

Atti del 9° Censimento generale
dell'industria e dei servizi e
Censimento delle istituzioni non profit

Il Censimento delle imprese

2



CENSIMENTO
DELL'INDUSTRIA
E DEI SERVIZI 2011



Atti del 9° Censimento generale
dell'industria e dei servizi e
Censimento delle istituzioni non profit

Il Censimento delle imprese

2



CENSIMENTO
DELL'INDUSTRIA
E DEI SERVIZI 2011



A cura della:

Direzione Centrale delle Rilevazioni Censuarie
e Registri Statistici (DCCR)

Atti del 9° Censimento generale dell'industria e dei servizi
e Censimento delle istituzioni non profit

2 - IL CENSIMENTO DELLE IMPRESE

© 2015

Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma

ISBN 978-88-458-1894-3 (elettronico)
ISBN 978-88-458-1893-6 (stampa)

Salvo diversa indicazione la riproduzione è libera,
a condizione che venga citata la fonte.

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



Indice

	Pag.
Presentazione dell'opera	7
Il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit	9
Introduzione	19
PARTE PRIMA	
IL CENSIMENTO VIRTUALE: IL SISTEMA DEI REGISTRI SULLE UNITÀ ECONOMICHE ASIA	
Capitolo 1 - Il censimento virtuale e il Sistema dei registri Asia	23
Capitolo 2 - Il registro delle imprese attive Asia	27
2.1 Il ruolo del registro nel sistema delle statistiche economiche	27
2.2 La copertura del registro e il suo contenuto in termini di unità e variabili	29
2.2.1 <i>La copertura del registro</i>	29
2.2.2 <i>Le unità del registro e le loro relazioni</i>	30
2.2.3 <i>Le variabili del registro</i>	35
2.2.4 <i>Il trattamento dei cambiamenti</i>	37
2.3 Il processo di aggiornamento delle unità nel registro Asia	40
2.4 La prima fase del processo di aggiornamento delle imprese: l'integrazione delle fonti amministrative e l'identificazione delle unità statistiche	43
2.4.1 <i>L'uso delle fonti amministrative per il registro</i>	43
2.4.2 <i>Il processo di integrazione delle fonti amministrative e l'identificazione delle unità</i>	47
2.5 La seconda fase del processo di aggiornamento delle imprese: la stima dei caratteri	51
2.5.1 <i>La stima dell'attività economica</i>	51
2.5.2 <i>La stima dello stato di attività e l'età dell'impresa</i>	54
2.5.3 <i>La stima dei principali eventi di trasformazione societaria</i>	58
2.5.4 <i>La stima degli altri caratteri</i>	61
2.6 La terza fase del processo di aggiornamento delle imprese: l'integrazione delle fonti statistiche	64
2.7 La quarta fase del processo di aggiornamento delle imprese: il piano di controllo e correzione	70
2.8 Il campo di osservazione Cis e le diverse modalità di utilizzo del registro	78
Capitolo 3 - L'occupazione di Asia: il nuovo database Asia-occupazione	83
3.1 Il framework concettuale e la revisione delle definizioni	83
3.1.1 <i>Le definizioni ILO-SNA e il framework concettuale di riferimento utilizzato</i>	83
3.1.2 <i>Le nuove definizioni dell'occupazione</i>	86
3.1.3 <i>Gli indipendenti: la declinazione della definizione</i>	88
3.1.4 <i>La classificazione degli outworkers: la scelta operativa effettuata</i>	90

	Pag.	
3.2	La nuova struttura informativa sui lavoratori e occupazione	91
3.3	La metodologia per la realizzazione di Asia-occupazione: gli occupati dipendenti	94
	3.3.1 <i>Le fonti amministrative utilizzate e la loro armonizzazione</i>	94
	3.3.2 <i>Il processo di integrazione</i>	98
	3.3.3 <i>La derivazione delle posizioni annue totali e retribuite</i>	99
	3.3.4 <i>La derivazione della variabile "Tipologie di beneficiari di sgravi contributivi"</i>	101
3.4	Gli occupati indipendenti	104
	3.4.1 <i>La procedura di identificazione dell'indipendente: la struttura informativa e le fonti amministrative</i>	105
	3.4.2 <i>La procedura di identificazione dell'indipendente: le regole deterministiche</i>	107
	3.4.3 <i>La stima probabilistica per un sottoinsieme di società di capitali senza indipendenti</i>	111
	3.4.4 <i>L'individuazione dell'indipendente per le società di capitali senza dipendenti</i>	114
	3.4.5 <i>Il calcolo delle posizioni lavorative dell'occupazione indipendente, in media annua e al dicembre</i>	116
3.5	L'occupazione esterna	117
	3.5.1 <i>I lavoratori parasubordinati: aspetti definitivi</i>	117
	3.5.2 <i>Le fonti amministrative per l'individuazione dei lavoratori parasubordinati</i>	118
	3.5.3 <i>La classificazione dei parasubordinati</i>	119
	3.5.4 <i>Il trattamento statistico</i>	121
	3.5.5 <i>I lavoratori temporanei (ex interinali)</i>	122
3.6	Le informazioni disponibili con la diffusione di Asia-occupazione	124
Capitolo 4 - Asia-unità locali		127
4.1	Il processo di realizzazione del registro delle unità locali	127
	4.1.1 <i>L'integrazione delle fonti</i>	127
	4.1.2 <i>L'individuazione delle unità e la stima degli addetti</i>	129
	4.1.3 <i>I controlli di qualità</i>	131
4.2	La stima del luogo di lavoro dei dipendenti presenti nel database di Asia-occupazione	133

PARTE SECONDA

LE RILEVAZIONI CAMPIONARIE – RILEVAZIONI SULLE UNITÀ LOCALI DELLE IMPRESE, MULTISCOPO SULLE IMPRESE, MULTISCOPO SULLE UNITÀ COMPLESSE

Capitolo 1 - La Rilevazione sulle unità locali delle imprese		141
1.1	Il quadro normativo e i tempi dell'indagine	141
1.2	Il campo di osservazione	142
1.3	Unità di rilevazione e unità di analisi	142
1.4	I criteri di selezione del campione	143
1.5	La comunicazione alle imprese	144
1.6	Il processo di lavorazione	145
1.7	I questionari respinti	145
1.8	Il sistema informatico	146
1.9	Il piano di controllo dei dati	146
1.10	Il contenuto del questionario	147
Capitolo 2 - La Rilevazione multiscopo sulle imprese		149
2.1	Gli obiettivi conoscitivi e la metodologia	149

	Pag.
2.2 Il questionario	150
2.2.1 <i>I contenuti informativi</i>	150
2.2.2 <i>Le sezioni del questionario</i>	151
2.2.3 <i>La struttura e le caratteristiche del questionario</i>	155
2.3 Il disegno di campionamento	157
2.3.1 <i>La popolazione di interesse e i domini di studio</i>	157
2.3.2 <i>L'allocazione del campione</i>	158
2.3.3 <i>Calcolo dei pesi finali</i>	161
2.4 Il campione delle imprese di dimensione inferiore a tre addetti	162
2.4.1 <i>Individuazione della popolazione di riferimento</i>	162
2.5 La fase di controllo e correzione dei dati	165
2.5.1 <i>Procedura di correzione per mancata risposta totale</i>	166
2.5.2 <i>Il trattamento delle mancate risposte parziali e delle incongruenze logiche</i>	168
2.6 In sintesi	170
Capitolo 3 - La Rilevazione multiscopo sulle unità complesse	173
3.1 Le finalità e gli obiettivi dell'indagine	173
3.2 Il questionario	174
3.3 La popolazione di interesse e i domini di studio	179
3.4 L'organizzazione della rilevazione	181
3.5 In sintesi	182
Riferimenti bibliografici	185
Elenco allegati disponibili in formato digitale	187

Presentazione dell'opera

Come avvenuto in occasione di tutti i precedenti censimenti, l'Istituto nazionale di statistica ha redatto gli "Atti" del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit. La pubblicazione descrive le caratteristiche tecniche, organizzative e metodologiche delle attività svolte.

Gli Atti sono editi esclusivamente in forma telematica e si articolano in sei fascicoli corredati dalla documentazione tecnica e dai documenti ufficiali del censimento. Considerata la rilevanza e la complessità degli argomenti trattati, l'opera è stata realizzata in maniera unitaria, essendo evidenti le interconnessioni fra le diverse fasi censuarie; tuttavia ciascun fascicolo ha propria autonomia, poiché tratta in maniera esaustiva un tema specifico.

I sei fascicoli che costituiscono gli Atti del Censimento generale dell'industria e dei servizi sono:

1. *Norme e organizzazione;*
2. *Il censimento delle imprese;*
3. *Il censimento delle istituzioni non profit;*
4. *Il censimento delle istituzioni pubbliche;*
5. *Le sezioni di censimento;*
6. *Il sistema informatico integrato.*

Il primo fascicolo *Norme e organizzazione* è strutturato in quattro parti. La prima parte, *Il quadro normativo*, illustra le norme in base alle quali è stato progettato e realizzato il censimento. Sono presentati e descritti nei loro aspetti principali la legge di indizione e finanziamento, il Piano Generale di Censimento, nonché le circolari operative emanate dall'Istat per definire e dare attuazione a tutte le fasi censuarie. La seconda parte riguarda l'impianto organizzativo della *Rilevazione sulle imprese* e della *Rilevazione sulle istituzioni non profit*, mentre la terza è relativa all'impianto organizzativo della *Rilevazione sulle istituzioni pubbliche*. La distinzione è motivata dalle profonde differenze metodologiche, organizzative e tecnologiche che hanno contraddistinto il processo di rilevazione delle tre tipologie di unità economiche. Sia nella seconda che nella terza parte si descrivono le linee guida adottate nell'organizzazione delle rilevazioni, la composizione e le caratteristiche della rete di rilevazione, le modalità di formazione degli operatori censuari, gli strumenti di supporto alle rilevazioni e di gestione delle criticità nel corso del processo produttivo. La quarta parte, *Comunicazione e diffusione*, presenta la campagna di comunicazione integrata realizzata dall'Istat per promuovere la condivisione degli obiettivi da parte dei rispondenti e dei potenziali utilizzatori dei dati. In particolare, illustra le azioni e gli strumenti specifici ideati e realizzati per ciascuna rilevazione censuaria, pur nell'unitarietà della strategia comunicativa. In essa è anche descritto il piano di diffusione, incentrato sul sistema di *data warehousing* dell'Istat I.stat e articolato in diversi prodotti.

Il secondo, terzo e quarto fascicolo trattano in dettaglio gli aspetti peculiari di ciascuna delle tre rilevazioni: sulle imprese, sulle istituzioni non profit e sulle istituzioni pubbliche. In ciascun fascicolo sono illustrate le caratteristiche generali e tecniche della rilevazione, percorrendo le principali tappe che hanno portato alla definizione delle liste precensuarie delle unità di rilevazione, dei contenuti informativi dei questionari, del piano di controllo e correzione dei dati, della diffusione dei risultati. Nel secondo fascicolo è inoltre descritto il processo metodologico e produttivo seguito per la realizzazione del sistema di registri sulle unità economiche Asia.

Il quinto fascicolo, *Le sezioni di censimento*, descrive la procedura di attribuzione delle sezioni di censimento alle unità locali, evidenziandone caratteristiche e peculiarità e presentando i risultati finali.

Il sesto fascicolo, *Il sistema informatico integrato*, descrive l'architettura e le soluzioni informatiche adottate per un censimento articolato e complesso che ha utilizzato differenti modalità di compilazione dei questionari e anche differenti modalità di restituzione dei questionari compilati, che ha previsto la registrazione dei questionari acquisiti in formato cartaceo presso gli uffici di rilevazione nonché un *paperless census* per la rilevazione sulle istituzioni pubbliche. Tutti i processi adottati, connessi e dialoganti tra loro, hanno garantito coerenza logica, velocità e trasparenza alle operazioni realizzate. Nel fascicolo sono descritti il Sistema di gestione della rilevazione (Sgr), l'impianto informatico di acquisizione dati e gli strumenti informatici di diffusione dei dati.

Il 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit

A cura di Andrea Mancini

Il tradizionale campo di osservazione del censimento economico italiano si presenta complesso e molto articolato. In esso sono comprese imprese che agiscono sul mercato, amministrazioni pubbliche e organizzazioni private non profit. Si tratta in ogni caso di unità economiche produttrici di beni e servizi ma operanti con forme giuridiche molto diversificate, scopi plurimi, strutture produttive assai differenti, dimensioni molto variabili. L'informazione statistica relativa a questo universo di riferimento è altrettanto articolata, ma in occasione delle decennali tornate censuarie il quadro strutturale delle unità economiche residenti nel nostro Paese è sempre divenuto oggetto di osservazione statistica generalizzata, simultanea e a grande dettaglio di livello territoriale. Tuttavia la progressiva realizzazione del sistema di registri statistici (Asia-imprese, Asia-gruppi, Asia-unità locali), tutti basati su informazioni desunte da una pluralità di archivi amministrativi, ha consentito di produrre annualmente parti sostanziali del quadro statistico strutturale prima demandato al censimento decennale. Di fatto il sistema dei registri Asia permette di avere annualmente il conteggio delle unità economiche attive e delle loro unità locali, insieme con alcune variabili fondamentali, quali l'attività economica prevalente, il numero di addetti, il fatturato. Ciò almeno per le imprese operanti sul mercato.

In questo contesto l'obiettivo più rilevante e peculiare del 9° Censimento generale dell'industria e dei servizi è stato di completare, con riferimento al 31 dicembre del 2011, il sistema di registri statistici basato su dati di fonte amministrativa e di fornire informazioni strutturali di approfondimento mirato sui singoli e differenti settori istituzionali compresi nell'universo delle unità economiche. A quest'ultimo fine il 9° Censimento ha fatto uso di tre liste precensuarie delle unità economiche (imprese, istituzioni non profit, amministrazioni pubbliche) per raccogliere informazioni mediante varie rilevazioni, sia campionarie che totali, ciascuna specializzata nella raccolta di informazioni e dati idonei a completare il quadro statistico di ciascun settore, tenendo conto delle loro peculiarità giuridico-istituzionali e operative.

Dunque il 9° Censimento ha presentato elementi di omogeneità e sinergia, ma anche differenziazioni tematiche e metodologiche rilevanti, che hanno reso la sua realizzazione particolarmente complessa. Di fatto si è trattato di un censimento suddiviso in differenti rilevazioni tematiche, rese omogenee da una unica *governance* che ha permesso la gestione unitaria sul piano organizzativo, metodologico, informatico e tecnico. Esso è stato caratterizzato da innovazioni, tra le quali rilevanza generale hanno assunto:

- ▶ l'ampio uso di dati amministrativi per la predisposizione delle liste precensuarie, per il controllo e la correzione delle informazioni rilevate (anche contestuali alla raccolta dei dati), per la realizzazione di un censimento virtuale sui dati strutturali delle imprese grazie alle informazioni residenti nel sistema di registri statistici Asia. Grazie all'utilizzo di dati amministrativi, gli obiettivi raggiunti sono stati la maggiore tempestività nel rilascio dei risultati, il contenimento del carico statistico sui rispondenti, il miglioramento della qualità dei dati prodotti, la riduzione dei costi. Gli esiti censuari hanno consentito anche di migliorare i processi metodologici e statistici che portano alla realizzazione dei registri

statistici di unità economiche. Questi costituiranno il fulcro sul quale ruoterà l'impianto organizzativo e metodologico delle rilevazioni post censuarie sulle unità economiche e dei censimenti economici continui;

- ▶ le novità sul piano dei contenuti informativi: tematiche inedite sono state approfondite, grazie alla realizzazione di specifiche rilevazioni con differenti questionari di settore, di tipo *long form*, che hanno permesso di trattare tematiche di interesse di ciascuno dei tre comparti;
- ▶ la realizzazione di una complessa e omogenea architettura informatica in grado di gestire sia il sistema di monitoraggio di tutte le operazioni censuarie, sia il sistema di acquisizione online dei questionari. Il nucleo portante dell'architettura informatica del censimento è stato il Sistema di gestione della rilevazione (Sgr), che ha consentito di seguire l'andamento di tutte le fasi del processo di produzione per singolo organo di censimento e operatore censuario, grazie alla realizzazione di oltre cinquanta funzioni componenti il sistema;
- ▶ l'uso massivo degli indirizzi di posta elettronica certificata (Pec), non solo nel rispetto di un obbligo dettato dalla normativa vigente ma anche come indispensabile strumento di riduzione dei costi di una indagine statistica. Di questo strumento si è fatto un uso il più ampio possibile nelle varie fasi del processo produttivo, sebbene il quadro complessivo presenti una situazione a "macchia di leopardo". Infatti, tale strumento è stato usato in via esclusiva per l'universo delle imprese e delle istituzioni pubbliche, mentre è stato usato in modo selettivo per l'universo delle istituzioni non profit, che non hanno l'obbligo normativo di dotarsi di indirizzi Pec;
- ▶ la valorizzazione del sistema di data warehousing *corporate* dell'Istat I.Stat (<http://dati-censimentoindustriaeservizi.istat.it>), grazie al quale l'utente può accedere liberamente e gratuitamente al rilevante patrimonio informativo censuario reso disponibile per la navigazione dei macrodati, per comporre tabelle ed effettuare confronti storici. Il censimento, con la sua volumetria, complessità e dettaglio di dati, ha usufruito della tecnologia adottata dall'Istat per il data warehouse I.Stat, seppur in una versione differente e resa più performante;
- ▶ una strategia di comunicazione e informazione che ha integrato le azioni della pubblicità classica con le attività tipiche delle relazioni pubbliche, con azioni ad hoc realizzate per target specifici e con l'uso dei nuovi media, al fine di spiegare l'importanza del censimento, informare sulle modalità di svolgimento delle operazioni, sollecitare la partecipazione, rassicurare sulla tutela della privacy.

Vista l'impostazione generale del 9° Censimento generale, nel seguito vengono illustrati brevemente gli aspetti peculiari delle rilevazioni per singoli comparti.

Le imprese

Consistente l'innovazione che riguarda le imprese per le quali l'Istat ha voluto superare la rilevazione censuaria tradizionale costituita dall'enumerazione esaustiva delle unità giuridico-economiche svolta sul campo, realizzando un "censimento virtuale" grazie all'utilizzo integrato di numerose fonti amministrative. Si è inteso realizzare così un progetto con approccio metodologico misto, che ha portato alla pubblicazione di dati desunti da archivi statistici e da rilevazioni campionarie tematiche assistite da archivi. I dati strutturali di base (natura giuridica, attività economica, occupazione, unità locali, localizzazione, fatturato) sono stati desunti dal sistema di registri Asia. Ciò ha consentito di fornire un quadro informativo statistico sulla struttura del sistema economico relativo alle imprese e alle unità locali fino al livello comunale e di sezione di

censimento e di garantire la confrontabilità dei censimenti economici, in particolare con quello del 2001. Oltre a questi dati strutturali sono state fornite ulteriori informazioni per l'analisi di specifici domini (in termini di settori di attività economica, tematiche di rilievo, eccetera), utilizzando informazioni amministrative finora parzialmente utilizzate e rilevazioni campionarie assistite da archivio. Il patrimonio informativo delle fonti amministrative è valorizzato in chiave statistica, restituendo all'utenza molte più informazioni rispetto a quelle prodotte tradizionalmente dal censimento, senza bisogno di farne oggetto di un questionario da compilare a cura di tutte le imprese italiane.

In particolare la valorizzazione del patrimonio informativo residente negli archivi amministrativi e le innovazioni metodologiche hanno consentito di realizzare un nuovo prodotto, denominato Asia-occupazione, che integra informazioni micro su impresa e occupati, realizzando un sistema informativo di tipo LEED (*Linked Employer Employee Database*). La disponibilità di nuove fonti amministrative con informazioni sul legame tra lavoratore e impresa e lo sviluppo di un quadro concettuale di definizioni e regole per il trattamento di tali dati a fini statistici hanno consentito lo sviluppo di un sistema di microdati integrati, dove è possibile identificare l'unità economica e l'unità lavoratore e dove ciascun individuo viene classificato in base alla tipologia occupazionale all'interno dell'impresa con cui, alla data di riferimento, ha un rapporto di lavoro. Anche queste nuove informazioni saranno d'ora in avanti prodotte e diffuse annualmente dall'Istituto grazie all'aggiornamento continuo dei dati amministrativi.

Per completare il quadro informativo sulle imprese sono state condotte tre rilevazioni campionarie: Indagine sulle unità locali delle grandi imprese (Iulgi), Rilevazione multiscopo sulle imprese, Rilevazione sulle unità complesse. Questa struttura di rilevazioni ha consentito di cogliere importanti elementi conoscitivi del sistema produttivo italiano, solo parzialmente rilevabili sulla base del tradizionale impianto delle indagini strutturali sulle imprese.

L'Indagine Iulgi è svolta dall'Istat con frequenza annuale per aggiornare l'Archivio statistico delle imprese attive (Asia) e delle unità locali ad esse afferenti (Asia-unità locali). Essa è nata nel 2004 per sopperire alle carenze di dati amministrativi sulle unità locali d'impresa, in particolare per quanto riguarda la distribuzione territoriale degli addetti. Con riferimento al 2011 la rilevazione annuale Iulgi è stata ampliata per migliorare il controllo di qualità del registro statistico Asia-unità locali, in termini sia di errore di copertura sia di errore di misura. La rilevazione è stata condotta su un campione probabilistico di circa 40 mila imprese (circa l'1 per cento delle imprese italiane, pari a tre volte la numerosità di Iulgi annuale), estratto da Asia-imprese, con probabilità di inclusione pari a 1 per le imprese con almeno 100 addetti.

Le altre due rilevazioni sono state progettate ed eseguite per la prima volta in occasione del 9° Censimento, dando la possibilità di cogliere la trama dei processi decisionali da cui dipendono le strategie aziendali, collocando le scelte su produzione, innovazione, capitale umano e sulle altre variabili competitive nel complesso delle relazioni di collaborazione o competizione che l'impresa intrattiene con il contesto che la circonda. Infatti, i recenti rapporti redatti da istituzioni internazionali, associazioni di categoria, network di ricerca rilevano come, dietro una apparente "staticità strutturale" delle imprese, in termini di dimensione media e settori prevalenti, il sistema industriale italiano nasconde mutamenti di grande rilevanza, che vanno verso una maggiore complessità e articolazione dei modelli di business. Le soluzioni e i percorsi evolutivi intrapresi dalle imprese presentano un'attenzione particolare alla flessibilità produttiva, all'innovazione, al

capitale umano e al controllo della tecnologia, che a sua volta richiede una conoscenza approfondita dell'intera filiera produttiva. Tali caratteristiche, peraltro, si compongono diversamente a formare nuovi modelli di attività e allo stesso tempo accomunano spesso tutte le classi dimensionali d'impresa. Mutamenti di questo tipo sono ampiamente diffusi nel settore produttivo italiano, ma l'informazione statistica ufficiale manca ancora di completezza rispetto a questi aspetti.

Più in particolare, mediante la Rilevazione multiscopo sulle imprese (rivolta a circa 260 mila imprese, pari al 6 per cento delle imprese italiane) si è inteso cogliere alcuni dei suddetti fenomeni, estendendo la disponibilità di dati relativi alle tecnologie ICT e all'Innovazione alle imprese con meno di 10 addetti, oltre a cogliere informazioni, alcune delle quali per la prima volta, come le strategie di *governance*, la gestione delle risorse umane, la competitività e la posizione sul mercato nazionale e internazionale, la capacità di innovazione, le scelte finanziarie, l'internazionalizzazione e le ragioni principali delle scelte di delocalizzazione¹. Questa rilevazione è stata realizzata sul campo dagli Uffici di censimento costituiti appositamente a livello provinciale o regionale presso le Camere di commercio competenti per territorio e con il supporto a livello nazionale dell'Ufficio di statistica di Unioncamere.

Infine, con la Rilevazione sulle unità complesse, il 9° Censimento ha effettuato un investimento mirato a cogliere le peculiari caratteristiche di queste unità economiche anche nella prospettiva di razionalizzare i loro rapporti con l'Istat nella conduzione delle numerose rilevazioni correnti. Queste particolari unità sono spesso costituite nella forma di gruppi di impresa di livello nazionale o multinazionale ed hanno una complessa organizzazione interna delle attività e funzioni di supporto aziendale; tendono però a non essere adeguatamente rappresentate dalla statistica ufficiale sulla base dei tradizionali criteri e strumenti di classificazione e rilevazione. È stata pertanto realizzata una approfondita attività di progettazione che ha coinvolto anche accademici ed esperti nel settore dell'organizzazione aziendale. L'impianto di questa indagine presenta numerosi aspetti innovativi nella selezione del campione censuario, nei contenuti informativi del questionario, nelle modalità di conduzione della rilevazione. I fenomeni sono stati rilevati rispetto all'intero perimetro delle singole unità economiche complesse e coprono tutti i principali aspetti dell'organizzazione e della performance economica, in modo coerente con l'effettiva articolazione delle attività economiche. La raccolta di informazioni ha riguardato, tra gli altri aspetti, l'articolazione dell'unità economica complessa per linee di attività economica e funzioni aziendali (con riferimento sia alle modalità di internalizzazione e/o esternalizzazione in Italia e/o all'estero), le competenze e le capacità manageriali, la formazione dei processi decisionali, il capitale umano, i rapporti e le scelte strategiche adottate nei confronti dei fornitori e dei clienti, i processi di internazionalizzazione, il ruolo della conoscenza esteso a tutti i fattori intangibili di competitività. La rilevazione ha interessato circa 3.500 vertici di gruppi di impresa di tipo nazionale ed internazionale, incluse anche imprese indipendenti con dimensioni economiche rilevanti, che rappresentano una quota rilevante di imprese ed occupati residenti in Italia e all'estero. La raccolta dei dati è stata condotta sul campo da rilevatori con una elevata esperienza professionale nei rapporti con il top management aziendale, selezionati mediante una gara internazionale.

¹ Questi argomenti, unitamente alle nuove informazioni prodotte con il sistema dei registri Asia, sono stati analizzati nel convegno "Check-up delle imprese italiane", tenutosi il 28 novembre 2013 a Milano. Sono stati presentati diversi lavori, articolati in due focus tematici su "Struttura, organizzazione, risorse umane" e "Performance e strategie", oltre a tre sessioni di lavoro parallele, rivolte a studiosi e ricercatori, dal titolo: "Strategie di competizione e cooperazione tra le imprese", "Globalizzazione e mercati", "Imprese, occupazione e territorio". Gli interventi sono disponibili all'indirizzo del sito Istat <http://www.istat.it/it/archivio/103067>.

Le istituzioni non profit

Il Censimento delle istituzioni non profit costituisce un'occasione di particolare rilevanza nel panorama della produzione statistica nazionale e internazionale. Con l'eccezione di alcune indagini annuali condotte dall'Istat tra la fine degli anni '90 e la metà degli anni 2000, riferite ad alcuni specifici comparti del non profit (organizzazioni di volontariato, cooperative sociali, fondazioni), non ci sono nel Paese informazioni di quadro generale come quelle prodotte dal censimento². In questo senso esso permette di colmare un vuoto informativo su quello che è ormai divenuto il "modello non profit" di produzione e di fornitura dei servizi, in grado di creare opportunità di lavoro anche in un periodo di grande difficoltà generale. Oggi molte attività rilevanti legate ai bisogni sociali della collettività sono svolte da soggetti istituzionali che non si pongono fini di lucro, in grado di offrire lavoro retribuito e in larga parte di operare grazie al lavoro volontario dei cittadini. Ancora poco conosciuto, il mondo del non profit in Italia è caratterizzato da diversità sia dei soggetti che lo compongono sia delle loro modalità di azione. Fortemente dicotomico, il settore è caratterizzato dalla presenza contestuale di poche istituzioni di grandi dimensioni con strutture organizzative complesse e - sia in termini di occupazione sia economici - e di tantissime organizzazioni di piccole dimensioni, spesso attive solo grazie all'esclusivo apporto volontario dei cittadini.

Anche a livello internazionale, la rilevazione censuaria colloca l'Italia in posizione di avanguardia nel panorama delle informazioni statistiche ufficiali e affidabili sul settore, in linea con le definizioni e le indicazioni fornite dal *System of National Accounts*, dal Sistema europeo dei conti economici e dal Manuale sulle istituzioni non profit³.

Il censimento è stato condotto a partire dalla lista precensuaria delle istituzioni non profit, la cui realizzazione si è rivelata particolarmente complessa. Il settore è, infatti, caratterizzato da una forte eterogeneità nella qualità delle fonti amministrative utili, alcune disponibili a livello nazionale ma con copertura parziale, altre gestite da enti territoriali sub nazionali con maggiori problemi di qualità. Sono state 17 le fonti amministrative utilizzate per realizzare la lista precensuaria, alcune di grandi dimensioni, altre costituite da poche specifiche unità.

La tecnica di rilevazione si è basata sull'invio postale dei questionari long form e sulla loro compilazione su supporto cartaceo o informatico, con intervento mirato dei rilevatori per la compilazione e il recupero dei questionari non ancora pervenuti. Come da tradizione, la rete di rilevazione è stata costituita, in cooperazione con l'Ufficio statistico di Unioncamere, dagli uffici di censimento di livello provinciale o regionale, operanti presso le Camere di commercio.

L'importanza del censimento è ben evidenziata anche dagli obiettivi colti: rispondere alle esigenze informative di *policy maker*, studiosi, esperti e operatori del settore riguardo a dimensioni economiche e sociali, caratteristiche organizzative e peculiarità delle istituzioni non profit nel nostro Paese; definire la consistenza del settore non profit in Italia e pervenire alla sua rappresentazione statistica in termini di grandezze confrontabili con quelle delle precedenti rilevazioni censuarie del 1999 e del 2001; rilevare informazioni utilizzabili per la predisposizione dei conti nazionali e in particolare per la costruzione del conto satellite sulle istituzioni non profit; costruire un sistema di

² Con riferimento al 1999 l'Istat ha effettuato la prima rilevazione censuaria delle istituzioni non profit (Cfr. "Istituzioni non profit in Italia", Istat, Collana Informazioni, n. 50, 2001) seguita dal 8° Censimento generale dell'industria e servizi nel cui campo di osservazione esse erano comprese (Cfr. *data warehouse* all'indirizzo del sito Istat <http://dwcis.istat.it/cis/index.htm>).

³ United Nations, Department of Economic and Social Affairs – Statistics Division, *Handbook on Non-profit Institutions in the System of National Accounts*, Studies in Methods, Series F., No. 91, New York 2003.

statistiche comparabili a livello internazionale; verificare la copertura e qualità della lista precensuaria in vista della realizzazione del registro statistico delle istituzioni non profit come parte integrante ed annualmente aggiornata del sistema dei registri sulle unità economiche.

Il patrimonio informativo raccolto con il censimento consente di analizzare la natura non profit delle unità istituzionali censite e di classificarle rispetto alle principali variabili di natura euristica, quali la tipologia di attività economica svolta (*market/non market*), l'orientamento (mutualistico/di pubblica utilità), la fonte di finanziamento prevalente (pubblico/privato); di rilevare le attività svolte (prevalenti e secondarie), i servizi erogati e i destinatari di tali servizi; di misurare l'entità delle risorse umane impiegate (sia retribuite sia volontarie) e le loro caratteristiche principali (categorie e profili professionali); di conoscere il numero e le caratteristiche dei soci; di rappresentare le dimensioni economico-finanziarie e le varie componenti di bilancio in termini di entrate e uscite.

La ricchezza informativa offerta si è rivelata pienamente in diverse occasioni seminariali e convegnistiche. La principale, organizzata dall'Istat, è stata il convegno nazionale "Il non profit in Italia. Quali sfide e quali opportunità per il Paese" svoltosi a Roma il 16 aprile 2014⁴. A livello internazionale, in occasione della Undicesima Conferenza Internazionale della *International Society for Third Sector Research* (ISTR), tenutasi a Muenster il 22-25 luglio 2014, è stato organizzato un Panel dal titolo "A statistical framework for the analysis of the Third Sector: the findings of the 2011 Italian Census", durante il quale sono stati presentati dei lavori realizzati da studiosi del settore e ricercatori Istat su diversi aspetti del non profit italiano e sull'esperienza statistica italiana nel settore.

Le istituzioni pubbliche

Nell'ambito del 9° Censimento, l'Istat ha condotto una rilevazione *ad hoc* sulle istituzioni pubbliche, con un questionario da compilare esclusivamente online dai referenti di ciascuna amministrazione, senza l'intervento di rilevatori e con l'uso di strumenti di *e-learning* disponibili nel portale del censimento quale supporto alla formazione dei referenti di amministrazione.

Le amministrazioni coinvolte nella rilevazione sono state inserite dall'Istat in una lista precensuaria predisposta sulla base dell'integrazione di 10 archivi amministrativi e di altre fonti statistiche. Come in tutti i precedenti censimenti, anche in occasione di questa edizione non sono rientrate nel campo di osservazione le informazioni relative alle Forze armate, alla Guardia di finanza, ai Corpi di polizia e di sicurezza.

L'impianto organizzativo è stato completamente nuovo. Infatti, la rilevazione è stata articolata in due fasi distinte: la prima fase, svolta tra il 26 luglio e il 20 settembre 2012, aveva lo scopo di aggiornare le informazioni anagrafiche dell'unità istituzionale presente nella lista precensuaria, individuare il nominativo del dirigente o funzionario incaricato del coordinamento della rilevazione per l'unità istituzionale, acquisire l'elenco delle unità locali di ciascuna unità istituzionale e i nominativi dei referenti di ciascuna unità locale. La seconda fase, iniziata il primo ottobre 2012, ha riguardato la compilazione dei questionari di unità istituzionale e di unità locale.

⁴ Cfr. gli interventi diffusi sul sito dell'Istat <http://www.istat.it/it/archivio/116695>.

I contenuti informativi del censimento sono stati fortemente innovativi. Numerose le tematiche specifiche che hanno consentito di fornire utili informazioni per una migliore conoscenza dell'organizzazione e funzionamento delle amministrazioni pubbliche italiane. Si evidenziano le informazioni sui percorsi formativi nella PA (quali istituzioni hanno formato il proprio personale, quali tematiche, quante ore, quante risorse umane coinvolte), sul grado di informatizzazione (dotazione e uso di ICT sia nelle unità istituzionali sia nelle singole unità locali), sull'amministrazione sostenibile intesa come tutela e sicurezza sul luogo di lavoro, sostenibilità ambientale, trasparenza e qualità dell'azione amministrativa (*accountability*).

Inoltre, il censimento ha consentito di tracciare la mappatura territoriale delle unità locali della pubblica amministrazione (cioè dei luoghi di lavoro e delle risorse umane ivi impegnate) e quella dei servizi pubblici offerti presso ciascuna unità locale (quali servizi, da chi, dove, quante risorse umane impegnate e il bacino d'utenza raggiunto). Queste informazioni sono rilevabili soltanto attraverso una operazione censuaria, che consente di effettuare rielaborazioni uniche in campo economico, grazie alla profondità informativa che si spinge fino ai dati a livello di categoria di attività economica e comune di ubicazione delle unità locali, cioè, appunto, dei luoghi di lavoro. Le analisi condotte utilizzando la variabile "unità locale" e "addetti alle unità locali" sono le più indicate per le analisi territoriali e settoriali perché l'informazione è quella di pertinenza, cioè quella che "insiste" direttamente sull'economia e sulle caratteristiche infrastrutturali e sociali del territorio di residenza dei cittadini e di ubicazione delle imprese e delle istituzioni.

In considerazione della crescente richiesta di informazione statistica sull'organizzazione e sul funzionamento delle amministrazioni pubbliche è necessario che l'Istituto si doti di strumenti innovati per fornire queste informazioni non più a carattere decennale, come è tradizione dei censimenti italiani, ma con scadenze più ravvicinate nel tempo, al fine sia di aggiornare periodicamente la mappa dei luoghi di lavoro e dei servizi offerti dalle istituzioni pubbliche sia di fornire informazioni statistiche di interesse per i decisori pubblici e gli esperti di settore. Anche da questa esigenza deriva la nuova strategia dell'Istituto volta alla realizzazione di censimenti economici continui, come evidenziato in sintesi nel seguito.

Il sistema informatico censuario

Centro nevralgico dell'organizzazione censuaria è stato il Sistema di gestione della rilevazione (Sgr), utilizzato autonomamente da ciascun organo di censimento, rilevatore o coordinatore, che ha avuto accesso alle funzioni e alle tabelle di pertinenza attraverso proprie chiavi di accesso (user id e password).

In generale, i sistemi realizzati per la gestione delle rilevazioni del 9° Censimento sono stati sviluppati utilizzando prevalentemente tecnologie software *open source*, librerie di sviluppo e schemi di progetto standard. Essi si sono basati su un'infrastruttura completamente centralizzata che ne ha consentito l'accesso e l'utilizzo a tutte le persone dotate di un personal computer o tablet, di una connessione a internet e di un browser per la navigazione, senza dover installare software o configurare *ad hoc* il proprio dispositivo elettronico. Di fatto si è costruito un sistema integrato, parametrizzato e standardizzato in grado di dare supporto a tutte le attività legate alla fase di acquisizione dei dati e in alcuni casi anche di loro controllo, minimizzando le attività di sviluppo di codice specifico e massimizzando il riuso dei software già realizzati per il Censimento dell'agricoltura e per il Censimento della popolazione, in particolare per i sistemi di gestione della rilevazione e i sistemi di generazione e presentazione dei questionari elettronici.

Per supportare tutte le attività di raccolta dei questionari compilati e di monitoraggio della rilevazione sono stati realizzati due sistemi web direttamente accessibili attraverso il portale principale del censimento: il primo sistema è stato dedicato alla Rilevazione sulle imprese e al Censimento delle istituzioni non profit; il secondo sistema è stato approntato per il Censimento delle istituzioni pubbliche. I sistemi hanno permesso di gestire una rete di rilevazione articolata composta di 18 Uffici Regionali di Censimento (Istat) e di 103 Uffici provinciali di censimento, dei quali 99 costituiti presso le Camere di commercio, 2 presso le Unioni regionali delle Camere di commercio e 2 costituiti presso i Servizi di statistica delle Province autonome di Trento e Bolzano. Gli Uffici Regionali di Censimento hanno impiegato 166 addetti, mentre gli Uffici provinciali di censimento hanno creato una rete composta di quasi tremila addetti al censimento, di cui 2.257 rilevatori. Il sistema di acquisizione telematico ha portato alla registrazione online via internet dei questionari di tutte le unità di rilevazione. In particolare, ha scelto di compilare il questionario online il 78,8 per cento delle imprese e il 58,9 per cento delle istituzioni non profit, mentre il resto dei questionari è stato compilato online direttamente dagli Uffici di censimento, che hanno registrato i questionari cartacei compilati dalle imprese e dalle istituzioni non profit tramite l'utilizzo dello stesso sistema di acquisizione predisposto per le unità di rilevazione. Il Censimento delle istituzioni pubbliche si è basato invece su un sistema di compilazione e acquisizione dei dati esclusivamente online (*paperless census*).

Conclusioni e prospettive

Le numerose innovazioni di prodotto e di processo descritte hanno permesso di accrescere la tempestività della diffusione dei dati definitivi anticipando rispetto al passato la restituzione delle informazioni. L'11 luglio del 2013 sono stati diffusi i principali dati a livello di unità istituzionali, a 4 mesi dalla data di chiusura delle operazioni censuarie, a 11 mesi dall'inizio delle rilevazioni sul campo e a 18 mesi dalla data di riferimento del censimento. Il 30 ottobre 2013 sono stati diffusi analoghi dati a livello di unità locali e i dati tematici sulle imprese desunti dalla Rilevazione multiscopo. Il 21 dicembre 2013 i principali dati tematici sulle istituzioni pubbliche e all'inizio del 2014 i dati relativi all'intero questionario delle istituzioni non profit.

Numerose le occasioni, nazionali e internazionali, nelle quali rappresentanti del mondo della politica, della ricerca, delle imprese e delle istituzioni, pubbliche e non profit, oltre a ricercatori Istat, hanno avuto modo di confrontarsi, discutere e analizzare la struttura produttiva del Paese così come emersa dai dati censuari. Oltre agli eventi citati in precedenza, l'Istat ha organizzato un ciclo di incontri svolti in tutte le regioni italiane, da maggio a luglio 2014, per diffondere sul territorio i risultati del 9° Censimento generale dell'industria e dei servizi. Ciascun appuntamento ha rappresentato un'opportunità di confronto e di dibattito sul contributo offerto dal censimento alla conoscenza del sistema produttivo locale.

Al termine della tornata censuaria 2010-2011 è sorta la necessità di dare continuità alle numerose e rilevanti innovazioni di metodo e di processo realizzate. Il 9° Censimento ha costituito un banco di prova significativo di processi produttivi articolati e complessi che ben si adattano alla fase successiva dei processi statistici connessi alla produzione di informazioni statistiche sulla struttura del sistema produttivo italiano.

L'Istituto nazionale di statistica è oggi impegnato nella progettazione di una nuova stagione dei censimenti continui i cui punti cardine riguardano l'uso integrato di indagini e di informazioni di fonte amministrativa. In particolare, per quanto riguarda le unità giuridico-economiche l'attività può essere divisa in due macro filoni: a) la realizzazione annuale di un sistema integrato di registri statistici per le diverse tipologie di unità economiche (imprese, aziende agricole, istituzioni pubbliche, istituzioni non profit); b) la realizzazione annuale o biennale di indagini necessarie non solo a verificare la qualità delle fonti amministrative e a garantire l'eshaustività dei registri da esse derivanti, ma anche a completare le informazioni in esse non presenti. Le esperienze acquisite e gli investimenti realizzati con il 9° Censimento costituiscono il punto di riferimento sia per l'aggiornamento annuale del sistema dei registri, sia per lo svolgimento delle rilevazioni periodiche.

Introduzione

La Rilevazione sulle imprese del 9° Censimento generale dell'industria e dei servizi è stata strutturata in maniera articolata e fortemente innovativa grazie sia all'uso statistico delle fonti amministrative locali e centrali e all'integrazione tra i dati in esse contenuti sia alla rilevazione di informazioni mai censite in passato.

Sul piano delle fonti amministrative è stato possibile realizzare un "censimento virtuale" grazie all'uso intensivo e massimale delle informazioni presenti nel sistema dei registri statistici imperniato sull'Archivio statistico delle imprese attive (Asia). Asia, che è realizzato annualmente dall'Istat, sia per le imprese che per le unità locali afferenti, è stato integrato per l'occasione censuaria grazie alla disponibilità di nuove fonti amministrative (ad esempio sulla struttura dell'occupazione delle imprese), che ha portato alla realizzazione di un nuovo prodotto, denominato Asia-occupazione. Il patrimonio informativo residente nei registri statistici ha consentito di pubblicare dati di base sulla struttura dell'universo delle imprese senza effettuare alcuna rilevazione di campo. È stato così possibile fornire il numero di imprese e di unità locali afferenti, la relativa occupazione (addetti, indipendenti e dipendenti, e personale esterno rappresentato dai collaboratori professionali e dal personale temporaneo – ex interinale), al massimo dettaglio territoriale (sezione di censimento) e di attività economica svolta (categoria, 5 *digit*). Inoltre, grazie ad Asia-occupazione, è stato possibile fornire informazioni sulla struttura socio demografica degli occupati classificati rispetto all'unità locale presso cui operano. Nel senso ora descritto, pertanto, il censimento è stato realizzato in modo virtuale ossia utilizzando interamente le informazioni contenute nei registri statistici e nelle fonti amministrative.

Nella Parte Prima del presente fascicolo sono descritte le caratteristiche definitorie, classificatorie e metodologiche del sistema dei registri statistici Asia. La prima parte del presente fascicolo è organizzata e suddivisa in quattro capitoli, nell'arco dei quali vengono approfonditi tre dei registri principali del sistema che contengono le imprese e che sono quelli innovati nell'operazione Cis2011: Asia-imprese, Asia-occupazione e Asia-unità locali.

Nella Parte Seconda del fascicolo sono invece descritte le caratteristiche metodologiche e di contenuto delle tre rilevazioni svolte nell'ambito del censimento ad integrazione del suo patrimonio informativo strutturale. Infatti, specifiche informazioni sono state rilevate dall'Istat presso le imprese mediante appositi questionari distribuiti a campioni di unità mediante rilevazioni condotte con riferimento a tre sottopopolazioni, per ciascuna delle quali è stata adottata un'organizzazione fondata su piani di campionamento distinti e questionari differenti; nel Capitolo 1 è descritta la Rilevazione sulle unità locali delle imprese, nel Capitolo 2 la Rilevazione multiscopo sulle imprese, nel Capitolo 3 la Rilevazione multiscopo sulle unità complesse.

PARTE PRIMA

IL CENSIMENTO VIRTUALE: IL SISTEMA DEI REGISTRI SULLE
UNITÀ ECONOMICHE ASIA

Capitolo 1

Il censimento virtuale e il sistema dei registri Asia⁵

Per la prima volta, nel 2011, il decennale Censimento dell'industria e dei servizi (Cis2011), inteso come strumento necessario a fornire una fotografia della struttura delle imprese attive e della relativa occupazione ad una certa data e al massimo dettaglio territoriale (sezione di censimento), viene realizzato in modo virtuale ossia utilizzando interamente le informazioni contenute nei registri statistici e nelle fonti amministrative. Si tratta di una profonda innovazione, realizzata attraverso un metodo di lavoro che ha previsto la collaborazione di oltre venti esperti appartenenti a diversi settori dell'Istat, che hanno condiviso gli aspetti concettuali/definitivi e la definizione delle metodologie per il trattamento dei dati amministrativi⁶.

La produzione di statistica ufficiale basata sull'uso massivo di dati provenienti dai registri statistici si è resa possibile grazie all'esistenza di un Archivio statistico delle imprese attive (Asia) aggiornato con cadenza annuale e consolidato nel tempo, e alla disponibilità di nuove fonti amministrative contenenti unità e variabili di base, (ad esempio sulla struttura dell'occupazione delle imprese), con un dettaglio informativo tale da poter rispondere alle necessità statistiche di classificazione.

Il ricorso all'uso quasi esclusivo di dati provenienti da fonti amministrative, opportunamente integrati con informazioni di natura statistica per descrivere caratteristiche strutturali di una popolazione di unità, presenta numerosi vantaggi, primo tra tutti il risparmio sia in termini di riduzione del fastidio statistico alle imprese sia in termini di costi sostenuti nel condurre rilevazioni censuarie.

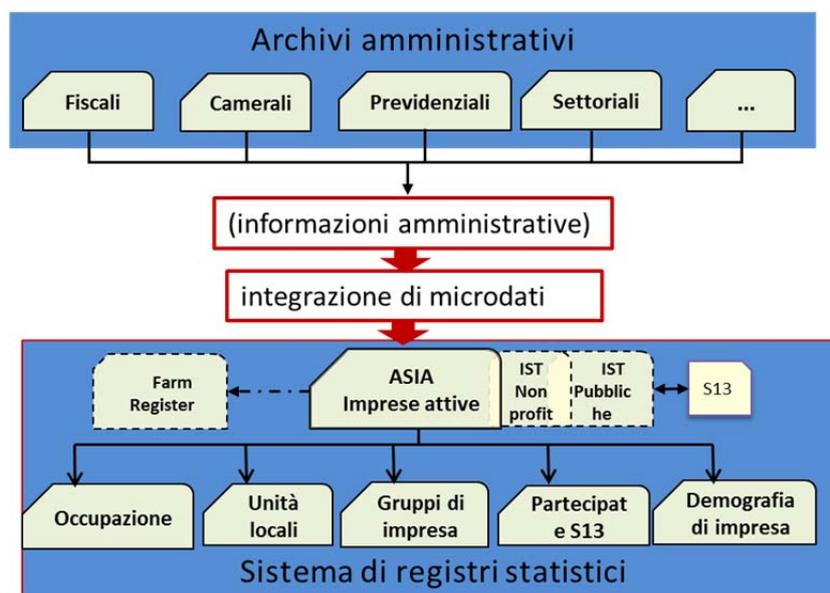
Il registro Asia produce dal 1996 il *frame* delle unità economiche attive (imprese e unità locali) e i relativi caratteri anagrafici (denominazione) e di struttura (attività economica, dimensione, forma giuridica), di localizzazione, e demografici (ad esempio date di inizio e fine attività).

La funzione dell'Archivio statistico delle imprese, che tradizionalmente è stata quella di fornire popolazioni di unità statistiche (*frame*) da cui estrarre campioni e di essere uno strumento a supporto quasi esclusivo delle indagini economiche, si è evoluta nel tempo. Oggi il registro ricopre altri importanti ruoli: in primo luogo è uno strumento fondamentale per la mobilitazione, integrazione e uso di dati provenienti da fonte amministrativa oltre che quella statistica; in secondo luogo contenendo sempre più variabili che descrivono la struttura delle imprese e delle altre unità statistiche, può essere utilizzato come fonte per la produzione diretta di statistiche economiche. Nuove analisi possono essere derivate dalla combinazione delle informazioni contenute nel Registro con informazioni provenienti da altri registri statistici che sono prodotti regolarmente (anche in relazione a richieste da regolamenti Ue), da fonti amministrative e dalle indagini dirette dell'Istituto, garantendo la coerenza tra i vari risultati statistici nell'ottica dello sviluppo di processi di produzione statistici più efficienti. Per questa sua centralità e per l'arricchito patrimonio informativo sviluppato nel corso del tempo è corretto oggi parlare di un sistema di registri Asia (Figura 1.1).

⁵ Parte Prima a cura di Caterina Viviano.

⁶ Task force Istat per la stima della struttura delle imprese e unità locali e della relativa occupazione per il Censimento 2011. DPTS, n. 150 del 29 luglio 2011.

Figura 1.1 – Il sistema dei registri statistici delle unità economiche



Il sistema di registri ad oggi sviluppato è costituito dai seguenti prodotti della statistica ufficiale: Asia-imprese, Asia-unità locali, Asia-occupazione, Asia-gruppi di impresa, lista delle unità istituzionali del settore S13⁷, registro delle unità partecipate da unità S13, dati di demografia di impresa, e registri non ancora definitivi quali Istituzioni non profit, Istituzioni pubbliche e *Farm Register*.

Il sistema di registri Asia ottenuto dallo sviluppo di procedure statistiche specifiche copre tutte le unità economiche e le loro relazioni e sono sostanzialmente basati sull'uso di dati amministrativi. Il ruolo principale è svolto da Asia-imprese che rappresenta il registro *base* del sistema con riferimento all'unità statistica Impresa e alla copertura della *target population*, come definita dai regolamenti Ue; ad esso sono interconnessi tramite un sistema di codici identificativi univoci tutti gli altri registri. In tale modo, ad esempio, le principali variabili per ciascuna unità sono derivate una sola volta, come ad esempio nel caso del numero di occupati, registrato dell'apposito registro Asia-occupazione e relativo alle imprese attive, aggiornate annualmente nel registro base Asia-imprese. Un fattore che ha facilitato l'utilizzo massivo di dati amministrativi, consentendo una loro integrazione a fini statistici è l'uso di sistemi di identificazione unificati. La presenza di identificativi univoci prima nelle fonti amministrative (il codice fiscale delle unità economiche e degli individui) e il conseguente sviluppo di un sistema unico di identificazione ossia una funzione che permette di assegnare un codice identificativo in maniera univoca a ciascuna unità del sistema, siano esse unità giuridiche o statistiche, imprese o individui, è un elemento cruciale per il funzionamento dell'intero sistema e quindi per la produzione di statistiche *register-based*.

⁷ Elenco delle unità istituzionali appartenenti al settore delle Amministrazioni Pubbliche (Settore S.13), i cui conti concorrono alla costruzione del Conto economico consolidato delle Amministrazioni Pubbliche. Tale elenco è redatto annualmente dall'Istat applicando le norme classificatorie e definitorie e i criteri di natura statistico-economica propri del Sistema europeo dei conti. L'elenco, ai sensi della L. 31 dicembre 2009, n. 196, art. 1, comma 3 (e successive modificazioni e integrazioni) è pubblicato in Gazzetta Ufficiale entro il 30 settembre di ogni anno.

Il Censimento generale dell'industria e dei servizi del 2011 (Cis2011) è stata occasione per passare, ufficialmente, alla produzione di statistiche “*register-based*”, per sostituire dati che tradizionalmente con i censimenti economici venivano collezionati tramite intervista del rispondente. Avendo questo obiettivo, il Cis2011 è stata anche l'occasione per introdurre significativi cambiamenti che hanno interessato il processo di produzione e aggiornamento del registro statistico Asia, con l'orientamento ad un significativo miglioramento nella tempistica di diffusione dei dati, una revisione complessiva del sistema delle definizioni statistiche adottate e ad un ampliamento delle informazioni prodotte, garantendo anche una maggiore coerenza per tutto il Sistema statistico nazionale. In particolare il processo per la realizzazione di un nuovo sistema informativo sull'occupazione si è realizzato attraverso un nuovo approccio logico e nuove metodologie che hanno permesso di: a) definire e identificare le differenti tipologie di lavoratori utilizzati dall'impresa; b) produrre una più ampia e più dettagliata quantità di informazione sugli occupati, sia rispetto alle caratteristiche demografiche del lavoratore (sesso, età, paese di nascita), sia sulle caratteristiche del rapporto di lavoro che qualificano il lavoratore (posizione nella professione, tempo determinato/indeterminato, full time/part time); c) revisionare le metodologie di stima e le regole operative per misurare gli occupati indipendenti; d) individuare nuove metodologie per misurare l'occupazione; e) rivedere e risolvere problemi di misura quali il calcolo dell'occupazione in media annua e al 31 dicembre; f) identificare il luogo di lavoro dei lavoratori utilizzando metodi statistici innovativi per l'allocazione del singolo lavoratore.

Le innovazioni introdotte nel processo produttivo di Asia e nel collegato sistema informativo sull'occupazione si sono rese possibili grazie alla disponibilità di un insieme di nuove fonti amministrative e allo sviluppo di un *framework* concettuale di definizioni e regole per il trattamento di tali dati a fini statistici. Tutto ciò ha consentito il passaggio da un sistema dove l'occupazione è considerato un attributo di tale unità, ad un sistema in cui ciascuna unità economica è collegata ad uno o più individui singolarmente identificati secondo una struttura concettuale di tipo LEED.

La struttura informativa di tipo LEED (*Linked Employer Employee Database*) permette di collegare quindi - attraverso un processo di integrazione di fonti amministrative (previdenziali, camerali, assicurative e fiscali) - ciascun individuo *lavoratore* con l'impresa in cui svolge l'attività lavorativa per tramite di un rapporto di lavoro (*job*), classificato secondo le forme occupazionali omogenee agli standard internazionali. In tale sistema ciascuna unità appartenente alla fonte amministrativa è collegata al registro statistico Asia grazie ad un sistema di codici identificativi: l'individuo può essere associato all'impresa presso cui svolge una attività lavorativa come dipendente oppure di cui ne è proprietario perché ne possiede quote societarie e al contempo vi lavora, eccetera. Tale struttura integrata contiene, relativamente a uno specifico periodo temporale *t* (settimana, mese, anno), le caratteristiche del rapporto di lavoro e alcune variabili demografiche del lavoratore.

Sulla base del *framework* concettuale/definitorio adottato, la classificazione delle tipologie di occupazione di una impresa è suddivisa in una componente cosiddetta interna all'impresa, rappresentata dagli addetti – dipendente e indipendente – ed una componente chiamata esterna costituita da quei lavoratori che partecipano al processo produttivo attraverso forme di lavoro così detto “parasubordinato”, non solo con contratti di collaborazione tipo Co.Co.Co e Co.Co.Pro con o senza partita Iva ma anche ad esempio i contratti con voucher o gli amministratori/manager, oppure tramite lavoro somministrato (lavoratori temporanei o ex interinali). Ciascuna forma occupazionale ha le proprie specificità, ad esempio, mentre per il lavoratore dipendente la qualifica professionale è un elemento rilevante per le analisi, per il lavoratore indipendente è rilevante capire se si tratta di titolare, socio o familiare. Tale approccio consente per la prima volta di misurare l'input di lavoro complessivo “utilizzato” da una singola unità economica, consentendo, ad esempio, una più significativa analisi della sua produttività.

Il presente fascicolo è organizzato e suddiviso in quattro capitoli, nell'arco dei quali vengono approfonditi tre dei registri principali del sistema che contengono le imprese e che sono quelli innovati nell'operazione Cis2011: Asia-imprese, Asia-occupazione e Asia-unità locali.

Il Capitolo 2 costituisce di fatto una sorta di manuale metodologico che descrive il processo per la produzione e aggiornamento del registro statistico delle imprese attive Asia, in cui si descrive: (i) il ruolo che ha assunto all'interno del sistema delle statistiche economiche dell'Istat, (ii) il processo di costruzione e aggiornamento fortemente basato sull'uso integrato di fonti amministrative, (iii) i contenuti in termini di unità, loro relazioni e variabili, (iv) le principali metodologie di stima dei caratteri e infine (v) le attività per la diffusione delle informazioni e in particolare per il rispetto dei criteri necessari a fornire dati puntuali a carattere censuario (alla data del 31 dicembre). Inoltre, considerando da un lato il fenomeno di globalizzazione economica ormai in atto da alcuni anni e che determina la costituzione di complesse strutture organizzative tra le imprese, e dall'altro la necessità di implementare in modo più consistente la definizione di impresa che i regolamenti europei impongono, è descritto, in sintesi, il processo di aggiornamento del Registro dei gruppi d'impresa.

Il Capitolo 3 è dedicato alla descrizione della metodologia sviluppata per la costruzione del nuovo sistema informativo sull'occupazione delle imprese. Dopo una panoramica sulle definizioni e gli standard nazionali e internazionali e sul *framework* concettuale adottato per la classificazione delle differenti componenti o tipologie occupazionali, i paragrafi successivi illustrano, con riferimento alle diverse tipologie occupazionali, le fonti amministrative utilizzate, le metodologie e il processo per l'integrazione di tali fonti e le variabili derivate.

Infine il Capitolo 4 è dedicato alle unità locali. Si descrive il processo per la realizzazione e aggiornamento del registro Asia-unità locali, le fonti amministrative utilizzate, i principali metodi di stima dei caratteri per definire le unità locali, la rilevanza dell'uso di una indagine di supporto al registro (Iulgi) necessaria a consentire l'aggiornamento delle unità più complesse con un livello di qualità difficilmente ottenibile dal solo utilizzo di informazioni di natura amministrativa. Grazie alla disponibilità della nuova base dati sull'occupazione, inoltre, si descrive anche la nuova procedura per la stima dei lavoratori dell'impresa allocati a livello di singola unità locale permettendo quindi l'identificazione del luogo di lavoro del singolo occupato.

Capitolo 2

Il registro delle imprese attive Asia⁸

2.1 Il ruolo del registro nel sistema delle statistiche economiche

All'interno del sistema statistico dell'Unione Europea, un ruolo di primaria importanza ha assunto la costruzione di un "registro delle imprese" che rappresenta l'universo di riferimento unico per tutte le statistiche economiche. La realizzazione e l'utilizzo di tale universo costituisce quindi, nel progetto europeo, il presupposto per rendere integrabili e comparabili le informazioni economiche relative alle unità che descrivono il sistema economico di un paese. A conferma dell'importanza di questo strumento si sottolinea che il regolamento che ne definisce le caratteristiche è uno dei primi emanati dalla Ue e risale al 1993⁹. L'Archivio statistico delle imprese attive (Asia) è lo strumento che l'Istat ha realizzato dal 1996 per rispondere a quanto richiesto da tale regolamento e costituisce il quadro informativo di riferimento per tutte le indagini sulle imprese effettuate dall'Istituto e anche dagli altri enti Sistan.

Nel corso del tempo il ruolo di tale registro è significativamente mutato, sia per iniziativa autonoma dell'Istat sia per l'approvazione di ulteriori regolamenti della Ue che ampliano il campo di osservazione e l'insieme di informazioni del registro. Si è, quindi, notevolmente arricchito il patrimonio informativo reso disponibile, tanto che oggi è più corretto parlare del sistema di registri Asia e non dell'archivio delle imprese Asia. Il registro ha un ruolo centrale nella produzione delle statistiche economiche, sia con riferimento al modo in cui le statistiche sono prodotte sia relativamente al loro contenuto e alla loro qualità. Tradizionalmente la sua funzione è quella di fornire la popolazione delle unità statistiche da cui poter estrarre i *frame* ed i campioni per le indagini, così come gli strumenti atti a supportare la fase di monitoraggio o per la riduzione dell'onere statistico. In aggiunta, oggi il registro è utilizzato direttamente per la produzione di statistiche sulla struttura della popolazione delle imprese per settore di attività economica, dimensione delle imprese e territorio. È la base per le analisi sull'evoluzione della struttura delle imprese e sulla loro demografia, ed è anche base per la realizzazione di nuovi prodotti statistici ottenuti come risultato della sua integrazione con altri registri statistici o archivi amministrativi.

⁸ A cura di Monica Consalvi. Autori del capitolo:

- Paragrafo 2.1: Monica Consalvi;
- Paragrafo 2.2: Monica Consalvi (Paragrafi 2.2.1 e 2.2.3), Monica Consalvi e Serena Migliardo (Paragrafo 2.2.2), Monica Consalvi e Mirella Morrone (Paragrafo 2.2.4);
- Paragrafo 2.3: Monica Consalvi, Serena Migliardo;
- Paragrafo 2.4: Monica Consalvi (Paragrafo 2.4.1), Monica Consalvi, Laura Maglione e Carla Runci (Paragrafo 2.4.2);
- Paragrafo 2.5: Simone Ambroselli (Paragrafo 2.5.1), Patrizia Cella (Paragrafo 2.5.2), Mirella Morrone (Paragrafo 2.5.3), Barbara Gentili e Arianna Speranza (Paragrafo 2.5.4);
- Paragrafo 2.6: Barbara Gentili e Arianna Speranza;
- Paragrafo 2.7: Francesca Alonzi e Alessandra Fiori;
- Paragrafo 2.8: Monica Consalvi e Barbara Gentili.

⁹ Il registro Asia nasce nel 1996 in base al Regolamento del Consiglio Europeo n. 2186/93 relativo al coordinamento comunitario dello sviluppo dei registri d'impresa utilizzati a fini statistici. Al fine di garantire lo sviluppo in un quadro armonizzato dei registri di imprese si è ritenuto opportuno adottare un nuovo Regolamento del Parlamento europeo e del Consiglio n. 177/2008 del 20 febbraio 2008, che istituisce un quadro comune per i registri di imprese utilizzati a fini statistici e abroga il Regolamento (CEE) n. 2186/93 del Consiglio. I cambiamenti di rilievo riguardano principalmente l'ampliamento del campo di osservazione in termini di settori di attività economica coperti e in termini di unità contenute nel registro.

Pertanto il registro è l'infrastruttura unica per assicurare la coerenza per la produzione di tutte le statistiche economiche e per lo sviluppo di processi di produzione più efficienti.

Il futuro del registro statistico prevede inevitabilmente un allargamento dell'orizzonte nazionale. La globalizzazione economica ormai in atto da alcuni anni pone delle nuove sfide alla produzione del registro, e rende più complessa e difficile la raccolta dei dati, anche per la presenza crescente di accordi di produzione globali tra unità residenti e unità estere e di complesse strutture organizzative tra le imprese (i gruppi d'impresa). In questo contesto in continuo cambiamento il registro deve essere in grado di raccogliere e fornire informazioni coerenti e rilevanti, soddisfacendo la crescente domanda di un'informazione comparabile a livello internazionale.

Il patrimonio informativo oggi disponibile dal sistema Asia è tale che per la prima volta sostituisce il tradizionale Censimento generale dell'industria e servizi dell'anno 2011, relativamente alle informazioni di base (struttura e occupazione), contribuendo in modo rilevante alla riduzione sia dei costi dell'operazione censuaria sia del fastidio statistico per le imprese.

I diversi ruoli del registro (Tavola 2.1) rendono necessaria la diffusione di opportune tipologie di prodotti confacenti al loro diverso utilizzo, tra le quali si distinguono le due principali, il registro *live* o *running* denominato anche "Asia corrente" e il registro *frozen* o congelato contenente la lista delle imprese attive al tempo t . In particolare, pur essendo possibile produrre più *frozen file* nel corso dell'anno, la scelta del momento in cui produrre una unica versione definitiva di tale file – che diventa la popolazione di riferimento – è legata sia ai tempi di realizzazione del processo di aggiornamento rispetto allo specifico anno di riferimento sia alle tempistiche richieste dagli utenti finali, principalmente di natura strutturale. Al fine di garantire la consistenza delle informazioni statistiche prodotte dalle molteplici indagini economiche è necessario fornire loro una base comune di riferimento, sia per l'estrazione dei campioni sia per il riporto all'universo delle informazioni campionarie. Accade però che il registro sia continuamente revisionato durante l'anno e che questa attività comporti l'introduzione di nuovi dati e/o la correzione di errori o l'inserimento di aggiornamenti su alcune variabili per specifiche unità. Il costante aggiornamento del registro potrebbe causare disallineamenti nella popolazione di riferimento delle indagini quando queste fossero realizzate in periodi diversi. Infatti le indagini Istat, anche se aventi lo stesso periodo di riferimento, sono avviate in momenti diversi e quindi potrebbero estrarre la lista di imprese da indagare (sia censuaria, per particolari sottoinsiemi, sia campionaria) da due diverse "foto" del registro. Tale situazione potrebbe indurre delle differenze nei risultati delle diverse indagini. La soluzione adottata per evitare tale disomogeneità è quella di produrre una edizione del registro ad una data prefissata, il cosiddetto *frozen file*, l'edizione "congelata" del registro, da considerarsi come una fotografia del registro scattata in un istante temporale; essa è solitamente diffusa alla fine del primo trimestre di ogni anno solare ($t+15$ mesi) e costituisce l'universo delle imprese attive per l'anno t . Rimanendo invariata nell'anno t essa rappresenta, quindi, la popolazione di riferimento per tutte le indagini sulle imprese che hanno tale anno come periodo di riferimento. Le correzioni e gli aggiornamenti dei dati sono inseriti nel corso dell'anno nella versione *live* o *running*, il database relazionale di Asia, denominata nel seguito "Asia corrente" ed entrano a far parte della nuova struttura delle imprese italiane a disposizione degli utilizzatori nella edizione successiva del file annuale *frozen* o congelato.

Tavola 2.1 – Panoramica dei ruoli del registro statistico

	RUOLO	SCOPO
1	Registro <i>live</i> o <i>running</i> (Asia corrente)	Detiene e costruisce le unità statistiche; ha il compito, concettuale e fisico, di identificare le unità per l'osservazione e l'analisi statistica, a partire dalle unità giuridiche e amministrative, e assegnare ad esse dei codici identificativi univoci. Consente il raccordo tra i dati provenienti dalle diverse fonti di input e l'unità statistica, fornendo i legami tra le unità statistiche e quelle amministrative. Su base giornaliera il registro è aggiornato con le informazioni provenienti da diverse fonti, utilizzate per creare o cessare le unità o aggiornarne le caratteristiche; si tiene traccia di tutti i cambiamenti, registrando le informazioni e gli opportuni metadati in un registro storico.
2	Lista unità attive anno <i>t</i>	Fornisce la popolazione delle unità statistiche valida per un determinato periodo di riferimento (anno <i>t</i>), l'insieme delle unità attive nel periodo e le informazioni necessarie per le indagini. Quindi rappresenta una popolazione coordinata di unità statistiche nello spazio e nel tempo. È la base di riferimento per il riporto all'universo dei dati campionari. È anche utilizzato per le stime di Contabilità Nazionale.
3	Registro <i>frozen</i> o congelato	Fornisce la base per la creazione dei <i>frame</i> per le indagini, un sottoinsieme di unità statistiche <i>ad hoc</i> per ogni indagine contenente le unità valide per un determinato periodo di riferimento con tutte le caratteristiche richieste.
4	<i>Frame</i> per le indagini economiche	Registrando le informazioni sullo stato delle imprese e sul loro onere di risposta alle indagini figura anche come supporto per l'ottimizzazione dei disegni di indagine, per il loro coordinamento e per il monitoraggio dello stato della compilazione. Soprattutto in caso di imprese non rispondenti fornisce un aiuto per la gestione dei solleciti e degli eventuali procedimenti sanzionatori.
5	Supporto per le indagini economiche	Fornisce delle statistiche basate direttamente su dati del registro, sia mediante <i>data warehouse</i> I.Stat sia su richiesta elaborando delle statistiche <i>ad hoc</i> per diverse tipologie di utenti esterni. Tale diffusione è vincolata dalle disposizioni di riservatezza applicabili.
6	Fonte informativa sulla struttura delle imprese (microdati)	Fornisce liste di imprese (ma anche unità locali e gruppi di imprese) prevalentemente per utenti del Sistan, con la sola finalità dell'esecuzione di trattamenti connessi all'attività istituzionale o all'ambito territoriale del richiedente, quindi solo relativamente alla porzione del registro per la quale possiedono le necessarie autorizzazioni.
7	Garanzia di coerenza per scambio di dati anche a livello internazionale	Promuove la coerenza e la comparabilità internazionale delle statistiche economiche. È collegato all' <i>EuroGroups Register</i> (EGR), mantenuto da Eurostat, per consentire la produzione di statistiche sulla globalizzazione. Facilita lo scambio di dati e metadati nei vari paesi, particolarmente importante per garantire la coerenza delle statistiche economiche transfrontaliere, ad esempio tra le statistiche <i>Inward</i> e <i>Outward</i> delle affiliate estere (FATS) di due paesi.
8	Presupposto per la modernizzazione della produzione e dei servizi statistici	Si occupa dell'integrazione e della gestione delle fonti più rilevanti, facilita l'utilizzo delle fonti amministrative e l'integrazione tra fonti tradizionali e i nuovi media, così come può rendere possibile un nuovo tipo di produzione delle statistiche, standardizzato e "industrializzato", sia a livello nazionale sia internazionale. Il ruolo centrale del registro all'interno del processo produttivo delle statistiche economiche è rafforzato quando è visto come supporto ai processi come descritto nel <i>Generic Statistical Business Process Model</i> (GSBPM).

2.2 La copertura del registro e il suo contenuto in termini di unità e variabili

2.2.1 La copertura del registro

Per garantire la completezza un registro statistico deve contenere tutte le unità istituzionali che svolgono un'attività economica e che contribuiscono alla formazione del prodotto interno lordo ai prezzi di mercato (Pil) secondo i limiti di produzione previsti dal Sec; deve garantire la più estesa copertura – proporzione della produzione economica nazionale che le unità rappresentano – e possibilmente un rilevante contenuto – il set di caratteristiche delle unità contenute nel registro.

