione embedded system. Il DBMS supporta l'SQL con un buon rispetto dello standard ANSI. MySql è distribuito con un interfaccia utente text-based, le interfacce utente web-based per MySql sono fornite da terze parti: il programma più diffuso e in uso in Istat per le funzioni di amministrazione con MySQL è phpMyAdmin. Inoltre, il linguaggio SQL di MySQL comprende numerose estensioni che sono tipiche di altri DBMS (PostgreSQL, Oracle e Sybase).

Le tabelle predefinite di MySql sono realizzate secondo la struttura ISAM (Indexed Sequential Access Method; metodo di accesso sequenziale indicizzato), questo metodo consente di immagazzinare dati, effettuare degli aggiornamenti e di definire indici su strutture separate per la gestione delle chiavi primarie ed esterne. Sono disponibili diversi criteri di ottimizzazione sui valori, non vengono supportati gli indici Bitmap mentre sono supportati gli indici Btree e Hash.

Da un punto di vista prestazionale, sono documentate esperienze di DWH MySql con oltre 200GB di informazioni elementari gestire, prevalentemente con tabelle sommarizzate e con prestazioni soddisfacenti. Nella versione 5.1 è stato introdotto il partizionamento delle tabelle, anche se una precedente funzione di marge su più tabelle risulta ancora largamente utilizzata. Gli strumenti di ottimizzazione sono disponibili per l'analisi del piano di esecuzione come per aggiornare le statistiche dell'ottimizzatore. E' presente una funzione benchmark per i test comparativi che fornisce i tempi di esecuzione delle funzioni da analizzare.

MySQL è scritto in linguaggi C e C++, è pertanto disponibile su i Sistemi Operativi in uso in Istat (AIX, GNU/Linux, Solaris, SunOS, Windows); sono inoltre presenti i driver di connessione per: ODBC, Java, .Net.”



### 4.3.2 Tecnologie OLAP

In letteratura sono descritte tre tipologie base di sistemi OLAP: Relazionale OLAP (ROLAP), Multidimensionale OLAP (MOLAP) e Hybrid OLAP (HOLAP).

- Il ROLAP utilizza il DB Relazionali come motore delle estrazioni ed è lo schema più utilizzato in Istituto, fa riferimento a strutture di dati basate su più schemi dimensionali sui diversi fatti d'interesse. Questa organizzazione delle informazioni si basa sostanzialmente sulle risorse del RDBMS per l'estrazione delle informazioni.

- Lo schema MOLAP utilizza un database di riepilogo che ha uno specifico motore per l'analisi multidimensionale, questo risulta particolarmente vantaggioso per DWH in cui il volume di dati non è molto elevato, indicativamente la performance ottima per uno schema MOLAP si ottiene per volumi di dati oggetto dell'analisi dell'ordine di grandezza o poco superiore dello spazio disponibile nella memoria primaria dell'elaboratore. I sistemi MOLAP non hanno tra le loro caratteristiche principali la scalabilità, sono statici e sono pensati espressamente per le funzioni di estrazioni delle informazioni.

- HOLAP utilizza tabelle relazionali per memorizzare i dati e strutture multidimensionali per le aggregazioni "speculative", in questo contesto le funzionalità del RDBMS per l'ottimizzazione delle estrazioni di informazioni sono utilizzate su strutture dati specializzate secondo modelli multidimensionali, con un intensivo utilizzo d'indici Btree e partizionamenti. Lo schema HOLAP si pone ad un livello di efficienza e scalabilità intermedio tra il ROLAP e il MOLAP; infatti rispetto al ROLAP risulta più efficiente nelle estrazioni a discapito della semplicità di definizione degli schemi di dati; l'HOLAP risulta invece più scalabile dello schema MOLAP pur non raggiungendo gli stessi livelli di specializzazione per le funzioni di estrazione dei dati.

### 4.3.3 Tecnologie per l'integrazione

#### XML<sup>13</sup>

«L'XML, acronimo di eXtensible Markup Language, ovvero «Linguaggio d'integrazione basato sulla marcatura estensibile» è un metalinguaggio creato e gestito dal World Wide Web Consortium (W3C) che permette di definire la grammatica di diversi linguaggi specifici derivati.

XPath è il linguaggio per la selezione e filtraggio di nodi nei documenti XML, utile per cercare contenuti all'interno di un file XML.

Tra le tecnologie attualmente e in uso in Istat viene evidenziata la famiglia di linguaggi XML non intesa solo ai sistemi web, ma in generale al trattamento massivo dei dati. La sua struttura gerarchica e l'utilizzo di marcatori definibili dall'utente lo rende adattabile a varietà di contesti ove devono essere rappresentati dati strutturati o semi-strutturati. Sui documenti XML è possibile eseguire, senza ledere alla compatibilità di quanto già sviluppato, diverse operazioni quali l'integrazione, la trasformazione, la validazione prescindendo dalla codifica dei singoli elementi. Queste caratteristiche hanno reso possibile la definizione di un certo numero di componenti standard da cui possono essere costruite applicazioni orientate all'utilizzo di documenti XML. L'XML presenta svariati vantaggi su altri formati di inter-scambio: supporta il formato Unicode o diversi formati alternativi di codifica per i caratteri rendendo i documenti altamente portabili ed in grado virtualmente di codificare la maggior parte dei dati che siano rappresentabili tramite stringhe. L'utilizzo di marcatori per delimitare i dati rende possibile la costruzione di archivi auto-esplicativi rendendoli non solo intelligibili dall'utente, ma anche

---

<sup>13</sup> Tratto da G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)

da applicazioni non inizialmente pensate per il loro utilizzo: un formato che è di fatto prevedibile, può consentire la lettura degli archivi da ogni tipo di programma, al di là dell'interpretazione della semantica che sottende il sistema di marcatori.

Tra gli svantaggi il più evidente è la dimensione degli archivi, che risultano tuttavia altamente comprimibili anche per la natura ripetitiva dei marcatori; in ogni caso, nella codifica di insiemi estesi di dati numerici rappresentati come stringhe l'archivio risultante sarà sempre di dimensioni superiori al corrispondente archivio in formato binario”.

La crescente disseminazione di dati semi-strutturati in formato XML ha reso indispensabile la definizione di modelli e di basi di dati particolarmente efficienti e scalabili per l'accesso organizzato ad una siffatta tipologia di archivi. L'approccio da seguire è sostanzialmente di due tipi: da una parte i sistemi per basi dati esistenti sono stati adattati per memorizzare dati esibiti nel formato XML utilizzando caratteristiche intrinseche o middleware esterni, in questa categoria referenziata come “XML-Enabled Databases” rientrano strumenti come Oracle 8i; alternativamente, sono stati disegnati Database Nativi XML come Oracle 10g). Nella versione 10g del database Oracle viene fornita a corredo la possibilità di memorizzare dati in formato nativo xml basandosi su uno specifico schema, che viene utilizzato per creare i collegamenti tra gli oggetti e la struttura di tipo relazionale: i documenti XML sono immagazzinati, opportunamente sezionati e mappati sulla struttura ad oggetti-relazionale, senza perdere la semantica dello schema originale. La totalità delle basi dati XML native non utilizza ancora un modello di dati comune; l'orientamento è comunque verso due tipi di approccio: l'utilizzo di liste di indici, che porta ad una esplosione dello spazio richiesto per le strutture o la memorizzazione separata con lo svantaggio che il contenuto di un documento è isolato in diverse tabelle e pertanto la sua ricostruzione completa diventa un'operazione computazionalmente onerosa. Un database XML-nativo è in grado di trattare l'intero spettro di sotto-linguaggi definibili in ambito XML: memorizzando tanto i collegamenti alle strutture, che l'ordine con cui l'informazione si presenta è possibile la ricostruzione di documenti scritti secondo il paradigma “document-centric”, che “data-centric”. L'XML ed i database nativi XML non rimpiazzeranno almeno nel futuro prossimo le basi dati di tipo relazionale, per cui uno degli aspetti cruciali sarà la possibilità di scambiare dati da e verso qualsiasi sorgente XML con sistemi di tipo RDBMS”.

