

L'INAIL NEL SISTEMA INFORMATIVO SULLE PROFESSIONI: DESCRIZIONE DEL SISTEMA E APPROCCIO STATISTICO METODOLOGICO ALLA CONDIVISIONE DELLA BASE DATI

ADELINA BRUSCO*

1. Introduzione

Negli anni più recenti si sta sviluppando sempre più l'idea di condividere informazioni detenute da vari soggetti e di renderle fruibili al maggior numero di utenti; ciò è possibile anche grazie agli sviluppi della tecnologia informatica che agevola il sistema di interscambio di flussi di dati partendo da un denominatore comune. Questo approccio metodologico è anche alla base di un progetto ampio e articolato denominato "Sistema Informativo sulle Professioni". L'obiettivo del lavoro di seguito esposto è descrivere sia il Sistema Informativo, sia le attività messe in atto da un gruppo di lavoro INAIL e Istat per consentire l'ingresso della base dati INAIL nel sistema stesso.

2. Il Sistema Informativo sulle Professioni

La logica che sottende il Sistema Informativo sulle Professioni è mettere insieme informazioni detenute da tutti gli Enti partecipanti, per ricostruire il contesto sociale ed economico e le condizioni in cui si svolgono tutte le professioni esercitate nel mercato del lavoro italiano. La prima difficoltà è rappresentata dal fatto che le informazioni messe in campo sono raccolte inizialmente dai vari soggetti partecipanti per finalità diverse e a vario titolo; la seconda è rappresentata dal fatto che le informazioni sono di fonte e natura differente. La condizione dettata dal sistema è che tutte le informazioni siano riorganizzate secondo una chiave comune che nello specifico è la Nomenclatura delle Unità Professionali dell'Istat (NUP); in particolare, l'edizione del 2006, che costituisce lo standard con la quale le informazioni sono connesse tra di loro e condivise.

* INAIL - Direzione Generale - Consulenza Statistico Attuariale.

I soggetti che contribuiscono al sistema sono diversi:

- l'Istat fornisce le stime degli occupati per categoria professionale della NUP e la descrizione delle stesse professioni;
- l'INAIL produce le statistiche sugli infortuni sul lavoro denunciati e sui casi mortali per gruppi e sottogruppi di unità professionali organizzati in una Banca Dati delle Professioni;
- l'Isfol mette a disposizione i dati sui fabbisogni occupazionali nel breve e medio periodo per classi di professioni;
- il Ministero del Lavoro e delle Politiche Sociali ha attivato un collegamento col portale pubblico per il lavoro;
- Unioncamere fornisce le previsioni di assunzione nel breve termine e il trend occupazionale dell'ultimo triennio disponibile;
- la Regione Basilicata pubblica informazioni sull'offerta regionale di formazione professionale;
- l'Albo professionale degli agrotecnici e degli agrotecnici laureati, fornisce le caratteristiche degli iscritti all'albo e la loro distribuzione territoriale.

Il Sistema è da considerare dinamico perché via via si stanno definendo ulteriori rapporti di collaborazione con altri soggetti istituzionali.

Il Sistema Informativo sulle Professioni poggia su due aspetti:

- le informazioni messe a disposizione dai vari Enti sono fisicamente sui siti web di chi li produce che è responsabile, gestore e titolare delle stesse;
- ogni sito rappresenta una porta di accesso al Sistema tramite la chiave di aggancio costituita dalla NUP.

In altri termini, si può pensare schematicamente ad una struttura nella quale al centro è collocata la NUP, attorno alla quale gravitano tutti i soggetti che così entrano in contatto tra di loro. La stessa classificazione è esplosa a un livello di dettaglio più o meno ampio a seconda delle informazioni che ognuno dei partecipanti riesce a mettere a disposizione.

