

***Il sistema informativo statistico
delle Comunicazioni Obbligatorie
SISCO***

Maggio 2014

Ministero del lavoro e delle politiche sociali

Gruppo di lavoro tecnico sulle C.O.

Sommario

1	Introduzione	3
2	Le Fonti Amministrative	5
3	Dal dato amministrativo al dato statistico: i flussi informativi delle Comunicazioni Obbligatorie	8
4	Il Nodo di Coordinamento Nazionale	9
5	L'archivio nazionale delle CO: XML Repository	11
6	Il database amministrativo (COB).....	13
7	Processo propedeutico alla costruzione di SISCO	14
	7.1 Definizione rapporti di lavoro.....	18
8	Il Sistema Informativo Statistico delle Comunicazioni Obbligatorie (SISCO).....	21
	8.1 Ricostruzione rapporti di lavoro validi: la chiave	22

1 Introduzione

Questo documento ha lo scopo di sintetizzare le azioni intraprese dal Ministero del lavoro e delle politiche sociali per il trattamento dei dati amministrativi delle Comunicazioni Obbligatorie a fini statistici. La valorizzazione del patrimonio informativo contenuto nei diversi archivi amministrativi realizzati dal Ministero del lavoro, dall'INPS e dall'INAIL, oltre a rispondere ai principi di sviluppo della statistica ufficiale¹ rappresenta, infatti, un elemento essenziale per la realizzazione di un sistema di monitoraggio e del mercato e delle politiche del lavoro alla cui definizione hanno concorso nel tempo diverse disposizioni normative.

L'Art. 17. del d.lgs. 10 settembre 2003. n.276 - Monitoraggio statistico e valutazione delle politiche del lavoro - fornisce una prima indicazione sull'uso degli archivi amministrativi laddove indica *“Le basi informative costituite nell'ambito della borsa continua nazionale del lavoro, nonché le registrazioni delle comunicazioni dovute dai datori di lavoro ai servizi competenti e la registrazione delle attività poste in essere da questi nei confronti degli utenti per come riportate nella scheda anagrafico professionale dei lavoratori costituiscono una base statistica omogenea e condivisa per le azioni di monitoraggio dei servizi svolte ai sensi del presente decreto legislativo e poste in essere dal Ministero del lavoro e delle politiche sociali, le regioni e le province per i rispettivi ambiti territoriali di riferimento. Le relative indagini statistiche sono effettuate in forma anonima”*.

Un richiamo alle attività di monitoraggio e di valutazione attraverso la valorizzazione dei dati amministrativi è contenuto nell'Articolo 1 della legge 92/2012, dove ai commi 5 e 6 si prescrive che *“Al fine di monitorare lo stato di attuazione degli interventi e delle misure di cui alla presente legge e di valutarne gli effetti sull'efficienza del mercato del lavoro, sull'occupabilità dei cittadini, sulle modalità di entrata e di uscita nell'impiego, e' istituito presso il Ministero del lavoro e delle politiche sociali, in collaborazione con le altre istituzioni competenti, un sistema permanente di monitoraggio e valutazione basato su dati forniti dall'Istituto nazionale di statistica (ISTAT) e da altri soggetti del Sistema statistico nazionale (Sistan)..... (Comma 5) Le banche dati di cui al comma 4 contengono i dati individuali anonimi, relativi ad età, genere, area di residenza, periodi di fruizione degli ammortizzatori sociali con relativa durata ed importi corrisposti, periodi lavorativi e retribuzione spettante, stato di disoccupazione, politiche attive e di attivazione ricevute ed eventuali altre informazioni utili ai fini dell'analisi di impatto e del monitoraggio.*

In questo contesto, il sistema informativo delle Comunicazioni Obbligatorie rappresenta, indubbiamente, una fonte primaria e la sua valorizzazione a fini statistici non può che costituire un obiettivo strategico del Ministero, sia nella prospettiva di un rafforzamento del Sistema Statistico Nazionale in materia di lavoro ed occupazione, sia per la realizzazione delle attività di monitoraggio e valutazione previste dalla legge.

Il quadro normativo che disciplina le comunicazioni obbligatorie discende, come è noto, dai commi dal 1180 al 1185 dell'articolo unico della legge 296/2006 (legge finanziaria per il 2007). In essi si disponeva che tutti i datori di lavoro pubblici e privati dovessero comunicare al servizio competente, esclusivamente in via telematica, l'ubicazione della sede di lavoro, l'instaurazione, la proroga, la trasformazione e la cessazione dei rapporti di lavoro. Le modalità di comunicazione, i tempi, le informazioni da comunicare sono stati successivamente definiti nel decreto del *Ministro del lavoro e della previdenza sociale* del 30 ottobre 2007 che ha adottato i modelli con i quali tutti i datori di lavoro pubblici e privati potessero comunicare, oltre alle

¹ ISTAT Decima Conferenza Nazionale di Statistica 2012 – Enrico Giovannini Statistica 2.0: the next level

informazioni sull'unità produttiva, tutte le informazioni riguardanti il lavoratore e il rapporto di lavoro oggetto della comunicazione stessa ed a partire dal 1° marzo 2008 la comunicazione di tali informazioni è realizzabile esclusivamente per via telematica.

Nel tempo i contenuti informativi della comunicazione – come chiarito nella documentazione tecnica messa a disposizione dal Ministero del lavoro² - sono stati progressivamente aggiornati alle trasformazioni normative che hanno interessato i contratti di lavoro dipendente e parasubordinato, dando vita ad un sistema informativo dinamico, che, grazie ad un sistema di interscambio telematico in tempo reale tra il livello nazionale e quello regionale consente di tracciare tutte le attivazioni, le proroghe, le trasformazioni e le cessazioni di tutti i rapporti di lavoro dipendente e parasubordinato, rilevando altresì le variazioni e le cessioni dei rami di azienda, al fine di tracciare tutti i flussi amministrativi legati alle attività di lavoro, ad esclusione di quelle autonome e professionali.

La trasformazione di tale patrimonio informativo e la sua valorizzazione a fini statistici, proprio per la natura dei flussi tracciati si è dimostrata assai complessa e per il suo sviluppo è stato costituito un apposito Gruppo di lavoro tecnico formato da esperti del Ministero del Lavoro, dell'Inps, dell'Isfol, dell'Istat e di Italia Lavoro che nel tempo ha sviluppato specifiche procedure per la normalizzazione e la messa in qualità di una prima parte degli archivi che compongono il sistema Informativo delle CO. Il presente documento, insieme alla documentazione tecnica, si propone di illustrare in forma sintetica gli aspetti chiave di tale processo evidenziando le principali scelte di natura metodologia che hanno portato alla realizzazione del Sistema Informativo Statistico delle Comunicazioni Obbligatorie (SISCO).

² Si veda <http://www.cliclavoro.gov.it/Aziende/Adempimenti/Pagine/Area-download.aspx>

2 Le Fonti Amministrative

Al fine di rendere operative le disposizioni normative che istituivano il sistema informativo delle CO il Ministero del Lavoro, in regime di cooperazione con le Regioni, ha realizzato una soluzione tecnologica per tutti i soggetti obbligati e abilitati, che permette ai differenti sistemi regionali di dialogare tra loro rendendo immediata l'applicazione del provvedimento su tutto il territorio nazionale.

Oggi, quindi, le informazioni sono contenute in un sistema gestionale che assicura la raccolta e la conservazione dei dati. Gli attori principali di tale processo amministrativo sono quindi da una parte i datori di lavoro o eventuali delegati (consulenti del lavoro, ...) le agenzie di somministrazione, cui spetta l'obbligo dell'invio, dall'altra le Regioni e i nodi periferici preposte alla ricezione delle comunicazioni e, come punto di raccordo, il nodo centrale identificato dal Ministero del Lavoro.

La comunicazione si dice unica perché è sufficiente inviarla al servizio informatico regionale ove è ubicata la sede di lavoro perché questa sia messa a disposizione, tramite il nodo di coordinamento nazionale, dei soggetti definiti per decreto (enti previdenziali e assistenziali, Prefetture in caso si tratti di rapporti riguardanti cittadini extracomunitari, ecc.).

Il Sistema delle comunicazioni obbligatorie è basato su un insieme di informazioni e dizionari standard, relativi ai modelli per la trasmissione dei dati, ai dizionari terminologici di riferimento per la classificazione delle informazioni e alle modalità tecnologiche per lo scambio dei modelli telematici inviati dai datori di lavoro o da loro delegati. Il sistema è dinamico, per consentire l'adeguatezza alle normative in vigore sui rapporti di lavoro, e prevede che i processi di adeguamento e manutenzione avvengano con regole precise, concordate dal Ministero del lavoro insieme alle Regioni. Gli standard utilizzati in modo uniforme sul territorio nazionale, consentono di gestire le comunicazioni attraverso il sistema di cooperazione applicativa delle Regioni e province Autonome e di condividere le informazioni in modo semplice ed efficace. Il vantaggio per gli utenti è di avere un sistema che garantisce, a livello nazionale, uniformità anche all'interno di differenti sistemi amministrativi territoriali.