## **WEB SERVICES<sup>14</sup>**

Web Service (servizio web) è un sistema software progettato per supportare l'interoperabilità tra diversi elaboratori su di una medesima rete; caratteristica fondamentale di un Web Service è quella di offrire un'interfaccia software (descritta in un formato automaticamente elaborabile quale, ad esempio, il WSDL (Web Services Definition Language) utilizzando la quale altri sistemi possono interagire con il Web Service stesso attivando le operazioni descritte nell'interfaccia tramite appositi "messaggi" scambiati tramite SOAP (Simple Object Access Protocol): tali messaggi sono, solitamente, trasportati tramite il protocollo HTTP e formattati secondo lo standard XML. Proprio grazie all'utilizzo di standard basati su XML, tramite un'architettura basata sui Web Service (Service oriented Architecture - SOA) applicazioni software scritte in diversi linguaggi di programmazione e implementate su diverse piattaforme hardware possono quindi essere utilizzate, tramite le interfacce che queste "espongono" pubblicamente e mediante l'utilizzo delle funzioni che sono in grado di effettuare (i "servizi" che mettono a disposizione) per lo scambio di informazioni e l'effettuazione di operazioni complesse.

### **4.3.4 I linguaggi per il WEB**

La tecnologia Web inizialmente nata su contenuti statici si è evoluta verso il contesto dinamico, la gestione delle informazioni visualizzate su un browser sono ora generalmente contenute in una

---

<sup>14</sup> Le definizioni sono tratte da Wikipedia (2009), <http://wikipedia>

base di dati e si parla comunemente di “Web e Basi di Dati” supportati da frameworks di sviluppo specifici (Eclipse, NetBeans, Zend Studio). In tale contesto, in Istat, sono adottati due linguaggi di programmazione per il Web: PHP, JSP.

Il PHP (PHP: Hypertext Preprocessor) è un linguaggio di scripting interpretato, con licenza open source, originariamente concepito per la realizzazione di pagine web dinamiche. Attualmente è utilizzato principalmente per sviluppare applicazioni web lato server, viene anche utilizzato per applicazioni standalone con interfaccia grafica.

JSP (Java Server Pages , letto anche talvolta come Java Scripting Preprocessor) è una tecnologia Java per lo sviluppo di applicazioni Web che fornisce contenuti dinamici in formato HTML o XML. Si basa su un insieme di speciali tag con cui possono essere invocate funzioni predefinite o codice Java. In aggiunta, permette di creare librerie di nuovi tag che estendono l'insieme dei tag standard. Le librerie di tag JSP si possono considerare estensioni indipendenti dalla piattaforma delle funzionalità di un Web server.

### 4.3.5 Strumenti di sviluppo

Come già evidenziato, l'analisi ha evidenziato soprattutto l'importanza e l'uso diffuso e ormai standard di strumenti di sviluppo e linguaggi di programmazione elencati a seguire organizzati per famiglie Suite di sviluppo Oracle (10AS), Java, PHP, R, C. Si cita inoltre Oracle Forms, parte di Oracle Developer Suite, come un ambiente Rapid Application Development (RAD) per sviluppare applicazioni per base dati, il linguaggio utilizzato è il PL/Sql. Consente di realizzare una interfaccia per la visualizzazione e l'editing di informazioni

L'utilizzo del linguaggio C può essere suggerito in particolari contesti applicativi e solo su specifiche esigenze di ottimizzazione.

#### **ECLIPSE**

**Eclipse**<sup>15</sup> è una open source community i cui progetti mirano a costruire una piattaforma per lo sviluppo aperto comprensiva di un "framework" espandibile, tools e moduli runtime per supportare, durante il ciclo di vita del software, la costruzione, l'implementazione e la gestione del software stesso e la gestione completa delle funzionalità del client CVS.

I progetti di Eclipse offrono tools e framework per tutto il ciclo di vita dello sviluppo dell' intero software, includendo la modellazione, lo sviluppo, tools di spiegamento, reportistica, manipolazione di dati, fase di test e profiling. I tools e i framework sono principalmente concentrati su come costruire JEE, i web services e applicazioni web. Eclipse offre anche supporto per altri linguaggi, come C / C++ PHP, ed altri.

#### **NETBEANS**<sup>16</sup>

“**Netbeans** (open source ): è un ambiente di sviluppo multi-linguaggio scritto interamente in Java da contrapporre al più diffuso Eclipse. La versione più recente è la 5.5.

A tutt'oggi risultano in uso i due prodotti: NetBeans IDE e NetBeans Platform.

NetBeans IDE è un ambiente di sviluppo – uno strumento destinato ai programmatori per scrivere, compilare ed eseguire il debug ed il deploy di programmi. E' scritto in Java ma può supportare qualsiasi linguaggio di programmazione. Esiste anche un gran numero di moduli utili per estendere le

---

<sup>15</sup> Per approfondimenti si faccia riferimento al sito <http://www.ossipa.cnipa.it>.

<sup>16</sup> Tratto da E. Calamani, S. De Francisci, C. Pellegrini, A. Toma, C. Vaccari, *Sistemi informativi, trasversali e amministrativi*, Documento interno, [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_6\\_sistemi\\_informativi\\_trasversali](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_6_sistemi_informativi_trasversali)

funzionalità di NetBeans IDE. NetBeans IDE è un prodotto gratuito senza alcuna restrizione riguardante il suo uso.

E' anche disponibile NetBeans Platform; una piattaforma modulare ed estensibile da usare come supporto per la creazione di grosse applicazioni desktop. Entrambi i prodotti sono open source e gratuiti".

## 5. L'acquisizione e la codifica dei dati

Nell'ambito della statistica ufficiale la funzione di acquisizione e di codifica dei dati riveste particolare importanza. Vista la notevole importanza del tema e la complessità della materia il sottogruppo<sup>17</sup> incaricato all'analisi degli standard per l'acquisizione e la codifica dei dati ha sviluppato ampiamente l'argomento, già per altro affrontato e gestito negli anni da strutture organizzative dedicate, e prodotto notevole documentazione di dettaglio a cui si rimanda per approfondimento<sup>18</sup>.

Nell'ambito dell'acquisizione dati è importante riportare la distinzione, riconosciuta nel contesto statistico ufficiale, tra acquisizione primaria, che si determina nella conduzione di indagini tradizionali e genera un evento di acquisizione per ogni intervista/modello di rilevazione, e acquisizione secondaria, con cui vengono acquisite raccolte di dati già disponibili a vario titolo presso fonti esterne.

Per l'acquisizione primaria le soluzioni software si caratterizzano diversamente a seconda della tecnica prescelta per la somministrazione del questionario:

CAPI (Computer Assisted Personal Interviewing); CATI (Computer Assisted Telephone Interviewing);

CAWI (Computer Assisted Web Interviewing), anche note come Web Survey;

CADI (Computer Assisted Data Inputing). L'acquisizione secondaria deve realizzare operazioni di Secure File Transfer.

Tradizionalmente le tecniche sono andate specializzandosi sulle diverse classi di rispondenti: CATI e CAPI sono indirizzate prevalentemente, ma non unicamente, alle famiglie; Web Survey esclusivamente a imprese e istituzioni.

Va evidenziato inoltre che, per necessità o comodità, i diversi strumenti, inclusi quelli tradizionali, spesso vengono associati nello stesso processo di acquisizione di una rilevazione statistica dando luogo alle tecniche note come mix-mode.

La tavola a seguire riassume gli standard individuati per la fase di acquisizione e codifica dei dati e in particolare evidenzia l'ambiente di riferimento per le fasi di acquisizione primaria e secondaria e per le fasi di codifica automatica ed assistita oltre, oltre ai linguaggi e ai software generalizzati ritenuti standard per questa fase.