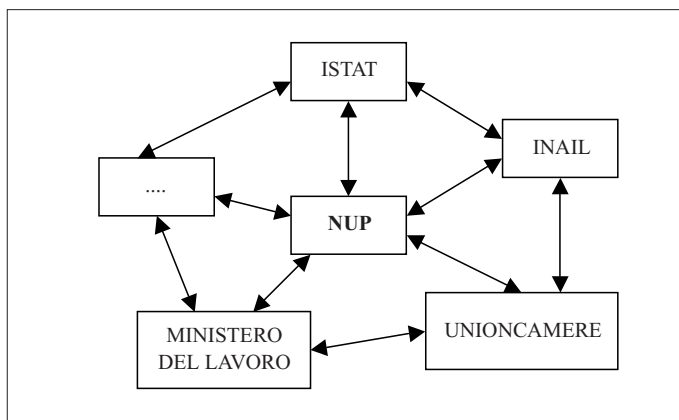


Grafico 1: *Schema del Sistema Informativo sulle Professioni.*

Un navigatore che entra da qualsiasi sito di uno dei soggetti partecipanti al Sistema Informativo sulle Professioni può avere non solo le informazioni messe a disposizione dall'Ente dal quale accede, ma conoscere le caratteristiche dell'unità professionale o del gruppo di professioni, accedere ad un altro Ente e avere informazioni sulle possibilità occupazionali, o sapere quanti sono gli occupati, conoscere l'iter formativo, può poi raggiungere il sito dell'INAIL e avere informazioni sui rischi lavorativi, informarsi, per date professioni, sugli iscritti all'Albo e conoscerne le varie distribuzioni e così via. Potrebbe anche accedere dall'ultimo punto del sistema descritto e informarsi, tramite operazioni identiche, dell'offerta formativa, dei rischi lavorativi, delle caratteristiche della professione, senza seguire un percorso definito o con una struttura rigida in cui il punto di partenza e di arrivo siano unici.

Il Sistema Informativo sulle Professioni diventa quindi l'insieme di tutti i percorsi di navigazione che si possono realizzare attorno ad una categoria, professione o gruppi di professioni omogenee. Le informazioni che sono assemblate dall'utente dipendono dalle scelte che lo stesso effettua durante il percorso di navigazione.

3. Il Nomenclatore delle Attività Professionali

Si è detto che il Sistema Informativo sulle Professioni poggia sulla Nomenclatura delle Unità Professionali dell'ISTAT (NUP) che rappresenta la base d'accesso dei soggetti partecipanti al Sistema. Una condizione necessaria è che gli Enti organizzino i propri dati amministrativi e statistici e le informazioni detenute secondo questo sistema di classificazione.

In particolare per l'avvio del progetto si è fatto riferimento all'edizione del 2006; va precisato, però, che la stessa è stata rivista dall'Istat nel corso del 2011 e che sono state realizzate delle tabelle di raccordo tra le due edizioni.

La NUP è organizzata con una struttura ad albero che prevede nove grandi raggruppamenti di professioni affini, di seguito indicate:

1. Legislatori, dirigenti e imprenditori;
2. Professioni intellettuali, scientifiche e di elevata specializzazione;
3. Professioni tecniche;
4. Impiegati;
5. Professioni qualificate nelle attività commerciali e nei servizi;
6. Artigiani, operai specializzati e agricoltori;
7. Conduttori di impianti e operai semiqualeficati di macchinari fissi e mobili;
8. Professioni non qualificate;
9. Forze Armate.

Ogni grande gruppo a sua volta si dettaglia attraverso una struttura ramificata che arriva ad individuare professioni simili (unità professionali).

Per comprendere la struttura si riporta, a titolo di esempio, il dettaglio semplificato del gruppo 5, che a livello successivo si distingue in:

- 5.1 - Professioni qualificate nelle attività commerciali;
- 5.2 - Professioni qualificate nelle attività turistiche ed alberghiere;
- 5.3 - Maestri di arti e mestieri, addestratori ed assimilati;
- 5.4 - Professioni qualificate nei servizi sanitari;
- 5.5 - Professioni qualificate nei servizi sociali, culturali, di sicurezza, di pulizia ed assimilati.

Le professioni qualificate nelle attività commerciali (5.1) a loro volta si suddividono in ulteriori rami ed esplodendo al maggior livello, sempre a titolo esemplificativo, solo il primo dei rami, si definiscono le seguenti classi e categorie:

- 5.1 - Professioni qualificate nelle attività commerciali;
 - 5.1.1 - Esercenti ed addetti alle vendite all'ingrosso;
 - 5.1.1.1 - Esercenti e gestori delle vendite all'ingrosso;
 - 5.1.1.1.0 - Esercenti e gestori delle vendite all'ingrosso.