Il sistema delle CO raccoglie i dati sui flussi occupazionali relativi ai rapporti di lavoro subordinato, associato, di tirocini e di altre esperienze professionali previste dalla normativa vigente³ che interessano cittadini italiani e stranieri anche solo temporaneamente presenti nel Paese, in possesso di regolare permesso di soggiorno (lavoro stagionale)⁴. Pertanto le fonti amministrative che alimentano il sistema amministrativo corrispondono ai diversi modelli di CO previsti dalla legge, di seguito elencati:

- **UNILAV:** modulo informativo mediante il quale tutti i datori di lavoro pubblici e privati, di qualsiasi settore (ad eccezione delle agenzie per il lavoro, relativamente ai rapporti di somministrazione) adempiono all'obbligo di comunicazione dell'assunzione dei lavoratori, della proroga, trasformazione e cessazione dei relativi rapporti di lavoro, direttamente o tramite i soggetti abilitati.
- **UNISOMM:** modulo mediante il quale le agenzie per il lavoro adempiono all'obbligo di comunicazione relativo a tutte le tipologie di rapporti di somministrazione.
- **UNIURG:** modulo informatico mediante il quale tutti i datori di lavoro pubblici e privati, di qualsiasi settore (ad eccezione delle agenzie per il lavoro relativamente ai rapporti di somministrazione) adempiono all'obbligo di comunicazione di assunzione dei lavoratori, effettuate in casi di urgenza

³ Art. 4-bis del D.Lgs. n. 181/2000, così come modificato dall'art. 1, comma 1184 della L.296/2006.

⁴ Il Sistema non registra i rapporti di lavoro delle forze armate e quelli che interessano le figure apicali quali, ad esempio, presidenti e amministratori delegati di società pubbliche e private.

connesse ad esigenze produttive, fermo restando l'obbligo di invio della comunicazione mediante il modulo Unificato Lav entro il primo giorno utile».

- **VARDATORI:** da utilizzare per le comunicazioni di variazione della ragione sociale del datore di lavoro, trasferimento d'azienda per cessione o per fusione, cessione di ramo d'azienda, cessione di contratto.
- **UNIMARE:** costituisce il modello per l'invio delle Comunicazioni Obbligatorie e riguarda tutti i rapporti di lavoro che coinvolgono soggetti iscritti alle liste della Gente di Mare».

La prima scelta di metodo, quando ci si trova davanti a diversi moduli amministrativi, è quella di decidere il momento della loro integrazione. L'opzione metodologica del Ministero è quella di operare tutte le tecniche di trattamento del dato dalle singole fonti amministrative fino alla costituzione di un archivio statistico interrogabile di ogni singola fonte, e solo successivamente lavorare sulle integrazioni fra diversi archivi.

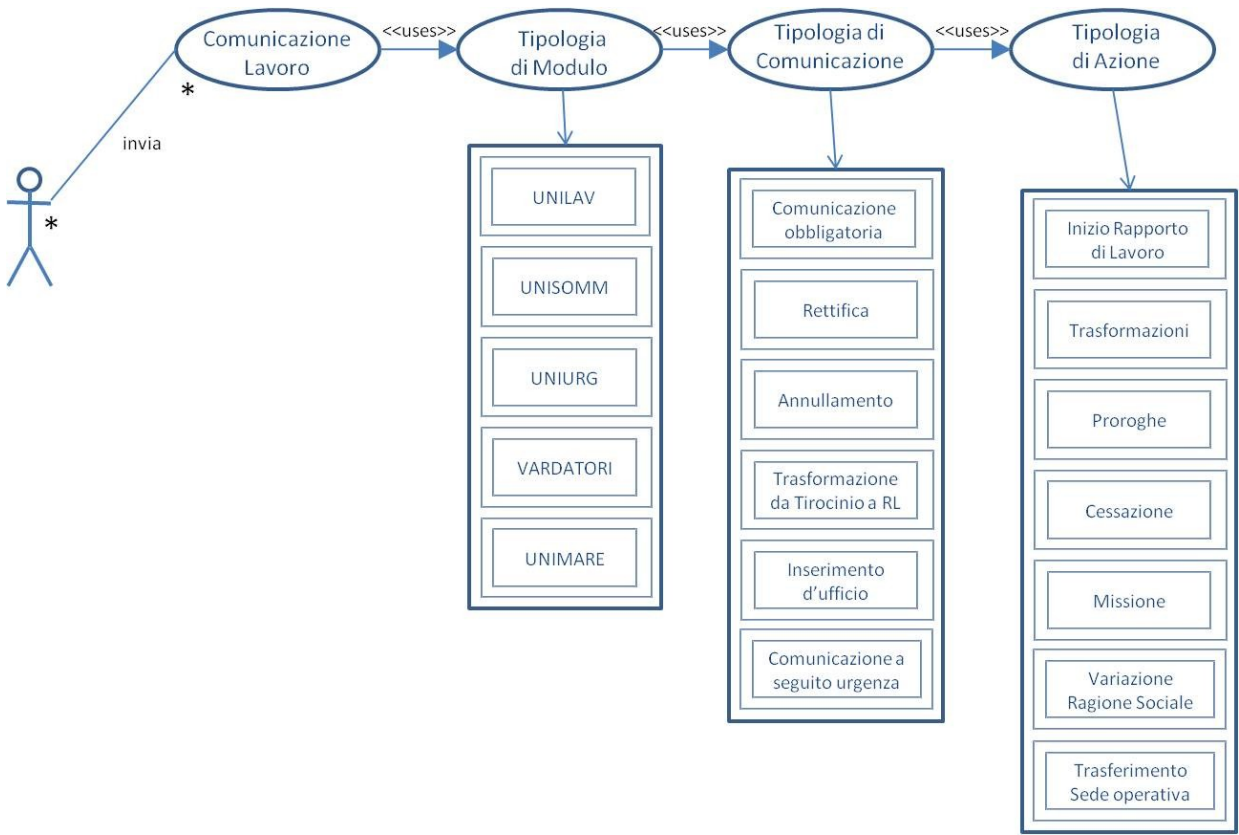
Esiste quindi un livello dati 'fonte amministrativa', costituito dalle CO strutturate attraverso un *xml repository*, e via via livelli informativi intermedi, fino a un livello dedicato alle aree informative statistiche denominate datamarts, su cui insistono le strutture dati specifiche (ipercubi) che permettono l'analisi delle unità statistiche (rapporti di lavoro, lavoratori, datori di lavoro).

I livelli indicati genericamente come 'intermedi' sono quelli che garantiscono il processo di 'ricostruzione' dell'informazione, gestendo comunque le operazioni di annullamento e rettifica delle singole comunicazioni e in modo tale che siano soddisfatti contemporaneamente i seguenti obiettivi:

- ✓ Minimizzazione della possibilità di acquisire errori presenti nella singola comunicazione, mediante un sistema di regole, che guida il caricamento delle comunicazioni obbligatorie, che ricostruisce la cronologia degli eventi e privilegia le informazioni strettamente inerenti all'azione che si sta comunicando. Per evento si intende ogni comunicazione di avviamento al lavoro, di trasformazione, proroga o cessazione. L'evento è l'elemento base su cui si fonda l'intero sistema informativo e di norma è caratterizzato da una data di inizio, eventualmente da una data di fine e da uno o più soggetti interessati (persone, imprese, ecc.).
- ✓ Massimizzazione della correttezza e coerenza delle informazioni inserite nel sistema da più comunicazioni tra loro collegate, attraverso l'implementazione di un processo di verifica ed eventuale eliminazione dell'informazione

Nel processo di trattamento delle CO, innanzitutto è necessario soffermarsi sull'insieme delle procedure propedeutiche alla trasformazione delle informazioni di natura amministrativa in dati statistici, per il controllo e la messa in qualità del patrimonio informativo delle CO. La figura 1 fornisce una prima rappresentazione del Sistema Informativo delle Comunicazioni obbligatorie evidenziando le azioni ed i contenuti che lo caratterizzano.