---

<sup>17</sup> Sottogruppo per l'analisi degli standard in merito all'acquisizione dati. Membri : Balestrino, Casale, Guarino, Losco, Moscara, Toma

<sup>18</sup> Per approfondimenti si faccia riferimento a R. Balestrino, D. Casale, A. Guarino, S. Losco, D. Moscara, A. Toma, *Standard per l'acquisizione dati*, Documento interno, <http://wiki/doku.php/standard:documenti:acquisizione>  
R. Balestrino e A. Gaucci, *Tecniche di cattura nei processi di produzione statistica*, Collana Documenti Istat, n°2/2009.

**Tavola 3 :** Standard per la fase di *Acquisizione e Codifica*

	Acquisizione				Codifica		
	Primaria				Secondaria	Automatica	Assistita
	web	cati	capi	cadi			
<b>Ambiente di riferimento</b>							
Architetture esposte web	indata				indata, ftp server		
Architetture interne distribuite		X	X	X		X	X
<b>Linguaggi</b>							
Suite Oracle			X				
Java	X		X				
PHP	X		X				
<b>SW generalizzati</b>							
	Teleform, exFolia, LimeSurvey	Blaise	Blaise, CsPro	Blaise, CsPro	DATISTAT	ACTR	Blaise

### 5.1 L'architettura per la cattura dati via WEB

L'architettura definita standard per l'Istituto in merito alla cattura dati via Web è quella attualmente disponibile sul sito Web ufficiale per il data capturing, INDATA (<https://indata.istat.it>).

Il sito è dedicato alle operazioni di acquisizione telematica primaria e secondaria ed è disponibile per tutti i settori di produzione statistica interessati ad adottare soluzioni innovative basate su Web.

La piattaforma INDATA, evidenziata nella tavola 2, offre ai rispondenti la possibilità di compilare online il modello di rilevazione accedendo a un sito visibile sulla rete Internet predisposto con prerequisiti di alta affidabilità e disponibilità.

Su tale piattaforma è disponibile un ambiente WEB uniforme e omogeneo che consente l'accesso a questionari in formato ipertestuale o altro formato elettronico; i rispondenti possono compilare e inviare il questionario accedendo per mezzo del loro browser all'indirizzo del web server di Istituto.

La compilazione può avvenire online e i dati immessi dal rispondente sono contestualmente resi disponibili nell'ambiente di produzione che viene alimentato quindi con tempestività massima.

In alternativa alla modalità online, per la specificità dell'indagine statistica di riferimento o per la complessità del questionario viene data la possibilità al rispondente di effettuare la compilazione offline. Tale tipo di compilazione può essere efficace anche per ridurre i tempi di connessione del rispondente o per consentire una compilazione più agevole a fronte di questionari molto complessi, scaricando sulla propria postazione client il questionario elettronico in formato eseguibile, e, in una successiva sessione, di restituire il questionario compilato.

L'architettura di riferimento per l'acquisizione dei dati via WEB è ampiamente descritta nei documenti riportati in bibliografia. Si riportano in questa sede alcuni elementi da intendersi come requisiti e vincoli del sistema ritenuti utili alla descrizione degli standard di riferimento per tale fase del ciclo di vita del dato statistico.

### ***Requisiti funzionali***

Il sito *indata.istat.it* costituisce il punto unico in cui sono disponibili tutti i servizi, gli strumenti e le informazioni relativi alle indagini che realizzano cattura dati via Web

Lo stile di presentazione è omogeneo per qualsiasi indagine;

Il sistema è concepito in maniera tale da integrare le applicazioni di acquisizione dati già per settore di competenza;

I rispondenti effettuano la registrazione sul sito e si autenticano tutte le volte che accedono al questionario elettronico, utilizzando user-id e password gestite conformemente alla norma vigente;

Il sistema offre ai rispondenti, come funzione primaria, diverse possibilità di trasmissione dei dati (questionari compilati in linea, questionari compilati fuori linea, prelievo/deposito di file, trasferimento di questionari via fax...);

Il sistema offre funzioni secondarie e servizi utili alle attività di raccolta sia per i rispondenti sia per gli utenti statistici (ad esempio documentazione relativa all'indagine, assistenza, monitoraggio degli arrivi e gestione degli archivi delle unità rispondenti).

### ***Requisiti architetturali***

Il sistema offre alti livelli di affidabilità, tali da limitare al massimo la presenza di punti unici di criticità;

l'architettura è disegnata per ridurre i momenti di non disponibilità dei servizi e per garantire la conservazione dei dati acquisiti.

### ***Vincoli***

L'acquisizione dei dati attraverso il sistema rispetta il "Codice in materia di riservatezza dei dati personali" (D.L. 196/03) ed in particolare tutte le misure di sicurezza definite nel relativo Allegato B e nel Documento Programmatico sulla Sicurezza dell'Istituto;

Tutte le soluzioni di cattura via Web saranno sviluppate secondo il modello definito e ospitate sul sito unificato;

Le indagini che acquisiscono dati per mezzo della posta elettronica migrano sul nuovo sistema di acquisizione;

Il sistema non impone vincoli stringenti sulla presenza di software sulle postazioni client dei rispondenti né impone l'acquisizione e/o installazione di software su di esse, riducendo al minimo il "disturbo statistico".

### ***Formato del questionario elettronico***

Il questionario elettronico costituisce la base su cui impiantare la cattura dati: a meno delle situazioni in cui è previsto un semplice trasferimento dati, va sviluppato un questionario elettronico che sarà generalmente autocompilato dal rispondente in modalità online oppure offline.

Sul sito *indata.istat.it* per i questionari elettronici i formati attualmente utilizzati sono PDF, HTML, EXCEL che risultano facilmente trattabili dai più diffusi browser, sia di mercato sia open source, e familiari agli utenti.

Il formato XHTML potrebbe essere raccomandabile se si vuole fare interagire i rispondenti con l'Istituto attraverso dispositivi (palmari o cellulari) diversi dai tradizionali computer.

I questionari in formato PDF vengono generati con il modulo Designer del software Teleform che è un sistema di mercato utilizzato dall'Istituto per la generazione e il trattamento di form cartacei ed elettronici.

I form PDF vengono pubblicati sul sito *indata.istat.it*, ed è prevista la funzione *accesso al modello*, che realizza il link al questionario, personalizzato sulla base dei dati identificativi del rispondente.

I questionari elettronici in formato HTML possono essere generati con il modulo Designer del software Teleform oppure possono essere creati con sviluppo ad hoc.

I questionari elettronici in formato EXCEL vengono scaricati, in forma personalizzata, dai rispondenti, dopo accesso a *indata.istat.it* e passo di autenticazione, e vengono restituiti con un'operazione di upload su una specifica area del sito.

### ***Modalità di applicazione e requisiti di accessibilità/usabilità***

Le applicazioni sviluppate per la cattura di dati via WEB devono rispettare lo schema funzionale previsto e implementato sul sistema di acquisizione.

Accanto alla funzione che permette l'accesso al questionario di rilevazione dell'indagine di riferimento devono essere disponibili le seguenti informazioni:

- breve illustrazione della Rilevazione;
- guide (anche scaricabili) alla compilazione del questionario e all'uso della procedura telematica, ed eventualmente altra documentazione utile al rispondente;
- recapiti (indirizzi e-mail e numeri telefonici/fax) dei referenti dell'indagine e del personale tecnico responsabile del sito.

Questa cornice funzionale è già disponibile sul sito e pertanto utilizzabile da tutte le rilevazioni che prevedono cattura di dati via WEB.

Nello sviluppo delle applicazioni, indipendentemente dal linguaggio adottato, si deve far riferimento ai requisiti di usabilità ed accessibilità dei siti Web della Pubblica Amministrazione espressi nella legge 9 gennaio 2004 n. 4, comunemente nota come "Legge Stanca".

La legge è relativa a qualsiasi tipologia di sito con particolare riferimento alla diffusione delle informazioni; per quanto attiene, in particolare, applicazioni Web per l'immissione dei dati, possono essere evidenziati i seguenti requisiti, a cui ispirarsi per garantire la massima usabilità:

1. le pagine Web devono essere visualizzabili su qualsiasi sistema indipendentemente dalla sua configurazione software;
2. se l'applicazione per l'inserimento dei dati è suddivisa in più pagine, esse devono essere collegabili tramite collegamenti visibili e comprensibili;
3. le pagine si devono poter usare anche senza mouse;
4. le pagine devono essere stampabili.