Inoltre il registro fornisce le versioni cosiddette "congelate", che come delle istantanee forniscono una fotografia del registro in un dato momento e contengono tutte le unità che nel registro *live* sono attive in un determinato periodo di riferimento. I *frame* per le indagini economiche corrispondono proprio a tali versioni congelate. Pertanto la copertura del registro per un determinato periodo di riferimento e per una specifica indagine è la misura in cui tale *frame* estratto dal registro rappresenti in maniera adeguata la popolazione *target* dell'indagine.

L'Archivio statistico delle imprese attive è costituito dalle unità economiche che esercitano arti e professioni nelle attività industriali, commerciali e dei servizi alle imprese e alle famiglie e fornisce

informazioni identificative (denominazione e indirizzo) e di struttura (attività economica, addetti dipendenti e indipendenti, forma giuridica, data di inizio e fine attività, fatturato) di tali unità.

Per la sua realizzazione sono raccolti ed elaborati i dati relativi a tutte le unità giuridiche presenti nel territorio italiano, sia residenti che non residenti ma aventi stabile organizzazione, le cui informazioni sono registrate nei principali archivi amministrativi utilizzati come input nel processo di aggiornamento del registro. Il campo di osservazione dei dati diffusi ogni anno si è adeguato inizialmente a quello delle statistiche economiche strutturali¹⁰. Pertanto, nonostante presenti nel sistema informativo del registro, dal campo di osservazione del registro delle imprese attive sono escluse le attività economiche relative a: Agricoltura, silvicoltura e pesca (sezione A della classificazione Nace Rev. 2); Amministrazione pubblica e difesa; assicurazione sociale obbligatoria (sezione O); Attività di organizzazioni associative (divisione 94); Attività di famiglie e convivenze come datori di lavoro per personale domestico; produzione di beni e servizi indifferenziati per uso proprio da parte di famiglie e convivenze (sezione T); Organizzazioni ed organismi extraterritoriali (sezione U). Inoltre, sono escluse le unità classificate come istituzioni pubbliche e istituzioni private non profit, a prescindere dall'attività svolta; insieme alle aziende agricole questi ultimi tre insieme sono organizzati in registri separati e non sono oggetto di approfondimento. In particolare, nel 2011 le istituzioni pubbliche e le istituzioni non profit sono state rilevate dai due specifici censimenti¹¹. La distinzione tra produzione di beni e servizi *market* (destinabili alla vendita) e *non market* è importante per la comprensione di quale sia la copertura assicurata dal registro. Le unità *market* sono imprese che vendono tutto o la maggior parte del loro output a prezzi economicamente significativi, definiti come i prezzi che producono un effetto considerevole sui quantitativi di prodotti che i produttori sono disposti a fornire e sui quantitativi di prodotti che gli acquirenti sono pronti ad acquistare. I produttori di beni e servizi destinabili alla vendita prendono decisioni su cosa e su quanto produrre in risposta ai livelli attesi di domanda e dei costi previsti di fornitura e sono esposti ai rischi associati a questa produzione. Sono incentivati ad adeguare l'offerta con lo scopo di realizzare un profitto a lungo termine o, come minimo, di coprire i costi di capitale e gli altri costi.

La copertura del registro statistico comprende le unità *market* la cui attività principale è la produzione di beni e servizi. È quella che si definisce la copertura di base di un registro in quanto comprende le unità più rilevanti da un punto di vista economico.

2.2.2 Le unità del registro e le loro relazioni

Scopo di un registro di imprese è la registrazione delle unità statistiche, delle loro relazioni e delle loro caratteristiche a un certo livello di qualità in un certo momento.

Le unità contenute nel registro sono le unità statistiche, ossia definite per scopi statistici, le unità per le quali si chiedono informazioni e si compilano le statistiche¹², e le unità giuridiche ed amministrative, le prime riconosciute per legge, le seconde create da unità giuridiche al fine di soddisfare delle norme amministrative o per la propria gestione. Le unità statistiche possono essere sia unità fisiche identificabili nel mondo reale sia unità teoricamente definite e costruite per specifiche esigenze statistiche. L'insieme delle unità contenute nel registro sono utili ad osservare i vari aspetti della realtà in maniera esaustiva e senza duplicazioni. È importante avere definizioni armonizzate per ciascuna delle unità statistiche, in modo tale da poter garantire statistiche economiche coerenti tra settori economici, paesi e aree geografiche, e nel corso del tempo. È

¹⁰ Regolamento Ue n. 295/2008 per le statistiche strutturali (SBS – *Structural Business Statistics*).

¹¹ Si rimanda ai Fascicoli 3 e 4 per la loro descrizione.

¹² Nel glossario Ocse (Organizzazione per la Cooperazione e lo Sviluppo Economico) le unità statistiche sono definite come "*the entities for which information is sought and for which statistics are ultimately compiled*".

anche importante avere un quadro concettuale ben chiaro e una conoscenza approfondita delle unità giuridiche e amministrative, considerando l'uso sempre più ampio che se ne fa per la riduzione dell'onere di risposta che si ottiene attraverso la raccolta di dati provenienti da fonti amministrative.

Le definizioni delle unità statistiche adottate sono conformi a quanto dispone il Regolamento del Consiglio Europeo (CEE) n. 696 del 15 marzo 1993, relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità. In questo regolamento per costituire l'unità impresa si è fatto ricorso all'utilizzazione delle unità giuridiche che esercitano totalmente o parzialmente un'attività produttiva. *“Le unità giuridiche sono sia persone giuridiche la cui esistenza è riconosciuta dalla legge indipendentemente dalle persone o dalle istituzioni che le possiedono o che ne sono membri; sia persone fisiche che esercitano un'attività economica come indipendenti. L'unità giuridica costituisce sempre, da sola o talvolta con altre unità giuridiche, il supporto giuridico dell'unità statistica impresa”*¹³.

Tale regolamento individua l'*impresa* come l'unità centrale per la realizzazione delle statistiche economiche e la definisce come *“la più piccola combinazione di unità giuridiche che costituisce un'unità organizzativa per la produzione di beni e servizi che fruisce d'una certa autonomia decisionale. In particolare per quanto attiene alla destinazione delle sue risorse correnti. Un'impresa esercita una o più attività in uno o più luoghi. Un'impresa può corrispondere ad una sola unità giuridica”*¹⁴. Secondo questa definizione sono considerati impresa anche i lavoratori autonomi e i liberi professionisti.

Requisito basilare della definizione in vigore è l'identificazione del concetto di autonomia dell'impresa in relazione alla sua capacità di decidere sull'allocazione delle risorse correnti, diversamente da quanto previsto per altre unità, quali i gruppi di impresa, dove l'autonomia è invece connessa alle decisioni strategiche e di lungo periodo.

Inoltre l'impresa è definita come un'entità economica che, in certe circostanze, può corrispondere al raggruppamento di più unità giuridiche, che rispondono dell'impresa e ne costituiscono sempre il supporto giuridico. La necessità di raggruppare, sotto certe circostanze, più di una unità giuridica in un'unica unità statistica è giustificato dall'esistenza di unità giuridiche che esercitano la loro attività esclusivamente a favore o a supporto di altre entità giuridiche da cui dipendono e la loro esistenza è dovuta unicamente a ragioni amministrative, ad esempio fiscali, senza assumere rilevanza dal punto di vista economico. Rientrano in questa categoria anche una grande parte delle unità giuridiche senza occupazione. Spesso le loro attività devono essere interpretate come attività ausiliarie dell'unità giuridica madre a cui esse appartengono e a cui devono essere ricollegate per costituire l'entità “impresa”. È questo il caso delle attività di gestione della contabilità, del personale, della logistica e dei magazzini, dei sistemi informativi quando le attività sono svolte a supporto esclusivo ed indispensabile per lo svolgimento del processo produttivo di altre unità.

Nel registro statistico attualmente esiste una relazione univoca tra unità giuridica ed impresa. Le informazioni sulla base delle quali combinare insieme unità giuridiche per formare un'impresa non sono in molti casi disponibili, ma il sistema informativo può attualmente supportare la registrazione separata delle unità giuridiche e delle imprese come unità distinte. Di fatto ci sono tutte le necessarie competenze per distinguere e trattare singolarmente i casi più importanti nei quali le imprese che fanno parte di grandi gruppi, sia nazionali sia multinazionali, possono non coincidere con le unità giuridiche. Sono numerosi i progetti internazionali che in questi ultimi anni l'Istat ha

¹³ Regolamento (CEE) n. 696/93 del Consiglio del 15 marzo 1993 relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità (Allegato, sezione II A).

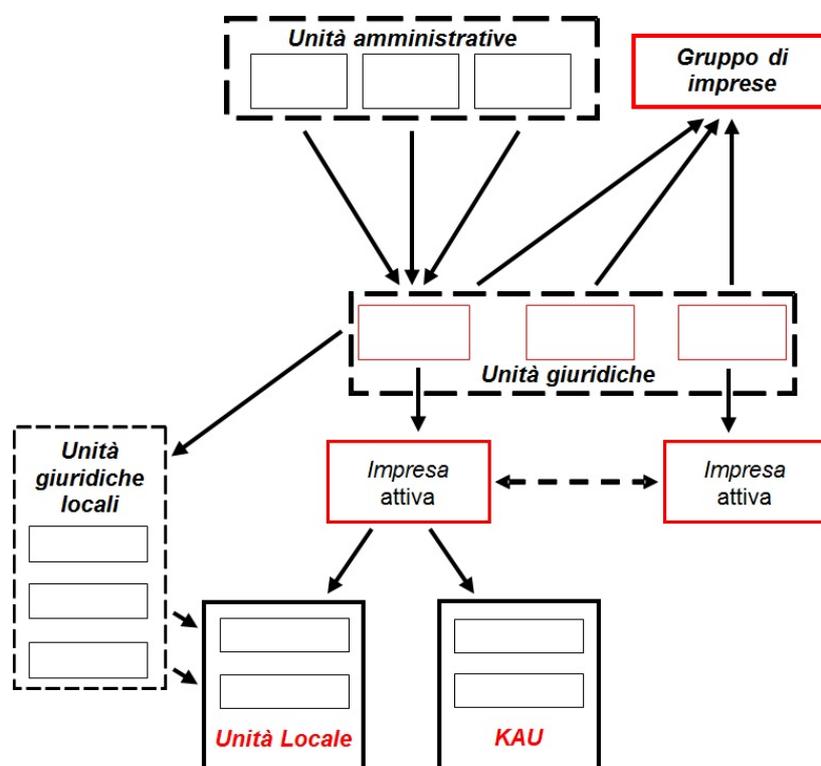
¹⁴ Regolamento (CEE) n. 696/93 del Consiglio del 15 marzo 1993 relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità (Allegato, sezione III A).

gestito o in cui è stata partecipe, nei quali si sono sottoposte a test le nuove procedure di analisi e trattamento dei dati per l'identificazione corretta dell'unità impresa¹⁵.

I vari tipi di unità statistiche non sono indipendenti, ma sono legati l'uno all'altro, secondo un modello delle unità statistiche, che è un quadro concettuale delle unità e delle loro relazioni. Il modello fornisce un'astrazione concettuale semplificata dell'insieme delle unità che esistono nel "mondo reale", il mondo economico che si intende misurare attraverso le unità statistiche.

Il modello utilizzato è basato sul Sec2010 e sul regolamento dei registri statistici del 2008¹⁶, che richiede che quattro unità – unità giuridica, impresa, unità locale e gruppo di imprese – devono essere contenute nei registri nazionali degli Stati membri dell'Ue e tenute aggiornate.

Figura 2.1 – Le unità e le loro relazioni nel sistema dei registri



¹⁵ Le definizioni delle unità statistiche, tra cui l'impresa e il gruppo di imprese sono attualmente in fase di revisione nell'Sse. L'obiettivo è quello di supportare lo sviluppo di statistiche che tengano conto della globalizzazione dell'economia e, in particolare, stabilire un regolamento trasversale che possa eliminare le cause di incoerenza tra paesi e tra settori statistici, permettendo l'integrazione delle statistiche sulle imprese negli Stati membri dell'Ue. Le definizioni revisionate prendono in considerazione il risultato di due programmi di lavoro quinquennali lanciati da Eurostat (*ESSnet on Profiling large and complex multinational Enterprise Groups* e *ESSnet on consistency of concepts and methods of business-related statistics*) e della "Task Force sulle unità statistiche" che hanno sviluppato definizioni, modelli e metodologie discussi in numerosi workshop internazionali, soprattutto con riferimento agli impatti organizzativi e metodologici delle definizioni proposte e ai loro costi per i singoli istituti di statistica. Le definizioni revisionate saranno incluse nel regolamento quadro di Eurostat *Framework Regulation Integrating Business Statistics* (FRIBS) per la loro applicazione negli anni a venire.

¹⁶ *Business Register (BR) Regulation* (EC) n. 177/2008.

Le unità sono gerarchicamente organizzate come illustrato nella Figura 2.1. Ciascuna impresa è collegata a (consta di) una o più unità giuridiche e ciascuna impresa è collegata a (consta di) una o più unità locali, dove svolge la sua attività¹⁷. È opportuno notare che talune unità giuridiche, ad esempio le unità inattive, non sono necessariamente collegate ad una qualche impresa, mentre le unità locali sono sempre collegate a un'impresa.

Il gruppo di imprese costituisce l'unità statistica di ordine gerarchico superiore, una struttura complessa che può includere diverse unità giuridiche e diverse imprese, collegate tra loro da legami funzionali o proprietari. Sempre secondo lo stesso regolamento si definisce come *“un'associazione di imprese retta da legami di tipo finanziario e non”* avente *“diversi centri decisionali, in particolare per quel che concerne la politica della produzione, della vendita, degli utili”* e in grado di *“unificare alcuni aspetti della gestione finanziaria e della fiscalità”*. Il gruppo si caratterizza come *“l'entità economica che può effettuare scelte con particolare riguardo alle unità alleate che lo compongono”*¹⁸.

Nel modello delle unità statistiche la relazione tra gruppo e unità giuridiche (o imprese) è di tipo *“uno a molti”*: ogni gruppo è sempre associato a una o più unità giuridiche (e di conseguenza a una o più imprese); ogni impresa può appartenere a un solo gruppo di imprese e non tutte le imprese fanno parte di un gruppo. È il caso, in particolare, delle imprese *“semplici”*, in cui la (sola) unità giuridica costituente non ha legami di controllo con nessun'altra unità giuridica¹⁹.

Il gruppo d'impresa è stato dichiarato obbligatorio per il registro in tempi successivi rispetto alle altre unità statistiche, con l'introduzione del nuovo Regolamento Comunitario n. 177/2008, quando si sono resi disponibili i criteri operativi per la sua individuazione e implementazione nel registro e soprattutto per concretizzare il legame di *“associazione”* previsto in tale regolamento²⁰.

Le unità giuridiche infatti possono essere caratterizzate da vari tipi di relazioni che derivano dai rispettivi diritti ed obblighi o che perlomeno sono coerenti con essi. Queste relazioni possono essere relazioni finanziarie – proprietà e controllo, contenute nel registro statistico – e relazioni commerciali²¹. Mentre le prime sono raccolte e mantenute nel registro, quelle commerciali non rientrano nel suo ambito.

La proprietà è una relazione forte tra una persona fisica o giuridica – entità proprietaria – e una persona giuridica e può essere totale o parziale. La facoltà di disporre liberamente di questo tipo di proprietà è talvolta limitata, ad esempio in alcune società le azioni non sono trasferibili.

Il controllo è un'altra relazione molto forte che intercorre sempre tra una persona fisica o giuridica controllante e un'unità giuridica controllata, che è sempre costituita da una persona giuridica. L'unità controllante influenza o ha il diritto di influenzare lo statuto, l'attività, l'uso delle risorse, la gestione o l'amministrazione dell'unità controllata. Solo il primo livello di controllo, diretto o indiretto, è registrato per ciascuna unità e l'intera catena di controllo può essere ricostruita

¹⁷ Per una trattazione estesa dell'*unità locale* si veda il Capitolo 4 di questo fascicolo.

¹⁸ Regolamento (CEE) n. 696/93 del Consiglio del 15 marzo 1993 relativo alle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità (Allegato, sezione III A).

¹⁹ Nel seguito sono analizzate solo le diverse tipologie di relazioni tra le unità giuridiche e si rimanda al Paragrafo 2.3 per una trattazione dettagliata del processo di realizzazione e aggiornamento del Registro gruppi d'impresa.

²⁰ A questa esigenza operativa ha risposto la metodologia armonizzata messa a punto da Eurostat in collaborazione con gli istituti statistici dei Paesi Membri nell'ambito del *Working Group “Enterprise Groups”* e pubblicata nel capitolo 21 del *Recommendation Manual on Business Registers* la cui redazione è stata curata dall'Istat nel 2002, dove sono state proposte le linee guida per il trattamento dei gruppi nei registri statistici in vista della revisione del regolamento, che ha introdotto l'obbligo della loro registrazione.

²¹ Tra le più importanti il subappalto, l'esternalizzazione o *outsourcing*, le alleanze strategiche, le collaborazioni a livello di marketing, gli accordi di licenza e gli accordi di affiliazione commerciale o franchising. La base delle relazioni commerciali può andare da un accordo informale a un accordo contrattuale.

utilizzando un algoritmo, sviluppato in Istat, in grado di risalire sino al controllante ultimo o vertice del gruppo.

Va sottolineato che, benché siano molto spesso legate tra loro, la relazione di proprietà, anche quando comporta una partecipazione maggioritaria, e la relazione di controllo non sono equivalenti. Ciascuna può esistere indipendentemente dall'altra. In genere, la proprietà del 50 per cento o più di una persona giuridica comporta la possibilità di controllo, ma non è sempre così, in particolare se le disposizioni legislative che ne determinano la forma giuridica consentono la formazione di minoranze di blocco riguardo a talune decisioni fondamentali della vita societaria (es. scioglimento, fusione con un'altra società, aumento di capitale, modifica dello statuto, trasformazione di obbligazioni in azioni, eccetera). Il controllo non necessariamente implica una partecipazione maggioritaria; a volte, nelle società per azioni, corrisponde a molto meno del 50 per cento delle quote azionarie.

Le relazioni suindicate non si escludono a vicenda. Al contrario, occorre spesso combinarle per rendere conto delle effettive relazioni che intercorrono tra le unità. Questa combinazione induce, a sua volta, a raggruppare le unità giuridiche proprio in base alle relazioni che le legano.

In particolare, il legame di associazione che è stato scelto per l'individuazione corretta dell'unità statistica gruppo d'impresa deve concretizzarsi a fini operativi in un effettivo legame di controllo, esercitato, in via diretta o indiretta, da parte del vertice del gruppo su tutte le altre imprese che lo compongono.

Il concetto di controllo è definito nel Regolamento Europeo n. 2223/1996 Sistema europeo dei conti (Sec95²²), che al paragrafo 2.26 recita: *“Per «controllo di una società» si intende la capacità di determinarne la politica generale, se necessario scegliendone gli amministratori. Una unità istituzionale — un'altra società, una famiglia o una unità delle amministrazioni pubbliche — esercita il controllo su una società quando detiene la proprietà di più della metà delle azioni con diritto di voto o controlla in altro modo più della metà dei voti degli azionisti. Inoltre, una amministrazione pubblica può esercitare il controllo su una società in forza di leggi o regolamenti che le danno il diritto di determinare la politica della società o di nominarne gli amministratori”*. Il Sec specifica inoltre che il concetto di controllo va inteso anche in senso indiretto.

I criteri operativi, improntati ai principi di oggettività, osservabilità e misurabilità delle informazioni utilizzate a fini statistici, prevedono che per poter istituire un legame di controllo tra imprese debba realizzarsi almeno una delle seguenti condizioni²³:

- una unità istituzionale controlla direttamente, attraverso il possesso del 50 per cento più uno dei diritti di voto, una società;
- una unità istituzionale controlla indirettamente, tramite altre società controllate, una terza società;
- una società consolida integralmente il bilancio di un'altra società, anche se possiede una quota inferiore al 50 per cento dei diritti di voto;
- una o più unità istituzionali dichiarano alla competente autorità di vigilanza di controllare, in accordo tra loro, una società quotata sul mercato regolamentato del paese membro.

Un'amministrazione pubblica esercita il controllo su una società in forza di leggi, decreti o regolamenti specifici che le danno il diritto di indirizzare la gestione della società.

²² Il 21 maggio 2013 è entrato in vigore il nuovo Regolamento Europeo n. 549/2013 (Sec2010) in cui il concetto di controllo è definito al paragrafo 2.35 e seguenti.

²³ Le società controllate possono essere: S.p.a., S.r.l., S.a.p.a., Cooperative, Consorzi di diritto privato, altre forme di cooperazione tra imprese, imprese costituite all'estero che svolgono attività economica in Italia (*branches*), autorità indipendenti, enti pubblici economici, aziende speciali e aziende pubbliche di servizi.

2.2.3 Le variabili del registro

Le variabili contenute nel registro possono essere classificate secondo quattro tipologie: variabili identificative (codici identificativi, ragione sociale, indirizzo e altri caratteri per l'esatta individuazione dell'unità sul territorio); variabili di stratificazione (attività economica dell'impresa classificata secondo la classificazione Ateco, forma giuridica, dimensione dell'impresa, in termini di addetti indipendenti e dipendenti medi annui e di fatturato); variabili demografiche (data di creazione e di cessazione dell'impresa, data di eventi quali scorpori, fusioni o procedure concorsuali, fallimenti, liquidazioni, eccetera); variabili di relazione, riguardanti le relazioni tra unità. Il registro contiene registrazioni e numeri di identificazione distinti per l'unità giuridica, l'impresa e l'unità locale, anche se queste coincidono. Tali codici rimangono gli stessi per tutta la vita dell'unità identificata. Di conseguenza, anche se le caratteristiche dell'unità possono cambiare durante la vita di quest'ultima, il numero di identificazione è indipendente da tali caratteristiche. La soluzione trovata per il registro Asia è stata quella di attribuire ad ogni unità un numero sequenziale non significativo dotato di una chiave di controllo informatica.

Riguardo la forma giuridica delle unità del registro, per la loro classificazione, è stata realizzata dall'Istat nel 2002²⁴ una specifica classificazione operando l'integrazione delle forme registrate nei principali archivi amministrativi, in particolare il Registro delle imprese della Cciaa e l'Anagrafe tributaria, per far fronte alla presenza di numerose duplicazioni esistenti nell'elenco utilizzato precedentemente, di forme giuridiche troppo generiche e per tener conto dell'evoluzione normativa degli anni più recenti che aveva determinato la nascita di nuove forme giuridiche e la modifica o cessazione di altre. La forma che i soggetti economici assumono è una caratteristica essenziale da cui non solo dipendono i loro diritti e doveri, ma anche il tipo di organizzazione, le regole di relazione fra i soggetti fisici associati, le finalità dell'organizzazione e, in ultima analisi, le modalità con cui il soggetto si relaziona con il mercato per la produzione, cessione o gestione di beni e servizi. La valenza e l'utilità di tale classificazione risiede nella possibilità, offerta dal criterio giuridico, di associare ad una determinata categoria di soggetti economici caratteristiche generali utili ad interpretare struttura e comportamenti organizzativi ed economici delle unità statistiche.

Ai fini della produzione dell'informazione statistica, le imprese sono classificate per attività economica, definita in base ad un livello specifico della nomenclatura Ateco²⁵. L'attività economica è la combinazione di risorse, quali attrezzature, manodopera, tecniche di fabbricazione, reti di informazione o di prodotti, che porta alla creazione di specifici beni o servizi. Le imprese sono distinte in base all'attività economica esclusiva o principale, secondo il criterio della prevalenza. Quando, nell'ambito di una stessa unità, sono esercitate più attività, la prevalenza è individuata sulla base della quota prevalente di valore aggiunto creata e in mancanza di tale dato, sulla base

²⁴ <http://www.istat.it/it/archivio/6523>.

²⁵ La nuova classificazione delle attività economiche Ateco 2007, in vigore dal 1 gennaio 2008, costituisce la versione nazionale della nuova classificazione europea delle attività economiche Nace Rev. 2, che corrisponde alla versione europea della nuova Isic Rev. 4, definita e adottata in ambito ONU. Essa rappresenta il risultato di un processo di aggiornamento e convergenza di classificazioni diverse (la Nace Rev. 1, impiegata principalmente in Europa e la Naics, adottata dai paesi del Nord America). La nuova versione nazionale della classificazione ha in comune con la Nace le prime quattro cifre e con la Isic Rev. 4 le prime due. I livelli di *categoria* (cinque cifre) e di *sotto-categoria* (sei cifre) sono propri della versione italiana. Attualmente, le informazioni statistiche diffuse dall'Istat e dagli altri enti del Sistan fanno riferimento solo al livello di categoria, mentre per fini amministrativi e fiscali viene utilizzata anche la sotto-categoria. Il valore aggiunto della nuova Nace e della nuova Ateco consiste nell'avere, per la prima volta, dati totalmente comparabili a livello internazionale. Informazioni specifiche sulla struttura, sulle novità della classificazione e sulle tavole di raccordo con le classificazioni precedenti sono disponibili sul sito dell'Istat all'indirizzo <http://www3.istat.it/strumenti/definizioni/ateco/>. Per la nuova metodologia di scelta del codice di attività economica prevalente per le imprese del registro si veda il Paragrafo 2.5.1.

del fatturato, del numero medio annuo di addetti, delle spese per il personale, delle retribuzioni lorde.

Nel caso in cui si disponga dell'informazione, anche le attività secondarie sono registrate, quando queste rappresentano una percentuale significativa dell'attività totale dell'impresa o dell'attività nazionale dello stesso tipo. Si tratta di informazioni importanti per varie statistiche, in particolare le statistiche strutturali sulle imprese. Esse possono essere utilizzate anche per stabilire indirettamente la probabile esistenza di unità di attività economica. Anche le fonti amministrative spesso contengono informazioni sulle attività secondarie, ma sono meno accurate.

Il registro contiene anche la misurazione della dimensione delle unità statistiche che è essenziale anche per la selezione dei campioni, per l'estrapolazione e la proiezione dei risultati sull'universo e per l'analisi di sottopopolazioni di unità. Per elaborare un indicatore di dimensione è possibile ricorrere a vari criteri, quali il numero di persone occupate o il fatturato dell'impresa. L'importante è disporre di un aggiornamento almeno annuale per tutte le unità, indipendentemente dal fatto che esse siano state selezionate o meno l'anno precedente per essere incluse nel campione di un'indagine. Di conseguenza, sono utilizzati soltanto questi indicatori perché possono essere aggiornati a partire da fonti amministrative esaustive. Questo vincolo ha limitato nel tempo la scelta dei possibili indicatori.

La dimensione dell'impresa in termini di addetti – lavoratori dipendenti e indipendenti – è calcolata come posizioni lavorative in media annua²⁶.

Il fatturato è l'ammontare complessivo derivante dalle cessioni di beni e dalle prestazioni di servizi effettuate nell'anno di riferimento.

Per quanto attiene le variabili demografiche queste rispondono a specifiche definizioni. La data di inizio attività dell'impresa corrisponde alla data della sua creazione. Dovrebbe essere in teoria la data in cui sono stati effettuati i primi investimenti finanziari, ma in pratica coincide con la data di registrazione o di inizio attività dichiarata nelle fonti amministrative.

La data di cessazione dell'impresa corrisponde alla data di cancellazione dell'unità negli archivi amministrativi, che avviene quando l'unità cessa di esistere, oppure quando tutti i mezzi di produzione vengono acquisiti da un'altra unità o trasferiti a un'altra unità.

I dati contenuti nel registro riflettono l'universo delle imprese attive, ossia unità che hanno svolto un'effettiva attività produttiva per almeno sei mesi nell'anno di riferimento, localizzate nel territorio nazionale²⁷. Le relative variabili di localizzazione si riferiscono alla sede amministrativa dell'impresa intendendo per sede amministrativa il luogo dove di fatto l'impresa esplica i propri affari e sono ubicati i principali uffici amministrativo-gestionali.

Le variabili di relazione sono fondamentali per la gestione del registro e per gli utilizzatori responsabili di indagini statistiche, al fine di collegare correttamente insieme le unità, come ad esempio nel caso degli eventi di trasformazione societaria. Sono memorizzate nel registro sotto la semplice forma di collegamento tra due codici identificativi, uno per ogni unità della relazione, ed ogni *link* include la data in cui la relazione ha avuto inizio ed eventualmente la data in cui si è interrotta.

Tutte le variabili relative alle unità statistiche del registro possono essere direttamente osservabili, nel caso in cui siano disponibili da fonte amministrativa con la qualità desiderata, o possono

²⁶ Per una trattazione esaustiva delle variabili che misurano l'occupazione delle imprese si veda il Capitolo 2.

²⁷ "Una unità è considerata unità *residente* di un paese quando essa ha il suo centro di interesse economico nel territorio economico di tale paese – ossia quando esercita per un lungo periodo (un anno o più) attività economiche su tale territorio". Pertanto, sono considerate residenti anche le unità residenti fittizie, definite come "quelle parti di unità non residenti che hanno un centro di interesse economico prevalente (ossia che effettuano operazioni economiche per un anno o più) nel territorio economico del paese o quelle proprietarie di terreni o fabbricati nel territorio economico del paese, per le sole operazioni connesse a tali terreni o fabbricati." (Sec2010 1.61-63).

essere derivate, suddividendo o raggruppando i dati delle unità di osservazione utilizzando metodi di stima statistica, o in alcuni casi possono essere fornite direttamente dal rispondente ad un'indagine, secondo il processo descritto nel Paragrafo 2.5.

2.2.4 Il trattamento dei cambiamenti

Il trattamento dei cambiamenti è inevitabilmente in relazione con quanto è riportato nel registro in un determinato momento, ossia le unità, le caratteristiche e le relazioni tra unità. In altre parole, il verificarsi di eventi nell'ambiente esterno che siano ritenuti rilevanti ha conseguenze sul registro in termini di modificazioni nell'esistenza delle unità, nelle relazioni e nelle loro caratteristiche nel tempo. Le variazioni registrate dovrebbero rispecchiare nella misura più fedele possibile le modifiche rilevate nella realtà osservabile, tenuto conto della domanda di informazioni, delle risorse disponibili, della necessità di coerenza tra le statistiche congiunturali e strutturali e degli effetti della rilevazione di informazioni sull'onere di risposta per le imprese.

In primo luogo i cambiamenti delle caratteristiche, o eventi non demografici, possono rispecchiare gli eventi della realtà, nel qual caso sono considerati degli aggiornamenti, o possono modificare delle informazioni errate, nel qual caso sono considerati correzioni. In questo contesto, il termine "errato" si riferisce a dati che non sono conformi alla visione della realtà convenuta con gli utenti quale parametro di riferimento per il registro e la loro individuazione e trattamento costituisce lo scopo del piano di controllo e correzione del registro descritto nel Paragrafo 2.7.

Riguardo solo al trattamento del primo tipo di cambiamenti nelle caratteristiche, si è scelto un approccio che tenga conto della qualità del registro in funzione delle esigenze degli utenti, seguito anche nel determinare la natura, la frequenza e la periodicità delle modifiche apportate alle caratteristiche, poiché non ha molto senso spendere risorse ed energie per aggiornare caratteristiche che non rivestono alcun interesse per gli utenti. Pertanto si è tenuto conto soprattutto dell'uso del registro come strumento per la preparazione, il coordinamento delle indagini ed il loro riporto all'universo, ma anche come fonte di informazioni per l'analisi statistica della popolazione delle imprese e della sua demografia.

Le caratteristiche non demografiche riguardano le variabili di identificazione e quelle di stratificazione, con differenze nelle esigenze di aggiornamento per gli utenti. I cambiamenti delle variabili di identificazione sono in genere registrati subito dopo essere stati comunicati dalla fonte amministrativa, senza necessità di controllo ulteriore. Per le variabili di stratificazione, principalmente attività, dimensione e ubicazione geografica, poiché possono interessare una particolare indagine, è importante decidere se e quando effettuare il loro aggiornamento; in alcuni casi infatti si sceglie di mantenerle invariate per un certo periodo in modo tale da ridurre l'impatto dei cambiamenti ai fini di assicurare coerenza nei risultati.

Il cambiamento della forma giuridica, essendo utilizzata anche come variabile di stratificazione, è tenuta sotto controllo soprattutto quando le modifiche comportano una uscita dal (o una entrata nel) campo di osservazione del registro, influenzando in questo modo l'appartenenza dell'unità alla popolazione di riferimento, in particolare quando si tratta di unità di grandi dimensioni. Ne è un esempio il trasferimento nel settore privato di una determinata attività esercitata precedentemente nel settore pubblico, ad esempio in settori quali la sanità e l'istruzione. Spesso alcune modifiche di forma giuridica sono accompagnate da un cambiamento della ragione sociale e questo consente di effettuare un controllo incrociato tra le due variabili, facilitando l'individuazione dei problemi di qualità²⁸.

²⁸ Se, per esempio, un'impresa senza personalità giuridica si trasforma in società per azioni, è normale attendersi l'aggiunta, al suo nome, della sigla S.p.a. che indichi tale trasformazione.

La segnalazione di un eventuale mutamento della forma giuridica di un'impresa proviene prevalentemente da fonti amministrative, nelle quali si registra il cambiamento del tipo di organizzazione sociale, che determina solo una variazione della "veste giuridica" di una impresa esistente, rimanendo invariati gli altri aspetti (patrimonio, soggetto economico, attività, rapporti con i terzi). Il cambiamento nell'ordinamento corporativo si ottiene con la modifica dell'atto costitutivo. Il regime può essere omogeneo, quando una società si trasforma in un'altra società dello stesso tipo (ad esempio "società di persone" in una "società di persone" e "società di capitali" in una "società di capitali") e disomogeneo quando la società si trasforma in un altro tipo. Nell'ipotesi che una "società di capitali" passi ad una "società di persone" la trasformazione è "regressiva", viceversa, se una "società di persone" passa ad una "società di capitali" la trasformazione è "progressiva". La riforma del diritto societario ha introdotto anche la trasformazione eterogenea disciplinata dall'art. 2500-*septies* all'art. 2500-*nonies* del Codice Civile. Con tale termine si intendono le operazioni di trasformazione che consentono a "società per azioni", "società in accomandita per azioni" e "società a responsabilità limitata" di dar vita a "consorzi", "società consortili", "società cooperative", "comunioni d'azienda", "associazioni non riconosciute" e "fondazioni"; e, similmente, le operazioni inverse che consentono di trasformare "consorzi", "società consortili", "società cooperative", "comunioni d'azienda", "associazioni non riconosciute" e "fondazioni" in "società di capitali".

Le informazioni relative ai cambiamenti che interessano le diverse attività economiche non sempre determinano una registrazione della modifica del codice di attività principale o secondaria nel registro. Tali cambiamenti possono essere improvvisi, a causa, per esempio, di un cambiamento della politica gestionale dell'impresa interessata oppure gradualmente, nel caso in cui l'importanza relativa delle singole attività svolte cambi nel tempo. In questi casi, le note esplicative della Nace Rev. 1.1 prevedono l'applicazione di una norma di stabilità secondo la quale "*prima di modificare la classificazione, l'attività secondaria dovrebbe superare per almeno due anni l'attività per la quale l'unità è stata classificata*". Questa norma consente di ridurre quelle fluttuazioni che rappresentano dei semplici fenomeni temporanei o che costituiscono degli artifici amministrativi.

Nel registro esiste un sistema di referenziazione delle modificazioni. Nell'aggiornare le variabili sono registrate sempre anche le informazioni relative alla data del cambiamento e alla fonte della nuova informazione, utili ai fini della valutazione della qualità delle variabili e per il controllo del processo di cambiamento. Le caratteristiche possono essere aggiornate utilizzando fonti amministrative, fonti statistiche come le indagini o una combinazione di entrambe. La politica di aggiornamento è determinata in gran parte dalla disponibilità di dati qualitativamente idonei. In genere, il tipo di politica adottata dipende dalle dimensioni delle unità, sia perché è più probabile che nei campioni di indagine vengano incluse le grandi unità piuttosto che le piccole unità, ma anche perché le grandi unità hanno una rilevanza economica maggiore.

Sicuramente il trattamento dei cambiamenti in conseguenza di eventi demografici assume una rilevanza ancora più ampia, in quanto la loro individuazione ha un'incidenza sull'esistenza delle unità e sulle relazioni tra queste. Un evento può produrre effetti su un unico tipo di unità statistica, ma potrebbe avere un'incidenza contemporaneamente su tipi diversi di unità statistiche e sulle loro relazioni, determinando ad esempio la costituzione di gruppi di imprese. Può inoltre influenzare le caratteristiche delle unità, quali l'attività economica principale o la dimensione.

Seguendo le raccomandazioni Eurostat sono state adottate delle soluzioni di tipo pragmatico per la collocazione temporale degli eventi e dei movimenti nel registro. In realtà i cambiamenti si verificano gradualmente, ad esempio l'avvio di un'impresa o la fusione di varie imprese può richiedere del tempo. Nel registro ogni cambiamento è "datato" in modo tale che si possa essere in grado, ad esempio, di sapere in quale momento una nuova impresa, che attraversa un lungo processo prima di iniziare l'attività, è considerata esistente. Inoltre, almeno per le imprese di grandi

dimensione, è possibile che in un singolo anno avvengano più eventi, nel qual caso sono tutti presi in considerazione.

La soluzione adottata al fine di garantire la consistenza delle informazioni statistiche prodotte è quella di registrare i movimenti nel corso di un anno parallelamente agli eventi reali, ossia continuamente, inserendo tali correzioni e/o aggiornamenti solo nella versione *live* o *running* del registro, il database relazionale denominato “Asia corrente”, in continua evoluzione e aggiornamento. Diversamente per l'estrazione dei campioni e per il riporto all'universo delle informazioni campionarie delle indagini strutturali e congiunturali si produce una edizione del registro ad una data prefissata, solitamente alla fine del primo trimestre di ogni anno solare ($t+15$ mesi), da utilizzare come popolazione di riferimento durante tutto l'anno anche per indagini aventi lo stesso periodo di riferimento ma realizzate in periodi diversi.

I principali eventi demografici, catalogati in base a quali sono gli effetti che determinano sulle imprese, devono essere tradotti in movimentazioni nel registro.

Cambiamenti inerenti all'esistenza (eventi che implicano l'esistenza di una sola impresa dopo l'evento e di nessuna impresa prima dell'evento, oppure di un'unica impresa prima dell'evento e di nessuna impresa successivamente ad esso):

- creazione;
- cessazione.

In relazione con le creazioni e le cessazioni di imprese è importante anche tener conto del concetto di riattivazione. Il caso di un'impresa che diventa inattiva ma riprende nuovamente la propria attività entro 24 mesi si configura, per convenzione, come una riattivazione e non quindi come una cessazione con conseguente creazione di un'impresa. Nel registro la creazione e la cessazione corrispondono rispettivamente alla creazione di un nuovo numero di identificazione o alla cancellazione di uno esistente.

Concentrazioni (eventi che implicano l'esistenza di più di un'impresa prima dell'evento e di una sola impresa dopo l'evento):

- fusioni;
- acquisizioni.

Le imprese possono integrarsi fino a ridurre il numero delle imprese esistenti e si verifica un fenomeno di concentrazione. L'integrazione consente di unificare compagini societarie e patrimoni di diverse società, che prima dell'operazione erano distinti. I vantaggi che ne derivano sono il miglioramento della capacità di compenetrazione del mercato e un maggior sfruttamento dei fattori produttivi. L'integrazione può assumere diverse forme: due imprese possono integrarsi completamente, perdendo la propria identità perché si fondono fino a non essere più riconoscibili nella nuova organizzazione, oppure una delle imprese può restare inalterata. In quest'ultimo caso l'altra impresa è generalmente molto più piccola: è semplicemente assorbita dall'impresa più grande che perlopiù non subisce variazioni. Se entrambe le imprese perdono la propria identità l'evento è denominato fusione, se una delle imprese conserva la propria identità si tratta di un'acquisizione. Nel primo caso più società (cosiddette fuse) trasferiscono l'intero patrimonio attivo e passivo in capo a una nuova società che si costituisce *ex novo*; nel secondo caso, una o più società (cosiddette incorporate) trasferiscono l'intero patrimonio attivo e passivo in capo a una società già esistente che le incorpora (cosiddetta incorporante).

In seguito ad un'acquisizione alcune caratteristiche dell'impresa che conserva la sua identità potrebbero modificarsi. Ad esempio, l'impresa potrebbe essere inclusa in una classe di dimensione diversa o svolgere un'attività economica principale differente. Tali effetti sono molto importanti e

per tenerne conto in maniera corretta il registro conserva i dati di tutti gli eventi in un registro satellite, il database degli eventi societari, a disposizione dei responsabili delle principali indagini economiche e diffuso insieme al registro²⁹.

Riguardo gli effetti che queste tipologie di eventi hanno sul registro, nel caso di una fusione tutti i codici identificativi delle imprese esistenti prima dell'evento sono cancellati ed è creato un codice identificativo dell'impresa di nuova costituzione. Nel caso di un'acquisizione l'impresa che effettua l'incorporazione conserva il proprio numero di identificazione e nel registro non si ha pertanto alcun inserimento bensì soltanto la cancellazione dei codici identificativi delle imprese incorporate. Dal punto di vista teorico la data in cui si ritiene abbia effetto la variazione è la data in cui le imprese oggetto della fusione o dell'acquisizione perdono la propria autonomia decisionale, citata nella definizione di impresa.

Smembramenti (eventi che implicano l'esistenza di una sola impresa prima dell'evento e di più di una impresa dopo l'evento):

- scissioni;
- scorpori.

Esigenze di riorganizzazione e/o di ristrutturazione aziendale possono indurre una unità (cosiddetta scindenda) ad assegnare l'intero suo patrimonio ad una o più società preesistenti o di nuova costituzione, in tal caso si assiste alla frammentazione di un unico patrimonio in più parti dello stesso. Lo smembramento può essere totale o parziale, nel primo caso la società trasferisce l'intero suo patrimonio a una o più società, estinguendosi; nel secondo caso si assiste, invece, al trasferimento di solo una parte del patrimonio della società scindenda a favore di una o più società, dunque, la società scissa rimane in vita con un patrimonio ridotto.

I due eventi sono dunque speculari a quelli di concentrazione: l'evento contrario a una fusione è la scissione e l'evento contrario a un'acquisizione è lo scorporo. Nel primo caso l'impresa è suddivisa in maniera tale che nessuna delle nuove imprese mantiene l'identità dell'impresa originaria. Nel caso di uno scorporo la nuova impresa o le nuove imprese sono generalmente più piccole e l'identità dell'impresa originaria è conservata dall'impresa più grande.

Poiché le scissioni e gli scorpori rappresentano gli eventi speculari rispettivamente delle fusioni e delle acquisizioni, la traduzione di questi eventi in movimenti del registro è analoga. In caso di scissione sono creati i codici identificativi di tutte le imprese dopo l'evento ed è cancellato quello dell'impresa originaria. In caso di scorporo un nuovo codice identificativo è attribuito soltanto all'impresa o alle imprese scorporate.

L'aggiornamento degli eventi si basa sulle informazioni ricevute da diverse fonti. Tali segnali riguardano modifiche amministrative e, in funzione della qualità, della rilevanza e della pertinenza dell'informazione amministrativa, può risultare necessario intervenire per aggiornare il registro. La politica adottata per decidere se aggiornare i dati si basa su criteri di rilevanza e quindi risulta diversa a seconda che le imprese siano di piccole, medie o grandi dimensioni³⁰.

2.3 Il processo di aggiornamento delle unità nel registro Asia

L'insieme delle unità statistiche che costituisce la base del registro Asia per l'anno di riferimento t è aggiornato annualmente, utilizzando una metodologia che fa uso di fonti amministrative e di fonti

²⁹ Per una trattazione più estesa del registro satellite si veda il Paragrafo 2.5.3.

³⁰ Per le imprese più rilevanti del registro l'individuazione della presenza di un evento e la stima dei caratteri delle imprese in esso coinvolte sono parte rilevante delle operazioni di aggiornamento del registro descritte nel Paragrafo 2.5.

statistiche. Ogni anno la base amministrativa del registro, ossia l'insieme delle unità giuridiche, è costruita facendo uso di tutte le fonti necessarie alla sua realizzazione. La maggior parte delle fonti amministrative sono in comune al processo di aggiornamento di tutte le unità statistiche in quanto utilizzate per l'identificazione delle unità, altre invece sono più specifiche e utilizzate per l'assegnazione delle variabili peculiari di ciascuna unità.

Il processo di aggiornamento delle *unità giuridiche* e delle *imprese*, con riferimento all'anno t , ha inizio nella seconda metà dell'anno $t+1$, quando si rendono disponibili le forniture annuali di dati provenienti dai principali archivi amministrativi, utilizzate per la costruzione della base delle unità giuridiche. Le informazioni, dopo essere state sottoposte a un processo di normalizzazione e standardizzazione, che trasforma le unità e i caratteri amministrativi in unità e variabili statistiche, sono integrate fra loro. L'insieme delle unità che ne deriva costituisce la base del registro Asia per l'anno di riferimento t . Sulle unità della base sono successivamente stimate le variabili strutturali ed identificative, utilizzate per l'identificazione delle imprese e dei propri caratteri. In particolare, utilizzando tutte le informazioni disponibili, per la totalità delle unità sono stimati i valori delle variabili attività economica, forma giuridica, i caratteri identificativi e di localizzazione, lo stato di attività e tutte le variabili che misurano l'occupazione. Tali informazioni sono integrate con le fonti statistiche e successivamente sottoposte ad un processo di controllo di qualità, la cui fase finale consiste nell'aggiornamento del sistema informativo di Asia, un database relazionale che include le informazioni storiche, su tutte le unità giuridiche e le imprese, a partire dal 1998 e contiene informazioni sulle modifiche subite dalle unità economiche nel corso degli anni (il cosiddetto "Asia corrente").

Le unità di piccola dimensione, corrispondenti a circa il 95 per cento delle imprese del registro, sono aggiornate annualmente mediante il solo processo di integrazione di archivi amministrativi, mentre le unità più rilevanti sono monitorate durante tutto l'anno da esperti di settore, utilizzando tutte le fonti disponibili amministrative e statistiche, e anche attraverso l'attività di *profiling* delle unità complesse³¹.

Per illustrare meglio le diverse operazioni necessarie per la realizzazione dell'insieme delle unità giuridiche e delle imprese dotate dei relativi caratteri statistici con riferimento ad un generico anno t , sono necessarie quattro macro fasi:

- 1) integrazione delle fonti amministrative, identificazione delle unità giuridiche e creazione della base dell'anno t ³²;
- 2) imputazione degli attributi principali e individuazione delle unità statistiche attive nell'anno t da sola fonte amministrativa³³;
- 3) aggiornamento da fonti statistiche e integrazione con la base delle unità giuridiche e delle imprese³⁴;
- 4) controllo e correzione dei dati, creazione della lista di imprese attive nell'anno t per la sua diffusione; aggiornamento di "Asia corrente" con l'anno t ³⁵.

³¹ L'attività di *profiling* è comunemente definita come "un metodo che serve ad analizzare la struttura giuridica, operativa e contabile di un gruppo di imprese a livello nazionale e internazionale, allo scopo di definire le unità statistiche presenti in tale gruppo, i legami che intercorrono fra loro e le strutture più efficienti per la raccolta dei dati statistici" (Eurostat 2010, *Business Register Recommendations Manual*). Il *profiling* si configura, dunque, come uno strumento idoneo per il trattamento di situazioni complesse in cui l'unità statistica "impresa" dovrebbe essere identificata correttamente come una combinazione di una o più unità giuridiche.

³² La prima fase del processo è descritta nel Paragrafo 2.4.

³³ La fase di stima dei caratteri è illustrata nel Paragrafo 2.5.

³⁴ Per le procedure di aggiornamento da fonti statistiche utilizzate negli ultimi anni si rimanda al Paragrafo 2.6.

³⁵ Il Paragrafo 2.7 chiarirà quale è il piano di controllo e correzione scelto per assicurare un livello opportuno di qualità.

In occasione del Censimento 2011 sono state introdotte importanti innovazioni in questo processo di aggiornamento e in particolare nella stima dei caratteri delle imprese – dal punto di vista definitorio e metodologico – utili ad aumentare la qualità ed ampliare la quantità delle informazioni diffuse, con un significativo miglioramento nella futura tempistica di diffusione dei dati. La restante parte del fascicolo sarà dedicata proprio ad illustrare questo processo e le novità che sono state apportate in tale occasione.

Il processo di aggiornamento dei *gruppi d'impresa* si svolge parallelamente a quello delle altre unità statistiche, una volta completato il processo di individuazione delle unità giuridiche della base dell'anno t , e si conclude con la realizzazione di una lista annuale di oltre 90 mila gruppi, diffusa autonomamente e in tempi diversi rispetto alle altre unità³⁶. Alla descrizione del processo è dedicato uno spazio minore in questo fascicolo, per la sola ragione di un suo minore utilizzo nella diffusione dei dati censuari rispetto alle altre unità. Si è fatto uso, infatti, del solo attributo "appartenenza ad un gruppo" per tutte le imprese nel campo di osservazione del Censimento, ed è stato utilizzato sia come variabile di classificazione dei dati diffusi sulle imprese, sia nella fase di creazione della lista per la Rilevazione sulle unità complesse³⁷ e per la successiva analisi delle informazioni rilevate con tale indagine.

Nondimeno è rilevante sottolineare come, nel sistema dei registri Asia, il Registro dei gruppi d'impresa, in linea con quanto stabilito dal Regolamento Comunitario n. 177/2008, segue i criteri operativi stabiliti nel *Recommendations Manual*³⁸. Nella metodologia indicata sono forniti concetti base e criteri operativi applicabili a prescindere dalle specifiche fonti disponibili e sufficienti a garantire comunque la confrontabilità internazionale delle statistiche prodotte. In particolare stabilisce che il legame di associazione, indicato dal Regolamento n. 696/1993 nella definizione di gruppo, debba concretizzarsi a fini operativi in un effettivo legame di controllo, esercitato, in via diretta o indiretta, da parte del vertice del gruppo su tutte le altre imprese che lo compongono.

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

unità partecipante → unità partecipata, quota di partecipazione

L'utilizzo di una base informativa così ampia mira a garantire la copertura totale del fenomeno sia dei gruppi a controllo privato sia dei gruppi a controllo pubblico³⁹.

La metodologia di integrazione delle diverse fonti amministrative, partendo dai dati elementari sulla struttura delle partecipazioni dirette di tutte le società, individua i legami di controllo, esercitati sia

³⁶ La realizzazione del Registro gruppi d'impresa nel sistema informativo di Asia è stata curata da Augusto Pallante e Franco Tulli, Istat.

³⁷ Cfr. Parte Seconda, Capitolo 3 del presente fascicolo.

³⁸ Regolamento (CE) n. 177/2008 del Parlamento Europeo e del Consiglio del 20 febbraio 2008 che istituisce un quadro comune per i registri di imprese utilizzati a fini statistici e abroga il Regolamento (CEE) n. 2186/93 del Consiglio. Eurostat 2010, *Business Register Recommendations Manual*.

³⁹ Le fonti amministrative utilizzate nel processo di aggiornamento dei gruppi sono: 1) le dichiarazioni sulla struttura delle partecipazioni rilevanti rese alla Commissione nazionale per le società e la Borsa (Consob); 2) le dichiarazioni degli elenchi dei soci delle società di capitale iscritti al Registro delle imprese, gestito dalle Camere di commercio; 3) le informazioni desumibili dai bilanci civilistici e consolidati delle società di capitale; 4) le dichiarazioni dell'elenco dei consorzi e delle società a totale o parziale partecipazione da parte delle amministrazioni pubbliche registrate nella banca dati CONSOC; 5) le dichiarazioni relative alle partecipazioni detenute dalle amministrazioni pubbliche al Dipartimento del Tesoro; 6) le dichiarazioni relative alle partecipazioni detenute da regioni, province e comuni alla Corte dei conti; 7) il Sistema Conti Pubblici Territoriali (Cpt) che rileva circa 4 mila soggetti (enti dipendenti, aziende, consorzi e società) partecipati dagli enti della Pubblica Amministrazione.

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

Il punto di forza di tale metodologia è costituito dalla struttura informativa di base, costruita in modo semplice e uniforme per tutte le fonti: codice fiscale dell'unità sussidiaria, codice fiscale dell'unità controllante prossima (se residenti, altrimenti la denominazione) e la quota di partecipazione. Alcune fonti forniscono informazioni sulle quote di tipo quantitativo, altre di tipo qualitativo garantendo così l'individuazione sia del controllo di diritto sia del controllo di fatto. La qualità del registro viene inoltre garantita dai controlli interattivi, effettuati su ogni singolo gruppo da revisori esperti, su un campione di gruppi selezionati ogni anno sulla base di criteri dipendenti dalla dimensione dei gruppi e dalla loro globalizzazione. In particolare la mappatura dei gruppi a controllo delle regioni viene effettuata ogni anno, al fine di individuare i soggetti controllati per i quali il controllo si esplica in forme diverse dalla partecipazione maggioritaria al capitale della società, ossia in forza di leggi o regolamenti che danno il diritto di determinare la politica dell'unità o di nominarne gli amministratori.

Nel caso di gruppi multinazionali a controllo ultimo estero la catena di controllo ricostruibile attraverso questa metodologia si interrompe alla prima controllante prossima non residente in quanto le fonti amministrative disponibili rilevano le controllate residenti. In questo caso si parla di gruppi multinazionali troncati presenti nel registro. Queste strutture costituiscono la popolazione di riferimento della Rilevazione sulle attività internazionali delle imprese (*Inward FATS*) svolta dall'Istat a partire da un campione selezionato dall'archivio, che rileva l'informazione sul controllo ultimo estero. Tale informazione, di natura statistica, viene utilizzata per aggiornare – relativamente alle imprese comprese nel campione e rispondenti alla rilevazione – il registro stesso per quanto riguarda il vertice ultimo dei gruppi e la loro nazionalità estera. Le relazioni con imprese estere vengono inoltre ottenute grazie all'integrazione dei dati del Registro europeo dei gruppi multinazionali (*EuroGroups Register* - EGR) alimentato anche dal Registro dei gruppi d'impresa Istat che rappresenta la fonte autentica per le relazioni tra imprese residenti nel territorio italiano.

Infine, il processo di aggiornamento delle *unità locali* con riferimento all'anno t ha inizio una volta individuate le unità giuridiche e le imprese, utilizzando queste come base di partenza alla quale sono integrate altre fonti amministrative e statistiche specifiche per tali unità. Alla sua descrizione è dedicato interamente il Capitolo 4 della Parte Prima del presente fascicolo, per l'importanza che riveste nella diffusione dei dati censuari.

2.4 La prima fase del processo di aggiornamento delle imprese: l'integrazione delle fonti amministrative e l'identificazione delle unità statistiche

2.4.1 L'uso delle fonti amministrative per il registro

Un aspetto che caratterizza fortemente il processo di realizzazione del registro Asia è l'utilizzo delle fonti amministrative. Il tipo di prodotto che deve essere realizzato, infatti, rende imprescindibile l'utilizzo di tali fonti come base informativa primaria. Anche da questo punto di vista la realizzazione di Asia assume un rilievo particolare in quanto è una delle prime esperienze compiute di uso di dati amministrativi integrati per fini statistici, determinando anche la necessità di definire metodologie innovative per consentire un corretto utilizzo delle informazioni rese disponibili da atti amministrativi.

Sono molti i vantaggi nell'utilizzare le fonti amministrative. Il principale è la riduzione nei costi rispetto alle indagini, soprattutto quando realizzate sotto forma di censimenti. Le fonti amministrative sono spesso gratuite, soprattutto se appartengono al settore pubblico, e anche nel caso in cui sia previsto il pagamento di un corrispettivo per la fornitura dei dati amministrativi, il costo è spesso inferiore a quello della raccolta delle stesse informazioni tramite indagini; le elaborazioni richiedono meno personale e non vi è alcuna necessità di sollecitare le risposte. Il loro uso consente anche di ridurre l'onere di risposta gravante sulle imprese. Mentre le imprese comprendono, di solito, i motivi della necessità di trasmettere i dati ad esempio ai fini fiscali, quello delle indagini è spesso considerato dalle imprese come un onere supplementare non necessario, ancor di più avendo già fornito la stessa informazione ad altre amministrazioni pubbliche. Un altro vantaggio è costituito dal fatto che le fonti amministrative offrono spesso una copertura completa, o quasi completa, della popolazione di riferimento, riducendo così notevolmente le mancate risposte e fornendo stime più accurate e dettagliate delle varie sottopopolazioni. Grazie ai dati amministrativi l'aggiornamento di alcune variabili è migliore, ricevendo solitamente informazioni recenti sull'anagrafica delle unità o sulla loro ubicazione. Oltre a migliorare la tempestività delle variabili, l'uso dei dati provenienti dalle fonti amministrative può, in alcuni casi, migliorare la tempestività delle statistiche ricavate dai dati del registro, direttamente o mediante campioni. Questo perché, in genere, le indagini statistiche richiedono tempo per l'elaborazione e la gestione dei formulari, per l'analisi della popolazione, per l'ottimizzazione dei campioni, eccetera. È il caso, in particolare, delle raccolte di dati annuali o *ad hoc*. Di conseguenza, l'accesso a un'ideale fonte amministrativa tramite il Registro delle imprese può essere una soluzione più efficace. Va sottolineato, tuttavia, che, in certi casi, l'uso delle fonti amministrative può anche comportare una riduzione della tempestività, soprattutto per quanto riguarda gli indicatori congiunturali.

L'uso delle fonti amministrative pone al contempo alcuni problemi che rendono necessario individuare opportune metodologie proprio per il loro trattamento a fini statistici. Oltre a differire nella definizione delle unità, le fonti amministrative e il registro presentano delle differenze anche a livello di definizione delle variabili. I dati presenti nelle fonti amministrative vengono in genere raccolti per uno specifico scopo amministrativo ed è comprensibile che le esigenze e le priorità inerenti a tale scopo siano diverse da quelle del sistema statistico. Analogamente, i sistemi di classificazione utilizzati nelle fonti amministrative spesso differiscono da quelli usati in ambito statistico, oppure sono applicati in modo diverso a seconda degli obiettivi primari della fonte amministrativa. Nel primo caso è in genere necessario elaborare delle tabelle di corrispondenza tra i codici utilizzati nella classificazione amministrativa e quelli utilizzati nel registro statistico. Le corrispondenze in questione non sono sempre di tipo "uno a uno", ma anche "uno a molti" o "molti a molti" ed è quindi necessario ricorrere ad assegnazioni di natura probabilistica. Ciò consente da un lato di avere una codifica accurata a livello aggregato, ma non necessariamente lo è anche a livello di singole unità del registro.

Un altro problema comune che si incontra nell'uso dei dati da fonte amministrativa riguarda la tempestività. Può succedere che i dati non siano disponibili in tempo per soddisfare le esigenze del registro o che facciano riferimento a un periodo che non coincide con quello necessario a fini statistici. Si registra in genere un certo *lag* temporale tra il verificarsi di un evento nella realtà e la sua registrazione nella fonte amministrativa, cui si aggiunge un ulteriore intervallo tra detta registrazione e la disponibilità dei dati per il registro statistico delle imprese. È noto che i *lag* temporali relativi alle creazioni e alle cessazioni di imprese rappresentano una delle maggiori fonti degli errori contenuti nel registro e la possibilità di misurarli consente di tenerne conto in qualsiasi statistica basata sui dati del registro.

L'assoluta dipendenza del registro dalle fonti amministrative ne rappresenta forse la maggiore criticità. Le fonti amministrative del settore pubblico sono in genere finalizzate alla riscossione delle imposte e al monitoraggio delle politiche pubbliche. Ne deriva che sono soggette ai cambiamenti normativi. Un eventuale cambiamento di politica può influire sulle fonti amministrative a livello di copertura, definizioni, soglie, eccetera, o persino comportare la totale soppressione delle fonti stesse. Si tratta di cambiamenti che in passato si sono verificati improvvisamente o con un preavviso insufficiente. Si è cercato in qualche modo di gestire tale rischio mediante contatti regolari con i responsabili delle fonti amministrative per cercare di anticipare eventuali cambiamenti, ma non è possibile eliminare del tutto le incertezze.

Il problema della coerenza che accompagna l'uso di più fonti è invece più facile da gestire ed è stato risolto in maniera differente per ogni caratteristica del registro. I dati forniti da una determinata fonte possono, apparentemente, contrastare con quelli provenienti da un'altra fonte, come effetto delle differenze di definizioni, classificazioni e momento di registrazione o può essere dovuto semplicemente alla presenza di errori in una delle fonti. Per risolvere tali conflitti è stato necessario prevedere delle regole di priorità, decidendo qual è la fonte più affidabile per una determinata variabile, così da garantire che i dati forniti da una fonte ad alta priorità non vengano sostituiti dai dati provenienti da una fonte a bassa priorità. L'informazione su quale sia stata la fonte scelta per assegnare il valore di una variabile ad ogni unità del registro è sempre archiviato insieme agli altri metadati del registro.

Per tutti questi motivi, per utilizzare correttamente a fini statistici i dati che si rendono disponibili dalle fonti amministrative, è stato necessario sviluppare delle metodologie aventi l'obiettivo di trasformare un'informazione amministrativa in un dato statistico. In generale tali metodologie si distinguono in:

- funzioni di normalizzazione (o standardizzazione);
- funzioni di determinazione (o specificazione).

Le prime sono utilizzate nella prima macro fase del processo, quando sono integrate le fonti amministrative e identificate le unità giuridiche, mentre le funzioni di determinazione sono utilizzate successivamente nella fase di stima dei caratteri.

In generale, con normalizzazione si intende un qualsiasi processo che modifica un oggetto per renderlo più conforme a qualche criterio di regolarità⁴⁰. Una funzione di normalizzazione è una regola R_N che, applicata ad una singola entità o carattere amministrativo (E_a) la trasforma in entità o carattere statistico (E_s):

$$R_N: E_a \Rightarrow E_s$$

In altre parole è un'applicazione che cambia i valori $x_j \in X_j$ in valori $x_i \in X_i$.

Una regola di normalizzazione, generalmente di tipo deterministico, che trasforma concetti e classificazioni amministrative in concetti e classificazioni statistiche, può essere distinta in tre tipologie:

- regole di codifica: trasformano le codifiche in valori appartenenti alle classificazioni statistiche ufficiali. In particolare si utilizzano per l'attività economica principale svolta dalle

⁴⁰ Per approfondimenti cfr. Garofalo G. (2002) "To exploit Administrative sources: a framework of concepts", 16th Roundtable on Business Survey Frames, Lisbon, Portugal, 2002.

imprese⁴¹, per la forma giuridica e per la localizzazione sul territorio. Ogni ente amministrativo può avere diverse modalità di registrazione dei dati relativi alle unità economiche contenute nel proprio archivio, che sono peculiari per i propri scopi. Ad esempio l'Agenzia delle Entrate codifica i dati relativi alle forme giuridiche assunte dalle unità nell'archivio dell'Anagrafe tributaria utilizzando la "*tabella generale di classificazione natura giuridica*", che tratta separatamente i soggetti residenti e quelli non residenti ed è più orientata esplicitamente all'universo delle forme giuridiche disciplinate dal diritto privato mentre l'universo di quelle disciplinate dal diritto pubblico è sottorappresentato in quanto soffre di una eccessiva aggregazione prevedendo termini troppo generici. È stato possibile predisporre un collegamento con la classificazione adottata dall'Istat, che si articola nel dettaglio con il fine di prendere in considerazione tutte le unità legali, ma il raccordo tra i due elenchi risulta problematico a causa dei diversi obiettivi per i quali sono state adottati dai rispettivi enti.

Allo stesso modo sono state costruite delle regole per convertire i codici catastali usati nell'Anagrafe tributaria in codici NUTS (Nomenclatura delle Unità Territoriali per le Statistiche dell'Italia);

- regole di *link*: permettono di raggruppare le differenti unità registrate in una fonte in modo tale da determinare unità statistiche; così i diversi record elementari di unità amministrative possono essere uniti insieme per formare una unità legale; oppure le unità legali in una fonte possono essere unite in modo tale da formare le unità statistiche "impresa" e "unità locale". Ad esempio nel registro Inps le unità amministrative sono le posizioni contributive e utilizzando opportune chiavi identificative quali la matricola Inps e il codice fiscale è possibile ricostruire le unità giuridiche, che possono utilizzare anche diversi centri di costo per pagare i contributi ai propri dipendenti.

In mancanza di un codice di identificazione comune, per registrare il collegamento si può adottare un approccio probabilistico, come avveniva in passato per alcune fonti, i cui risultati vanno verificati attraverso controlli qualitativi che utilizzano variabili ausiliarie. Nel caso in cui queste variabili non siano coerenti potrebbero essere necessari ulteriori controlli, soprattutto per le unità più grandi. È utile, in generale, controllare periodicamente le unità non abbinata nel tentativo di stabilire ulteriori collegamenti o di determinare il motivo del mancato abbinamento, che potrebbe essere dovuto a differenze di momento di registrazione o di copertura tra la fonte amministrativa e il registro statistico delle imprese;

- regole di trasformazione: permettono di ottenere variabili statistiche da caratteri amministrativi. Si tratta di procedure che riguardano anche il pretrattamento delle fonti per adeguare i dati alle esigenze del registro statistico e comprendono, tra l'altro, l'applicazione delle definizioni statistiche relative alle unità e alle variabili. Sono regole di tipo complesso che non possono prescindere da uno studio iniziale della fonte e degli eventuali cambiamenti nel tempo, come nel caso delle procedure di trasformazione degli eventi amministrativi di fonte fiscale per la costruzione del database degli eventi societari oppure il complesso insieme di regole per l'utilizzo dei questionari degli Studi di settore ai fini dell'assegnazione dell'attività economica delle imprese.

⁴¹ Di particolare rilevanza era nel passato l'applicazione delle regole di codifica che, fino all'anno 2007, trasformavano l'attività economica dichiarata nelle fonti amministrative in valori appartenenti alla classificazione Ateco 2002, versione italiana della Nace Rev 1.1. A partire dal 1° gennaio 2008 infatti, con l'entrata in vigore della nuova Nace Rev. 2, si è pervenuti ad un'unica classificazione Ateco 2007 e per la prima volta il mondo della statistica ufficiale, il mondo fiscale e quello camerale adottano la stessa classificazione delle attività economiche.

Una funzione di determinazione è una regola che, associata alle determinazioni di uno o più caratteri individuati in un insieme di fonti, permette di stimare, con una data probabilità, una determinazione di un carattere statistico. Permette quindi di individuare o stimare un dato statistico a partire dall'insieme dei dati disponibili nelle diverse fonti utilizzate. Tale regola è sempre di tipo probabilistico e può essere di due tipologie:

- regole di scelta: tramite l'utilizzo di metodologie probabilistiche permettono di individuare, fra i differenti attributi presenti nelle varie fonti, l'attributo "ottimale". È questo il caso dell'attività economica prevalente delle imprese a partire dalle informazioni sui codici presenti nelle varie fonti amministrative;
- regole di stima: utilizzando in maniera strumentale le informazioni contenute in altri attributi delle unità presenti nelle fonti amministrative, permettono di stimare il valore di una variabile con un prefissato livello di errore. È questo il caso dello stato di attività delle imprese stimato a partire dai segnali di attività presenti nelle varie fonti, quali principalmente la presenza di dipendenti e il volume di affari dichiarato.

L'uso generalizzato di informazioni amministrative per fini statistici impatta in modo considerevole anche sull'organizzazione dell'Istat. Il loro utilizzo a fini statistici e la numerosità fortemente aumentata delle fonti che l'Istituto acquisisce ha spinto l'Istat verso l'accentramento di alcune funzioni per l'acquisizione, l'archiviazione, l'integrazione e la valutazione della qualità dei dati amministrativi, creando un sistema interconnesso di fonti amministrative, realizzato proprio a partire dall'esperienza censuaria, per massimizzare il vantaggio derivante dall'aver a disposizione una quantità di informazioni così ampia.

Per garantire la trasversalità e la condivisione delle fonti amministrative in tutti i processi produttivi, garantendone il loro utilizzo in maniera sistematica, è stata istituita una specifica Direzione centrale che costituisce il punto di riferimento di tutte le attività connesse agli archivi amministrativi gestiti dall'Istituto, per regolare il rapporto con gli enti, l'accesso e l'uso condiviso degli archivi amministrativi e per assicurare l'adozione di standard comuni di trattamento dei dati, al fine di mantenere la coerenza e la qualità complessiva della produzione statistica.

A partire dal 2013 nel nuovo Sistema integrato di microdati (Sim) si realizza l'integrazione di tutti i microdati acquisiti dalle diverse fonti amministrative, non più solo quelle utilizzate dai registri economici. Di conseguenza si allarga l'ambito nel quale si concretizza l'attribuzione dei codici identificativi univoci per le unità economiche, ampliando tale funzione anche agli individui, i luoghi e le relazioni tra unità economiche e individui. I microdati di ciascuna fonte amministrativa inserita nel sistema sono pertanto identificati in maniera univoca e risultano essere automaticamente integrati o, comunque, integrabili non solo con le altre fonti in cui sono presenti gli stessi oggetti di analisi ma anche con tutti i possibili output statistici che usano il Sim come punto di partenza per i loro processi. Il sistema dei registri Asia diventa dunque un "utente" del Sim come uno dei processi di produzione statistica dell'Istat di cui il Sim è infrastruttura di base.

2.4.2 Il processo di integrazione delle fonti amministrative e l'identificazione delle unità

L'obiettivo della prima macro fase è l'integrazione tra gli archivi amministrativi e la conseguente creazione dei gruppi e dei grappoli di unità elementari che formano le unità giuridiche. In sintesi, dopo aver sottoposto le informazioni ad un processo di normalizzazione, che trasforma le unità e i caratteri amministrativi in unità e variabili statistiche, queste sono integrate fra loro in modo tale da formare l'insieme delle unità giuridiche che costituisce la base amministrativa del registro Asia per l'anno di riferimento *t*.