Figura 1 . Rappresentazione del contenuto del Sistema delle CO

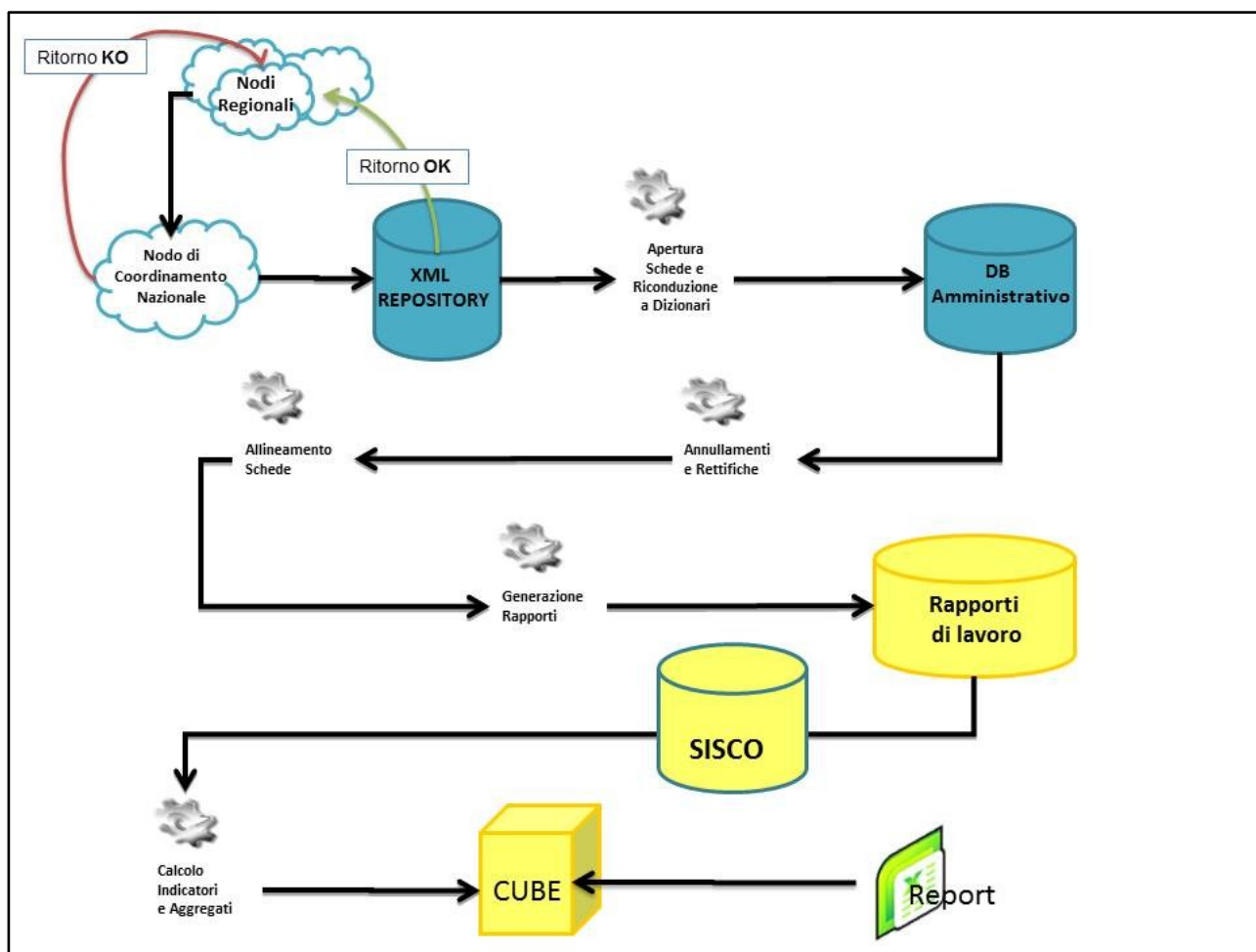


Fonte: MLPS - Il sistema di Data Warehouse e Business Intelligence delle Comunicazioni Obbligatorie 2009

3 Dal dato amministrativo al dato statistico: i flussi informativi delle Comunicazioni Obbligatorie

Le CO inviate dai datori di lavoro (o loro intermediari) vengono raccolte sul Nodo di Coordinamento Nazionale, e successivamente elaborate per la costruzione del database amministrativo (COB), da cui parte il processo di trattamento e trasformazione del dato statistico (SISCO) come rappresentato nella **figura 2**.

Figura 2. Il processo di costruzione dell'archivio amministrativo e di quello statistico



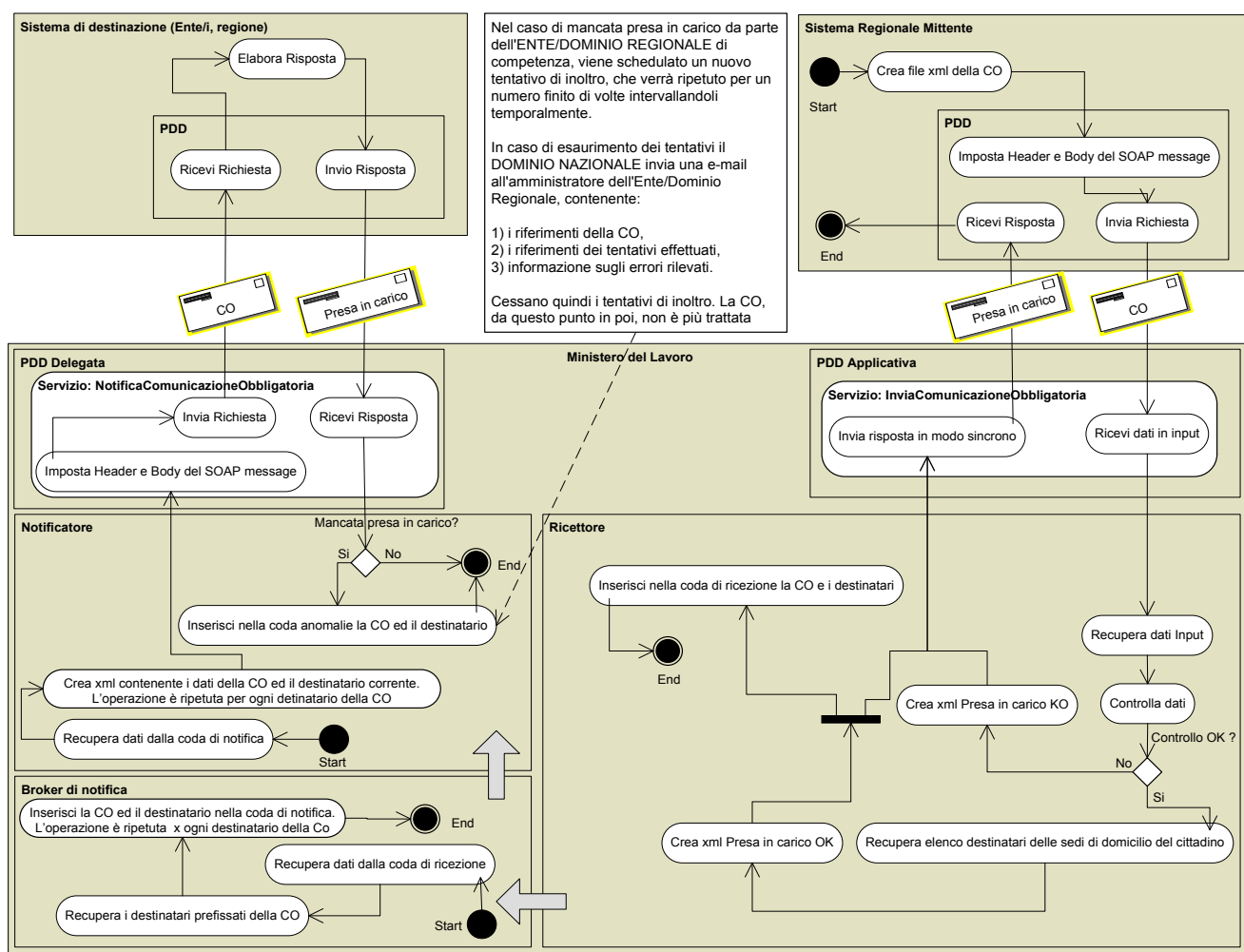
Fonte: Ministero del lavoro - Il sistema di Data Warehouse e Business Intelligence delle Comunicazioni Obbligatorie 2009

Ad ogni caricamento dei dati amministrativi pervenuti il processo viene rieseguito interamente riconsiderando tutte le CO arrivate, comprese quelle scartate nei flussi precedenti. Quindi il sistema COB e quello SISCO vengono rigenerati interamente ad ogni caricamento. Nelle pagine seguenti sono illustrati i singoli processi che alimentano i diversi database dei dati.

4 Il Nodo di Coordinamento Nazionale

Al fine di evitare C.O. affette da errori che ne impediscano l'utilizzo, la porta di dominio del Nodo di Coordinamento Nazionale (NCN) applica dei controlli, definiti da decreti ministeriali, che impediscono alle CO affette da errori provenienti dai nodi regionali di entrare nel sistema nazionale. Queste C.O. inutilizzabili vengono restituite al nodo regionale con un messaggio che ne segnala il tipo di errore. La complessa struttura di tale processo è rappresentata nella figura 3, tratta dalla documentazione tecnica messa a disposizione da Ministero del Lavoro (Allegato G).

Figura 3 - Diagramma di attività: controlli e funzioni previste per il dialogo tra le porte di dominio



Fonte: Allegato G Activity Diagram

Inoltre sempre negli standard tecnici (Allegato H) di trasmissione delle CO dai nodi di raccolta regionali verso il Nodo di Coordinamento Nazionale (NCN) è definito un livello minimo di controlli aggiuntivi per verificare la correttezza formale (es: verifica della compilazione dei dati obbligatori, formato dei campi, ecc.) e relazionale (es: controllo delle caratteristiche del lavoratore coinvolto nel rapporto di lavoro, controllo delle date di fine rapporto in caso di contratto a TD, ecc.) delle Comunicazioni Obbligatorie; i

controlli effettuati sulla porta di dominio del NCN del Ministero sono indicati dettagliatamente nell'Allegato H agli standard e hanno ripercussioni sui seguenti moduli:

- Interfacce Web delle piattaforme CO
- Import massivo
- Cooperazione applicativa

Sulla base degli stessi decreti vengono anche definiti gli standard di classificazione delle variabili; cfr. <http://www.cliclavoro.gov.it/Aziende/Adempimenti/Pagine/Area-download.aspx> per il dettaglio dei controlli e degli standard.