### ***Tecnologie adottate***

Il sito ufficiale *indata.istat.it* utilizza come DBMS le tecnologie definite come standard per l'Istituto e in particolare ORACLE 10g e MYSQL 5.0, specializzando l'uso del DBMS Oracle l'archiviazione dei microdati rilevati, mentre il DBMS MySql per le tabelle gestionali utilizzate per la definizione ed il controllo degli accessi, per il log delle connessioni e per tutte le informazioni di supporto.

E' previsto l'uso del software generalizzato EX-FOLIA per di trattare in modalità automatica fogli Excel utilizzati come questionari per l'acquisizione dei dati e trasmessi dai rispondenti.

## **5.2 La rilevazione con tecniche CATI**

La tecnica di rilevazione CATI (Computer Assisted Telephone Interview) consiste nella conduzione assistita da computer di interviste telefoniche.

La tecnologia standard di supporto è il software BLAISE che consente la realizzazione del sistema CATI su tre moduli principali: il questionario elettronico, il sistema di schedulazione delle chiamate ed il sistema di indicatori di monitoraggio delle interviste.

L'Istituto adotta, come strategia primaria di conduzione delle rilevazioni CATI, quella di affidare il servizio a società specializzate di mercato, che forniscono gli intervistatori e tutta la strumentazione tecnica necessaria (telefonia, cuffie microfonate, hardware, software di base e d'ambiente, etc). E' invece a carico dell'Istituto fornire alle società il sistema CATI (Blaise), corredato dell'applicazione che implementa il questionario elettronico.

In casi particolari, dovuti alla necessità di rispettare le tempistiche d'indagine, la strategia utilizzata può essere quella di assegnare a società specializzate la completa conduzione della rilevazione CATI. In questi casi le società dovranno disporre di un loro sistema CATI, in grado comunque, di garantire i requisiti funzionali descritti in precedenza.

## **5.3 La rilevazione con tecnica CAPI**

La tecnica di rilevazione CAPI (Computer Assisted Personal Interview) consiste nella conduzione assistita da personal computer (generalmente laptop o palmari) di interviste faccia a faccia. Le rilevazioni possono riguardare interviste dirette ad individui oppure rilevazioni specifiche e sono sempre condotte da rilevatori. L'organizzazione descritta a seguire prende spunto dalla realizzazione del sistema di supporto all'indagine sulle Forze di Lavoro considerata una buona pratica per l'Istituto e standard de facto per l'Istituto stesso.

La realizzazione delle interviste viene effettuata tramite l'utilizzo di un sistema software applicativo, sistema CAPI, composto da alcune componenti riassunte a seguire:

- *Banca Dati CAPI*: contiene i dati relativi al campione di riferimento, alle unità di rilevazione, alle interviste;
- *Funzionalità di gestione delle interviste* : il sistema consente la gestione territoriale delle interviste (gestione dei rilevatori, degli uffici territoriali, comuni e Istat); la schedulazione delle interviste. la gestione delle attivazioni, interruzioni definitive e/o temporanee, rifiuti, sostituzioni dei rilevatori, assegnazione secondo scadenza prefissata (settimanale, sub-settimanale o giornaliera) delle unità di rilevazione ai rilevatori attivi, Riassegnazioni in termini di controllo e modifica delle assegnazioni dei nominativi ai rilevatori da parte degli uffici territoriali.;ridistribuzione (Gestione della redistribuzione del carico di lavoro dei rilevatori già assegnato), preparazione dei file per i rilevatori.
- *Funzionalità di gestione e monitoraggio delle trasmissioni* :il sistema inoltra ai rilevatori, in modo automatico e a scadenza prefissata le unità di rilevazione da intervistare. Tale inoltra avviene durante le sessioni di connessione da parte dei rilevatori.



- *Funzionalità a disposizione del rilevatore* :Gestione dell'agenda del rilevatore e dei tentativi di contatto.
- *Esecuzione dell'intervista*. Consiste nell'esecuzione del questionario elettronico.
- *Invio/ ricezione dei dati dai rilevatori*.

## **5.4 La rilevazione con tecnica CADI**

La tecnica di rilevazione dei dati di tipo CADI prevede la raccolta tramite questionari cartacei tradizionali. I dati dei modelli di rilevazione vengono acquisiti in formato digitale tramite successiva registrazione in service, eseguita di norma da società esterne legate all'Istituto da appositi contratti.

Attualmente l'Istituto invia secondo piani di registrazione prestabiliti i modelli di rilevazione delle singole indagini alla Società che eroga il servizio di registrazione. Gli invii, ripartiti in lotti successivi, sono eseguiti in base ad una programmazione concordata con i referenti di indagine e inviata anticipatamente alla ditta.

La Società effettua la registrazione dei modelli secondo tecniche stabilite dalla società stessa anche mediante programmi finalizzati e predisposti dalla Società stessa sulla base del piano di registrazione e delle specifiche tecniche fornite dall'Istat al momento dell'inizio dei lavori.

Per esigenze specifiche dell'indagine di riferimento, la registrazione potrà essere assistita e svolta con l'ausilio di programmi forniti dall'Istat. In questo caso le applicazioni dovranno essere realizzate dall'Istituto tramite il software Blaise (ver 4.7), largamente in uso presso le società di registrazione. Il personale preposto alla registrazione imparerà ad usare tali applicazioni tramite apposita formazione eseguita dal personale dell'Istituto stesso.

### ***Linguaggi di sviluppo e specifiche architetturali***

L'ambiente di sviluppo per la realizzazione dei software di data-entry è il Blaise ver 4.7.

Lo sviluppo della maschera Blaise di registrazione deve avvenire in base al piano di registrazione previsto per i modelli di rilevazione e comunque dopo aver predisposto e concordato gli opportuni tracciati record. L'utilizzo del software di registrazione sviluppato dall'Istituto, da un lato garantisce la certezza della completa applicazione del piano di registrazione con chiara ricaduta sulla qualità del dato registrato, consentendo nel contempo un collaudo più analitico dei dati registrati.

E' importante sottolineare, per quanto riguarda l'impostazione grafica del sistema di data-entry, l'opportunità di adottare una impostazione di pagina che sia il più possibile identica al cartaceo in modo da consentire all'operatore l'analoga visione del modello e ottimizzare la registrazione in service dei dati.

### ***Modalità di scambio dati per via telematica***

I file prodotti in fase di registrazione in service, raggruppati in lotti prestabiliti dal piano di registrazione, vengono inviati all'Istat per i controlli di qualità e per l'attivazione dei processi di controllo e correzione necessari prima della codifica e della fase di produzione.

Vengono inviati all'Istat su supporto digitale alla struttura Istat preposta unitamente ai modelli cartacei e/o utilizzando il sito per la trasmissione sicura di file <https://indata.istat.it>, con una utenza ed una password fornite dall'Istat.

I file dati ed i programmi eventualmente forniti dall'Istat necessari alla Società per svolgere l'attività di registrazione, sono resi disponibili sullo stesso sito (<https://indata.istat.it>), e trattati dalla Società secondo le specifiche tecniche che vengono concordate all'inizio dei lavori.

### ***Controllo della qualità dei dati***

Obiettivo del controllo di qualità è quello di verificare che il materiale elettronico in termini di file di dati consegnato dalla Società sia conforme ai livelli di qualità definiti nel contratto.

I controlli sono effettuati a campione su ciascun lotto di materiale consegnato. I campioni di controllo sul materiale consegnato dalla Società sono dimensionati in modo tale che il test sul rispetto dei livelli di qualità sottoponga la Società ad un rischio di prima specie<sup>19</sup> non superiore a 0,05. Per bilanciare tale rischio la potenza del test<sup>20</sup> verrà comunque mantenuta al di sotto del complemento a 1 del rischio di prima specie e sarà comunque non inferiore a 0,90.