Le attività necessarie per la prima macro fase del processo di aggiornamento del registro sono svolte su un insieme di fonti amministrative in continua evoluzione. Una delle attività interne al processo di Asia è, infatti, l'analisi continua sulle caratteristiche e sulla disponibilità di fonti informative. Questa si sostanzia in due processi:

- verifica della continuità, in termini di tempistica, variabili, classificazioni e definizioni utilizzate, delle fonti presenti nel processo per l'anno $t-1$. La presenza di una qualche discontinuità in una fonte comporta la revisione delle funzioni di normalizzazione. Tale revisione può essere di diversa rilevanza e può anche determinare la necessità di eliminare la fonte in analisi dal processo, se e dove le variazioni introdotte rendessero praticamente impossibile la traduzione delle informazioni da amministrative a statistiche;
- ricerca di nuove fonti che possono migliorare le stime delle variabili statistiche già presenti in Asia o implementare l'insieme di informazioni rese disponibili per descrivere, a livello micro, l'universo delle imprese. Questa attività ha avuto nuovo impulso una volta messa a regime la produzione del nucleo base delle informazioni di Asia-imprese e ha costituito il presupposto per la realizzazione di nuovi prodotti, i più significativi dei quali sono il registro delle unità locali (Asia-unità locali), che di fatto fornisce annualmente le informazioni prima disponibili solo con i censimenti generali, e il registro Asia-occupazione.

Le principali fonti amministrative, gestite da enti pubblici o da società private, oggetto della prima macro fase di integrazione sono:

- gli archivi gestiti dall'Agenzia delle Entrate per il Ministero dell'Economia e delle Finanze, quali l'Anagrafe tributaria delle persone giuridiche e delle persone fisiche con partita Iva, le dichiarazioni annuali delle imposte indirette, le dichiarazioni dell'imposta regionale sulle attività produttive (Irap), gli Studi di settore, i dati del modello Unico, quadro RH;
- i registri delle imprese delle Camere di commercio, industria, artigianato e agricoltura e gli archivi collegati dei soci delle società di capitale e delle "persone" con cariche sociali;
- gli archivi dell'Istituto Nazionale della Previdenza Sociale: le denunce retributive mensili EMens per gli occupati dipendenti; le dichiarazioni trimestrali della manodopera agricola (modello Dmag); la Cassa integrazione a pagamento diretto; le posizioni contributive degli imprenditori artigiani e commercianti; la Gestione separata parasubordinati; l'archivio delle denunce contributive lavoratori dello sport e dello spettacolo (ex Enpals);
- l'archivio dell'Inail, delle assicurazioni per i lavoratori con contratto di somministrazione;
- l'archivio delle utenze telefoniche gestito dalla Consodata S.p.a. (Gruppo SEAT-Pagine Gialle);
- l'archivio dei bilanci consolidati e di esercizio delle Camere di commercio;
- l'archivio degli istituti di credito gestito dalla Banca d'Italia;
- l'archivio delle società di assicurazioni gestito dall'Isvap.

L'Anagrafe tributaria e il Registro delle imprese sono le uniche fonti utilizzate per l'identificazione delle unità giuridiche presenti nella base amministrativa valida per l'anno t , tra le quali identificare le unità statistiche "imprese". Tutte le altre fonti sono utilizzate per individuare le unità giuridiche del registro e, in maniera esclusiva o in concomitanza con le precedenti, per la stima dei caratteri o per il controllo di particolari sottoinsiemi.

All'interno della prima macro fase possono essere individuate due attività eseguite in modo parallelo:

- 1) integrazione concettuale: acquisizione delle fonti e loro normalizzazione o standardizzazione;

- 2) integrazione fisica: *link* intra-archivio, *link* inter-archivi, identificazione delle unità giuridiche appartenenti alla base amministrativa e identificazione delle imprese.

L'obiettivo dell'integrazione concettuale è l'acquisizione delle forniture annuali o semestrali degli archivi amministrativi e un loro primo trattamento per renderli utilizzabili come fonti di input del processo. Le informazioni sono acquisite e conservate in apposite tabelle di un database relazionale in Oracle. Tutte le operazioni di pulizia dei dati grezzi, l'individuazione dei dati mancanti o di quelli non ammissibili, così come l'applicazione delle regole di normalizzazione, producono informazioni di tipo statistico che devono sempre essere conservate in aggiunta a quelle originali, così da garantire che ci sia una registrazione di ciò che è stato ricevuto indipendentemente dalle modifiche apportate durante la successiva lavorazione. Oltre ad essere importante ai fini di un *audit* o per eventuali comunicazioni con i fornitori in caso di problemi sulla fonte, la conservazione delle informazioni consente anche di poter replicare il processo in caso si verificassero dei problemi anche di natura informatica e soprattutto permette di poter utilizzare i dati amministrativi non trattati statisticamente come output informativo da diffondere a prescindere dal loro uso per il registro⁴². È dunque essenziale memorizzare le informazioni ricevute storicizzandole, quindi mantenendo la storia dei cambiamenti subiti da tutti gli attributi. Per ogni fonte, pertanto, il database contiene una tabella partizionata per anno o per semestre e in ogni partizione è conservata l'informazione ricevuta dall'archivio per l'anno o il semestre di riferimento.

A questo punto le informazioni sono sottoposte ad un insieme di regole e di procedure di normalizzazione al fine di esaminare i record elementari delle fonti amministrative di input acquisite e trasformarli in dati statistici. Per fare questo sono create, e tenute aggiornate nel tempo, delle matrici di conversione per mappare i codici presenti nelle classificazioni amministrative su quelli necessari per il registro, per le principali variabili quali la forma giuridica e i codici di localizzazione (regione, provincia e comune). Particolarmente importante è la storicizzazione di questi ultimi, per avere la possibilità di applicare alle sedi amministrative del registro solo le variazioni territoriali avvenute in un particolare periodo di riferimento⁴³.

In stretta connessione con le continue revisioni stabilite in sede internazionale e comunitaria, occorre aggiornare continuamente sia il linguaggio statistico economico (definizioni e classificazione delle unità e delle attività), sia le "traduzioni" dei diversi linguaggi amministrativi. Per questo motivo, questa prima fase è in continua evoluzione, anche tenendo conto dell'altrettanto rapida evoluzione delle organizzazioni produttive, delle norme amministrative e fiscali e dei sistemi informativi degli enti fornitori.

L'obiettivo dell'integrazione fisica è l'individuazione delle unità statistiche e quindi l'attribuzione di un codice identificativo univoco ad ognuna di esse, sia unità giuridica sia impresa, abbinando fra loro le unità elementari degli archivi di input. In occasione del Censimento anche questa fase ha subito delle modifiche sostanziali. In precedenza le operazioni di *link* erano eseguite in sequenza.

⁴² La sperimentazione più rilevante è stata il Progetto ARCHivio Integrato di Microdati Economici e Demografici (Archimede) che prevede di ampliare l'offerta informativa dell'Istat mediante la produzione di collezioni di dati elementari di tipo longitudinale e *cross section* da rendere disponibili all'utenza, utili alla ricerca sociale ed economica, alla programmazione territoriale e settoriale, alla valutazione delle politiche pubbliche a livello nazionale, regionale e locale. (2014) Garofalo G., Il progetto Archimede, obiettivi e risultati sperimentali, Istat Working Papers n. 9/2014 <http://www.istat.it/archivio/140232>.

⁴³ L'Istat acquisisce tutte le variazioni territoriali e amministrative verificatesi sul territorio nazionale e documentate dal relativo provvedimento legislativo (esempio: legge regionale) pubblicato su un documento istituzionale ufficiale (esempio: Gazzetta Ufficiale, bollettino regionale). Tutte le variazioni intercorse e pervenute all'Istat sono registrate negli appositi elenchi, pubblicati con cadenza semestrale il 30 giugno e il 31 dicembre di ogni anno. Ulteriori aggiornamenti vengono pubblicati in periodi diversi da quelli standard, per garantire la tempestiva divulgazione delle novità di tipo amministrativo, rilevanti per gli enti istituzionali e di interesse per tutti gli utenti, pubblici e privati.

Prima il *link* intra-archivio, ossia il confronto di ogni fornitura amministrativa con quanto acquisito l'anno precedente per la stessa fonte, al fine di costruire i nuovi gruppi o aggiornare quelli già esistenti, individuando le unità che si presentano per la prima volta, eliminando quelle non più presenti, e infine segnalando le eventuali variazioni nei principali caratteri tra l'anno $t-1$ e l'anno t per le unità presenti in entrambe le forniture. Si definisce gruppo l'insieme dei record elementari, o unità amministrative, di una fonte di input che afferiscono ad una stessa unità individuata dal codice fiscale. Solo successivamente si eseguisce il *link* inter-archivi dove ogni record elementare di ciascuna fonte era accoppiato con tutte le altre fonti, mediante codice fiscale o partita Iva (in alcuni casi nel passato si è fatto uso anche della ragione sociale e dell'indirizzo), per la costruzione dei grappoli (o *cluster*) di record, costituiti da un insieme finito di gruppi. Si definisce grappolo l'insieme dei record elementari presenti nelle varie fonti che afferiscono ad una stessa entità individuata da un codice fiscale.

A partire dall'anno censuario le stesse operazioni sono eseguite per ogni archivio amministrativo, anche quelli non utilizzati dal registro, all'interno del sistema Sim. Appositamente per questo scopo si è adottata la soluzione di creare, e tenere aggiornata nel tempo, una tabella in cui sono registrati tutti i codici fiscali che nel tempo l'Istat ha ricevuto da tutte le fonti di input che gestisce, contrassegnando ognuno di essi con un codice identificativo assegnato in maniera univoca (codice di unità giuridica). Quando l'Istat riceve una fornitura di dati da un qualsiasi ente, ad ogni unità amministrativa, ossia ad ogni record elementare ricevuto dalla fonte, è assegnato il codice identificativo di unità giuridica che corrisponde al record della precedente tabella che si abbina utilizzando in maniera opportuna le chiavi di codice fiscale e/o partita Iva. In caso di mancato abbinamento si tratta di un nuovo codice fiscale che contribuirà ad incrementare la tabella storica con un nuovo codice di unità giuridica. In questa maniera tutte le unità amministrative di ogni fonte, aventi ciascuna un proprio codice identificativo interno della fonte, ricevono un codice identificativo di unità giuridica.

In questa maniera con un unico passaggio sono creati i gruppi e i *cluster* di unità. Ogni unità giuridica – individuata dal corrispondente codice di unità giuridica assegnato univocamente e a cui è abbinato un codice fiscale e una o più partite Iva che la stessa unità ha utilizzato nel tempo – costituisce un cluster di record elementari che provengono dalle varie fonti amministrative che forniscono informazioni circa la stessa unità. Il sottoinsieme di record elementari, o unità amministrative, proveniente da una stessa fonte costituisce un gruppo.

In questo modo, l'arrivo di una nuova fonte comporta un aumento di informazione: i record che si presentano per la prima volta, se accoppiati ad un grappolo preesistente, aggiungono a questo una possibile informazione nuova; se non si accoppiano con un grappolo preesistente, sono utilizzati per costruire un nuovo grappolo.

I microdati di ciascuna fonte amministrativa inserita nel sistema, quindi, sono identificati in maniera univoca e risultano essere automaticamente integrati o, comunque, integrabili non solo con le altre fonti in cui sono presenti gli stessi oggetti di analisi ma anche con tutti i possibili output statistici che usano il Sim come punto di partenza per i loro processi. Per i fini peculiari del registro, al termine della fase di integrazione, occorre identificare i grappoli validi, ossia le unità giuridiche che appartengono alla base amministrativa del registro Asia dell'anno t , utili per la costruzione dell'unità statistica "impresa". Si tiene conto della provenienza dell'informazione del codice fiscale e la scelta ricade su quelle fonti che possono assegnare un codice fiscale alle nuove imprese e che permettono di intercettare i legami tra partite Iva e codici fiscali, quali l'Anagrafe tributaria delle persone giuridiche e delle persone fisiche con partita Iva e il Registro delle imprese delle Camere di commercio. Di conseguenza si attribuisce il codice identificativo dell'unità economica (codice unità) solo alle unità giuridiche che nell'anno t sono presenti in una di queste due fonti. Mentre il

processo di alimentazione del codice unità giuridica è continuo, l'attribuzione del codice unità è relativa all'anno di riferimento del registro e basata sulla presenza dell'unità in almeno una delle due fonti di riferimento del processo ma con procedure che garantiscono la stabilità nel tempo del codice identificativo. Ogni anno le unità statistiche individuate da questo tipo di relazione sono oltre 12 milioni, e costituiscono la base amministrativa del registro Asia-imprese e Asia-unità locali⁴⁴.

2.5 La seconda fase del processo di aggiornamento delle imprese: la stima dei caratteri

2.5.1 La stima dell'attività economica

Il processo di stima della variabile "attività economica" delle imprese, codificata utilizzando la classificazione Ateco 2007⁴⁵, si basa su un approccio micro. Il valore del codice è, infatti, determinato unità per unità, sfruttando il contenuto informativo di cui si dispone per ciascuna di esse fornito dalle fonti amministrative. La metodologia di attribuzione in uso è stata sviluppata utilizzando quattro fonti: Agenzia delle Entrate - Studi di settore (Sds); Archivio dichiarazioni annuali delle imposte indirette (Iva); Registro delle imprese delle Camere di commercio (Cciaa); Anagrafe tributaria delle persone giuridiche e delle persone fisiche con partita Iva (At). Il processo di assegnazione privilegia le fonti Sds e Iva, in quanto permettono di avere informazioni sui valori di fatturato e/o volume d'affari, in termini assoluti o percentuali, prodotti nelle diverse attività economiche svolte, in modo da poter anche ordinare tali attività per questi valori. Meno prioritarie sono le fonti di tipo "dichiarativo" come Cciaa e la fornitura At, utilizzata solo ad integrazione per le unità non coperte dalle altre fonti.

Le fonti sono prima di tutto analizzate singolarmente e poi integrate per definire le regole di scelta tra le possibili diverse informazioni, tenendo conto della scala di priorità delle fonti, come sarà descritto in dettaglio nel seguito. Si sottolinea che la procedura utilizzata è operativa dall'anno di riferimento 2007 quando, in concomitanza con l'introduzione della nuova classificazione Ateco 2007, la metodologia e le fonti utilizzate sono state completamente rivisitate⁴⁶.

Il trattamento a fini statistici degli oltre 200 modelli che costituiscono gli Studi di settore è l'elemento centrale della metodologia. Annualmente, la fonte contribuisce all'aggiornamento della variabile attività economica per oltre il 60 per cento delle imprese del Registro. Questo strumento fiscale è stato introdotto nel 1993 con l'obiettivo di valutare la capacità di produrre ricavi o conseguire compensi dalle singole attività economiche svolte dai soggetti interessati. In sintesi, le informazioni ottenute dalla fonte possono essere raggruppate in due insiemi: gli elementi comuni a tutti i tipi di impresa (occupazione, localizzazione, valori di bilancio); un insieme di dati legati alla

⁴⁴ In questa fase avviene anche l'individuazione provvisoria dell'unità statistica "unità locale". Ogni grappolo viene infatti suddiviso in sottoinsiemi simili per attributi di localizzazione, solitamente suddividendo in più parti uno stesso gruppo e attribuendo a ciascuno un proprio codice identificativo di unità locale. I sottoinsiemi così formati costituiscono la base di partenza per il successivo abbinamento con altre fonti amministrative e statistiche specifiche delle unità locali, nel processo di produzione del registro Asia-unità locali a cui è dedicato il Capitolo 4 della Parte Prima del presente fascicolo.

⁴⁵ Versione nazionale della classificazione europea Nace Rev. 2 in *Classificazione delle attività economiche - Ateco 2007*. Istat, Metodi e norme n. 40 Anno 2009.

⁴⁶ A partire dal 2007 si è abbandonato l'approccio probabilistico che non privilegiava *a priori* una fonte rispetto ad un'altra considerata meno affidabile, bensì utilizzava una metodologia per selezionare, con criteri probabilistici, fra i diversi valori forniti dalle fonti amministrative, avvalendosi di opportuni indicatori di qualità delle fonti ricavati utilizzando gli stessi dati a disposizione (probabilità *a priori*), aggiornati successivamente effettuando tutti i possibili confronti tra coppie di codici attribuiti ad una stessa unità (probabilità *a posteriori*). La stima probabilistica del carattere, nell'ipotesi di base che ogni fonte contenga degli errori, era ottenuta scegliendo il codice per il quale la misura di questo errore era minima, quello cui risultava associata la più piccola probabilità di errore *a posteriori*. Per approfondimenti su tale metodo, Abbate, C. 1995. *Una metodologia per la definizione ottimale degli attributi*, in "Verso un sistema statistico integrato delle imprese in Europa", Franco Angeli, Milano.

specifica attività svolta (input, output, processi produttivi, organizzazione e beni materiali). I dati di questo secondo sottoinsieme costituiscono il punto di partenza per la definizione di regole deterministiche che legano variabili, anche eterogenee, compilate in diverse sezioni del modello, dalle quali derivare contestualmente sia i codici di ogni attività economica svolta, principale o secondaria, sia le percentuali di fatturato che ciascuna di queste attività produce.

L'idea di base della metodologia sviluppata è stata quella di legare l'attribuzione del codice di attività economica all'analisi dei singoli "blocchi" che compongono la definizione di attività economica stessa presente nelle linee guida della Nace Rev. 2: "*un'attività economica ha luogo quando risorse come capitali, lavoro, conoscenze tecnologiche o prodotti intermedi sono combinati per produrre specifici beni o servizi*". Un'attività economica è, dunque, caratterizzata da input di risorse, processi produttivi e output di prodotti (beni o servizi) e le regole sono state costruite sulla base di "domande chiave" nei questionari Sds da cui ricavare informazioni su questi tre elementi caratterizzanti la definizione. Non avendo a disposizione i dati in termini di valore aggiunto, il fatturato (valori percentuali) è stato utilizzato come sua *proxy* per garantire l'individuazione dell'attività prevalente mediante applicazione del metodo dell'ordine decrescente, riuscendo ad ottenere in molti casi anche i codici delle attività secondarie svolte⁴⁷.

Anche questa fonte amministrativa è in continua evoluzione pertanto la fase di normalizzazione assume grande rilevanza, sia in termini di complessità sia di tempo necessario per il suo svolgimento. Ogni modello è analizzato singolarmente, in modo da riconoscere se il set di informazioni rilevanti ai fini della codifica sia rimasto lo stesso oppure se occorra modificare le regole di trattamento. L'analisi viene svolta sulla base dei metadati disponibili per l'anno di riferimento t , confrontandoli con quelli dell'anno precedente $t-1$. I metadati sono disponibili alcuni mesi prima (settembre dell'anno $t+1$) rispetto ai micro dati (dicembre $t+1$) e questo consente di anticipare la fase di aggiornamento delle regole per i modelli che hanno subito delle variazioni rispetto all'anno precedente, in modo che questa sia terminata prima di dover applicare le regole al momento della ricezione dei micro dati. Annualmente i modelli modificati, per i quali è necessario verificare se le regole definite sono ancora valide, sono circa il 35 per cento (quasi 70 modelli). Il confronto tra i metadati e la conseguente creazione e/o aggiornamento delle regole avviene tramite l'uso di un'applicazione web (*AtecoRegole*) che permette di popolare, modificare e tenere aggiornata la base dati Oracle con le regole da utilizzare, riducendo al minimo la possibilità di commettere errori di digitazione nella formalizzazione delle stringhe e nel calcolo dei valori percentuali⁴⁸. Occorre sottolineare come ogni anno siano solo dodici i modelli che non sono utilizzabili per la stima dell'attività economica.

Infine, per alcuni casi limitati, meno di dieci Studi di settore ogni anno, la codifica non passa attraverso la costruzione delle regole deterministiche ma si avvale dell'utilizzo dei gruppi omogenei (*cluster*) che l'Agenzia delle Entrate fornisce all'Istat. Si tratta di una elaborazione che applica tecniche statistiche quali l'Analisi in Componenti Principali e la *Cluster Analysis* e per ogni Studio di

⁴⁷ Dal volume Ateco 2007 si traggono le regole di base per la classificazione: "*ad ogni unità registrata nei registri statistici delle imprese si associa un codice Ateco in base all'attività economica principale, cioè quella che contribuisce maggiormente al valore aggiunto dell'unità*"; "*nel caso complesso in cui un'unità svolge più attività classificate in più di due posizioni Ateco, e nessuna delle attività incide per oltre il 50 per cento del valore aggiunto, la classificazione per attività dell'unità dovrà essere stabilita secondo il metodo top-down*" che "*segue il principio gerarchico per cui la classificazione di un'unità al livello più basso della classificazione deve essere coerente con la classificazione dell'unità al livello più alto della struttura*". Inoltre il processo "*identifica la sezione con la percentuale di valore aggiunto più alta; identifica, all'interno della sezione, la divisione con la percentuale di valore aggiunto più alta; identifica, all'interno della divisione, il gruppo con la percentuale di valore aggiunto più alta; identifica, all'interno del gruppo, la classe con la percentuale di valore aggiunto più alta*".

⁴⁸ L'applicazione è stata sviluppata da Luciano Cappai, Istat.

settore dà luogo alla creazione di gruppi omogenei basandosi sugli aspetti strutturali, cioè gli stessi quadri del modello che sono utilizzati nella procedura di attribuzione del codice Ateco.

Nella seconda fonte, l'Archivio dichiarazioni annuali delle imposte indirette (Iva), sono presenti i titolari di partita Iva tenuti alla presentazione della dichiarazione annuale. In questo caso, si dispone di dichiarazioni di volume d'affari prodotto per ogni attività economica svolta da ciascuna impresa e, dopo un'accurata fase di trattamento volta ad evitare soprattutto duplicazioni di valori, è possibile ordinare tali codici di attività economica sulla base del volume d'affari (*proxy* del valore aggiunto) potendo così distinguere tra attività principale e secondarie.

La terza fonte, il Registro delle imprese delle Camere di commercio, è un archivio pubblico in cui i soggetti giuridici sono obbligati per legge all'iscrizione (art. 2188 del Codice Civile). Riguardo il campo di osservazione, il settore agricolo è incluso solo parzialmente, mentre per lavoratori autonomi e liberi professionisti non sussiste l'obbligo di iscrizione. Inoltre, è da considerare come le iscrizioni siano un atto dovuto a fini amministrativi e non costituiscano un segnale reale di attività, come invece richiede la definizione di impresa prevista nei regolamenti europei. Al momento della registrazione l'impresa dichiara, tra le altre informazioni, quale attività economica svolge, tale comunicazione viene codificata da personale esperto del sistema camerale assegnando codici di attività principale e secondaria in classificazione Ateco 2007, per le attività svolte sia nella sede sia nelle eventuali unità locali.

La quarta fonte, l'Anagrafe tributaria, è un archivio gestito dall'Agenzia delle Entrate per conto del Ministero dell'Economia e delle Finanze in cui sono registrate le persone fisiche e giuridiche tenute alla presentazione della dichiarazione per il pagamento delle imposte dirette o indirette. Per tutte le unità giuridiche, identificate dal codice fiscale, il codice Ateco 2007 è una delle variabili presente nella fonte e consiste in una autodichiarazione dell'attività svolta dal contribuente. Nella dichiarazione di inizio attività/variazione dati da presentare all'Agenzia delle Entrate deve essere indicato il codice dell'attività svolta in via prevalente (con riferimento al volume d'affari) desunto dalla classificazione delle attività economiche vigente al momento della presentazione del modello.

Dopo aver trattato singolarmente i quattro archivi, i risultati sono integrati in modo da associare a ciascuna impresa i codici, principali e secondari, forniti da ciascuna fonte e, ove presenti, i relativi valori percentuali di fatturato e/o di volume d'affari. Nella base amministrativa valida per l'anno t , è presente anche il codice Ateco attribuito per l'anno $t-1$ ad ogni unità dallo stesso processo. Il passo successivo prevede la scelta dell'attributo ottimale tra tutti i possibili codici di attività economica presenti nel vettore di informazioni proprio di ogni unità contenuta nella base.

Come già evidenziato, la metodologia accorda una maggiore preferenza alle fonti che generano codici Ateco ordinabili sulla base di variabili quantitative, quali Sds e Iva. Inoltre, gli Sds sono posti al vertice della gerarchia delle fonti data la disponibilità di informazioni riconducibili alla definizione stessa di attività economica. Quando il codice di attività è ricavato da dichiarazioni rese in sede di iscrizione dell'unità non si dispone di dati quantitativi e le relative fonti (Cciaa e At) occupano le posizioni più basse della scala gerarchica all'interno della metodologia di scelta.

In base alla presenza/assenza di informazioni nelle tre fonti principali (Sds, Iva e Cciaa) e nel registro Asia con riferimento all'anno precedente – Asia($t-1$), si ottengono sedici sottoinsiemi per fonti disponibili. Nel caso limite di assenza di dati in tutte e quattro le fonti, si assegna il codice Ateco 2007 registrato in At, se presente. Per i tre sottoinsiemi in cui è presente un'unica fonte, il codice Ateco principale in essi stimato è scelto automaticamente e lo stesso avviene per eventuali situazioni residuali in cui è presente il solo codice del registro per l'anno $t-1$.

Per i sottoinsiemi in cui sono presenti più fonti, invece, si procede con una prima serie di regole di scelta (livello 1) basate su due criteri: uso del *rank* delle fonti stabilito (Sds>Iva>Cciaa); confronto con i codici del registro all'anno $t-1$, al fine di garantire stabilità nel tempo alla codifica. Il rispetto del principio di stabilità implica la registrazione del cambio di codice Ateco nel registro solo nel caso in cui la fonte scelta fornisca, per l'unità economica in esame, lo stesso tipo di variazione per almeno due anni consecutivi. L'applicazione dei due criteri determina l'adozione delle seguenti scelte:

- per i sottoinsiemi dove sono presenti i dati Sds, Iva e Asia($t-1$) si è preferito utilizzare una regola in grado di garantire stabilità nel tempo alla codifica: se il codice Sds è uguale a quello presente nel registro l'anno precedente allora è da preferire indipendentemente dal codice Iva;
- per i sottoinsiemi in cui non è presente Iva, il codice Sds è scelto se uguale al valore del registro;
- per i sottoinsiemi caratterizzati dall'assenza della fonte Sds, il codice Iva è scelto se uguale al valore presente nel registro;
- nei casi in cui non è presente alcun dato nel registro, si è scelto di attribuire automaticamente il codice Ateco di provenienza Sds o Iva rispetto a quello delle Cciaa;
- la contemporanea presenza di Iva e Sds in assenza di Asia($t-1$), determina l'assegnazione del codice Ateco solo nei casi di uguaglianza tra le fonti;
- nei casi in cui è presente la sola Cciaa contestualmente al registro, sulla base del principio di stabilità, è stato confermato il codice già presente nel registro.

L'applicazione di queste regole determina l'attribuzione del codice Ateco a circa l'80 per cento delle unità interessate. Per il resto si procede applicando delle ulteriori regole (livello 2) che tengono conto di indicazioni qualitative sugli Sds (minore qualità, codifiche incomplete) e di un'informazione "terza" rispetto ai codici Ateco: lo stato di attività. In quest'ultimo caso il criterio adottato è quello di preferire il codice legato all'anno in cui l'impresa si attiva (o riattiva) rispetto a quelli eventualmente già registrati negli anni precedenti per l'unità giuridica.

Infine, per le unità con tutte le fonti a disposizione discordanti, a cui ancora non risulta assegnato un codice Ateco mediante regole precedenti, si è preferito estendere il principio di stabilità nell'ambito di una stessa fonte. In generale, la regola adottata è quella di scegliere un codice, anche diverso da quello presente nel registro, nei casi in cui sia confermato dalla stessa fonte per almeno due anni. Di conseguenza, in assenza di fonti concordi per due anni è confermato il valore del registro mentre si deroga al principio della stabilità solo quando le due fonti principali (Sds e Iva) contemporaneamente convergono verso lo stesso valore.

2.5.2 La stima dello stato di attività e l'età dell'impresa

Il processo di stima dello stato di attività di un'impresa si basa su una metodologia che integra il contenuto informativo di differenti fonti amministrative di input. In particolare, tale procedura si compone di vari passaggi al termine dei quali, per ogni unità della base amministrativa – l'insieme delle unità giuridiche dell'anno t – risulta assegnato un valore (pari ad uno se attiva o a zero se non attiva) che indica lo stato di attività dell'unità nell'anno di riferimento t , insieme alle corrispondenti date di inizio attività e di cessazione e al calcolo della età dell'impresa.

Occorre ricordare che, in occasione del Censimento dell'industria e dei servizi 2011, sono state introdotte importanti innovazioni nel processo di aggiornamento del registro e nella stima dei caratteri delle imprese sia dal punto di vista definitorio che metodologico. Con riferimento allo stato di attività di un'impresa, la nuova metodologia abbandona la procedura basata su modelli di natura

probabilistica, utilizzata precedentemente, per adottare criteri di tipo deterministico che fanno uso in via prioritaria dei segnali di attività “reale” quali la presenza di occupazione o di una dichiarazione di volume di affari di valore adeguato a far supporre l'unità come attivamente operante sul mercato⁴⁹.

In sintesi sono considerate attive:

- tutte le imprese che hanno occupazione interna (gli addetti) e/o esterna (ad esempio i collaboratori);
- le imprese senza occupazione interna o esterna che però dichiarano un volume di affari superiore ad almeno il 50 per cento di un certo valore soglia, che equivale al costo medio di un lavoratore nello stesso settore e nella stessa ripartizione geografica, calcolato utilizzando la media delle retribuzioni degli occupati dipendenti, per la sola tipologia di occupati “operai” e in condizione di full time, presenti nelle denunce retributive mensili EMens delle imprese appartenenti allo stesso strato, omogeneo per ripartizione geografica e attività economica prevalente svolta.

Il processo di imputazione dello stato di attività, che si compone di quattro passi principali, si applica alla base amministrativa di unità giuridiche identificate per l'anno t , ossia un insieme di oltre 12 milioni di grappoli di unità amministrative considerati validi per l'anno t , come descritto nel Paragrafo 2.4.2.

Il primo passo consiste nella scelta di tutte le informazioni utili per assegnare lo stato di attività, ossia quelle che rappresentano un possibile segnale di attività per ogni unità della base amministrativa:

- data di inizio e di fine attività che le unità giuridiche presentano nell'Anagrafe tributaria delle persone giuridiche e delle persone fisiche con partita Iva (At);
- data di inizio e di fine attività che le unità giuridiche presentano nel Registro delle imprese delle Camere di commercio (Cciaa);
- date di inizio di procedure concorsuali (fallimento, amministrazione straordinaria, liquidazione);
- lista delle partite Iva provenienti dalle statistiche con il commercio con l'estero che fornisce un utile patrimonio informativo sui flussi commerciali import/export dell'Italia con il resto del mondo delle imprese (Coe);
- lista degli eventi amministrativi di acquisizione, nascita, cessione e cessazione, con le relative date di validità dell'evento, selezionati dal database degli eventi societari;

⁴⁹ Per la realizzazione delle prime versioni del registro, lo stato di attività è stato determinato attraverso l'applicazione di un modello logistico basato sull'individuazione di alcune variabili registrate nelle singole fonti amministrative che potevano essere considerate oggettivi segnali di attività. L'ipotesi base era quella di classificare una unità come attiva o non attiva sfruttando l'informazione congiunta proveniente da tali segnali. Fino al 1998 il modello era costruito sulla base dell'informazione ricavata dal Censimento intermedio dell'industria e dei servizi (Viviano C., *The determination of the state of activity of a statistical unit: a probabilistic approach*, in 'Proceedings of session 51th', Bulletin of ISI – August 18-19, Istanbul, 1997, pp. 625-626). La rapida obsolescenza delle informazioni raccolte e della struttura della popolazione di riferimento, nonché l'impossibilità, per motivi di costi, di realizzare una indagine periodica mirata e particolarmente controllata, ha reso necessario, a partire dall'anno di riferimento 1999, l'utilizzo di una diversa base di informazione su cui ancorare le stime dei parametri del modello, scegliendo opportuni sottoinsiemi di informazioni delle trasformazioni giuridiche di impresa registrate nell'archivio dell'Anagrafe tributaria (Garofalo G., Porcu M., Viviano C. 2003. *Lo Stato di Attività delle Imprese. L'impiego delle fonti amministrative per l'adattamento di un modello probabilistico*, in Quaderni di ricerca Istat n. 3/2003, e nel volume “L'integrazione di dati di fonti diverse”, n. 365.397, Franco Angeli, 2005). In occasione del Censimento per l'anno di riferimento 2011 del registro, si abbandona definitivamente l'approccio probabilistico.

- volume affari attribuito alle imprese del registro Asia (dichiarazioni Iva annuali integrate con la fonte dell'Agenzia delle Entrate dei "Contribuenti minimi", ovvero le dichiarazioni di quei contribuenti la cui attività d'impresa o professionale abbia determinati requisiti specifici secondo la normativa vigente).

Allo scopo di considerare in maniera differenziata le unità con e senza occupazione, si recuperano tutte le informazioni relative all'occupazione nell'anno t a livello di impresa (dal sistema informativo LEED) e si agganciano alla base amministrativa per codice unità⁵⁰. Nello specifico, per individuare un segnale di occupazione sono considerati, per ogni unità giuridica:

- presenza di almeno un lavoratore dipendente – fonte EMens;
- presenza di almeno un lavoratore in Cig – fonte EMens;
- presenza di lavoratori esterni all'impresa (parasubordinati e/o lavoratori temporanei, ex interinali) – fonte EMens;
- per le imprese individuali: presenza di almeno un familiare a carico, desunta dall'archivio delle posizioni contributive degli imprenditori artigiani e commercianti;
- presenza di occupazione in particolari categorie contrattuali, di norma riferite al lavoro autonomo, quale ad esempio quella desunta dall'albo dei medici con reddito maggiore di zero e con Irap valorizzata.

Il secondo passo consiste nell'assegnazione delle date di inizio attività e di cessazione tra quelle esistenti nelle fonti amministrative.

La data di inizio attività è assegnata nel modo seguente:

- se l'unità è presente nella base amministrativa di Asia($t-1$), si conferma la data di inizio $t-1$;
- se l'unità è nuova la data di inizio attività è scelta come la minima (la meno recente) tra quelle esistenti nei due archivi At e Cciaa;
- se l'unità è coinvolta in eventi di trasformazione (acquisizione, cessione, eccetera) nell'anno t si sostituisce la data inizio di tale evento alla data di inizio attività precedentemente scelta.

La data di fine attività è assegnata nel modo seguente:

- se nell'archivio At l'unità presenta un particolare *flag* di cessazione con la relativa data di cessazione, si sceglie quest'ultima come data di fine attività;
- altrimenti, se valorizzata, si sceglie la data di fine attività presente nell'archivio Cciaa;
- se l'unità è coinvolta in un evento amministrativo che implica la cessazione dell'unità stessa, si sceglie proprio la data dell'evento come data di fine attività;
- se l'unità ha una procedura concorsuale in corso (ossia con una data di inizio procedura valida, minore o uguale all'anno t), in Asia($t-1$) non era stata stimata come unità attiva e in aggiunta nell'anno t l'unità non presenta né volume d'affari né occupazione, allora si impone come data di fine attività quella relativa alla procedura concorsuale;
- in tutti i casi precedenti, a meno del caso in cui l'unità è coinvolta in eventi, la data di fine attività non viene comunque valorizzata se l'unità ha occupazione nell'anno t ;

Non viene imputato alcun valore alla data di fine attività quando non sussistono le condizioni elencate nei punti precedenti, oppure quando l'unità è presente nei dati Coe dell'anno t .

⁵⁰ Nel Capitolo 3 della Parte Prima di questo fascicolo è descritto in maniera estesa il sistema informativo sull'occupazione.

Il terzo passo consiste nell'applicazione di regole deterministiche che assegnano in maniera certa, e coerentemente alle date di inizio e fine attività già assegnate, lo stato di attività alle unità quando sussistono alcune condizioni nelle fonti amministrative:

- se l'unità ha effettuato nell'anno t importazioni e/o esportazioni – l'unità è presente nel $Coe(t)$ – questo è considerato un segnale forte di attività e dunque lo stato è posto uguale a 1 (l'unità è considerata attiva);
- se l'unità nello stesso anno ha subito un evento che attesti la sua attività (es. evento di nascita o di acquisizione), questo è considerato un segnale forte di attività e dunque lo stato è posto uguale a 1 (l'unità è considerata attiva);
- se l'unità ha subito un evento di cessazione precedente all'anno di riferimento, questo attesta la sua inattività e dunque lo stato è posto uguale a 0 (l'unità è considerata non attiva);
- come per la data di fine attività, se l'unità ha subito una procedura concorsuale nell'anno t e lo stato di attività nell'anno $t-1$ era pari a 0 (unità non attiva nell'anno $t-1$), il suo volume d'affari è zero o mancante, non ha occupazione e non sussistono le condizioni precedenti di esistenza di eventi e/o import/export, allora anche per l'anno t lo stato è posto uguale a 0 (l'unità è considerata non attiva).

Il quarto passo consiste nell'individuare due sottoinsiemi distinti in cui suddividere la base per utilizzare il set informativo ottenuto:

- 1) unità con occupazione: unità con presenza di dipendenti o con lavoratori esterni collaboratori e/o somministrati e altre unità secondo le specifiche evidenziate nella descrizione del passo uno. Per tutte queste unità lo stato è assegnato come 1 (tutte le unità con una qualsiasi forma di occupazione sono considerate attive). Fanno eccezione solo le unità coinvolte in eventi di cessazione per le quali lo stato è posto uguale a 0 (l'unità è considerata non attiva);
- 2) unità senza occupazione (assenza di dipendenti e di lavoratori esterni collaboratori e somministrati).

Per stabilire lo stato di attività di queste unità è fondamentale la presenza/assenza del volume d'affari. In generale sono considerate attive tutte quelle imprese che dichiarano un volume di affari superiore ad almeno il 50 per cento di uno specifico "valore soglia", che equivale al costo medio di un lavoratore che opera nello stesso settore e nella stessa ripartizione geografica, calcolato utilizzando la media delle retribuzioni degli occupati dipendenti, per la sola tipologia di occupati "operai" e in condizione di full time, presenti nelle denunce retributive mensili EMens.

L'idea di fondo è che per poter svolgere una attività produttiva il fatturato debba essere sufficiente almeno a remunerare il 50 per cento del costo medio di un lavoratore nello stesso settore e nella stessa ripartizione geografica.

Pertanto tutte le imprese senza occupazione vengono suddivise in strati omogenei per ripartizione geografica e attività economica prevalente svolta. Ad ogni strato così individuato viene assegnato un valore soglia di fatturato.

Sono considerate attive tutte le unità dello strato con volume d'affari superiore al valore soglia. Fanno eccezione naturalmente tutte le imprese alle quali lo stato viene assegnato secondo le regole descritte nel terzo passo.

Esistono delle eccezioni a tale regola, relativamente a due categorie particolari:

- nel caso di imprese individuali, se l'unità ha volume di affari inferiore al valore soglia è comunque considerata attiva se risulta avere un fatturato maggiore di zero e se è presente nella fonte Artigiani e Commercianti (Inps) con un familiare a carico;
- nel caso della categoria dei medici, se l'unità ha volume di affari inferiore al valore soglia è comunque considerata attiva se risulta presente nella fonte Irap con una dichiarazione maggiore di zero.

Una diretta conseguenza dell'assegnazione della data di inizio attività e dello stato di attività è la stima dell'età dell'impresa.

In generale l'età dell'impresa è calcolata come differenza tra l'anno di riferimento t e la data inizio attività presente in Asia(t). Tenuto conto che la data di inizio attività è una variabile di tipo amministrativo proveniente o dalla fonte At o da Cciaa, non sempre tale data coincide con l'effettiva data di "inizio attività" di un'impresa presente nel registro.

Secondo la definizione della Commission Regulation (EC) No 2700/98 per nascita (reale) di un'impresa si intende "la creazione di una nuova attività i cui fattori produttivi scaturiscono dal nulla, ossia dove nessuna altra impresa è coinvolta nell'evento. Sono escluse tutte quelle creazioni dovute a modificazioni strutturali quali fusione, scorporo, ecc ... o che risultino collegate secondo specifici criteri di continuità ad unità già esistenti". La creazione di una impresa può essere considerata una nascita (reale), se implica la creazione di nuovi fattori produttivi, e in particolare nuovi posti di lavoro.

Il processo di identificazione delle imprese reali nate è basato sull'applicazione di una complessa procedura costituita da una serie di passaggi automatici e manuali che mirano ad escludere, ovvero ad identificare, le componenti "non reali" dall'insieme delle nuove imprese (Entrate) del Registro delle imprese attive con riferimento ad un periodo (anno) t .

Per ogni anno, a partire dal 1999 (primo anno in cui i dati di demografia d'impresa sono stati prodotti), sono disponibili gli insiemi delle imprese reali nate. Per tali imprese l'età è calcolata come differenza tra l'anno t e l'anno di nascita "reale" dell'impresa.

Ciò implica che l'età dell'impresa e la differenza tra l'anno t e la data di inizio attività potrebbero non coincidere.

2.5.3 La stima dei principali eventi di trasformazione societaria

Una impresa può subire nel corso della sua vita eventi di trasformazione organizzativa che impattano sulla sua identità⁵¹. Gli eventi demografici, ossia quelli che determinano cambiamenti sostanziali di esistenza o di modificazione organizzativa, devono essere tradotti in movimenti nel registro. Nel sistema Asia, considerata la rilevanza di tale fenomeno e l'impatto sulla struttura economica che ne consegue, il trattamento e la registrazione di tali informazioni segue un processo produttivo e di aggiornamento *ad hoc*, che termina con la loro registrazione in un archivio apposito denominato database degli eventi societari.

Alla base del processo di costruzione di tale database si utilizza principalmente l'archivio fiscale degli "eventi societari" gestito dall'Agenzia delle Entrate, denominato file delle modificazioni, che viene successivamente integrato con le fonti statistiche a disposizione. Le dichiarazioni che le unità economiche utilizzano per comunicare all'Agenzia delle Entrate le operazioni di riorganizzazione aziendale poste in essere avvengono sostanzialmente per due finalità:

- per una migliore distribuzione delle strutture societarie e produttive all'interno di un determinato gruppo di società;

⁵¹ Per una descrizione dei principali eventi demografici si veda il Paragrafo 2.2.4.

- per consentire il passaggio di strutture produttive dal controllo di un gruppo societario ad un altro, o comunque in presenza di modificazioni della compagine sociale, con l'ingresso di nuovi partner/soci.

In seguito all'analisi del contenuto informativo del file delle modificazioni si è scelto di utilizzare i dati relativi solo ad alcune "operazioni straordinarie", tra le quali rientrano tutte le categorie di trasformazioni esaminate in precedenza⁵².

Per ciascuna operazione il contenuto informativo della fonte amministrativa di cui si dispone riguarda le unità coinvolte, la tipologia di evento e la data in cui esso avviene, che rappresentano i tre elementi costitutivi del database degli eventi societari:

- unità giuridiche coinvolte: l'evento registrato nella fonte coinvolge due o più codici fiscali ai quali sono assegnati i codici identificativi di unità giuridica. In tale modo le unità giuridiche identificate possono essere collegate al registro Asia. Le informazioni sugli eventi interessano qualunque tipologia di unità giuridica, indipendentemente dalla natura e dal settore di attività. Inoltre possono riguardare unità che sebbene si costituiscano da un punto di vista fiscale/giuridico potrebbero non svolgere alcuna reale attività economica;
- tipologia di evento: le tipologie di operazione sono distinte tra quelle che comportano la cessazione/la nascita del soggetto da quelle che invece comportano solo una modificazione nell'esistenza delle unità e nelle loro caratteristiche strutturali. Le relazioni possono coinvolgere unità dello stesso tipo, cioè concretizzarsi/perfezionarsi tra persone fisiche o tra persone giuridiche, ma anche tra unità di tipo diverso, cioè persone fisiche con persone giuridiche e viceversa;
- data dell'evento: si riporta la data in cui l'evento di trasformazione è stato registrato, non disponendo di alcuna informazione riguardo la data in cui ha effetto la trasformazione. A tal proposito, il comma 2 dell'art. 2504-*bis* prevede per le sole fusioni per incorporazione, la possibilità di postdatare tutti gli effetti della fusione rispetto alla data dell'ultima iscrizione dell'atto di fusione, mentre il comma 3 del citato art. 2504-*bis* individua i casi di retrodatazione degli effetti della fusione.

La prima fase del processo di alimentazione e di aggiornamento del database degli eventi societari consiste nella normalizzazione dei dati amministrativi ricevuti con particolare riferimento alla tipologia dell'evento. In sintesi vengono applicate tabelle di transcodifica che convertono i codici evento di fonte fiscale in codici statistici utilizzati nel registro Asia e a ciascun legame si attribuisce un indicatore che specifica se la conversione è stata possibile e corretta, possibile e modificata, oppure impossibile⁵³.

Nell'ottica dell'aggiornamento annuale, ogni anno t la struttura informativa ricevuta è confrontata con quella dell'anno precedente con l'obiettivo di:

- integrare le nuove relazioni;

⁵² Gli eventi di trasformazione societaria descritti nel Paragrafo 2.2.4 sono tutti compresi nelle seguenti tipologie di evento selezionate dalla fonte fiscale: fusione per incorporazione; fusione propria; scissione parziale; scissione totale; conferimento (solo per società), cessione e donazione d'azienda; conferimento (solo per società), cessione e donazione di ramo d'azienda; conferimento (solo per società) o cessione d'azienda con mantenimento del codice fiscale; conferimento (solo per ditta), cessione o donazione d'azienda con mantenimento della partita Iva; successione ereditaria; modifica di ditta in società o viceversa.

⁵³ Le tipologie di evento da statistica sono: cessione per trasferimento in nuova impresa (04001); cessazione per trasferimento in nuova impresa (04002); cessione per trasferimento in impresa già esistente (04003); cessazione per trasferimento in impresa già esistente (04004); cessazione per cambio di natura giuridica (04005); cessazione per successione ereditaria (04011); cessazione per subentro di attività (04012); cessione per affitto di azienda (04013); affitto di ramo di azienda impresa esistente (04014).

- validare quelle già consolidate;
- escludere o correggere le relazioni non confermate.

Al fine di validare le tipologie di evento rispetto alla data dell'evento vengono inoltre utilizzate le informazioni riguardanti le date di inizio e fine attività di fonte fiscale delle unità coinvolte nell'evento, in particolare per accertare se l'entità era già esistente o si è costituita alla data in cui si perfeziona l'operazione o viceversa se essa cessa di esistere.

Il database degli eventi societari viene successivamente integrato con le ulteriori informazioni acquisite nell'ambito della continua attività di aggiornamento delle imprese di grandi dimensione da revisori esperti e archiviate nel database relazionale di Asia, il cosiddetto "Asia corrente"⁵⁴.

Nelle operazioni di integrazione tra dati di provenienza amministrativa e dati di natura statistica vengono svolti controlli per accertare se tipo di evento e data (a livello di anno e mese) coincidono. Nel caso di non concordanza si attribuisce maggiore peso alla data ed alla tipologia dell'evento di provenienza statistica ossia registrati in modo interattivo da personale qualificato.

Tutte le unità coinvolte in eventi di diversa natura vengono raggruppate mediante l'attribuzione di un codice identificativo del *cluster*, un codice progressivo numerico che permette di osservare in modo unitario tutte le unità collegate fra loro direttamente o indirettamente. In questa maniera è possibile ricavare facilmente informazioni sull'intero gruppo di unità collegate fra loro, come il numero di legami, cioè di coppie, che lo costituiscono, oppure il numero di unità giuridiche coinvolte o la storia della trasformazione di specifiche unità.

Alla fine del processo di aggiornamento si assegnano alcuni indicatori utili a valutare la qualità delle relazioni ottenute:

- indicatore della fonte di provenienza della relazione;
- indicatore di sequenza;
- indicatore sintetico di qualità attribuito alla coppia di unità.

L'indicatore di provenienza permette di distinguere la fonte di ogni legame, ossia se questo è presente solo nella fonte amministrativa, solo nella fonte statistica oppure in entrambe le fonti.

L'indicatore di sequenza è un modo di rappresentare e riassumere in un'unica stringa la sequenza di eventi che riguardano una stessa unità. Solo alcuni dei valori che assume sono considerati coerenti, in quanto si riferiscono a sequenze ammissibili. Pertanto può essere attribuita l'ammissibilità della sequenza solo dopo aver adottato delle regole di coerenza e compatibilità tra gli eventi. I legami possono dar luogo a sequenze "unidirezionali" o "bidirezionali". Nel primo caso, si possono configurare due diverse situazioni: l'unità, denominata parte cedente o passiva, perfeziona solo operazioni di cessione/cessazione di attività a favore di altre unità; oppure l'unità, denominata parte beneficiaria o attiva, perfeziona solo operazioni di nascita/acquisizione da parte di altre unità. Nel caso di sequenze bidirezionali, invece, l'unità può essere al contempo attiva e passiva.

Nel caso di sequenze unidirezionali, le sequenze di legami considerate ammissibili per una unità passiva sono:

⁵⁴ I dati statistici per l'integrazione provengono dalle principali indagini economiche dell'Istituto o sono dati amministrativi disponibili in tempo reale nel Registro delle Camere di commercio accessibile via web relativi a visure storiche, bilanci d'esercizio e consolidati delle unità interessate. Per le imprese più rilevanti del registro le operazioni necessarie per l'individuazione della presenza di un evento e per la stima dei caratteri delle imprese in esso coinvolte sono parte rilevante delle operazioni di aggiornamento del registro descritte nel Paragrafo 2.6. Nel periodo 2008-2013 i legami registrati in modalità interattiva sono stati oltre 3.200 e hanno riguardato, nella maggior parte dei casi, imprese che rivestono un peso economico rilevante sia in termini di addetti che di fatturato.

- l'unità può cedere parte dell'azienda e successivamente, o al massimo alla stessa data in cui si perfeziona il conferimento di parte dell'azienda, cessare l'attività;
- alla medesima data dell'evento possono coesistere conferimenti e cessazioni nei confronti di diverse unità;
- alla medesima data dell'evento possono coesistere più cessazioni;
- a date differenti possono perfezionarsi più operazioni di cessione, dunque, l'evento di cessazione deve essere l'evento conclusivo;
- in seguito ad un evento di cessazione non è ammesso alcun evento, eccetto i casi in cui l'unità coinvolta sia una persona fisica.

Nel caso di sequenze unidirezionali, le sequenze di legami considerate ammissibili per una unità attiva sono:

- le operazioni di nascita devono precedere qualunque altra operazione, tuttavia, l'unità può perfezionare operazioni di nascita e acquisizione alla stessa data;
- le operazioni di acquisizione possono perfezionarsi alla stessa data ma anche a date differenti.

Nel caso di sequenze bidirezionali, l'unità, può essere sia soggetto conferente sia beneficiario, pertanto le uniche sequenze di legami considerate ammissibili sono quelle dove si conserva questo ordine: l'unità deve prima nascere, poi acquisire, infine cedere e/o cessare.

Con l'assegnazione dell'indicatore sintetico di qualità si assegna un *rank* all'evento che coinvolge le unità; tale *rank* è utilizzato sia allo scopo di monitorare la qualità del processo sia per segnalare all'utente finale la qualità dell'informazione prodotta.

Con riferimento alla periodicità dell'aggiornamento, il database degli eventi societari è alimentato annualmente per quanto riguarda la fonte amministrativa, invece la fase di integrazione con "Asia corrente" è di carattere continuo. Generalmente l'ultima fase di aggiornamento è allineata al momento in cui avviene la diffusione finale del registro Asia, in modo tale da accompagnare la diffusione della lista delle imprese attive al tempo *t* con una versione del database degli eventi societari che contenga le stesse informazioni recepite dal registro e quindi coerente con esso, a disposizione per gli utenti delle principali indagini strutturali e congiunturali dell'Istituto.

2.5.4 La stima degli altri caratteri

La procedura di stima degli attributi di classificazione e di identificazione, come la ragione sociale, la forma giuridica, il carattere artigiano e la localizzazione della sede d'impresa, si effettua ogni anno per tutte le unità giuridiche del registro utilizzando esclusivamente informazioni di tipo amministrativo ottenute da diverse fonti di input. Solo successivamente, nella fase finale del processo, le stime da fonte amministrativa sono integrate con quelle di fonte statistica registrate continuamente nel database storico (Asia corrente).

In occasione del Censimento dell'industria e dei servizi 2011 il processo è stato riorganizzato in maniera tale da non distinguere più tra unità entrate per la prima volta nel registro, sottoinsieme per il quale erano stimati tutti i caratteri, e unità già esistenti dove i caratteri erano stimati esclusivamente laddove fosse stata registrata una variazione rispetto all'anno precedente. La variazione poteva riguardare la composizione del grappolo costituente l'unità giuridica, sia in termini di modifica del valore fornito da uno stesso archivio amministrativo nei due anni, sia in termini di variazione nel numero di archivi di input che contribuivano con le proprie informazioni alla stima del carattere⁵⁵. Per tutte le altre unità statistiche i processi di stima non erano eseguiti,

⁵⁵ Per la definizione di grappolo si veda il Paragrafo 2.4.2.

pertanto, considerando come invariati i caratteri, il loro valore si faceva corrispondere alla stima eseguita l'anno precedente. A partire dall'anno di riferimento 2011, potendo anche beneficiare di importanti innovazioni dal punto di vista tecnologico che hanno permesso un notevole risparmio di tempo di esecuzione delle procedure, i processi di stima sono operati su tutte le unità giuridiche della base amministrativa.

La metodologia di assegnazione della *forma giuridica* consiste nell'assegnare ad ogni unità giuridica un codice appartenente alla classificazione realizzata dall'Istat nel 2002⁵⁶. Nel campo di osservazione del registro sono comprese tutte le imprese individuali e familiari, le società di persone (semplici o di fatto, in nome collettivo, in accomandita semplice, gli studi associati), le società di capitale (società per azioni, società a responsabilità limitata, società in accomandita per azioni), le cooperative, i consorzi, le aziende autonome, le società consortili, le associazioni in partecipazione, le imprese costituite all'estero che svolgono una attività economica in Italia.

Il processo si basa su una metodologia che integra il contenuto informativo delle fonti amministrative di input: l'archivio Anagrafe tributaria gestito dall'Agenzia delle Entrate per il Ministero dell'Economia e delle Finanze, e il Registro delle imprese delle Camere di commercio, industria, artigianato e agricoltura. Si articola in diverse fasi e prevede innanzitutto la normalizzazione dei codici utilizzati dalle due amministrazioni in uno o più codici che fanno riferimento alla classificazione ufficiale delle forme giuridiche Istat. In questa fase di standardizzazione sono ulteriormente esaminati i risultati ottenuti di tipo "uno a molti", ossia i casi in cui il codice utilizzato dalla fonte amministrativa è troppo generico e corrisponde a più codici statistici possibili. In questi casi, distintamente per le due fonti, sono applicate delle procedure di scelta che si basano sulla ricerca, all'interno della stringa di ragione sociale, di eventuali indicazioni sulla natura giuridica dell'impresa. A tale scopo si utilizzano opportuni dizionari di parole chiave a cui corrispondono codici di forma giuridica, che nel tempo sono stati creati e tenuti costantemente aggiornati appositamente per questo tipo di procedure di riconoscimento testuale.

La scelta tra le informazioni derivanti dalle due fonti, che possono anche essere discordanti, per l'attribuzione di un codice univoco, è operata distinguendo tra le unità aventi codice fiscale alfanumerico (le imprese individuali) e quelle per cui il codice fiscale è numerico (le società). Per le prime, si escludono dalla possibile scelta tutti i valori non compresi tra quelli attribuibili a imprese individuali e familiari, quindi le prime due cifre del codice di forma giuridica sono valorizzate in automatico per tutti i codici fiscali alfanumerici⁵⁷. Solo successivamente, in fase di check con opportune regole deterministiche, è possibile attribuire le ultime due cifre specificando la tipologia di forma giuridica individuale, quando si rendono disponibili gli altri caratteri del processo, eseguendo delle regole di scelta che tengono conto della registrazione del codice fiscale nell'archivio delle Camere di commercio, della presenza o meno di dipendenti, dell'eventuale carattere artigiano dell'impresa e della tipologia di attività economica svolta.

Per le unità del secondo tipo, aventi codice fiscale numerico, la scelta è possibile solo fra i codici compatibili con tutte le altre forme societarie, ad esclusione delle imprese individuali, contribuendo a garantire in questo modo la compatibilità tra le due informazioni. La selezione avviene inizialmente tra i codici giuridici dichiarati dall'impresa e registrati nei diversi archivi provinciali delle Camere di commercio, essendo considerata la fonte prioritaria per questa informazione, scegliendo tra questi il codice che si presenta con maggior frequenza. In caso si verifichi la coesistenza di due o più codici diversi aventi uno stesso valore di frequenza massimo, la scelta ricade su quello che si presenta con frequenza maggiore nel secondo archivio di input. Solo nel

⁵⁶ <http://www.istat.it/it/archivio/6523>.

⁵⁷ Il codice corrispondente è il 1.1 - "Imprenditore individuale, libero professionista e lavoratore autonomo".

caso in cui non esista alcun codice dichiarato alle Camere di commercio, la scelta ricade sull'informazione proveniente dall'Anagrafe tributaria.

Al termine del processo di attribuzione del codice di forma giuridica, sono calcolati, per ogni unità statistica, una variabile di controllo che segnala l'archivio di provenienza dell'informazione, insieme ad un indicatore della modalità con cui è stata attribuita la stima che permette di distinguere i casi di attribuzione univoca da quelli in cui l'attribuzione è multipla.

La stima del carattere *ragione sociale* è effettuata utilizzando come input esclusivamente la base informativa fornita dagli archivi principali di riferimento per l'individuazione delle unità del registro, ossia l'archivio delle Camere di commercio e l'Anagrafe tributaria dell'Agenzia delle Entrate. Nel dettaglio, la scelta viene eseguita in più fasi, diversificando il procedimento secondo la tipologia di codice fiscale che identifica l'unità, se numerico o alfanumerico. Se l'unità presenta codice fiscale numerico, la scelta utilizza come prioritaria l'informazione proveniente dal Registro delle imprese delle Camere di commercio. Al contrario, nel caso di impresa individuale o familiare, nella scelta si privilegia il valore registrato nell'Anagrafe tributaria. Il motivo è principalmente di opportunità processuale, infatti in questo caso si rende disponibile il nome e cognome dell'imprenditore registrato in maniera standardizzata. In entrambi i casi se in un archivio non fosse presente alcuna informazione valida relativamente alla ragione sociale, la ricerca si spingerebbe comunque verso l'altro archivio.

Al termine del processo di stima, anche per la ragione sociale, nella fase conclusiva del processo che porta all'individuazione del valore corretto, si tiene memoria, registrandone l'informazione nel registro, della fonte di provenienza della stima finale del carattere, in modo da consentire tutti i controlli che si dovessero ritenere necessari successivamente.

In occasione della revisione del processo nella tornata censuaria 2011 sono state unificate di fatto le due procedure di scelta del codice di forma giuridica e di ragione sociale. Rispetto a quanto avveniva precedentemente, tale innovazione si concretizza in una inversione nella priorità tra le due fonti quando per una stessa unità non si dispone di un'informazione completa per entrambe le variabili. In tale maniera per ogni unità la stessa fonte fornisce contemporaneamente le informazioni per entrambe le variabili. In realtà le unità per le quali i due metodi porterebbero a risultati differenti sono solo un numero irrisorio, nondimeno si è scelto di applicare tale modifica per evitare di incorrere nel rischio di incompatibilità tra i valori scelti per le due variabili.

Infine, il processo di stima del *carattere artigiano* delle imprese è strettamente connesso alla natura dello stesso che è per definizione esclusivamente di tipo amministrativo. Le caratteristiche di un'impresa artigiana, infatti, in termini di dimensione, di attività economica svolta e di natura giuridica, sono identificate sia sulla base di specifiche disposizioni di legge sia sulla base dell'iscrizione negli appositi albi provinciali.

A livello legislativo, la normativa di riferimento è rappresentata dalla L. 443/85 che definisce come artigiana l'impresa esercitata dall'imprenditore nei limiti dimensionali definiti dall'art. 4, che abbia per scopo prevalente lo svolgimento di un'attività di produzione di beni, anche semilavorati, o di prestazioni di servizi. Inoltre in ciascuna provincia, presso le Camere di commercio, è istituito l'Albo delle imprese artigiane, al quale sono tenute a far domanda d'iscrizione, entro trenta giorni dall'inizio dell'attività, tutte le imprese aventi i requisiti previsti da tale legge. Proprio questi albi provinciali rappresentano gli unici detentori dell'informazione legale ritenuta ufficiale per il carattere artigiano, in quanto l'iscrizione agli stessi assume natura costitutiva ed è condizione per la concessione delle agevolazioni previste per tali tipologie di imprese.

Nonostante la natura amministrativa, non esiste, però, a livello nazionale ed europeo una definizione statistica unitaria di "impresa artigiana"; questo comporta che qualsiasi confronto a

livello internazionale sia puramente fuorviante perché inficiato dalle specifiche disposizioni legislative.

Tuttavia, la notevole pressione informativa connessa alla produzione di statistiche sulle imprese artigiane, ha fatto sì che all'interno del registro, fino all'anno 2004 di riferimento dei dati, si sia tentato di produrne delle stime, considerandolo uno dei tanti caratteri statistici rilevanti. Fino a tale data la metodologia di stima era basata su un processo di integrazione di più fonti amministrative. In particolare i dati desunti dalle Camere di commercio venivano integrati con quelli dell'Inps e dell'Anagrafe tributaria. La decisione di non utilizzare più questo metodo è stata determinata dal riscontro di una sovrastima nel numero delle imprese artigiane e nella relativa occupazione perché la cancellazione dagli albi provinciali non si accompagnava ad una modifica delle informazioni Inps e ancor meno in quelle dell'Anagrafe tributaria, non esistendo alcun vincolo o sanzioni di tipo fiscale a riguardo.

Pertanto dal 2004 in poi si utilizza come unica fonte l'archivio delle Camere di commercio, che contiene informazioni riguardo l'iscrizione e la cancellazione delle imprese dagli albi artigiani. In seguito a tale modifica il carattere artigiano non si considera al pari delle altre variabili statistiche del registro ma come una variabile amministrativa.

2.6 La terza fase del processo di aggiornamento delle imprese: l'integrazione delle fonti statistiche

Le informazioni presenti nel registro sono in larga parte il risultato di un complesso processo di integrazione di dati provenienti da archivi amministrativi tuttavia, pur rappresentando questi la base informativa primaria per la realizzazione del registro, non sono di per sé sufficienti per la stima definitiva dei caratteri identificativi delle unità.

Tutto ciò è dovuto ad una pluralità di ragioni, che rendono quindi indispensabile lo svolgimento di una continua attività di manutenzione ed aggiornamento del registro stesso. Vi sono delle limitazioni nell'utilizzo delle fonti amministrative a fini statistici, tra cui in primo luogo il *lag* temporale che intercorre tra la data in cui le fonti sono utilizzabili e l'anno di riferimento del registro. La mancata tempestività nella ricezione delle fonti comporta una perdita informativa che deve essere sanata in altro modo. Pertanto si ricorre all'utilizzo di fonti di tipo "statistico", mediante le quali sia possibile aggiornare in tempo reale almeno una parte, quella ritenuta più rilevante, delle imprese contenute nel *running file*, il database relazionale "Asia corrente", contenente le informazioni storiche di tutte le imprese diffuse a partire dal 1996, incluse tutte le modifiche strutturali subite da tali unità nel corso del tempo in conseguenza di cambi dei caratteri identificativi o di localizzazione, di attività economica svolta o del verificarsi di eventi come scorpori e/o fusioni. In linea generale, mentre il processo di integrazione dei dati amministrativi è sufficiente per l'aggiornamento annuale delle unità di piccole dimensioni, che costituiscono circa il 90 per cento delle imprese del registro, per le unità più rilevanti è fondamentale il monitoraggio da parte di revisori esperti di settore.

I dati riferiti alle grandi imprese assumono una significativa rilevanza ai fini della costruzione di indicatori strutturali e per questo devono essere monitorati e aggiornati con particolare attenzione. L'attività economica prevalente, il numero medio di occupati dipendenti e la sede amministrativa, caratteri fondamentali dell'unità statistica impresa, sono dati che non possono essere aggiornati esclusivamente con metodi di attribuzione automatici ma necessitano di controlli manuali da parte di revisori dotati di vasta esperienza e sensibilità nel valutarli. Per questo gli aggiornamenti di tipo online sono considerati qualitativamente superiori a qualunque altro tipo di aggiornamento, in quanto sono frutto di un costante monitoraggio da parte del revisore che, confrontando le

informazioni provenienti da diverse fonti, individua il corretto valore delle variabili da attribuire all'unità statistica esaminata.

Generalmente tra le unità sottoposte a questo tipo di aggiornamento rientrano soprattutto quelle di grandi dimensioni in termini di addetti che spesso risultano parte di un unico gruppo d'impresa; cosa che accade ad esempio tra le società di capitali con oltre 500 addetti, per le quali si osserva che il 95 per cento risulta appartenere a gruppi. In questi casi la struttura organizzativa di tali unità risulta particolarmente complessa, per cui le fonti amministrative non sono da sole in grado di fornire informazioni esaustive e adeguate per stimarne correttamente le variabili strutturali ed identificative. Situazioni del genere devono essere sottoposte ad un approfondito processo di controllo di qualità e richiedono una valutazione delle unità e dei loro legami proprio all'interno del gruppo d'impresa di cui fanno parte; è molto frequente, infatti, che alcune unità rientranti in tali categorie non assumano rilevanza dal punto di vista economico in quanto la loro esistenza è dovuta esclusivamente a ragioni puramente amministrative e fiscali; sono unità che spesso svolgono attività esclusivamente a favore delle imprese da cui risultano controllate senza una collocazione diretta sul mercato.

L'aggiornamento continuo mediante fonti statistiche del registro "Asia corrente" è quindi il risultato di processi differenti che vedono coinvolti i revisori responsabili di tale attività durante tutto l'anno. Durante un generico anno $t+2$ sono aggiornati i caratteri di alcune tipologie di impresa considerate "rilevanti" per gli utenti, con riferimento agli anni t e $t+1$. Più raramente si rendono anche disponibili informazioni per l'anno $t+2$. La rilevanza è definita in base alla dimensione delle imprese, in termini di occupazione e/o fatturato, oppure in quanto di particolare importanza ai fini di una specifica indagine o perché attive in particolari settori di attività economica. L'aggiornamento e la navigazione interattiva del sistema dei registri Asia avviene tramite una *web application* "Asi@web" realizzata durante la reingegnerizzazione del processo di aggiornamento del registro in occasione della tornata censuaria⁵⁸.

Gli input per l'attività di aggiornamento possono derivare sia da fonte interna sia esterna al registro.

Tra gli input ricevuti da fonte interna rientra innanzitutto l'attività del *profiling* e l'aggiornamento del Registro dei gruppi d'impresa; durante tale attività, il revisore analizza la struttura legale, operativa e contabile di un gruppo di imprese, identifica le unità statistiche, controlla il corretto perimetro del gruppo, la nazionalità del vertice, determina l'attività economica prevalente e le eventuali unità ancillari.

Anche l'attività di controllo e correzione dei dati richiesta in fase di check può essere considerata come input da fonte interna; questa è prevista nell'ambito della quarta macro fase del processo di aggiornamento del registro, descritta nel Paragrafo 2.7, durante la quale sono prodotte delle liste di imprese da sottoporre a validazione da parte dei revisori. A questo punto, facendo opportuno uso di alcune banche dati amministrative online, si procede all'esame dei valori sospetti che vengono individuati in seguito all'applicazione di opportune regole di compatibilità. Per effettuare tali analisi è necessaria un'attenta valutazione da parte dei revisori

⁵⁸ Asi@web è nata per sostituire il software NavigaAsia. Trattandosi di una *web application* è residente su un server web e utilizzabile lato client tramite browser da qualunque PC interno all'Istituto, senza dover installare alcun software sul PC degli utenti. Tra le più rilevanti funzioni attualmente implementate: visualizzazione e modifica dei dati relativi a imprese e unità giuridiche; inserimento di nuove imprese o unità giuridiche; associazione di unità giuridiche a imprese esistenti; gestione delle connessioni o eventi strutturali tra unità (visualizzazione, inserimento, modifica, cancellazione); gestione delle note (visualizzazione, inserimento, modifica, cancellazione, eccetera). L'applicazione è stata realizzata utilizzando il *framework Struts2*, con linguaggi Java e *JavaScript* e tecnologie *Spring*, XML, JSP, *jQuery*, *Tiles*, *display:table*. Asi@web può essere utilizzato connettendosi all'URL <http://asia.istat.it>. L'applicazione è stata sviluppata da Massimiliano Amarone, Adele Bianco e Giampaolo De Matteis, Istat.

che non devono limitarsi a verificare la correttezza formale dei codici di classificazione ma devono anche analizzare la compatibilità dei valori di una stessa variabile nei due anni di riferimento del registro ($t-1$ e t).

Tra gli input di aggiornamento da fonte esterna al registro rientrano, invece, le informazioni che vengono acquisite mediante lo scambio dati con le principali indagini statistiche strutturali e congiunturali dell'Istat⁵⁹. In passato tale scambio avveniva in maniera non strutturata utilizzando diverse modalità di contatto tra responsabili di indagine e del registro, quali l'invio di e-mail, fax o contatti telefonici diretti. Dalle indagini che l'Istat effettua sulle imprese venivano trasmesse alla struttura che gestisce il registro le informazioni raccolte in apposite sezioni dei questionari di indagine, in larga parte precompilate con dati del registro stesso, tra le quali la denominazione, l'indirizzo, la forma giuridica, il telefono, una breve descrizione dell'attività economica, le date di inizio e fine attività. I responsabili di ogni indagine inviavano al registro solo le eventuali modifiche comunicate dall'unità intervistata. Una particolare attenzione è sempre stata riservata ai dati provenienti dall'Indagine sulle unità locali delle grandi imprese (Iulgi), che sono utilizzati in maniera prioritaria per implementare il registro di un nuovo livello informativo sulle unità locali, nel rispetto della normativa comunitaria e quindi per l'aggiornamento del registro Asia-unità locali, ma sono particolarmente utili anche ai fini dell'individuazione di variazioni nei caratteri delle imprese⁶⁰. Di fatto forniscono un segnale, anticipato rispetto alla tempistica delle fonti amministrative, del verificarsi di eventi che mutano la "vita" dell'impresa e che comportano cambiamenti sostanziali di stato, dimensione o attività svolta.