Giunti al quinto livello, quello delle unità professionali, sono elencate le professioni che fanno parte del raggruppamento; in questo caso specifico:

agente distributore di giornali
distributore di acque gassate
distributore di ghiaccio
gestore di agenzia di distribuzione merci
gestore di magazzino di deposito
gestore di magazzino di vendita
grossista
conduttore di magazzino di deposito
conduttore di magazzino di vendita
consegnatario responsabile di magazzino all'ingrosso

Come si può notare ad ogni gruppo/classe/unità professionale corrisponde un codice univoco al massimo di 5 digit.

In totale sono descritte circa 800 unità professionali all'interno delle quali ricadono le professioni presenti sul mercato del lavoro nazionale. La NUP adottata non ha la pretesa di essere considerata esaustiva di tutte le professioni svolte, ma ne rappresenta la quasi totalità.

La logica adottata per accorpate le professioni nello stesso gruppo è quella di considerare due aspetti: il livello di competenza, inteso in termini di responsabilità, complessità e compiti svolti e il campo di competenza, inteso come bene o servizio prodotto e quindi materiali e attrezzature utilizzati.

Per poter accedere al Sistema, l'INAIL ha dovuto uniformarsi a questa variabile descrittiva delle professioni italiane e nei prossimi paragrafi sarà descritta l'attività di analisi e di adattamento della decodifica che individua le mansioni svolte dai lavoratori infortunati, mansioni che rispondono ad una logica completamente diversa, quella della rischiosità connessa alla lavorazione.

Nell'ambito del Sistema il valore aggiunto dato da INAIL è stato rilevante ed è consistito nel dare un'indicazione sulla rischiosità lavorativa tramite la fornitura del numero di infortuni sul lavoro e dei casi mortali denunciati in un triennio, con riferimento alle categorie professionali, distribuiti rispetto ad una serie di variabili di interesse.

4. La base dati dell'INAIL

Prima di spiegare l'impianto metodologico che sottende l'aggancio della NUP

alle qualifiche professionali dell'INAIL è necessario descrivere brevemente la base dati.

L'INAIL dispone delle informazioni relative agli infortuni sul lavoro occorsi ai soggetti per i quali vige l'obbligo all'assicurazione. Il datore di lavoro nel compilare il modulo di denuncia di infortunio sul lavoro inserisce una serie di informazioni che riguardano sia le modalità di accadimento dell'evento lesivo (dove e in che modo si è verificato, ...), sia le caratteristiche dell'infortunato (età, sesso, professione, qualifica, ...), sia quelle dell'azienda (codice che identifica il rapporto con l'Ente, attività svolta, ...).

Occorre precisare che alcune delle informazioni devono essere comunicate obbligatoriamente all'INAIL per consentire di connotare il rischio, l'azienda, la lavorazione e, soprattutto, per avviare l'iter amministrativo e sanitario che porterà, se ne sussistono i presupposti, all'erogazione della prestazione a seguito dell'infortunio. Alcune informazioni sono, invece, facoltative e tra queste rientra anche la descrizione della professione o mansione svolta.

Per quanto riguarda la descrizione della professione al momento dell'infortunio, l'INAIL fino a luglio del 2008 ha adottato un mansionario, composto da 533 voci, che individuava le specifiche attività, distinte per la gestione Agricoltura e per la gestione Industria e Servizi. Costruito nell'ottica di differenziare rispetto al rischio erano indicate precise professioni: muratore, infermiere, tornitore, mezzadro, ecc. e non era prevista nessuna possibilità di accorpamenti e aggregazioni per omogeneità. Questo primo mansionario presente nell'Istituto dagli anni '70 è stato rivisto qualche anno fa, per tener conto dell'evoluzione in campo lavorativo, ma ha lasciato intatta l'impalcatura. Il nuovo mansionario consta di 676 mansioni/professioni; ad ogni qualifica corrisponde un codice numerico di 4 digit (erano 3 digit nel vecchio mansionario).