I controlli effettuati sul materiale consegnato riguardano:

- caratteri numerici
- stringhe alfabetiche
- codici assegnati alle stringhe
- codici preassegnati per i quali non è stata suggerita una sostituzione
- codici assegnati a sostituzione di quelli pre-assegnati dall'Istat ritenuti errati.

Al termine delle operazioni di controllo, relativamente a tutti i record dei campioni, saranno conteggiati gli errori riscontrati per il calcolo delle stime dei parametri di qualità.

I livelli di errore, verificati sulla base dell'analisi dei campioni di controllo, saranno calcolati come rapporto tra eventi sfavorevoli (errori) ed eventi totali. In particolare, il livello di errore verrà calcolato, per i caratteri numerici, come rapporto tra il numero di caratteri registrati in modo errato ed il totale dei caratteri presenti nel campione; per le stringhe alfabetiche, come rapporto tra il numero delle stringhe registrate in maniera errata ed il totale delle stringhe presenti nel campione; per i codici numerici assegnati alle stringhe, come rapporto tra codici attribuiti in modo errato ed il totale delle stringhe da codificare; per i codici preassegnati in fase di codifica automatica, come mancate segnalazioni su totale codici preassegnati in maniera errata; per i codici suggeriti a sostituzione di quelli apposti in fase di codifica automatica, come segnalazioni errate su totale segnalazioni.

Le stringhe alfabetiche vengono considerate registrate in maniera errata quando, a giudizio di un operatore esperto dell'Istat, risultino significativamente alterate in quanto a significato, collocazione e forma.

Il controllo di qualità della codifica sarà eseguito da operatori esperti dell'Istat, attribuendo i codici a partire dalle stringhe alfabetiche originali e utilizzando lo stesso software messo a disposizione della Società. Nel caso in cui si verifichi un errore nei campi identificativi del record verrà considerato errato l'intero record. Per le diverse tipologie di dati non devono essere superati i livelli massimi di errore descritti nella tabella che segue.

---

<sup>19</sup> Probabilità di rifiutare un lotto conforme ai requisiti di qualità.

<sup>20</sup> Complemento a 1 della probabilità di accettare un lotto non conforme ai requisiti di qualità.

**Tavola 4:** Riepilogo del livello massimo di errore per tipologia di dato

<b>TIPOLOGIA DI DATI</b>	<b>LIVELLO MASSIMO DI ERRORE</b>
Caratteri numerici	5‰
Stringhe alfabetiche	2%
Codici assegnati alle stringhe	2%
Codici preassegnati e lasciati invariati	2%
Codici proposti a sostituzione di quelli preassegnati dall'Istat ritenuti errati	2%

Qualora i risultati dei controlli rilevino livelli di qualità inferiori a quelli stabiliti, l'intero lotto viene rimandato alla Società per essere rielaborato, senza oneri per l'Istat, nelle parti del lavoro che non hanno rispettato gli standard di qualità richiesti. Sulle successive consegne di materiale rielaborato l'Istat ripeterà il controllo di qualità

## **5.5 Acquisizione secondaria**

La produzione statistica ufficiale, nell'ottica della cooperazione tra Amministrazioni pubbliche, del riutilizzo dei patrimoni informativi già disponibili, e del contenimento del disturbo statistico verso i rispondenti, si pone sempre più l'obiettivo di utilizzare fonti esterne, anche di carattere amministrativo, e ridurre, quando è possibile, la costosa attività di rilevazione diretta.

E' prevedibile pertanto un incremento delle attività di acquisizione secondaria dei dati, soprattutto per lo scambio di fonti amministrative e dati raccolti all'esterno dell'Istituto.

Le operazioni di acquisizione secondaria, regolare o occasionale, vengono risolte in una delle due modalità descritte ai paragrafi seguenti, già regolarmente utilizzate.

### ***Datistat***

Datistat è un'applicazione che realizza trasferimento di file; è scritta in linguaggio PHP e disponibile sulla home page del sito unificato *indata.istat.it*.

Tale funzione consente il trasferimento di file con operazioni di UPLOAD/DOWNLOAD guidate per l'utente da una cornice applicativa che le rende di facile utilizzo. Le operazioni si svolgono in protocollo HTTPS garantendo così la protezione dei dati nel transito dal client dell'utente al server di Istituto. Le funzioni di Datistat, come tutte le altre funzioni di cattura disponibili su *indata.istat.it*, sono accessibili con garanzie reciproche, a chi trasmette e a chi riceve i dati, dell'identità del sito (che è regolarmente certificato) e degli utenti che operano sul sito (attraverso la gestione di USER-ID e PASSWORD secondo le norme vigenti).

### ***File Transfer Protocol***

Nei casi generali di interscambio dati (dati, archivi registri, fonti amministrative) con altri enti nell'ambito della Pubblica Amministrazione è da utilizzare il protocollo *ftp* tramite collegamento

offerto da SPC ( il Sistema Pubblico di Connettività) o, ove esso non sia disponibile, tramite collegamenti dedicati.

In particolare il trasferimento dei dati individuali o sensibile dovrà avvenire per mezzo del protocollo SFTP (Secure File Transfer Protocol), che garantisce la cifratura dell'intero canale di trasmissione. E' da tenere presente, comunque, che, generalmente ogni nuova procedura di scambio dati con enti esterni, passa attraverso la stesura di accordi bilaterali che definiscano il mezzo e il protocollo più opportuno a supportarlo.

L'Istituto dispone di una architettura esposta con funzioni di ftp server (hostname "Ungheria"), raggiungibile attraverso la rete SPC o attraverso collegamenti diretti.

Allo stato attuale alcune Pubbliche Amministrazioni (Banca d'Italia e la società SOGEI per il Ministero delle Finanze (Dogane)) utilizzano regolarmente questa modalità di trasmissione ma con protocollo non cifrato (FTP), che viene supportato per compatibilità con le attuali applicazioni e nell'attesa di concordare con le parti il passaggio al FTP sicuro.

Per ragioni di sicurezza l'uso del SFTP o del FTP andrebbe limitato alla sola modalità "Put", per non consentire ai corrispondenti l'accesso in lettura al FTP server.

## 6. Controllo e correzione

Il sottogruppo<sup>21</sup> incaricato, alla cui documentazione di dettaglio si rimanda<sup>22</sup>, ha individuato nella fase di controllo e correzione rientrano tutti i processi relativi alla individuazione, nell'ambito dei dati raccolti, dei casi di mancata risposta parziale e dei casi di errore, alla integrazione delle mancate risposte e alla correzione degli errori. Le procedure che implementano i processi descritti possono essere interattive (effettuate da personale esperto su dati individuali), automatiche (intervengono in modalità batch su insiemi di dati), o miste.

Per applicazioni di controllo e correzione di tipo individuale, dove quindi sono necessarie maschere di visualizzazione/inserimento/modifica, è possibile realizzare applicazioni web-based (che hanno il vantaggio di non richiedere installazione di software sulle postazioni client) con accesso a database relazionali. Nell'ottica del contenimento dell'impegno di sviluppo e dell'utilizzo di strumenti open source si segnala che il software statistico "R" offre diverse funzioni utili per lo sviluppo di procedure automatiche di controllo e correzione dei dati. Sono inoltre disponibili presso l'Istituto numerosi software generalizzati per il controllo e correzione dati, sviluppati in casa o presso altri Istituti Nazionali di Statistica, che implementano approcci diversi (sistematico o probabilistico) e sono specializzati su tipologie di variabili.