Le CO che non superano i controlli tecnici del NCN vengono rimandate al nodo da cui provengono con un messaggio "KO" e il dettaglio dell'errore individuato. Tali CO dovranno essere corrette dal soggetto mittente e reinviare, con l'inizio del nuovo percorso di controllo sulle porte di dominio del NCN. Di seguito viene riportato il dettaglio delle CO accettate ('OK') dal nodo nazionale, di quelle rifiutate ('KO'), e la relativa percentuale sul totale per ciascun anno.

Tabella 1. Comunicazioni Obbligatorie accettate/rifiutate dalla porta di dominio del Nodo di Coordinamento Nazionale. Anni 2009 – 2013

ANNO	Comunicazioni OK	Comunicazioni KO	Totale CO	% rifiuti sul totale CO
2009	20.712.884	4.791.238	25.504.122	18,8
2010	22.461.765	3.556.548	26.018.313	13,7
2011	22.552.002	2.772.586	25.324.588	10,9
2012	23.976.148	1.728.045	25.704.193	6,7
2013	21.475.985	1.278.639	22.754.624	4,4

Fonte: MLPS Nodo di Coordinamento Nazionale

Grazie al progressivo adeguamento dei nodi regionali agli standard di cooperazione applicativa, la percentuale dei rifiuti è passata dal 18,8% nell'anno 2009 al 4,4% nel 2013.

5 L'archivio nazionale delle CO: XML Repository

Le CO che superano i vincoli del NCN entrano sotto forma di *file xml* all'interno dell'archivio dedicato (*XML Repository*). Da qui le CO vengono inviate ai seguenti destinatari⁵:

- ENTI:
 - ✓ INPS (destinatario fisso)
 - ✓ INAIL (destinatario fisso)
- SISTEMI REGIONALI DI DESTINAZIONE:
 - ✓ Il sistema regionale di competenza per la sede di lavoro del datore di lavoro
 - ✓ Il sistema regionale di competenza per la sede di domicilio del lavoratore
 - ✓ Il sistema regionale di competenza per la sede di lavoro precedente (se presente nella CO), comunicata quando si effettua una trasformazione del rapporto di lavoro o Il sistema regionale di competenza per la sede di lavoro del datore presso il quale il lavoratore è stato distaccato (se presente nella CO), comunicata quando si effettua una trasformazione del rapporto di lavoro;
- ALTRI DESTINATARI:
 - ✓ L'ente previdenziale comunicato nel rapporto di lavoro. Tale ente deve essere diverso da INPS e deve aver aderito al sistema delle comunicazioni obbligatorie
 - ✓ Direzione Territoriale del Lavoro
 - ✓ UTG se il lavoratore è di nazionalità extracomunitaria

I destinatari restituiscono al Ministero del Lavoro la notifica dell'avvenuta ricezione.

Il processo di lavorazione inizia sui *file xml*, dove viene effettuato un primo *check* che permette di individuare le CO non valide, per i seguenti motivi:

1. Quando la DATACOMUNICAZIONE è precedente al 11/01/2008, la CO è ritenuta non valida e il suo **IDSTATO = C**;
2. Quando ci sono comunicazioni duplicate, a parità di DATACOMUNICAZIONE e CODICECOMUNICAZIONE, la CO riceverà lo stato:
 - **IDSTATO = A** per l'occorrenza più recente
 - **IDSTATO = D** per tutte le altre;
3. Tutte le rimanenti Comunicazioni (cioè CO non Cancellate e senza duplicati a parità di DATACOMUNICAZIONE e CODICECOMUNICAZIONE) sono ritenute valide e il loro stato sarà contrassegnato da **IDSTATO = A**.

Sulla base di questi controlli vengono considerate ai fini del processo di definizione del DB amministrativo (COB) circa il 98,4% delle CO ricevute dal 1/gennaio/2009 al 31/12/2013. Negli anni il processo di raccolta è stato perfezionato negli aspetti applicativi e questo ha portato ad una radicale riduzione degli scarti.

Tabella 2. Comunicazioni Obbligatorie per IDSTATO. Anni 2009 – 2013 (valore assoluto e %)

ANNO / ID STATO	IDSTATO A	IDSTATO C	IDSTATO D	Totale CO	% IDSTATO = A	%IDSTATO = C,D
2009	21.197.147	28.740	127.905	21.353.792	99,27	0,73
2010	22.478.145	132.742	1.891	22.612.778	99,40	0,60
2011	23.458.073	6.367	1.483	23.465.923	99,97	0,03
2012	22.771.753	205.015	497.470	23.474.238	97,01	2,99
2013	21.539.550	5	8.848	21.548.403	99,96	0,04

⁵ Standard Tecnici_CO_20071106_v5.3, Allegato G (<http://www.cliclavoro.gov.it/Aziende/Adempimenti/Pagine/Area-download.aspx>)

Tabella 3. Numero di CO, per regione mittente e per anno, in valore assoluto e % di co scartate per IDSTATO diverso da A . Anni 2009 – 2013

Regione Mittente	CO inviate(v.a.)					Comunicazioni con IdStato = C,D (v.%)				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
Piemonte	1.035.632	1.040.630	1.038.804	995.569	917.466	0,0	0,0	0,0	1,7	0,0
Valle d'Aosta	51.247	52.762	39.004	63.488	43.801	0,1		0,0	2,2	0,0
Lombardia	3.712.695	3.335.098	3.454.112	3.447.442	3.238.555	2,4	0,0	0,0	2,4	0,0
Bolzano	276.265	286.412	295.731	294.402	278.336	0,0	0,0	0,0	2,9	-
Trento	253.510	256.245	255.427	251.262	239.681	0,2	0,0	0,0	1,6	0,0
Veneto	1.413.486	1.465.707	1.517.603	1.521.897	1.371.705	0,1	0,0	0,0	3,0	0,0
Friuli Venezia Giulia	318.757	305.013	321.529	314.042	269.048	0,1	0,0	0,0	2,7	0,0
Liguria	533.348	513.573	533.149	598.947	478.456	0,1	0,0	0,0	11,9	1,0
Emilia-Romagna	2.735.790	3.071.087	3.167.993	3.049.702	2.859.730	0,2	0,0	0,0	2,4	0,0
Toscana	1.145.596	1.179.551	1.169.542	1.175.085	1.067.760	0,9	0,0	0,0	2,9	0,0
Umbria	286.086	323.836	296.919	311.710	278.809	0,0	0,0	0,0	2,9	-
Marche	475.823	483.641	505.794	502.939	441.858	0,1	0,2	0,0	2,6	0,0
Lazio *	2.399.286	3.187.573	3.344.278	3.428.034	3.242.516	0,5	0,0	0,2	3,7	0,0
Abruzzo	488.054	536.309	566.653	571.124	515.315	2,7	0,0	0,0	2,0	0,0
Molise	312.756	362.286	396.556	363.783	333.634	0,1	0,9	0,0	0,8	0,0
Campania	1.563.844	1.692.635	1.666.147	1.799.941	1.628.905	0,2	2,6	0,0	3,3	0,1
Puglia (b)	1.589.991	1.620.991	1.669.590	1.794.355	1.656.348	0,1	0,0	0,0	2,8	0,0
Basilicata	241.879	233.847	236.225	250.246	216.043	2,3	0,0	0,0	2,0	-
Calabria	541.902	573.134	601.833	614.404	565.820	0,1	0,0		2,2	0,0
Sicilia	1.474.125	1.600.816	1.891.711	1.620.664	1.462.919	0,2	5,3	0,0	2,9	0,1
Sardegna	502.918	491.632	497.323	505.202	441.698	1,5	0,0	0,0	3,6	0,0
N.D.	802	-	-	-	-	48,3	-	-	-	-
Totale	21.353.792	22.612.778	23.465.923	23.474.238	21.548.403	0,7	0,6	0,0	3,0	0,0

Fonte: MLPS xml repository

*La regione Lazio risulta essere la regione mittente anche per le Comunicazioni Obbligatorie che provengono da INPS sui datori di lavoro domestico.

6 Il database amministrativo (COB)

Il database amministrativo è costituito da una serie di tabelle del COB per ogni singolo modulo alimentate dalle procedure di estrazione dall'XML Repository. La fase di estrazione ("Parsing") interessa tutte e sole le CO che passano la validazione e si trovano quindi con IDSTATO=A. La fase di estrazione avviene giornalmente e prende in considerazione i 3 giorni precedenti ⁶.