Oggi la maggior parte degli aggiornamenti viene acquisito tramite la condivisione delle informazioni raccolte da "Statistica & Imprese", il Portale delle imprese gestito dall'Istat⁶¹. Si tratta di un innovativo sistema di servizi statistici dedicato alle imprese e agli operatori economici per la raccolta e la restituzione di informazioni nell'ambito di tutte le rilevazioni condotte dall'Istat. Attraverso questo sistema, l'impresa può assolvere in modo semplice e veloce ai propri obblighi informativi e può monitorare lo stato degli adempimenti statistici in tempo reale; inoltre accedendo ad una sezione "Anagrafica" può consultare i propri dati contenuti nel registro Asia ed eventualmente comunicare variazioni anagrafiche e strutturali, come gli eventi di scorporo e fusione. Tale strumento costituisce il punto di raccolta unitario sia delle comunicazioni provenienti dalle imprese durante la compilazione dei questionari d'indagine, sia di informazioni acquisite tramite altri canali dal personale coinvolto nelle operazioni di indagine che, comunicandole sul Portale, permette la loro condivisione con tutti i potenziali interessati.

Lo scambio di informazioni tra diverse direzioni dell'Istituto è un processo ancora solo in parte standardizzato, in termini di fasi e di tempi, necessario al fine di: permettere un aggiornamento il più possibile completo e tempestivo del registro; permettere alle altre direzioni dell'Istituto di avere liste aggiornate a cadenza infrannuale per le proprie indagini congiunturali e strutturali; garantire la condivisione, all'interno del sistema delle indagini dell'Istituto, delle informazioni anagrafiche e strutturali delle imprese, qualunque sia la loro provenienza, in tempi anticipati rispetto al rilascio annuale del registro.

⁵⁹ Le fonti statistiche comprendono l'Indagine sulle unità locali delle grandi imprese (Iulgi) e le indagini strutturali e congiunturali che l'Istat effettua sulle imprese, in particolare l'indagine sul sistema dei conti delle imprese con più di 100 addetti, l'indagine campionaria sulle piccole e medie imprese, le indagini sul commercio con l'estero, l'indagine annuale sulla produzione industriale, nonché le indagini congiunturali sulla produzione, sul fatturato e ordinativi delle attività manifatturiere, sul fatturato del commercio e dei servizi, sui prezzi alla produzione e sull'occupazione delle grandi imprese.

⁶⁰ L'utilizzo dell'indagine Iulgi ai fini dell'individuazione delle unità locali del registro è descritto nel Capitolo 4.

⁶¹ <https://imprese.istat.it/>.

Il punto di partenza per lo svolgimento dell'attività di aggiornamento del registro è, pertanto, costituito proprio dalle segnalazioni di variazioni anagrafiche e strutturali acquisite attraverso il *form* "Anagrafica" del sistema del Portale. Quando la segnalazione acquisita è recepita dal back office del Portale, si attiva un'istanza amministrativa identificata da un specifico protocollo, che un revisore del registro prende in carico, procedendo alla sua verifica, anche utilizzando le fonti amministrative disponibili online; al termine della sua analisi il valore aggiornato della variabile "stato della lavorazione", visibile anche all'esterno dall'impresa stessa, fornirà elementi per distinguere i casi di chiusura dell'istanza per accettazione della variazione segnalata, da quelli in cui l'istanza è stata rifiutata perché non considerata plausibile e in tal caso è anche fornita una specifica motivazione per chiarire il rifiuto.

Qualunque sia il canale da cui provenga il segnale che un carattere del registro ha subito una modifica, la verifica della autenticità dell'informazione segue una consueta prassi per l'accertamento, che solitamente ha inizio con la consultazione di alcune fonti di informazione messe a disposizione in seguito a convenzioni e accordi con altri enti, disponibili online. La più utilizzata tra le banche dati amministrative online è il servizio "Telemaco" che consente l'accesso alle banche dati del Registro delle imprese e dell'Albo delle imprese artigiane delle Camere di commercio. Attraverso questo sistema si possono richiedere visure di tutti gli atti camerale e scaricare i bilanci delle imprese registrate. Questo sistema garantisce agli utenti l'accesso immediato ai documenti e ai dati, sia a carattere informativo sia con valore legale, contenuti negli archivi camerale. La visura storica consente di selezionare in modo puntuale blocchi di informazioni che permettono di valutare e validare i cambiamenti anagrafici e strutturali di un'impresa. In particolare consente l'aggiornamento della ragione sociale, della forma giuridica e dell'esistenza di eventuali procedure concorsuali, di cui fornisce date di delibera e esecuzione. Per tutte le iscrizioni e le comunicazioni contenute nella visura storica è sempre indicata la data dell'operazione; questo permette di monitorare le imprese nel tempo e quindi di aggiornare correttamente il registro in tempo reale.

La visura storica fornisce inoltre informazioni sulla *localizzazione*, distinguendo la sede legale da quella amministrativa e, se presenti, anche le diverse tipologie di unità locali. L'impresa può avere, infatti, una sede legale, una sede amministrativa e una o più sedi operative, eventualmente tutte coincidenti nello stesso luogo fisico e uno stesso indirizzo. La sede legale è il luogo dove fiscalmente è stata registrata l'impresa, l'indirizzo dove i terzi sanno essere presente, tramite pubblicità al Registro delle imprese, quella specifica persona giuridica. Dall'atto costitutivo, rappresenta quindi il luogo in cui l'impresa risulta avere il centro dei propri affari. Il concetto di "centro degli affari" può essere ricondotto al mero luogo di notifica della corrispondenza legale e la sede legale può essere anche una semplice casella postale. Nella pratica professionale molte società di capitale o cooperative indicano nell'atto costitutivo come sede legale lo studio di un professionista e in tale luogo vengono tenute le riunioni del consiglio di amministrazione o anche le assemblee dei soci. Frequente è il caso della domiciliazione di società presso consulenti, commercialisti, avvocati, pertanto si può verificare il caso di più società domiciliate su uno stesso indirizzo. Essa funge da supporto giuridico dell'impresa, quindi a volte non è una unità locale, così come si definisce l'unità statistica nel Regolamento (CEE) n. 696/1993, ma un semplice indirizzo senza attività reale o lavoratori. La sede amministrativa, invece, è la sede "effettiva", quella dove si trovano gli uffici dell'impresa, nella quale vengono svolte le attività politiche e gestionali, il luogo dove hanno concreto svolgimento le attività amministrative e di direzione dell'ente e dove operano i suoi organi amministrativi o i suoi dipendenti. Non sempre è possibile distinguere, con procedure automatiche, tra i diversi indirizzi forniti dalle fonti amministrative per assegnare correttamente i valori delle variabili di localizzazione delle unità del registro. Questi dovrebbero essere sempre

riferiti alla sede amministrativa dell'impresa, ossia il luogo dove di fatto l'impresa esplica i propri affari e sono ubicati i principali uffici amministrativo-gestionali. Per le unità più rilevanti l'informazione si può ottenere solo mediante ricorso alla consultazione della visura storica, pertanto il registro contiene sicuramente la sede amministrativa nei casi in cui si ha la possibilità di esaminare tale fonte e validarla da un punto di vista statistico (grandi imprese). Comunque nella maggioranza dei casi, e principalmente in piccole e medie imprese la sede legale coincide con quella amministrativa.

Mentre non sussistono particolari difficoltà nel determinare la sede amministrativa delle imprese, una particolare attenzione è posta nell'aggiornamento dello stato di attività delle imprese, dell'attività economica svolta prevalentemente nell'anno e della dimensione in termini di dipendenti medi, viste le loro implicazioni ai fini statistici.

Per quanto riguarda l'attribuzione dello *stato di attività*, la stima da fonti amministrative⁶² necessita a volte di una revisione, sia per le conseguenze di *lag* temporali esistenti dal periodo di riferimento dei dati al momento in cui sono disponibili le fonti, sia per una diversa interpretazione delle date di validità degli eventi strutturali che intercorrono nella vita delle imprese. Anche per queste analisi si fa ricorso alla banca dati Telemaco e in particolare alle informazioni contenute in visura, la data di iscrizione e l'eventuale data di cancellazione dal Registro delle imprese o l'esistenza di un evento strutturale. In quest'ultimo caso il registro è aggiornato tenendo conto della data di efficacia dell'evento, desumibile dall'atto allegato. L'accesso ai documenti ottici di Telemaco, come atti di cessione e/o acquisizione, costituisce uno strumento informativo efficace e rilevante, in quanto permette di individuare le operazioni di carattere straordinario, quali scorpori o fusioni, che legano due o più imprese, il passaggio di dipendenti dall'unità cedente e l'eventuale variazione dell'attività economica principale che l'evento può comportare.

L'attribuzione del codice di *attività economica* segue una propria prassi ed è, in generale, una delle operazioni più complesse durante l'aggiornamento manuale del registro.

Dal punto di vista definitorio, l'attività principale di un'unità statistica è determinata sulla base del valore aggiunto da essa conseguito mediante i processi produttivi o i servizi resi. Il valore aggiunto è quindi il concetto di base ai fini dell'assegnazione del codice che quantifica la differenza tra il valore della produzione e il consumo intermedio, definendo una misura aggiuntiva del contributo di ogni unità economica al prodotto interno lordo (Pil).

In linea generale, nel caso in cui un'attività rappresenti oltre il 50 per cento del valore aggiunto, questa attività determina la classificazione dell'unità; nel caso, invece, più complesso in cui un'unità svolga più di due attività, appartenenti a due voci dell'Ateco nessuna delle quali rappresenti oltre il 50 per cento del valore aggiunto, la classificazione va determinata utilizzando il metodo dell'ordine decrescente. Il metodo consiste nel seguire il principio gerarchico (*top-down*) per cui la classificazione di un'unità al livello più basso della classificazione deve essere coerente con la classificazione dell'unità al livello più alto della struttura. Nella pratica il processo parte identificando la sezione con la percentuale di valore aggiunto più alta, successivamente si scende a livello di divisione, di gruppo, di classe e alla fine al suo interno viene individuata la categoria con la percentuale di valore aggiunto più alta; in questo modo viene assegnato all'unità il codice Ateco che ne rappresenta l'attività prevalente.

Dal punto di vista operativo, però, difficilmente si hanno a disposizione informazioni sulla ripartizione del valore aggiunto tra le varie attività svolte dall'unità e per questo spesso si utilizzano dei criteri alternativi al valore aggiunto, come il valore delle vendite o il fatturato oppure il numero di lavoratori impiegati nelle varie attività economiche dell'unità. Tali criteri permettono di assegnare

⁶² Il processo di stima da fonti amministrative è descritto nel Paragrafo 2.5.2.

all'impresa il codice di attività prevalente e le eventuali attività secondarie e possono essere applicati nell'aggiornamento online grazie all'analisi dei dati economici deducibili dal bilancio, consultando in particolare la nota integrativa e la relazione sulla gestione. Quest'ultima è un documento redatto dagli amministratori della società e analizza la situazione della stessa nel settore di mercato in cui opera, con particolare riguardo ai costi, ai ricavi e agli investimenti, nonché fornisce una descrizione dei principali rischi ed incertezze a cui la società stessa si è esposta. La nota integrativa documenta i dati sintetici dei prospetti contabili (stato patrimoniale e conto economico); fornisce quindi le informazioni necessarie per una interpretazione più chiara ed esplicativa dei dati quantitativi riportati nelle tavole di sintesi.

Infine il conto economico espone il risultato dell'esercizio attraverso la rappresentazione in forma scalare dei costi e degli oneri sostenuti, nonché dei ricavi e degli altri proventi conseguiti nell'esercizio. Sicuramente questo prospetto costituisce la fonte principale per l'assegnazione del codice Ateco prevalente, in quanto molto spesso sono presenti i valori dei ricavi delle vendite e delle prestazioni ripartiti per le varie attività svolte dall'impresa; in questo modo è possibile assegnare il codice non solo per l'attività prevalente ma anche le eventuali attività secondarie.

Per quanto riguarda, invece, la variabile *dipendenti medi*, le notevoli innovazioni apportate in occasione del Censimento dell'industria e dei servizi 2011 hanno sostituito completamente le modalità di aggiornamento online delle variabili occupazionali. In precedenza gli occupati dipendenti delle imprese di media e grande dimensione erano continuamente sottoposti a verifica e aggiornati quasi esclusivamente online, utilizzando le informazioni presenti all'interno del bilancio d'esercizio e quelle dichiarate dall'impresa stessa nei questionari delle principali indagini economiche⁶³. I dati erano messi a confronto per la ricerca delle eventuali incompatibilità e per la loro correzione si utilizzava come *benchmark* la banca dati online gestita dall'Inps, dove era possibile consultare i dati dei modelli DM10, compilati dal datore di lavoro per denunciare all'Inps le retribuzioni mensili corrisposte ai dipendenti, i contributi dovuti e l'eventuale conguaglio delle prestazioni anticipate per conto dell'Inps, delle agevolazioni e degli sgravi.

L'innovazione censuaria è stata possibile grazie alla disponibilità di nuove fonti amministrative sulla struttura dell'occupazione che si sono rese disponibili negli ultimi anni, nelle quali sono derivabili informazioni sul legame tra lavoratore e impresa⁶⁴. La rilevanza del contenuto informativo dei nuovi dati forniti dall'Inps, rende questa fonte sostanzialmente esclusiva per l'aggiornamento del numero medio dei dipendenti del registro.

Solo in alcuni casi particolari si fa ancora uso di controlli interattivi per la variabile occupazione, quando la fonte UniEMens è sottostimata oppure quando non tiene conto correttamente delle date di validità in cui avvengono gli eventi strutturali che comportano un passaggio di dipendenti tra unità coinvolte in un evento.

Il primo caso si verifica in particolari sottoinsiemi di unità, dove non esiste la copertura totale di tutti i lavoratori perché non sussiste l'obbligo dei versamenti retributivi e contributivi presso l'Inps come avviene ad esempio per le imprese che operano nei settori dei trasporti marittimi e aerei. Nel secondo caso rientrano tutte le situazioni in cui le variazioni strutturali sono registrate nella fonte UniEMens alla data in cui l'evento è stato iscritto nel Registro delle imprese delle Camere di commercio, che può essere differente dalla data di efficacia dell'evento, ricavabile solo da un attento esame della visura camerale. Diventa essenziale per il revisore valutare, specialmente nel caso di fusione, quale sia la data reale di efficacia, a partire dalla quale si hanno gli effetti contabili

⁶³ La procedura di individuazione dei dati anomali e di attribuzione del numero di dipendenti validato, utilizzata nel registro per oltre un decennio fino al 2011, è stata messa a punto da Roberto Palumbo, Istat.

⁶⁴ Nel Capitolo 3 sono trattate in maniera estesa le nuove fonti ed il loro utilizzo per la stima dell'occupazione nel registro.

e fiscali dell'evento, verificando quale sia il momento in cui le poste economiche dell'impresa incorporata vengono imputate a bilancio dell'incorporante. Nella maggior parte dei casi tali effetti risultano retrodatati al primo gennaio dell'anno di riferimento, pertanto sarà modificata nello stesso modo la data inizio dell'evento e se ne dovrà tener conto nella fase finale di stima dell'occupazione (ossia al momento di ripartire i dipendenti tra le imprese coinvolte da uno stesso evento), quando lo spostamento dei dipendenti nell'impresa che incorpora sarà anch'esso anticipato al primo gennaio dell'anno di riferimento.

2.7 La quarta fase del processo di aggiornamento delle imprese: il piano di controllo e correzione

La fase di controllo e correzione dei dati consiste nell'individuazione e nel trattamento degli errori presenti nei dati, allo scopo di garantire risultati finali con determinati livelli di qualità. Il processo di verifica della qualità dei dati si è sviluppato insieme alle prime realizzazioni del registro nel 1996, come insieme di attività di tipo manuale-interattivo, in cui cioè entrambe le operazioni di individuazione e di correzione degli errori erano effettuate manualmente da esperti mediante uso di informazioni ausiliarie. Successivamente si è scelto un approccio di tipo misto, in cui le modalità di controllo e correzione interattiva sono impiegate in combinazione con quelle automatizzate. L'utilizzo di procedure informatiche in fase di individuazione degli errori e/o in fase di correzione ha prodotto da un lato una notevole riduzione dei tempi e dei costi connessi con questa fase, dall'altro un generale aumento della qualità dei risultati finali. Nell'attività di verifica di dati, infatti, le attività più costose, in termini di tempo e di risorse, sono quelle per mettere a punto gli strumenti idonei e per portare a termine le operazioni connesse al controllo e correzione dei dati. Per limitare il tempo speso nelle operazioni di controllo e correzione si è investito nell'informatizzazione del processo di verifica della qualità dei dati amministrativi già in fase di acquisizione della fonte, per ottenere un maggior controllo e un miglior livello di documentazione del processo di controllo e correzione, e soprattutto un trattamento più oggettivo ed omogeneo delle situazioni di errore. Alla parte automatizzata del processo di controllo e correzione sono affidate le operazioni di individuazione degli errori e di correzione dei casi con minore impatto sulle stime finali dei dati, mentre la fase di controllo interattivo è utilizzata per la selezione e la correzione degli errori con maggior influenza sulla struttura delle imprese. In questo modo, da un lato sono eliminate le gravi distorsioni che tali errori possono provocare nei risultati, e dall'altro è stato possibile ridurre i tempi e l'impiego di risorse connessi alla fase di controllo interattivo dei dati.

Nella quarta macro fase del processo non si esegue solo il controllo e la correzione dei dati, ma tre azioni distinte: i dati ottenuti dal processo di stima dei caratteri da fonte amministrativa sono inizialmente associati con le informazioni provenienti dalle fonti statistiche; la lista che ne deriva è sottoposta al vero e proprio processo di controllo di qualità ed eventuale correzione dei dati anomali; la lista finale di unità e caratteri riferiti al tempo t viene preparata per la sua diffusione e per l'aggiornamento del sistema informativo di Asia.

Le tre fasi sono da considerarsi distinte solo dal punto di vista concettuale, poiché, in occasione del Censimento dell'industria e dei servizi 2011, la procedura informatica alla base di ognuna di esse è stata rivista con lo scopo di definire un nuovo sistema in grado di gestire l'intera macro fase e in futuro anche le altre fasi del processo di aggiornamento del registro. Tali innovazioni di processo rientrano nel più ampio obiettivo di reingegnerizzazione dei sistemi per la costruzione e la fruizione dei registri economici, che hanno comportato anche la scelta di abbandonare l'utilizzo di diversi strumenti consolidati in favore di procedure e programmi più snelli.

Tra questi rientra il software Gestione regole per l'analisi dei dati (Granada), utilizzato negli anni precedenti la tornata censuaria per il controllo e la correzione degli accertamenti e degli errori del

registro e per il successivo aggiornamento del sistema informativo di Asia con le informazioni riferite all'anno t^{65} . Si tratta di un software appositamente predisposto all'interno dell'Istat, particolarmente adeguato per l'ottimizzazione della definizione delle regole, l'individuazione dei record anomali, l'imputazione deterministica e apprezzabile per la gestione delle correzioni, stocastiche o manuali. Il piano di controllo era stato sviluppato come un insieme di progetti del software Granada operanti come moduli separati da eseguire in sequenza, suscettibili di modifiche nella loro composizione e nel loro numero. Ogni progetto era quindi caratterizzato da un insieme di regole accomunate da una struttura simile e dall'aver uno stesso comportamento nel modificare i dati.

Il nuovo Sistema di gestione Asia (Sga), che sostituisce il precedente, è un'applicazione che facilita l'esecuzione di questa fase finale del processo di aggiornamento del registro e semplifica il controllo di gestione durante tutte le fasi, dall'acquisizione dei dati da diversi ambienti fino alla loro validazione finale⁶⁶. In generale l'applicazione, accessibile tramite web, ha consentito una riduzione dei costi e dei tempi di gestione del processo, una gestione "distribuita" delle fasi di check e soprattutto una maggiore qualità del dato finale. La procedura di gestione delle regole di controllo realizzata presenta numerosi vantaggi ed innovazioni, tra i quali il riuso del software presente in Istituto: la progettazione strutturata modularmente ha determinato un importante miglioramento delle performance in termini di giornate/uomo necessarie allo sviluppo del software, in quanto, con bassi costi, è stato utilizzato un software realizzato in altri sistemi. Solo gli algoritmi delle regole devono infatti essere riscritti di volta in volta, secondo le esigenze dei diversi processi. Dal punto di vista logico, il nuovo processo ricalca la struttura prevista nel precedente, prevedendo diversi pacchetti progettati per essere eseguiti in sequenza ma in grado di operare anche separatamente, organizzati in gruppi di regole di livelli successivi. Differentemente da quanto avveniva in passato, però, la nuova procedura consente di svolgere anche la fase d'integrazione con le fonti statistiche, oltre al controllo e la correzione dei dati, la predisposizione dei dati per la diffusione e, per ultimo, l'aggiornamento del sistema informativo di Asia con le informazioni riferite all'anno t .

Nello specifico, esistono due diverse tipologie di regole:

- le regole di *edit-imputation*, regole deterministiche costruite in modo tale che al verificarsi di particolari condizioni (*edit*) corrisponda una correzione di tipo automatico (*imputation*); sono spesso utili per evitare errori logici nella determinazione dei dati;
- le regole di compatibilità, che possono generare errori e/o accertamenti: sono usate solitamente per accertare eventuali incongruenze fra le principali variabili attribuite ad una stessa unità o tra due diversi valori attribuiti per una stessa unità in due anni diversi ($t-1$ e t).

Questi due tipi di regole svolgono funzioni diverse a seconda del livello di applicazione che occupano. Il livello di applicazione indica l'ordine nel quale applicare le regole: regole di livello di applicazione inferiore vengono applicate prima di regole con livello di applicazione superiore; ad esempio le regole con livello di applicazione pari ad 1 hanno la precedenza rispetto a quelle di livello 2.

⁶⁵ Granada è stato sviluppato da E. Riccini Margarucci, P. Floris, R. Ciacci e T. Buglielli dell'Istat.

⁶⁶ La realizzazione di Sga è parte del processo di messa in sicurezza delle procedure di aggiornamento del registro. È anche utilizzato per la procedura di stima dell'attività economica prevalente e in futuro lo sarà per la gestione del processo di aggiornamento dell'intero sistema di registri Asia. La realizzazione dei moduli finora esistenti è stata curata da Annalisa Cesaro, Laura Maglione e Lucia Toti; il modulo per la stima dell'attività economica è a cura di Michele Calabria; la manutenzione e i nuovi sviluppi Java di Sga sono a cura di Marco Passacantilli. Istat.

Nel dettaglio, è possibile individuare regole di *edit-imputation* apparentemente simili ma con differenti livelli di applicazione, pensate, quindi, per soddisfare diverse finalità. In generale, queste regole, che permettono di modificare in maniera automatica i valori di alcuni caratteri in conseguenza del verificarsi di specifiche condizioni, interessano le fasi di integrazione con fonti statistiche (fase 1) e controllo e correzione dei dati (fase 2).

A seconda del diverso livello di applicazione che occupano, è possibile individuare le seguenti regole di tipo *edit-imputation*:

- a) regole di *edit-imputation* per la determinazione del valore “finale” in fase di integrazione con le fonti statistiche (regole di scelta);
- b) regole di *edit-imputation* per l'individuazione e la correzione di diverse tipologie di errori: errori di coerenza interna delle variabili, errori di incompletezza del codice di classificazione, errori di codifica di alcuni caratteri (regole deterministiche vere e proprie);
- c) regole di *edit-imputation* da eseguire successivamente alle regole di compatibilità per la correzione di eventuali errori ricorrenti individuati dalle precedenti regole (regole di correzione);

Sono poi previsti due insiemi di regole di compatibilità a due diversi livelli di applicazione:

- regole di compatibilità vere e proprie per l'individuazione di errori e/o accertamenti (regole di errore e/o di accertamento);
- regole di compatibilità preparatorie alla fase di aggiornamento del sistema informativo di Asia per evitare errori logici nel caricamento dei dati (regole preparatorie alla fase di aggiornamento).

Entrambi gli insiemi di regole interessano la fase di controllo e correzione dei dati; tuttavia, le regole preparatorie alla fase di aggiornamento hanno un impatto significativo anche sulla successiva fase di aggiornamento del sistema informativo di Asia con le informazioni riferite all'anno t poiché individuano la presenza di situazioni irregolari che renderebbero del tutto impossibile l'aggiornamento e indicano quali siano le incoerenze da sottoporre a correzione.

Una categoria a parte è costituita dalle cosiddette regole di aggiornamento, che non ha senso classificare nelle due tipologie *edit-imputation* o compatibilità in quanto si tratta di istruzioni che permettono il corretto caricamento nel database di valori riferiti a variabili complesse, come lo stato di attività.

Di conseguenza, nell'analizzare le varie regole, è utile tener conto del diverso ruolo che sono chiamate a ricoprire, anche in considerazione della fase in cui sono applicate. Lo schema che segue sintetizza quanto detto finora evidenziando anche l'ordine di applicazione delle regole rispettato dalla procedura informatica creata *ad hoc* per la gestione della macro fase (Tavola 2.2).

Tavola 2.2 – Le regole di integrazione e controllo: tipologia e ordine di applicazione

FASE DI APPLICAZIONE	REGOLE	TIPOLOGIA DI REGOLE
1 Integrazione con fonti statistiche	Regole di scelta	<i>Edit – Imputation</i>
	Regole deterministiche vere e proprie	<i>Edit – Imputation</i>
2 Controllo e correzione dei dati	Regole di errore e/o di accertamento	Compatibilità
	Regole di correzione	<i>Edit – Imputation</i>
	Regole preparatorie alla fase di aggiornamento	Compatibilità
3 Aggiornamento del sistema informativo di Asia con le informazioni riferite all'anno t	Regole di aggiornamento	-

Per attuare la prima fase di integrazione con le fonti statistiche si utilizzano le cosiddette regole di scelta che occupano il primo livello di applicazione. Si tratta, infatti, di regole di tipo *edit-imputation* che hanno la funzione di confrontare i valori stimati di ciascun carattere (fonte amministrativa) con quelli presenti nel database (fonte statistica), riferiti allo stesso anno t , con l'obiettivo di identificare il valore da scegliere per essere sottoposto al successivo processo di controllo. In pratica, l'utilizzo di criteri di priorità fra i differenti attributi (quelli che risultano come output del processo di stima da sola fonte amministrativa e quelli eventualmente presenti per le stesse unità nel database relazionale "Asia corrente" da fonte statistica), permette di individuare il dato da considerarsi ottimale. Tali regole, quindi, portano alla determinazione dei valori che saranno in seguito sottoposti alla successiva fase di controllo e correzione per le variabili stato di attività, codice di forma giuridica, codice di attività economica, numero di occupati dipendenti medi annui, numero di occupati indipendenti. In aggiunta in questa prima fase sono anche inserite variabili da utilizzare come ausiliarie per il processo di correzione (come le stime degli anni precedenti) e tutte quelle variabili stimate per l'anno t nel processo di aggiornamento che non necessitano di integrazione con fonti statistiche, che dovranno essere preparate per la diffusione finale e per aggiornare il database per l'anno t (è il caso ad esempio del *flag* che individua le imprese artigiane o il volume affari).

Con le regole di livello 1 (regole di scelta) l'integrazione tra output da fonte amministrativa e output da fonte statistica produce un primo insieme provvisorio delle imprese attive nell'anno t completo di tutti i caratteri. In questa maniera si tiene conto di tutti i dati relativi all'anno t che sono eventualmente già presenti nel sistema informativo di Asia perché aggiornati dai revisori esperti del registro⁶⁷. Al momento dell'integrazione, infatti, per la totalità delle imprese il database risulta aggiornato con le informazioni relative all'anno di riferimento $t-1$, tuttavia per le imprese che hanno usufruito di aggiornamenti da fonti statistiche saranno eventualmente presenti caratteri riferiti anche ad anni successivi ad esso. Grazie alle regole di scelta, per ogni unità elementare sono confrontati i valori stimati di ciascun carattere con quelli presenti nel database, riferiti allo stesso anno t . Il valore da scegliere per essere inserito nella tabella di input del processo di controllo sarà quello, tra i due, che ha la priorità rispetto al processo che lo ha attribuito: le stime più recenti hanno sempre la prevalenza, ad eccezione dei casi in cui il valore presente nel database sia stato inserito online negli ultimi due anni, o provenga da archivi satellite⁶⁸. Per determinare i criteri di priorità si fa uso di alcuni metadati contenuti nel sistema informativo di Asia e in particolare della terna di variabili associata a ogni valore di ciascun carattere, adatta per fornire l'informazione sul tipo di attribuzione e sulla fonte che lo ha determinato: processo di produzione, modalità di attribuzione e provenienza.

Il risultato del confronto è sintetizzato dai valori di un indicatore che è disponibile per ogni variabile sottoposta al processo di correzione. Sono stati creati indicatori per lo stato di attività, la forma giuridica, l'attività economica principale, i dipendenti medi annui e gli indipendenti. L'indicatore assume i valori:

- "01" se il carattere è presente per la prima volta nella stima al tempo t ; negli anni precedenti nessuna informazione era presente quindi non si trova alcun valore nel database da utilizzare per il confronto;

⁶⁷ Il processo è stato descritto nel Paragrafo 2.6.

⁶⁸ Si fa riferimento in particolare a quelli del settore finanziario e assicurativo, creati utilizzando i dati provenienti dagli organi di vigilanza quali la Banca d'Italia e l'Ivass, oppure l'insieme delle holding operative, holding finanziarie e le *special purpose entities*.

- “02” se la stima al tempo t ha la priorità sul valore presente nel database e i due valori sono identici, pertanto si tratta di una conferma;
- “03” se la stima al tempo t ha la priorità sul valore presente nel database e i due valori sono diversi, pertanto si tratta di una variazione;
- “04” se il valore presente nel database ha la priorità sulla stima al tempo t in quanto inserito online da controlli interattivi;
- “05” se il valore presente nel database ha la priorità sulla stima al tempo t in quanto inserito da fonte statistica (archivi satellite, liste, eccetera);
- “06” se il valore presente nel database è l'unico presente al tempo t , in quanto non sono state fornite informazioni dalle fonti amministrative per assegnare la stima.

Durante tutte le successive fasi del processo di controllo, ogniqualvolta si modifica il valore di un carattere anche il valore del corrispondente indicatore è aggiornato in maniera coerente alle azioni che si compiono per la correzione, nell'ottica del suo utilizzo finale per l'aggiornamento del sistema informativo di Asia. Gli indicatori sono utilizzati, infatti, sia per distinguere il tipo di aggiornamento da operare per ciascun carattere, sia per assegnare i metadati relativi al tipo di processo che ne ha prodotto il valore.

La successiva fase consiste nel vero e proprio controllo e correzione, che riguarda solo le variabili di maggiore interesse, quali lo stato di attività, la forma giuridica, l'attività economica, il numero di occupati dipendenti medi annui e il numero di occupati indipendenti. L'obiettivo è quello di accertare eventuali incongruenze rispetto alle principali variabili attribuite ad una stessa unità o tra due diversi valori che, per una stessa unità, sono stati attribuiti in due anni diversi ($t-1$ e t), sottoponendo a verifica tutte le unità che compongono l'insieme costruito nella fase precedente, mediante l'applicazione di opportune regole.

Il livello 2 (regole deterministiche vere e proprie) contiene regole di tipo *edit-imputation* che vengono applicate successivamente alle regole di scelta per correggere diverse tipologie di errori:

- errori di coerenza interna delle variabili;
- errori di incompletezza del codice di classificazione;
- errori di codifica di alcuni caratteri.

In particolare, gli errori di coerenza interna delle variabili vengono individuati mediante regole che analizzano contemporaneamente i valori assunti da più variabili. Solitamente si controlla che il codice di attività economica attribuito ad un'unità sia coerente con i valori assunti per la stessa unità da altre variabili, quali lo stato di attività, il codice di forma giuridica, il numero di occupati dipendenti medi annui. La correzione automatica che ne deriva, trattandosi di regole di *edit-imputation*, è la modifica del codice di attività economica. Sono regole create appositamente per risolvere delle situazioni che si presentano ripetutamente nelle fonti amministrative che danno luogo ad assegnazioni di codici non conformi all'attività economica realmente svolta⁶⁹.

Gli errori di incompletezza del codice di classificazione, invece, si verificano quando le variabili codificate utilizzando le classificazioni ufficiali dell'Istituto assumono dei valori incompleti, che possono essere in alcuni casi completati in maniera automatica. È questo il caso, ad esempio, del codice di classificazione Ateco che dovrebbe essere composto da cinque cifre ma che talvolta, a causa di informazioni limitate delle fonti amministrative a disposizione, è possibile stimare solo a

⁶⁹ A titolo di esempio, per le imprese individuali a cui è stato assegnato il codice attività economica corrispondente alla “fabbricazione di autoveicoli” (29100), che risultano attive per l'anno di riferimento t impiegando meno di cinque addetti, si corregge il codice automaticamente in “riparazioni meccaniche di autoveicoli” (45201); in questo caso le variabili coinvolte sono il codice di attività economica, sul quale agisce la regola, la forma giuridica, lo stato di attività e il numero di addetti medi annui.

livello di classe, gruppo o divisione, rispettivamente di quattro, tre, due cifre. In caso si disponga, per la stessa unità, di una stima completa per l'anno precedente $t-1$ le cui cifre iniziali sono proprio le stesse del codice incompleto, si utilizza la stima $t-1$ per completare le cifre mancanti nell'anno t . Le regole che individuano gli errori di codifica di alcuni caratteri servono a modificare i valori di una o più variabili che siano stati definiti in modo non appropriato. Si utilizzano, ad esempio, per assegnare il codice di forma giuridica completo fino alla quarta cifra a tutte le imprese individuali, per le quali, infatti, il processo che utilizza le informazioni delle fonti amministrative per assegnare il codice può solo attribuire la divisione "1.1 - Imprenditore individuale, libero professionista e lavoratore autonomo"⁷⁰. Solo al termine della fase di stima degli altri caratteri si dispone delle informazioni necessarie a discernere tra le forme giuridiche tipiche dell'attività d'impresa in forma individuale, che sono specificate in questa fase sotto forma di regole deterministiche. Le regole di scelta tengono conto della registrazione del codice fiscale nell'archivio delle Camere di commercio, della presenza o meno di dipendenti, dell'eventuale carattere artigiano dell'impresa e della tipologia di attività economica svolta, così da distinguere se si tratta di un "Imprenditore individuale agricolo" (1.1.10), un "Imprenditore individuale non agricolo" (1.1.20), un "Libero professionista" (1.1.30) o un "Lavoratore autonomo" (1.1.40).

Il successivo livello 3 (regole di errore e/o di accertamento) contiene regole di compatibilità, che individuano gli accertamenti e gli errori veri e propri.

Gli accertamenti sono controlli da effettuare su insiemi di dati da tenere sotto osservazione, oppure riguardo a situazioni anomale per le quali occorre verificare se si tratta di valori reali o errori. Rientrano nel primo caso le regole che individuano le unità di alcune fasce dimensionali rilevanti che presentano una variazione di uno o più attributi (principalmente l'attività economica) da un anno all'altro (da $t-1$ a t). Gli insiemi così definiti necessitano di una analisi approfondita da parte di revisori esperti per determinare quale sia la reale attività economica svolta nell'anno t . Una impresa che svolge un'attività economica che necessita di un grande numero di addetti (ad esempio, "Estrazione di carbone (esclusa torba)", oppure "Estrazione di petrolio greggio e di gas naturale)", in presenza di un numero esiguo di questi (nel caso precedente, fino a venti addetti) è, invece, un esempio di situazione anomala. In questi casi, l'intervento del revisore può portare ad una conferma o ad una correzione dell'attività economica svolta dall'impresa.

Differentemente da quanto previsto per gli accertamenti, gli errori devono essere necessariamente corretti. La rettifica può essere di tipo puntuale da parte di revisori esperti, con l'apertura di una istruttoria per grandi imprese o unità rilevanti, o di tipo automatico; la correzione automatica è possibile per il trattamento di casistiche simili, una volta individuate *ad hoc* una o più regole di correzione, ossia di *edit-imputation*. Logicamente quindi, le regole di correzione hanno un livello di applicazione successivo a quelle che individuano l'errore (Tavola 2.2).

Le regole di errore permettono di individuare diverse tipologie di errore da sottoporre a controllo e a successiva correzione; riguardano il controllo:

- dell'esattezza formale dei codici di classificazione;
- della completezza dei caratteri;
- della compatibilità tra valori della stessa variabile in due anni diversi;
- della compatibilità tra variabili di una stessa impresa;
- della correttezza di codici di attività economiche "delicate", tramite liste di imprese.

In particolare, sono previste regole che permettono il controllo di esattezza formale dei codici di classificazione della forma giuridica e dell'attività economica; in pratica si tratta di appurare che le

⁷⁰ Il processo di assegnazione della forma giuridica è trattato nel Paragrafo 2.5.4.

variabili in questione assumano solo modalità appartenenti alle relative classificazioni ufficiali. In caso contrario, i valori devono essere corretti. Anche la completezza dei caratteri è verificata tramite apposite regole. Un esempio ricorrente è quello del codice di attività economica che dovrebbe essere composto da cinque cifre. Nei casi in cui tale codice è incompleto a livello di divisione (due cifre), gruppo (tre cifre) o classe (quattro cifre), nei casi in cui le regole deterministiche del passo precedente non sono riuscite a determinare le cifre mancanti in maniera automatica, apposite regole segnalano l'errore che necessariamente dovrà essere corretto, manualmente da parte di revisori esperti, in caso di unità rilevanti, o automaticamente per tutte le imprese attive del registro⁷¹.

Per quanto riguarda la verifica della compatibilità tra i valori di una stessa variabile in due anni diversi, sono previste regole che evidenziano le trasformazioni teoricamente non ammissibili da sottoporre ad esame. È il caso delle variabili relative all'attività economica e alla forma giuridica. Ad esempio, alcune specifiche regole individuano le situazioni in cui la forma giuridica stimata per l'anno t è incompatibile con il codice attribuito nell'anno $t-1$ come nel caso di società di persone o di capitali che si trasformano in enti privati o forme disciplinate dal diritto pubblico, o viceversa.

Spesso tali errori riguardano anche il carattere stato di attività, che riveste una grande importanza soprattutto per le conseguenze sulla demografia delle imprese. La sua importanza fa sì che il tipo di correzione da effettuare sia, spesso, di tipo manuale, da effettuarsi a cura di revisori esperti che siano in grado di valutare con esattezza i valori delle rispettive date di validità. Un'attenzione particolare è riservata agli accertamenti di riattivazione delle imprese: si verificano quando per un'impresa è stimato uno stato di attività "attivo" nell'anno di riferimento t , mentre negli anni precedenti era stata registrata la sua cessazione, avvenuta con data precedente a quella di inizio del nuovo stato. In questi casi gli accertamenti si differenziano per le imprese individuali e fanno uso di informazioni ausiliarie esterne al processo, quali le dichiarazioni di cambio di partita Iva.

Come le regole deterministiche, anche le regole di errore individuano errori di compatibilità tra i valori assunti da più variabili per una stessa impresa. Se però le regole deterministiche sono in grado di modificare in modo automatico alcuni di questi valori, le regole di errore necessitano di un'analisi più approfondita. Nello specifico, sono state previste delle regole che verificano la coerenza tra il codice fiscale di un'impresa e la sua forma giuridica, tra carattere artigiano e attività economica svolta, tra quest'ultima e il numero di dipendenti.

Alcune regole di errore sono definite prendendo in considerazione archivi satellite o liste di imprese da utilizzare per il controllo dell'attività economica svolta da alcune particolari unità. Rilevanti sono le regole che controllano la sovra-copertura del settore finanziario. In particolare, per questo settore sono utilizzati degli archivi satellite la cui fonte di provenienza ha un elevato grado di attendibilità, anche per il controllo degli altri caratteri principali. L'insieme di regole riguarda in particolar modo le banche e le assicurazioni, ma anche altre tipologie di imprese finanziarie, per le quali è stato realizzato un archivio settoriale di supporto al registro Asia. Per questi settori di particolare interesse, infatti, sono raccolte informazioni presso gli enti, le amministrazioni o le associazioni preposti al controllo delle imprese operanti nel settore di pertinenza. Le caratteristiche di esaustività delle informazioni sono garantite solo per l'archivio delle banche, le cui informazioni sono fornite dalla Banca d'Italia e per quello delle assicurazioni, fornito dall'Ivass, Istituto per la vigilanza sulle assicurazioni, con l'esclusione delle imprese di assicurazione estere che hanno in

⁷¹ Per la correzione di tipo automatico e per quella interattiva è stata predisposta una applicazione che permette sia al revisore di visualizzare e aggiornare in modo interattivo i codici incompleti, sia di eseguire il completamento automatico utilizzando un algoritmo che assegna le cifre mancanti in maniera proporzionale alla distribuzione dei codici completi nelle imprese attive. La sua realizzazione è stata curata da Annalisa Cesaro e Marco Passacantilli, Istat.

Italia una loro rappresentanza. Queste ultime sono da considerarsi imprese a tutti gli effetti che, per la parte che opera sul territorio nazionale, sono da inserire nel registro Asia. Il controllo di copertura di Asia relativo a questi particolari settori è gestito operando delle integrazioni specifiche fra i singoli archivi settoriali e l'output standard del processo di costruzione del registro Asia. I revisori sono ampiamente coinvolti in questa attività.

Il successivo livello 4 (regole di correzione) contiene regole di tipo *edit-imputation* utilizzate quando, tramite le regole di compatibilità del livello precedente, si individuano errori ricorrenti che è possibile correggere in modo automatico. Sono inizialmente utilizzate per risolvere situazioni che si presentano in un particolare anno, ma in caso si ripresentino anche nei successivi, si possono utilizzare come test per predisporre nuove regole deterministiche, da applicare l'anno successivo al livello iniziale 2, per ampliare l'insieme delle correzioni automatiche. In generale, tutte le correzioni e le verifiche effettuate con riferimento ad uno specifico anno aggiornano l'archivio delle correzioni, da utilizzare negli anni successivi. In questo modo, grazie al suo continuo sviluppo, la fase di correzione puntuale può essere considerevolmente ridotta con il passare degli anni.

L'ultimo livello contiene le regole preparatorie alla fase di aggiornamento del sistema informativo di Asia, successiva a quella di controllo. Sono utili per evitare errori logici nel caricamento dei dati, ad esempio per controllare e successivamente correggere eventuali errori di coerenza nelle date di validità dei caratteri. La maggior parte di queste regole riguarda la variabile stato di attività e ha lo scopo di mettere in luce i casi che, non essendo stati corretti in precedenza, risultano impossibili da aggiornare. Un esempio concreto è la situazione di un'impresa stimata attiva nell'anno t ma per la quale è stata anche stimata una data di inizio attività successiva al 31 dicembre dello stesso anno t .

Al termine della fase di controllo e correzione dei dati si ottiene la lista definitiva delle imprese attive con riferimento all'anno t , con la quale si aggiorna integralmente il database relazionale di Asia. A questo punto, infatti, si dispone di una lista di imprese aggiornata non solo con i valori corretti e definitivi per l'anno t ma anche con tutti i metadati utili alla realizzazione della fase successiva di aggiornamento e caricamento dei dati nel sistema informativo di Asia, tra cui fondamentali sono:

- il processo che ha determinato il dato valido, per distinguere se è stato ottenuto come risultato del processo di integrazione di archivi amministrativi, oppure successivamente da archivi statistici o aggiornato online da controlli interattivi;
- la modalità di attribuzione del valore, se attribuito come stima, o da correzione di check o come informazioni acquisite da indagini dell'Istituto;
- la provenienza dell'informazione, ossia la fonte dell'informazione (il particolare archivio amministrativo, la specifica indagine, eccetera).

Nella terza e ultima fase (regole di aggiornamento) sono eseguite le regole che creano opportuni indicatori di caricamento che, per ogni carattere e per ogni unità giuridica, forniscono istruzioni che saranno interpretate dal successivo programma di aggiornamento del database storico, così che è compito dello statistico di volta in volta indicare la necessaria azione da compiere, sfruttando tutte le funzioni elementari già predisposte.

Rispetto ai sei valori dell'indicatore descritto in precedenza, sono quindi inseriti nel database solo i valori in aggiornamento, ossia i valori di quei caratteri il cui indicatore assume valori "01", "02" e "03". Nello specifico:

- in presenza di un indicatore con valore "01", saranno inseriti nel database il valore del carattere, il processo che lo ha determinato, la sua modalità di attribuzione, la provenienza e la data di riferimento del processo;
- in caso di indicatore pari a "02", è necessario aggiornare solo la data di riferimento del processo, così da segnalare che il valore è confermato anche dal processo dell'anno in corso, e la terna di variabili processo, modalità di attribuzione e provenienza;
- se l'indicatore assume valore "03", il valore del carattere deve essere modificato con il nuovo valore, a partire dalla sua data di inizio validità, a cui saranno associate le tre variabili processo, modalità di attribuzione e provenienza e la relativa data di riferimento del processo.

Di conseguenza, non sono aggiornati i valori dei caratteri a cui è associato un indicatore con valore "04", "05" oppure "06" in quanto il valore già presente nel database è quello che sarà diffuso con la lista delle attive nell'anno *t*.

Nel database così aggiornato ogni carattere del registro è referenziato. Per ogni unità del registro e per ogni suo attributo sono specificati: il processo che lo ha assegnato, la modalità di attribuzione, la provenienza del dato, le sue date di validità, di aggiornamento/conferma, e quelle di riferimento.

Uno strumento utile per il monitoraggio e per un'attenta e puntuale gestione del processo di controllo è costituito dai rapporti riassuntivi. Nella macro area del nuovo software Sga a essi dedicata, si hanno a disposizione report di sintesi che permettono di analizzare vari aspetti, quali l'esito delle regole, l'andamento del processo di correzione, il lavoro degli operatori e soprattutto i risultati della fase di controllo e correzione. Al termine di ogni ciclo di check si ha a disposizione un innovativo riepilogo dinamico che contiene diversi livelli che si possono espandere a piacimento, fino ad arrivare all'estrazione delle liste di imprese i cui dati soddisfano una particolare regola. Inoltre tutti i prospetti e i relativi microdati in essi contenuti possono essere esportati nei formati *Excel* in modo da poter essere elaborati anche localmente sul proprio personal computer.

2.8 Il campo di osservazione Cis e le diverse modalità di utilizzo del registro

Nel 2011 il registro Asia ha rappresentato il contenuto informativo strutturale del Censimento, che per questo è stato definito censimento virtuale. Infatti i dati pubblicati relativi alla numerosità delle unità, alla loro forma giuridica, attività economica, localizzazione, occupazione sono derivati direttamente dal registro Asia e non da una rilevazione esaustiva. Con l'occasione sono state introdotte in Asia importanti innovazioni nel processo di aggiornamento del registro e nella stima dei caratteri delle imprese, dal punto di vista definitivo e metodologico, utili ad aumentare la qualità ed ampliare la quantità delle informazioni diffuse, con un significativo miglioramento nella tempistica di diffusione dei dati. A partire da tale data le innovazioni di processo e di prodotto introdotte sono replicate annualmente e costituiscono la struttura informativa del cosiddetto Censimento continuo delle imprese: le nuove informazioni sono disponibili nel registro ogni anno e diffuse a partire dal 2012 in aggiunta alle tradizionali variabili del registro, con una pubblicazione apposita⁷². Tali novità garantiscono una migliore coerenza dell'intero sistema informativo delle statistiche economiche e della Contabilità Nazionale e una migliore coerenza con gli standard definiti in ambito Ue, adottando un *framework* concettuale comune e metodi armonizzati di stima.

⁷² I principali elementi innovativi introdotti, grazie anche alla disponibilità di nuove fonti amministrative, sono relativi alle descrizioni delle diverse tipologie di occupazione con cui le imprese utilizzano il fattore lavoro. Tale argomento è approfondito nel Capitolo 3 del presente fascicolo.

La pubblicazione del registro Asia come dato censuario (censimento virtuale) ha richiesto specifiche rielaborazioni dei dati del registro poiché le due popolazioni presentano specifiche differenze, nel campo di osservazione e nel riferimento temporale dei dati (data di "rilevazione").

Il campo di osservazione storico del censimento differisce da quello del registro per vari motivi.

Con riferimento all'attività economica delle imprese, nei dati censuari sono incluse unità che svolgono alcune attività economiche nella Sezione A: "Agricoltura, silvicoltura e pesca" dell'Ateco 2007, che sono invece assenti nel campo di osservazione del registro Asia-imprese, e che non erano incluse nel 6° Censimento dell'agricoltura 2010. Pertanto nei dati censuari, rispetto a quelli presenti nel registro, si trovano anche le unità che svolgono:

- attività di supporto all'agricoltura e attività successive alla raccolta, classificate nel gruppo 01.6 della Nace Rev. 2, ad esclusione delle attività di conservazione del territorio agricolo al fine di mantenerlo in buone condizioni agricole ed ecologiche e della manutenzione del terreno al fine di mantenerlo in buone condizioni ambientali per uso agricolo;
- attività di caccia, cattura di animali e servizi connessi classificate nel gruppo 01.7;
- attività di silvicoltura ed utilizzo di aree forestali classificate nella divisione 02;
- pesca e acquacoltura classificate nella divisione 03.

Con riferimento alla forma giuridica delle imprese, invece, le cooperative sociali, presenti come imprese nel registro Asia, sono state censite come istituzioni non profit; si è preferito somministrare loro un questionario che cogliesse le peculiarità proprie del settore non profit rispetto alla sua struttura organizzativa, l'attività svolta e i servizi resi.

Non va dimenticato, infine, che lo scostamento fra il Censimento dell'industria e dei servizi e il registro statistico dipende anche dal loro diverso riferimento temporale. Il registro, aggiornato annualmente, considera attive le unità che nel corso dell'intero anno di riferimento hanno dato un segnale di attività. Il Censimento, invece, fornisce un'istantanea della situazione alla data di riferimento della rilevazione. Per definizione, dunque, le unità del 9° Censimento generale dell'industria e dei servizi attive al 31 dicembre 2011 costituiscono un sottoinsieme delle unità attive secondo Asia nel 2011.

Le modalità di diffusione dei dati del registro per gli utenti esterni all'Istat, fino all'anno censuario, erano di più tipi:

- dati elementari per tutti gli uffici di statistica del Sistema statistico nazionale, come stabilito dalla normativa vigente; il richiedente, responsabile dell'ufficio di statistica o titolare/responsabile del trattamento, per gli enti e gli organismi pubblici di informazione statistica, può richiedere la lista delle imprese attive nell'anno t per le sole finalità di attuazione del Programma statistico nazionale oppure per l'esecuzione di trattamenti connessi all'attività istituzionale o all'ambito territoriale del richiedente; la fornitura di dati corredati di identificativi è limitata ai casi di assoluta e stretta necessità, per impossibilità di raggiungere l'obiettivo prefissato senza i dati identificativi, nel qual caso il sottoscrittore della richiesta assume piena responsabilità;
- dati aggregati tramite note per la stampa e tavole standard inserite nel sito web dell'Istat⁷³;
- elaborazioni *ad hoc* su richiesta degli utenti, acquisita tramite il *Cont@ct Centre* dell'Istituto⁷⁴;

⁷³ Tali prodotti sono consultabili sul sito Istat, in particolare negli argomenti *Imprese* e *Industria e costruzioni*. A titolo di esempio, i dati diffusi delle imprese attive nel 2012, <http://www.istat.it/it/archivio/131578>.

⁷⁴ Il *Cont@ct Centre* è il servizio a cui rivolgersi per richiedere dati, pubblicazioni e file di microdati accessibile tramite il sito web dell'Istituto <https://contact.istat.it/> previa registrazione.

- rapporto di qualità per Eurostat, sotto forma di questionario annuale, per adempiere agli obblighi comunitari, in modo tale che Eurostat possa verificare la conformità al Regolamento (CE) n. 177/2008 e monitorare l'avanzamento del lavoro di armonizzazione delle statistiche europee. A seguito delle disposizioni di cui all'art. 9 del suddetto regolamento, il *Business Registers Annual Inquiry* misura ogni anno la qualità dei registri e identifica le problematiche che gli Stati membri possono incontrare nell'attuazione del regolamento. Sulla base delle risposte al questionario, viene poi redatto un rapporto che Eurostat presenta ogni anno al *Working Group on Business Registers and Statistical Units*⁷⁵. I risultati sono utilizzati per monitorare la situazione dei registri europei, i loro contenuti, la loro copertura in termini di unità e variabili e il trattamento dei dati, per lavorare nella direzione di registri europei armonizzati. Il questionario, che si articola in diverse parti, nelle prime raccoglie per sezione di attività economica, secondo la classificazione Nace Rev. 2, informazioni quantitative, in termini di unità, addetti occupati e volume d'affari, relative a tutte le unità statistiche contenute nel registro (unità giuridiche, imprese, unità locali e gruppi d'impresa, distintamente per i gruppi composti da unità giuridiche residenti sul territorio nazionale e gruppi multinazionali che hanno almeno due unità giuridiche localizzate in paesi diversi). In aggiunta, in una sezione dedicata, Eurostat raccoglie informazioni sulle fonti amministrative utilizzate per l'identificazione delle unità e l'assegnazione dei rispettivi caratteri. Infine, per ciascuna tipologia di unità è verificata la completezza dei caratteri obbligatori e di quelli opzionali; nel caso in cui una variabile non fosse disponibile è richiesta la precisa motivazione insieme ad una previsione della sua eventuale futura introduzione nel registro.

Con la redazione del piano di diffusione di tutti i dati rilevati al Censimento dell'industria e dei servizi 2011 è avvenuta una significativa evoluzione anche nel dettaglio delle informazioni diffuse da Asia annualmente, anche per gli anni successivi al 2011. Infatti, nel rispetto della normativa nazionale in materia di segreto statistico e di protezione dei dati personali, nelle elaborazioni di dati diffuse da Asia fino a tale data era valida la limitazione circa il numero minimo unità per cella (almeno tre), come per tutte le informazioni normalmente diffuse dall'Istat⁷⁶. In considerazione del fatto che le informazioni di Asia risultano analoghe a quelle raccolte con le operazioni censuarie e che per queste, dato il loro carattere e la importanza per le analisi economiche, non vigeva tale limitazione⁷⁷, l'Istat si è fatto promotore di una iniziativa presso il Garante per la tutela dei dati personali al fine di ottenere per le informazioni prodotte da Asia annualmente la stessa deroga valida per il Censimento dell'industria e dei servizi. Il garante ha concordato sulla proposta, come testimoniato dal Psn 2008-2010⁷⁸. Inoltre, l'Istat ha implementato una ulteriore modalità di

⁷⁵ Il gruppo di lavoro è formato da responsabili dei registri di tutti i paesi europei e tiene riunioni annuali sulle problematiche connesse con l'attuazione del regolamento europeo sui registri e per garantire comuni interpretazioni delle definizioni delle unità statistiche.

⁷⁶ D.Lgs 6 settembre 1989, n. 322. Inoltre l'Istat opera in conformità al Codice di deontologia e di buona condotta per i trattamenti di dati personali per scopi statistici e scientifici (all. A.4 del D.Lgs. 30 giugno 2003, n. 196 - Codice in materia di protezione dei dati personali) e aderisce al Codice di condotta per le statistiche europee, adottato come Raccomandazione della Commissione nell'ambito della Comunicazione al Consiglio e al Parlamento Europeo del 25 maggio 2005 sull'indipendenza, integrità e responsabilità delle autorità statistiche nazionali e dell'autorità statistica comunitaria.

⁷⁷ D.L. 31 maggio 2010, n. 78 convertito con modificazioni in L. 30 luglio 2010, n. 122, art.50, comma 2, c).

⁷⁸ Dal testo del Psn 2008-2010, il garante afferma che "Un secondo gruppo di progetti per i quali si chiede la diffusione in forma disaggregata riguarda il settore 'Struttura e competitività delle imprese' e in particolare gli archivi 'Asia-imprese (IST-00806)' e 'Asia-unità locali, realizzazione con utilizzo di indagini dirette e fonti amministrative (IST-01685)'. Tale proposta nasce dalla constatazione che le informazioni presenti nei suddetti archivi costituiscono un

diffusione delle informazioni presenti in Asia, che consiste nella realizzazione annuale di un *data warehouse*, mediante il sistema *corporate* dell'istituto *I.stat*, così come realizzato per il censimento. È evidente che questo passo è di fondamentale importanza per un utilizzo adeguato alle esigenze delle diverse tipologie di utilizzatori, in particolare il mondo della ricerca e i *policy makers*.

Pertanto, analogamente a quanto fatto per i dati censuari 2011, è stata realizzata una banca dati online dalla quale l'utenza può estrarre l'informazione desiderata a partire dall'anno di riferimento 2012, accessibile dal sito dell'Istat, che consente la costruzione di tavole statistiche personalizzate attraverso una interrogazione guidata, con possibilità di scaricare sul proprio PC l'informazione visualizzata⁷⁹. In generale è possibile costruire vari incroci tra le principali informazioni del registro, ottenendo dati disaggregati fino al livello comunale per attività economica, classe di addetti, forma giuridica, carattere artigiano e altre variabili.

La possibilità di ottenere annualmente informazioni esaustive sulla struttura produttiva delle imprese dell'industria e dei servizi evitando di svolgere tradizionali censimenti economici, quale è stata l'esperienza del Censimento del 2011, testimonia la ricchezza del contenuto delle fonti amministrative e il risparmio in termini economici e di fastidio statistico che un utilizzo appropriato a fini statistici di tale patrimonio consente.

aggiornamento dei dati di fonte censuaria relativi alla struttura economica ed occupazionale, la cui conoscenza è stata ritenuta di particolare interesse dal legislatore che, con l'art. 37, comma 2, lett. e), della L. 17 maggio 1999, n. 144, esplicitamente ne ha autorizzato la diffusione fino al dettaglio comunale, anche nei casi di frequenza inferiore a tre unità."

⁷⁹ *I.Stat* è la banca dati delle statistiche correntemente prodotte dall'Istituto nazionale di statistica, accessibile al link <http://dati.istat.it/>. I dati sono organizzati in modo coerente e omogeneo, presentati sotto forma di tavole multidimensionali che gli utenti possono esportare in formato *xls*, *csv* e *sdmx*. È inoltre possibile creare tabelle e grafici personalizzati agendo sulle variabili, il periodo di riferimento e la disposizione di testate e fiancate. Al suo interno le statistiche sono articolate per tema e ogni tema è suddiviso in più argomenti. I dati del registro sono contenuti nell'argomento *Struttura* all'interno del tema *Imprese*.

Capitolo 3

L'occupazione di Asia: il nuovo database Asia-occupazione⁸⁰

3.1 Il framework concettuale e la revisione delle definizioni

3.1.1 Le definizioni ILO-SNA e il framework concettuale di riferimento utilizzato

La realizzazione del Censimento “virtuale” dell'industria e dei servizi si è basata sull'utilizzo di nuove e più precise fonti amministrative soprattutto per le stime dell'occupazione. Il processo si è realizzato attraverso un nuovo approccio logico (e nuove metodologie) che individua l'esatta relazione che esiste fra singolo occupato, l'unità (e luogo) in cui lavora, la tipologia di contratto, il periodo di occupazione in quella unità; approccio noto in letteratura come LEED (*Linked Employer Employee Database*). Questo nuovo approccio oltre a permettere la produzione di una quantità maggiore di informazioni (es. quelle sulle caratteristiche dei lavoratori), riduce i rischi di duplicazioni e introduce maggiore dettaglio nell'informazione statistica diffusa. Inoltre rende possibile l'individuazione di criteri operativi atti ad individuare con maggiore precisione gli “individui” che lavorano effettivamente in una impresa isolando, ad esempio, i soci lavoratori percettori di una forma di reddito misto (che remunera lavoro e capitale) dai restanti soci percettori esclusivamente di reddito da capitale e che quindi non apportano lavoro all'impresa.

Infine in occasione del Censimento è stata realizzata una revisione complessiva del sistema classificatorio dell'occupazione al fine di garantire maggiore coerenza per tutto il Sistema statistico nazionale. Tale nuovo sistema è stato applicato alle stime dell'occupazione del Cis2011, e di conseguenza, alle successive stime annuali, provvisorie e definitive, del sistema dei registri economici Asia, che individua, a sua volta, l'universo delle unità e dell'occupazione a cui devono convergere le stime delle indagini sulle imprese, a loro volta base per le stime di Contabilità Nazionale. Per questa ragione il sistema definitorio e classificatorio è stato costruito per avere le proprietà, da un lato, di completezza, dall'altro di coerenza con le strutture classificatorie di riferimento per le stime dell'occupazione dell'indagine sulle Forze di lavoro (che utilizza il sistema *International Classification of Status in Employment – ICSE93*) o per le stime dei conti nazionali (che utilizzano a livello internazionale il *System of National Accounts - SNA2008*⁸¹) e il connesso Sistema europeo dei conti economici (*Sec/ESA2010*⁸²).

Un sistema completo, implica una esaustiva definizione e classificazione di tutte le componenti dell'occupazione. Gli esempi più rilevanti sono l'individuazione e quantificazione di tutte le figure di lavoro così detto “parasubordinato” (non solo Co.Co.Co e Co.Co.Pro con o senza partita Iva, ma

⁸⁰ A cura di Danila Filippini e Giuseppe Garofalo. Autori del capitolo:

- Paragrafo 3.1: Giuseppe Garofalo;
- Paragrafo 3.2: Danila Filippini;
- Paragrafo 3.3: Ciro Baldi, Davide Di Cecco, Danila Filippini e Silvia Pacini (Paragrafi 3.3.1, 3.3.2, e 3.3.3), Grazia Di Bella (Paragrafo 3.3.4);
- Paragrafo 3.4: Caterina Viviano (Paragrafi 3.4, 3.4.1, 3.4.2 e 3.4.5), Irene Rocchetti (Paragrafo 3.4.3), Carla Troccoli (Paragrafo 3.4.4);
- Paragrafo 3.5: Francesca Cuppone e Roberta Rizzi (Paragrafi 3.5.1, 3.5.2, 3.5.3 e 3.5.4), Davide Di Cecco e Roberta Rizzi (Paragrafo 3.5.5);
- Paragrafo 3.6: Caterina Viviano.

⁸¹ European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations and World Bank, *System of National Accounts 2008*, New York, 2009.

⁸² Regulation (EU) No 549/2013 of the European Parliament and of the Council of 21 May 2013 on the European system of national and regional accounts in the European Union.

anche, ad esempio, i contratti con voucher o gli amministratori/manager) e del lavoro somministrato (ex interinale). Il fine ultimo è quello di rappresentare in maniera compiuta il complesso dell'input di lavoro - e le sue componenti - di una impresa, di un territorio, di un settore produttivo.

Un sistema coerente non implica una uguale allocazione di tutte le figure occupazionali in tipologie omogenee per tutti i domini statistici (es. classificare alcune figure sempre come dipendenti o indipendenti, nei dati sia di Contabilità Nazionale – CN – sia di SBS⁸³). La coerenza è stata invece ricercata in un processo in cui l'esatta identificazione delle figure e loro relativa quantificazione sia realizzata, in termini di definizioni e metodologie di stima, avendo presente le differenti esigenze dei diversi domini statistici (continuando l'esempio in CN e SBS). In concreto, ciò è stato realizzato con un approccio di *building blocks* ovvero attraverso una sufficiente disaggregazione che permetta l'identificazione esatta di tutte le figure occupazionali rilevanti per l'insieme del Sistema statistico nazionale, al fine di produrre aggregazioni differenti per differenti domini statistici. In questa maniera, se classificare differenti figure in differenti tipologie per differenti output statistici può avere un senso da un punto di vista concettuale, viste le differenti richieste dei regolamenti internazionali, le differenze esistenti possono essere sempre esattamente individuate, garantendo qualità a tutto il sistema statistico e maggiore trasparenza nei confronti con l'utenza.

Il framework concettuale utilizzato

La necessità di produrre un sistema di definizioni (e regole operative) che non sia solo limitato al Censimento, ma che coinvolga tutto il sistema delle statistiche economiche, ha reso necessario ancorare tale sistema a riferimenti concettuali di base sufficientemente generali, in particolare: l'*European System of Account* – ESA2010 e l'*International Classification of Status in Employment* – ICSE93.

La scelta del primo riferimento nasce dal fatto che l'ESA rappresenta il "... quadro centrale di riferimento per le statistiche sociali ed economiche dell'Unione Europea e dei suoi Stati membri"⁸⁴. Per quanto riguarda l'ICSE: "*The ICSE classifies jobs held by persons at a point in time. A job is classified with respect to the type of explicit or implicit contract of employment of the person with other persons or organizations. The basic criteria used to define the groups of the classification are the type of economic risk, an element of which is the strength of the attachment between the person and the job, and the type of authority over establishments and other workers which the job incumbents have or will have*"⁸⁵.

Accanto a queste due basi classificatorie si è tenuto conto delle definizioni adottate dal Regolamento (CE) n. 250/2009 sulle *Structural Business Statistics* (SBS).

Le definizioni presenti nei riferimenti internazionali sono state ricondotte a tre tipologie: *employees*, *self-employed persons*, *grey area*.

Employees: le definizioni adottate sono sostanzialmente simili. Questa tipologia di occupazione è caratterizzata dai seguenti elementi:

⁸³ I risultati economici delle imprese sono prodotti conformemente a quanto disposto dal Regolamento Ue n. 295/2008 per le statistiche strutturali (SBS - *Structural Business Statistics*). Il regolamento SBS definisce un quadro comune per la raccolta, l'elaborazione e la trasmissione dei dati allo scopo di disporre, annualmente, di statistiche armonizzate per valutare la struttura, l'attività e la competitività delle imprese nell'Unione europea.

⁸⁴ ESA2010, Capitolo 1 "Caratteristiche generali e principi di base", punto 1.05.

⁸⁵ *Resolution concerning the International Classification of Status in Employment (ICSE), adopted by the Fifteenth International Conference of Labour Statisticians (January 1993).*

- 1) sono *“those persons who work for an employer”*. *“The relationship of employer to employee exists when there is an agreement, which may be formal or informal, between the employer and a person”*;
- 2) esiste *“an explicit (written or oral) or implicit contract of employment”*;
- 3) il contratto deve essere *“on a continuous basis: implies a period of employment which is longer than a specified minimum determined according to national circumstances”*;
- 4) essi *“receive remuneration in cash or in kind or their labour”*.

Il concetto di remunerazione *“in cash or in kind”* presente in ESA e SBS, sembra più ampio di quello previsto dall'ICSE in cui ci si riferisce in maniera esplicita al *“payroll”*. Infatti, ad esempio, in ESA e SBS rientrano in questa tipologia gli *“students who have a formal commitment”* o possono rientrare gli *“owners of corporations and quasi-corporations”*. L'ICSE non fa riferimento agli studenti, mentre per gli *“owner-managers”* non specifica come debbano essere classificati. Inoltre sia l'ESA sia l'SBS escludono in maniera esplicita i lavoratori volontari, non citati in ICSE.

Self-employed persons: maggiori differenze fra ESA e ICSE si evidenziano per questa tipologia (nel regolamento SBS non viene direttamente identificata ma rientra nella definizione di *“employment”*). In particolare:

- 1) l'ESA esplicitamente li identifica come le *“persons who are the sole or joint owners of the unincorporated enterprises”*;
- 2) sono caratterizzati dal fatto che *“the compensation for self-employment is included in mixed income because it is not possible to observe separately the return to labour from the return to any capital used in the unincorporated enterprise”*;
- 3) escludono *“those unincorporated enterprises that are classified as quasi corporations”*. In particolare si escludono i soci delle società di persone e di capitale;
- 4) l'ICSE, al contrario, include tra i *self-employed*, tutti i proprietari delle *unincorporated enterprises*, non identificando quelle che possono essere definite come *quasi corporations*.

Un esempio richiamato per questa differenza è quello dei *“members of producers' cooperatives”* classificati come *self-employed* per l'ICSE ma come *employees* per l'ESA poiché: *“the remuneration of these people is included in compensation of employees rather than in mixed income”*.

Grey area: rientrano in questa tipologia gli *outworkers* (Co.Co.Pro e figure simili) e i *owner-managers* delle *corporations*.

Per quanto riguarda gli *outworkers*, ESA e ICSE propongono la stessa definizione (il regolamento SBS non considera queste tipologie di occupazione):

- 1) *“An outworker is a person who agrees to work for a particular enterprise or to supply a certain quantity of goods and services to a particular enterprise by prior arrangement or contract with that enterprise”*;
- 2) *“but whose place of work is not within any of the establishments that make up that enterprise”*.

Una condizione vincolante nell'individuazione di questa tipologia di occupazione è che: *“il luogo di lavoro non sia all'interno di una unità locale dell'impresa con la quale è in essere un contratto”*. Tale vincolo è legato alla necessità che l'*outworker* abbia costi di produzione (affitto locali, attrezzature) e che questi quindi non siano a totale carico dell'impresa. Nello stesso tempo svolgere l'attività all'interno di una unità locale dell'impresa implica che il lavoratore sia sottoposto

ad un controllo sull'attività svolta. Sembra quindi che le tipologie di lavoro caratterizzate dall'assenza della condizione 2) debbano essere classificate come "dipendenti".

Inoltre tali figure, se non lavorano all'interno di una unità locale dell'impresa con la quale è in essere un contratto, possono comunque possedere sia le caratteristiche di dipendenti sia quelle di lavoratori in proprio: "*they may be classified as being in 'paid employment' or in 'self-employment' according to the specific terms of their contract*". Il criterio di classificazione è legato a come il lavoro sia remunerato. Se il reddito dell'*outworker* deriva da remunerazione dell'ammontare del lavoro effettuato all'interno del processo produttivo dell'impresa, il lavoratore deve essere classificato tra gli *employees*; se il reddito è in funzione del valore dell'output prodotto deve essere classificato tra gli *own-account workers*.

Sia l'ESA sia l'ICSE mettono in evidenza la difficoltà oggettiva nella classificazione di questa tipologia di occupazione, con il suggerimento di individuarli in maniera separata dalle altre tipologie.

Per quanto riguarda i manager delle *corporations* mentre per SNA essi sono classificati come *employees*, l'ICSE (che parla per la precisione di *owner-managers*) non specifica come debbano essere classificati.

Occupazione e residenza

L'occupazione identificata dal Censimento virtuale 2011 fa riferimento a quella impiegata dalle unità economiche residenti sul territorio nazionale. L'ESA2010 specifica come debbano essere considerate come residenti tutte quelle unità che, indipendentemente dalla nazionalità e dalla natura giuridica, abbiano sul territorio nazionale un centro di interesse economico⁸⁶.