---

<sup>21</sup> Sottogruppo di analisi di controllo e correzione. Membri: Barcaroli, Fanfoni, Fazio, Pagliuca, Spagnuolo, Torelli

<sup>22</sup> G. Barcaroli, L. Fanfoni, L. Fazio, D. Pagliuca, F. Spagnuolo, R. Torelli, *Controllo e correzione dei dati*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_3\\_correzione](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_3_correzione)

**Tavola 5** : Standard per la fase di Controllo e Correzione

<b>TIPOLOGIE DELLA FASE</b>		
	<b>Interattiva</b>	<b>Automatica</b>
<b>Ambiente di riferimento</b>		
Architetture esposte web		
Architetture interne distribuite	X	X
Architetture interne web	X	X
<b>Linguaggi</b>		
Suite Oracle	X	X
Java	X	X
PHP	X	X
C		X
R		X
<b>SW generalizzati</b>		
CONCORD		
CANCEIS		
DIESIS		

## 7. Produzione di stime e tabulazioni

Il sottogruppo<sup>23</sup> incaricato, alla cui documentazione di dettaglio si rimanda per approfondimenti<sup>24</sup>, di analizzare la fase di produzione delle stime e tabulazione definisce il processo di stima e tabulazione come quel complesso di elaborazioni che, partendo dai microdati validati di una o più indagini, produce distribuzioni di dati aggregati, che vengono opportunamente tabulati.

Il processo di aggregazione, prioritario alla tabulazione, nel caso di indagini censuarie si esaurisce nell'applicazione di operatori statistici (conteggi, somme, medie, varianze, ...) a variabili di misura, definendo nel contempo le variabili di classificazione rispetto alle quali effettuare tali operazioni. Nel caso di indagini campionarie, a tali elaborazioni occorre far precedere una fase di calcolo dei coefficienti di riporto, o pesi campionari. Tale calcolo deve tener conto, da una parte, del particolare disegno campionario adottato, e, dall'altra, dei vincoli rappresentati da totali noti nella popolazione di interesse.

Per realizzare stime e tabulazioni è possibile sviluppare applicazioni web, con accesso a db relazionali, basate su generazione dinamica delle interrogazioni e sulla disponibilità dei metadati di indagine organizzati in forma di base dati. Per le funzioni caratteristiche di questa fase sono disponibili sia software statistici proprietari che funzioni e librerie di software open source (software statistico "R"). Sono infine disponibili diversi software generalizzati sia di mercato sia di produzione interna.

<sup>23</sup> Sottogruppo di analisi di stime e tabulazione. Membri: Antonio Laureti Palma, Claudia Albergamo, Giulio Barcaroli, Daniela Pagliuca

<sup>24</sup> C. Albergamo, G. Barcaroli, A. Laureti Palma, D. Pagliuca, V. Patrino, *Stime e tabulazioni*, Documento interno [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_4\\_stime](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_4_stime)

**Tavola 6 : Standard per la fase di Stima e Tabulazione**

	Tipologie della fase	
	Stime	Tabulazione
<b>Ambiente di riferimento</b>		
Architetture esposte web		
Architetture interne distribuite	X	X
Architetture interne web	X	X
<b>Linguaggi</b>		
Suite Oracle		X
Java	X	X
PHP	X	X
R	X	X
<b>SW generalizzati</b>	GENESEES, R-Survey	SMOL, IMPAGINA, R-EVER, BO, Oracle Discoverer

## 8. Produzione di file standard e Data Warehouse

Nell'ambiente statistico è frequente la necessità di produrre file standard e Data Warehouse. Il sottogruppo<sup>25</sup> incaricato all'analisi, alla cui documentazione si rimanda per approfondimenti<sup>26</sup>, ha analizzato il problema facendo riferimento alle collezioni campionarie rese anonime, note come "File di microdati", o "File standard", che possono essere rilasciate per fini di studio e di ricerca su richiesta motivata e previa autorizzazione del Presidente dell'Istituto.

Per "File standard" si intendono collezioni campionarie di microdati resi anonimi che possono essere rilasciate a utenti esterni per fini di studio e di ricerca. Per processi di "data warehouse" si intende l'insieme dei processi di trattamento del dato statistico, basati sulle tipiche attività di riconciliazione, staging, integrazione, aggregazione, costruzione di data mart tematici, funzionalità OLAP. Queste attività sono distribuite lungo il ciclo di vita del dato, dalla conclusione del processo di validazione fino al momento della diffusione su Web.

Negli ultimi anni è maturata in Istituto la visione di un ambiente web di Data Warehouse per la gestione integrata ed omogenea dell'output, sia a livello di microdati sia di dati aggregati. Tale orientamento è nato dall'esigenza di far fronte con sempre maggior frequenza a sistemi informativi statistici caratterizzati da una molteplicità di fonti, tematiche e tipologie di output. Per questo, la proposta di adozione di un unico ambiente dedicato a questa funzione è al momento implementata tramite una architettura generalizzata orientata a gestire il ciclo di vita dei dati statistici validati e rendere disponibili all'utenza dati a progressivi gradi di integrazione ed aggregazione. Questo avviene tramite istanze di data warehouse implementate in modo autonomo dalle strutture di produzione con gli strumenti (toolkit e workflow) del sistema generalizzato. Di base per uno sviluppo armonico di DW sono l'assunzione, come fonte privilegiata, del contenitore centrale dei microdati validati (Armida), la possibilità di costruire specifici strati semantici di natura statistica (metadati) tramite strumenti di modellazione dimensionale, lo sviluppo di funzioni di staging e ETL integrate, la

<sup>25</sup> Sottogruppo di analisi di file standard e data warehouse. Membri : De Francisci, Barcaroli, Laureti Palma, V. Patruno, Rotondi.

<sup>26</sup> G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno, [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)

possibilità di interscambio e fusione tra i vari output prodotti. Particolare attenzione va prestata al tema della riservatezza. Sono disponibili software generalizzati per le principali funzionalità, dalla modellazione alla pubblicazione dei dati.

La procedura standard per la produzione dei file dei microdati di una data indagine ddeve prevedere:

- l'estrazione dei dati dal file dei microdati validati contenuto in ARMIDA
- la valutazione del rischio di violazione della riservatezza e l'eventuale manipolazione dei dati onde contenere tale rischio al di sotto di una soglia prefissata
- l'associazione dei relativi metadati operativi e statistici
- la collocazione del file su opportuni supporti (CD) o in ambienti web.

**Tavola 7 : Standard per DataWarehouse e File Standard**

<b>Tipologie della fase</b>				
	<b>Disegno</b>	<b>Staging e ETL</b>	<b>Query, Analysis, Reporting</b>	<b>Navigazione Web</b>
<b>Ambiente di riferimento</b>				
Architetture esposte web				X
Architetture interne distribuite			X	
Architetture interne web	X	X	X	
<b>Linguaggi</b>				
Java	X	X	X	X
PHP	X	X	X	X
Suite Oracle		X		
<b>SW generalizzati</b>	Foxtrot.Meta, Foxtrot.PD, Foxtrot.MD	Foxtrot.Meta, Foxtrot.PD, Foxtrot.MD	SMOL	Web.PD, Web.MD

## 9. Sistemi informativi trasversali e sistemi di supporto allo sviluppo

La tavola seguente riassume gli strumenti standard di supporto allo sviluppo e in particolare alla gestione generale dei progetti, alle fasi di analisi e progettazione, allo sviluppo e gestione dei Data Base, allo sviluppo di applicazioni software e di work flow. Si fornisce inoltre indicazione delle soluzioni standard scelte dall'Istituto per la gestione di contenuti dei siti WEB (CMS) e per la gestione delle configurazioni software (CCM).