Il nuovo mansionario rispetto al precedente ha accorpato attività simili (esempio banconiere e barista, oppure operatore turistico e guida turistica), eliminato professioni ormai obsolete o troppo generiche (es. bagnatore, operatore) e introdotto altre nuove (es. pony express, addetto al taglio laser).

La classificazione INAIL differisce dalla nomenclatura dell'ISTAT perché diverse sono le logiche e le finalità per le quali sono state realizzate. La NUP dell'Istat, come si è già detto, attraverso una strutturazione ad albero arriva a dettagliare gruppi di mansioni affini ed è molto specifica per alcune professioni (dettaglia per esempio le varie specializzazioni mediche, mentre l'INAIL distingue, in generale, se i medici sono esposti a radiazioni ionizzanti o meno). Il mansionario INAIL è realizzato in un'ottica di analisi del rischio e quindi individua singole professioni, ma resta talvolta generico per alcune attività come quella dell'impiegato amministrativo e dell'operatore (qualifica presente fino a metà anno 2008).

Occorre precisare che talvolta la classificazione INAIL riesce ad arrivare ad un livello di dettaglio più approfondito disponendo, in un certo senso, di un sesto livello rispetto al nomenclatore Istat. Ad esempio, mentre l'INAIL distingue la

professione dell'Orefice da quella del Salumiere o del Fornaio, considerando il diverso rischio di infortunio, l'Istat le accomuna nella stessa unità professionale (5.1.2.2.0 - Esercenti delle vendite al minuto), dal momento che svolgono la medesima attività lavorativa.

Il nuovo mansionario ha comportato, dal punto di vista statistico, un disallineamento delle serie storiche per via, appunto, delle nuove professioni e per la soppressione di alcune.

Come si diceva, la professione dell'infortunato non è un'informazione obbligatoria e infatti, una quota non trascurabile, pari al 15% circa delle denunce, ha qualifica non determinata.

Rilevante la percentuale di casi per i quali è indicata la mansione generica di operatore; si tratta di circa 173mila casi nel 2007 e di oltre 76mila nel 2008, ultimo anno in cui la professione è presente nel mansionario.

L'introduzione della nuova tabella che descrive le professioni secondo la logica INAIL ha prodotto due effetti: ha ridotto la quota di casi non determinati e ha popolato le nuove professioni introdotte.

Tabella 1

Infortuni sul lavoro denunciati all'INAIL - Anni evento 2007-2009.

	2007	2008	2009
Infortuni denunciati	912.979	875.648	790.361
di cui:			
qualifica professionale non determinata	123.033	161.242	112.190
% non determinati	13,5	18,4	14,2

Nella tabella 1 sono indicati i valori della base dati di partenza, presa in esame per il lavoro. Le percentuali indicate fanno riferimento al peso dei soli casi per i quali la qualifica professionale è mancante.

Occorre precisare che la mancanza dell'informazione sulla professione è plausibile per gli infortuni denunciati e definiti in franchigia (assenza di durata inferiore ai 3 giorni lavorativi), per i quali le informazioni sono minime e spesso desunte da un certificato medico.

Fatte queste premesse sulla base dati INAIL, nei paragrafi successivi si descriverà il lavoro di attribuzione del codice NUP ai codici INAIL delle professioni.

Il procedimento consta di due passi: assegnare in maniera deterministica un codice NUP ad ogni codice INAIL, quando è possibile; procedere attraverso un'imputazione di tipo probabilistico in tutti i casi dove ciò non è possibile e in tutti i casi nei quali l'informazione manca o è troppo generica.

Il lavoro è stato effettuato su una base dati triennale relativa al periodo 2007-2009 e considerando solo le professioni presenti in entrambe le classificazioni. Sono state eliminate, infatti, le categorie di soggetti assicurate all'INAIL e non presenti nella NUP, come per esempio gli studenti (categoria assicurata e assimilata ai lavoratori nello svolgimento di attività tecnico-scientifiche) e le professioni relative al gruppo delle Forze Armate presenti nella NUP e non assicurate all'INAIL.