I principi che governano questa fase dunque sono:

- ✓ estrazione e scrittura, su campi dedicati, di tutte le informazioni obbligatorie, obbligatorie condizionali e opzionali previste dai modelli come da indicazione dei documenti allegati ai decreti;
- ✓ ripulitura di tutti i campi testuali da caratteri sporchi (caratteri non stampabili, accenti, etc.) e uniformazione di maiuscole e minuscole;
- ✓ ripulitura di tutti i campi di testo contenenti una data gestendo tutte le possibili eccezioni e riconducendole al formato gg/mm/aaaa;
- ✓ i dati archiviati non subiscono alcun trattamento di riclassificazione.

Dal database amministrativo è possibile studiare la qualità delle C.O. trasmesse per anno, sistema mittente, modalità di invio e analizzare la stabilizzazione dei dati su COB in merito alla distanza fra la data dell'evento, i vincoli di obbligo temporali delle comunicazioni e la data di ricezione e popolamento di COB.

E' da sottolineare che al momento il MLPS prende in considerazione solo il modulo UNILAV per l'analisi statistica, sfruttando, quindi, solo una parte delle informazioni contenute nel sistema amministrativo.

⁶ La verifica sui 3 giorni precedenti consente di evitare perdita di informazioni dovuta a blocchi temporanei o malfunzionamenti del sistema.

7 Processo propedeutico alla costruzione di SISCO

U Nel processo di costruzione del Sistema Informativo Statistico Comunicazioni Obbligatorie (SISCO), nella prima fase vengono identificati i motivi di annullamento e di rettifica delle informazione elementari attraverso una serie di passi operativi

1) Gestione di annullamenti e rettifiche

Il *primo step* di questo processo è relativo all'individuazione nel database amministrativo di quelle CO che non possono essere contabilizzate nel database statistico attraverso la gestione di annullamenti e rettifiche. Nei diversi moduli CO (ad eccezione di UniUrg), oltre al codice comunicazione è presente anche il codice di un'eventuale comunicazione precedente ad essa collegata, che permette di verificare CO collegate tra loro. Risalendo, attraverso il codice comunicazione, alla CO precedente univocamente identificata, si procede con l'esclusione delle: 1) CO Annullate; 2) CO di Annullamento; 3) CO Rettificate.

Tabella 4 – Comunicazioni UNILAV (tutti gli eventi) per regione sede di lavoro, per anno evento, valori assoluti e percentuali di scarti amministrativi (annullamenti e rettifiche) Anni 2009 - 2013

Regione sede Di lavoro	CO entrate nel NCN (v.a.)					CO di annullamento, annullate e rettificate (%)				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
Piemonte	1.098.523	1.094.459	1.111.372	1.051.401	974.984	6,9	6,5	6,0	5,1	5,4
Valle d'Aosta	57.809	59.734	56.490	61.627	50.782	7,3	6,9	6,4	6,2	6,3
Lombardia	2.829.847	2.920.188	3.043.940	3.003.446	2.768.150	6,6	6,5	6,6	6,7	7,0
Bolzano	282.004	293.093	299.947	290.075	284.008	8,8	9,3	8,8	8,3	8,9
Trento	258.971	266.152	266.700	260.770	248.678	7,8	8,2	8,3	8,0	8,5
Veneto	1.404.533	1.422.948	1.490.462	1.449.630	1.332.327	7,9	7,7	7,6	7,0	7,5
Friuli Venezia Giulia	349.212	343.483	350.702	341.829	303.152	8,1	7,8	7,8	7,0	7,2
Liguria	467.552	463.650	482.003	474.685	420.196	7,4	7,4	7,5	7,0	6,9
Emilia-Romagna	1.531.683	1.549.358	1.636.298	1.612.421	1.468.181	6,3	6,1	6,0	5,5	5,9
Toscana	1.234.438	1.235.941	1.235.931	1.226.532	1.139.120	8,9	8,5	8,1	7,5	8,0
Umbria	261.046	267.739	269.272	263.823	239.975	5,2	5,1	5,4	5,2	5,4
Marche	488.390	495.088	513.642	497.956	441.602	9,2	8,7	8,4	8,0	8,6
Lazio *	2.184.255	2.342.116	2.453.794	2.425.970	2.319.032	6,2	6,4	7,1	7,4	8,0
Abruzzo	422.923	442.633	464.163	463.687	420.285	7,4	7,9	7,8	8,1	8,6
Molise	87.239	89.316	93.650	95.512	84.587	7,8	8,7	8,5	9,3	9,2
Campania	1.340.300	1.406.058	1.441.820	1.527.751	1.413.865	7,7	8,1	8,2	8,5	9,1
Puglia (b)	1.313.592	1.668.528	1.687.968	1.785.930	1.691.645	5,2	5,6	5,7	5,7	5,9
Basilicata	232.305	241.408	243.950	251.868	223.174	6,6	7,0	7,3	7,1	7,4
Calabria	549.857	578.167	603.792	604.842	560.284	7,4	7,9	8,6	8,6	8,6
Sicilia	1.456.182	1.505.743	1.605.555	1.563.538	1.450.079	6,7	7,0	7,4	7,0	7,8
Sardegna	544.026	548.617	572.675	565.008	524.297	7,9	8,3	8,6	7,9	9,5
N.D.	4.573	5.524	5.334	5.658	5.804	7,7	7,9	8,6	7,9	8,6
Totale	18.399.260	19.239.943	19.929.460	19.823.959	18.364.207	7,0	7,0	7,1	6,9	7,4

2) Allineamento agli standard tipologici

Il *secondo step* prevede il controllo e l'eventuale correzione della classificazione di alcune informazioni all'interno delle singole CO. In altre parole questa fase prevede l'abbinamento dei campi obbligatori ad identificativi interni al sistema di classificazione indicato dal decreto. In particolare viene effettuato un controllo sul codice contenuto nella CO rispetto a quello previsto nel sistema di classificazione vigente.⁷ Operativamente la procedura prevede l'inserimento del codice *Non disponibile* (N.D.):

- nel caso di dato obbligatorio mancante;
- nel caso di dato obbligatorio classificato in maniera errata.

In alcuni casi, quando i sistemi di classificazione vengono aggiornati e sopraggiungono CO in ritardo, cioè precedenti alla modificazione del sistema di classificazione si utilizza il Codice *Not available* (N.A) .

3) Scarti logici

Il *terzo step* comprende controlli destinati a classificare e ad escludere CO con palesi errori ed incongruenze. Quelle affette da errori vengono evidenziate con particolari codici in modo da NON essere considerate ai fini delle successive elaborazioni statistiche ma rimangono comunque all'interno del database, isolate in tabelle separate, in modo poterle utilizzare nel moneto in cui sopraggiungono altre CO che le rendono nuovamente utilizzabili ai fini statistici. Tali errori sono classificati in tre tipologie:

A. **Errori "amministrativi"**: tra questi rientrano quelli relativi a CO:

- ✓ "duplicate" di CO già esistenti;
- ✓ "non dovute", cioè quelle di cessazioni alla data prevista dal contratto ("cessazioni per zelo").

B. **Errori "sintattici"**: sono quelli di congruenza su campi che rendono impossibile identificare univocamente un contratto di lavoro

Schema errori sintattici per tipo di evento

Descrizione	Evento
Codice Fiscale Lavoratore Uguale al Codice Fiscale datore*	Tutti
Campi (cf_datore, cf_lavoratore, data_inizio_rapporto) non valorizzati*	Tutti
Data cessazione mancante	Cessazione
Data fine proroga mancante	Proroga
Datore lavoro vecchio = Datore lavoro nuovo	Vardatori(**)
Tipo_trasformazione o data_trasformazione mancanti	Trasformazione

*Tali controlli riguardano le comunicazioni acquisite dal nodo nazionale precedentemente alla data di implementazione dei controlli più stringenti sul nodo nazionale introdotte recentemente dal decreto

** Il ricorso al Modulo Vardatori si rende necessario per verificare appunto errori dovuti al cambio di datore di lavoro per contratti a seguito di variazione o cessione di rami di azienda

⁷ I sistemi di classificazione, come è noto, sono dinamici nel senso che possono essere soggetti a modifiche nel tempo.

- C. **Errori “semantici”**: sono errori che controllano la coerenza interna di alcune informazioni presenti nelle singole comunicazioni. Viene di seguito riportato uno schema riepilogativo in cui sono elencati i vincoli minimali che devono essere soddisfatti affinché una serie di eventi riferiti a uno stesso rapporto di lavoro possano essere considerati validi per la ricostruzione del rapporto stesso e per il calcolo della durata del rapporto stesso.