**Tavola 8** : *Standard per gli strumenti di supporto*

---

<b>Gestione Progetti</b>	Ganttproject, DotProject
<b>Analisi e progettazione</b>	ErWin/BpWin, DBDesigner
<b>Gestione/Sviluppo DB</b>	TORA, SQL Developer
<b>Sviluppo del software (IDE)</b>	Netbeans, Zend, Eclipse, Oracle Developer
<b>Gestione della configurazione (CCM)</b>	Harvest, Subversion
<b>Gestione dei contenuti (CMS)</b>	Typo3
<b>WorkFlow (WfMS)</b>	Oracle WorkFlow, JBPM

---

### **9.1 Sistemi informativi trasversali**

Premesso che i sistemi informativi orientati alla diffusione dell'informazione statistica sono trattati secondo gli standard di cui al punto precedente, restano molte possibilità di applicazione dei sistemi informativi in ambito statistico: sistemi di documentazione dei processi e dei contenuti di indagine (metadati statistici), di archiviazione di microdati, sistemi amministrativi e gestionali, sistemi per la integrazione di altri sistemi. E' evidente che tutti i sistemi informativi già in essere di carattere trasversale sono passibili di riuso. In generale l'esigenza di sviluppo di un nuovo sistema va verificata con l'offerta di soluzioni di sistemi informativi già disponibili. Per lo sviluppo ex-novo di un sistema informativo vanno utilizzati gli strumenti standard definiti nel presente documento e relativamente alle specifiche architetturali si suggeriscono applicazioni web-based, supportate preferibilmente da un sistema distribuito su tre livelli fisici (presentazione, logica applicativa, dati), che risulta più sicuro relativamente alla accessibilità del livello dati.

Il sottogruppo<sup>27</sup> ha analizzato i sistemi informativi trasversali o sistemi a supporto delle indagini di produzione ed i sistemi amministrativi-contabili.

Si riporta a seguire la loro definizione e classificazione rimandando alla documentazione tecnica<sup>28</sup> di dettaglio prodotta dal sottogruppo di analisi per eventuali approfondimenti<sup>29</sup>.

- “Sistemi informativi di documentazione, archiviazione e strumentali: nell'ambito di questa tipologia vengono collocati i sistemi di metadati statistici (SIDI, SDOSIS), i sistemi centralizzati di archiviazione dei dati (ARMIDA) e i sistemi strumentali alle indagini (ASIA, SISTER, Sistema delle classificazioni)”
- “Sistemi informativi di supporto: sono quei sistemi trasversali alle fasi del processo produttivo delle indagini (SIGIF, SIRC)”
- “Sistemi informativi gestionali: sono i sistemi amministrativi (SIGED, LYBRA, TIMESHEET)”
- “Sistemi informativi direzionali: sono i sistemi di pianificazione e controllo (PSN, PST)”

---

<sup>27</sup> Sottogruppo di analisi dei Sistemi informativi, trasversali e amministrativi. Membri: Calamani, De Francisci, Pellegrini, Toma, Vaccari

<sup>28</sup> C. Pellegrini, *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza di SIDI*, Collana Documenti Istat, n°4/2006.

<sup>29</sup> La classificazione è tratta da E. Calamani, S. De Francisci, C. Pellegrini, A. Toma, C. Vaccari, *Sistemi informativi, trasversali e amministrativi*, Documento interno, [http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_6\\_sistemi\\_informativi\\_trasversali](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_6_sistemi_informativi_trasversali)



- “Sistemi informativi integrati: sistemi dove viene attuata la interoperabilità o cooperazione (INTERMETA, INCIPIIT)”.

## **9.2 Strumenti di gestione dei progetti e di supporto allo sviluppo del software**

Per la gestione e il controllo di progetti complessi così come per la progettazione e lo sviluppo del software esiste una grande varietà di strumenti. Gli strumenti adottano metodi, tecniche, formalismi e standard riconosciuti, favorendo così comunicazione e trasferimento delle esperienze, standardizzazione dei processi di sviluppo e riuso di soluzioni già adottate.

Sono state individuate le seguenti tipologie di strumenti:

- strumenti per la gestione dei progetti (Project management)
- strumenti di analisi e progettazione
- strumenti per lo sviluppo del software e per la gestione DB
- strumenti per la gestione della configurazione software
- strumenti di workflow management systems (WfMS).

Avvalersi di tali strumenti nelle diverse fasi di pianificazione, progettazione e sviluppo di sistemi software, assicura il rispetto di standard qualitativi per i processi di produzione statistica ed in generale per i processi gestionali. Tra gli strumenti disponibili, a parità di funzionalità offerte, è preferibile adottare quelle operanti in architetture WEB, Open Source e supportate da repository relazionali.

Il sottogruppo<sup>30</sup> di analisi, alla cui documentazione tecnica di dettaglio si rimanda per approfondimenti<sup>31</sup>, ha analizzato gli strumenti che sono a supporto della definizione e del governo di un processo di progettazione e sviluppo di sistemi software.

Nella tabella a seguire vengono sintetizzati in particolare gli strumenti di Project management attualmente utilizzati in Istituto e analizzati dal sottogruppo; lo schema mette in evidenza la modalità operativa (stand-alone e cooperativa). Tutti i prodotti Open sono Free.

---

<sup>30</sup> Sottogruppo di analisi degli Strumenti di progettazione, project management, workflow e monitoraggio. Membri : Cappella, Casale, Spagnuolo, Pellegrini, Toma, Vaccari

<sup>31</sup> M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno, [http://wiki/doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki/doku.php/standard:documenti:sottogruppo_7_strumenti)

**Tavola 9 : Schema di sintesi degli strumenti di project management<sup>32</sup>**

Strumento per project managment	Modalità operativa	Ambiente Operativo	Open Source (si/no)	Funzioni supportate	Riferimenti	Note
Win_Project	Stand-alone	Window	No	Avanzate		
Ganttproject	Stand-alone	Window, Linux	Sì	Avanzate	<a href="http://ganttproject.biz">ganttproject.biz</a>	In sperimentazione
OpenWorkBench	Stand-alone	Window	Sì	Avanzate	<a href="http://www.openworkbench.org">www.openworkbench.org</a>	
dotProject	Cooperativa (via WEB)	Server: PHP / MySQL / Apache Interfaccia: Browser	SI	Avanzate e lavoro cooperativo	<a href="http://www.dotproject.net/">www.dotproject.net/</a>	In sperimentazione
PHProjekt	Cooperativa (via WEB)	Server: PHP / MySQL / Apache Interfaccia: Browser	SI	Base e lavoro cooperativo	<a href="http://www.PHProjek.it">www.PHProjek.it</a>	

<sup>32</sup> Fonte: M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno, [http://wiki/doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki/doku.php/standard:documenti:sottogruppo_7_strumenti)

La tabella di sintesi a seguire riassume gli strumenti di gestione dei database standard e in uso in Istat.

**Tavola 10** : Schema di sintesi degli strumenti per la gestione dei DB<sup>33</sup>

Strumento per la gestione DB	Forniti dal DBMS o strumenti ad hoc	Ambiente Operativo ed Architettura	Open Source (si/no)	Funzioni e/o caratteristiche tecniche	Note
TORA	Strumento ad hoc fornito da Oracle	Linux e Windows Client Server	-----	<ul style="list-style-type: none"> <li>• Supporto per connessioni multiple</li> <li>• Editor sql dotato di explain plan e statistiche</li> <li>• Editor con supporto della sintassi dei linguaggi SQL e PL/SQL</li> <li>• Schema browser</li> <li>• Storage manager</li> <li>• Session manager</li> <li>• SGA trace</li> <li>• Rollback manager</li> <li>• PL/SQL Debugger</li> <li>• Security manager</li> </ul>	Free
TOAD	Strumento ad hoc	Windows Client Server	No	<ul style="list-style-type: none"> <li>• Generazione ed esecuzione query</li> <li>• Creazione e modifica oggetti di database</li> <li>• Sviluppo ed debug di codice SQL e PL/SQL.</li> <li>• Import e export di dati</li> <li>• Confronto tra schemi e aggiornamento statistiche</li> <li>• Funzionalità integrate di SQL tuning e revisione automatica del codice PL/SQL</li> </ul>	E' un tool da fornire, viste le potenzialità, solo a utenti tecnici esperti
SQLDeveloper	Fornito direttamente da DBMS Oracle	Windows Client Server	No	<ul style="list-style-type: none"> <li>• Sviluppo, collaudo, esecuzione di debug</li> <li>• Ottimizzazione degli Gli Stored Program In Oracle PL/SQL come (Package, Trigger e altro)</li> <li>• Tuining</li> </ul>	E' incluso nel DBMS ORACLE
SQLyog	Strumento ad hoc	Linux e Windows Client Server	Sì	<ul style="list-style-type: none"> <li>• Gestione e la Configurazione di server MySQL e database</li> <li>• Import dati da database ODBC</li> <li>• Sincronizzazione dei dati tra database</li> <li>• Esecuzione</li> <li>• Copia database tra host diversi</li> <li>• Generazione di schemi relazionali di db.</li> </ul>	
PhpMyAdmin per MySql	Strumento ad hoc	LAMP e WAMP - WEB	Sì	<ul style="list-style-type: none"> <li>• Creazione e drop di db</li> <li>• Create/drop/alter tables</li> <li>• Delete/edit/add fields</li> <li>• Esecuzione SQL statement,</li> </ul>	In sperimentazione
Enterprise Manager Oracle WEB	Strumento ad hoc fornito da Oracle	WEB-based	No	Funzionalità di gestione del database e di sviluppo, debugging PL/SQL stored program	