Per centro di interesse economico si intende l'esistenza di un luogo, all'interno del territorio economico di un paese, in cui una unità esercita attività economica in misura significativa per un periodo di tempo indeterminato o per un periodo limitato ma relativamente lungo (un anno o più)⁸⁷.

L'occupazione definita fa quindi riferimento a tutti gli individui, residenti o non residenti in Italia, che lavorano in un'unità definite come residenti, quindi comprende anche⁸⁸:

- a) i lavoratori frontalieri non residenti;
- b) i lavoratori stagionali non residenti;
- c) i cittadini appartenenti al personale di basi scientifiche nazionali insediate fuori dal territorio economico;
- d) i membri di equipaggi di navi, aeromobili e piattaforme galleggianti gestite da unità residenti.

3.1.2 Le nuove definizioni dell'occupazione

Le definizioni adottate nell'ambito del Cis2011, e poi nelle successive versioni del sistema di registri Asia, fanno riferimento al *framework* concettuale descritto precedentemente. In alcuni casi, ad esempio gli indipendenti, è stato necessario, però, identificare una declinazione puntuale che tenesse conto delle specificità delle realtà italiane. In altri casi, ad esempio gli *outworkers*, sono state prese decisioni *a priori* nella classificazione per considerare la realistica applicabilità delle definizioni internazionali.

Le definizioni adottate specificano due tipologie di occupazione, il personale direttamente addetto ad una impresa e il personale esterno ad esso che però opera nel processo di produzione

⁸⁶ ESA2010, paragrafi 1.61 e 2.04.

⁸⁷ ESA2010, paragrafo 2.07.

⁸⁸ ESA2010, paragrafo 11.17.

dell'impresa stessa. L'insieme di queste tipologie determina il lavoro totale impiegato da una impresa.

Personale addetto di una impresa:

I concetti di base utilizzati per l'identificazione di questa tipologia di occupazione di una impresa sono:

- a) *remuneration: reward for employment in the form of pay, salary, or wage, including allowances, benefits (such as company car, medical plan, pension plan), bonuses, cash incentives, and monetary value of the noncash incentives*⁸⁹;
- b) la remunerazione deve essere relativa a *"amount of labour that is contributed as an input into some process of production, irrespective of the value of the output produced or the profitability of the production process"*⁹⁰;
- c) *mixed income: ... has characteristics of wages and salaries, and characteristics of profit due to work carried out as an entrepreneur*⁹¹.

Il personale addetto è suddiviso in dipendente ed indipendente.

Dipendenti: personale che lavora, con vincoli di subordinazione, per conto di una impresa in forza di un contratto di lavoro, esplicito o implicito, e che percepisce per il lavoro effettuato una remunerazione in forma di salario, stipendio, onorario, gratifica, pagamento a cottimo o remunerazione in natura. Rientrano fra i dipendenti di una impresa:

- a) i soci di una impresa (anche cooperativa) se iscritti nelle gestioni ordinarie dell'Inps (compreso il Dmag);
- b) i dirigenti, i quadri, gli impiegati e gli operai, a tempo pieno o parziale con contratto a tempo indeterminato o determinato;
- c) i lavoratori a domicilio;
- d) i lavoratori stagionali;
- e) gli apprendisti;
- f) i lavoratori con contratto di formazione lavoro o di inserimento;
- g) i lavoratori con contratto di lavoro ripartito;
- h) gli stagisti e i tirocinanti. Cioè giovani che, assolto o meno il ciclo di studi, svolgono una attività, solo parzialmente remunerata con danaro, in cambio di formazione.

I lavoratori somministrati o intermittenti sono classificati come dipendenti della società fornitrice di lavoro.

Sono considerati lavoratori dipendenti anche le persone temporaneamente non al lavoro purché mantengano un legame formale con la loro posizione lavorativa. Tale legame è determinato secondo uno o più dei seguenti criteri:

- a) la percezione in via continuativa di un salario o di uno stipendio;
- b) la garanzia di riprendere il lavoro al termine del periodo di inattività o l'esistenza di un accordo circa la data della ripresa del lavoro;
- c) il periodo di assenza dal lavoro trascorso che, quantunque rilevante, può corrispondere al periodo per il quale i lavoratori hanno diritto a percepire una indennità senza essere obbligati ad accettare altri lavori.

Si tratta di persone temporaneamente non al lavoro per motivi di malattia o infortunio, ferie o vacanze, sciopero o serrata, assenza per ragioni di studio o di formazione, maternità o congedo

⁸⁹ <http://www.businessdictionary.com/definition>.

⁹⁰ ESA2010, paragrafo 8.19.

⁹¹ SNA2008, paragrafo 7.35.

parentale, riduzione dell'attività economica (in Cassa integrazione guadagni ordinaria o straordinaria – anche in deroga), sospensione temporanea del lavoro per vari motivi, quali maltempo, guasti tecnici o elettrici, mancanza di materie prime.

Indipendenti: personale che svolge la propria attività lavorativa in una impresa, senza vincoli formali di subordinazione, e la cui remunerazione abbia natura di reddito misto (capitale/lavoro).

Rientrano fra gli indipendenti:

- a) gli imprenditori individuali, i liberi professionisti e i lavoratori autonomi;
- b) i familiari coadiuvanti se prestano lavoro nell'impresa senza il corrispettivo di una prefissata retribuzione contrattuale e versano i contributi per le assicurazioni previdenziali o di infortunio sul lavoro;
- c) i professionisti, con o senza partita Iva individuale, che partecipano, indipendentemente dalla quota di partecipazione, a studi associati;
- d) i soci delle società di persone o di capitali (comprese le cooperative) se non iscritti nella Gestione ordinaria dell'Inps (compreso il Dmag) e a condizione che effettivamente lavorino nella società. In questo caso il socio, proprio per il suo ruolo, svolge l'attività in autonomia avendo la responsabilità dell'organizzazione e della gestione dell'impresa e percepisce un flusso di reddito che remunera, senza possibilità di distinzione, il lavoro svolto, il capitale investito, il rischio d'impresa, cioè un cosiddetto reddito misto.

Personale esterno di una impresa

Rientrano nel personale esterno di una impresa tutti gli individui, non classificati come personale addetto, che per mezzo di un contratto volontario, di tipo formale o informale, prestano servizi in maniera continuativa per l'impresa stessa. Rientrano in tale tipologia i collaboratori e i manager con o senza partita Iva (non soci dell'impresa), che hanno le caratteristiche di lavoratori in proprio in quanto il loro reddito dipende dal valore dei beni e dei servizi forniti all'impresa.

Nel personale esterno rientrano anche i lavoratori somministrati e utilizzati dall'impresa.

Lavoro totale impiegato

Il lavoro totale impiegato da una impresa è definito come l'ammontare complessivo del lavoro effettivamente prestato in maniera continuativa in una impresa per la produzione di beni e servizi. Rientrano:

- a) il personale addetto all'impresa con esclusione dei dipendenti comandati in altra impresa;
- b) i dipendenti comandati da altra impresa;
- c) il personale esterno di una impresa.

3.1.3 Gli indipendenti: la declinazione della definizione

Con riferimento alla coerenza con gli schemi definitivi internazionali non esistono particolari evidenze per quello che riguarda i dipendenti. Maggiori problemi si presentano per gli indipendenti a causa delle specifiche peculiarità della situazione italiana.

Come precedentemente evidenziato l'ESA definisce indipendenti le persone che lavorano per se stesse, ovvero i titolari/soci di un'impresa non distinta dalla persona fisica (o dalle persone fisiche) che la possiede. In sostanza il lavoro indipendente viene individuato in quelle unità produttive in cui il proprietario/socio in virtù del suo ruolo svolge l'attività in autonomia, avendo la responsabilità dell'organizzazione e gestione dell'impresa e percepisce un flusso di reddito che remunera, senza possibilità di distinzione, il lavoro svolto, il capitale investito, il rischio d'impresa, cioè un cosiddetto "reddito misto" che è dipendente dal risultato dell'attività produttiva e non dalla quantità di lavoro prestato. Da questa definizione generale discende che la figura dell'indipendente è individuata solo laddove il titolare/socio lavora nell'impresa e il suo lavoro non è remunerato né come

salario/stipendio (in tal caso rientrerebbe tra i dipendenti), né come prestazione di un servizio reso da terzi all'impresa (in tal caso la remunerazione del servizio rientrerebbe nei costi intermedi dell'impresa e non nel suo valore aggiunto e il lavoratore sarebbe quindi esterno all'impresa) bensì viene remunerato dal "reddito misto".

Due sono gli elementi della revisione - concettuale ed operativa - della definizione di addetti indipendenti introdotta a partire dal Censimento del 2011.

Il primo elemento è stato quello di considerare la peculiarità della struttura produttiva italiana che presenta un alto numero di imprese di piccola e media dimensione tra le società di persone e di capitali. Per tali imprese, in alcuni casi anche senza dipendenti, è molto diffusa la presenza di soci che lavorano nell'impresa stessa e che pertanto presentano tutte le caratteristiche del lavoro indipendente.

Il secondo elemento è legato alla necessità di superare l'approccio, considerato non più adeguato, che si basava sul concetto di "funzione di coordinamento e gestione", secondo il quale tale funzione deve sempre essere presente in una impresa. La conseguenza di tale approccio è stata quella di imputare, nei censimenti, nelle statistiche strutturali sulle imprese e nel registro Asia, almeno un addetto indipendente per tutte le imprese. Tale approccio ha introdotto nel corso degli anni delle distorsioni nelle stime, come, ad esempio, l'imputazione che coinvolge imprese (quali società a controllo pubblico, imprese con tutti soci persone giuridiche, eccetera) che per loro caratteristica non possono possedere addetti indipendenti.

Questi due elementi comportano due rilevanti conseguenze:

- 1) è possibile la presenza di lavoro indipendente anche nelle società di capitali;
- 2) non tutte le imprese devono avere come input di lavoro quello definito indipendente.

Sulla base di quanto precedentemente evidenziato la scelta operativa fatta a livello censuario è stata quella di declinare regole specifiche per specifiche tipologie di impresa:

- a) per le imprese individuali la figura dell'indipendente corrisponde al titolare dell'impresa; qualora questi si avvalga di prestazioni da parte di parenti o affini, alla figura dell'indipendente si aggiungono quelle dei familiari coadiuvanti;
- b) in imprese con forma giuridica di consorzio, ente pubblico economico, o in *branch* di imprese estere, in coerenza con le regole applicate già a partire dal 2001 dalla Contabilità Nazionale, non è ammissibile lavoro indipendente.

Più complessa risulta essere l'individuazione di lavoro indipendente nelle società di persone e di capitali. L'inserimento fra gli indipendenti dei soci di società è in contraddizione con quanto definito dall'ESA2010 dove si specifica come i *self-employed* "are persons who are the sole or joint owners of the unincorporated enterprises" e pertanto possono essere presenti solo nelle unità produttive classificabili nel settore delle famiglie. È importante sottolineare che questa scelta è condizionata dalle caratteristiche delle unità produttive operanti nel nostro paese e influenza l'intero Sistema statistico nazionale.

- a) Nelle società di persone (S.a.s., S.n.c.) prevale l'elemento soggettivo, rappresentato dai soci, rispetto al capitale. Esse, per la legislazione italiana, non sono persone giuridiche, sono dotate di autonomia patrimoniale imperfetta, i soci rispondono con il loro patrimonio delle obbligazioni assunte dalla società. Per queste loro caratteristiche, a metà tra l'impresa individuale e le società di capitali, la presenza del lavoro indipendente è del tutto plausibile e quindi esso deve essere stimato per queste tipologie di imprese;
- b) le società di capitali sono definite tali in quanto l'elemento del capitale ha una prevalenza concettuale e normativa rispetto all'elemento soggettivo rappresentato dai soci. Sono caratterizzate da avere personalità giuridica e una autonomia patrimoniale perfetta. La

responsabilità dei soci per le obbligazioni sociali è limitata. Infine il potere di amministrazione è svincolato dalla qualità di socio. Per queste tipologie di società la presenza dell'occupazione indipendente è ipotizzabile solo in presenza del *mixed income* evidenziabile da elementi certi che identifichino l'effettivo lavoro effettuato dal socio nella società.

3.1.4 La classificazione degli outworkers: la scelta operativa effettuata

Come messo in evidenza nel Paragrafo 3.1.1, l'area grigia è identificata da tutti quei lavoratori che possono avere le caratteristiche sia di indipendente che di dipendente. Dal punto di vista concettuale la classificazione, in una o nell'altra tipologia di lavoro, può essere determinata sulla base di alcune caratteristiche del contratto che intercorre fra l'*employer* e l'*outworker*, ad esempio: il livello di autonomia, il possesso dei mezzi di produzione, il luogo in cui si esplica l'attività.

La scelta operativa effettuata in occasione del Censimento dell'industria e dei servizi 2011 è stata quella di classificare tale tipologia di lavoro, indentificato a livello nazionale dai cosiddetti lavoratori parasubordinati, nel personale esterno di una impresa.

In Italia sono due le tipologie di lavoro che rientrano nel concetto di *outworkers*:

- a) i professionisti che offrono prestazioni remunerate con fattura (titolari di partita Iva);
- b) i parasubordinati, di cui fanno parte - fra gli altri - i titolari di contratti di collaborazione coordinata e continuativa a progetto e i collaboratori occasionali.

Per quanto riguarda i professionisti titolari di partita Iva, nell'ambito del Censimento (e quindi sia nel registro Asia sia nelle statistiche strutturali d'impresa) questi sono considerati titolari di un'impresa e pertanto figurano nelle statistiche sulle imprese così come nei conti nazionali come indipendenti della propria impresa. In quanto tali, rientrano quindi nel conteggio nel numero complessivo degli addetti.

Per quanto riguarda i parasubordinati, invece, si rileva un differente trattamento di questi lavoratori nelle statistiche sulle imprese rispetto a quanto avviene nelle stime di Contabilità Nazionale. L'apertura di una posizione contributiva come parasubordinato non costituisce di per sé elemento sufficiente per dare vita ad una impresa. Nell'ambito delle statistiche sulle imprese quindi il collaboratore viene classificato come personale esterno all'impresa e sfugge quindi al conteggio nel numero complessivo degli addetti.

Per contro la Contabilità Nazionale, seguendo l'approccio della misura dell'input di lavoro complessivo, è svincolata dall'individuazione e misurazione del numero di imprese mentre è piuttosto orientata alla misura di tutte le unità istituzionali (incluse le famiglie nella loro veste di produttori). In quest'ottica la CN include i collaboratori nelle proprie stime, classificandoli come indipendenti della propria unità istituzionale classificata nell'attività economica che meglio rappresenta l'attività svolta dal collaboratore stesso (e non in quella dell'impresa committente).

La scelta di classificare il lavoro parasubordinato come "lavoro esterno" nasce quindi dalla necessità sia di mantenere coerenza con quanto operato nell'ambito delle statistiche strutturali delle imprese, sia di produrre informazioni utili ai conti nazionali per operare in maniera corretta la loro classificazione.

Resta il fatto che la possibilità di riclassificare l'input di lavoro rappresenta, in ogni caso, una risorsa importantissima e un fondamentale avanzamento nella comprensione della struttura ed evoluzione dei principali comparti dell'economia italiana sotto un profilo analitico e sperimentale: avere a disposizione la possibilità di simulare l'internalizzazione di una parte dei subordinati può aprire, tra l'altro, alla possibilità di proporre misurazioni alternative dell'input di lavoro e degli indicatori su di esso basati (anzitutto la cosiddetta produttività, comunque la si definisca).

3.2 La nuova struttura informativa sui lavoratori e occupazione

La disponibilità di nuove fonti amministrative in concomitanza con la tornata censuaria del 2011, in aggiunta all'opportunità di revisionare ed armonizzare le definizioni sull'occupazione del sistema statistico, ha consentito una revisione complessiva dei metodi di stima dell'occupazione al fine di produrre informazioni più dettagliate sulla caratteristiche dei lavoratori, identificare le diverse tipologie di lavoratori e risolvere problemi di misurazione dell'occupazione. In questo contesto l'Istat ha sviluppato una base informativa sull'occupazione, denominata Asia-occupazione, con una struttura del tipo *Linked Employer Employee Database* (LEED).

In generale, un archivio LEED è una base informativa che lega i dati relativi ai due lati del mercato del lavoro: le imprese (o più in dettaglio il luogo di lavoro) e gli individui. Per far ciò sono identificate e messe in relazione tre differenti unità statistiche:

- gli individui presenti nella popolazione delle famiglie e degli individui;
- le unità locali presenti nella popolazione delle imprese e delle unità locali;
- i rapporti di lavoro presenti nella popolazione dei rapporti di lavoro.

Tale struttura dati permette analisi congiunte sulle caratteristiche dei lavoratori, delle imprese e sul tipo di rapporto di lavoro esistente tra l'impresa e il lavoratore.

Esistono due modelli di LEED: (i) longitudinali e (ii) *cross-section*. I LEED longitudinali sono usualmente costruiti utilizzando informazioni di natura amministrativa. Una struttura LEED di tipo longitudinale permette di ottenere informazioni longitudinali e quindi di analizzare e valutare nel tempo l'impatto delle politiche del lavoro sulle imprese e sui lavoratori. I LEED *cross-section* sono invece basati su dati di indagine. Essendo costruiti a partire da dati campionari contengono un numero ridotto di unità, ma sono ricchi di covariate caratterizzanti il rapporto di lavoro. Solitamente i LEED *cross-section* contengono informazioni dettagliate dal lato impresa (ad esempio politiche del lavoro) e dal lato individuo (ad esempio attitudini del lavoratore). Molti paesi del nord Europa come Norvegia, Danimarca, Germania e Francia hanno sviluppato LEED *cross-section* e longitudinali, integrando le informazioni di natura amministrativa con quelle di indagine. Questa struttura informativa permette di valutare l'impatto delle politiche del lavoro sfruttando un ampio numero di covariate.

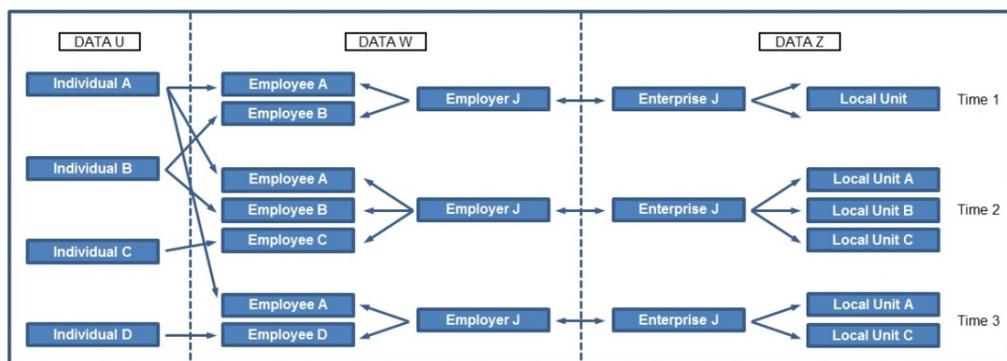
Utilizzando dati amministrativi, principalmente di natura previdenziale, le informazioni anagrafiche sugli individui e il registro delle imprese Asia, l'Istat ha sviluppato un LEED di tipo longitudinale con la seguente struttura informativa. Si definiscono con U , Z e W tre matrici di dati dove:

- U contiene i dati sulle caratteristiche degli N individui della popolazione. Una riga di U denotata con $[\mathbf{u}_i]_{i \in \{1, 2, \dots, N\}}$ contiene le caratteristiche anagrafiche dell'individuo i e si assume che siano invarianti nel tempo;
- Z contiene i dati sulle caratteristiche delle M imprese per periodo t . Una riga di Z denotata con $[\mathbf{z}_{jt}]_{j \in \{1, 2, \dots, M\}, t \in \{t_{j1}, \dots, t_{jT_j}\}}$ contiene le caratteristiche dell'impresa j nell'intervallo di tempo t , dove T_j è il numero totale di intervalli per l'impresa j con le caratteristiche \mathbf{z}_{jt} ;
- W contiene i dati sulle caratteristiche dei rapporti di lavoro esistenti per l'individuo i nell'intervallo di tempo t . Una riga di W denotata con $[\mathbf{w}_{ikt}]_{i \in \{1, 2, \dots, N\}, k \in \{1, \dots, k_i\}, t \in \{t_{ik1}, \dots, t_{ikT_i}\}}$ contiene le caratteristiche del rapporto di lavoro k per il lavoratore i per il periodo t , dove k_i è il numero totale di rapporti di lavoro per il lavoratore i nel periodo t e T_j è il numero totale di intervalli per il rapporto di lavoro k del lavoratore i col le caratteristiche \mathbf{w}_{ikt} .

I dati U e W sono abbinati tramite un identificativo dell'individuo, mentre i dati Z e W sono abbinati tramite un identificativo dell'impresa. Il *link* tra i dati Z e U può essere concettualizzato tramite una

funzione $link\ j = J(i; t)$ dove j indica le imprese presso le quali il lavoratore lavora nell'intervallo di tempo t . La Figura 3.1 rappresenta la struttura del LEED e le relazioni tra i dati U , Z e W .

Figura 3.1 – Struttura del LEED



Nello specifico di Asia-occupazione le caratteristiche degli individui presenti nella matrice $[U]$ sono ad ora quelle demografiche e invariante nel tempo, quali sesso, età e luogo di nascita.

Le caratteristiche delle imprese presenti nella matrice $[Z]$ sono quelle presenti in Asia-imprese e Asia-unità locali e sono attività economica, forma giuridica, classe di fatturato, età dell'impresa, appartenenza a gruppi di impresa, localizzazione. Queste caratteristiche possono variare nel tempo e l'intervallo di validità dell'informazione è l'anno coerentemente con il periodo di aggiornamento di Asia.

Le caratteristiche dei rapporti di lavoro esistenti per l'individuo e presenti nella matrice $[W]$ sono quelle, opportunamente trattate, presenti nelle fonti amministrative utilizzate per la realizzazione di Asia-occupazione. I domini di interesse sono: tipologia di lavoratori (indipendenti, dipendenti, lavoratori esterni), qualifica, regime orario, tipo di contratto, beneficiari di sgravi contributivi e classe di compenso. Il carattere longitudinale di Asia-occupazione deriva dal dettaglio temporale con cui viene fornito ogni rapporto di lavoro in essere nell'anno. In particolare, l'unità temporale di Asia-occupazione è la "settimana nel mese", ovvero la settimana, o parte di settimana, inclusa in un mese. La scelta di adottare questa struttura temporale è stata dettata sia dalla volontà di sfruttare al meglio il dettaglio minimo disponibile in tutte le fonti amministrative utilizzate, sia dalla convenienza computazionale. La convenienza computazionale deriva dal fatto che la "settimana nel mese" permette di gestire agevolmente anche l'unità temporale mese, rispetto alla quale sono riferite molte delle variabili disponibili d'interesse. Insieme all'informazione settimanale è inoltre riportato il numero di giorni lavorativi per la settimana t nell'anno di riferimento.

Le informazioni e le relazioni tra le matrici U , Z e W permettono facilmente di:

- misurare il numero di posizioni lavorative in un preciso punto nel tempo (ad esempio alla data censuaria), oppure il numero di posizioni lavorative medie in un determinato intervallo di tempo (ad esempio posizioni medie nel mese o nell'anno), per domini caratterizzanti non solo l'impresa ma anche il lavoratore e il rapporto di lavoro;
- misurare il numero di transizioni di posizioni lavorative in un determinato intervallo di tempo per stato di origine e stato di destinazione, dove gli stati possono essere descritti da caratteristiche dell'impresa, dell'individuo o del rapporto di lavoro.

Per calcolare le misure descritte è utile definire l'esistenza di una posizione lavorativa dell'individuo i nell'impresa j nell'intervallo di tempo t come:

$$l_{ijt} = 1_{\{j \in J(i;t)\}}$$

dove $1_{\{j \in J(i;t)\}}$ è una funzione indicatrice che assume valore 1 se $j \in J(i;t)$, cioè se j appartiene all'insieme delle imprese presso le quali il lavoratore i lavora nell'intervallo di tempo t , e 0 altrimenti. L'intervallo tempo t è la settimana nel mese. Inoltre per ciascuna settimana è preso in considerazione un peso g_t pari al numero di giorni lavorativi nella settimana t nell'anno di riferimento. A partire dalla posizione lavorativa è possibile definire diverse misure dell'occupazione:

- 1) numero totale di lavoratori nell'impresa j nella settimana t :

$$l_{.jt} = \sum_{i=1}^N l_{ijt}$$

- 2) numero medio di lavoratori nell'impresa j nell'intervallo $[t_1 - t_T]$:

$$\bar{l}_{.jt} = \sum_{t=t_1}^{t_T} g_t l_{.jt} / \sum_{t=t_1}^{t_T} g_t$$

- 3) numero totale di posizioni lavorative per l'individuo i al tempo t :

$$l_{i.t} = \sum_{j=1}^M l_{ijt}$$

Altri esempi di misure calcolabili a partire da Asia-occupazione sono misure di transizione come "Employer-to-Employer transition" oppure "jobs destruction/jobs creation". A tal fine è utile definire le seguenti funzioni:

$$EEl_{ijkt} = 1_{\{l_{ik(t_1)}=1 \& l_{ij(t_T)}=0 \& l_{ik(t_1)}=0 \& l_{ik(t_T)}=1\}}$$

che individua l'individuo i che si sposta dall'impresa j nell'impresa k nell'intervallo di tempo $[t_1-t_T]$ (se esistono posizioni lavorative multiple nella posizione di origine o in quella di destinazione è necessario identificare il flusso primario);

$$Dl_{ijt} = 1_{\{l_{ij(t_1)}=1 \& \sum_{k=1}^M l_{ik(t_T)}=0\}}$$

$$Cl_{ijt} = 1_{\{\sum_{k=1}^M l_{ik(t_1)}=0 \& l_{ij(t_T)}=1\}}$$

che individuano rispettivamente distruzione e creazione di lavoro nell'intervallo di tempo $[t_1-t_T]$ per l'individuo i nell'impresa j .

Misure di “*Employer-to-Employer transition*” e “*jobs destruction/jobs creation*” sono:

- 1) Numero di transizioni dall'impresa j all'impresa k nell'intervallo $[t_1-t_T]$:

$$EEL_{.jkT} = \sum_{i=1}^N EEL_{ijkT}$$

- 2) Numero di transizioni tra imprese nell'intervallo $[t_1-t_T]$:

$$EEL_{..T} = \sum_{j=1}^M \sum_{\substack{k=1 \\ K \neq j}}^M EEL_{.jkT}$$

- 3) Numero di posti di lavoro persi nell'impresa j nell'intervallo $[t_1-t_T]$:

$$DL_{.jT} = \sum_{i=1}^N DL_{ijT}$$

- 4) Numero totale di posti di lavoro persi nell'intervallo $[t_1-t_T]$:

$$DL_{..T} = \sum_{j=1}^M DL_{.jT}$$

- 5) Numero di posti di lavoro creati nell'impresa j nell'intervallo $[t_1-t_T]$:

$$Cl_{.jT} = \sum_{i=1}^N Cl_{ijT}$$

- 6) Numero totale di posti di lavoro creati nell'intervallo $[t_1-t_T]$:

$$Cl_{..T} = \sum_{j=1}^M Cl_{.jT}$$

3.3 La metodologia per la realizzazione di Asia-occupazione: gli occupati dipendenti

Il registro Asia-occupazione è stato realizzato e sarà aggiornato attraverso un processo di integrazione di informazioni provenienti da archivi amministrativi e fonti statistiche. Centrale in questo processo produttivo è l'identificazione dell'unità statistica rapporto di lavoro tra un individuo i e un'impresa j e la stima dei relativi attributi.

La metodologia sviluppata prevede l'utilizzo di differenti fonti amministrative e differenti trattamenti dei dati a seconda della tipologia di rapporto di lavoro, dove le differenti tipologie di rapporto di lavoro sono: indipendenti, dipendenti, lavoratori esterni.

In questo paragrafo sono descritte le fonti amministrative, la loro integrazione e l'armonizzazione utilizzate per la stima dei rapporti di lavoro dipendenti, mentre la stima dei rapporti di lavoro indipendente ed esterni è discussa rispettivamente nei Paragrafi 3.4 e 3.5.

3.3.1 Le fonti amministrative utilizzate e la loro armonizzazione

Le fonti utilizzate per l'identificazione delle unità statistiche rapporto di lavoro sono le fonti amministrative di carattere contributivo e fiscale con informazioni sul singolo lavoratore in relazione

all'impresa datore di lavoro. Per la stima dei rapporti di lavoro dipendente le fonti principali sono quelle Inps che coprono la maggior parte dei lavoratori dipendenti del settore privato.

A partire dal 1 gennaio 2010, tutte le imprese con dipendenti devono consegnare all'Inps mensilmente una dichiarazione, denominata UniEMens, con tutte le informazioni retributive e contributive a livello di singolo lavoratore, oltre che di impresa. L'introduzione dell'UniEMens ha unificato due distinti moduli e procedure di acquisizione delle informazioni:

- le dichiarazioni individuali EMens, relative ai dati retributivi dei singoli lavoratori;
- il modello aziendale DM10, relativo ai dati contributivi aziendali (matricole Inps).

La trasmissione dei flussi informativi, attraverso il nuovo modello di denuncia telematica, riguarda i lavoratori dipendenti, nonché i lavoratori iscritti alla Gestione separata e gli associati in partecipazione. L'invio è mensile e deve avvenire entro l'ultimo giorno del mese successivo a quello di competenza.

Questa transizione, pur rappresentando una occasione per l'Istat dal momento che si dispone di informazioni a livello di singolo lavoratore, ha determinato un break strutturale nel processo di Asia-imprese che ha utilizzato, fino al 2010, il DM10 come fonte principale per l'aggiornamento delle posizioni lavorative (numero di addetti dipendenti).

Dal 2010 l'Inps fornisce all'Istat un'estrazione di informazioni retributive, dalle dichiarazioni mensili UniEMens corrispondente sostanzialmente, alla ex dichiarazione EMens. Le unità presenti nell'UniEMens sono univocamente individuate dalle seguenti variabili (variabili chiave):

- Identificativo del lavoratore e dell'impresa - codice fiscale del lavoratore, codice fiscale dell'impresa;
- Identificativo del mese;
- Qualifica del lavoratore, tipo di contratto (tempo determinato/indeterminato), regime orario (full time/part time);
- Tipo di contribuzione.

Per ciascuna unità sono inoltre presenti le seguenti variabili di interesse:

- Giorno Assunzione: giorno del mese in cui è intervenuta l'assunzione o la ripresa di attività;
- Giorno Cessazione: giorno del mese in cui è intervenuta cessazione o la sospensione di attività;
- Comune: codice catastale del comune luogo di lavoro;
- Imponibile: imponibile previdenziale complessivo del mese;
- Giorni retribuiti: numero dei giorni retribuiti nel mese per i lavoratori full time;
- Settimane utili: per i lavoratori a tempo parziale e per i lavoratori "intermittenti", viene indicato il numero di ore complessivamente retribuite nel mese espresso in settimane equivalenti di un full time (determinate dividendo il numero di ore per l'orario contrattuale settimanale del corrispettivo lavoratore a tempo pieno);
- Tipo Copertura: un vettore che indica la copertura retributiva per ciascuna settimana (da domenica a sabato o frazione) del mese cui si riferisce la denuncia. Per ciascuna settimana la variabile può assumere i valori: X - settimana totalmente retribuita, 1 - settimana totalmente NON retribuita a causa di eventi tutelati dalla legge, 2 - settimana parzialmente retribuita a causa di eventi tutelati dalla legge, 0 o *blank* - Nessuna copertura;
- Codice evento: per le settimane in cui non sia dovuta la retribuzione o in cui la retribuzione abbia subito una riduzione a causa di eventi tutelati dalla legge (codici 1 o 2 del Tipo Copertura) è indicato il tipo evento (es. maternità, malattia, Cassa integrazione, eccetera).

Il Tipo Copertura e il Codice evento sono alla base dell'algoritmo di calcolo delle posizioni lavorative totali (dipendenti). È da sottolineare che i dati a disposizione forniscono un record per ciascun contratto di lavoro dipendente che un individuo iscritto all'Inps ha avuto nell'anno di riferimento, e che un individuo può avere più contratti con la stessa impresa o con più imprese. Per il dettaglio temporale di ciascuna fornitura UniEMens valgono le seguenti osservazioni:

- la fornitura è relativa ad un anno specifico – cambiamenti nelle leggi vigenti e nell'organizzazione delle fonti amministrative possono cambiare il significato e le modalità di alcune variabili da un anno all'altro;
- per ogni rapporto di lavoro in essere nell'anno viene specificata l'estensione temporale con un dettaglio pari alla settimana nel mese. La definizione di settimana segue quella dell'Inps per la quale il primo giorno è domenica, l'ultimo è il sabato. Per settimana nel mese si intende l'intera settimana o parte di essa inclusa nel mese di riferimento per un totale massimo di sei settimane al mese e 72 settimane nell'anno. Per ogni settimana, o parte di settimana, inclusa in ciascun mese la variabile Tipo Copertura definisce l'assenza/presenza del rapporto;
- alcune variabili (quali imponibile o numero di giorni retribuiti) sono relative ad un mese specifico. Possono quindi assumere fino a 12 valori per ciascun rapporto di lavoro.

Pur essendo la fonte largamente predominante sull'occupazione alle dipendenze delle imprese nel settore privato, l'UniEMens non permette ancora una copertura esaustiva. Per raggiungere questo obiettivo è stato necessario integrare le informazioni dell'UniEMens con altre fonti: l'archivio della Cassa integrazione guadagni (Cig) a pagamento diretto, quello delle dichiarazioni Dmag per i lavoratori agricoli, quello delle dichiarazioni ex Enpals per i lavoratori dello spettacolo.

Il fenomeno della Cassa integrazione infatti, che dal 2008 con l'insorgere della crisi economica è divenuto particolarmente rilevante, viene registrato solo in parte nell'UniEMens ossia relativamente alla Cig cosiddetta "a conguaglio". Si tratta dei casi in cui l'azienda anticipa il pagamento della Cig per conto dell'Inps e, dopo aver ricevuto l'autorizzazione, chiede il rimborso delle somme anticipate. In questi casi quando l'impresa ottiene l'autorizzazione della Cig registra le informazioni nell'UniEMens segnalando il mese di competenza relativo, il tipo di copertura settimanale e l'evento. Nessuna informazione viene dichiarata invece nell'UniEMens in caso di Cig il cui pagamento viene effettuato direttamente dall'Inps al lavoratore. In tali situazioni si distinguono due casi: Cig non zero ore e Cig a zero ore. Nel primo caso l'impresa continua a mandare il relativo flusso UniEMens ma senza alcuna indicazione relativa alla Cig. Nel caso di Cig a zero ore, invece, quando il lavoratore viene temporaneamente sospeso dall'attività lavorativa, l'azienda deve sospendere anche l'invio dell'UniEMens per questi lavoratori. In tutti e due i casi l'impresa invia all'Inps attraverso un altro flusso mensile (modello SR41) le informazioni utili al calcolo della prestazione di Cig e, in buona parte, coincidenti con quelle previste, ma omesse, dal flusso UniEMens⁹². Tali informazioni mancanti sono state pertanto recuperate e integrate nei dati utilizzati a fini statistici per il corretto computo dei lavoratori dipendenti. Il file della Cig a pagamento diretto contiene informazioni su:

- anno e mese di competenza;
- codice fiscale del lavoratore e matricola aziendale;

⁹² Il flusso UniEMens è in continua evoluzione e le modalità di denuncia della Cassa integrazione guadagni stanno subendo delle modifiche finalizzate a far confluire nell'unico flusso UniEMens tutte le informazioni, riducendo così possibilità di errori e incongruenze. Qui è stata descritta la situazione al momento del trattamento delle informazioni disponibili e utilizzate per la produzione di Asia e del Cis relativi al 2011.

- numero di ore di Cig fruite (oltre al tipo di Cig, ossia ordinaria, straordinaria, in deroga, solidarietà).

Mentre in tutti i settori gli obblighi di invio della denuncia contributiva all'Inps sono assolti mediante l'UniEMens, i datori di lavoro agricoli sono obbligati a inviare i dati per gli operai mediante una dichiarazione trimestrale utilizzando il modello Dmag Unico. La contribuzione dovuta all'Inps per Cassa integrazione, Cassa integrazione straordinaria, e mobilità per gli operai a tempo indeterminato dipendenti da cooperative che trasformano e commercializzano i prodotti agricoli e zootecnici propri segue le regole ordinarie (UniEMens).

Il modello Dmag è composta di due parti:

- quadro 1: Azienda - comprende le informazioni relative alla posizione anagrafica e contributiva dell'azienda;
- quadro 2: Dati anagrafici e retributivi dei lavoratori - riepiloga tutti i dati relativi alla manodopera occupata nel trimestre di competenza.

La struttura del modello è molto complessa e sono stati necessari trattamenti al fine di ottenere una informazione armonizzata con quella presente nell'UniEMens. In particolare, il dettaglio con cui viene specificata l'estensione temporale di ciascun rapporto di lavoro è stato ricostruito in linea con l'informazione presente nell'UniEMens a partire dall'informazione sul numero di giorni lavorati in ciascun mese. Per quanto riguarda gli attributi di ciascun rapporto di lavoro, questi sono stati armonizzati, dove possibile, con quelli presenti nell'UniEMens. Ovviamente, alcune variabili, come Tipo di contribuzione, possono assumere significati e modalità diverse a seconda della fonte. Le unità provenienti dalla fonte Dmag trasformate sono posizione lavorative mensili univocamente identificate dalle seguenti variabili:

- Identificativi del lavoratore e dell'impresa;
- Identificativo del mese;
- Qualifica del lavoratore, tipo di contratto (tempo determinato/indeterminato), regime orario (full time/part time);
- Tipo di contribuzione.

È utile sottolineare che la Qualifica del lavoratore, per i lavoratori provenienti dal Dmag è stata sempre considerata uguale a operaio, inoltre il Tipo di contribuzione presenta modalità differenti da quelle provenienti dall'UniEMens. Per ciascuna unità, inoltre, sono presenti o sono state ricavate le seguenti variabili di interesse ai fini di Asia-occupazione:

- Giorno Assunzione: giorno del mese in cui è intervenuta l'assunzione o la ripresa di attività;
- Giorno Cessazione: giorno del mese in cui è intervenuta cessazione o la sospensione di attività;
- Retribuzioni: retribuzioni corrisposte nel mese;
- Giorni retribuiti: numero dei giorni retribuiti nel mese;
- Ore: ore lavorate per i lavoratori part time;
- Ricostruzione delle settimane retribuite o parti di settimana nel mese sulla base di giorni retribuiti;
- Ricostruzione dei codici eventi di malattia o Cassa integrazione.

L'ultima fonte amministrativa utilizzata per coprire il campo di osservazione dei lavoratori dipendenti del settore privato è l'Enpals. L'Enpals è l'ente di previdenza per lavoratori dello spettacolo e gli sportivi professionisti. L'attività lavorativa, le retribuzioni ed i compensi degli assicurati sono caratterizzati dalla saltuarietà. Infatti, è la qualifica/categoria del lavoratore che dà

luogo all'obbligo assicurativo indipendentemente dalla natura subordinata o autonoma del rapporto di lavoro. L'archivio Enpals a partire dal 2012 confluirà nell'UniEMens e nel 2011 molti dei lavoratori dello spettacolo sono già inclusi nell'archivio Inps.

I dati Enpals sono suddivisi in due parti. La prima relativa alle attività d'impresa e la seconda è relativa agli iscritti dello spettacolo e dello sport. La fornitura è organizzata per denunce mensili, quindi mensilmente per ciascun lavoratore oltre a conoscere l'impresa con la quale esso ha sottoscritto un tipo di rapporto di lavoro, si dispone dei giorni lavorati, della retribuzione e dei contributi.

La struttura Enpals è stata trattata al fine di ottenere una informazione armonizzata con quella presente nell'UniEMens. In particolare, così come per il Dmag, il dettaglio con cui viene specificata l'estensione temporale di ciascun rapporto di lavoro è stato ricostruito in linea con l'informazione presente nell'UniEMens a partire dall'informazione sul numero di giorni lavorati in ciascun mese. Per quanto riguarda gli attributi di ciascun rapporto di lavoro, anche questi sono stati armonizzati, dove possibile, con quelli presenti nell'UniEMens.

Le unità provenienti dalla fonte Enpals sono posizioni lavorative mensili univocamente identificate dalle seguenti variabili:

- Identificativi del lavoratore e dell'impresa;
- Identificativo del mese;
- Qualifica del lavoratore, tipo di contratto (tempo determinato/indeterminato), regime orario (full time/part time);

Per ciascuna unità, inoltre, sono presenti o sono state ricavate le seguenti variabili di nostro interesse:

- Retribuzioni: retribuzioni corrisposte nel mese;
- Giorni retribuiti: numero dei giorni retribuiti nel mese;
- Ricostruzione delle settimane retribuite o parti di settimana nel mese sulla base di giorni retribuiti.

3.3.2 Il processo di integrazione

La realizzazione di un sistema informativo del tipo LEED richiede un complesso processo di *linkage* finalizzato ad integrare tutte le informazioni necessarie per analizzare congiuntamente le caratteristiche dei lavoratori, delle imprese e del tipo di rapporto di lavoro esistente tra l'impresa e il lavoratore. Nello specifico il processo di integrazione coinvolge le differenti unità statistiche presenti nel sistema – posizioni lavorative, imprese ed individui – ed è realizzato abbinando:

- i rapporti di lavoro presenti nelle singole fonti amministrative utilizzate;
- i datori di lavoro presenti nei rapporti di lavoro e il registro delle imprese;
- i lavoratori presenti nei rapporti di lavoro e gli individui.

L'abbinamento tra i rapporti di lavoro presenti nelle differenti fonti amministrative è finalizzato alla copertura del campo di osservazione di interesse – occupazione dipendente delle imprese del settore privato – mentre l'abbinamento tra i rapporti di lavoro, le imprese e gli individui è finalizzato all'integrazione di informazioni presenti rispettivamente nel registro delle imprese (Asia) e negli archivi sugli individui (ad esempio Censimento della popolazione, Liste Anagrafiche Comunali). Chiaramente l'accuratezza degli abbinamenti ha un impatto significativo sulla qualità delle statistiche prodotte a partire dal sistema integrato.

Il codice fiscale dell'individuo ed il codice fiscale dell'impresa sono rispettivamente codici identificativi dell'individuo e dell'unità giuridica presenti in tutte le fonti amministrative a

disposizione. Attraverso un accurato processo di normalizzazione, a ciascun codice fiscale individuo viene assegnato un codice statistico dell'individuo e a ciascun codice fiscale dell'impresa un codice statistico impresa. La presenza di chiavi statistiche permette di effettuare esatti abbinamenti tra le fonti amministrative e statistiche a disposizione. In particolare il codice impresa permette l'abbinamento con il registro Asia ed il codice individuo l'abbinamento con gli archivi statistici e amministrativi sugli individui presenti in Istituto.

Più complesso è invece il processo di integrazione tra le differenti fonti amministrative sui rapporti di lavoro, dove il processo di integrazione deve essere gestito in modo da (i) evitare possibili ridondanze informative e (ii) integrare informazioni differenti provenienti dalle diverse fonti.

L'integrazione dei rapporti di lavoro presenti nell'UniEMens, nel Dmag e nell'Enpals è effettuata per codice impresa, codice individuo e mese di riferimento. Una volta identificata la stessa unità statistica la scelta dei caratteri del rapporto di lavoro è effettuata sulla base delle informazioni disponibili. Vista la migliore qualità e completezza della fonte UniEMens, per le dichiarazioni mensili presenti sia nell'UniEMens che nel Dmag o nell'Enpals vengono scelte tutte le variabili caratterizzanti il rapporto di lavoro nel mese presenti nell'UniEMens, altrimenti quelle delle altre due fonti.

L'integrazione delle informazioni presenti nella Cig con quelle del flusso UniEMens è stata differente a seconda che ci fosse solo il modello della Cig a pagamento diretto oppure la compresenza di quest'ultimo con la dichiarazione mensile UniEMens. Le questioni da affrontare in generale sono state quelle della conversione delle ore di Cassa integrazione a pagamento diretto in tipo di coperture settimanali coerenti con le altre informazioni presenti nell'UniEMens; la valutazione della coerenza tra le informazioni presenti eventualmente nei due file e la risoluzione di eventuali incoerenze; l'attribuzione delle informazioni anagrafiche e delle caratteristiche del contratto di lavoro alle posizioni lavorative presenti soltanto nel file della Cig a pagamento diretto. Più nello specifico sono state distintamente trattati i casi che seguono:

- informazioni presenti, per un certo mese e un certo lavoratore all'interno di una data impresa, solo nel flusso della Cig a pagamento diretto: in tal caso si è provveduto a aggiungere un record, imputare a 1 il Tipo Copertura settimanale, imputare i relativi Codice evento settimanali con indicazione del tipo di Cig (ordinaria, straordinaria, deroga o solidarietà) a pagamento diretto, imputare le informazioni anagrafiche al lavoratore e il tipo di contratto attraverso un recupero delle stesse da altri mesi dello stesso anno o, se necessario, dell'anno precedente iniziando dai mesi più prossimi a quello di stima;
- informazioni presenti, per un certo mese e un certo lavoratore all'interno di una data impresa, sia nel flusso della Cig a pagamento diretto sia nell'UniEMens. In questo caso si è provveduto innanzitutto a valutare la coerenza dei dati disponibili e solo nei casi in cui erano palesemente incoerenti le informazioni sulla Cig a pagamento diretto sono state fatte prevalere rispetto alle altre. Le coperture settimanali, coerenti con le ore di Cig fruite, sono state imputate a 1 se erano zero o mancanti nei dati UniEMens; sono state lasciate inalterate se nei dati di origine erano state dichiarate a 1 o 2, mentre nei casi in cui erano state dichiarate a X queste sono state imputate a 2. In ogni caso si è provveduto a modificare il relativo Codice evento.

3.3.3 La derivazione delle posizioni annue totali e retribuite

Dopo avere identificato il rapporto di lavoro dipendente, si pone il problema della misura della posizione lavorativa media (mensile o annua) nell'anno di riferimento, principalmente finalizzata alla stima degli occupati dipendenti nel registro dell'impres.

Per definire formalmente le posizioni medie mensili o annuali, è stato introdotto *in primis* il concetto di presenza della posizione lavorativa settimanale. La presenza settimanale è calcolata sulla base della variabile Tipo Copertura che è presente nell'UniEMens ed è stata ricostruita, coerentemente con le definizioni presenti nell'UniEMens, per i rapporti di lavoro provenienti dalle altre fonti amministrative. La variabile Tipo Copertura può assumere i valori 1, 2, X, 0 o *missing*, dove gli ultimi due valori indicano che non risulta alcuna dichiarazione o che è stata annullata una dichiarazione precedente, mentre i primi tre indicano come presente la dichiarazione, e quindi il rapporto di lavoro, e differiscono per la modalità di retribuzione:

- 1 indica la settimana come totalmente non retribuita;
- 2 indica la settimana come parzialmente non retribuita;
- X indica la settimana come regolarmente retribuita.

Per totalmente non retribuita si intende la settimana in cui non sia dovuta la retribuzione a causa di eventi tutelati dalla legge, per l'accredito dei contributi figurativi. Per parzialmente retribuita si intende la settimana in cui la retribuzione abbia subito una riduzione a causa di eventi tutelati dalla legge, per l'accredito dei contributi figurativi. Infine, per totalmente retribuita si intende la settimana in cui sia dovuta la retribuzione e questa non abbia subito una riduzione a causa di eventi tutelati dalla legge (ovvero, nei casi di riduzione di retribuzione per eventi diversi da quelli tutelati dalla legge, per esempio, nei casi di eventi di malattia di durata inferiore ai sette giorni, è presente una X).

Sulla base dell'informazione presente nella variabile Tipo Copertura è possibile derivare due tipi di presenze settimanali:

- presenze totali, nelle quali si considera qualsiasi rapporto di lavoro esistente nella settimana di riferimento indipendentemente dalla retribuzione;
- presenze retribuite, nelle quali si considerano unicamente i rapporti di lavoro retribuiti nella settimana di riferimento.

Indichiamo con $T^{l_{ijsm}}$ la presenza totale dell'individuo i in una data impresa j nella settimana o frazione di settimana s del mese m dell'anno T . $T^{l_{ijsm}}$ è pari a uno se Tipo Copertura nell' s -esima settimana nel mese m è pari a 1, 2 o X e pari a zero altrimenti. Indichiamo con $R^{l_{ijsm}}$ l'equivalente presenza retribuita che è pari a uno se Tipo Copertura nell' s -esima settimana nel mese m è pari a 2 o X e zero altrimenti.

La definizione di posizione (totale o retribuita) media in un mese tiene conto di quante e quali settimane corrispondono ad una presenza (totale o retribuita rispettivamente). Ciascun mese può avere un massimo di sei settimane, che, nel calcolo della posizione media, sono state considerate per la loro durata in giorni. Sia $g_{s,m}$ il numero di giorni dell' s -esima settimana o frazione di settimana nel mese m , ($g_{s,m} \in \{0; 1; \dots; 7\}$). Allora la Posizione Totale media $T^{pos_{i,m}}$, e la Posizione Retribuita media $R^{pos_{i,m}}$ nel mese m per un individuo i nell'impresa j sono definite come:

$$T^{pos_{ij,m}} = \frac{\sum_{s=1}^6 T^{l_{ijsm}} g_{s,m}}{(\sum_{s=1}^6 g_{s,m})} \quad R^{pos_{ij,m}} = \frac{\sum_{s=1}^6 R^{l_{ijsm}} g_{s,m}}{(\sum_{s=1}^6 g_{s,m})}$$

Nell'UniEMens sono presenti alcuni casi di record multipli relativi ad uno stesso individuo, nella stessa impresa (dovuti principalmente a cambi di qualifica/contratto nella stessa impresa) che indicano la compresenza di più rapporti lavorativi con Tipo Copertura pari a 1, 2 o X in uno stesso

intervallo temporale. Si ritiene sbagliato conteggiare una presenza unitaria per ciascuno di questi rapporti sovrapposti, secondo la logica generale per la quale, in un qualunque intervallo temporale, in una stessa impresa, per ciascun dipendente (codice fiscale di lavoratore) si può avere al massimo una posizione unitaria. Questa logica vale anche nel caso in cui lo stesso individuo nella stessa impresa appaia come lavoratore dipendente e indipendente, mentre la duplicazione è stata considerata ammissibile per individui che ricoprono la posizione di dipendente e lavoratore esterno (interinale o parasubordinato). Nel caso di una o più sovrapposizioni per l'individuo i nell' s -esima settimana nel mese m , il valore della presenza totale settimanale $T^{l_{ijsm}}$ sarà ricalcolato come frazione dell'unità pari al reciproco del numero di presenze totali sovrapposte nella settimana s . Allo stesso modo, il valore della presenza retribuita settimanale $R^{l_{ijsm}}$ sarà ricalcolato sulla base del numero di presenze retribuite sovrapposte.

Una volta definite le posizioni medie mensili totali e retribuite, si definisce una posizione totale media nell'anno a , $T^{pos_{ij,a}}$ ed una posizione retribuita media nell'anno, $R^{pos_{ij,a}}$, come semplici medie aritmetiche delle relative posizioni mensili:

$$T^{pos_{ij,a}} = \sum_{m=1}^{12} \frac{T^{pos_{ij,m}}}{12} \quad R^{pos_{ij,a}} = \sum_{m=1}^{12} \frac{R^{pos_{ij,m}}}{12}$$

La posizione lavorativa totale alla data del censimento è quindi misurata dalla presenza dell'individuo i in una data impresa j nell'ultima settimana di dicembre.

3.3.4 La derivazione della variabile “Tipologie di beneficiari di sgravi contributivi”

La variabile Tipologie di beneficiari di sgravi contributivi per politiche del lavoro classifica i lavoratori dipendenti delle imprese che alla data censuaria sono beneficiari di sgravi contributivi in relazione all'applicazione di incentivi all'assunzione. Quindi gli interventi di politica del lavoro osservati sono le misure che prevedono incentivi all'assunzione in forma di sgravi contributivi. Alla data di produzione dei dati censuari non sono disponibili dati relativi agli altri tipi di incentivi. Gli sgravi contributivi sono definiti dalla normativa in vigore nel periodo di riferimento. Ciascun intervento è classificato rispetto alla popolazione *target* destinataria della politica a cui il lavoratore beneficiario appartiene.

L'attività di definizione della variabile statistica e della sua classificazione è stata condotta in collaborazione con lo Staff di Statistica, Studi e Ricerche sul Mercato del Lavoro di Italia Lavoro, integrando le competenze giuridiche connesse al mercato del lavoro con le competenze statistiche dell'Istat.

I microdati utilizzati per la costruzione della variabile derivano dalla combinazione di dati amministrativi e dati statistici. I dati amministrativi riguardano il “profilo contributivo” del lavoratore dipendente dalla singola impresa definito come l'insieme delle caratteristiche che determinano l'ammontare dei contributi previdenziali (posizione nella professione, regime orario, eccetera). Ciascun rapporto di lavoro può avere più profili contributivi in relazione alle modifiche nel tempo delle caratteristiche di riferimento. Tra le variabili amministrative utilizzabili a fini statistici per descrivere il rapporto di lavoro dipendente nell'ambito delle imprese private, alcune di queste permettono di individuare i profili contributivi associati a sgravi contributivi in relazione all'applicazione di misure di politica del lavoro.

Per definire rigorosamente la variabile statistica derivata da fonte amministrativa sono state utilizzate le indicazioni normative di riferimento delle politiche del lavoro e le informazioni contenute

nelle corrispondenti circolari Inps, che rendono operativa l'applicazione della politica e definiscono nel dettaglio i codici che i datori di lavoro devono esporre nelle dichiarazioni. In particolare per individuare e descrivere ciascuna politica sono stati considerati i seguenti elementi (Martini A., Sisti M., 2009):

- obiettivi della politica;
- tipi di misure/strumenti;
- popolazione *target* a cui è indirizzata la politica sia in termini di lavoratori beneficiari che in termini di imprese co-beneficarie;
- strumenti di implementazione previsti ed eventuali gradi di implementazione;
- possibili vincoli temporali: numero massimo di anni per usufruire della sgravio contributivo, data di attivazione della politica, eventuale data di termine di applicazione della politica (ad esempio in riferimento a modifiche normative);
- possibili vincoli territoriali: restrizioni su specifiche aree geografiche, eventuali differenziazioni delle misure applicate sul territorio anche in relazione alle disponibilità economiche.

Occorre sottolineare che le modifiche normative, proprie degli atti amministrativi, ma ancora più rilevanti in riferimento alle politiche del lavoro che vengono spesso riformate nel tempo, determinano una necessaria attività di monitoraggio continuo. Tale attività permette di seguire e descrivere le novità normative in materia ma è anche necessaria a garantire gli standard di qualità dei dati statistici prodotti. In dettaglio, le questioni su cui porre l'attenzione sono le seguenti.

A seguito di cambiamenti normativi che modificano gli elementi descrittivi della politica (ad esempio la specificazione della popolazione *target* dei lavoratori o delle imprese, possibili variazioni dell'aliquota contributiva o della durata dell'incentivo, eccetera) si determinano dei cambiamenti dei metadati amministrativi (descrizione dei codici nelle istruzioni di compilazione della modulistica amministrativa); tali cambiamenti devono essere individuati e riportati nei metadati statistici al fine di riposizionare correttamente il significato della variabile statistica.

Le nuove politiche determinano nuove variabili e nuove modalità amministrative da analizzare e codificare nella variabile statistica.

Alcune politiche non più applicabili, in seguito ad abrogazioni o a modifiche, possono continuare a presentarsi nei dati amministrativi ad esaurimento delle coorti di lavoratori il cui beneficio è stato avviato negli anni passati. Per questi codici non più attivi in termini di flusso in entrata, si determina una progressiva riduzione della frequenza associata.

Alcune modifiche possono riguardare il tipo di misura di implementazione della politica o comunque possono determinare una variazione nella generazione del dato amministrativo. Ad esempio, qualora alla misura degli sgravi contributivi venga sostituita una misura di trasferimento diretto alle imprese, il segnale amministrativo che traccia il lavoratore beneficiario e l'impresa co-beneficaria non è più presente. In tal caso l'impatto potrebbe determinare una sottostima della popolazione osservata. Occorre allora modificare i metadati di descrizione e attivare la ricerca di altri possibili segnali amministrativi da utilizzare. È questo il caso delle politiche per l'assunzione dei disabili⁹³.

⁹³ La L. n. 247 del 2007 ha introdotto modifiche alla L. n. 68/1999 in materia di incentivi per l'assunzione di soggetti con disabilità: per le imprese che assumono disabili a partire dal 1 gennaio 2008 non è più prevista la fiscalizzazione totale o parziale degli oneri previdenziali e assistenziali, ma la concessione di un contributo diretto commisurato ad una determinata percentuale del costo salariale annuo del lavoratore disabile variabile in funzione del grado di riduzione della capacità lavorativa del soggetto.

Le attività di trattamento dei dati amministrativi hanno quindi riguardato l'aspetto dell'interpretazione e riconduzione dei concetti amministrativi in concetti statistici e la verifica della qualità in termini di coerenza dei dati. In fase di validazione dei dati integrati, alcuni errori di compilazione delle dichiarazioni sono state individuati e corretti in modo deterministico utilizzando la combinazione di alcune variabili amministrative con variabili anagrafiche dei lavoratori; infine la valutazione dell'errore statistico in termini di casualità e di bassa incidenza ha portato alla produzione della variabile statistica dei "Beneficiari di sgravi contributivi per politiche del lavoro".

All'interno della base informativa sull'occupazione, la variabile individua, quindi, i lavoratori e le imprese che usufruiscono delle varie tipologie di incentivi all'occupazione. Si determina così un potenziale informativo di microdati, sia dal lato delle imprese che dal lato dei lavoratori, molto interessante per le analisi. In particolare queste analisi possono collocarsi nell'ambito dell'Analisi di implementazione della politica che *"consiste nell'analizzare il processo di attuazione della politica allo scopo di descrivere cosa è realmente accaduto durante la messa in atto dell'intervento, giudicare se, e fino a che punto, quello che è accaduto è in linea con quanto era previsto nel disegno originario dell'intervento e spiegare le ragioni di eventuali discrepanze"* (Martini, Sisti, 2009). Per quanto riguarda le politiche del lavoro le possibili analisi delle popolazioni beneficiarie delle politiche possono essere effettuate dal lato lavoratori dipendenti considerando le distribuzioni per genere, per età, per tipo di impresa presso cui lavorano, per tipo di contratto, per distribuzione territoriale, dal lato delle imprese per tutte le variabili strutturali presenti nel registro delle imprese Asia. La produzione di questa struttura di dati è certamente propedeutica alla definizione degli elementi conoscitivi necessari all'applicazione di metodi di valutazione.

Il sistema di classificazione

La classificazione dei lavoratori beneficiari in modalità più aggregate ha lo scopo di permettere la pubblicazione dei dati in tabelle e, allo stesso tempo presentare categorie rilevanti per la produzione dell'informazione statistica.

In generale, considerando gli elementi descrittivi sopra elencati, le tipologie di incentivi all'occupazione possono essere classificati in relazione alla specifica tipologia di azione, ovvero nel modo in cui l'intervento agisce per raggiungere il suo obiettivo; al livello del beneficiario; alle tipologie di gruppi *target* (destinatari) a cui l'intervento è rivolto. La classificazione utilizzata considera questo ultimo criterio.

La scelta dei livelli di aggregazione per la pubblicazione della variabile pur ancorata il più possibile alle categorie individuate dal manuale Eurostat è necessariamente vincolata agli specifici gruppi *target* connessi alle politiche emanate in Italia. Quindi, con lo scopo di mantenere una chiave di lettura informativa delle tabelle di diffusione dei dati aggregati, si è adottata una classificazione gerarchica a due livelli. Il primo livello individua dei gruppi *target* ed il secondo permette di accedere ad un dettaglio maggiore in termini di tipologia di intervento o di sottogruppi *target* più omogenei.

La classificazione adottata è dunque la seguente:

Tipologie di beneficiari di sgravi contributivi per politiche del lavoro

- 1) Giovani;
 - 1.1) Giovani in apprendistato diritto - dovere;
 - 1.2) Giovani in apprendistato professionalizzante;
 - 1.3) Giovani in apprendistato per acquisizione di un diploma o per percorsi di alta formazione;
 - 1.4) Altri giovani;

- 2) Fasce deboli;
 - 2.1) Detenuti o internati;
 - 2.2) Disabili;
 - 2.3) Disoccupati over 50;
 - 2.4) Donne svantaggiate;
 - 2.5) Altri in fasce deboli;
- 3) Disoccupati o beneficiari di ammortizzatori sociali;
 - 3.1) Disoccupati di lunga durata;
 - 3.2) Disoccupati iscritti nelle liste di mobilità;
 - 3.3) Lavoratori in Cigs;
 - 3.4) Altri disoccupati o beneficiari di ammortizzatori sociali.

Per le categorie 2 e 3 è stato possibile dettagliare il *target* in relazione al loro *status* socio-demografico, mentre per la categoria 1 dei giovani ciò non è stato possibile. Si è tentato di dettagliare rispetto alle classi di età ma, l'analisi della normativa porta a concludere che non è possibile classificare gli interventi con questo criterio poiché si creerebbero classi non disgiunte⁹⁴. Per la categoria 3 necessariamente si è dovuto considerare l'insieme non omogeneo disoccupati o beneficiari di ammortizzatori sociali, poiché sono previste politiche rivolte a lavoratori disoccupati da almeno ventiquattro mesi o sospesi dal lavoro e beneficiari di trattamento straordinario di integrazione salariale e quindi non è possibile distinguere i lavoratori provenienti da una o dall'altra categoria⁹⁵. I beneficiari di questa misura sono riportati nella modalità 3.4 - Altri disoccupati o beneficiari di ammortizzatori sociali, mentre i beneficiari ascrivibili in modo omogeneo a ciascun gruppo sono riportati nelle altre modalità: 3.1 - Disoccupati di lunga durata; 3.2 - Disoccupati iscritti nelle liste di mobilità; 3.3 - Lavoratori in Cig straordinaria.

In genere la modalità "Altro" è necessaria poiché oltre a contenere interventi disomogenei, è utile a classificare nel tempo elementi non altrove classificabili. Infatti uno degli obiettivi della classificazione è di rimanere stabile nel tempo per permettere confronti temporali. Occorre ovviamente tener presente che la variabilità temporale delle politiche del lavoro mette a rischio la tenuta delle possibili ipotesi di classificazione che, necessarie per la diffusione aggregata dei dati, devono comunque mantenere un livello informativo di dettaglio. Nel corso degli aggiornamenti successivi, si verificherà la robustezza della classificazione e si valuteranno eventuali ipotesi di adattamento.

3.4 Gli occupati indipendenti

La componente degli occupati indipendenti riveste un ruolo significativo per la misura dell'occupazione complessiva di un sistema di imprese sia perché rappresenta una delle due componenti che contribuiscono a definire gli addetti dell'impresa determinandone quindi la sua

⁹⁴ Giovani in apprendistato diritto – dovere: Giovani da 15 a 18 anni (D.Lgs n. 276/03) oppure giovani da 15 a 25 anni (D.Lgs n. 167/11); Giovani in apprendistato professionalizzante e Giovani in apprendistato per acquisizione di un diploma o per percorsi di alta formazione: Giovani da 18 (o 17) a 29 anni (sia D.Lgs n. 276/03 che D.Lgs n. 167/11) <http://www.nuovoapprendistato.gov.it/cosa.html>.

⁹⁵ L. n. 407/1990 art. 8 comma 9, agevolazioni contributive per i datori di lavoro che assumono con contratto a tempo indeterminato di lavoratori disoccupati da almeno ventiquattro mesi o sospesi dal lavoro e beneficiari di trattamento straordinario di integrazione salariale da un periodo uguale a quello suddetto.

dimensione, sia perché la figura del lavoratore indipendente approssima (sia per difetto che per eccesso) quella dell'imprenditore, uno degli elementi al centro degli studi sull'imprenditorialità⁹⁶.

Secondo la definizione adottata, per occupato indipendente si intende il personale che svolge la propria attività lavorativa in una impresa, senza vincoli formali di subordinazione, e la cui remunerazione abbia natura di reddito misto (capitale e lavoro).

L'approccio concettuale e operativo utilizzato fino al 2010 era basato sul concetto che *"il lavoro indipendente esiste sempre in una impresa poiché rappresenta la funzione di coordinamento e gestione dell'impresa stessa"*. La revisione adottata oltre che di carattere definitorio ha di conseguenza interessato il metodo per l'identificazione e la misura di tale componente. Il nuovo approccio si basa sull'identificazione dell'indipendente come colui che *"effettivamente svolge una attività lavorativa all'interno dell'impresa in cambio di una remunerazione"*.

L'identificazione dell'indipendente si pone in modo differente a seconda della complessità organizzativa dell'impresa; nell'ambito delle variabili di classificazione presenti nel registro Asia si fa quindi riferimento alle diverse tipologie di forme giuridiche dell'impresa. Mentre per alcune tipologie, quali le imprese individuali, la figura dell'indipendente è abbastanza semplice da identificare, in quanto corrisponde al titolare dell'impresa e ad eventuali familiari coadiuvanti (non iscritti a libro paga); all'opposto per le imprese a complessa organizzazione quali quelle sotto forma di consorzio, ente pubblico economico, o quelle di *branch* di imprese estere, per la natura della forma giuridica, non è ammissibile la presenza del lavoro indipendente.

Per le società di persone, prevalendo nella struttura organizzativa e giuridica l'elemento soggettivo, vengono riconosciute nel socio che svolge una attività lavorativa all'interno dell'impresa (ossia non risulta iscritto a libro paga o remunerato esclusivamente attraverso l'utile distribuito) le caratteristiche del lavoratore indipendente. Più complessa risulta, invece, l'individuazione dell'indipendente nelle società di capitali, nelle quali il capitale ha una prevalenza concettuale e normativa rispetto all'elemento soggettivo rappresentato dai soci, e distinguere tra i soci coloro che vengono direttamente coinvolti nelle attività di gestione, organizzazione e lavoro all'interno della società risulta concettualmente più complesso.

Per la stima dell'occupazione indipendente, a differenza delle altre tipologie di occupati, sono state sviluppate procedure statistiche *ad hoc* che utilizzano le fonti amministrative, integrate a livello di impresa e di singola persona/lavoratore, in modo indiretto, nel senso che nessuna fonte, da sola o integrata, consente una misura diretta del fenomeno ma è solo attraverso un sistema di regole operative, che interpretano e traducono le assunzioni definitorie, che è possibile identificare e classificare il rapporto di lavoro come indipendente.

3.4.1 La procedura di identificazione dell'indipendente: la struttura informativa e le fonti amministrative

La procedura sviluppata per la stima dell'occupazione indipendente si basa sull'utilizzo di informazioni provenienti da fonti amministrative, gestite da enti pubblici, integrate con quelle da fonti statistiche. Si possono distinguere cinque principali fasi operative:

- 1) l'integrazione di fonti amministrative per costruire la struttura informativa di base che contiene tutti i legami tra impresa e individuo che potenzialmente consentono di identificare i rapporti di lavoro di tipo indipendente. Si tratta di legami di proprietà del capitale sociale – l'individuo è socio dell'impresa –, legami in termini di cariche sociali

⁹⁶ Il programma di Indicatori sull'Imprenditorialità "OECD/Eurostat Entrepreneurship Program" è stato proposto inizialmente nel 2006 dall'Ocse al fine di dare linee guida sulle definizioni standard dell'imprenditore e dell'imprenditorialità e il calcolo di indicatori armonizzati a livello internazionale.

ricoperte – l'individuo è ad esempio amministratore dell'impresa –, legami di partecipazione agli utili della società (semplice) o all'associazione – l'individuo dichiara la quota del reddito (o della perdita) della società;

- 2) l'integrazione di fonti amministrative per costruire un set di informazioni associate al legame individuo-impresa utilizzate come segnali *proxy* per misurare un qualsiasi flusso remunerativo dalla società ai suoi membri o che evidenzia una qualunque attività lavorativa svolta;
- 3) l'assegnazione di criteri di tipo deterministico basati sui segnali costruiti come al punto 2 precedente per l'identificazione del lavoratore indipendente. Il processo è articolato e segue regole parzialmente differenti a seconda della forma giuridica delle imprese;
- 4) la stima e l'identificazione del lavoratore indipendente, se non identificato nello step precedente, sulla base di un metodo probabilistico;
- 5) il processo di stima del lavoratore indipendente utilizzando informazioni aggiuntive per le imprese che risultano senza addetti.

Con riferimento alla fase 1, le fonti amministrative utilizzate per la costruzione della struttura informativa di base dei legami impresa-individui sono:

- i registri delle imprese delle Camere di commercio, industria, artigianato e agricoltura e in particolare gli archivi collegati dei soci delle società di capitale e delle "Persone con cariche sociali";
- gli archivi gestiti dall'Agenzia delle Entrate per il Ministero dell'Economia e delle Finanze, e in particolare le dichiarazioni fiscali del modello Unico/PF, quadro RH.

L'archivio soci fornisce *"l'elenco dei soci (persone fisiche e giuridiche, residenti e non), riferito alla data di approvazione del bilancio con l'indicazione del numero delle azioni o quote possedute, nonché dei soggetti diversi dai soci che sono titolari di diritti o beneficiari di vincoli sulle azioni medesime"*, che gli amministratori sono tenuti a depositare presso le Camere di commercio. Le principali informazioni utilizzate si riferiscono all'elenco dei soci, identificati dal codice fiscale e alla quota di partecipazione al capitale. Tale fonte fornisce annualmente circa 3,9 milioni di legami.

L'archivio delle "persone con carica sociale" registra, nell'ambito di un'impresa (in particolare delle società di persone) le persone che assumono determinate qualifiche (socio, socio amministratore, eccetera) o che possono essere nominate a determinate cariche (presidente, consigliere delegato, eccetera). Una persona può essere titolare di più cariche e qualifiche. La carica ricoperta riflette la regolamentazione giuridica della sottostante struttura organizzativa. Con periodicità annuale la fonte fornisce circa 6,5 milioni di legami; le principali informazioni riguardano appunto i legami tra l'impresa e le persone con cariche, identificati dal codice fiscale, le date di validità della posizione assunta, che vengono referenziate all'anno di riferimento, il tipo di carica, la quota di partecipazione agli utili.