Schema errori semantici per tipo di evento

Descrizione Errore	Evento
Data fine presunta precedente alla data inizio	Attivazioni TD
Differenza tra data inizio e data fine presunta > 100 anni	Attivazioni TD
Data fine presunta valorizzata per contratti a tempo indeterminato	Attivazioni
Data fine presunta nulla per contratti a tempo determinato	Attivazioni
Data fine distacco errata	Trasformazioni
Data fine trasformazione precedente alla data inizio trasformazione	Trasformazioni

*Tali controlli riguardano le comunicazioni acquisite dal nodo nazionale precedentemente alla data di implementazione dei controlli più stringenti sul nodo nazionale (cfr. Decreti)

Schema errori semantici per tipo di evento con necessità della ricostruzione del rapporto di lavoro per la loro corretta valutazione

Descrizione Errore	Evento
Data fine presunta precedente alla data inizio*	Proroghe –Trasformazioni - Cessazioni
Differenza tra data inizio e data fine presunta > 100 anni*	Proroghe –Trasformazioni - Cessazioni
Data fine presunta valorizzata per contratti a tempo indeterminato*	Proroghe –Trasformazioni - Cessazioni
Data fine presunta nulla per contratti a tempo determinato*	Proroghe –Trasformazioni - Cessazioni
Cessazione di contratto già cessato	Cessazioni
Data cessazione effettiva successiva alla cessazione a termine	Cessazioni
Differenza tra data inizio e data cessazione effettiva > 100 anni	Cessazioni
Data cessazione effettiva precedente all'inizio contratto	Cessazioni
Data fine proroga precedente alla cessazione definita nel contratto	Proroghe
Differenza tra data inizio contratto e data di proroga > 100 anni	Proroghe
Proroga su contratto a tempo indeterminato	Proroghe
Data fine distacco non coerente	Trasformazioni
Differenza tra data inizio contratto e fine trasformazione > 100 anni	Trasformazioni
Data inizio trasformazione più di 10 giorni fuori dalla durata del contratto	Trasformazioni
Orario contratto non trasformabile con codice trasformazione corrente	Trasformazioni
Contratto non trasformabile	Trasformazioni

Nella tabella seguente viene riportata la numerosità degli scarti effettuati sul modulo UNILAV: si precisa infatti che ai fini della ricostruzione dei rapporti di lavoro vengono considerate solo le CO che superano i controlli appena descritti.

Tabella 5 – Attivazioni di rapporti di lavoro per regione sede di lavoro, per anno, valori assoluti e percentuale di scarti logici, Anni 2009 - 2013

Regione Sede di lavoro	Attivazioni complessive (v.a.)					% scarti logici				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
Piemonte	537.365	532.372	538.486	502.487	466.052	1,4	0,9	1,2	1,6	0,9
Valle d'Aosta	31.351	33.504	32.534	34.442	27.833	0,8	0,5	1,2	0,9	0,4
Lombardia	1.434.374	1.461.138	1.507.042	1.447.101	1.321.195	1,4	0,9	1,3	1,4	0,8
Bolzano	137.928	144.404	148.033	141.065	139.212	4,4	1,9	1,3	1,4	1,2
Trento	132.435	134.513	134.642	131.673	125.102	0,9	0,9	0,9	0,9	1,3
Veneto	670.021	679.668	711.196	677.295	617.303	0,9	0,5	1,0	1,2	0,7
Friuli Venezia Giulia	172.050	166.013	170.844	165.350	143.162	1,1	0,9	1,2	1,3	0,8
Liguria	232.844	225.364	237.015	227.940	200.713	3,8	2,4	3,1	3,4	2,3
Emilia-Romagna	796.241	811.306	856.382	833.282	762.499	1,4	0,9	1,5	1,5	0,6
Toscana	640.296	647.011	641.561	630.687	589.444	1,7	1,1	1,4	1,8	1,0
Umbria	141.170	144.060	145.289	140.961	128.580	1,8	0,9	1,2	1,7	0,8
Marche	239.679	246.259	253.502	244.697	214.361	2,1	1,7	1,9	2,0	1,2
Lazio	1.364.176	1.460.208	1.503.348	1.461.924	1.416.627	2,4	2,0	2,1	2,5	1,5
Abruzzo	218.881	226.854	240.051	241.040	218.641	2,2	1,5	1,6	1,5	1,0
Molise	41.787	42.882	45.078	46.353	41.622	1,4	0,7	1,1	1,1	0,8
Campania	754.009	783.326	794.125	837.634	779.062	1,3	2,4	1,4	1,1	1,0
Puglia (b)	800.838	977.324	983.072	1.034.438	1.002.996	0,8	0,7	1,0	0,7	0,5
Basilicata	129.806	135.597	136.478	140.551	127.269	0,8	0,5	0,8	0,7	0,7
Calabria	335.701	347.450	351.833	344.754	330.469	1,2	0,9	1,1	1,0	1,4
Sicilia	822.202	827.439	862.385	827.829	783.585	1,0	0,9	1,1	1,0	1,0
Sardegna	282.351	285.418	294.856	292.202	269.559	1,4	0,9	1,4	1,9	0,8
N.D.	2.363	2.716	2.676	2.585	2.877	1,1	0,5	1,2	0,7	0,9
Totale	9.917.868	10.314.826	10.590.428	10.406.290	9.708.163	1,5	1,2	1,4	1,5	1,0

Fonte: MLPS

Tabella 6 – Cessazioni di rapporti di lavoro per regione sede di lavoro, per anno, valori assoluti e percentuale di scarti logici, Anni 2009 - 2013

Regione sede di lavoro	Comunicazioni inviate (v.a.)					% scarti logici				
	2009	2010	2011	2012	2013	2009	2010	2011	2012	2013
Piemonte	563.694	547.776	555.660	532.613	499.486	7,3	6,7	6,3	5,3	4,8
Valle d'Aosta	32.437	33.362	32.988	34.988	29.449	5,7	4,4	4,2	3,3	2,9
Lombardia	1.470.379	1.517.646	1.563.491	1.548.839	1.420.490	7,5	7,4	7,4	6,6	5,9
Bolzano	131.879	141.051	146.284	144.144	141.695	10,0	3,8	3,3	2,7	2,3
Trento	135.789	138.321	138.815	137.365	129.921	6,2	5,6	5,2	4,5	4,3
Veneto	698.089	695.173	730.166	719.666	660.784	6,9	5,8	5,9	5,1	4,4
Friuli Venezia Giulia	176.032	174.026	177.276	176.139	157.401	8,4	6,5	6,0	5,0	4,5
Liguria	226.921	226.429	240.250	239.835	211.735	9,0	6,8	6,8	5,9	5,0
Emilia-Romagna	803.120	821.669	858.804	864.480	793.183	6,3	5,1	4,7	3,9	3,3
Toscana	650.888	649.379	648.619	660.152	610.822	6,8	5,6	5,3	4,5	4,0
Umbria	143.124	147.685	148.123	147.685	137.938	8,2	7,7	5,0	4,2	3,4
Marche	249.888	251.842	257.552	258.621	233.189	6,9	7,2	5,2	4,1	3,8
Lazio *	1.333.614	1.455.436	1.514.522	1.489.385	1.453.534	6,3	5,1	4,4	3,9	3,0
Abruzzo	222.556	233.242	244.447	249.054	231.217	7,5	6,3	4,3	3,4	2,8
Molise	44.875	44.882	47.573	49.745	44.766	8,4	7,0	5,8	4,9	4,0
Campania	765.688	799.613	817.949	849.865	794.203	5,8	6,0	4,4	3,5	2,9
Puglia (b)	817.023	1.008.710	1.026.125	1.079.575	1.047.566	6,6	5,8	5,4	4,3	3,3
Basilicata	136.109	140.657	143.285	145.101	131.411	5,7	4,7	4,7	3,1	2,8
Calabria	339.330	351.292	360.371	355.893	336.182	3,7	3,1	2,6	2,5	2,0
Sicilia	835.063	846.574	897.794	860.445	818.088	4,6	3,9	4,3	3,1	2,2
Sardegna	300.746	299.933	309.502	302.788	281.976	9,4	7,5	6,3	5,3	4,1
N.D.	2.261	2.707	2.574	2.500	2.777	4,6	5,0	3,7	3,9	5,2
Totale	10.079.505	10.527.405	10.862.170	10.848.878	10.167.813	6,7	5,8	5,3	4,5	3,7