5. La metodologia applicata

Al fine di rendere conformi i due archivi e agevolare così l'accesso degli utenti alla rete del Sistema informativo, i dati presenti nell'archivio Inail sono stati sottoposti a uno specifico trattamento per ottenere, da una lato la codifica delle professioni secondo la NUP Istat e, dall'altro per correggere le mancate risposte parziali.

L'obiettivo è stato raggiunto anche attraverso una fase che ha previsto un percorso formativo reciproco, caratterizzato da una parte interamente teorica seguita da una fase di "training on the job". La prima ha trattato le conoscenze teorico-applicative sulla logica e i contenuti della classificazione delle professioni e sui metodi di correzione delle mancate risposte parziali, con riferimento, in particolare, alle variabili qualitative; la seconda è servita a definire e ad implementare un appropriato modello di controllo e correzione per le risposte mancanti dell'archivio INAIL.

5.1 *L'imputazione deterministica*

La prima fase del lavoro è consistita nell'imputazione deterministica che si propone di assegnare a tutte le qualifiche INAIL una o più codifiche professionali della classificazione delle professioni.

Occorre precisare che poiché l'analisi ha previsto un triennio di dati infortunistici (2007-2009) sono state assegnate le qualifiche Istat sia alle mansioni del vecchio che a quelle del nuovo mansionario.

Fortunatamente nella maggior parte dei casi si è riusciti ad attribuire un codice univoco senza ambiguità, facendo corrispondere ad ogni mansione INAIL una unità professionale Istat. A fine lavoro il 91% delle mansioni INAIL aveva associato un codice delle professioni fino al massimo livello di dettaglio (quinto digit). Il restante 9% di qualifiche professionali ha richiesto un lavoro più lungo e articolato. Per poter meglio discriminare la professione, soprattutto in presenza di alcune qualifiche generiche che potevano rientrare in più ambiti lavorativi, è stata effettuata un'analisi approfondita che ha previsto l'esame anche di altre variabili aggiuntive come la lavorazione INAIL (voce di tariffa) e il settore di

attività economica (Ateco). Con queste ulteriori variabili a quasi tutte le professioni sono stati attribuiti dei codici, anche se non sempre si è riusciti ad arrivare al massimo livello di dettaglio, ma ci si è dovuti fermare a livelli superiori. Per poter meglio discriminare rispetto all'attività professionale sono state analizzate per questi casi specifici le distribuzioni degli infortuni sul lavoro per qualifiche professionali relativamente ad un quinquennio, in modo da disporre di un bacino di "donatori" più ampio.

Sono rimaste escluse dall'imputazione deterministica alcune professioni molto generiche quali "operatore" e "responsabile" - in tutto il 2% del totale delle professioni, ma con un'incidenza ben superiore in termini di denunce infortunistiche - per le quali non è stato possibile definire alcuna regola di imputazione deterministica e la loro codifica è stata rinviata alla fase successiva di trattamento delle mancate risposte.

Come si può capire, l'imputazione deterministica della NUP non si è basata su una semplice operazione di transcodifica, ma ha previsto una fase di analisi e di confronto all'interno del gruppo di lavoro Istat e INAIL finalizzata ad orientare la scelta dei raccordi più adeguati tra le figure professionali coinvolte. Il raccordo delle qualifiche INAIL ai codici della classificazione Istat delle professioni è risultato particolarmente delicato alla luce delle diverse logiche classificatorie sottese dalle due tassonomie

A titolo esemplificativo nella tabella che segue sono riportati due casi di qualifiche professionali INAIL, quella del levigatore e quella del tubista, per le quali l'imputazione deterministica ha portato ad attribuire un codice NUP differente a seconda del settore di attività economica attribuito all'infortunio.

Tabella 2

Sintesi dell'imputazione deterministica per levigatore e tubista.