Con le informazioni provenienti dal quadro RH dei modelli Unico/PF, i contribuenti che compilano questo quadro dichiarano la quota di reddito derivante da partecipazione in società semplici o associazioni fra artisti e professionisti. I redditi o le perdite sono imputati a ciascun socio proporzionalmente alla propria quota di partecipazione. Il dichiarante, identificato dal codice fiscale, indica attraverso l'uso di uno o più moduli del modello Unico una o più società partecipata indicandone il codice fiscale, la quota di partecipazione espressa in percentuale, la quota di reddito o perdita. In termini di struttura informativa, i dati fiscali di base necessitano di un'attività di pretrattamento che consiste nel ricostruire per la società partecipata, identificata dal codice fiscale,

i legami con tutti i dichiaranti che vi partecipano con le relative quote. Vengono così ricostruiti, annualmente, circa 2,6 milioni di legami.

Al secondo gruppo di fonti amministrative utilizzate – fase 2 – appartengono quelle da cui si derivano i “segnali” utili a classificare la natura del legame e in particolare a misurare la presenza di flussi remunerativi tra la società e i suoi membri. Le fonti utilizzate sono:

- gli archivi dell'Istituto Nazionale della Previdenza Sociale (Inps) e in particolare:
 - a) la gestione Artigiani e Commercianti, dove sono registrati i contributivi di tipo previdenziale versati dal titolare, artigiano e commerciante, per se stesso e per i familiari (coadiuvanti e collaboratori). L'importo versato si calcola in base al reddito d'impresa che è costituito dalla totalità dei redditi di impresa denunciati ai fini Irpef per l'anno al quale i contributi si riferiscono;
 - b) la gestione autonomi agricoli che riguarda i versamenti effettuati dai lavoratori autonomi (coltivatori diretti, imprenditori agricoli, eccetera) in agricoltura e quelli dei familiari;
 - c) la Gestione separata dei parasubordinati collaboratori, che registra i rapporti di collaborazione tra il lavoratore e la struttura organizzativa del datore di lavoro committente, senza vincolo di subordinazione. L'informazione rilevante è il compenso percepito quale indicatore del flusso di remunerazione;
 - d) la Gestione separata dei parasubordinati-professionisti che contiene i lavoratori autonomi - liberi professionisti - per i quali non è stata prevista una specifica cassa previdenziale in quanto “soggetti Iva”;
 - e) l'archivio delle dichiarazioni assicurative di quelle imprese che conferiscono incarichi di collaborazione coordinata e continuativa in attività classificate a rischio di infortuni sul lavoro (iscrizioni all'Inail).
- la base dati LEED sull'occupazione dipendente ottenuta dall'integrazione delle fonti specifiche (come descritto nel Paragrafo 3.3);
- gli archivi gestiti dall'Agenzia delle Entrate per il Ministero dell'Economia e delle Finanze, quali l'Anagrafe tributaria per la popolazione di soggetti con partita Iva e gli Studi di settore in particolare il quadro A sul personale addetto all'attività;
- il registro Asia-imprese e il registro Asia-gruppi d'impresa, utilizzati rispettivamente per identificare i titolari di partita Iva attiva (e quindi i titolari di una impresa) e le imprese “complesse” ossia quelle legate da legami di controllo.

Il processo di integrazione utilizza le chiavi certe del codice fiscale dell'impresa e dell'individuo, disponibili in queste fonti, per l'abbinamento alla struttura integrata di base, costruita nella fase precedente. Tra le fonti utilizzate una novità che ha caratterizzato il nuovo processo di individuazione dell'occupazione indipendente consiste nell'aver impiegato per la classificazione come lavoratori indipendenti, parte dei parasubordinati (amministratori e Co.Co.Co) se soci dell'impresa, fonte che rappresenta la base informativa primaria per la classificazione dei lavoratori esterni (Paragrafo 3.5).

3.4.2 La procedura di identificazione dell'indipendente: le regole deterministiche⁹⁷

Per l'identificazione dell'indipendente all'interno di una impresa inteso come colui/colei che effettivamente svolge una attività all'interno dell'impresa in cambio di una remunerazione (reddito misto), sono stati adottati alcuni criteri operativi. La procedura di identificazione si basa su alcuni criteri di base ed è poi sviluppata in maniera differente con riferimento alle diverse tipologie di

⁹⁷ Ha collaborato alla fase di sperimentazione delle regole operative per la scelta dell'indipendente Mirella Morrone.

forme giuridiche. Le regole deterministiche adottate per tutte le forme giuridiche si basano sulle seguenti condizioni relative all'individuo:

- a) deve essere socio;
- b) deve possedere una carica;
- c) deve partecipare all'attività lavorativa della società.

Se le condizioni a) e b) sono misurabili utilizzando direttamente l'informazione della struttura base, la condizione c), considerata una misura *proxy* del reddito misto, è più complessa da accertare e quindi da misurare; per tale ragione le regole adottate fanno uso della struttura integrata dei segnali.

Di seguito viene descritto più nel dettaglio il processo di identificazione e stima degli occupati indipendenti per forma giuridica delle imprese.

Le imprese individuali

Nel caso delle imprese individuali, l'indipendente corrisponde alla figura del titolare e, laddove presenti, anche ai familiari coadiuvanti. In particolare l'indipendente in senso stretto coincide esattamente con il titolare dell'impresa attiva. L'elemento rilevante per la classificazione è quindi conoscere se l'impresa è attiva, informazione che deriva da procedure di assegnazione dello stato di attività nel processo del registro Asia⁹⁸.

Le società di persone - società in accomandita semplice (S.a.s.), società in nome collettivo (S.n.c.) e studi associati

Tra tutte le società di persone, le società in accomandita semplice e quelle in nome collettivo rappresentano le forme più numerose (rispettivamente poco più di 320 mila e 410 mila). La società in accomandita semplice (S.a.s.) è una società di persone caratterizzata dal fatto che di essa fanno parte due diverse categorie di soci: gli accomandanti (SOC) e gli accomandatari (SOR). Gli accomandatari partecipano alla gestione della società e rispondono solidalmente e illimitatamente per le obbligazioni sociali, anche con il patrimonio personale. Di fatto essi partecipano all'amministrazione della società. Gli accomandanti non partecipano alla gestione della società, sono obbligati unicamente al conferimento e rispondono nei confronti dei terzi nei limiti della quota conferita.

La struttura organizzativa della società in nome collettivo (S.n.c.) è caratterizzata prevalentemente dalla presenza di un socio amministratore (SOA) e dagli altri soci (COM); nella S.n.c. tutti i soci sono responsabili in modo illimitato e l'amministrazione della società compete a uno o più soci. La struttura societaria di una S.a.s. e anche di una S.n.c. è sostanzialmente formata mediamente da due-tre persone di cui almeno una è il SOR per la S.a.s. e il SOA per la S.n.c. Tuttavia, la presenza di almeno un SOR o SOA, che di fatto partecipano all'amministrazione della società, non è sufficiente di per sé, secondo i nuovi criteri definitori⁹⁹, ad identificare almeno questa tipologia di persone come indipendenti dell'impresa. Infatti, per la stima dell'indipendente è necessario sviluppare una procedura per identificare che la persona svolga un'attività lavorativa all'interno dell'impresa e che questa sia remunerata.

La procedura di identificazione sviluppata si basa sull'assunto di partenza, secondo il quale nella società di persone la figura dell'indipendente deve esistere. La persona all'interno dell'impresa viene identificata come indipendente in base alla struttura dei segnali provenienti dalle fonti integrate. I segnali *proxy* utilizzati per verificare i criteri di base sono:

⁹⁸ Per la stima dello stato di attività si veda il Paragrafo 2.5.2.

⁹⁹ Si veda il Paragrafo 3.4.

- il pagamento di contributi alla gestione Artigiani e Commercianti Inps da parte dello stesso lavoratore, o del versamento di contributi da autonomi in agricoltura (senza partita Iva);
- il pagamento da parte della società con cui ha un rapporto di lavoro di una remunerazione per il lavoro prestato (imponibile dichiarato presso la Gestione separata Inps dei parasubordinati-collaboratori - senza partita Iva).

Per le società dove non è stato identificato l'indipendente utilizzando i suddetti segnali, e non vi sono ragioni valide per assumere la non esistenza dell'indipendente (quando ad esempio tutti i membri della società sono classificati come dipendenti), si ricorre a regole aggiuntive. Queste prendono in esame alcune caratteristiche dell'individuo ritenute, in un ordine di rilevanza, migliori in termini di requisiti per l'eleggibilità: l'individuo non ha partita Iva, ha una carica SOR o SOA, ha una quota di partecipazione più elevata, eccetera. A questi elementi si associano anche i risultati ottenuti da tecniche di *record linkage*¹⁰⁰ che basandosi sul confronto tra la denominazione della società e il nome del socio consente di assegnare dei punteggi che variano in base alla bontà dell'abbinamento dei nomi; tali punteggi aggiungendosi agli altri requisiti per l'eleggibilità sono utilizzati per l'identificazione dell'indipendente.

In assenza di informazioni sui legami impresa-individuo nella struttura informativa base, dovuti a vari problemi di sotto-copertura delle fonti o di classificazione della forma giuridica, ma in presenza di una impresa attiva dal registro Asia, l'indipendente viene imputato e posto pari ad uno. In quest'ultimo caso non saranno disponibili i caratteri anagrafici quali età, sesso e paese di nascita dell'indipendente.

Per quanto riguarda gli studi associati, la struttura informativa è costituita dalle persone che, munite dei necessari titoli di abilitazione professionale si associano per l'esercizio delle professioni o delle altre attività per cui sono abilitate o autorizzate. Gli studi associati debbono usare, nella denominazione del loro ufficio e nei rapporti coi terzi, esclusivamente la dizione di "studio tecnico, legale, commerciale, contabile, amministrativo o tributario", seguito dal nome e cognome, coi titoli professionali, dei singoli associati (L. 1815/39, art. 1). Da un punto di vista fiscale i redditi prodotti dallo studio associato sono redditi di lavoro autonomo. Un professionista può esercitare la propria attività in parte quale associato in uno studio e in parte individualmente. Per l'esercizio di talune professioni la legge prevede espressamente l'iscrizione in appositi albi o elenchi professionali.

Le persone che partecipano allo studio sono identificate dalla fonte fiscale modello Unico persone fisiche, quadro RH, relativo ai redditi di partecipazione in società di persone e assimilati. Per la tipologia di società in esame la struttura dei segnali secondo la quale classificare ciascuna persona non riveste una particolare rilevanza, né tanto meno il fatto che un associato possa avere una partita Iva differente dallo studio con cui possa svolgere un'altra attività. Tutti gli associati dello studio, per le caratteristiche giuridiche che possiedono, sono considerati indipendenti dello stesso.

L'assegnazione del familiare

Il familiare viene riconosciuto utilizzando la fonte Artigiani e Commercianti dell'Inps e, in modo più limitato relativamente al campo di osservazione delle imprese in esame, quella sui lavoratori autonomi in agricoltura; la sua presenza è subordinata all'esistenza di un indipendente in senso stretto. L'esistenza del familiare è ammessa solo per le imprese individuali e per le società di persone; se lo stesso familiare si presenta nella struttura informativa sia in una impresa individuale sia in una società di persone, al fine di evitare doppi conteggi, il suo apporto lavorativo principale si ammette solo nella impresa individuale. Analogamente se lo stesso familiare è presente in più

¹⁰⁰ Le tecniche di *record linkage* sono state adattate ai fini dell'identificazione dell'indipendente da Patrizia Cella.

società di persone si ammette solo in una tra quelle che hanno forma organizzativa più semplice e con un numero minore di persone-soci.

Le società di capitali

Tra tutte le società di capitali quelle organizzate come società a responsabilità limitata (S.r.l.) rappresentano la forma più numerosa annoverando oltre 640 mila imprese, seguono le S.r.l a socio unico (124 mila) e quindi le società per azioni (S.p.a.) che sono 29 mila. All'interno delle S.r.l. si ritrova in prevalenza la figura dell'amministratore unico (AUN), nel caso di impresa con una sola persona con carica, dell'amministratore (AMM) e dei consiglieri (CON) al crescere del numero dei soci. Nelle S.r.l. con unico socio le due cariche prevalenti sono quello del socio unico (SOU) e della figura dell'amministratore unico (AUN). Mentre nelle S.p.a. in generale, le cariche più diffuse sono quelle di consiglieri (CON) e di controllo come il sindaco effettivo (SIE) e il sindaco supplente (SIS).

Il processo di identificazione dell'indipendente nelle società di capitali è articolato in più fasi. Nella prima vengono adottate regole deterministiche analoghe a quelle utilizzate per le società di persone: l'indipendente corrisponde al socio, amministratore o meno, senza partita Iva, che si remunera con un reddito misto ossia si versa i contributi per il lavoro che svolge (presso la gestione Artigiani e Commercianti) e/o percepisce un compenso per la carica che svolge all'interno della società (contributi versati alla Gestione separata Inps). I criteri elencati si seguono per tutte le persone collegate a ciascuna impresa; in alcuni casi marginali, in assenza dell'indicazione certa del socio, si adottano regole di inclusione selezionando l'indipendente tra coloro, senza partita Iva, che ricoprono cariche specifiche riconoscibili come quella di socio o che da tecnica di *record linkage* risultano abbinata tramite il nome alla denominazione dell'impresa. La logica con cui però si viene a valutare la presenza dell'indipendente in una società cambia quando essa appartiene ad un gruppo d'impresе; in tali casi l'identificazione avviene osservando le informazioni congiuntamente per tutte le imprese del gruppo, non viene duplicata e si tiene in considerazione la struttura organizzativa del vertice dello stesso. Alcune assunzioni che escludono l'assegnazione di indipendente sono a) imprese con tutti soci persone giuridiche (imprese/istituzioni), b) *branches*, imprese a controllo pubblico, c) gruppi a controllo estero, gruppi con oltre tre livelli, gruppi con al vertice imprese individuali o società di persone o cooperative.

Per le società non appartenenti a gruppi per le quali è assente la fonte Archivio soci, se dalle fasi precedenti non viene identificato nessun indipendente si procede a stima probabilistica da modello (Paragrafo 3.4.3). Laddove dai risultati provenienti anche dalla fase di stima probabilistica non è stato identificato nessun indipendente e al contempo non vi è altra occupazione (né dipendente né esterna), pur essendo l'impresa attiva (ossia con fatturato), si procede alla verifica di ulteriori criteri, di natura economica, stabiliti sia da alcune attività di revisione e controlli manuale sia da informazioni provenienti da ulteriori fonti (Studi di settore - quadri contabili, bilanci, archivio del commercio con l'estero Istat, eccetera) per assegnare/imputare un indipendente alla società (Paragrafo 3.4.4).

Sul complesso delle circa 800 mila società di capitali, la figura dell'indipendente è stata riconosciuta per il 51 per cento delle imprese, tale percentuale è differente tra le unità in gruppi d'impresa (30 per cento) dove è maggiormente rappresentata la forma giuridica delle S.p.a. e quelle che non vi appartengono (53 per cento). Per quanto riguarda le imprese non in gruppi dove non sono stati stimati gli indipendenti è da sottolineare che per circa 51 mila imprese i soci sono mancanti per un non abbinamento della fonte e quindi è stata successivamente applicato il metodo di stima da modello probabilistico; ulteriori 55 mila sono società che hanno tutti i soci o dipendenti o persone giuridiche o sono partecipate e a controllo di società estere oppure i soci non sono

amministratori e quelli amministratori sono invece dipendenti. Per la parte rimanente con indipendenti a zero e senza lavoro dipendente, pari a circa 137 mila imprese, poco più del 50 per cento sono classificate nei settori delle Attività immobiliari e delle Costruzioni.

Le società cooperative

L'elemento che ha destato maggiore discussione in ambito di revisione delle definizioni adottate riguarda il tema della riclassificazione dei soci di cooperative di produzione e lavoro. Laddove il socio non sia già tra i dipendenti, la logica per la stima dell'indipendente è analoga a quanto fatto nel caso delle società di capitali. La difficoltà maggiore da un punto di vista operativo riguarda l'assenza di informazione da fonte amministrativa sulla struttura dei soci. Di conseguenza l'informazione di socio viene approssimata con quella di alcune cariche sociali. Sulla base di alcune evidenze giuridiche ed empiriche si può assumere che la maggior parte dei consiglieri/amministratori – carica prevalente nelle cooperative – sono scelti nella notevole maggioranza dei casi tra i soci; inoltre, il socio è direttamente derivato dalle informazioni presenti nella fonte Inps sui parasubordinati-collaboratori dove sono esplicitamente indicati per alcune tipologie contrattuali, i soci lavoratori di cooperative.

Come per le società di capitali il processo di identificazione è articolato separatamente per le imprese appartenenti a gruppi e per quelle non in gruppo, dove le prime però sono numericamente molto poche. Le regole deterministiche sono analoghe a quelle utilizzate per le società di capitali: l'indipendente corrisponde al socio, stimato o identificato, senza partita Iva, che si remunera con un reddito misto ossia che percepisce un compenso per la carica che svolge all'interno della società (contributi versati alla Gestione separata Inps). Se invece il legame è attraverso il segnale di versamento dei contributi da parte della persona come artigiano o commerciante allora la sua identificazione come indipendente è vincolata all'accertamento che la cooperativa sia l'unica impresa in cui lavora. Ne risulta che sul complesso delle cooperative, pari a poco più di 50 mila, siano stati attribuiti circa 33 mila indipendenti.

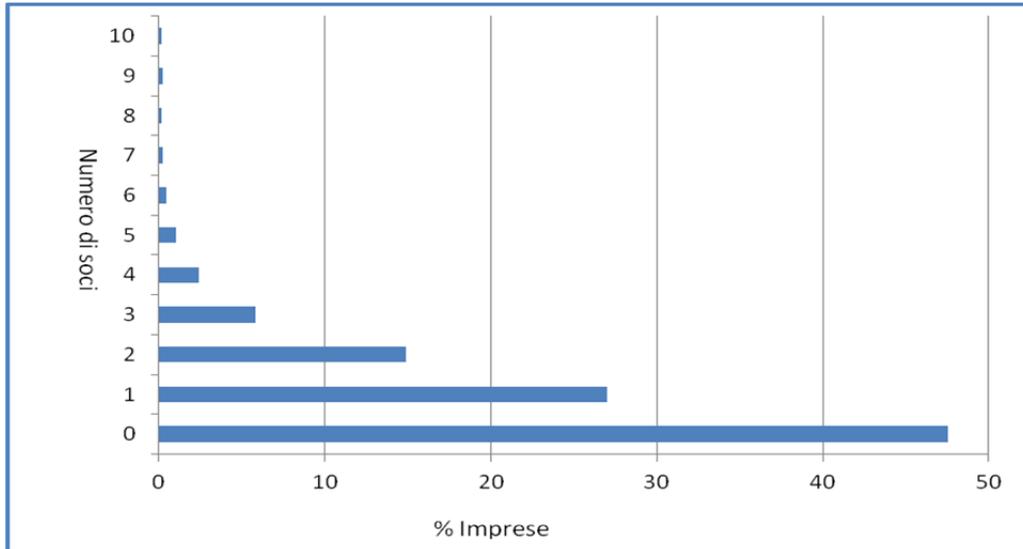
3.4.3 La stima probabilistica per un sottoinsieme di società di capitali senza indipendenti

Per superare il problema della mancata individuazione dell'indipendente nel sottoinsieme di società di capitali è stato sviluppato un approccio di tipo probabilistico.

L'idea consiste nello stimare il numero di indipendenti per le imprese senza informazioni sulla struttura societaria, utilizzando modelli di tipo regressivo e sotto l'ipotesi di affidabilità dell'informazione proveniente dagli Studi di settore (Sds) in particolare sul numero di soci (sia quelli con occupazione prevalente nell'impresa sia quelli con occupazione non prevalente, amministratori o non). Sulla base della stima probabilistica del numero di soci è stata successivamente ricostruita l'informazione sui lavoratori indipendenti attraverso l'applicazione delle stesse regole deterministiche di identificazione del lavoratore indipendente utilizzate negli altri insiemi considerati della popolazione delle società di capitali.

All'interno delle sottopopolazioni con mancanza di informazione sulla struttura societaria, una predizione basata su un modello di Poisson standard potrebbe portare a stime non affidabili a causa della eccessiva iperdispersione dovuta alla elevata quantità di zeri nel numero di soci registrato dalla fonte Sds (Figura 3.2).

Figura 3.2 - Distribuzione percentuale delle imprese per numero di soci dichiarati da Studi di settore (Sds)



Fonte: Istat, elaborazioni su dati Sds.

Una condizione importante del modello di Poisson è l'assunzione di equidispersione; tuttavia, la varianza è spesso maggiore della media, questo implica che la variabile dipendente sia iperdispersa. Seguendo Greene (1994), la extra variabilità dei dati può essere dovuta alla presenza di eterogeneità non osservata che può avere origine da eventuali covariate omesse. Il modello di Poisson consente di considerare solo l'eterogeneità osservata, sintetizzata dalle covariate osservate (in questo caso le variabili di segnale). Al fine di considerare l'iperdispersione presente nei dati sarebbe più opportuna l'implementazione di un modello di regressione binomiale negativo, in quanto tale distribuzione ha una media che corrisponde a quella della distribuzione di Poisson, ma varianza diversa in funzione della media e del cosiddetto parametro di dispersione. Poiché nel caso presente gli effetti di alcune variabili di segnale fondamentali per la predizione non vengono stimati dal modello binomiale negativo *zero inflated* (ZINB) per ridotte dimensioni campionarie, e dopo aver constatato che la differenza nei parametri stimati, in intensità e direzione, e nei valori predetti (numero di soci delle società di capitali) sono trascurabili, si è scelto di utilizzare un modello di Poisson *zero inflated* (ZIP).

I modelli per i dati di conteggio *zero inflated* rappresentano un approccio coerente a formalizzare un modello di regressione per una variabile di conteggio in presenza di un eccesso di zeri. In questo caso, si ipotizza che per ogni osservazione ci siano due processi di generazione del dato; il risultato di un esperimento bernoulliano determina quale dei due processi è "attivo". Il primo processo determina solo conteggi pari a zero, il secondo processo genera conteggi da un modello di Poisson. Un modello *zero inflated* di Poisson (ZIP) è definito da una mistura di una variabile casuale di Poisson e una variabile casuale con distribuzione degenerata a zero; la funzione di densità di probabilità è definita come segue:

$$P(Y = y) = \begin{cases} p_0 + (1 - p_0) \exp(-\mu) & \text{per } y = 0 \\ (1 - p_0) \frac{\mu^y \exp(-\mu)}{y!} & \text{per } y = 1, 2, \text{ etc.} \end{cases}$$

In dettaglio, il processo bernoulliano assume valore zero con probabilità p_0 ; in questo caso è $y=0$. Se il processo assume valore 1, con probabilità $(1 - p_0)$ allora Y si distribuisce come una variabile casuale di Poisson di media μ .

Il parametro p_0 può essere considerato funzione delle caratteristiche degli individui ed il corrispondente modello può essere definito da un modello di regressione logistica o probit.

Sette modelli ZIP, uno per ogni gruppo di attività economica, sono stati implementati. I gruppi di attività economica considerati sono Agricoltura, Industria, Costruzioni, Commercio, Trasporti, HORECA (hotel, ristorazione e catering), Altri servizi. Le variabili esplicative considerate sono state costruite utilizzando tre informazioni riferite alle singole persone che, tramite la base integrata da fonti amministrative, risultano collegate a ciascuna impresa. Le prime due informazioni riguardano la condizione di socio e/o carica ricoperta secondo il seguente dettaglio: quella persona è un socio (variabile indicizzata con $_a$), non è socio ma ricopre una carica amministrativa ($_b$), non è socio né ricopre carica amministrativa ($_c$). La terza informazione fa riferimento ai segnali di presenza/assenza derivante dall'abbinamento con le fonti amministrative; il numero di persone per tipo di segnali, a livello di impresa, costituisce le variabili indicizzate come g_1 che indica assenza di segnali, g_2 per quelle persone che hanno anche una partita Iva attiva, g_3 se i segnali provengono dalla fonte artigiani/commercianti o autonomi agricoli (e anche da parasubordinati Inps), g_5 se provenienti da parasubordinati Inps.

La bontà di adattamento dei modelli è stata valutata confrontando la devianza dei modelli correnti e quella dei modelli nulli (senza covariate), allo scopo di verificare se siano presenti differenze significative (in tal caso il modello corrente è preferibile al modello nullo) attraverso un test del chi-quadrato con gradi di libertà pari al numero delle covariate nel modello. Tale test è stato effettuato su modelli distinti per ogni gruppo di attività economica ed ha portato a risultati soddisfacenti. Il numero di parametri del test dipende dal numero di variabili considerate; in questo caso, le variabili di segnale sono state utilizzate in entrambi i processi considerati, quello che genera conteggi nulli e quello che genera conteggi da una distribuzione di Poisson.

Un'analisi di sensibilità è stata effettuata attraverso la curva ROC al fine di determinare il valore soglia (cut-off) che fornisce la migliore specificità senza perdita di sensibilità. Tale soglia permette l'individuazione di un valore massimo oltre il quale il valore predetto di un conteggio (numero di soci) positivo è pari a zero e al di sotto del quale invece è possibile avere conteggi nulli o positivi derivati dal modello di Poisson. Considerando le stesse covariate per il modello per conteggio nullo e per il processo di Poisson si ottiene, sulla base della soglia selezionata, un valore AUC (*area under the curve* ROC) di poco inferiore a 0,70; ciò indica che il modello ha un potere di "discriminare" non molto elevato e la sua accuratezza potrebbe quindi essere migliorata attraverso uno studio più dettagliato mirato all'identificazione di ulteriori altre variabili che descrivono in modo più efficace la probabilità di un numero di soci pari a zero. Si è deciso di utilizzare tutte le variabili di segnale a disposizione in quanto queste entrano in gioco nella stima deterministica del numero di lavoratori indipendenti delle società di capitali, e sulla base di queste si è voluta stimare la probabilità di un conteggio a zero.

Per capire come si comportano le variabili di segnale sui modelli di Poisson *zero inflated* utilizzati per stimare il numero di soci degli Sds e per la stima del numero di soci nelle sottopopolazioni delle società di capitali su cui non si ha l'informazione, occorre analizzare i risultati dei modelli. Per quanto riguarda l'Industria, le Costruzioni, il Commercio e l'HORECA le variabili g_3_b e g_3_c hanno un peso maggiore e statisticamente significativo nella determinazione del numero di soci. Nel sottogruppo dei Trasporti invece è la variabile g_2_b a giocare un ruolo determinante nella stima del conteggio, mentre negli Altri Servizi tale ruolo è ricoperto dalle variabili g_3_c e g_5_c .

Le variabili di segnale si comportano diversamente nei vari sottogruppi di attività economica per quanto concerne l'influenza sulla probabilità di avere un conteggio pari a zero. Il loro contributo è molto più variabile rispetto al modello di Poisson; per esempio nel settore Industria, all'aumentare di g1_b, g1_c e g5_c aumenta significativamente e positivamente il logaritmo del rischio di avere un conteggio pari a zero, mentre tale rischio diminuisce all'aumentare di g5_b g34_c, e nel settore del Commercio tutte le variabili considerate contribuiscono positivamente all'aumentare del logaritmo del rischio.

3.4.4 L'individuazione dell'indipendente per le società di capitali senza dipendenti

Dalle procedure deterministiche e probabilistiche di stima per l'assegnazione dell'indipendente e dal processo di integrazione per la stima dell'occupazione dipendente, ne deriva che circa 137 mila società di capitali, non appartenenti a gruppi d'impresa, non presentano addetti. Questo risultato ha portato ad approfondire lo studio di tali imprese che sono rappresentate per la più alta percentuale da società operanti nel settore delle Attività immobiliari (circa il 36 per cento), cui segue il settore delle Costruzioni (18 per cento), Commercio (14 per cento) e Attività manifatturiere (6 per cento).

Al fine di verificare la reale assenza di occupazione sono state svolte attività di revisione manuale per un campione di tali aziende. L'analisi ha messo in evidenza che si verificano situazioni in cui alcune imprese avrebbero cessato l'attività economica nel periodo successivo all'anno di riferimento, o che avevano in corso procedure concorsuali (in liquidazione, fallimento, eccetera) o al contrario manifestavano flussi di commercio con l'estero.

Se da un lato è plausibile che ad esempio un'azienda in fallimento o che stia per cessare non abbia addetti, dall'altro si è ritenuto necessario indagare ulteriormente sulle imprese che effettuavano import/export. Infatti, nell'eventualità avessero effettuato scambi con l'estero nell'anno in esame, ciò poteva portare ad ipotizzare la presenza di almeno un addetto. La generalizzazione di tale analisi ha portato alla integrazione delle informazioni relative alle date di cessazione e alle procedure concorsuali derivanti dalle Camere di commercio e presenti nel registro Asia-imprese, e quelle della banca dati Istat del commercio con l'estero (Coe) che rileva l'attività d'import/export dell'impresa.

Il secondo tipo di approccio ha riguardato l'analisi della gestione economica dell'impresa, in particolare per alcune imprese attive in Asia, che sebbene non presentino occupazione, conseguono un volume d'affari, per alcune anche superiore a 100 mila euro. La fonte utilizzata per l'analisi è quella dei bilanci d'esercizio dell'anno 2011 redatti secondo gli artt. 2424 e 2425 del Codice Civile. Le voci del conto economico prese in esame sono le variabili di ricavi concorrenti alla formazione del volume d'affari quali:

- 1) A.1 - Ricavi delle vendite e delle prestazioni;
- 2) A.5 - Totale altri ricavi e proventi con separata indicazione dei contributi in conto esercizio.

L'attenzione viene posta sulla voce "Ricavi delle vendite e delle prestazioni" che, qualora fosse stata nulla o molto bassa, evidenziava che i ricavi concorrenti alla formazione del volume d'affari erano quelli relativi alle voci "altri ricavi e proventi" e/o "contributi in c/esercizio".

I primi riguardano la gestione accessoria dell'impresa cioè non l'attività principale quali ad esempio: plusvalenze derivanti dalla vendita di beni strumentali, sopravvenienze attive che non derivano da errori di registrazione, eccetera; mentre i secondi si riferiscono invece ai sussidi che lo Stato o le Regioni, gli enti locali o altri enti pubblici erogano alle imprese. La valorizzazione esclusivamente della voce A.5 non presuppone necessariamente la presenza dell'addetto nell'impresa.

Qualora fosse stata valorizzata la voce A.1 “Ricavi delle vendite e delle prestazioni” riguardante la gestione principale dell’impresa, non si poteva escludere la presenza dell’addetto.

Per la natura delle voci sopra menzionate si potrebbe ipotizzare che è plausibile possano esistere imprese con alto volume d’affari ma con ricavi delle vendite e delle prestazioni nulli e senza addetti.

Anche le voci di costo del conto economico sono state analizzate e in particolare la voce di bilancio B.7 “Costi per servizi”, che per moltissime imprese evidenziava valori molto alti. Tale voce comprende oltre alle spese di gestione anche i costi relativi ai compensi erogati ad esempio ai collaboratori, ai professionisti per consulenze tecniche, legali, fiscali, amministrative e commerciali e revisione contabile; costi per lavorazioni eseguite da terzi, amministratori, sindaci, costi per il personale distaccato presso l’impresa e dipendente da altre imprese; prestazioni di personale esterno, provvigioni e rimborsi spese ad agenti e rappresentanti, eccetera. Purtroppo dai dati di bilancio forniti, non è possibile estrarre e analizzare singolarmente le parti sopra menzionate, in particolare quelle che si riferiscono ai compensi erogati per prestazioni di lavoro. Pertanto la variabile “Costi per servizi”, utile come *proxy*, qualora fosse stata nulla simultaneamente a quella che si riferisce ai “Costi della produzione per materie prime, sussidiarie, di consumo e di merci” (voce B.6 del conto economico) poteva eventualmente confermare l’assenza dell’addetto e poteva giustificare il mancato utilizzo di personale per la produzione o l’acquisto di beni e servizi nell’esercizio oggetto di analisi perché realmente non ha prodotto o acquistato merci.

Una ulteriore voce di bilancio utile a segnalare eventuali prestazioni di lavoro nell’azienda è quella del “Totale costi personale (B.9)”; che riguarda le retribuzioni erogate ai dipendenti nell’esercizio di competenza. La valorizzazione di tale variabile indica la presenza di personale dipendente (evidenziando quindi indirettamente ritardi o sotto coperture di altre fonti amministrative di natura contributiva).

L’analisi ha poi preso in esame l’attività economica svolta dall’impresa: se da un lato per le imprese nel commercio al minuto l’addetto non può mancare, sulle imprese del settore manifatturiero, costruzioni ed estrazione, è stato calcolato un indicatore, dato dal rapporto tra i costi per materie prime, sussidiarie, di consumo e merci e i ricavi delle vendite e delle prestazioni. Per queste attività un rapporto minore o uguale ad una certa soglia percentuale (fissata pari a 60 sulla base di evidenze empiriche) indicherebbe la necessità di dover avere addetti per svolgere l’attività principale; al contrario un rapporto superiore alla soglia rende ammissibile l’assenza dell’addetto in virtù dell’elevata incidenza di questa tipologia di costi rispetto al livello dei ricavi.

Alla luce dei risultati ottenuti dalle analisi svolte, i criteri individuati per stabilire la probabilità di esistenza dell’addetto (e in particolare almeno dell’indipendente) sono stati generalizzati per tutte le società di capitali senza addetti (e non in gruppo) articolando il processo in tre fasi.

Nella fase I si integrano le informazioni sulle date di cessazione e sulle procedure concorsuali (dal registro Asia); sulle variabili “Costi per il personale”, “Costi per materie prime, sussidiarie, di consumo e merci”, “Costi per servizi”, “Ricavi delle vendite e delle prestazioni” (dai bilanci d’esercizio); sulla presenza dei flussi d’import/export (dal Coe).

Nella fase II vengono delineate le regole di eleggibilità dell’impresa ai fini della presenza/assenza dell’addetto. Le regole operano a cascata e si basano sulla verifica delle informazioni integrate nella fase I secondo la sequenza descritta nella Tavola 3.1.

Tavola 3.1 - Regole di eleggibilità all'addetto

VINCOLI IMPOSTI	Risultato
Regola 1 - Imprese cessate	No addetto
Regola 2 - Imprese con procedure concorsuali in atto	No addetto
Regola 3 - Imprese che hanno i ricavi delle vendite e delle prestazioni nulli	No addetto
Regola 4 - Imprese che hanno simultaneamente costi per materie prime, sussidiarie, di consumo e merci nulle e costi per servizi nulli	No addetto
Regola 5 - Imprese senza deposito bilancio d'esercizio	No addetto
Regola 6 - Imprese che hanno i costi del personale > 100 euro	Si addetto
Regola 7 - Imprese che svolgono commercio al minuto	Si addetto
Regola 8 - Imprese con presenza dei flussi d'import/export (dati Coe)	Si addetto
Regola 9 - Imprese manif. e costruz. con indicatore rapp_mat* > 60%	No addetto
Regola 10 - Imprese con indicatore rapp_mat* ≤ 60%	Si addetto

* rapp_mat = (Costi per materie prime, sussidiarie, di consumo e di merci/Ricavi delle vendite)*100

La procedura genera tre insiemi:

- a) imprese a cui assegnare l'addetto (SI addetto);
- b) imprese a cui confermare l'assenza di addetto (NO addetto);
- c) imprese residue sulle quali non si possono fare ipotesi (Imprese residue).

Nella fase III, sulla base dell'appartenenza delle imprese ai tre insiemi è stata successivamente ricostruita l'informazione sui lavoratori indipendenti attraverso l'applicazione di regole deterministiche di identificazione del lavoratore indipendente analoghe a quelle utilizzate negli altri insiemi considerati della popolazione delle società di capitali. In particolare ai fini dell'identificazione dell'indipendente all'interno della struttura informativa disponibile, in assenza di segnali di flussi remunerativi tra la società e i suoi membri, gli insiemi di cui al punto a) e c) sono stati ulteriormente approfonditi utilizzando due tipi di informazione. Da un lato è stato rafforzato il segnale dell'esistenza dell'indipendente grazie alla presenza di soci amministratori dichiarati nella fonte integrata Studi di settore-quadro A; dall'altro la selezione ha escluso i soci con partita Iva. La procedura applicata ha permesso di assegnare l'indipendente a circa il 35 per cento delle iniziali società di capitali prese in esame.

3.4.5 Il calcolo delle posizioni lavorative dell'occupazione indipendente, in media annua e al dicembre

Al fine di garantire omogeneità di misura rispetto alla struttura informativa di tipo LEED, dopo avere identificato l'occupato indipendente, si pone il problema della misura della sua posizione lavorativa. A differenza delle altre tipologie occupazionali, caratterizzate da un rapporto di lavoro temporalmente conosciuto in maniera certa (data inizio e fine contratto per alcune tipologie di lavoro, soprattutto quelle esterne, copertura contributiva/retributiva settimanale o mensile per i dipendenti), per definizione la posizione occupata come indipendente si può associare strettamente alla durata di esistenza dell'impresa (sotto l'ipotesi che ad esempio per le società la struttura societaria si mantenga stabile nell'anno).

In generale, per assegnare una posizione media annua in un anno di riferimento è possibile misurare la durata dell'impresa a partire dalle date inizio e fine attività assegnate dal processo di Asia. Per coerenza con la misura adottata per le altre tipologie occupazionali, il metodo trasforma le date in settimane, ottenendo vettori di presenza/assenza nella settimana per ciascun mese. Da tali presenze si calcolano le posizioni mensili e poi la posizione in media annua come definita nel

Paragrafo 3.3.3. Per il calcolo della posizione media viene utilizzato lo stesso metodo per tutte le forme giuridiche delle imprese, sia per il calcolo relativo all'indipendente in senso stretto che per il familiare. Fa eccezione, per le imprese individuali e per parte delle società di persone, la misura della posizione lavorativa del familiare che si fa corrispondere al periodo in cui sono stati versati i contributi dal titolare presso la gestione Artigiani e Commercianti dell'Inps, secondo un algoritmo di quadratura per il quale è valida l'espressione: la durata della posizione lavorativa del familiare \leq la durata della posizione lavorativa del titolare = la durata dell'impresa (Asia)¹⁰¹.

Per la misura di una posizione puntuale quale quella del Censimento (31 dicembre 2011) il problema della misura si semplifica basandosi sulla presenza dell'indipendente identificato e sull'esistenza dell'impresa alla data censuaria.

Alcune peculiarità riguardano l'eventualità che lo stesso individuo abbia avuto nel corso dell'anno una posizione lavorativa sia come dipendente che come indipendente; in tali casi non è ammessa la contemporanea doppia posizione, entrambe le posizioni esistono se sono occupate in settimane/mesi differenti in modo tale che in totale nell'anno la posizione lavorativa occupata da un addetto non possa superare 1. Al dicembre, invece, la posizione non potrà che essere solo di un tipo o come dipendente o come indipendente, dove il primo tipo ha priorità sul secondo. Una ulteriore peculiarità riguarda il caso di imprese organizzate in gruppo dove la dicotomia (indipendente sì/no) è assicurata rispetto al gruppo e non rispetto alla singola impresa del gruppo; per consentire comunque di potere fare statistiche rispetto all'impresa, si assegna un peso a ciascun individuo presente in più imprese all'interno del gruppo dove tale peso è pari all'inverso del numero di imprese in cui questi ricoprirebbe la posizione di indipendente (peso=1/n dove n è il numero di imprese nel gruppo in cui si trova).

3.5 L'occupazione esterna

3.5.1 I lavoratori parasubordinati: aspetti definitivi

I lavoratori parasubordinati si collocano in un'area grigia fra l'occupazione dipendente e l'occupazione indipendente, presentando caratteristiche intermedie tra quelle del lavoro subordinato e quelle del lavoro autonomo. Essi anche dal punto di vista amministrativo risultano a metà strada fra i lavoratori dipendenti e autonomi: fiscalmente (a partire dal 2001) i loro redditi sono assimilati a quelli da lavoro dipendente, giuridicamente hanno lo stesso regime del lavoro autonomo¹⁰².

Questo concetto di lavoro parasubordinato rientrante nella categoria del personale esterno, già presente nel Censimento 2001, nel Censimento 2011 viene ridefinito con l'obiettivo di seguire i mutamenti nel mercato del lavoro intervenuti nell'ultimo decennio, soprattutto connessi all'espansione e all'articolazione a livello normativo delle figure contrattuali dei parasubordinati. Il lavoro di ridefinizione ha coniugato l'esigenza di sfruttamento delle potenzialità offerte dalle fonti amministrative con quella di armonizzazione delle statistiche economiche che utilizzano fonti amministrative sull'occupazione.

Nelle definizioni amministrative dei lavoratori parasubordinati fornite dagli enti per chiarire l'universo dei lavoratori oggetto dell'atto amministrativo (ad esempio i lavoratori rientranti nella Gestione separata Inps), si rintracciano alcune caratteristiche peculiari di tale tipologia di lavoratori già presenti nelle definizioni statistiche internazionali degli "outworkers" (cfr. Paragrafo 3.1).

¹⁰¹ L'algoritmo è stato sviluppato da Dario Ercolani.

¹⁰² Non è applicabile il principio di automaticità delle prestazioni.

Dalla lettura congiunta degli aspetti che emergono dalle definizioni statistiche e amministrative, è possibile delineare le caratteristiche principali dei rapporti di lavoro di tipo parasubordinato.

Essi possono essere identificati in tutte quelle "collaborazioni", intese come lo "svolgimento di ogni attività finalizzata al raggiungimento di scopi determinati da altri"¹⁰³, che sono:

- regolate da un contratto concluso fra le parti in forma scritta o comunque da un incarico formale conferito dal committente ed accettato dal lavoratore;
- svolte in modo autonomo dal collaboratore, che è libero di organizzare la propria attività determinando i luoghi, i tempi e i modi di svolgimento dell'incarico;
- sostenute dal lavoratore prendendo in carico alcuni eventuali costi di produzione (vitto, alloggio, spostamenti);
- svolte in modo coordinato con il committente, il quale può fornire indicazioni funzionali alle esigenze della sua organizzazione aziendale nei limiti dell'autonomia professionale del collaboratore, che non è inserito strutturalmente nell'organizzazione gerarchica dell'impresa;
- remunerate in relazione al risultato, indipendentemente dal tempo impiegato per realizzarlo;
- remunerate con retribuzione periodica e prestabilita.

Poste queste caratteristiche generali, la collaborazione può essere continuativa nel tempo (Co.Co.Co) oppure assumere un carattere più occasionale (mini Co.Co.Co), può essere vincolata ad un progetto (Co.Co.Pro) o meno (Co.Co.Co della PA), presentare caratteristiche anche molto diverse quanto ad attività lavorative svolte e tuttavia tali da non poter essere considerate "interne" ad un'impresa né tali da dar vita ad un'attività di tipo indipendente.

3.5.2 Le fonti amministrative per l'individuazione dei lavoratori parasubordinati

La ricognizione delle fonti che registrano dati riferiti all'universo dei parasubordinati ha portato all'individuazione degli archivi amministrativi di maggiore rilievo, quali: la Banca dati Parasubordinati dell'Inps, che raccoglie i dati dei lavoratori che hanno avuto almeno un versamento contributivo per lavoro parasubordinato nel corso dell'anno, la Banca dati integrata sui parasubordinati dell'Inail che registra gli eventi di assunzione, proroga, trasformazione e cessazione dei lavoratori parasubordinati, e l'archivio del modello 770 dell'Agenzia delle Entrate, per la sezione relativa ai collaboratori coordinati e continuativi, in cui confluiscono le dichiarazioni fiscali dei lavoratori che nell'anno precedente a quello di dichiarazione fiscale hanno avuto un reddito da lavoro come Co.Co.Co.

Fra queste è stata privilegiata la fonte Inps, in quanto le altre pur fornendo un contenuto informativo comparabile in termini di universo di riferimento e parzialmente complementare in termini di variabili, presentano dei limiti: da una parte l'archivio Inail non permette una disaggregazione fine dei diversi tipi di contratti inclusi nell'ampia tipologia dei parasubordinati, dall'altra l'archivio dell'Agenzia delle Entrate presenta problemi di tempestività dovuti alla strutturale tempistica delle dichiarazioni fiscali.

La Banca dati Parasubordinati dell'Inps è il frutto dell'integrazione di diversi archivi: l'archivio del modello UniEMens, ovvero delle denunce retributive mensili, l'archivio del modello F24, ovvero dei versamenti contributivi mensili effettuati dai committenti, e l'archivio delle domande di iscrizione alla Gestione separata.

Gli eventi amministrativi che costituiscono l'universo sono tutti i versamenti contributivi mensili ricevuti nell'anno t dai lavoratori contribuenti alla Gestione separata, fondo pensionistico che nasce

¹⁰³ Cass. 21 febbraio 1998, sent. n. 1897.

con la L. 335/95 come gestione assicurativa a carattere “residuale”, rivolta a quei redditi da lavoro non assicurati presso altre gestioni pensionistiche Inps o presso casse professionali.

Occorre precisare che la banca dati individua due tipi di contribuzioni distinte, che danno vita a due archivi differenti: quello dei contribuenti collaboratori e quello dei contribuenti professionisti. L'Inps definisce contribuenti collaboratori quei lavoratori che hanno avuto almeno un versamento contributivo per lavoro parasubordinato nell'anno, dove il versamento dei contributi è effettuato dal committente, persona fisica o soggetto giuridico con cui il lavoratore collabora, mentre considera contribuenti professionisti quei lavoratori che hanno avuto almeno un versamento contributivo per lavoro parasubordinato nell'anno dove il versamento è effettuato dal lavoratore stesso.

Per l'individuazione del personale esterno parasubordinato delle imprese è stato sfruttato solo il primo archivio, composto da circa 10 milioni di denunce mensili per fornitura annuale.

L'archivio, di tipo LEED, permette di ricostruire informazioni sul collaboratore e sull'impresa committente (e sul loro legame) attraverso la disponibilità dei codici fiscali, e fornisce numerose variabili caratterizzanti il rapporto di lavoro quali: il tipo di rapporto, il tipo di attività, la data inizio e la data fine attività a cui si riferiscono i compensi, l'imponibile previdenziale, la contribuzione, l'anno e il mese di competenza.

In particolare il tipo di rapporto si è rivelata una variabile strategica a fini classificatori, in quanto ha permesso di individuare nel dettaglio le singole forme contrattuali in cui si presenta l'occupazione parasubordinata studiarne le caratteristiche e procedere a raggruppamenti omogenei.

3.5.3 La classificazione dei parasubordinati

La variabile tipo di rapporto di lavoro, che assume numerose modalità in base alle diverse figure tenute all'iscrizione alla Gestione separata Inps per il versamento dei propri contributi, si è rivelata fondamentale per la classificazione dei parasubordinati presente nella fonte Inps.

A partire dallo studio normativo, si sono potute delineare le caratteristiche principali delle singole figure riportate nel Prospetto 3.1, e rintracciare nell'ambito del profilo generale dell'occupazione parasubordinata, caratterizzato da autonomia nell'organizzazione del lavoro e coordinamento con il committente, alcune caratteristiche peculiari di ognuna di esse, anche sotto il profilo fiscale e previdenziale.

Prospetto 3.1 - Tipi di rapporto di lavoro

TIPO DI RAPPORTO	TIPO DI RAPPORTO
1A - Amministratore di società	07 - Venditore porta a porta
1B - Sindaco di società	08 - Collaboratori occasionali (mini Co.Co.Co)
1C - Revisore di società	09 - Rapporti occasionali autonomi
1D - Liquidatore di società	10 - Co.Co.Co dei titolari di pensione
1E - Amministratore e legale rappresentante	11 - Co.Co.Co presso la PA
02 - Collaboratore di giornali	12 - Rapporti Co.Co.Co prorogati
03 - Partecipazione a collegi e commissioni	13 - Associati in partecipazione
04 - Amministratore di enti locali	14 - Medici in formazione specialistica
05 - Dott. di ricerca, assegno, borsa di studio	16 - Lavoro occasionale accessorio – voucher
06 - Co.Co.Co (con contratto a progetto)	17 - Consulente parlamentare

Le evidenze empiriche hanno mostrato la prevalenza di due figure principali, quella dei collaboratori coordinati continuativi con contratto a progetto (codice 06) e quella degli

amministratori di società (codice 1A), che complessivamente pesano per circa l'80 per cento dei rapporti di lavoro nel campo d'osservazione (cfr. Paragrafo 3.5.4).

La prima figura, più numerosa e maggiormente rispondente da un punto di vista concettuale alla definizione di personale parasubordinato, comprende i rapporti di collaborazione coordinata e continuativa che, a partire dal 2003, a seguito della cosiddetta riforma Biagi, devono obbligatoriamente *“essere riconducibili a uno o più progetti specifici o programmi di lavoro o fasi di esso determinati dal committente e gestiti autonomamente dal collaboratore in funzione del risultato, nel rispetto del coordinamento con la organizzazione del committente e indipendentemente dal tempo impiegato per l'esecuzione della attività lavorativa”* (D.Lgs 10 settembre 2003, n. 276, art. 61). Inoltre, dal punto di vista fiscale i redditi da collaborazione coordinata e continuativa, considerati fino al 2000 redditi da lavoro autonomo, da gennaio 2001 sono redditi assimilati a quelli da lavoro dipendente.

La figura degli amministratori di società comprende gli incarichi di gestione di un'impresa assunti da chi, all'interno di una società, ha funzioni di dirigenza e organizzazione delle attività necessarie per l'attuazione dell'oggetto sociale, nel ruolo di socio o meno, a seconda dei diversi tipi di società. Le due figure si distinguono anche in base ad alcuni risultati emersi dalle analisi preliminari svolte per testare le ipotesi classificatorie: pur essendo entrambe caratterizzate soprattutto da monocommittenza¹⁰⁴, gli amministratori rispetto ai collaboratori lavorano per periodi più lunghi, con maggior continuità¹⁰⁵ e con compensi nettamente superiori (più del 50 per cento lavora per 12 mesi contro circa il 20 per cento dei collaboratori, circa l'80 per cento per periodi continui superiori a 6 mesi contro il 35 per cento dei collaboratori, più del 30 per cento riceve compensi superiori a 30 mila euro contro il 5 per cento dei collaboratori).

La lettura congiunta delle differenze riscontrate sia rispetto alla natura del contratto sia alle caratteristiche emerse dalle analisi, ha portato a tenere distinte queste due figure principali e a ricondurre ad esse le restanti tipologie di rapporti presenti nella classificazione originaria della fonte Inps (Prospetto 3.1).

Nella macro tipologia degli “Amministratori” sono state ricondotte anche le figure dell'amministratore degli enti locali (codice 04) e dell'amministratore legale rappresentante (codice 1E), in considerazione dell'analoga funzione di responsabilità interna al processo di produzione dell'impresa.

Nella macro tipologia dei “Collaboratori” sono confluite molte delle restanti figure, tutte accomunate dalla caratteristica di fornire un contributo di collaborazione più circoscritto nel tempo e nelle attività svolte. Sono dunque incluse tutte le collaborazioni coordinate e continuative: oltre ai Co.Co.Pro (06), i Co.Co.Co dei titolari di pensione di vecchiaia (10), i Co.Co.Co della PA (11), i Co.Co.Co prorogati (12), i Collaboratori di giornali (02), le collaborazioni occasionali (08) (anche dette mini Co.Co.Co, ovvero collaborazioni di durata inferiore ai trenta giorni e con un compenso non superiore ai cinquemila euro con lo stesso committente). Inoltre nella stessa macro tipologia rientrano i rapporti occasionali autonomi (09), le figure del sindaco, revisore e liquidatore di società (1B, 1C, 1D), i partecipanti a collegi e commissioni (03), i venditori porta a porta (07) e i consulenti parlamentari (17).

¹⁰⁴ Alcune analisi preliminari hanno portato a definire una variabile “committenza”, che indica se il lavoratore nell'anno di riferimento ha avuto un solo committente o più committenti; essa è costruita con riferimento alla testa ma può essere attribuito anche della posizione lavorativa e permettere di caratterizzare i diversi tipi di rapporto.

¹⁰⁵ Alcune analisi preliminari hanno portato a definire una variabile “continuità del periodo di attività dei contratti nell'anno”, che indica se il periodo di attività del collaboratore con la stessa impresa e lo stesso tipo di rapporto presenta interruzioni o meno (a partire dalla presenza/assenza di attività nei 365 giorni dell'anno, un periodo è definito non continuo quando vi è almeno un giorno di assenza di attività che separa due periodi di presenza).

Altre tipologie di rapporti, dalle connotazioni ben definite, sono state mantenute distinte da questi due macro raggruppamenti. Si tratta dei contratti degli associati in partecipazione, lavoratori che in linea generale partecipano agli utili di un'impresa dietro il corrispettivo di un apporto di lavoro, di capitale o misto, e che nel caso dei contribuenti alla Gestione separata è esclusivamente costituito da una prestazione lavorativa; essi presentano una discreta consistenza numerica nell'archivio (poco meno del 5 per cento nel campo d'osservazione censuario) e assumono un certo interesse per le analisi economiche sul mercato del lavoro. Sono stati tenuti distinti anche i buoni voucher ovvero forme di lavoro inquadrato nella tipologia del lavoro occasionale di tipo accessorio, introdotte per regolamentare attività lavorative meramente occasionali, prevalentemente utilizzate nel settore agricolo e domestico, che facilmente possono confluire nel lavoro sommerso.

Queste ultime due figure, insieme ad altre del tutto residuali considerando l'universo di riferimento (dottorato di ricerca, assegno, borsa di studio - codice 05, medici in formazione specialistica - codice 14) sono confluite nella macro tipologia di "Altro lavoratore esterno". La classificazione finale adottata per la diffusione è quindi costituita da tre tipologie di lavoratori esterni: gli amministratori, i collaboratori e gli altri lavoratori esterni.

3.5.4 Il trattamento statistico

Il trattamento della fonte amministrativa a fini statistici ha riguardato diversi aspetti: la definizione dell'universo di riferimento in termini temporali e di armonizzazione con il sistema classificatorio dell'occupazione; la definizione dell'unità di analisi (posizione lavorativa) necessaria alla misura dell'occupazione; la costruzione delle variabili di diffusione.

Definizione dell'universo di riferimento

La Banca dati Parasubordinati Inps registra le comunicazioni retributive mensili pervenute all'Inps - Gestione separata - in ciascuno dei dodici mesi dell'anno di riferimento secondo un criterio di cassa, ovvero al pagamento del compenso al collaboratore (è previsto che la comunicazione sia trasmessa dal committente entro il mese successivo al pagamento).

Ogni evento di comunicazione retributiva è caratterizzato dal mese in cui la comunicazione avviene e dalle date puntuali di inizio e fine attività cui si riferisce il compenso erogato. Sfruttando le due date è possibile individuare il periodo delle attività di collaborazione, che può non coincidere con i mesi di pagamento e collocarsi anche fuori dall'anno.

Per individuare le collaborazioni realizzate esattamente nell'anno di riferimento, quindi, si è deciso, di perseguire il criterio di competenza, considerando solo le collaborazioni che cadono, in tutto o in parte, nell'anno t . Una volta verificata la qualità delle date presenti nella fonte, si sono esclusi i periodi di collaborazione antecedenti e successivi all'anno t .

A partire dalle comunicazioni mensili trattate in base al criterio di competenza, si sono derivati i singoli rapporti di lavoro svolti nell'anno, individuati dalla chiave codice fiscale del lavoratore, codice fiscale dell'impresa e tipo di rapporto di lavoro (nelle modalità indicate nel Prospetto 3.1), e si sono ricondotte le variabili dell'archivio a questa unità di analisi. Le date delle attività di collaborazione che, retribuite con *tranche* soprattutto mensili, ma anche bimestrali, trimestrali o altro, sono presenti nell'archivio suddivise nelle singole comunicazioni retributive e in alcuni casi ripetute anche se relative alla stessa chiave, sono state accorpate e deduplicate per singolo rapporto di lavoro, dando luogo ad un unico periodo. Le altre variabili sono state ridotte per somma (nel caso dell'imponibile) o per prevalenza (nel caso del tipo di attività).

Il rapporto di lavoro così ottenuto può essere anche interpretato come una *proxy* del contratto di lavoro. Dalla conversione delle denunce retributive in rapporti di lavoro si ottiene una base dati intermedia di circa 1,75 milioni di unità.

Procedendo da qui nell'individuazione dell'universo di riferimento, coerentemente con il sistema classificatorio del Censimento che distingue l'occupazione in dipendente, indipendente ed esterna, è stata effettuata una selezione dell'universo dei rapporti di lavoro parasubordinati presenti nell'archivio in base al criterio di competenza.

Sono stati esclusi i contratti dei lavoratori titolari di partita Iva dal momento che rappresentano imprese individuali incluse tra le imprese attive del registro Asia; i contratti con committenti non appartenenti al campo di osservazione del registro Asia, ovvero sono state escluse le collaborazioni nella PA, nel settore dell'agricoltura e nelle unità non profit; infine sono stati esclusi i contratti di quei parasubordinati già classificati come indipendenti delle imprese nel sistema dell'occupazione del Censimento, in quanto soci delle società con cui hanno un rapporto di lavoro. Complessivamente la numerosità dei rapporti di lavoro si riduce di circa il 50 per cento rispetto al valore iniziale.

Definizione dell'unità di analisi

In linea con la misura dell'occupazione dipendente ed indipendente, anche la misura dell'occupazione parasubordinata si basa sul calcolo delle posizioni lavorative. In questo caso, una posizione lavorativa è identificata univocamente dalle variabili: codice fiscale del lavoratore, codice fiscale dell'impresa e tipo di rapporto di lavoro (nelle modalità indicate nel Prospetto 3.1).

Ogni posizione lavorativa è caratterizzata dal periodo in cui si svolge il rapporto di lavoro, ottenuto dalle date di inizio e fine attività presenti in ogni denuncia, espresso in giorni esatti nell'anno; quindi in ogni mese si ha la presenza del rapporto di collaborazione nei singoli giorni.

Da qui, per armonizzare il calcolo delle posizioni rispetto ai restanti tipi di occupazione (dipendenti e indipendenti), si è ricondotta la misura da giorni in settimane, ovvero alla struttura temporale dell'UniEMens, fonte da cui si stima la parte prevalente dell'occupazione. Ogni posizione lavorativa media mensile è ottenuta, quindi, in base alla presenza del rapporto di collaborazione in ogni settimana del mese come definita nel Paragrafo 3.3.3.

In base a questo si sono potute calcolare le posizioni medie annue, che costituiscono la misura dell'occupazione nel registro Asia, e per le esigenze specifiche del Censimento, che deve fotografare l'occupazione al 31 dicembre 2011, si sono conteggiati per convenzione i collaboratori il cui contratto risulta attivo nell'ultima settimana del mese di dicembre 2011.

3.5.5 I lavoratori temporanei (ex interinali)

Un rapporto di lavoro interinale è legato a due imprese: un'azienda di lavoro interinale (impresa fornitrice) e l'impresa a cui il lavoro è somministrato (impresa utilizzatrice). Coerentemente, in Asia-occupazione, sono presenti sia i rapporti di lavoro con le imprese fornitrici con forma occupazionale "Dipendente", sia quelli con le imprese utilizzatrici con forma occupazionale "Somministrato (ex interinale)".

Analogamente a quanto effettuato per i lavoratori parasubordinati, anche per l'individuazione dei lavoratori interinali si è compiuta una ricognizione delle fonti amministrative in grado di cogliere tale importante tipologia di lavoratori esterni. Le fonti utilizzate sono la Banca dati integrata sugli interinali dell'Inail e l'archivio delle denunce mensili UniEMens dell'Inps. Elemento comune alle fonti individuate è la disponibilità dei codici fiscali dell'impresa fornitrice, di quella utilizzatrice e del lavoratore.

La Banca dati integrata sugli interinali dell'Inail combina dati provenienti dalle Comunicazioni obbligatorie del Ministero del Lavoro inoltrate all'Inail e dalla fonte Denuncia Nominativa degli Assicurati (Dna) dell'Inail. Peculiarità di questa fonte è la possibilità di mettere in relazione il

lavoratore interinale sia all'impresa fornitrice, con la quale si instaura un rapporto di somministrazione, sia a quella utilizzatrice, con la quale viene attivata una missione.

Le unità che costituiscono l'universo della banca dati sono le posizioni lavorative interinali per cui è presente una copertura assicurativa presso l'Inail nell'anno t di riferimento.

Considerando come unità d'analisi il lavoratore, nel caso dei dati provenienti dalle Comunicazioni obbligatorie (modello UniSomm), l'universo è composto da tutti i lavoratori interinali soggetti nell'anno ad almeno un evento fra: assunzione, proroga, trasformazione e cessazione del rapporto di lavoro, opportunamente trattato al fine di ricostruire l'intera posizione lavorativa; nel caso dei dati provenienti dalla Dna, da tutti i lavoratori interinali che hanno un rapporto di lavoro che, in base alla data inizio e/o fine, ricade nell'anno.

Se il dato proviene dalle Comunicazioni obbligatorie è possibile individuare delle macro tipologie di lavoratori interinali attraverso lo sfruttamento della variabile "Tipo contratto". Essa può assumere le seguenti modalità: "Lavoro interinale o a scopo di somministrazione a tempo indeterminato" (A.06.00), "Lavoro interinale o a scopo di somministrazione a tempo determinato" (A.06.01), "Contratto di agenzia a tempo indeterminato" (M.01.00) e "Contratto di agenzia a tempo determinato" (M.01.01).

Con riferimento all'unità di rilevazione, la copertura degli archivi oggetto di integrazione nella Banca dati interinali dell'Inail è esaustiva rispetto al campo di osservazione del registro Asia. Sono infatti tenuti a compilare le Comunicazioni tutti i datori di lavoro privati (e pubblici) appartenenti a tutti i settori di attività economica (compresa l'agricoltura, la PA, il settore domestico ed escluse le Forze Armate).

Il contenuto informativo della fonte caratterizzato dalla presenza del codice fiscale del lavoratore interinale, di quello dell'impresa fornitrice e utilizzatrice, dalla data di inizio e fine del rapporto di somministrazione (intercorrente tra agenzia e lavoratore), dalla data di inizio e fine missione (intercorrente tra impresa utilizzatrice e lavoratore), oltre che dal tipo contratto, dalla Posizione assicurativa territoriale (Pat) e dal comune della sede di lavoro è stato oggetto del seguente trattamento statistico.

A partire dalle informazioni temporali presenti nella fonte Inail, in Asia-occupazione è stata ricostruita la variabile Tipo Copertura per ciascuna settimana nel mese (valorizzata unicamente come "X" o "0"), coerentemente con la struttura del database, attraverso un confronto dei dati UniEMens e dei dati dell'Inail.

La costruzione delle variabili Tipo Copertura prevede tre step per ciascun lavoratore interinale:

- 1) costruzione delle variabili Tipo Copertura per ogni rapporto di lavoro definito dalla coppia impresa utilizzatrice e fornitrice. Per ogni rapporto di lavoro presente sull'Inail possono esserci più intervalli di date di inizio e fine rapporto non necessariamente disgiunti. Il Tipo Copertura per una data settimana sarà valorizzato come "X" se la settimana risulta compresa in almeno uno degli intervalli di date;
- 2) confronto con i dati UniEMens relativi alle imprese fornitrici. L'assenza del rapporto di lavoro, *in toto* o in parte, dal lato dell'impresa fornitrice, annulla di conseguenza il rapporto dal lato dell'utilizzatrice. In pratica:
 - ogni record dell'Inail relativo ad una coppia (lavoratore, impresa fornitrice) non presente nell'UniEMens è stato tolto;
 - per le coppie (lavoratore, impresa fornitrice) presenti nell'UniEMens, ogni settimana che risulti non coperta (Tipo Copertura='0') nell'UniEMens sarà non coperta anche in tutti i record dell'Inail relativi a quella coppia;

- 3) costruzione delle variabili Tipo Copertura per ogni rapporto di lavoro definito solamente dall'impresa utilizzatrice. Sono state accorpate le informazioni relative alla coppia (lavoratore-utilizzatrice) nel caso di più aziende fornitrici valorizzando come "X" le settimane in cui ci sia almeno un rapporto.

Una volta costruite le variabili Tipo Copertura, il calcolo delle posizioni lavorative è avvenuto nello stesso modo dei lavoratori con altre forme occupazionali.

3.6 Le informazioni disponibili con la diffusione di Asia-occupazione

La struttura informativa di tipo LEED, aggiornata con cadenza annuale, diventa la base informativa in cui sono registrate le informazioni relative ai due operatori del mercato del lavoro: le imprese e gli individui lavoratori. Possono essere definiti tre livelli informativi: (i) livello Impresa, dove le informazioni di struttura sono perfettamente allineate con quelle aggiornate e diffuse dal registro Asia-imprese; (ii) livello Individuo, per il quale possono essere assegnate variabili demografiche e (iii) variabili di tipo occupazionale più strettamente collegate al rapporto di lavoro dell'individuo in relazione ad una data impresa. La relazione impresa-occupato utilizza due livelli di classificazione: la tipologia di lavoratori e le caratteristiche del rapporto di lavoro che variano al variare della tipologia di lavoro svolto.

Sono state classificate quattro tipologie di lavoratori: indipendenti, dipendenti, lavoratori esterni senza partita Iva, lavoratori somministrati (ex interinali). In particolare, gli indipendenti e i dipendenti, nel loro complesso rappresentano il personale direttamente addetto ad una impresa (interno al processo produttivo), le altre due componenti rappresentano il personale esterno che contribuisce comunque all'input di lavoro dell'impresa.

Per quanto riguarda le variabili demografiche relative al lavoratore e presenti per tutte le tipologie di lavoratori si fa riferimento: all'età, al sesso e al paese di nascita del lavoratore. I dati vengono diffusi secondo le seguenti modalità: "sesso" (M, F), "classe di età" (15-29 anni, 30-49 anni, 50 anni e più), e "paese di nascita" (Italia, Paesi EU27 esclusa l'Italia, Paesi extra-EU27). Le tre variabili demografiche sono state derivate tramite l'applicazione di un algoritmo che si applica su sottostringhe estratte dal codice fiscale della persona fisica (del lavoratore). Tali modalità non sono presenti per tutti i lavoratori; per una percentuale molto piccola di posizioni lavorative (circa lo 0,2 per cento) non è stato possibile assegnare nessuna di queste variabili per una mancata identificazione dell'individuo dovuta o ad errore di codice fiscale o perché si tratta di lavoratori imputati dal processo di stima.

Le informazioni demografiche, a regime, saranno arricchite con ulteriori informazioni, al momento non disponibili, derivabili dal Sistema integrato di microdati (Sim) disponibili in Istat: il titolo di studio e la residenza.

Per quanto riguarda le caratteristiche del rapporto di lavoro, esse variano a seconda della tipologia di lavoratore.

Per i *lavoratori dipendenti* sono state diffuse le seguenti quattro variabili:

- tipologia contrattuale, che si articola nelle seguenti modalità: dirigente, quadro, operaio, impiegato, apprendista e altro dipendente (ad esempio i viaggiatori o piazzisti, i lavoratori autonomi dello spettacolo, i giornalisti professionisti, eccetera);
- regime orario di lavoro, che si articola nelle seguenti modalità: tempo pieno, tempo parziale (di tipo orizzontale, verticale e misto);

- carattere occupazionale, che si articola nelle seguenti modalità: tempo indeterminato, tempo determinato (che comprende i contratti a tempo determinato e quelli di tipo stagionale);
- tipologia di beneficiari di politiche attive¹⁰⁶.

Per i *lavoratori indipendenti* è stata differenziata la tipologia di indipendente “in senso stretto” che corrisponde al titolare nel caso delle imprese individuali o ai soci per le società¹⁰⁷, e di indipendente familiare/coadiuvante presente esclusivamente nelle imprese individuali e, in parte meno rilevante, nelle società di persone.

Tra i *lavoratori esterni*, senza partita Iva, sono stati classificati due sottoinsiemi provenienti da due fonti differenti: i parasubordinati-collaboratori che provengono sostanzialmente dalla fonte Inps-Gestione separata e gli autonomi che provengono da fonte Enpals e che presentano i tipi di contratto di lavoro a progetto, di associato in partecipazione, di lavoratore autonomo e che non risultano nella Gestione ordinaria Inps dei lavoratori dipendenti. Per i lavoratori esterni sono state diffuse le seguenti tre variabili:

- il tipo di lavoratore esterno che si articola nelle seguenti modalità: amministratore, collaboratore e altro lavoratore esterno (ad esempio i contratti degli associati in partecipazione, i buoni voucher, i contratti derivanti dalla fonte Enpals, eccetera)¹⁰⁸;
- la classe di compenso totale percepito nell'anno: il compenso totale deriva dalla variabile imponibile previdenziale ed indica l'importo dei compensi percepiti nel mese di denuncia, sul quale si applica l'aliquota di contribuzione, ed include i contributi a carico del lavoratore (1/3), non quelli a carico del datore di lavoro (2/3). Il compenso annuo percepito, ottenuto dalla somma dei compensi mensili, è stato raggruppato in sei classi: “fino a 1.000 euro”, “1.000 – 5.000 euro”, “5.000 – 10.000 euro”, “10.000 – 25.000 euro”, “25.000 – 50.000 euro”, “50.000 euro e più” (classi chiuse a sinistra);
- il periodo di attività dei contratti nell'anno, calcolato sulla base dei giorni di inizio e fine contratto a cui si riferiscono i rapporti di lavoro; i giorni sono convertiti in mesi e quindi in classi di mesi come segue: “fino a un mese” (0-31 giorni), “2-6 mesi” (32-186 giorni), “7-11 mesi” (187-337 giorni), “12 mesi” (338-365 giorni). Occorre tenere presente che tale variabile indica appunto l'arco temporale entro il quale, per contratto, il collaboratore è tenuto a svolgere il progetto/collaborazione, e non il reale numero di giorni lavorati che, come è noto, non è determinabile, essendo discrezionale per il collaboratore la modalità di svolgimento dell'attività lavorativa all'interno del periodo di collaborazione.