7.1 Definizione rapporti di lavoro

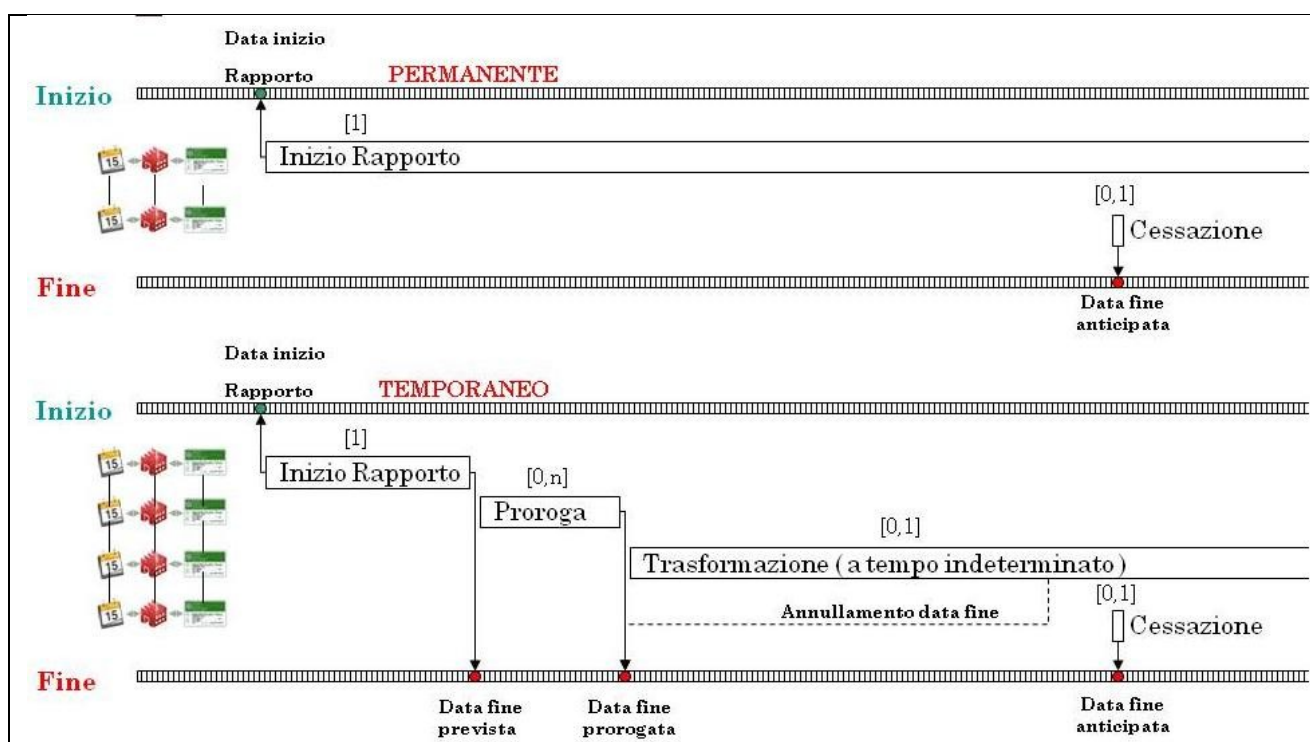
Sino a questo punto la ricostruzione dei contenuti informativi ha riguardato singoli eventi. Infatti, come già detto, ogni comunicazione obbligatoria – ossia l'informazione elementare - è definibile *come un evento osservato in un certo momento temporale* di un certo tipo: un avviamento al lavoro, una trasformazione, una proroga, una cessazione. L'evento è l'elemento base su cui si fonda l'intero sistema informativo e di norma è caratterizzato da una data di inizio, eventualmente da una data di fine e da uno o più soggetti interessati (persone, imprese, ecc.).

Tali eventi, al fine di aumentare il loro contributo informativo, possono essere aggregati in rapporti di lavoro, considerando cioè tutti gli eventi successivi e contigui che legano due soggetti (lavoratore e datore di lavoro, ad esempio la filiera avviamento, proroga, trasformazione, cessazione) e che, appunto,

concorrono alla creazione di un unico rapporto di lavoro. Il rapporto rappresenta il massimo livello di aggregazione degli eventi e il punto di partenza per tutte le aggregazioni successive. In questo senso esso rappresenta una nuova unità statistica che, appunto è la combinazione di più eventi.

Un **rapporto di lavoro** viene, quindi, definito dalla relazione fra un datore di lavoro e un lavoratore rispetto ad una stessa data inizio: la data inizio di un rapporto di lavoro, infatti, è un'informazione sempre presente in qualsiasi movimento. Da esso si possono analizzare le durate effettive dei rapporti di lavoro, oltre a ricostruire le storie occupazionali dei soggetti e la domanda dei datori di lavoro. Il modello logico di ricostruzione dei rapporti di lavoro è rappresentato nella figura 4 che illustra tale processo di concatenazione di eventi.

Figura 4. Rappresentazione dei rapporti di lavoro come connessione di eventi.

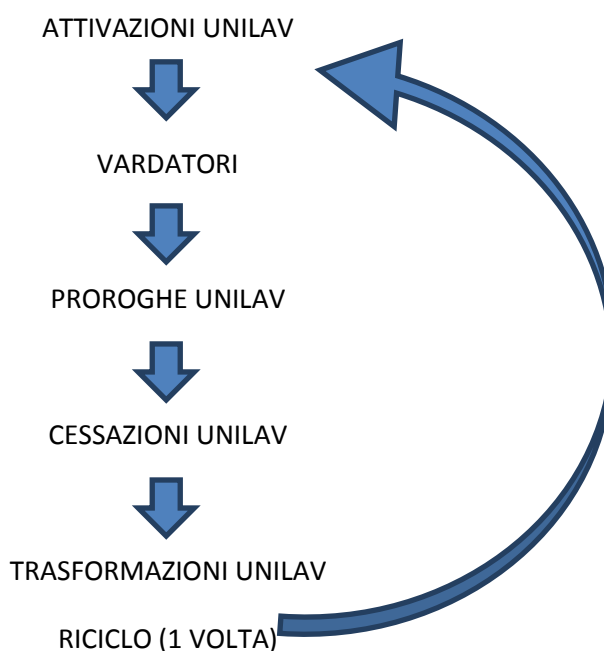


In questa fase vengono integrate le informazioni provenienti dal modulo VARDATORI in modo da non perdere riferimenti a rapporti di lavoro per cui sia cambiato il datore di lavoro in seguito a trasferimenti o cessioni di rami di azienda. Quindi le CO del modulo VARDATORI, sebbene non considerate in termini numerici ai fini dell'analisi, hanno comunque un impatto sulle CO di UNILAV in termini di correttezza delle ricostruzioni dei rapporti di lavoro.

In questa procedura vengono fatte la maggior parte delle attività di validazione delle CO e di ricostruzione dei rapporti di lavoro. Le CO vengono elaborate in sequenza in base all'ordine di arrivo e processate attraverso tutto il flusso. Al termine vengono riprocessati gli scarti nella fase di RICICLO.

Per quanto riguarda il modulo UNILAV le fasi del processo sono le seguenti:

Figura 5 Ricostruzione dei rapporti di lavoro.



L'algoritmo che determina la data di cessazione di un rapporto di lavoro si può descrivere con una serie di controlli a cascata. Se l'algoritmo entra in un ramo valorizza la data fine effettiva del contratto e esce dal ciclo. I controlli sono per semplicità di esposizione riportati di seguito (per il dettaglio degli algoritmi utilizzati cfr. documento tecnico):

La data fine effettiva viene valorizzata con:

- ✓ la data di fine rapporto se presente un movimento di cessazione
- ✓ la data di trasferimento di contratto in caso di VARDATORI (chiusura del rapporto di lavoro per modifica della datore di lavoro),
- ✓ con nessuna data se il contratto sin dalla sua attivazione è a tempo indeterminato
- ✓ con nessuna data se l'ultimo movimento disponibile è una trasformazione a tempo indeterminato
- ✓ con la data di scadenza dell'ultima proroga se presente almeno un proroga
- ✓ con la data fine prevista se non ci sono ulteriori movimenti associati al movimento di inizio rapporto e il rapporto di lavoro è di carattere temporaneo

8 Il Sistema Informativo Statistico delle Comunicazioni Obbligatorie (SISCO)

Una volta definito il concetto di rapporto di lavoro come connessione di più eventi elementari è ora possibile descrivere la struttura del nuovo *Sistema Informativo Statistico delle Comunicazioni Obbligatorie* (SISCO) che consente di eseguire analisi su dati con un'architettura tipica di un Data Warehouse. SISCO è alimentato dagli stessi dati che viaggiano sul nodo nazionale delle comunicazioni obbligatorie attraverso la rete SCPCOOP attraverso repliche periodiche dal sistema di produzione.

SISCO permette di organizzare sia le comunicazioni (eventi), sia i rapporti di lavoro ampliando significativamente il potenziale informativo, che grazie al processo di ricombinazione degli eventi elementari permette di soddisfare tre classi di esigenze analitiche: un monitoraggio congiunturale tempestivo dell'andamento occupazionale del mercato del lavoro; l'analisi approfondita di politiche del lavoro e dei comportamenti degli soggetti su tale mercato; analisi della struttura della domanda di lavoro:

- a. Monitoraggio congiunturale
 - ✓ Tempestività
 - ✓ Dettaglio territoriale e settoriale
 - ✓ Indicatori di interesse (attivazioni, cessazioni, trasformazione a tempo indeterminato, durata dei contratti, ecc...)
 - ✓ La relazione con altre fonti statistiche sul mercato del lavoro
- b. Analisi delle politiche del lavoro e dei comportamenti
 - ✓ L'efficacia delle politiche
 - ✓ Effetti delle agevolazioni
 - ✓ La transizione verso forme di lavoro più stabile
 - ✓ Dinamiche di porzioni specifiche di forza lavoro (immigrati, apprendisti, ...)
- c. Analisi di composizione della domanda di lavoro
 - ✓ Chi assume?
 - ✓ Chi sono gli assunti?
 - ✓ Quali sono le caratteristiche dei rapporti di lavoro creati?