<sup>33</sup> Fonte: M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno, [http://wiki/doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki/doku.php/standard:documenti:sottogruppo_7_strumenti)

In merito agli strumenti IDE (*integrated development environment*) di sviluppo del software, la tabella di sintesi a seguire riporta gli strumenti in uso con dettaglio delle funzionalità offerte e caratteristiche principali.

**Tavola 11** : Schema di sintesi degli strumenti IDE<sup>34</sup>

IDE	Linguaggi supportati	Ambiente Operativo ed Architettura	Open Source (si/no)	Funzioni e/o caratteristiche tecniche	Note
Eclipse	JAVA, PHP, C, C++	Java Client/Server Windows e Linux	Sì	<ul style="list-style-type: none"> <li>• "framework" espandibile</li> <li>• tool e moduli runtime per supporto allo sviluppo, implementazione, gestione del software</li> <li>• gestione completa delle funzionalità del client CVS</li> </ul>	<ul style="list-style-type: none"> <li>• I tools e i framework sono principalmente concentrati su come costruire JEE, web services ed applicazioni web</li> <li>• In sperimentazione</li> </ul>
Netbeans IDE	Multi-linguaggio include JAVA, PHP, C, C++	Java Client/Server Windows e Linux	Sì	<ul style="list-style-type: none"> <li>• Editing</li> <li>• Compilazione</li> <li>• Debugging</li> <li>• Deploy</li> </ul>	<ul style="list-style-type: none"> <li>• Esiste anche un gran numero di moduli utili per estendere le funzionalità di NetBeans IDE</li> <li>• In sperimentazione</li> </ul>
Netbeans Platform	Multi-linguaggio include JAVA, PHP, C, C++	Java Client/Server Windows e Linux	Sì	<ul style="list-style-type: none"> <li>• Gestione di progetti di medie dimensioni che richiedano l'uso di servlet e jsp</li> <li>• Fornito internamente di un motore Tomcat che permette il test delle applicazioni sviluppate senza la necessità di dover installare a parte il servlet-engine Tomcat</li> </ul>	<ul style="list-style-type: none"> <li>• Supporto per la creazione di grosse applicazioni desktop</li> <li>• In sperimentazione</li> </ul>
Zend Studio	PHP	LAMP e WAMP (con sever Linux)	Sì	Tool completo di sviluppo per applicazioni WEB in PHP	<ul style="list-style-type: none"> <li>• Ricco di funzionalità</li> <li>• Non è distribuibile gratuitamente</li> </ul>
Oracle JDeveloper 10g	Java, Xml, Jsp, Pl/Sql, Sql	Client/Server ed WEB Windows e Linux	No	Approccio visuale denominato "Oracle Application Development Framework" (Oracle ADF) per: <ul style="list-style-type: none"> <li>• Modellazione</li> <li>• Codifica</li> <li>• Debugging</li> <li>• Fasi Di Test</li> <li>• Installazione</li> <li>• Integrazione</li> <li>• Fine-Tuning</li> </ul>	
Oracle Developer Suite	Oracle/Form Oracle/Report Pl/Sql, Sql	Client/Server Windows e Linux	NO	Ambiente di sviluppo completo per ambiente Oracle in particolare per Forms e Report	

<sup>34</sup> Fonte : M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno, [http://wiki/doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki/doku.php/standard:documenti:sottogruppo_7_strumenti)

## Riferimenti

- R. Balestrino and L. Fanfoni, The Quality of Production Processes and the Electronic Survey System: The Experience of ISTAT's Structural Business Statistics, European Conference on Quality in Survey Statistics. Proceedings of Q2006.
- G. Barcaroli, Il processo di produzione dell'Informazione statistica e l'opzione open source, Convegno Il software per la statistica ufficiale: dai sistemi proprietari all'open source, <http://adamsoft.caspur.it>.
- C. Ghezzi, A. Fuggetta, S. Morasca, A. Morzenti, M. Pezze, Ingegneria del Software, Mondadori Informatica, II edizione.
- Institute of Electrical and Electronics Engineers. IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries. New York, NY: 1990.
- R. Balestrino e A. Guacci, Tecniche di cattura nei processi di produzione statistica, Collana Documenti Istat, n°2/2009.
- D. Pagliuca, R. Cianchetta, M. Broccoli, T. Buglielli, R. di Giuseppe, D. Zardetto, *L'Osservatorio Tecnologico per i Software Generalizzati (OTS) nel 2008*, Collana Documenti Istat, n°4/2009.
- C. Pellegrini, *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza di SIDI*, Collana Documenti Istat, n°4/2006.
- CNIPA (2009), <http://www.ossipa.cnipa.it>.
- Wikipedia (2009), <http://wikipedia>
- Istat (2009), <http://www.istat.it/strumenti/metodi/software/>
- Istat (2009), <http://www.istat.it/strumenti/metodi/software/osservatorio.html>
- R. Balestrino, G. Barcaroli, S. De Francisci, L. Fanfoni, S. Losco, C. Vaccari, *Relazione finale del Comitato degli Standard*, Documento interno  
[http://wiki.istat.it/doku.php/standard:documenti:relazione\\_finale](http://wiki.istat.it/doku.php/standard:documenti:relazione_finale)
- A. Guarino, S. Losco, D. Moscara, *Infrastruttura: standard di software di base e reti*, Documento interno, <http://wiki/doku.php/standard:documenti:architettura>
- R. Balestrino, D. Casale, A. Guarino, S. Losco, D. Moscara, A. Toma, *Standard per l'acquisizione dati*, Documento interno, <http://wiki/doku.php/standard:documenti:acquisizione>
- G. Barcaroli, L. Fanfoni, L. Fazio, D. Pagliuca, F. Spagnuolo, R. Torelli, *Controllo e correzione dei dati*, , Documento interno  
[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_3\\_correzione](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_3_correzione)
- C. Albergamo, G. Barcaroli, A. Laureti Palma, D. Pagliuca, V. Patruno, *Stime e tabulazioni*, Documento interno  
[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_4\\_stime](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_4_stime)
- G. Barcaroli, S. De Francisci, A. Laureti Palma, V. Patruno, G. Rotondi, *File Standard e DWH Tecnologie per l'integrazione*, Documento interno  
[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_5\\_dwh](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_5_dwh)
- E. Calamani, S. De Francisci, C. Pellegrini, A. Toma, C. Vaccari, *Sistemi informativi, trasversali e amministrativi*, Documento interno,  
[http://wiki.istat.it/doku.php/standard:documenti:sottogruppo\\_6\\_sistemi\\_informativi\\_trasversali](http://wiki.istat.it/doku.php/standard:documenti:sottogruppo_6_sistemi_informativi_trasversali)
- M. Cappella, D. Casale, F. Spagnuolo, C. Pellegrini, A. Toma, C. Vaccari, *Strumenti di project management, analisi, progettazione e sviluppo software, workflow*, Documento interno,  
[http://wiki/doku.php/standard:documenti:sottogruppo\\_7\\_strumenti](http://wiki/doku.php/standard:documenti:sottogruppo_7_strumenti)