I *lavoratori somministrati* o temporanei (ex interinali) sono lavoratori somministrati da aziende fornitrici e presenti (utilizzati) in aziende utilizzatrici in qualità di lavoro esterno sulla base di un contratto di lavoro di durata variabile. La variabile diffusa è il periodo di attività svolto nell'impresa utilizzatrice, calcolato sulla base dei giorni di inizio e fine contratto in cui si è realizzata la somministrazione del lavoratore da parte dell'impresa fornitrice, tale periodo di attività espresso in giorni è stato convertito in mesi e quindi in classi di mesi come segue: “fino a un mese” (0-31 giorni), “2-6 mesi” (32-186 giorni), “7-11 mesi” (187-337 giorni), “12 mesi” (338-365 giorni).

¹⁰⁶ Per i dettagli si veda il Paragrafo 3.3.4.

¹⁰⁷ Per i dettagli si veda il Paragrafo 3.4.2.

¹⁰⁸ Per una dettagliata descrizione si veda il Paragrafo 3.5.3.

Capitolo 4

Asia-unità locali¹⁰⁹

4.1 Il processo di realizzazione del registro delle unità locali

Il registro Asia-imprese aggiorna con criteri statistici le variabili demografiche e di stratificazione delle imprese attive nell'anno di riferimento. La normativa comunitaria in materia di registri di impresa utilizzati a fini statistici (n. 177/2008)¹¹⁰ richiede tuttavia, per le imprese che esercitano le proprie attività in più luoghi, anche la registrazione e l'aggiornamento di un archivio delle singole unità locali. L'unità locale corrisponde ad una impresa, o ad una sua parte, situata in un luogo topograficamente identificato. In tale luogo, o a partire da esso, una o più persone lavorano (a tempo pieno o parziale) per conto della medesima impresa, esercitando una o più attività economiche.

Il registro statistico delle unità locali delle imprese attive, denominato Asia-unità locali, è stato realizzato nel dicembre del 2006 e a partire da tale data è aggiornato annualmente. Così come per il registro delle imprese, la realizzazione e l'aggiornamento del registro delle unità locali avviene utilizzando informazioni di natura amministrativa e statistica. In particolare, il processo prevede la creazione di una base informativa di unità locali, ottenuta attraverso un'attività di normalizzazione e di integrazione di record presenti nei singoli archivi amministrativi, in parte già utilizzati per la costruzione di Asia-imprese, e in fonti statistiche.

L'individuazione delle unità locali attive alla data di riferimento e la determinazione dei relativi caratteri ha richiesto il superamento di difficoltà essenzialmente riconducibili alla scarsità di informazioni direttamente riferibili alla definizione di unità locale, negli archivi amministrativi disponibili. La carenza informativa è stata compensata con l'implementazione di un'indagine annuale per le unità locali di imprese di grandi dimensioni (Iulgi), finalizzata a verificare sul campo lo stato di attività e gli altri caratteri delle unità locali. La stima dei caratteri del registro è quindi effettuata attraverso due distinti processi produttivi: (i) l'utilizzo delle informazioni dell'indagine per le unità locali delle imprese di grande dimensione; (ii) un trattamento statistico delle informazioni presenti negli archivi amministrativi per le unità locali delle imprese di piccola e media dimensione e delle imprese non rispondenti all'indagine.

4.1.1 L'integrazione delle fonti

La base informativa per la costruzione del registro delle unità locali è realizzata integrando informazioni di natura amministrativa e statistica. I principali archivi amministrativi che forniscono informazioni sugli indirizzi nei quali l'impresa svolge la propria attività sono:

- gli archivi gestiti dall'Agenzia delle Entrate per conto del Ministero dell'Economia e delle Finanze, quali l'Anagrafe tributaria, le dichiarazioni annuali delle imposte indirette, le dichiarazioni dell'imposta regionale sulle attività produttive (Irap), eccetera;
- gli archivi gestiti dalle Camere di commercio, industria, artigianato e agricoltura, quali il Registro delle imprese e il Modello unico di dichiarazione ambientale (Mud);

¹⁰⁹ A cura di Simonetta Cozzi e Danila Filipponi. Autori del capitolo:

- Paragrafo 4.1: Danila Filipponi (Paragrafi 4.1 e 4.1.2), Carla Runci (Paragrafo 4.1.1), Simonetta Cozzi (Paragrafo 4.1.3);

- Paragrafo 4.2: Davide Di Cecco, Danila Filipponi e Irene Rocchetti.

¹¹⁰ Il registro Asia è descritto nel Capitolo 2 del presente fascicolo.

- l'Archivio delle dichiarazioni retributive e contributive (UniEMens) gestito dall'Istituto Nazionale della Previdenza Sociale;
- l'Archivio delle utenze telefoniche gestito dalla Consodata S.p.a. (Gruppo SEAT-Pagine Gialle);
- l'Archivio degli istituti di credito gestito dalla Banca d'Italia;
- l'Archivio della Grande Distribuzione gestito dalla società AC-Nielsen.

Come fonti statistiche utilizzate per definire il registro, è utilizzata l'Indagine sulle unità locali delle grandi imprese (Iulgi), implementata proprio per supportare la realizzazione del registro. Il campo di osservazione dell'indagine varia di anno in anno secondo le esigenze conoscitive che emergono anche in relazione alle fonti amministrative disponibili e alla loro qualità.

Tutti i record presenti nelle diverse fonti di input sono abbinati attraverso chiavi comuni (codice fiscale ed indirizzo) al fine di ottenere grappoli di record che identificano la stessa unità locale. Tali grappoli costituiscono la base informativa per la successiva scelta e assegnazione dei caratteri statistici di stato di attività e di struttura (addetti e attività economica).

Nella Tavola 4.1 è riportato il numero di indirizzi presenti nei diversi archivi a disposizione e relativi alle imprese attive in Asia 2011.

Tavola 4.1 – Numero di indirizzi negli archivi amministrativi

ARCHIVI	Numero indirizzi
Anagrafe tributaria (Iva)	4 milioni
Dichiarazione nominativa assicurati (Inail)	2 milioni
Camere di commercio (Cciaa)	4 milioni
Utenze telefoniche (Seat)	2 milioni
Modello unico di dichiarazione ambientale (Mud)	400 mila
Archivio UniEmens	1,7 milioni
Archivio degli istituti di credito (Banca Italia)	30 mila
Archivio della Grande Distribuzione (Nielsen)	20 mila
Indagine Iulgi	125 mila
Totale indirizzi elaborati	16 milioni
Indirizzi diversi	6 milioni

I 16 milioni circa di indirizzi presenti nei diversi archivi, si riducono a circa 6 milioni di indirizzi differenti, generati da circa 4,4 milioni di imprese attive. Ciascuna impresa è classificata in unilocalizzata o plurilocalizzata a seconda della presenza di uno o più indirizzi negli archivi amministrativi. Per le imprese unilocalizzate, avendo a disposizione un solo indirizzo, non sono necessarie ulteriori analisi e si considerano quindi coincidenti i caratteri dell'unità locale e dell'impresa, già definiti in Asia-imprese.

In fase di costruzione della base informativa per il registro delle unità locali è stata effettuata una scelta deterministica, al fine di ridurre il numero delle imprese plurilocalizzate (e delle relative unità locali) soggette a successiva stima probabilistica: le imprese con un solo addetto, anche se plurilocalizzate secondo gli archivi di base, sono considerate unilocalizzate, scegliendo come unità attiva quella con indirizzo coincidente con la sede di impresa. Per le restanti imprese plurilocalizzate è possibile suddividere quelle per le quali la costruzione del registro delle unità

locali è effettuato attraverso modelli probabilistici di stima e quelle per le quali le informazioni di stato di attività e caratteri strutturali delle unità locali provengono da indagini statistiche (Iulgi).

4.1.2 L'identificazione delle unità e la stima degli addetti

I caratteri delle unità locali delle imprese plurilocalizzate non appartenenti al campo di osservazione Iulgi sono determinati attraverso l'utilizzo di modelli che prevedono:

- la stima dello stato di attività delle unità locali delle imprese che nelle fonti amministrative presentano almeno un indirizzo;
- la stima degli addetti per le unità locali attive;
- la determinazione del codice Ateco.

A partire dall'elenco degli indirizzi associati ad un'impresa sulla base delle informazioni amministrative, le unità locali dell'impresa vengono individuate scegliendo, tra gli indirizzi a disposizione, quelli più probabilmente corrispondenti ad unità effettivamente attive secondo la definizione di unità locale. Pertanto, il problema della determinazione dello stato di attività dell'unità locale può essere ricondotto a un classico problema di classificazione di unità statistiche, dove l'assegnazione delle unità ad una delle due classi (attiva/non attiva) è effettuata analizzando in che misura il comportamento di una o più variabili esplicative riesce a spiegare l'appartenenza di ciascuna unità ad una delle due classi.

Il modello utilizzato per la stima dello stato di attività di un'unità locale al tempo t è un modello lineare generalizzato ad effetti misti (GLMM)¹¹¹ che permette di tener conto delle informazioni longitudinali disponibili sulla stessa unità locale e, quindi di modellare la correlazione positiva esistente tra lo stato di attività di una stessa unità locale in anni diversi. Le informazioni amministrative, cioè la presenza/assenza dell'indirizzo nelle diverse fonti disponibili nel tempo, sono utilizzate come covariate.

Le variabili esplicative prese in considerazione per gli anni 2004-2011 sono:

- presenza dell'indirizzo nell'archivio dell'Anagrafe tributaria (At);
- presenza dell'indirizzo nell'Archivio delle utenze telefoniche (SEAT);
- presenza dell'indirizzo nell'archivio del Registro delle imprese (Cciaa);
- presenza dell'indirizzo nell'archivio delle Camere di commercio relativo al Modello unico di dichiarazione ambientale (Mud);
- presenza di lavoratori nell'archivio UniEMens (questa informazione non è riferibile a un singolo indirizzo, ma è un indicatore di presenza di unità locali a livello comunale);
- presenza dell'indirizzo nell'Archivio degli istituti di credito (Banca Italia);
- presenza dell'indirizzo nell'Archivio della Grande Distribuzione (Nielsen).

Le variabili di presenza assumo le seguenti modalità:

- 1 = l'indirizzo relativo all'unità locale è presente nell'archivio nell'anno t ;
- 0 = l'indirizzo relativo all'unità locale non è presente nell'archivio nell'anno t , ma esistono altri indirizzi relativi ad altre unità locali della stessa unità giuridico-economica;
- 9 = l'unità giuridico-economica è assente nell'archivio nell'anno.

La probabilità che l'indirizzo sia attivo al tempo t è stimata sulla base della curva di previsione:

$$P(Y_{it} = 1 / \gamma_i) = \frac{\exp(\mathbf{x}_{it}\boldsymbol{\beta} + \gamma_i)}{1 + \exp(\mathbf{x}_{it}\boldsymbol{\beta} + \gamma_i)}$$

¹¹¹ GLMM - Generalized Linear Mixed Model.

dove:

Y_{it} è la risposta per l'indirizzo $i, i=1, \dots, N$, al tempo $t, t=1, \dots, T$;

\mathbf{x}_{it} è il vettore p -dimensionale di covariate;

$\gamma_i, i=1, \dots, N$, è l'intercetta casuale. Si assume che i γ_i sono indipendenti e si distribuiscono secondo una normale con varianza σ^2 .

Una volta determinata la probabilità di attività di ciascun indirizzo $p_i = \widehat{p}(y=1/x)$ occorre:

- 1) definire un valore p_o tale che se $p_i \geq p_o$ allora l'indirizzo è classificato come unità locale attiva, come unità locale non esistente o cessata. Il valore p_o scelto è pari a 0,5.
- 2) per le unità locali definite attive, sulla base delle fonti esistenti, di cui al punto 1, occorre: stimare gli addetti \hat{a} , scegliere il numero di addetti di ciascuna unità locale dell'impresa in modo tale che il loro totale coincida con gli addetti complessivi d'impresa (informazione proveniente da Asia-impresе).

Per la stima degli addetti \hat{a} ¹¹², per le unità locali delle piccole e medie imprese plurilocalizzate per le quali non si dispone di fonti che forniscano direttamente questa informazione, o comunque l'informazione non esiste per tutte le unità di analisi, le indicazioni utili sono:

- il numero di addetti presente nell'archivio UniEMens al tempo t ;
- il numero di addetti presente nell'archivio Mud al tempo t ;
- il numero di addetti presenti nell'archivio Nielsen al tempo t ;
- il numero di addetti presente nell'archivio Inail al tempo t ;
- il numero di addetti rilevato o stimato al tempo $t-1$ (per le unità attive al tempo $t-1$).

Per ciascuna impresa plurilocalizzata, la scelta del valore soglia p_o e la stima degli addetti delle relative unità locali \hat{a} è effettuata risolvendo un problema di programmazione non lineare vincolata, considerando che (i) è nota l'informazione sul totale degli addetti dell'impresa (proveniente da Asia-impresе), (ii) è stata attribuita a ciascuna unità locale una probabilità di attività p_i ed (iii) è stato "scelto" per ciascuna unità locale un numero di addetti tra le fonti su elencate a_i . Si vuole, cioè, minimizzare la differenza tra gli addetti scelti inizialmente (perché provenienti da qualche indagine statistica o fonte amministrativa) e gli addetti finali, considerando le sole unità locali, ordinate per probabilità di attività, con un numero di addetti "necessari" ad ottenere il totale degli addetti di impresa. La differenza tra addetti iniziali e addetti finali è ponderata sulla base dell'attendibilità della fonte iniziale. Il problema della minimizzazione è vincolato alla condizione che la somma dei valori ottenuti sia pari al totale noto degli addetti dell'impresa. In altri termini, si tratta di risolvere il seguente problema:

$$\min \sum_{i=1}^N \lambda_i (\hat{a}_i - a_i)^2 \delta(p_i \geq p_o)$$

¹¹² Naturalmente, la determinazione degli addetti per unità locale è immediata e non ha bisogno di stima per le imprese plurilocalizzate che sono oggetto dell'indagine Iulgi e rappresentano circa il 35 per cento delle imprese plurilocalizzate in termini di addetti.

con il vincolo

$$\sum_{i=1}^N (\hat{a}_i) \delta(p_i \geq p_o) = A$$

$$\hat{a}_i \geq 0 \quad i = 1 \dots N$$

$$0 \leq p_o \leq 1$$

Dove i dati di input sono:

- N pari al numero di unità locali dell'impresa;
- a_i numero di addetti inizialmente scelti per unità locale i -esima da una delle fonti disponibili;
- λ_i peso attribuito alle singole unità locali sulla base dell'attendibilità della fonte utilizzata per gli addetti;
- A è il totale (noto) degli addetti dell'impresa;
- p_i la probabilità *a posteriori* che l'unità locale sia attiva, precedentemente stimata;

e i valori di output sono:

- \hat{a}_i variabile incognita che rappresenta la stima degli addetti per l'unità locale i -esima;
- p_o valore soglia stimato per l'impresa in analisi.

I pesi λ_i attribuiti alle unità locali sono calcolati in funzione dell'attendibilità della fonte utilizzata, la quale è inversamente proporzionale all'errore misurato nel confronto fra i risultati dell'indagine diretta e le risultanze delle diverse fonti. Un peso elevato corrisponde, pertanto, a una buona attendibilità della fonte e quindi determina una minima modificazione del dato scelto, mentre un peso basso indica una scarsa attendibilità della fonte utilizzata e può determinare una consistente modificazione del dato.

4.1.3 I controlli di qualità

L'analisi ed il controllo delle informazioni contenute nel registro Asia-unità locali costituisce l'ultima fase del processo di realizzazione prima del rilascio definitivo dei dati.

Lo sviluppo di un piano di controllo delle informazioni è necessario, poiché le metodologie statistico-probabilistiche che hanno permesso di individuare i caratteri in base alle informazioni provenienti dagli archivi di input, vengono applicate a tutte le unità, anche a quelle per le quali esistono scarse informazioni di base.

La procedura di controllo è stata sviluppata in tre macro-fasi, ognuna riferita ad un differente livello di analisi:

- Fase I - controlli generali. In questa fase i controlli sono stati finalizzati a verificare la correttezza formale delle variabili relative alle singole unità locali delle imprese plurilocalizzate (localizzazione, stato attività, sede d'impresa e addetti) e la loro reciproca compatibilità;
- Fase II - controlli specifici imprese plurilocalizzate. In questa fase sono stati sviluppati alcuni controlli di compatibilità tra i caratteri propri dell'impresa e quelli delle relative unità locali, in modo da controllare la coerenza dell'impresa nel suo complesso. I controlli sono stati orientati, in primo luogo, alla verifica della coerenza tra il numero di addetti risultanti per l'impresa nel suo complesso e il numero risultante dalla somma degli addetti nelle

single unità locali, e, in secondo luogo, all'analisi della compatibilità tra l'attività economica di ciascuna unità locale e quella della relativa impresa. In questa fase sono stati implementati anche dei controlli specifici su determinate tipologie di imprese plurilocalizzate, (imprese per le quali non si dispone di una sufficiente informazione in termini di addetti dalle fonti amministrative, imprese soggette ad eventi di trasformazione strutturale, imprese che svolgono specifiche attività economiche);

- Fase III - controlli macro e validazione dei risultati. Nella fase finale dell'analisi di qualità del registro sono effettuati confronti macro tra i dati del registro all'anno t e quelli degli anni precedenti in relazione alle principali variabili: localizzazione territoriale (provincia e comune), attività economica e numero di addetti.

Le procedure per l'identificazione di valori anomali sviluppate nei controlli macro, si basano su un approccio parametrico e generalizzano il concetto di *outlier* presentato da Davies e Gather (1993). L'idea è quella di definire, sulla base della distribuzione teorica ipotizzata, delle "*outlier regions*", in modo tale che ciascuna osservazione che risulta compresa in tali regioni sia considerata un'osservazione anomala. Questo approccio può essere esteso alla ricerca dei valori anomali nelle tabelle di contingenza (Kuhnt, 2004).

Considerando T variabili categoriali con possibili modalità $\{1, \dots, I_t\}$, $t \in \{1, \dots, T\}$, a ciascuna osservazione di un dato campione corrisponde una particolare combinazione di modalità $i = (i_1, \dots, i_T)$, con $i \in I$ e ciascuna combinazione i definisce una cella di una tabella di contingenza. Le frequenze di celle sono indicate con n_i , $i \in I$, e possono essere rappresentate come un vettore $(n_i)_{i \in I}$. Assumendo il modello loglineare di Poisson $Poi(\lambda)$ come modello distribuzionale teorico per l'analisi delle tabelle di contingenza, la determinazione della *outlier region*, che deriva da una generalizzazione della definizione di Davies e Gather, è la seguente:

$$out(\alpha, \lambda) = \left\{ n \in N : \frac{\lambda^n}{n!} e^{-\lambda} < k(\alpha) \right\},$$

dove N indica l'insieme degli interi non negativi e

$$k(\alpha) = \sup \left\{ k > 0 : \sum_n \frac{\lambda^n}{n!} e^{-\lambda} 1_{[0,k]} \left(\frac{\lambda^n}{n!} e^{-\lambda} \right) \leq \alpha \right\}$$

Sulla base di questa formulazione, una cella n_i della tabella di contingenza è un α_i -*outlier* rispetto al modello loglineare di Poisson con vettore dei parametri λ_t , se giace nella α_i -*outlier region* della distribuzione di Poisson con parametro λ_t , $i \in I$. Il valore α_i dovrebbe essere scelto in modo che la probabilità che si verifichino uno o più valori anomali in una tabella non ecceda un determinato valore α , cioè in modo tale che:

$$P(\exists N_i, i \in I : N_i \in out(\alpha_i, \lambda_i)) = 1 - \prod_{i \in I} P(N_i \notin out(\alpha_i, \lambda_i)) \leq \alpha$$

Usando lo stesso α_i per tutte le celle $i \in I$, si ha:

$$1 - \prod_{i \in I} P(N_i \notin \text{out}(\alpha_i, \lambda_i)) \leq 1 - \prod_{i \in I} (1 - \alpha_i) \leq \alpha$$

e α_i può essere scelto pari a $\alpha_i = \alpha_I = 1 - (1 - \alpha)^{1/I}$, dipendente quindi solo dalla dimensione del campione I e da α .

Sulla base di questa procedura, ciascuna cella sarà classificata come *outlier* o come *inlier*. Se viene assegnato il valore 1 ad ogni *outlier* e il valore 0 ad ogni *inlier*, la procedura di identificazione porta ad associare a ciascun vettore n_i della tabella di contingenza, un vettore della stessa dimensione costituito da zero e uno.

Per l'applicazione della procedura di identificazione è necessario che il vettore dei parametri $\lambda_i \in I$, e il valore α_i siano noti.

Fissato il valore di α , il corrispondente α_i è facilmente determinabile sulla base dell'espressione $\alpha_i = 1 - (1 - \alpha)^{1/I}$, mentre il vettore dei parametri deve essere stimato.

Il classico stimatore utilizzato per modelli di tipo loglineare sopra descritti è quello di massima verosimiglianza (ML). Lo stimatore ML tuttavia, non è uno stimatore robusto, e risulta molto sensibile alla presenza di valori anomali. Una procedura che fornisce stime robuste nell'ambito dell'analisi di tabelle di contingenza a doppia entrata è quella proposta da Mosteller e Parunak (1985), basata sul metodo di "median polish". Questo metodo è un metodo non parametrico che opera iterativamente sulla tabella, calcolando e sottraendo mediane di riga e colonna e termina quando tutte le righe e le colonne hanno mediana pari a zero.

L'ipotesi su cui si basa è che la relazione tra ciascun valore osservato y_{ij} di una tabella di contingenza i due fattori (riga e colonna) è esprimibile attraverso un modello del tipo:

$$y_{ij} = g(i, j) + e_{ij}$$

Qualora si assuma un modello semplice di tipo additivo, il valore y_{ij} è esprimibile come somma di un termine costante comune m , un effetto riga relativo alla modalità i del fattore riga, un effetto colonna b_j relativo alla modalità j del fattore colonna e un termine casuale:

$$y_{ij} = m + a_i + b_j + e_{ij}$$

Applicando la metodologia sopra esposta, per l'individuazione di valori anomali nelle tabelle di contingenza, è stato possibile individuare le combinazioni di variabili (provincia, comune e divisione di attività economica) con valori non congruenti rispetto agli anni precedenti, in genere pari al 15 per cento del totale delle combinazioni analizzate.

Gli incroci di variabili considerati anomali, sono singolarmente esaminati, in modo da accertare l'esistenza di reali mutamenti produttivi e/o territoriali, o invece, la presenza di errori nelle stime occupazionali.

4.2 La stima del luogo di lavoro dei dipendenti presenti nel database di Asia-occupazione

Come ampiamente descritto nel Capitolo 3, il Censimento dell'industria e dei servizi 2011 ha permesso la realizzazione di un registro di tipo LEED (*Linked Employer Employee Database*) in cui è possibile determinare il rapporto diretto tra lavoratori ed impresa e, in particolare, dal lato unità locali, determinare il luogo di lavoro per i tutti i lavoratori dipendenti presenti in Asia-occupazione.

La determinazione del luogo di lavoro è immediata per le imprese unilocalizzate (tutti i lavoratori dell'impresa sono assegnati a un'unica unità), mentre è più complessa per le imprese plurilocalizzate.

Il processo di determinazione del luogo di lavoro avviene attraverso l'utilizzo di informazioni provenienti da più fonti amministrative.

La fonte principale utilizzata è l'archivio UniEMens che registra le dichiarazioni mensili dei datori di lavoro. Ogni record si riferisce a un unico contratto di lavoro e fornisce una serie di informazioni sul tipo di contratto (tempo determinato, indeterminato, eccetera), così come eventi rilevanti che si sono verificati durante l'anno (numero di giorni lavorati per mese, numero di giorni di congedo di maternità, congedo di malattia, eccetera). In particolare, il numero dei giorni di lavoro è utilizzato per assegnare a ciascun contratto di lavoro un peso $q \in [0,1]$ che è definito come la percentuale di giorni lavorati durante l'anno. L'informazione disponibile per l'identificazione del posto di lavoro è il comune in cui lavora il dipendente.

Un'altra fonte amministrativa utilizzata è l'archivio dell'Inail, che registra le dichiarazioni fatte dalle imprese relativamente ai propri dipendenti, in quanto beneficiari di assicurazione contro gli infortuni e le malattie professionali. L'Inail fornisce l'indirizzo (posto di lavoro) in cui ogni lavoratore è assicurato.

Infine, l'Anagrafe tributaria che fornisce le residenze delle persone in possesso di un codice fiscale italiano.

Riassumendo, le variabili amministrative d'interesse per la determinazione del luogo di lavoro sono le seguenti:

- il comune del posto di lavoro (fonte UniEMens);
- l'indirizzo del singolo posto di lavoro (fonte Inail);
- l'indirizzo delle residenze (fonte Anagrafe tributaria).

Un'analisi esplorativa, effettuata confrontando UniEMens con Inail, ha rivelato l'esistenza di molte discordanze per quanto riguarda il luogo di lavoro. In particolare, si riscontrano:

- problemi di copertura: un gran numero di dipendenti del registro Asia non è presente in nessuna delle due fonti. Nell'archivio Inail, in particolare, sono presenti solo circa il 60 per cento dei lavoratori registrati da Asia;
- problemi di coerenza: si registrano discordanze del 40 per cento tra i luoghi di lavoro delle due fonti.

Il processo di allocazione dei lavoratori delle imprese plurilocalizzate a ciascuna unità locale dell'impresa, si basa non solo sulle informazioni provenienti dai suddetti archivi amministrativi, ma utilizza anche i dati del registro Asia-unità locali. In particolare, poiché il registro fornisce il numero medio di lavoratori per ciascuna unità, il processo di allocazione è vincolato ai totali già forniti dal registro.

La metodologia sviluppata è la seguente: considerata un'impresa plurilocalizzata, il registro Asia fornisce l'elenco dei lavoratori di ciascuna impresa, mentre il registro delle unità locali fornisce l'elenco delle unità locali attive. Consideriamo tutte le possibili coppie, lavoratore-unità locale potenziale di appartenenza, si tratta di scegliere quella "giusta" (diremmo i *link* corretti in un contesto di *record linkage*).

Per formalizzare, data un'impresa plurilocalizzata con n dipendenti e k unità locali attive, indichiamo con:

- $U_j, j=1, \dots, k$, le sue unità locali;

- $P_i, i=1, \dots, n$, i suoi dipendenti;
- $q_i, i=1, \dots, n$, i pesi medi annuali associati ai dipendenti.

Per ogni impresa, si costruisce il prodotto cartesiano di $\{P_i\}_{i=1, \dots, n}$ e $\{U_j\}_{j=1, \dots, k}$ e per ciascuna delle $n \times k$ coppie (P_i, U_j) , registriamo i valori dei seguenti tre indicatori:

- $D_1(i, j)$ - indica se il comune del luogo di lavoro del dipendente P_i , registrato in UniEMens, coincide con l'indirizzo del comune dell'unità locale U_j registrato nel registro delle unità locali. In dettaglio, $D_1(i, j) \in \{1, 2, 3\}$ dove $D_1(i, j) = 1$ se il comune UniEMens è mancante, 2 se il comune è diverso, e 3 se il comune è lo stesso;
- $D_2(i, j)$ - misura la coerenza tra l'indirizzo del posto di lavoro del dipendente, come riportato da Inail, e l'indirizzo di unità locale, come individuato dal registro delle unità locali. $D_2(i, j) \in \{1, 2, 3, 4\}$, dove 1 indica che l'informazione Inail è mancante, 2 indica che il comune è differente, 3 indica che il comune è lo stesso, e 4 indica che l'indirizzo dell'unità locale coincide con il posto di lavoro dall'Inail;
- $D_3(i, j)$ - misura la distanza tra la residenza del dipendente geocodificato e l'indirizzo dell'unità locale geocodificato. In questo caso, si classifica la distanza in quattro classi corrispondenti ai valori 1, 2, 3, 4 dalla distanza più lontana a quella più vicina.

La probabilità del dipendente i -esimo P_i di appartenere all'unità locale j -esima U_j , cioè, la probabilità della coppia (P_i, U_j) sia corretta, è stata stimata attraverso un modello a classi latenti. In questo modello, ogni coppia (P_i, U_j) , rappresenta un'unità. Il numero di classi latenti è stato fissato a 2, sotto l'ipotesi che una struttura latente bidimensionale sia alla base del fenomeno analizzato, una classe che indica l' "appartenenza" e l'altra che indica la "non appartenenza". Si poteva scegliere un diverso numero di classi latenti, ma nel caso in esame, il quadro decisionale del problema è in qualche modo diverso da un tipico problema *record linkage*, come descritto nella teoria di Fellegi-Sunter. Infatti, non si deve definire una "zona critica", come non si ha la possibilità di rivedere i cosiddetti "link incerti".

Occorre assegnare ciascun dipendente ad un'unità locale in modo tale che il numero totale medio annuo dei dipendenti, dato dalla somma dei pesi dei lavoratori associati a tale unità locale, coincida con quello riportato nel registro unità locali.

Per questa ragione, una volta che le probabilità di ogni coppia di essere corretta è stata stimata, il problema di allocazione è stato affrontato come un problema di ottimizzazione, in cui la quantità da massimizzare è la somma delle probabilità delle coppie accettate (P_i, U_j) .

Se si indica con:

- L la variabile latente con due classi $c = 1, 2$;
- γ_c la probabilità *a priori* di appartenere alla classe c latente;
- $\rho_{s, r_s | c}$ la probabilità per la variabile $D_s, s=1, 2, 3$ di assumere la modalità r_s condizionatamente a $L=c$,

allora la probabilità *a posteriori* $p_{i, j}(c)$ per la coppia (P_i, U_j) , di appartenere alla classe c condizionatamente ai valori osservati $d(i, j) = (d(i, j)_1, d(i, j)_2, d(i, j)_3)$ delle variabili $D=(D_1, D_2, D_3)$ è:

$$p_{i, j}(c) = P(L = c | D = d(i, j)) = \frac{\left(\prod_{s=1}^3 \prod_{r_s=1}^{R_s} \rho_{(s, r_s | c)}^{I\{d_s(i, j) = r_s\}} \right) \gamma_c}{\sum_{c=1}^2 \left(\prod_{s=1}^3 \prod_{r_s=1}^{R_s} \rho_{(s, r_s | c)}^{I\{d_s(i, j) = r_s\}} \right) \gamma_c}$$

Ogni tripla di possibili valori per gli indicatori D_1, D_2, D_3 è chiamata *response pattern*. Poiché i tre indicatori hanno rispettivamente un numero di modalità pari a 3, 4, e 4, abbiamo 48 *response pattern*. Il numero totale di parametri da stimare è uguale al numero di probabilità di risposta (parametri ρ) per ciascuna classe latente, più il numero di probabilità *a priori* (parametri γ), cioè, nel modello:

$$(2 + 3 + 3) \cdot 2 + 1 = 17$$

Poiché il registro Asia ha milioni di unità locali e milioni di lavoratori, si registrano diversi miliardi di possibili coppie (P_i, U_j) . Per questo motivo, si estrae un campione di imprese e si stima il modello sul sottoinsieme di coppie provenienti da imprese di tale campione. Il modello è stato stimato varie volte su diversi campioni di dimensioni diverse per testare la robustezza delle stime e queste sono risultate essere molto stabili in differenti campioni. È stato usato un algoritmo EM (*Expectation-Maximization*) per trovare le stime ML di massima verosimiglianza dei 17 parametri del modello.

Analizzando le probabilità *a posteriori* ottenute per ogni modello di risposta si può vedere che i valori più elevati corrispondono alla classe latente 1. Cioè, ogni volta che si ha una coerenza tra il posto di lavoro riportato nelle fonti e quello dell'unità locale in Asia (D_1 e D_2), e l'unità locale in Asia è vicina alle residenze dei lavoratori in Inail (D_3), abbiamo una probabilità *a posteriori* superiore per tale classe. Inoltre, le probabilità stimate in molti casi sono molto vicino a 0 o molto vicino a 1. Questo fatto denota una buona "separazione della classe latente". Un approccio comune per quantificare la separazione delle classe latenti si basa su indici di entropia. Uno gli indici più comuni proposto è definito come la media ponderata delle singole probabilità *a posteriori*:

$$E = 1 - \frac{\sum_{i=1}^n \sum_{c=1}^C -p_i(c) \log p_i(c)}{n \log C}$$

dove $p_i(c)$ è la probabilità *a posteriori* dell'unità i nella classe latente c , e C è il numero di classi latenti. E è compreso tra 0 e 1, valori maggiori indicano una migliore separazione tra le due classi latenti. Nel caso del modello descritto, si registra $E = 0,9$, che indica un ottimo grado di separazione tra le due classi.

Inoltre, i valori $\hat{p}_{i,j}(1)$ sono perfettamente coerenti con le informazioni presenti nelle fonti. Infatti, le coppie di individui e unità confermate dall'UniEMens e dall'Inail hanno probabilità *a posteriori* superiore di appartenere alla classe latente 1, mentre la probabilità diminuisce quando le due fonti sono discordanti. Maggiore è la distanza tra l'unità locale e la residenza del lavoratore, più bassa è la probabilità.

L'importanza di tale distanza, cioè di D_3 , è molto inferiore a quella di D_1 e D_2 , coerentemente a quello che ci si aspettava. Queste evidenze hanno permesso di identificare la classe latente 1 come quella caratterizzante l' "appartenenza" del dipendente all'unità locale. Quindi, la stima della probabilità $\hat{p}_{i,j}(1)$ di una coppia (P_i, U_j) , di appartenere alla classe latente 1 è stata interpretata come la probabilità che il lavoratore P_i appartenga all'unità locale U_j , cioè come la probabilità di (P_i, U_j) di essere una coppia corretta.

Formalmente, l'assegnazione di ciascun lavoratore ad una singola unità locale avviene come segue: per ogni impresa, si tende a trovare le

$$x_{i,j} \in \{0,1\}, i = 1, \dots, n, j = 1, \dots, k$$

che massimizzano

$$\sum_{j=1}^k \sum_{i=1}^n p_{i,j} x_{i,j}$$

sotto i vincoli:

$$\begin{aligned} \sum_{i=1}^n x_{i,j} q_i &= n_j, \forall j = 1, \dots, k \\ \sum_{j=1}^k x_{i,j} &= 1, \forall i = 1, \dots, n \end{aligned}$$

dove $x_{i,j} = 1$ se il dipendente P_i è assegnato all'unità locale U_j e n_j rappresenta il numero fisso di dipendenti da assegnare all'unità locale j .

Il problema è stato risolto separatamente per ciascuna delle circa 200 mila imprese plurilocalizzate utilizzando una specifica procedura sviluppata in SAS.

Le imprese più grandi portano ad un prodotto cartesiano di miliardi di possibili coppie (P_i, U_j) , ciò ha richiesto lo sviluppo di una procedura di *blocking*: cioè, l'ottimizzazione è stata ristretta alle unità locali di una singola regione (o provincia o aree più piccole quando necessario) e procedendo in maniera iterativa.

In particolare, indicando con n_{r1}, n_{r2}, \dots il numero totale di dipendenti da assegnare alle unità locali della regione di 1, 2, ... e indicando con k_1, k_2, \dots il corrispondente numero di unità locali per regione, si risolve il problema di ottimizzazione per la prima regione con le seguenti limitazioni:

$$\begin{aligned} \sum_{i=1}^n x_{i,j} q_i &= n_{r1}, \forall j = 1, \dots, k_1 \\ \sum_{j=1}^k x_{i,j} &\leq 1, \forall i = 1, \dots, n \end{aligned}$$

L'algoritmo di assegnazione di ciascun lavoratore dipendente all'unità locale dove svolge principalmente l'attività lavorativa è stato utilizzato in Asia-occupazione 2011 ed ha coinvolto circa 6 milioni di lavoratori dipendenti a circa 600 mila unità locali.

PARTE SECONDA

LE RILEVAZIONI CAMPIONARIE – RILEVAZIONI SULLE UNITÀ
LOCALI DELLE IMPRESE, MULTISCOPO SULLE IMPRESE,
MULTISCOPO SULLE UNITÀ COMPLESSE

Capitolo 1

La Rilevazione sulle unità locali delle imprese¹¹³

1.1 Il quadro normativo e i tempi dell'indagine

L'Istituto nazionale di statistica conduce annualmente, nell'ambito dell'attività di aggiornamento dell'archivio statistico delle imprese attive (Asia), l'Indagine sulle unità locali delle grandi imprese (Iulgi)¹¹⁴ nel rispetto della normativa comunitaria in materia di *Business Register*.

Il Regolamento del Consiglio Europeo n. 177/2008 in materia di registri di impresa utilizzabili ai fini statistici richiede, oltre all'aggiornamento dell'universo delle imprese e dei relativi caratteri statistici, l'aggiornamento delle diverse localizzazioni presso le quali un'impresa esercita una o più attività.

Per l'edizione 2011, come definito dal Piano Generale del 9° Censimento generale dell'industria e dei servizi (Cis2011), la rilevazione Iulgi è rientrata tra quelle adottate per ottenere un quadro informativo statistico sulla struttura delle imprese italiane.

La rilevazione, realizzata nel periodo dicembre 2011 - aprile 2012, ha riguardato quasi 37 mila imprese a fronte delle circa 12 mila normalmente interpellate, alle quali è stato spedito il questionario cartaceo contenente le indicazioni per la compilazione e la restituzione/invio del modello all'Istat, corredato dalle user id e password per l'accesso al questionario online personalizzato.

I soddisfacenti risultati, quantitativi e qualitativi, conseguiti nelle precedenti edizioni hanno consentito di caratterizzare ulteriormente l'indagine Iulgi che in occasione del censimento è stata "ampliata" sia sul piano qualitativo, sia focalizzando l'attenzione sulla corretta attribuzione e codifica delle attività economiche svolte in ciascuna delle singole unità locali, alla luce della classificazione Ateco 2007. Come elemento di novità, rispetto alle passate edizioni, e al fine di avere un quadro sempre più esteso e completo delle imprese italiane, il questionario è stato ampliato con una nuova sezione, (Unità di attività economica), in cui l'impresa ha potuto specificare, oltre alla principale attività economica, altre eventuali attività economiche che godono di una significativa autonomia economica ed organizzativa (indicate anche con termini quali divisione operativa, *business unit*, *business line*, area di business o divisione).

Per l'indagine del 2011 è stata confermata l'organizzazione predisposta per le precedenti edizioni, nelle quali le Sedi territoriali Istat, l'Astat di Bolzano, l'Ufficio di Statistica della Provincia Autonoma di Trento e i revisori della Direzione centrale dei dati amministrativi e dei registri statistici (Dcar) hanno svolto il ruolo di veri e propri Centri di rilevazione, autonomi nell'organizzazione e gestione dell'indagine, nell'ambito degli aspetti generali, organizzativi, metodologici e tecnici concordati.

L'esecuzione dell'indagine svolta "in periferia", per l'area di pertinenza della sede centrale dell'impresa, ha consentito di valorizzare la migliore conoscenza della realtà locale e di instaurare un contatto diretto con i rispondenti, nonché con associazioni di categoria o di rappresentanza o con enti locali, con l'obiettivo di migliorare la qualità intrinseca dell'indagine e fidelizzando i rapporti diretti con le imprese.

¹¹³ A cura di Massimo Piaggese. Autori del capitolo:

- Paragrafi 1.1, 1.2, 1.3, 1.4, 1.10: Simonetta Cozzi;

- Paragrafi 1.5, 1.6, 1.7, 1.8, 1.9: Massimo Piaggese.

¹¹⁴ Informazioni dettagliate sulle caratteristiche organizzative, tecniche, metodologiche ed informatiche della rilevazione sono presenti in "Indagine delle unità locali delle imprese: la flessibilità organizzativa e il ruolo degli uffici regionali come strategia per la riduzione del disturbo statistico e il raggiungimento di elevati tassi di risposta" (a cura di F. Lorenzini), Documenti Istat, n. 1/2008.

Il ruolo attivo che le Sedi territoriali hanno svolto nelle precedenti indagini si è manifestato, tra l'altro, nella proficua collaborazione che è stata realizzata tra i Centri di rilevazione e la struttura organizzativa centrale che ha consentito di interagire nell'ambito di un sistema di lavorazione integrato. Gli scambi di informazioni con i responsabili e i referenti delle Sedi territoriali hanno portato a chiarimenti e implementazioni delle procedure e delle varie fasi di lavorazione in una sorta di *work in progress* che ha costituito uno degli aspetti più qualificanti dell'intero processo produttivo.

1.2 Il campo di osservazione

Il campo di osservazione dell'edizione 2011 dell'indagine, ampliato in termini di numerosità in occasione del 9° Censimento generale dell'industria e dei servizi, era rappresentato dall'universo delle imprese unilocalizzate e plurilocalizzate, che risultavano tali dalla presenza di uno o più indirizzi negli archivi amministrativi utilizzati dall'Istat ai fini della realizzazione dell'archivio statistico delle unità economiche Asia-unità locali.

Rientravano nel campo di osservazione le imprese che svolgevano la loro attività economica principale nei settori dell'industria e dei servizi. Tra queste erano escluse le attività agricole (aziende agricole e forestali), quelle riguardanti l'amministrazione pubblica e la difesa, l'assicurazione sociale obbligatoria (divisione 84 dell'Ateco 2007), le attività di organizzazioni associative (divisione 94), svolte da famiglie e convivenze (divisioni 97-98) e quelle riguardanti le organizzazioni ed organismi extraterritoriali (divisione 99).

Con riferimento alla forma giuridica, erano incluse le imprese individuali e familiari (compresi liberi professionisti), le società di persone, le società di capitali, le società cooperative (ad esclusione delle cooperative sociali), i consorzi di diritto privato, le società consortili, le aziende locali, regionali, provinciali e comunali (ex municipalizzate) e altre forme giuridiche d'impresa. Facendo ancora riferimento alla forma giuridica non rientravano nel campo di osservazione tutte le tipologie relative alle istituzioni pubbliche e non profit (rispettivamente, sezione 2 - Forme disciplinate dal diritto pubblico, e divisioni 1.7 e 1.8 - Enti privati con e senza personalità giuridica della classificazione Istat delle forme giuridiche delle unità legali).

1.3 Unità di rilevazione e unità di analisi

L'unità di rilevazione dell'indagine era costituita dall'impresa, definita come *“unità giuridico-economica organizzata per la produzione di beni e servizi, dotata di autonomia decisionale, in particolare per quanto attiene alla destinazione delle sue risorse correnti e che, in base alle leggi vigenti o a proprie norme statutarie, ha facoltà di distribuire i profitti realizzati ai soggetti proprietari, siano essi pubblici o privati. Un'impresa esercita una o più attività in una o più unità locali”*. La definizione deve essere intesa in senso statistico e si riferisce a qualsiasi soggetto fisico o giuridico che svolge una attività con rilevanza economica; pertanto sono considerate imprese anche i lavoratori autonomi ed i liberi professionisti.

Facevano parte dell'indagine le imprese costituite all'estero ma aventi nel territorio italiano una o più sedi con rappresentanza stabile (la principale delle quali veniva considerata sede centrale dell'impresa estera residente in Italia), mentre non sono rilevate le sedi o le unità locali con rappresentanza stabile operanti all'estero e appartenenti ad imprese costituite in Italia.

L'unità di analisi era costituita dall'unità locale, definita come unità giuridico-economica o parte di essa e situata in una località topograficamente identificata. In tale località, o a partire da tale

località, una o più persone svolgono (lavorando eventualmente a tempo parziale) delle attività economiche per conto di una stessa unità giuridico-economica.

Il questionario è stato inviato alla sede centrale dell'impresa, luogo dove sono ubicati i principali uffici amministrativi e/o direzionali, alla quale competeva la compilazione del modello e l'aggiornamento dell'elenco delle unità locali da essa gestite.

1.4 I criteri di selezione del campione

In considerazione dell'evento censuario, il campione selezionato era costituito dal seguente *panel*:

- 1) imprese con almeno 50 addetti;
- 2) imprese plurilocalizzate con 30-49 addetti;
- 3) imprese plurilocalizzate con 20-29 addetti, identificate nell'Ateco 2007 nella Sezione "G" (Commercio all'ingrosso e al dettaglio e riparazione di autoveicoli e motocicli).

Di conseguenza la numerosità delle imprese interessate alla rilevazione è stata pari a 36.917 unità per un numero complessivo di 193.508 unità locali (Tavole 1.1 e 1.2).

Tavola 1.1 – Numerosità di imprese e unità locali per classe di addetti – valori assoluti

CLASSE DI ADDETTI	Imprese	Unità locali
20 – 29	1.429	3.690
30 – 49	9.900	28.390
50 – 99	14.368	35.770
100 – 249	7.569	29.907
250 e oltre	3.651	95.751
Totale	36.917	193.508

Fonte: Iulgi 2011

Tavola 1.2 – Numerosità di imprese e unità locali per classi di unità locali – valori assoluti

NUMERO UNITÀ LOCALI	Imprese	Unità locali
1	10.789	10.789
2 – 5	21.227	55.492
6 – 10	2.632	19.682
11 – 20	1.328	19.269
21 – 35	455	12.015
36 – 50	156	6.547
oltre 50	330	69.714
Totale	36.917	193.508

Fonte: Iulgi 2011

I questionari delle imprese selezionate sono stati suddivisi e assegnati al Centro di rilevazione competente (Sedi territoriali Istat, Astat Bolzano, Ufficio di Statistica della Provincia Autonoma di Trento e Dcar) in base all'area regionale in cui ricadeva la sede centrale dell'impresa (Tavola 1.3).

Tavola 1.3 – Numerosità delle imprese appartenenti al campo di osservazione di Iulgi 2011 per Centro di rilevazione competente – valori assoluti

	Imprese
Istat - Sede territoriale per il Piemonte e la Valle d'Aosta	900
Istat - Sede territoriale per la Lombardia	1.296
Istat - Sede territoriale per il Veneto	1.000
Istat - Sede territoriale per il Friuli Venezia Giulia	600
Istat - Sede territoriale per la Liguria	250
Istat - Sede territoriale per l'Emilia Romagna	600
Istat - Sede territoriale per la Toscana	600
Istat - Sede territoriale per l'Umbria	525
Istat - Sede territoriale per le Marche	600
Istat - Sede territoriale per l'Abruzzo	600
Istat - Sede territoriale per il Molise	135
Istat - Sede territoriale per la Campania	800
Istat - Sede territoriale per la Puglia	600
Istat - Sede territoriale per la Basilicata	225
Istat - Sede territoriale per la Calabria	200
Istat - Sede territoriale per la Sicilia	400
Istat - Sede territoriale per la Sardegna	600
Astat - Bolzano	391
Ufficio di Statistica della Provincia Autonoma di Trento	381
Totale Sedi territoriali	10.703
Istat - Direzione centrale dei dati amministrativi e dei registri statistici (Dcar)	26.214
Totale generale	36.917

1.5 La comunicazione alle imprese

La rilevazione ha avuto inizio nel mese di dicembre 2011 con l'invio della lettera informativa e del questionario mediante raccomandata A/R.

Tramite questa comunicazione l'impresa prendeva atto di essere stata inclusa nel campione e veniva informata sulle finalità dell'indagine, sui riferimenti normativi in materia di segreto statistico, di obbligo di risposta e di sanzionabilità in caso di mancata risposta. Nella stessa lettera era prestampato altresì il nome della persona registrata come compilatore dell'impresa del questionario Iulgi per l'edizione dell'indagine più recente per cui tale informazione era disponibile, al fine di facilitare un indirizzamento corretto del questionario all'interno dell'impresa stessa.

Nel rispetto della trasparenza della Pubblica Amministrazione, erano inoltre indicati i riferimenti del Centro di rilevazione che avrebbe seguito la lavorazione del questionario (Sedi territoriali Istat, Astat Bolzano, Ufficio di Statistica della Provincia Autonoma di Trento e Dcar) e al quale l'impresa poteva rivolgersi per chiarimenti e assistenza.

Erano inoltre illustrate le diverse modalità di restituzione del questionario: mediante posta ordinaria, fax, collegandosi al sito <http://iulgi.istat.it/UL> (questionario online) o tramite file Excel (modalità offline), nonché le istruzioni e le note definitorie per facilitare la corretta compilazione delle varie sezioni del modello Iulgi.

Si evidenzia inoltre che il questionario, spedito all'impresa, era personalizzato e parzialmente precompilato sulla base di informazioni presenti nell'archivio Asia e, per le imprese che avevano preso parte all'edizione 2010 dell'indagine, sulla base delle informazioni rilevate in quell'occasione.

1.6 Il processo di lavorazione

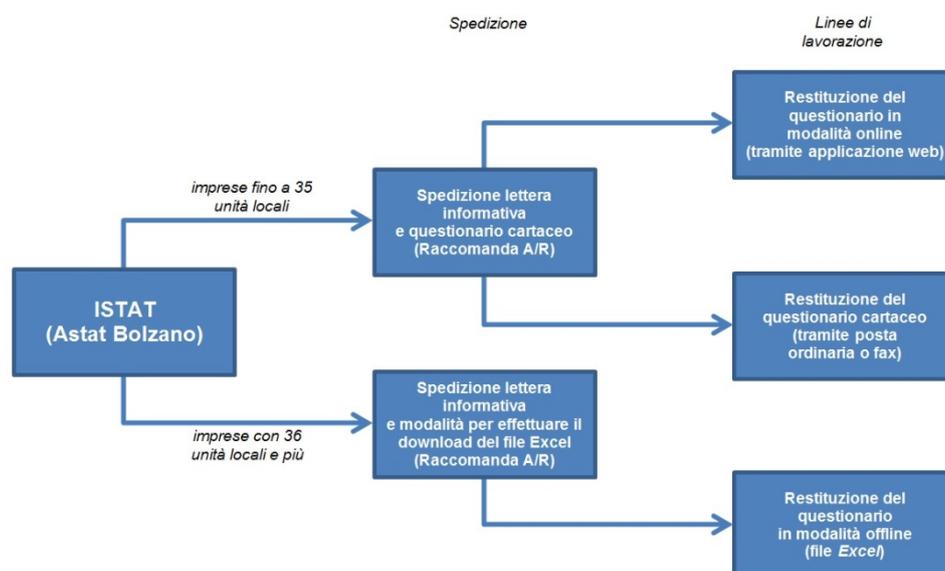
Il processo di lavorazione della rilevazione era costituito da tre linee parallele che prevedevano organizzazione e standard peculiari (Figura 1.1).

La linea principale consisteva nella spedizione del questionario cartaceo, nella sua verifica, correzione e compilazione da parte delle imprese, nella restituzione a ciascun Centro di rilevazione, nella registrazione online a cura di ciascun Centro di rilevazione del questionario cartaceo in *data entry* controllato.

La seconda linea di lavorazione contemplava la possibilità da parte delle imprese di compilare online il questionario direttamente via internet.

La terza linea prevedeva la compilazione da parte delle imprese, con 36 unità locali dipendenti o più, di un file in formato *Excel* che riproduceva il questionario cartaceo (imprese offline).

Figura 1.1 – Spedizione del questionario e linee di lavorazione



1.7 I questionari respinti

Il mancato recapito delle lettere informative e dei questionari cartacei alle imprese (Respinti) è stato gestito centralmente dalla Dcar e dall'Astat di Bolzano, che provvedevano alla registrazione nel sistema di monitoraggio di tali errori; mentre la lavorazione del respinto era cura di ciascun Centro di rilevazione.

Al fine di risolvere i casi di "questionario respinto" si è proceduto all'accertamento di alcune informazioni riguardanti l'impresa ricorrendo alla consultazione di fonti amministrative e statistiche disponibili, alle associazioni di categoria, ai siti internet d'impresa o ad altri elenchi in possesso.

I dati da controllare, causa del mancato recapito del questionario, erano essenzialmente:

- errato indirizzo postale (in questo caso occorreva riuscire ad ottenere l'indirizzo corretto della sede d'impresa);
- modifica ragione sociale (nel caso quella originale fosse imprecisa, o nel caso di fusioni, scorpori, eccetera);
- cambio di *status* (impresa non più attiva).

Una volta verificate ed eventualmente aggiornate tali informazioni, il Centro di rilevazione poteva autonomamente:

- ripristinare il questionario respinto, al fine di una nuova spedizione postale;
- chiudere la posizione dell'impresa e il relativo questionario (casi di inattività, cessazione, procedure concorsuali o imprese fuori campo di osservazione).

1.8 Il sistema informatico

Come nel corso delle precedenti edizioni, il sistema di rilevazione informatico consentiva a ciascun Centro di rilevazione di gestire le varie fasi dell'indagine ed era organizzato essenzialmente in tre macro funzioni:

- Monitoraggio, per accedere al sistema di monitoraggio e ai report riepilogativi riguardanti questionari respinti, contatti con le imprese, questionari arrivati e rispettivo stato di lavorazione;
- *Data Entry*, per procedere all'inserimento dei dati relativi alle imprese provenienti dai questionari cartacei e accedere ai dati relativi ai questionari compilati online dalle imprese;
- Check, per avviare le procedure di check dei dati raccolti.

1.9 Il piano di controllo dei dati

Il piano complessivo di controllo dei dati dell'indagine si divideva in tre diverse fasi:

- prima fase: inserimento dati, effettuato online dalle imprese o dai Centri di rilevazione sulla base dei questionari cartacei ricevuti (*data entry* controllato);
- seconda fase: controlli micro di competenza dei Centri di rilevazione, relativamente alle imprese di pertinenza (check);
- terza fase: controlli macro di competenza della Dcar che riguardavano l'universo delle imprese rilevate per ciò che attiene, ad esempio, al numero di imprese e addetti, per Ateco e/o per localizzazione territoriale, o i confronti con altre informazioni statistiche o amministrative disponibili.

Le regole di prima fase prevedevano solo controlli all'interno di ogni singola sezione del questionario; non prevedevano cioè controlli incrociati tra sezioni. Non era possibile, infatti, implementare controlli incrociati perché il questionario online era organizzato per singoli *frame* di sezione. Inoltre, si era scelto di non appesantire eccessivamente, anche in termini di tempo, la compilazione da parte delle imprese. Poteva accadere, quindi, che le imprese restituissero questionari parzialmente non corretti. Nel *data entry* online alcune delle regole di controllo erano di tipo "bloccante", nel senso che se non venivano inserite alcune informazioni ritenute fondamentali non si poteva procedere all'immissione di ulteriori dati. Altre regole producevano solo un "avvertimento" sulla mancata immissione del dato o sull'incongruenza o non correttezza dei dati inseriti all'interno di una stessa sezione. Era compito dei Centri di rilevazione correggere o completare i questionari in modo da renderli statisticamente significativi.

Le regole di seconda fase (attivati attraverso la funzione check dell'applicazione informatica) comprendevano quelle di prima fase più altre che implementavano controlli più completi e complessi, come ad esempio quelli che incrociavano informazioni tra diverse sezioni e quelli che prevedevano confronti con i dati di partenza. Il check di seconda fase poteva essere attivato dal Centro di rilevazione per singola impresa o contemporaneamente per tutte le imprese per le quali era stata completata la compilazione del questionario.

Gli errori relativi alla seconda fase (check) potevano anch'essi essere di due tipi: "bloccanti" (necessariamente da correggere) o di solo "accertamento" (con possibilità di confermare o modificare il dato).

Le regole di terza fase sono di natura centralizzata e riguardano le principali variabili a livello macro di disaggregazione. Come strumenti di verifica sono state utilizzate tavole di spoglio a doppia o tripla entrata in relazione al numero di imprese e di addetti, per attività economica e/o per localizzazione territoriale. Inoltre, sono stati applicati, allo stesso livello di disaggregazione, intervalli di accettazione per le stesse informazioni rispetto ad altre informazioni statistiche o amministrative disponibili presso altre fonti.

1.10 Il contenuto del questionario

Il questionario Iulgi 2011 si articolava in sei sezioni:

- Sezione 1: Dati strutturali dell'impresa;
- Sezione 2: Notizie dell'impresa;
- Sezione 3: Quadro occupazionale dell'impresa;
- Sezione 4: Elenco delle unità locali gestite dall'impresa da aggiornare;
- Sezione 5: Unità locali nuove o mancanti rispetto all'elenco della Sezione 4;
- Sezione 6: Unità di attività economica.

Le prime tre sezioni riguardavano la verifica dei dati strutturali dell'impresa (ragione sociale, codice fiscale, telefono, forma giuridica, attività economica principale) prestampati sul questionario, sulla base dei caratteri contenuti nell'archivio Asia, e la richiesta di alcune notizie dell'impresa nel suo complesso, tra le quali il quadro occupazionale con riferimento alla media 2011 e al dato puntuale del 31 dicembre 2011.

Le Sezioni 4 e 5 erano relative alle unità locali gestite dall'impresa, operanti sul territorio nazionale. In particolare, la Sezione 4 era volta all'aggiornamento delle notizie riferite alle unità locali già presenti nella lista di partenza (localizzazione, attività economica) da completare con l'informazione del numero di addetti a dicembre 2011. La Sezione 5 permetteva di integrare l'elenco delle unità locali già in possesso dell'Istat con quelle nuove o mancanti.

La nuova Sezione 6, da compilare solo da parte delle imprese con unità di attività economica significative oltre la principale, richiedeva per ciascuna di queste attività: la quota di dipendenti (media 2011) e di fatturato (anno 2011), la localizzazione e altri quesiti qualitativi relativi al valore della produzione, costi della produzione, costo del lavoro, valore delle immobilizzazioni. Inoltre per le attività ausiliarie di supporto ai processi di produzione di beni e servizi era richiesta esclusivamente la quota dei dipendenti. Nell'ultimo punto della Sezione 6 doveva essere indicata l'ubicazione regionale delle attività economiche.

Capitolo 2

La Rilevazione multiscopo sulle imprese¹¹⁵

2.1 Gli obiettivi conoscitivi e la metodologia

Il 9° Censimento generale dell'industria e dei servizi si è svolto in una particolare fase di evoluzione del sistema produttivo italiano. La lunga fase di crisi iniziata nel 2008 ha accentuato l'importanza di elementi che già nel recente passato la globalizzazione aveva fatto emergere tra i fattori chiave di competitività aziendale, come gli investimenti in conoscenza, la flessibilità produttiva e organizzativa, il rafforzamento della *governance*, il presidio dei mercati e la capacità di ridisegnare le catene del valore, in Italia e all'estero. Oggi questi aspetti appaiono strumenti indispensabili per l'avvio di un nuovo percorso di crescita e di salvaguardia della propria forza concorrenziale.

Negli ultimi anni il dibattito su tali tematiche ha ripreso vivacità: i recenti rapporti redatti da istituzioni internazionali, associazioni di categoria, network di ricerca rilevano come, dietro una apparente "staticità strutturale" delle imprese, in termini di dimensione media e settori prevalenti, il sistema industriale nazionale nasconde mutamenti di grande rilevanza, che vanno verso una maggiore complessità e articolazione dei modelli di business. Le soluzioni e i percorsi evolutivi intrapresi dalle imprese presentano un'attenzione particolare alla flessibilità produttiva, all'innovazione, al capitale umano e al controllo della tecnologia, che a sua volta richiede una conoscenza approfondita dell'intera filiera produttiva. Tali caratteristiche, peraltro, si compongono diversamente a formare nuovi modelli di attività, e allo stesso tempo accomunano spesso tutte le classi dimensionali d'impresa.

Mutamenti di questo tipo sono ampiamente diffusi nel settore produttivo italiano, ma l'informazione statistica ufficiale manca ancora di completezza. Il 9° Censimento generale dell'industria e dei servizi aveva quindi un'esigenza e un'opportunità in più rispetto ai censimenti passati, ovvero essere in grado di cogliere la trama dei processi decisionali da cui dipendono le strategie aziendali, collocando le scelte su produzione, innovazione, capitale umano e sulle altre variabili competitive nel complesso delle relazioni di collaborazione o competizione che l'impresa intrattiene con il contesto che la circonda.

Mediante la rilevazione dei suddetti fenomeni ci si proponeva di sviluppare il quadro informativo sul sistema delle imprese integrando le nuove informazioni con quelle di carattere economico e strutturale raccolte a cadenza regolare dall'Istituto. Ci si proponeva, inoltre, di estendere la disponibilità di dati relativi alle tecnologie ICT e all'innovazione all'insieme delle imprese minori di dieci addetti.

¹¹⁵ A cura di Mauro Masselli. Autori del capitolo:

- Paragrafo 2.1: Mauro Masselli;
- Paragrafi 2.2.1 e 2.2.2: Stefano Costa e Valeria Mastrostefano;
- Paragrafo 2.2.3: Flavio Verrecchia;
- Paragrafo 2.3.1: Salvatore Filiberti;
- Paragrafo 2.3.2: Stefano Falorsi;
- Paragrafo 2.3.3: Paolo Righi;
- Paragrafo 2.4.1: Giovanni Seri e Roberta Varriale;
- Paragrafo 2.5.1: Ugo Guarnera e Roberta Varriale;
- Paragrafo 2.5.2: Maria Teresa Buglielli e Francesca Romana Pogelli;
- Paragrafo 2.6: Mauro Masselli.

Per ridurre il disturbo statistico presso le imprese e, allo stesso tempo, per motivi di economicità si è scelto di adottare, per le imprese sotto una determinata soglia, una rilevazione su base campionaria. D'altro canto poiché la rilevazione era finalizzata alla raccolta di caratteristiche che presuppongono una complessità organizzativa e gestionale che in larga misura non è presente nelle microimprese e poiché l'elevato numero di tali unità nel sistema economico del nostro Paese avrebbe reso inefficiente il campione, si è deciso di fissare anche una soglia minima di inclusione nella popolazione di riferimento. Dopo alcune prove di sensitività tale popolazione è stata determinata come l'insieme delle imprese con una dimensione inferiore a venti addetti e superiore o uguale a tre. Le stime campionarie ottenute dai risultati della rilevazione sono quindi comparabili con quelli relativi alle altre rilevazioni della tornata censuaria per le imprese con almeno tre addetti. Tuttavia, poiché il taglio dimensionale in termini di addetti può considerarsi solo un indicatore *proxy* di esistenza di complessità organizzativa e gestionale, si è deciso di studiare anche quelle microimprese che a dispetto del numero di addetti possono essere equiparate a imprese di dimensione superiore. A tal fine, sulla base di differenti indicatori di complessità organizzativa e funzionale, elaborati da fonti statistiche ed amministrative, è stata determinata una sottopopolazione di microimprese da cui è stato selezionato un campione; la definizione di tale sottopopolazione è stata effettuata con criteri sperimentali e pertanto le stime ottenute sono da considerarsi materiale di studio e necessarie di ulteriori approfondimenti. Inoltre, dato il metodo utilizzato, i risultati non sono comparabili con i consueti raggruppamenti settoriali e territoriali utilizzati nelle analisi censuarie e nelle indagini correnti sulle imprese. Malgrado gli svantaggi in termini di comparabilità e di possibili errori di classificazione ("falsi positivi e negativi"), tale approccio innovativo, ovvero definire una sottopopolazione di interesse mirata specificatamente agli obiettivi conoscitivi della Rilevazione sulle imprese e non mediante i consueti parametri (classificazione di attività economica, addetti, fatturato), in alcuni casi potrebbe costituire un metodo vantaggioso in termini di efficienza della rilevazione e di riduzione del disturbo statistico. Dopo una prima fase di controllo - effettuata durante la rilevazione - a livello locale dagli Uffici provinciali di censimento istituiti presso le Camere di commercio, industria, artigianato e agricoltura (Cciaa), alcune Unioni Regionali delle Cciaa, l'Istituto Provinciale di Statistica (Astat) della Provincia Autonoma di Bolzano e presso il Servizio di Statistica della Provincia Autonoma di Trento si è provveduto presso l'Istituto alla messa a punto definitiva dei risultati dell'indagine, mediante l'identificazione e la correzione delle mancate risposte totali e parziali e l'elaborazione dei pesi finali per il calcolo delle stime. Per quanto concerne il trattamento delle mancate risposte totali, è stata adottata una duplice strategia. Per le imprese appartenenti alla parte censuaria le informazioni relative alla singola impresa non rispondente sono state imputate mediante modelli statistici mentre per le imprese della parte campionaria si è provveduto alla correzione dei pesi campionari.

2.2 Il questionario

2.2.1 I contenuti informativi

Gli obiettivi conoscitivi sopra delineati si declinano in modo diverso a seconda della posizione delle imprese nel sistema produttivo italiano, caratterizzato da una presenza massiccia di imprese di dimensioni particolarmente ridotte - almeno sulla base del numero di addetti - un'area diffusa di piccole e medie imprese, molte delle quali estremamente dinamiche e da un numero contenuto di grandi unità interessate da importanti fenomeni di ristrutturazione. Ciò ha implicato la definizione di un questionario "base", in grado di cogliere le caratteristiche d'impresa e i fattori di competitività,

per il quale prevedere varianti dimensionali e settoriali, mirate alla misurazione di alcune specificità di particolare rilevanza per le esigenze conoscitive cui il censimento doveva rispondere.

Ai fini di massimizzare l'informazione sulla complessità dei rapporti produttivi e, al contempo, abbassare l'onere statistico della rilevazione per le imprese, il questionario base si è concentrato su domande qualitative, in una posizione idealmente complementare al contenuto informativo già ricavabile da altre fonti (di natura amministrativa e non), che sono state utilizzate in un secondo momento per l'elaborazione e integrazione dei dati d'indagine, potenziando in misura significativa la fase di analisi e diffusione delle informazioni raccolte.

Il questionario base è stato articolato in sette sezioni, strutturate sui temi di maggiore rilievo per osservatori, *stakeholders* ed esperti della materia:

- 1) Imprenditorialità, controllo e *governance*;
- 2) Capitale umano;
- 3) Relazioni tra le imprese;
- 4) Mercato;
- 5) Innovazione;
- 6) Finanza;
- 7) Internazionalizzazione.

Altri aspetti, a cominciare dalla dotazione e utilizzo di tecnologie informatiche, sono stati affrontati attraverso singoli quesiti collocati nelle varie sezioni, coerentemente con le tematiche principali.

Come si è detto, per cogliere con maggiore dettaglio le caratteristiche del sistema produttivo italiano, la struttura e i contenuti del questionario base sono stati riprodotti in due varianti dimensionali, che pur rispettando l'impianto concettuale generale consentissero alla rilevazione di cogliere le specificità legate alle diverse dimensioni delle imprese. In particolare, sono stati predisposti due distinti questionari: uno per le microimprese (unità con meno di dieci addetti), l'altro per le imprese con dieci addetti e oltre. Come si vedrà meglio tra breve, il primo riportava i quesiti generali in una forma per certi aspetti "semplificata" (ad esempio limitando il numero di elementi rilevati su *governance* e relazioni interaziendali), approfondendo invece gli aspetti più legati all'imprenditorialità (ad esempio genere, età, background dell'imprenditore). Il questionario destinato alle imprese più grandi, a sua volta, ha affrontato questi aspetti in misura più generica, privilegiando piuttosto la ricostruzione di alcuni dettagli di complessità dell'attività aziendale, dalla *governance* (es. grado di accentramento decisionale delle strategie) alla pianificazione (gestione del personale e degli investimenti in formazione, gestione finanziaria).

Infine, alcune sottosezioni del questionario hanno introdotto piccole varianti settoriali – per i comparti manifattura, commercio, finanza – in modo da calibrare le modalità di risposta a seconda del settore economico di appartenenza dell'impresa. È il caso, ad esempio, dei quesiti sulla diversificazione produttiva – somministrati alle sole imprese manifatturiere – o di quelli sulle strategie di distribuzione dei prodotti e servizi, che ovviamente escludevano le imprese commerciali. Di seguito si presentano in dettaglio i contenuti delle sezioni del questionario.

2.2.2 Le sezioni del questionario

Sezione 1: "Imprenditorialità, controllo e *governance*" - La sezione intende cogliere l'identità dell'impresa attraverso un insieme ricco e articolato di informazioni sull'assetto proprietario, sulla distribuzione delle responsabilità gestionali e sui meccanismi che regolano i processi decisionali e strategici. In particolare, i quesiti si concentrano su:

- le principali caratteristiche della struttura proprietaria delle imprese italiane in termini di tipologia e nazionalità del socio principale;
- i meccanismi e le relazioni di controllo e governo societario dell'impresa con l'obiettivo di comprendere come si formino le decisioni strategiche in azienda, sia che queste si originino all'interno di un unico nucleo familiare sia che contemplino l'intervento di figure esterne alla compagine proprietaria. Per le imprese di maggiore dimensione (quelle con almeno dieci addetti), questa sotto-sezione si è arricchita di un quesito volto a valutare il grado di accentrato decisionale per funzione aziendale;
- le finalità delle strategie aziendali (difesa, accesso a nuovi mercati, ampliamento della gamma di prodotti offerti, riequilibrio attività in Italia e all'estero, eccetera);
- le forme di sostegno (in termini di servizi di informazione, consulenza e assistenza) ricevute dall'esterno (Camere di commercio, associazioni di categoria, PA, eccetera).

Per le microimprese (unità con meno di dieci addetti) il quadro informativo si completa con un'ampia sotto-sezione sulla figura dell'imprenditore e sull'imprenditorialità volta a indagare:

- le caratteristiche dell'imprenditore (età, sesso, cittadinanza, formazione, esperienza lavorativa precedente);
- il coinvolgimento (avvenuto o prevedibile) in almeno un passaggio generazionale;
- le conseguenze del passaggio generazionale sulla compagine proprietaria (rafforzamento, mantenimento o indebolimento del ruolo della famiglia proprietaria);
- l'eventuale presenza di ostacoli al passaggio generazionale (burocrazia/fisco, conflitti familiari, difficoltà di trasferire competenze, assenza di eredi qualificati, difficoltà economico-finanziarie).

Sezione 2: "Capitale umano" - In questa nuova fase di transizione economica, la capacità di competere dell'impresa fa decisamente leva anche sulla dotazione e la qualità del capitale umano. In particolare, l'investimento permanente nel capitale umano diventa un importante strumento di differenziazione e un *asset* strategico ai fini della stessa sopravvivenza/evoluzione dell'impresa. La sezione intende valutare il ruolo del capitale umano come risorsa strategica dell'impresa e della conoscenza come fattore chiave di competitività. In particolare, la sezione è articolata in tre sotto-aree tematiche.

Un primo gruppo di quesiti riguarda la dotazione e gli investimenti in capitale umano a elevata qualifica professionale (*high skills*), considerato uno degli elementi determinanti per lo sviluppo competitivo del sistema produttivo italiano. In questo ambito, sono raccolte informazioni su:

- l'incidenza del personale a elevata qualifica professionale (dirigenti, professionisti di elevata specializzazione, tecnici);
- gli investimenti effettuati in risorse *high skilled*;
- gli ostacoli all'acquisizione di nuove risorse *high skilled* (domanda, offerta, logistica, costo del lavoro, budget, burocrazia, eccetera).