Codice INAIL	Descrizione INAIL	Codice Ateco	Descrizione Ateco	Codice NUP	Descrizione NUP
558	Levigatore	F 45.4	Lavori di completamento edifici	6.1.3.2.2	Rifinitori di pavimenti
558	Levigatore	DN 36.1	Fabbricazione di mobili	6.5.2.2.3	Ebanisti
871	Tubista	DJ 27.2	Fabbricazione di tubi	7.1.2.4	Trafilatori ed estrusori metalli
871	Tubista	DM 35.1	Industria cantieristica	6.2.5.2.0	Carpentieri navali
871	Tubista	F 45.3	Installazione di servizi in fabbricati	6.1.3.6.1	Idraulici nelle costruzioni civili

Nel complesso sono state definite 955 regole deterministiche, che si possono sintetizzare come segue: il 90,7% di regole attribuisce il codice completo a 5 digit, il restante 9,3% attribuisce un codice parzialmente completo di 1 (1,5%), 2 (0,3%), 3 (6,3%), 4 digit (1,2%).

In termini di denunce di infortuni sul lavoro ciò significa che per il 2009 l'82,2% degli infortuni ha avuto l'attribuzione di un codice NUP e quasi il 70% ha avuto un codice a 5 digit.

Come si evince dalla tabella sottostante la quota di eventi denunciati da imputare probabilisticamente è del 34,1% nel 2007, del 29,4% nel 2008 e del 17,8% nel 2009.

Tabella 3

Distribuzione degli infortuni sul lavoro per professione Istat imputata in maniera deterministica - Anni evento 2007-2009.

	2007	%	2008	%	2009	%
Professione a 5 digit	450.732	49,4	496.120	56,7	543.678	68,8
Professione a 4 digit	8.310	0,9	9.060	1,0	9.720	1,2
Professione a 3 digit	83.024	9,1	69.297	7,9	64.188	8,1
Professione a 2 digit	-	0,0	515	0,1	1.097	0,1
Professione a 1 digit	59.915	6,6	43.336	4,9	30.750	3,9
Totale determinati	601.981	65,9	618.328	70,6	649.433	82,2
Non determinati	310.998	34,1	257.320	29,4	140.928	17,8
Totale	912.979	100,0	875.648	100,0	790.361	100,0

La fase di imputazione deterministica ha consentito di raccordare le due classificazioni delle professioni (Istat ed INAIL) e di generare un set di donatori, particolarmente robusto e accurato, indispensabile per la successiva fase di trattamento delle mancate risposte.

5.2 L'imputazione probabilistica

La seconda fase consiste nell'imputazione probabilistica che ha lo scopo di attribuire ai casi non determinati e alle qualifiche professionali troppo generiche un codice NUP mediante un procedimento di tipo probabilistico.

Di seguito si riporta l'elenco delle qualifiche professionali INAIL per le quali l'attribuzione del codice NUP è avvenuto probabilisticamente.

Tabella 4

Professioni imputate probabilisticamente.

Codice INAIL	Descrizione
102	ABBOZZATORE
131	AUSILIARIO LAV.MECC/AGR< 17 ANNI
132	AUSILIARIO LAV.MECC/AGR.>17ANNI
253	CALDAISTA
354	DIGROSSATORE
407	FIAMMELLISTA
499	IMPIGNONATORE
561	LIZZATORE
608	MISURATORE
652	OPERATORE
715	PARASUBORDINATO
828	SPORTIVI NON PROFESSIONISTI
950	ZIGRINATORE
1321	ADDETTO ATTIVITA' OPERATIVE
1325	VENDITORE
1326	RESPONSABILE NON DIRIGENTE

Lo stesso procedimento viene, inoltre, utilizzato per attribuire un codice al quinto digit in tutti i casi per i quali, sulla base del procedimento deterministico, si è riusciti ad attribuire il codice al massimo fino al quarto digit.

A tale scopo è stato adottato un metodo molto diffuso nell'ambito delle rilevazioni statistiche che consiste in una procedura probabilistica basata sul metodo di FELLEGI e HOLT¹ (1976).

Sono state sperimentate più strategie di imputazione e utilizzati due software prodotti e sviluppati entrambi dall'Istat:

- SCIA (Sistema di Controllo di Imputazione Automatica) - Concord (Controllo e Correzione dei Dati) in ambiente Sas
- SCIA - Concord in ambiente Java
scaricabili dal sito dell'Istat all'indirizzo:
<http://www.istat.it/strumenti/metodi/software/>

In questa prima fase sperimentale sono state seguite due diverse strade al fine di verificare quale procedimento fosse più idoneo:

- l'imputazione "step by step" di ogni singolo livello del codice NUP, parzialmente o totalmente mancante;
- l'imputazione di gruppi di livello del codice NUP, in particolare, in base a questa strategia, il codice NUP mancante o non determinato viene imputato tutto in una volta.