Le analisi svolte sui tempi di arrivo delle CO hanno permesso di monitorare il tempo di stabilizzazione del dato e di fissare a 50 giorni dalla chiusura del periodo di riferimento il momento migliore per effettuare lo scarico utile per la produzione dei report di monitoraggio e di pubblicazione dei rapporti sull'andamento del mercato del lavoro. I dati vengono quindi scaricati la notte del 20 del primo mese successivo la chiusura del trimestre di riferimento.

Nella tabella che segue è riportato il numero di CO relative al IV trimestre 2012 in relazione con il periodo di tabella 8 ricezione delle relative CO* divise per tipologia contrattuale.

Tabella 7 – Tempi di ricezione delle attivazioni per tipologia di rapporto di lavoro: IV trimestre. 2012

Periodo	Tipologia Rapporto di lavoro					Totale complessivo	Distribuzione %
	Tempo Indeterminato	Tempo Determinato	Apprendistato	Contratti di Collaborazione	Altro		
Prima del 1/10/12	66.053	325.085	8.941	107.037	49.403	556.519	11,37
Periodo (1/10/12 - 31/12/12)	1.259.750	2.220.963	111.665	248.282	225.876	4.066.536	83,10
Da 1/1/13 a 10/1/13	108.723	41.213	4.400	10.495	9.522	174.353	3,56
Dal 11/1/13 al 28/2/13	38.476	17.060	712	2.589	1.018	59.855	1,22
Dal 1/3/13 al 31/5/13	13.099	3.891	305	705	450	18.450	0,38
dopo il 1/6/13	9.499	6.311	451	856	544	17.661	0,36
Totale complessivo	1.495.600	2.614.523	126.474	369.964	286.813	4.893.374	100,00

Fonte: MLPS sistema SISCO

8.1 Ricostruzione rapporti di lavoro validi: la chiave

La ricostruzione dei rapporti di lavoro avviene a partire dagli eventi elementari messi in relazione attraverso una **chiave tripartita**⁸ composta dall'identificativo lavoratore, l'identificativo del datore di lavoro e data inizio del rapporto di lavoro.

La *rottura della chiave* costituisce uno dei problemi maggiori nella ricostruzione dei rapporti di lavoro. Le casistiche sono diverse ma la fenomenologia più ricorrente è rappresentata da **Rapporti di lavoro senza un movimento di inizio rapporto**. Si tratta di rapporti di lavoro che possono generare un movimento di inizio rapporto a partire dalle informazioni contenute nelle comunicazioni di proroga, trasformazione e cessazione; questi vengono distinti da quelli generati a partire da una comunicazione di inizio rapporto.⁹ In questo caso i movimenti di proroga, trasformazione e cessazione vanno ricondotti al relativo movimento di inizio rapporto a meno di ulteriori motivi di scarto logico della comunicazione (vd. Dettaglio scarti logici). E' noto che i movimenti non ricondotti ad un movimento di inizio rapporto sono la principale causa di instabilità del dato. L'assunzione di ricostruire un rapporto di lavoro a partire dalla data di inizio dichiarata in una comunicazione di proroga, attivazione o cessazione può creare un numero crescente di attivazioni per uno stesso periodo di tempo osservato a lag successivi (mancata stabilizzazione del dato) a causa della generazione di attivazioni da eventi temporalmente successivi.

Di seguito si riporta l'impatto delle *comunicazione improprie* sul sistema.

⁸ L'identificativo datore di lavoro, l'identificativo lavoratore e la data di inizio rapporto dichiarate sono state sottoposte in via sperimentale a procedure di controllo e correzione con archivi UNIEMENS e ASIA

⁹ A partire dal primo marzo 2008 nasce l'obbligo di legge relativo alla comunicazione telematica ma per sopperire ad eventuali lacune informatiche del sistema vengono ricostruire attivazioni non presenti ("attivazioni improprie"). Eventuali problemi di linkage devono essere risolti al punto precedente.

Tabella 9 – Attivazioni di rapporti di lavoro per anno, distinti fra attivazione proprie e attivazioni improprie con dettaglio dell'evento, Anni 2009 - 2013 (valori assoluti e percentuali)

Evento	2009		2010		2011		2012		2013	
	v.a.	%	v.a.	%	v.a.	%	v.a.	%	v.a.	%
Inizio Rapporto Lavoro	9.431.882	96,60	9.935.210	97,50	10.257.911	98,26	10.110.991	98,63	9.524.224	99,07
Attivazioni Improprie	332.360	3,40	254.962	2,50	181.605	1,74	140.392	1,37	89.766	0,93
Delle quali generate da										
<i>Proroghe</i>	<i>109.193</i>	<i>1,12</i>	<i>94.635</i>	<i>0,93</i>	<i>86.958</i>	<i>0,83</i>	<i>73.668</i>	<i>0,72</i>	<i>52.373</i>	<i>0,54</i>
<i>Trasformazioni</i>	<i>47.975</i>	<i>0,49</i>	<i>44.041</i>	<i>0,43</i>	<i>31.956</i>	<i>0,31</i>	<i>27.757</i>	<i>0,27</i>	<i>18.713</i>	<i>0,19</i>
<i>Cessazioni</i>	<i>175.192</i>	<i>1,79</i>	<i>116.286</i>	<i>1,14</i>	<i>62.691</i>	<i>0,60</i>	<i>38.967</i>	<i>0,38</i>	<i>18.680</i>	<i>0,19</i>
Totale Eventi	9.764.242	100,00	10.190.172	100,00	10.439.516	100,00	10.251.383	100,00	9.613.990	100,00

Fonte: MLPS sistema SISCO

Per quanto detto, i rapporti di lavoro, ricostruiti a partire dagli identificativi dichiarati nella comunicazione, generano un errore costante che ha impatto sulla stabilizzazione dei dati.

Per correggere tale errore è stato fatto uno studio per valutare strategie di correzione. Il primo studio ha analizzato tutte le comunicazioni che avevano un codice comunicazione precedente valorizzato, sebbene non fosse obbligatorio (ad esempio in caso di proroghe, trasformazioni o cessazioni). Il codice comunicazione precedente risulta essere una potenziale chiave univoca di ricostruzione dei rapporti di lavoro, ma questa è obbligatoria solo nei casi di annullamento o rettifica.

Il 2,5% degli eventi complessivi dopo le esclusioni riporta il codice comunicazione valorizzato. A partire da questo sottoinsieme è possibile stimare l'errore della chiave tripartita. Infatti il 98,7% delle comunicazioni in analisi si aggancerebbe utilizzando la chiave tripartita dichiarata al momento della compilazione della comunicazione.

Le comunicazioni che non si agganciano possiamo ora analizzarle verificando il tipo di errore per singola chiave. Da questa analisi di evidenza che sebbene le due comunicazioni siano effettivamente l'una riferita all'altra (attraverso il codice co precedente), si verifica quanto segue

1. Un solo termine della chiave è sbagliato (83,2%)
 - a. in uno dei due casi è stata inserita solo la data inizio differente (55,3%),
 - b. in uno dei due casi è stato inserito solo il codice fiscale del lavoratore in modo lievemente difforme(17,9%)
 - c. in uno dei casi solo il codice fiscale/partita iva del datore di lavoro è diverso (10,0%)
2. più di un termine della chiave è sbagliato (16,7%)

Da ciò ne consegue che è possibile correggere l'errore di chiave per una quota di oltre l'80 per cento dell'errore se si adottassero strategie di correzione che comportano la riduzione della chiave a due termini dichiarati abbinati ad un termine calcolato.

In particolare la seconda sperimentazione ha riguardato la realizzazione di un algoritmo di correzione dei problemi di chiave dei rapporti di lavoro. Tale algoritmo di tipo deterministico si basa sulle sole risorse interne alla fonte co. In particolare ogni termine della chiave può avere un suo sostituto nella stessa fonte. Il codice fiscale del lavoratore può essere sostituito con il codice fiscale ricalcolato, il codice fiscale/partita

iva del datore di lavoro può essere sostituito dalla posizione territoriale inail (PAT inail) e, infine, la data inizio può essere sostituita con un algoritmo di riposizionamento che tiene conto della data comunicazione, del tipo di evento e della sua casistica di invio più ricorrente.

Tale procedura ancora in fase sperimentale consentirebbe di abbattere l'errore di circa il 75%, rendendo pertanto residuale l'effetto da esso determinato (Baldi et al, 2011; Di Bella et al, 2011).

Riferimenti bibliografici

Baldi C., De Blasio G., Di Bella G., Lucarelli A., Rizzi R. (2011), *Turning the compulsory communication data into a statistical system: preliminary issues to set up an editing strategy*, Convegno intermedio SIS 2011, Nord America, maggio 2011, disponibile in <http://meetings.sis-statistica.org/index.php/sis2011/2011/paper/view/1311>;

Di Bella G., De Blasio G., Callori M., Lucarelli A. (2011), *A new administrative source on employment flows: aspects of the data consolidation process for statistical use*, SIS-VSP workshop on Enhancement and Social Responsibility of Official Statistics, Roma, 28-29 aprile 2011;