¹ FELLEGI, I.P., HOLT, D.: *A Systematic Approach to Automatic Edit and Imputation*, in *J. Am. Stat. Assoc.*, 1976, 71, 17-35.

Per l'imputazione probabilistica sono state considerate diverse combinazioni di variabili "chiave", vale a dire variabili correlate alla variabile oggetto di imputazione, utili per la scelta del record donatore.

Sulla base delle sperimentazioni effettuate si è osservato che i risultati migliori si ottenevano nel caso di imputazione effettuata per passi (step by step) dei codici NUP parzialmente o totalmente mancanti. Con questo procedimento, infatti, si riesce ad orientare meglio il codice da attribuire, evitando di assegnarlo in un unico passaggio (seconda strategia), in tutti quei casi che presentano una qualifica da imputare probabilisticamente (per le percentuali vedere tabella 3).

Come variabili chiave sono state adottate: il sesso, la tipologia di lavoratore distinta tra dipendente e autonomo, la ripartizione territoriale (considerando le 5 ripartizioni geografiche Nord-Est, Nord-Ovest, Centro, Sud, Isole) e l'anno.

Va precisato che è stata effettuata anche un'altra sperimentazione che ha coinvolto: il sesso, la tipologia di lavoratore distinta in autonomi, dipendenti, apprendisti, la ripartizione geografica a 4 (Nord-Est, Nord-Ovest, Centro e Mezzogiorno) e l'anno. Per disporre di un serbatoio più ampio e vario di record donatori si è considerato il triennio 2007-2009 come unico insieme piuttosto che i singoli anni.

Il software più adatto a questa prima sperimentazione è risultato Concord in ambiente Java.

Va precisato che per effetto dell'imputazione si sono generate nuove professioni, vale a dire professioni non presenti nell'archivio dati originario.

Per valutare il procedimento che meglio rispondeva, tra i vari sperimentati, sono state effettuate delle analisi statistiche sulle matrici di transizione, prima e dopo l'imputazione, volte a verificare l'indice di dissomiglianza e la variazione delle frequenze associate ad ogni NUP. Sono state analizzate quindi, le distribuzioni marginali prima e dopo l'imputazione, sia per gli infortuni denunciati che per i casi mortali. È stato scelto il procedimento che presentava una differenza in termini percentuali delle frequenze più accettabile e che non determinasse forti disallineamenti tra il prima e il dopo imputazione.

È risultato che alcune professioni sono state maggiormente alimentate dopo il processo di imputazione probabilistica; trattasi, in particolare, di:

- 4.1 - "Impiegati di ufficio";
- 6.1 - "Artigiani e operai specializzati dell'industria estrattiva e dell'edilizia";
- 6.4 - "Agricoltori e operai specializzati dell'agricoltura, delle foreste, della zootecnia, della pesca e della caccia".

6. Conclusioni

Il procedimento di imputazione probabilistica rappresenta un esempio di trattamento statistico di dati amministrativi che si avvale dell'uso di metodologie Istat per ottenere un insieme completo e coerente sulle professioni. Chiaramente que-

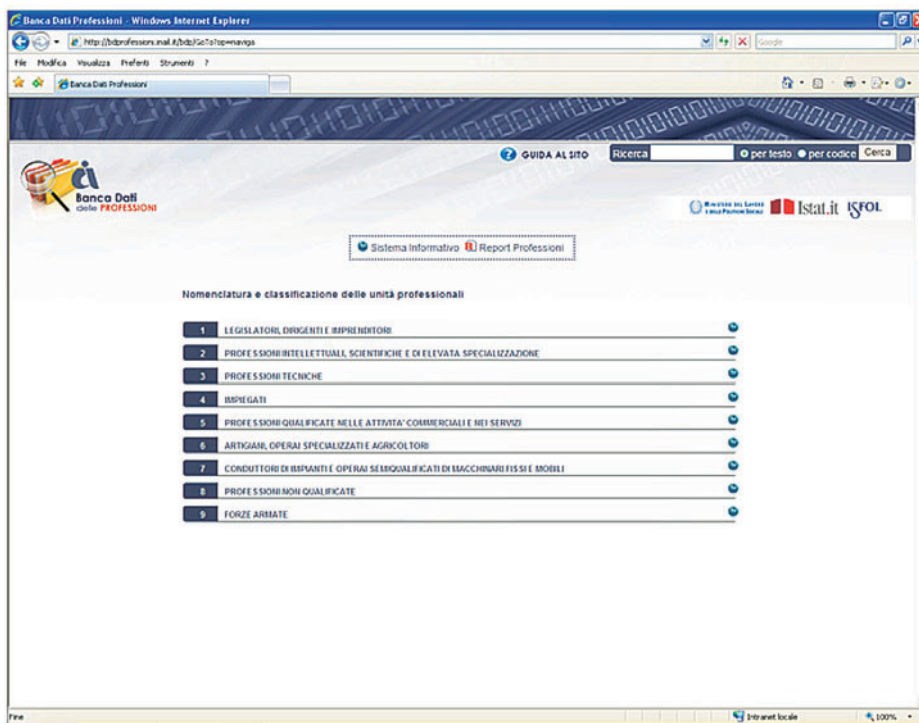
sto metodo può essere generalizzato per trattare anche altri tipi di dati che rispondono alla stessa esigenza.

Il file di dati, ottenuto come risultato del processo di imputazione, pur essendo un prodotto distinto dal corrispondente archivio INAIL, consente il dialogo tra INAIL e le altre Banche Dati del Sistema informativo.

L'obiettivo è stato pienamente raggiunto e si è concretizzato nella realizzazione di una Banca Dati delle Professioni, presente nella sezione delle Statistiche del portale dell'Istituto.

In questa Banca Dati, per ogni categoria professionale, sono effettuati i confronti con il complesso degli infortuni del grande gruppo di professioni di appartenenza e presentate le distribuzioni rispetto alle principali variabili di interesse (sesso, età, tipologia di infortunio, territorio, natura e sede della lesione). Le statistiche riportano un triennio di dati di infortuni sul lavoro denunciati e di casi mortali e sono aggiornate annualmente. Oltre alla serie storica triennale è presente un approfondimento per l'ultimo anno di riferimento. La navigazione è molto intuitiva, le statistiche sono visibili in corrispondenza del maggior livello di dettaglio dove sono presenti anche i link agli Enti che detengono informazioni per la sezione che si sta visualizzando.

Tabella 5: Home page della Banca Dati delle Professioni.



RIASSUNTO

Il lavoro descrive la collaborazione tra INAIL e Istat nell'ambito del progetto del Sistema Informativo sulle Professioni. L'obiettivo del Sistema è mostrare le diverse informazioni che provengono da varie fonti per ricostruire il contesto sociale ed economico in cui si sviluppa il mercato del lavoro italiano. Per facilitare gli utenti che accedono al Sistema Informativo sulle Professioni tutti i dati devono essere resi comparabili. Per questo motivo i dati dell'INAIL sono stati sottoposti ad un trattamento statistico che ha consentito il raccordo con la Nomenclatura delle Unità Professionali (NUP) dell'Istat. Il procedimento seguito si è composto di due fasi: nella prima, i dati sono stati sottoposti a regole deterministiche per attribuire alle qualifiche professionali INAIL il codice NUP Istat; nella seconda fase è stato adottato un procedimento di imputazione di tipo probabilistico delle mancate risposte da applicare in tutti i casi in cui la professione non è determinata o è troppo generica.

SUMMARY

The work describes the collaboration between Istat and INAIL for the project of the Occupational Information System. The aim of the System is to share different information from various sources for reconstructing social and economic context, in which occupations are developed. In order to facilitate users' access to the the Information System all data included in it need to be comparable. For this reason INAIL data have been subjected to statistical treatment to get occupation code according to the NUP and to correct item non-responses. This procedure is structured in two steps: in the first one a set of deterministic rules is applied to match the code from the NUP with the corresponding INAIL one when if condition is true; in the second step it needs to proceed through a probabilistic method of imputation when information is either missing or too generic.