Un secondo set di quesiti è dedicato al tema della gestione e valorizzazione del capitale umano. Al centro di questa sotto-sezione si trova il tema della formazione aziendale dei lavoratori; in questo specifico ambito, sono investigati:

- la presenza/assenza del fenomeno;
- la tipologia di attività svolte: corsi o altro;

- le finalità delle attività di formazione (formazione d'ingresso o formazione continua nel corso del rapporto di lavoro in direzione sia dell'aggiornamento sulle mansioni svolte ordinariamente che della riqualificazione della professionalità posseduta);
- gli ostacoli incontrati (costi elevati, insostituibilità dei lavoratori, assenza di offerta formativa, eccetera).

Inoltre, per le sole imprese con almeno dieci addetti, si è richiesto un dettaglio informativo sulle materie oggetto dei corsi di formazione per qualifica professionale.

Altre informazioni contenute in questa sotto-sezione riguardano le competenze tecnico-professionali (o specialistiche) ritenute importanti per la competitività (progettazione, R&S, informatica, pianificazione, finanza, marketing, eccetera). Infine, per le imprese con almeno dieci addetti sono state raccolte altre informazioni su:

- competenze trasversali rilevanti per lo sviluppo, ossia quell'insieme di conoscenze/capacità/abilità individuali (cognitive e relazionali) di carattere generale che possono essere applicate in ambiti lavorativi differenti (*problem solving*, adattamento, comunicazione, creatività, eccetera);
- modalità di gestione del personale (sistemi incentivanti individuali e collettivi, rotazione, autonomia decisionale, coinvolgimento dei dipendenti, eccetera);
- gli strumenti tecnologici, le reti e le applicazioni web utilizzate per facilitare la comunicazione e lo scambio e condivisione di informazioni all'interno dell'impresa.

Sezione 3: "Relazioni tra imprese" - La sezione colloca l'impresa all'interno delle relazioni di collaborazione interaziendale – interne o esterne alla filiera produttiva, nazionali o internazionali – che possono condizionarne le strategie competitive e di crescita.

I quesiti riguardano sia le caratteristiche dei rapporti attivati (tipologia, estensione, durata, eccetera) sia la posizione dell'impresa nell'ambito di tali relazioni (paritetica o all'interno di una gerarchia decisionale), sia infine le funzioni aziendali interessate dall'attivazione delle collaborazioni. Più in dettaglio, i quesiti riguardano i seguenti aspetti:

- tipologia delle relazioni: commessa, subfornitura, consorzio, contratto di rete, altri accordi formali (*joint venture*, Ati, eccetera), franchising, accordi informali;
- ampiezza delle relazioni (per tipologia di relazione): numero, tipologia (imprese, università, PA) e localizzazione delle controparti (area locale, nazionale, internazionale);
- intensità delle relazioni (per tipologia di relazione): funzione aziendale interessata; "peso" delle relazioni con il committente (subfornitore) principale in termini di percentuale sul totale di vendite (acquisti); durata delle relazioni, in termini di anni; ruolo dell'impresa nelle relazioni (leader o *follower*), per funzione aziendale;
- motivi: riduzione dei costi, sviluppo di prodotti/processi, accesso a nuovi mercati, accesso a nuove conoscenze/tecnologie, internazionalizzazione, maggiore flessibilità produttiva, eccetera;
- ostacoli all'attivazione di relazioni: localizzazione sfavorevole dell'impresa, dimensioni ridotte, burocrazia, timori di perdere autonomia, scarsità di controparti, eccetera.

Sezione 4: "Mercato" - La sezione intende valutare la posizione concorrenziale dell'impresa nei mercati di riferimento (locale, nazionale o internazionale) attraverso quesiti che si riferiscono sia alle caratteristiche del mercato stesso sia alla propria capacità competitiva.

I rapporti di competizione dell'impresa sono ricostruiti attraverso informazioni sulle principali strategie e variabili concorrenziali adottate, sui concorrenti, sui punti di forza competitiva e sulla

presenza di eventuali ostacoli ad un miglioramento della capacità concorrenziale dell'impresa. In particolare, vengono rilevate le strategie di posizionamento sul mercato in termini di azioni intraprese con riguardo all'offerta produttiva, alla gestione del marchio, alle scelte distributive e alla definizione dei prezzi. Particolare riguardo è posto anche al tema del commercio elettronico.

Nello specifico i quesiti riguardano i seguenti aspetti:

- le caratteristiche del mercato in termini di area di operatività (locale, nazionale, internazionale), di tipologia dei clienti principali (famiglie, imprese, PA), del settore di appartenenza) e dei concorrenti (localizzazione);
- la formazione del prezzo (*mark up*, *price taking*, prezzi amministrati, eccetera);
- la capacità competitiva, espressa dal posizionamento relativo rispetto ai concorrenti, dai punti di forza (prezzo, qualità, flessibilità, innovazione, eccetera), dagli ostacoli alla competitività (finanza, contesto socio-ambientale, problemi di fornitura, burocrazia, infrastrutture, domanda, eccetera).

La sezione si conclude con due sotto-aree dedicate a due specifiche tipologie di imprese. La prima interessa le imprese non commerciali e raccoglie informazioni sulla copertura distributiva (intensiva, selettiva o esclusiva) e i canali distributivi (reti indipendenti o di proprietà, agenti, fiere, eccetera). La seconda è rivolta alle sole imprese manifatturiere e contiene quesiti sulla diversificazione produttiva; in particolare, sulle modalità (tecnologia o domanda) e i vantaggi (espansione, riduzione del rischio, eccetera).

Sezione 5: "Innovazione" - La sezione si propone di cogliere la complessità dei meccanismi innovativi adottati dalle imprese. In particolare, si cerca di fornire elementi informativi su fenomeni poco esplorati, quali: l' "innovazione sommersa" (cioè l'innovazione svolta mediante attività solo parzialmente strutturate); la complementarità delle diverse tipologie di innovazione (l'innovazione non investe più separatamente i processi e i prodotti, ma è realizzata in forma integrata secondo un processo circolare di continui *feedback*); la rilevanza della dimensione relazionale nei processi innovativi (l'innovazione è, infatti, sempre più spesso il risultato di un processo di tipo collaborativo, che interessa una pluralità di soggetti anche dispersi geograficamente).

I quesiti della sezione seguono il seguente schema:

- la tipologia di innovazione introdotta (di prodotto, di processo, organizzativa, di marketing);
- le attività svolte a sostegno delle innovazioni (R&S; acquisizione di licenze o brevetti; marketing, eccetera);
- gli ostacoli all'attività innovativa (finanza, scarse informazioni su mercati o tecnologie, scarsità di personale qualificato, eccetera);
- le azioni a tutela della originalità dei prodotti (brevetti e altre strumenti di tutela della proprietà intellettuale/industriale);
- per le sole imprese con almeno dieci addetti è stato incluso un quesito diretto a valutare la rilevanza delle tecnologie ICT per tipologia di innovazione introdotta, mentre per le sole microimprese (non interessate alla rilevazione statistica ufficiale sulle tecnologie dell'informazione e comunicazione - ICT) è stata inserita una sotto-sezione sull'utilizzo delle tecnologie ICT (tipologia di connessione e di utilizzo di internet, commercio elettronico, eccetera).

Sezione 6: "Finanza" - La sezione rileva le diverse fonti di finanziamento all'attività d'impresa: interne, quali il reinvestimento di profitti, ed esterne, dal credito bancario al ricorso al capitale di rischio. A seguire si approfondisce la natura del rapporto banca-impresa, per meglio cogliere le

modalità con le quali, in un sistema banco-centrico quale quello italiano, i rapporti creditizi condizionano l'impresa nell'approvvigionamento di risorse finanziarie e nello svolgimento della propria attività. È noto, d'altra parte, che sia il multiaffidamento (frequente soprattutto presso le imprese maggiori) sia i rapporti stretti ed esclusivi con una banca principale (particolarità delle imprese minori) da un lato contribuiscono ad attenuare l'asimmetria informativa caratteristica dei mercati bancari, agevolando l'accesso a risorse esterne, ma dall'altro presentano rigidità potenzialmente vincolanti per molte attività aziendali. Gli ultimi quesiti della sezione si concentrano invece sulle modalità di gestione finanziaria dell'impresa, per coglierne la maggiore o minore complessità. I quesiti riguardano dunque i seguenti aspetti:

- importanza delle diverse fonti di finanziamento dell'impresa: autofinanziamento, credito bancario a breve e a medio-lungo termine, credito commerciale, *equity*, eccetera;
- informazioni sui rapporti con le banche, in termini di estensione (numero di banche da cui si ottengono fidi), intensità (presenza di una banca principale, rilevanza e durata del rapporto creditizio), condizioni (fattori di scelta della banca, elementi richiesti dalla banca per valutare il merito di credito);
- durata e importanza dei debiti e crediti commerciali vantati presso diverse tipologie di soggetti (imprese, famiglie, amministrazioni pubbliche, estero);
- informazioni sull'utilizzo di agevolazioni e finanziamenti pubblici (fondi statali e/o regionali, contributi a fondo perduto, finanziamenti a tasso agevolato, ricorso al Fondo centrale di garanzia, altro) e loro finalità (avvio attività, ricerca, formazione, investimenti, eccetera);
- informazioni sulla gestione finanziaria dell'impresa: strutture interne, banca principale, altri intermediari (per le sole unità con almeno dieci addetti).

Sezione 7: "Internazionalizzazione produttiva" - Obiettivo della sezione è rilevare l'eventuale partecipazione dell'impresa a filiere internazionali, tramite relazioni di mercato o accordi. Nell'evitare sovrapposizioni con altre indagini dell'Istat, quali *Outward FATS* e *Inward FATS*, i quesiti della sezione hanno riguardato i seguenti aspetti:

- tipologia dell'eventuale delocalizzazione (Ide o accordi/contratti);
- destinazione geografica degli impianti delocalizzati e della produzione in essi realizzata (mercati locali, esportazioni in paesi terzi, riesportazione in Italia), per tipologia della delocalizzazione;
- assistenza eventualmente ricevuta all'internazionalizzazione produttiva: da chi (enti pubblici italiani, enti pubblici esteri, organizzazioni private italiane, organizzazioni private estere) e per quale area geografica;
- motivazioni della decisione di delocalizzare, per area in cui si è delocalizzato (riduzione dei costi, accesso a nuovi mercati, necessità di seguire clienti e/o concorrenti, ragioni fiscali, eccetera);
- ostacoli che hanno impedito o limitato la delocalizzazione: barriere legali/amministrative, barriere linguistiche/culturali, valori sociali d'impresa contrari a questa strategia, incertezza degli standard internazionali, eccetera.

2.2.3 La struttura e le caratteristiche del questionario

Come già descritto, la progettazione del questionario ha comportato, oltre alla definizione di un set di quesiti "base", validi per tutte le imprese, la necessità di misurare caratteri specifici attraverso la diversificazione di gruppi di quesiti e specifiche sottosezioni in relazione alla dimensione e all'attività dell'impresa. In termini operativi, si trattava di integrare diverse esigenze informative in

funzione di sottoinsiemi di unità di rilevazione in modo tale da determinare un impatto positivo in termini di:

- organizzazione: risparmi di costi e tempi di realizzazione grazie alla semplificazione dell'implementazione del questionario cartaceo, del questionario elettronico, del piano di controllo dei dati, del materiale informativo e delle istruzioni per la rilevazione, del piano di spoglio. Risparmi anche per la stampa dei modelli cartacei con un numero ridotto di pagine (pari a sedici);
- formazione: semplificazione della formazione a cascata della rete di rilevatori;
- rilevazione: semplificazione della rilevazione sul campo;
- informatica: contenimento, ad un migliaio, delle variabili da archiviare nella base di dati primaria;
- qualità e analisi del dato: contenimento di errori non campionari connessi al *data entry* e/o codifica e/o tabulazione.

A tal fine sono stati definiti due modelli base in relazione alla dimensione dell'impresa ulteriormente diversificati al loro interno in funzione dell'attività. I due questionari sono stati diversificati mediante:

- differenti sigle identificative in copertina (modello Istat CIS2011_QS per le imprese con meno di dieci addetti; modello Istat CIS2011_QX_Bis per le imprese con almeno dieci addetti);
- adozione di diversi colori per le copertine dei due diversi questionari di rilevazione per renderli immediatamente distinguibili.

La struttura finale dei questionari di rilevazione conta sette sezioni tematiche contraddistinte da un progressivo numerico. Ad esse si aggiungono - in prima pagina - una sezione informativa contenente anche le credenziali di accesso per la compilazione online, in seconda pagina un riquadro preliminare relativo allo stato di attività dell'impresa e, in ultima pagina, oltre alle notizie sul compilatore è presente un riquadro riservato al rilevatore per l'attribuzione degli esiti della rilevazione (unità inattiva, cessata, irreperibile, eccetera).

La sequenza dei quesiti è stata resa il più lineare possibile, al fine di rispecchiare i legami logici tra le variabili oggetto di studio. Ogni sezione tematica è contraddistinta da un flusso interno alle sezioni stesse senza rimandi tra sezioni. Il flusso è gestito da quesiti filtro che guidano la compilazione all'interno delle sezioni.

I quesiti, prevalentemente di tipo qualitativo, sono posti in forma diretta in quanto riferiti all'attività d'impresa. Non sono state previste domande aperte. Si sono preferite le risposte pre-codificate che hanno il pregio di essere più rapidamente trattabili nelle fasi successive di registrazione e di revisione dei dati. Nella formulazione dei quesiti è stata considerata la necessità di esclusività e esaustività delle modalità (prevedendo la voce "Altro").

La razionalizzazione delle sezioni tematiche dei questionari ha tenuto conto di aspetti relativi all'organizzazione, alla formazione, alla rilevazione, all'informatica, alla qualità e analisi del dato.

L'obiettivo di una maggior armonizzazione tra i questionari è stato colto attraverso un processo di sintesi e razionalizzazione dei quesiti e delle modalità nei questionari e tra questionari. Le modifiche effettuate prima di arrivare alla redazione definitiva sono state numerose, in particolare in ordine alla:

- reciproca esclusività delle modalità di risposta;
- esaustività delle modalità di risposta;
- comprensibilità dei quesiti in relazione al linguaggio ed alla terminologia utilizzata;

- adeguatezza dei quesiti proposti in relazione alla sequenza e alla struttura dei questionari;
- difficoltà di reperimento delle informazioni richieste;
- tempi di compilazione;
- complessità dovuta alla molteplicità dei modelli di rilevazione.

Tale verifica è stata utilizzata per modificare, eliminare e armonizzare o diversificare i quesiti fino alla predisposizione della versione definitiva. I questionari seppur rimasti inalterati nella struttura delle sezioni e nei contenuti principali, risultano più snelli e più *user friendly*.

In particolare, sono state armonizzate le modalità di quesiti identici nei due modelli conservando la diversificazione per i restanti quesiti (ad esempio limitando il numero di elementi rilevati su *governance* e relazioni interaziendali in un modello, approfondendo invece gli aspetti più legati all'imprenditorialità - genere, età, background dell'imprenditore - nell'altro modello).

I due modelli sono stati predisposti con un impianto unitario, passando da una impostazione "*short/long form*" ad un approccio ad "incastro" per i quesiti presenti in base alla dimensione d'impresa. L'armonizzazione tra i due modelli ha quindi riguardato per i quesiti identici (i quesiti base):

- medesime modalità (e quindi medesime istruzioni nella guida alla compilazione e nelle istruzioni per il rilevatore);
- medesima numerazione;
- medesima posizione (quando possibile).

2.3 Il disegno di campionamento

2.3.1 La popolazione di interesse e i domini di studio

A fronte della opportunità di cogliere l'occasione della tornata censuaria per integrare le informazioni sui fenomeni sopra esposti con quelle tradizionalmente raccolte, si è deciso, per non aggravare ulteriormente il carico statistico sulle imprese e per vincoli di carattere economico ed organizzativo, di rilevare tali informazioni non su base censuaria ma solo su un campione di imprese.

Gli obiettivi conoscitivi hanno influito anche sulla definizione della popolazione di riferimento, escludendo l'ipotesi che questa fosse costituita da tutte le imprese presenti nei registri sulle unità economiche Asia e limitando la rilevazione alle unità maggiormente investite dai fenomeni di interesse. Questo approccio, rispetto a quello tradizionale utilizzato per i campioni delle indagini strutturali sulle imprese, permette di concentrare le risorse economiche ed organizzative solo sulla rilevazione delle unità economiche potenzialmente "informative". Si è quindi deciso di non includere nella popolazione obiettivo le microimprese, che rappresentano una rilevante quota del totale delle imprese e, per allineare le stime campionarie agli usuali domini di analisi del censimento, di individuare la sub-popolazione da escludere adottando una "soglia" sulla base della sola dimensione "numero di addetti".

Tuttavia, poiché anche le imprese sotto soglia possono presentare caratteristiche e comportamenti simili a quelle sopra soglia, si è deciso di individuare la sottopopolazione delle microimprese, con una struttura organizzativa "complessa" nella quale i fenomeni di interesse hanno una probabilità elevata di manifestarsi e di selezionare da essa un campione con gli stessi principi metodologici del campione principale. La definizione di tale sottopopolazione è stata effettuata con criteri e indicatori sperimentali, pertanto le stime ottenute, dato il meccanismo di selezione, non possono essere messe in relazione con i consueti domini di studio censuari. I dati del campione delle microimprese sono quindi da considerarsi materiale di analisi e sperimentazione.

Sulla base di tali considerazioni sono state effettuate alcune simulazioni sugli errori di campionamento attesi in funzione di due differenti ipotesi di popolazione che hanno portato a definire la popolazione obiettivo come quella costituita dalle imprese con un numero di addetti pari o superiore al valore soglia di tre e da quelle con meno di tre addetti ma con struttura organizzativa complessa. In tal modo il numero di imprese costituenti la popolazione *target* è stata ridotta a circa 1,625 milioni di imprese.

Per quanto riguarda i domini di studio sono state considerate due tipologie di interesse:

- domini di stima definiti a livello di classi della classificazione Ateco (codici a 4 *digit*) per classe addetti;
- domini di stima definiti a livello di gruppi della classificazione Ateco (codici a 3 *digit*) per classe addetti e per ripartizione geografica.

Tali domini sono detti “domini pianificati”, in quanto l’allocazione del campione è tale che:

- ogni dominio ha una numerosità campionaria stabilita in fase di progettazione del disegno (in assenza di mancate risposte totali);
- gli errori campionari attesi delle stime *target* riferite ai domini di interesse sono inferiori a delle soglie prefissate in fase di progettazione del disegno campionario.

Il disegno di campionamento è progettato per garantire la copertura campionaria a livello provinciale. A tale livello, tuttavia, le dimensioni campionarie non sono state fissate secondo criteri di rispetto di vincoli sugli errori di campionamento attesi.

La lista delle unità riguardanti la popolazione d’interesse, da utilizzare per le operazioni relative al disegno dell’indagine (stratificazione e allocazione del campione, calcolo degli errori attesi di campionamento e selezione delle unità campionarie) e per le operazioni di rilevazione sul campo, è stata ottenuta mediante un aggiornamento dell’archivio Asia 2010, per tenere conto degli eventi di cessazione, nascita e trasformazione delle imprese, intervenuti nel periodo. Lo scopo di tale lista, costruita appositamente per le operazioni censuarie, era quello di approssimare al meglio i dati di censimento non solo per migliorare l’efficienza del campione ma anche per ridurre il fastidio statistico sulle imprese e agevolare le operazioni sul campo.

2.3.2 L’allocazione del campione

Il disegno di campionamento progettato è di tipo stratificato. Gli strati sono ottenuti dalla combinazione delle modalità delle variabili strutturali considerate per la definizione delle due tipologie di dominio di interesse descritte nel paragrafo precedente: classi di attività economica (codici a 4 *digit* della classificazione Ateco), classi di addetti e ripartizioni geografiche. In ambedue le tipologie di dominio, sulla popolazione obiettivo sono state considerate le seguenti classi di addetti: minore del valore soglia (per il campione delle microimprese con struttura organizzativa complessa), soglia 3-9 addetti, 10-19 addetti, 20 e più addetti. Il campione provinciale¹¹⁶ è stato ottenuto allocando proporzionalmente le dimensioni campionarie calcolate a livello di ripartizione geografica in ragione del numero di imprese per provincia.

Gli strati relativi alle imprese con almeno venti addetti sono stati censiti, secondo il criterio correntemente seguito per le indagini strutturali sulle imprese. Tale soglia è stata definita a seguito di un processo di analisi che ha considerato diverse ipotesi alternative basate anche su misure di “ampiezza delle imprese” differenti dal numero di addetti.

¹¹⁶ Indispensabile per la rilevazione di campo svolta a livello provinciale da ciascuna Camera di commercio costituita quale Ufficio provinciale di censimento (Upc).

L'allocazione del campione in ciascuno strato è compiuta con una procedura multivariata e multi dominio che definisce dimensioni campionarie di strato in modo tale che gli errori campionari attesi di una o più variabili (aspetto multivariato) per uno o più domini di interesse (aspetto multi dominio) siano inferiori a delle soglie prefissate. Le variabili utilizzate nell'allocazione sono note per tutte le unità della popolazione obiettivo e si assume che siano correlate con i fenomeni oggetto di studio. Sono state considerate quali variabili guida esplicative dei fenomeni oggetto di studio il rapporto fatturato su addetti medi (f/a) e il rapporto valore aggiunto su fatturato (v/f), assumendo che valori elevati di queste due variabili indichino elevati livelli di competitività dell'impresa. A partire da questi due rapporti sono state costruite otto variabili guida dicotomiche (elencate nel seguito) secondo la seguente procedura:

- per ciascuna classe di attività economica (codici a 4 *digit* della classificazione Ateco) sono stati calcolati i percentili 50°, 75°, 90° e 95° delle distribuzioni dei rapporti (f/a) e (v/f);
- per ciascun percentile è stata costruita una variabile dicotomica {0;1} che assume valore pari a uno se per l'impresa il valore del rapporto considerato è superiore al percentile di riferimento e zero altrimenti.

Per quanto riguarda la definizione delle distribuzioni su cui calcolare i percentili sono state anche valutate le distribuzioni dei rapporti (f/a) e (v/f) a livello nazionale, per macro settori (Industria, Costruzioni, Commercio e Altri Servizi) e per strato.

Assumendo che i due rapporti (f/a) e (v/f) siano delle *proxy* della competitività dell'impresa, le otto variabili sono indicatori di caratteristiche e/o comportamenti legati alla competitività secondo diversi livelli di probabilità di essere osservate.

Nel processo di allocazione, alle otto variabili guida descritte sopra, si sono aggiunte ulteriori tre variabili presenti nell'archivio della popolazione, relative a fatturato, addetti medi e valore aggiunto. In base alla dimensione campionaria totale pari a circa 260 mila unità sono stati calcolati gli errori prefissati attesi delle stime sulle undici variabili guida nelle due ipotesi di popolazione di riferimento.

Sulla base dell'analisi dei coefficienti di variazione delle stime e dell'allocazione del campione nei differenti domini di interesse, è stata quindi individuata la popolazione obiettivo. Nella Tavola 2.1 è illustrata la distribuzione della popolazione e del campione per regione e classe di addetti calcolata sulla base della lista precensuaria¹¹⁷; per gli strati con numero di addetti minore di tre viene riportata la popolazione calcolata secondo la procedura delineata nel successivo Paragrafo 2.4.

¹¹⁷ I dati della Tavola 2.1, poiché derivati da una popolazione obiettivo appositamente individuata per le operazioni di censimento, non coincidono con quelli ottenibili dagli archivi Asia 2010, Asia 2011 o dal Censimento stesso.

Tavola 2.1 - Distribuzione del numero di imprese del campo d'osservazione e del campione per regione e per classe di addetti - valori assoluti

REGIONE	Numero di imprese del campo d'osservazione					Numero di imprese del campione				
	Classe addetti					Classe addetti				
	<3	3-9	10-19	20+	Totale	<3	3-9	10-19	20+	Totale
Piemonte	39.572	62.227	9.549	5.605	116.953	905	6.687	3.211	5.605	16.408
Valle d'Aosta/Vallée d'Aoste	1.251	2.722	353	150	4.476	334	264	71	150	819
Lombardia	131.725	153.180	28.724	18.091	331.720	3.881	27.445	11.002	18.091	60.419
Trentino-Alto Adige	11.862	18.603	3.419	1.982	35.866	1.145	2.175	747	1.982	6.049
- Bolzano/Bozen	5.713	9.958	1.893	1.081	18.645	577	1.240	436	1.081	3.334
- Trento	6.149	8.645	1.526	901	17.221	568	935	311	901	2.715
Veneto	62.622	76.488	14.351	9.054	162.515	1.457	10.402	5.041	9.054	25.954
Friuli-Venezia Giulia	10.689	18.166	3.128	1.813	33.796	705	1.950	843	1.813	5.311
Liguria	15.294	25.345	3.348	1.506	45.493	677	2.367	941	1.506	5.491
Emilia-Romagna	56.566	72.494	12.347	7.549	148.956	1.325	10.582	4.805	7.549	24.261
Toscana	47.805	64.540	10.293	4.677	127.315	1.189	9.544	3.822	4.677	19.232
Umbria	9.268	13.368	2.160	1.148	25.944	732	1.511	486	1.148	3.877
Marche	18.621	25.537	4.453	2.441	51.052	800	3.346	1.271	2.441	7.858
Lazio	67.904	71.174	11.068	5.755	155.901	2.495	11.300	4.222	5.755	23.772
Abruzzo	10.021	18.626	2.643	1.425	32.715	794	2.548	672	1.425	5.439
Molise	1.864	3.810	494	223	6.391	441	503	92	223	1.259
Campania	36.347	58.689	8.659	4.653	108.348	1.188	9.529	2.913	4.653	18.283
Puglia	20.651	48.277	6.596	2.947	78.471	788	6.514	1.773	2.947	12.022
Basilicata	2.706	6.309	857	411	10.283	527	811	172	411	1.921
Calabria	7.693	18.709	2.139	1.048	29.589	681	2.485	480	1.048	4.694
Sicilia	20.614	53.057	6.713	2.969	83.353	783	6.506	1.669	2.969	11.927
Sardegna	9.480	22.195	2.783	1.193	35.651	554	2.810	557	1.193	5.114
Totale	582.555	833.516	134.077	74.640	1.624.788	21.401	119.279	44.790	74.640	260.110

Le Tavole 2.2 e 2.3 riportano le distribuzioni dei coefficienti di variazione (CV) attesi per le stime delle undici variabili guida per i domini di interesse.

Tavola 2.2* - CV attesi per i domini di interesse Ateco 4 digit - valori percentuali

STATISTICHE	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9	CV10	CV11
Minimo	0,3	0,6	0,8	1,0	0,0	0,0	1,5	2,4	0,1	0,0	-33,7
1° quartile	4,1	7,2	11,0	14,8	6,0	10,7	18,0	24,3	4,7	1,0	5,0
Mediana	6,7	11,2	18,1	25,5	7,6	13,3	22,2	30,8	9,1	1,8	8,2
Media	6,5	11,0	18,2	25,9	8,2	13,5	21,9	30,1	10,7	2,2	10,2
3° quartile	8,8	15,5	26,7	38,3	10,2	16,5	26,8	37,4	15,2	3,4	13,2
Massimo	14,1	19,7	30,0	44,5	14,7	24,0	34,4	44,3	29,9	5,0	34,8

Tavola 2.3* - CV attesi per i domini di interesse Ateco 3 digit - valori percentuali

STATISTICHE	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8	CV9	CV10	CV11
Minimo	0,4	1,1	1,4	1,1	0,0	1,7	3,7	3,9	0,0	0,0	-37,7
1° quartile	5,9	10,2	15,7	21,0	8,3	14,7	24,2	32,7	6,5	1,4	6,5
Mediana	9,4	16,3	26,5	37,1	10,9	19,4	32,6	43,6	11,9	2,4	10,8
Media	9,3	15,9	26,7	41,3	11,7	19,2	31,4	47,6	14,3	3,0	14,3
3° quartile	12,7	22,3	39,0	57,2	14,4	24,0	39,7	57,6	19,9	4,8	18,5
Massimo	19,7	24,9	44,9	199,0	24,7	29,7	44,8	209,0	39,9	9,3	45,0

*Legenda variabili:

- CV1 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 50° percentile della distribuzione della variabile Z (fatturato/addetti medi) calcolata a livello di Ateco a 4 digit. X=0 altrimenti;
- CV2 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 75° percentile della distribuzione della variabile Z (fatturato/addetti medi) calcolata a livello di Ateco a 4 digit. X=0 altrimenti;
- CV3 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 90° percentile della distribuzione della variabile Z (fatturato/addetti medi) calcolata a livello di Ateco a 4 digit. X=0 altrimenti;
- CV4 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 95° percentile della distribuzione della variabile Z (fatturato/addetti medi) calcolata a livello di Ateco a 4 digit. X=0 altrimenti;
- CV5 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 50° percentile della distribuzione della variabile Z (valore aggiunto/fatturato) calcolata a livello di Ateco a 3 digit per classe addetti. X=0 altrimenti;
- CV6 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 75° percentile della distribuzione della variabile Z (valore aggiunto/fatturato) calcolata a livello di Ateco a 3 digit per classe addetti. X=0 altrimenti;
- CV7 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 90° percentile della distribuzione della variabile Z (valore aggiunto/fatturato) calcolata a livello di Ateco a 3 digit per classe addetti. X=0 altrimenti;
- CV8 = coefficiente di variazione (%) della variabile $X = \{0;1\}$. X=1 se l'impresa ha un valore superiore al 95° percentile della distribuzione della variabile Z (valore aggiunto/fatturato) calcolata a livello di Ateco a 3 digit per classe addetti. X=0 altrimenti;
- CV9 = variabile fatturato;
- CV10 = variabile addetti medi;
- CV11 = variabile valore aggiunto.

2.3.3 Calcolo dei pesi finali

Negli strati formati dalle imprese con un numero di addetti inferiore a venti, l'allocatione del campione ha determinato il valore dei pesi iniziali da attribuire alle singole unità; tali pesi sono uguali all'inverso delle probabilità di inclusione delle singole imprese campionate, calcolate come rapporto tra la dimensione del campione e la dimensione della popolazione nello strato. Dopo la fase di rilevazione dei dati, i pesi iniziali sono stati corretti per tenere conto delle mancate risposte totali adottando la tecnica delle celle di ponderazione o di omogeneità. In ciascuna cella il correttore rappresenta una stima della probabilità di risposta delle unità campionarie appartenenti a quella particolare sottopopolazione. I pesi diretti delle unità rispondenti nella cella di omogeneità sono stati quindi moltiplicati per il correttore, ottenendo i pesi corretti per mancata risposta totale. Sono state condotte diverse analisi statistiche per individuare in modo opportuno le celle di omogeneità, attraverso l'applicazione di modelli di tipo logit e alberi di regressione. Oltre a quelle utilizzate per definire la stratificazione, sono state considerate anche altre variabili, disponibili dagli archivi amministrativi, al fine di verificare la loro capacità predittiva in relazione alla mancata risposta totale. Dalle analisi condotte, tuttavia, è risultato che l'utilizzo di tali variabili non aumenta la bontà di adattamento dei modelli considerati e quindi per la costruzione delle celle di ponderazione, si è scelto di considerare gli strati definiti dal disegno di campionamento.

Per determinare i pesi finali di riporto all'universo i suddetti pesi sono stati ulteriormente corretti per imporre che le stime campionarie di alcune distribuzioni marginali del numero di imprese secondo le più importanti caratteristiche strutturali (classificazione Ateco, classe di addetti e territorio) coincidano con le corrispondenti distribuzioni conosciute a livello di popolazione. Tale procedura, nota in letteratura come *calibration estimator* è stata applicata nel caso dell'indagine per riprodurre il numero di imprese a livello di (i) classificazione Ateco a 4 digit per le due classi di addetti (3-9

addetti; 10-19 addetti), (ii) classificazione Ateco a 2 *digit* per le due classi di addetti e regione, (iii) quattro macro settori (Industria, Costruzioni, Commercio, Altri Servizi) per le due classi di addetti e provincia. La calibrazione del campione è stata effettuata utilizzando come “totali noti” i dati calcolati dall’archivio Asia 2011, disponibile al momento delle operazione di stima. Nel caso invece degli strati formati dalle imprese con almeno venti addetti, dove la rilevazione è stata di natura censuaria, la mancata risposta totale è stata trattata diversamente, adottando il metodo dell’imputazione delle variabili delle imprese non rispondenti mediante donatore (vedi Paragrafo 2.5).

2.4 Il campione delle imprese di dimensione inferiore a tre addetti

2.4.1 Individuazione della popolazione di riferimento

Nel definire la soglia per l’eleggibilità delle imprese è stata scelta la variabile “numero di addetti” sia per utilizzare una dimensione di analisi consueta e comune con le altre rilevazioni censuarie e con le indagini correnti sulle imprese, sia sulla base dell’ipotesi che le imprese sotto soglia non fossero in gran parte consonanti con gli obiettivi conoscitivi dell’indagine. Tuttavia il “numero di addetti”, in un insieme significativo di microimprese, poteva non rappresentare la variabile discriminante per escluderle dal campo di osservazione. In altri termini si è ipotizzato che, in tali microimprese, la complessità organizzativa ed economica risultasse potenzialmente superiore a quanto attestato dalla loro dimensione, misurata unicamente in termini di numero di addetti. Si è pertanto ritenuto opportuno riservare parte delle risorse per indagare le loro caratteristiche e comportamenti. Per determinare tale sub-popolazione, è stata adottata una procedura, basata su numerosi indicatori calcolati a livello di singola impresa e derivati da molteplici fonti. Tale procedura, inoltre, mirava anche a ridurre il disturbo statistico, evitando di coinvolgere nella rilevazione unità statistiche non interessate dal fenomeno oggetto di studio.

Le microimprese cui risultano associati uno o due addetti nell’archivio Asia (convenzionalmente si assume che il limite corrisponda ad un numero di addetti inferiore strettamente a 2,5 in media annua) sono circa 3,4 milioni. Ai fini di selezionare tra queste le imprese “rilevanti” per l’indagine ovvero quelle imprese che manifestano un certo livello di capacità organizzativa, produttività e dimensione economica, sono state acquisite per le singole imprese le informazioni derivate dagli archivi statistici e di fonte amministrativa disponibili presso l’Istituto.

Oltre agli archivi statistici relativi al registro delle imprese (Asia) e al commercio con l’estero (Coe), dai quali sono state tratte le informazioni relative alla classificazione per attività economica, numero di addetti, fatturato ed esportazioni, sono stati utilizzati i dati dei bilanci civilistici (BIL) per le variabili del conto economico, e degli Studi di settore (Sds - Agenzia delle Entrate) per le variabili del conto economico e per quelle sul lavoro.

In particolare, circa 3,8 milioni di imprese con fatturato compreso tra 30 mila e 7,5 milioni di euro compilano uno Studio di settore e tra le microimprese registrate in Asia, circa il 76 per cento sono presenti nell’archivio Sds, il cui questionario è diversificato in funzione del settore economico (sono utilizzati oltre 200 Sds diversi). Le informazioni relative alle persone che lavorano nell’impresa sono espresse in termini di giornate lavorative retribuite per alcune posizioni (dirigenti, quadri, dipendenti, eccetera) e come numero di persone, insieme alla relativa percentuale di tempo lavorativo impegnato nell’attività dell’impresa, per le altre (soci, collaboratori esterni, eccetera). In particolare gli Sds, che permettono di identificare quantità e forme di lavoro differenti rispetto agli “addetti”, sono stati elaborati per calcolare un indicatore di input di lavoro a livello di singola impresa analogo, concettualmente, a quello usualmente presente nella letteratura statistico-

economica non a livello micro ma come aggregato, determinato ai fini della Contabilità Nazionale per rappresentare la quantità di lavoro effettivamente erogata nell'attività economica.

Per il computo dell'input di lavoro a livello di singola impresa sono stati adottati due diversi criteri basati:

- 1) sulla unità di lavoro "equivalente a tempo pieno" per cui un anno uomo corrisponde a 312 giorni di lavoro retribuito;
- 2) sul numero di persone coinvolte nell'attività dell'impresa a prescindere dal tempo effettivamente impegnato da ciascuno.

Secondo il primo criterio, l'input di lavoro è stato calcolato come somma delle seguenti tre componenti: totale del numero di giorni retribuiti (per tutte le posizioni per cui è prevista tale misura del tempo lavorato) diviso per 312; totale del numero di collaboratori e degli amministratori non soci per i quali, in mancanza di specifiche, si è assunto un impiego a tempo pieno; la somma delle percentuali di tempo impegnato nell'attività dell'impresa per le posizioni lavorative per cui è riportata tale informazione (ad esempio i soci). In accordo con il secondo criterio, l'input di lavoro in termini del numero di persone coinvolte nell'attività d'impresa è stato calcolato come somma di due componenti: il totale del numero di giorni retribuiti (per tutte le posizioni per cui è prevista questa misura del tempo lavorato) diviso per 312; il totale del numero di persone riportate per tutti gli altri tipi di posizione lavorativa.

I criteri utilizzati per includere o meno le imprese nella sottopopolazione di interesse sono riportati di seguito:

- imprese con un fatturato superiore a 7,5 milioni di euro o con valore aggiunto superiore a 200 mila euro;
- imprese che presentano il bilancio civilistico;
- società a responsabilità limitata o cooperative con una dimensione economica minima;
- impresa esportatrice con una dimensione economica minima;
- imprese con "input di lavoro" in termini di unità di lavoro equivalenti a tempo pieno superiore a due con una dimensione economica minima;
- imprese con "input di lavoro" in termini di numero di persone coinvolte nell'attività dell'impresa superiore a due con una dimensione economica minima;
- imprese capo gruppo con una dimensione economica minima;
- imprese con un rapporto tra valore aggiunto e fatturato superiore a uno.

Ciascun criterio ha agito indipendentemente dagli altri, ovvero era sufficiente che una impresa soddisfacesse uno solo dei criteri citati per essere inclusa nella popolazione di interesse.

Ad alcuni dei criteri utilizzati è stata associata una dimensione economica minima. Più precisamente, sono definite di dimensione minima le imprese che presentano contemporaneamente un fatturato superiore a 60 mila euro o un valore aggiunto superiore a 30 mila euro o un rapporto tra valore aggiunto e fatturato superiore a 0,8. Tali limiti sono stati fissati sulla base del terzo quartile delle rispettive distribuzioni sulla popolazione di microimprese per cui il dato era disponibile.

Le variabili utilizzate nella procedura di selezione sono riportate nel Prospetto 2.1, unitamente alle relative fonti, riportate in ordine di priorità di acquisizione.

Prospetto 2.1 - Variabili utilizzate per i criteri di inclusione e relative fonti

VARIABILE	Fonte
Addetti, Ateco, forma giuridica, capo gruppo	Asia
Impresa esportatrice	Coe
Fatturato, valore aggiunto	BIL, Sds, Asia
Giorni retribuiti, numero Lavoratori, % tempo impiegato	Sds

La procedura sopra descritta, funzionale all'obiettivo di selezionare le microimprese rilevanti rispetto agli obiettivi conoscitivi dell'indagine, ha individuato una sottopopolazione di interesse, alla quale è stata attribuito l'etichetta "star", pari a 582.555 unità che rappresentano circa il 17 per cento della popolazione complessiva di microimprese.

Il sottoinsieme delle imprese "non rilevanti" è stato suddiviso in due gruppi di grandezza pressoché simile: (i) il gruppo dei cosiddetti "non star - contribuenti minimi" costituito da circa 1,5 milioni di imprese e (ii) il gruppo delle altre imprese "non star", costituito da circa 1,3 milioni di unità. Nelle Figure 2.1 e 2.2 sono rappresentate tramite *boxplot* le caratteristiche delle distribuzioni di fatturato e valore aggiunto nelle tre sottopopolazioni descritte.

Figura 2.1 - Boxplot delle distribuzioni di fatturato nelle sottopopolazioni di imprese "Non star - contribuenti minimi", "Non star" e "Star" - percentili 1 e 99, quartili 1 e 3, mediana; valori in migliaia di euro

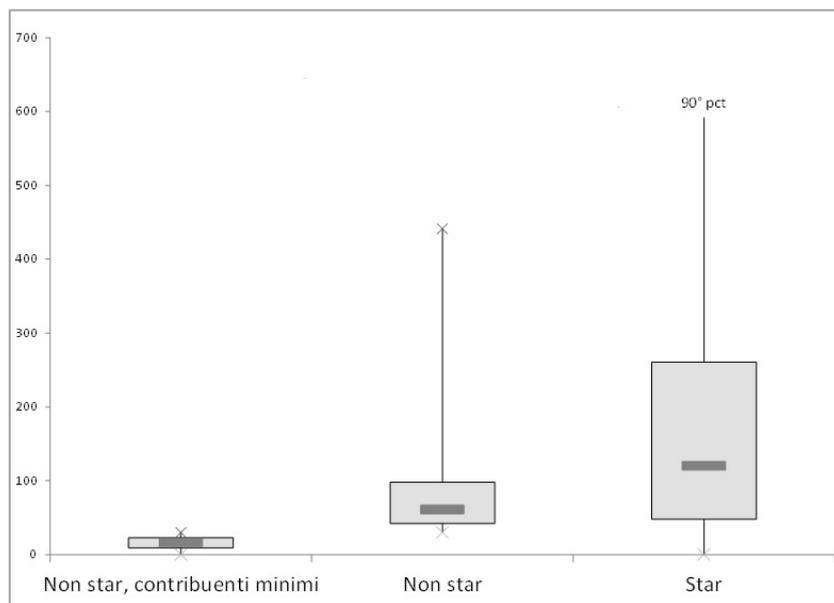
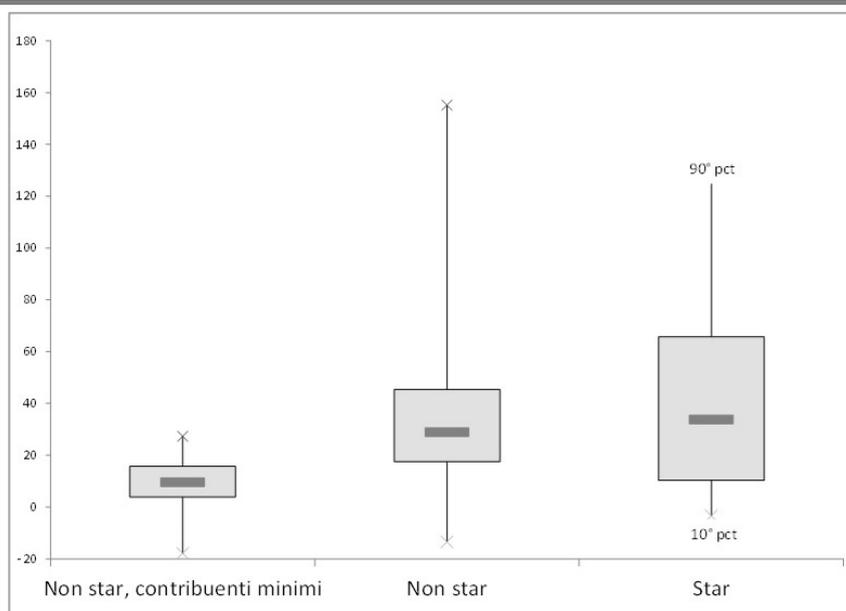


Figura 2.2 - Boxplot delle distribuzioni di valore aggiunto nelle sottopopolazioni di imprese "Non star - contribuenti minimi", "Non star" e "Star". Percentili 1 e 99, quartili 1 e 3, mediana – valori in migliaia di euro



La metodologia sopra riportata aveva lo scopo di individuare, mediante indicatori *proxy* degli obiettivi conoscitivi dell'indagine, una sottopopolazione di riferimento per indagare fenomeni particolari rilevabili solo su un sottogruppo di unità della popolazione originaria. Tale approccio, nel definire la popolazione di riferimento di un'indagine statistica sulle imprese, è innovativo rispetto ai criteri usualmente adottati, basati sulle "consuete" variabili (la classificazione di attività economica, gli addetti, il fatturato) e può rappresentare un vantaggio in termini di efficienza della rilevazione, in quanto il fenomeno viene rilevato solo su imprese potenzialmente informative, e di riduzione del disturbo statistico. D'altro canto, un simile approccio presenta alcuni inconvenienti perché dipende fortemente dalla definizione di criteri che "approssimano" gli obiettivi informativi. Tra gli inconvenienti di maggior rilievo vi sono sia la mancanza di comparabilità dei risultati, nel tempo oltre che tra rilevazioni condotte non solo da differenti istituti nazionali di statistica ma anche all'interno della stessa istituzione, sia la possibilità di errori di classificazione delle unità (falsi positivi e negativi).

2.5 La fase di controllo e correzione dei dati

La lista precensuaria, che è stata alla base della selezione del campione (oltre che delle altre operazioni di progettazione dell'indagine) non ha potuto tenere conto degli eventi di demografia di impresa e degli errori e inconsistenze, generati nel periodo intercorso tra la sua compilazione e l'effettivo svolgersi dalla rilevazione. Nelle operazioni sul campo sono stati quindi riscontrati errori di vario tipo, attribuibili alla lista: imprese cessate, imprese inattive, unità duplicate e irreperibili, eccetera. Gli errori di lista hanno costituito circa il 5 per cento delle 260 mila unità selezionate mentre le imprese intervistate, al netto di tali errori, costituiscono il 95 per cento delle totale delle

unità eleggibili (la somma dei gruppi “a” e “c” della Tavola 2.4). Le mancate risposte totali ammontano complessivamente a 13.084 unità.

Tavola 2.4 - Risultato della fase di rilevazione sul campo – valori assoluti e in percentuale

	Valori assoluti			Valori percentuali		
	Imprese con meno di 10 addetti	Imprese con almeno 10 addetti	Totale imprese	Imprese con meno di 10 addetti	Imprese con almeno 10 addetti	Totale imprese
Unità rilevate (a+b)	131.090	115.936	247.026	94,9	95,1	95,0
- intervistate (a)	122.696	111.318	234.014	88,8	91,3	90,0
- errori di lista (b)	8.394	4.618	13.012	6,1	3,8	5,0
Unità non rilevate (c)	7.100	5.984	13.084	5,1	4,9	5,0
Totale lista (a+b+c)	138.190	121.920	260.110	100,0	100,0	100,0

Tali positivi risultati sono stati ottenuti anche grazie all’impegno profuso, durante la fase di rilevazione sul campo dal Servizio censimenti economici, dagli Uffici provinciali di censimento, istituiti presso le Camere di commercio, industria, artigianato e agricoltura (Cciaa), da alcune Unioni regionali delle Cciaa, dall’Istituto Provinciale di Statistica (Astat) della Provincia Autonoma di Bolzano e presso il Servizio di Statistica della Provincia Autonoma di Trento che hanno provveduto sollecitare le imprese inadempienti per limitare il tasso di mancata risposta totale¹¹⁸. Tali uffici hanno inoltre provveduto a un primo controllo dei questionari pervenuti per rimuovere almeno parte delle mancate risposte parziali e delle incongruenze logiche tra variabili. Tuttavia alla fine della fase delle operazioni sul campo permanevano nel materiale raccolto mancate risposte totali e parziali che sono state individuate e “corrette” presso l’Istituto secondo le metodologie correnti.

2.5.1 Procedura di correzione per mancata risposta totale

La procedura di correzione per mancata risposta totale è stata effettuata con due metodi diversi a seconda che le unità mancanti appartenessero all’insieme delle imprese campionate (imprese con un numero di addetti inferiore a venti) o a quello delle imprese censite (imprese con un numero di venti e più addetti). Per le prime si è scelto di operare correggendo i pesi diretti di riporto all’universo secondo la procedura già illustrata nel Paragrafo 2.3, mentre per le seconde si è proceduto all’imputazione dei valori di tutte le variabili di interesse secondo la metodologia sotto riportata. In ambedue i casi è stata utilizzata come popolazione obiettivo quella rappresentata dall’archivio Asia 2011, disponibile al momento delle operazioni di trattamento dei dati. È stato quindi necessario riallineare la popolazione obiettivo adottata in fase di disegno d’indagine con quella usata per la calibrazione del campione e per le tecniche di trattamento della mancata risposta totale. In particolare, il riallineamento è stato operato a livello micro per trattare i casi di imprese con differente classe di addetti e/o differente attività economica riscontrate dopo la conclusione delle operazioni censuarie.

Tra le imprese con venti e più addetti, il tasso di mancata risposta totale è stato pari a circa il 15 per cento. I valori mancanti sono stati trattati con tecniche di imputazione in quanto si è voluta sfruttare la disponibilità per tutte le unità di diverse variabili ausiliarie, soprattutto di natura

¹¹⁸ Per approfondire vedi il Capitolo 3 della Parte Seconda del Fascicolo 1 della presente pubblicazione.

continua, che permettevano un utilizzo flessibile e mirato a livello micro e che non sarebbero state utilizzate appieno dall'applicazione di metodi di calibrazione.

A causa dell'elevato numero di variabili da imputare e della difficoltà di utilizzare un modello statistico esplicito in grado di tenere in considerazione le complesse relazioni tra le variabili, il metodo più adatto di imputazione è sembrato essere il metodo del donatore di minima distanza (Dmd) che consiste nel trasferire al corrispondente ricevente i valori delle variabili da imputare osservati sul donatore, dopo aver associato ogni record incompleto (ricevente) ad un record completo (donatore) sulla base di una specifica funzione di distanza.

Tra i metodi di imputazione Dmd, è stato utilizzato il metodo di *Predictive Mean Matching* (PMM) dove, differentemente dai metodi che utilizzano una funzione di distanza "propria" (come la distanza euclidea o di Manhattan), l'influenza delle informazioni ausiliarie per la selezione dei donatori dipende dal loro potere predittivo rispetto alle variabili che devono essere imputate (variabili *target*).

Il PMM, frequentemente utilizzato per variabili *target* continue, è composto da due fasi principali. Nella prima fase, viene stimato un modello di regressione delle variabili *target* su un insieme di covariate osservate su tutte le unità e, sulla base dei parametri del modello, vengono calcolate le medie predittive delle variabili *target* sia delle unità riceventi che delle unità donatrici. Nella seconda fase, per ogni unità ricevente viene selezionato il donatore più vicino rispetto alle medie predittive, dove la distanza è calcolata utilizzando la metrica di Mahalanobis basata sulla matrice di covarianza residua dal modello di regressione. Intuitivamente, la matrice di Mahalanobis associa pesi più grandi alle medie predittive cui corrispondono errori di previsione più piccoli.

Uno dei vantaggi del PMM consiste nel fatto che, nell'ipotesi di un modello sottostante adeguato, questo metodo consente di "pesare" le variabili ausiliarie disponibili in base al loro grado di associazione con un set di variabili *target* predefinito.

In questo contesto specifico, le tre variabili *target* - utili per identificare i diversi profili di impresa - sono state indicate dagli esperti di settore: tipologia e nazionalità del management (persona fisica o famiglia italiana, altra impresa italiana, management estero, altra tipologia di management italiana), tipo di relazioni dell'impresa (presenza di accordi formali e informali, presenza di soli accordi informali, assenza di accordi), aree di mercato dell'impresa (solo comune di localizzazione dell'impresa, in Italia ma non solo nel comune di localizzazione dell'impresa, all'estero).

Come covariate sono state utilizzate le variabili: numero di addetti, valore aggiunto, fatturato, appartenenza ad un gruppo, attività di import/export; l'informazione sull'attività economica dell'impresa (classificazione Ateco) è stata utilizzata per creare strati all'interno dei quali procedere all'associazione tra unità riceventi e donatori.

Data la natura non continua delle variabili *target*, il metodo PMM è stato opportunamente adattato. In particolare, è stato sviluppato un metodo innovativo di PMM che utilizza una variabile latente continua (fattore latente) che mette in relazione in maniera opportuna le covariate e le variabili *target*. Il PMM con un fattore latente sostituisce, nella prima fase del PMM tradizionale, l'utilizzo di un modello ad equazioni strutturali al modello di regressione. Il modello utilizzato risulta essere composto da due parti: un modello fattoriale che mette in relazione il fattore latente con le variabili *target* osservate - indicatori - e un modello di regressione che mette in relazione le covariate con il fattore latente. In altre parole, il modello è un modello a equazioni strutturali con Q variabili osservate esogene e una variabile latente continua endogena misurata da P indicatori. È importante notare che gli indicatori osservati possono essere di natura sia categorica che continua. Nella seconda fase del PMM, per ogni unità ricevente viene selezionato un donatore sulla base della distanza tra le rispettive medie predittive del fattore latente.

Questa procedura ha il vantaggio di poter essere applicata con un numero elevato di variabili *target* e di poter essere utilizzata per variabili di diversa natura (continua e/o categorica). Il metodo

descritto è stato valutato mediante uno studio di simulazione basato sulla introduzione artificiale di valori mancanti. La sperimentazione ha mostrato una buona performance del metodo rispetto ad altri metodi classici come il donatore di minima distanza basato sulle covariate o metodi propriamente parametrici basati su modellazione esplicita dei dati.

2.5.2 Il trattamento delle mancate risposte parziali e delle incongruenze logiche

Come già accennato, un primo piano di controllo dei dati è stato introdotto durante la fase di registrazione delle informazioni, avvenuta sia via web sia con l'ausilio dei revisori/rilevatori degli Uffici provinciali di censimento, in cui la fase di *data entry* poteva terminare soltanto se alcuni quesiti venivano compilati in modo corretto rispetto a predeterminate regole di controllo. Per evitare un'eccessiva pressione sui rispondenti, non tutte le relazioni anomale tra variabili davano luogo a blocco del questionario ma solo a "*warning*", rispetto ai quali il rispondente poteva correggere l'informazione fornita oppure confermarla. In questo modo, al termine della compilazione, potevano permanere mancate risposte e incongruenze logiche che sono state trattate direttamente dall'Istat, utilizzando tecniche automatiche di individuazione e correzione degli errori. A tal fine è stato utilizzato il software generalizzato Scia (Sistema controllo e imputazione automatici) da tempo comunemente impiegato nell'Istituto in numerosi contesti operativi di indagine per il trattamento di variabili categoriali. L'algoritmo presente in Scia è di tipo stocastico basato sulla metodologia di Fellegi e Holt¹¹⁹.

A differenza di un approccio di tipo deterministico, in cui per ogni situazione di errore, si devono definire le varie azioni da intraprendere, con lo strumento generalizzato Scia, l'esperto statistico deve limitarsi a definire le situazioni di errore (*edit*), demandando al metodo automatico presente nel software, il compito di riportare l'unità osservata ad una situazione di correttezza.

La procedura si divide concettualmente in tre fasi. La prima consiste nell'individuazione di un sistema di "regole" (o *edit*) non contraddittorio e completo (ovvero che garantisca sempre una soluzione all'errore), sia esplicitate direttamente dal responsabile della procedura sia individuate dal sistema stesso perché implicitamente contenute nelle regole del primo tipo. Nella seconda fase, Scia individua i valori incoerenti presenti in un record in base a tale insieme di regole e al principio del minimo cambiamento, secondo il quale l'insieme di variabili da correggere in un record errato, deve essere il minore numero possibile. Ciò significa che se per portare un record errato in una situazione di correttezza vi sono per esempio, due soluzioni possibili, una correggendo i valori di due variabili e un'altra i valori di tre, la procedura sceglie come soluzione quella con cardinalità minima. Se almeno un *edit* viene soddisfatto (ovvero viene riscontrata una violazione di una regola), il record in esame è considerato "errato" e confluisce nell'insieme delle unità da correggere, mentre se non manifesta nessuna incompatibilità, viene considerato esatto ed entra a far parte del bacino dei "record donatori" da cui attingere nella fase di imputazione.

Nella terza fase, quella dell'imputazione per ripristinare la coerenza tra i valori di tutte le variabili del record in esame, i dati errati sono trasformati in dati corretti, scegliendo nell'insieme dei record esatti il miglior (o i migliori) donatore/i possibile/i per ogni unità e variabile da imputare, nel rispetto delle regole e cercando di preservare il più possibile le distribuzioni delle variabili, riscontrate nell'insieme dei record esatti.

Le regole di Scia sono espressioni di relazioni logiche tra variabili di tipo categoriale e si possono definire come:

¹¹⁹ La metodologia di Fellegi e Holt è esposta nell'articolo "*A systematic approach to automatic edit and imputation*", pubblicato dagli autori I. Fellegi e D. Holt nel 1976 sul *Journal of the American Statistical Association*.

- relazioni di incompatibilità o incoerenza, ossia prefissate combinazioni di valori assunti da alcune variabili rilevate in una stessa unità devono soddisfare certi requisiti (alcuni esempi nel questionario: se si risponde che si sono acquisite risorse nel quesito 2.02, ma non si indica la tipologia di risorse acquisite nel quesito successivo 2.03; all'interno dello stesso quesito si presenta un'incoerenza se è stata indicata come risposta "nessuna delle precedenti" e contemporaneamente sono state indicate anche una o più delle altre modalità previste);
- valori fuori dominio, ossia i valori assunti da una certa variabile devono essere interni all'intervallo di definizione della variabile stessa (un esempio nel questionario: la quota percentuale di spesa per innovazioni deve essere inclusa tra lo zero e il cento per cento).

La presenza o meno di valori mancanti, può essere indicata nella fase di scrittura delle regole, a seconda che si desideri un'informazione completa sul record (il dato mancante non è ammesso) oppure nel caso in cui vi siano risposte consecutive a domande filtro (il dato mancante è ammesso in quanto l'utente non deve fornire una risposta).

Come già descritto (Paragrafo 2.2), la rilevazione ha utilizzato due modelli di questionari distinti a seconda della dimensione d'impresa: un questionario per le imprese con meno di dieci addetti, denominato modello CIS2011_QS, e un altro questionario per le rimanenti imprese, denominato modello CIS2011_QX_Bis. I due modelli hanno quesiti comuni e altri quesiti con sezioni e sottosezioni specifiche per modello. Sono inoltre strutturati in sette sezioni tematiche, oltre ad una domanda iniziale sullo stato di attività dell'impresa, presente in entrambi i modelli. Tali blocchi informativi risultano, generalmente, tra loro logicamente indipendenti ovvero i quesiti della singola sezione non sono logicamente correlati con quelli delle altre. È stato quindi possibile, in larga parte, implementare il piano di controllo e correzione complessivo come sommatoria di singoli piani relativi alle singole sezioni del questionario. Ne è quindi risultata una notevole semplificazione della procedura, sia dal punto di vista logico/statistico sia dal punto di vista informatico/applicativo. Gran parte dei "warning" previsti per il controllo nella fase di compilazione del questionario sono stati inseriti come *edit* nel trattamento con il software Scia, insieme ad altri *edit* stabiliti dagli esperti d'indagine.

In sintesi, i passi del piano di controllo e correzione comuni a tutte le sezioni dei questionari, sono stati:

- 1) pre-trattamento dei dati, ossia correzioni di tipo deterministico, che hanno riguardato in particolare alcune variabili quantitative e preparazione dei formati per trattare i dati con il software generalizzato;
- 2) individuazione dell'errore presente nelle variabili categoriche, secondo l'approccio probabilistico implementato da Scia;
- 3) imputazione tipo "hot deck" dei valori errati localizzati nella fase precedente, svolta con Scia;
- 4) analisi dei risultati di Scia e preparazione finale dei dati.

Di seguito, nella Tavola 2.5 sono riportati i principali risultati del piano di controllo e correzione distinti per sezione di questionario.

Tavola 2.5 - Risultati della fase di controllo e correzione dei dati - valori percentuali

	Record considerati (%)		Record considerati errati per numero di violazioni (%)			
	"esatti"	"errati"	1	2	3	Oltre 3
Sezioni 1, 2	78,0	22,0	64,8	7,2	7,6	20,3
Sezione 3	79,0	21,0	34,3	16,7	14,6	34,4
Sezione 4	86,0	14,0	33,3	40,3	14,9	11,5
Sezione 5	78,0	22,0	47,8	35,9	10,6	5,7
Sezione 6	84,0	16,0	58,8	15,6	7,9	17,7
Sezione 7	86,0	14,0	97,6	2,1	0,3	..

2.6 In sintesi

Nella progettazione e nella realizzazione dell'Indagine multiscopo sulle imprese sono state introdotte notevoli innovazioni non solo dal lato dei contenuti ma anche sul versante metodologico, oltre a quello informatico, per il quale si rimanda al Fascicolo 6 – Il sistema informatico integrato.

A livello di contenuti sono state raccolte “nuove” informazioni utili alla valutazione degli elementi di forza e di debolezza del sistema delle imprese, informazioni non rilevate correntemente dall'Istat e comunque disponibili nel nostro Paese in modo frammentario e parziale. È stata estesa, inoltre, la disponibilità di dati relativi alle tecnologie ICT e all'Innovazione, al sotto-universo delle imprese con meno di dieci addetti, non coperto dalle indagini correnti condotte sugli stessi temi dall'Istituto.

Tali risultati sono integrabili con quelli delle altre rilevazioni censuarie, degli archivi statistici e delle indagini correnti sulle imprese concorrendo a delineare un quadro informativo sulle imprese sempre più ricco e dettagliato.

La novità dei contenuti informativi si è riflessa sulle innovazioni metodologiche che hanno riguardato le fasi di identificazione della popolazione di riferimento, il metodo di allocazione del campione nei differenti strati e la fase di trattamento dei dati.

Nell'ipotesi che i fenomeni considerati riguardassero solo una quota ridotta delle microimprese, era necessario definire una soglia che, con il vincolo di comparabilità con i dati del censimento, escludesse tali imprese per concentrare le risorse nella raccolta dei dati nelle aziende con maggiore probabilità di essere toccate dai suddetti fenomeni. Tale soglia è stata definita a seguito di un processo di analisi che ha considerato ipotesi alternative basate anche su misure di “ampiezza delle imprese” differenti dal numero di addetti.

Ai fini di tenere conto anche di quelle microimprese, così definite in base al numero di addetti, che hanno una complessità organizzativa ed economica pari se non superiore alle imprese sopra soglia, e quindi a tutti gli effetti unità da intervistare, sono stati utilizzati archivi amministrativi e statistici (bilanci civilistici, Studi di settore, Asia e Coe) da cui sono stati elaborati indicatori che hanno permesso di selezionare il suddetto sottogruppo di “micro-star”. Tale procedura è fortemente innovativa sia poiché taglia i consueti strati e domini di analisi (attività economica, addetti) sia per l'integrazione di basi di dati per il calcolo di indicatori specifici in funzione degli obiettivi conoscitivi. La procedura sopra descritta può risultare vantaggiosa, in termini di efficienza della rilevazione e di riduzione del “fastidio statistico” (limitando la raccolta alle sole unità interessate dal fenomeno) laddove debbano essere analizzati fenomeni non uniformemente distribuiti su tutte le imprese e che hanno bassa correlazione con le variabili “attività economica” e “addetti” usualmente utilizzate come criteri di stratificazione e come principali domini di studio. Tale procedura, tuttavia, oltre a presupporre l'esistenza di archivi che abbiano disponibilità di informazioni su tutte le unità della popolazione in esame, presenta, almeno al momento, una forte soggettività nella selezione degli indicatori *proxy* per l'identificazione della sotto-popolazione di

riferimento e quindi può generare falsi positivi e negativi che potrebbero condizionare l'analisi, e, infine, non permette la comparabilità dei risultati con le indagini correnti sulle imprese. Per tali motivi il campione derivato dalla popolazione di microimprese "star" è stato considerato come sperimentale e oggetto di ulteriori studi.

Per il trattamento delle mancate risposte totali sono state utilizzate tecniche diverse nel caso degli strati in cui era stata progettata la rilevazione di tutte le unità ovvero nel caso degli strati in cui si è proceduto ad un campionamento. Per compensare l'effetto della mancata risposta, nel primo caso si è scelto di procedere all'integrazione dei non rispondenti con unità ricostruite mediante variabili acquisite dal rispondente più "simile" mentre, nel secondo caso, sono stati utilizzati "correttori" dei pesi iniziali (il cui calcolo si basa sull'inverso delle probabilità di inclusione nel campione). Per il trattamento delle mancate risposte parziali e delle incongruenze logiche, nell'ottica di massimizzare l'uso dei rilevatori/revisori degli Uffici provinciali di censimento, si è dedicata molta attenzione, nella fase di progettazione, all'integrazione tra la procedura di controllo supportata dagli Upc e quella condotta centralmente dall'Istituto per evitare ridondanze e contraddizioni. Tali procedure sono ampiamente utilizzate nelle rilevazioni correnti sulle imprese; tuttavia nel caso dell'indagine multiscopo sono state apportate significative innovazioni e sperimentazioni che hanno prodotto miglioramenti ai metodi standard che si riverseranno sulla qualità delle indagini correnti; per tale motivo l'indagine ha rappresentato un investimento per il futuro sia in termini di patrimonio informativo che di esperienze metodologiche acquisite.

Capitolo 3

La Rilevazione multiscopo sulle unità complesse¹²⁰

3.1 Le finalità e gli obiettivi dell'indagine

La Rilevazione multiscopo sulle unità complesse completa il quadro informativo fornito dal 9° Censimento generale dell'industria e dei servizi secondo un approccio di classificazione, rilevazione ed analisi dei dati di impresa radicalmente innovativo. In linea con la strategia generale del censimento, che privilegia l'utilizzo dei dati amministrativi, la raccolta dati sul campo ha interessato un numero molto limitato di imprese, circa 3.500 unità economiche residenti in Italia accuratamente selezionate in considerazione della loro posizione nell'ambito di gruppi di imprese nazionali o multinazionali. Queste imprese sono state tuttavia chiamate a rispondere per l'intero perimetro dell'unità economica complessa di cui sono poste al vertice, rappresentando così una quota molto rilevante di attività economiche realizzate in Italia e all'estero.

La rilevazione, per sua natura di carattere sperimentale, risponde ad un triplice obiettivo.

In primo luogo, misurare in modo più accurato e coerente le peculiari caratteristiche di quel particolare segmento di imprese residenti in Italia che, a prescindere dai tradizionali schemi di classificazione di natura settoriale e dimensionale, presentano un notevole grado di complessità nei modelli organizzativi e nei processi decisionali. L'internazionalizzazione dei mercati e dei processi produttivi, nonché il crescente ruolo dei fattori immateriali, per la competitività delle imprese hanno infatti stimolato molte aziende ad evolvere verso modelli organizzativi più complessi ed articolati ed a dotarsi di capacità manageriali avanzate per rispondere meglio alle sfide della globalizzazione. L'individuazione di queste imprese, che rappresentano la componente più dinamica del sistema produttivo italiano, è molto difficile impiegando i tradizionali schemi di classificazione, quali il settore di attività economica e la dimensione economica dell'impresa considerata come unità legale autonoma. È stato pertanto necessario realizzare un'approfondita ed accurata progettazione dell'indagine sulla base di schemi di classificazione e modalità di rilevazione dei dati più aderenti alla complessa realtà organizzativa e manageriale di queste imprese.

In secondo luogo, un altro importante obiettivo è stato quello di progettare e sperimentare sul campo la raccolta dati – invero limitata ad alcune informazioni di base sulle caratteristiche dei sistemi informativi aziendali - di alcune variabili economiche in modo coerente con il nuovo insieme di definizioni internazionali sulle unità economiche. È, infatti, in corso da alcuni anni un intenso dibattito a livello europeo sulle nuove definizioni di impresa nell'ambito delle attività preparatorie per l'approvazione del nuovo regolamento quadro europeo che integra vari domini delle statistiche d'impresa denominato FRIBS (*Framework Regulation Integrating Business Statistics*). Sperimentazioni sono in corso in vari paesi dell'Ue e la produzione di statistiche SBS¹²¹ secondo nuove linee guida per l'individuazione e la corretta classificazione delle imprese entrerà a breve in vigore. Le informazioni raccolte dalla Rilevazione multiscopo sulle unità complesse sono essenziali

¹²⁰ A cura di Stefano Menghinello. Autori del capitolo:

- Paragrafi 3.1, 3.2 e 3.3: Stefano Menghinello;

- Paragrafi 3.4 e 3.5: Domenico Moro.

¹²¹ I risultati economici delle imprese sono prodotti conformemente a quanto disposto dal Regolamento Ue n. 295/2008 per le statistiche strutturali (SBS - *Structural Business Statistics*). Il regolamento SBS definisce un quadro comune per la raccolta, l'elaborazione e la trasmissione dei dati allo scopo di disporre, annualmente, di statistiche armonizzate per valutare la struttura, l'attività e la competitività delle imprese nell'Unione Europea.

per acquisire indicazioni più accurate sulla capacità delle unità economiche più complesse di fornire dati secondo nuovi schemi di classificazione.

In terzo luogo, la rilevazione è stata utilizzata quale veicolo comunicativo per proporre alle unità economiche più complesse – grandi imprese, gruppi di imprese nazionali ed internazionali – un utilizzo più ampio e articolato dei servizi offerti dal Portale delle imprese, il *single entry point* recentemente attivato dall'Istat per la raccolta dati e che interesserà tutte le imprese residenti in Italia con almeno venti addetti che sono state selezionate per almeno una rilevazione corrente dell'Istat. Nel caso delle unità economiche più complesse, caratterizzate da un numero spesso elevato di imprese controllate e da una elevata frammentazione delle informazioni economiche nell'ambito del proprio perimetro di controllo, è stato possibile acquisire tramite il questionario di indagine indicazioni puntuali sulle modalità migliori secondo cui desiderano essere intervistate tenendo conto delle proprie caratteristiche organizzative e del grado di integrazione dei sistemi informativi aziendali.

Le principali innovazioni introdotte dalla Rilevazione multiscopo sulle unità complesse fanno parte di un quadro integrato per la progettazione e l'implementazione dell'indagine che ha coinvolto un numero elevato di esperti di settore sia interni che esterni all'Istat, tra cui si segnala l'essenziale contributo fornito dal Prof. Gianluca Spina del Politecnico di Milano prematuramente scomparso.

Le principali innovazioni introdotte dalla rilevazione sono state:

- a) utilizzo di unità di analisi e di schemi di classificazione più coerenti nell'individuare il perimetro effettivo delle unità complesse e più accurati nel descriverne l'articolazione interna e le relazioni esterne con clienti e fornitori (Paragrafo 3.2) a partire da un approccio di selezione della popolazione obiettivo di tipo *register-based* (Paragrafo 3.3).
- b) progettazione ed implementazione del questionario con quesiti articolati in sezioni tematiche che realizzano un progressivo approfondimento su aspetti molto complessi dell'organizzazione e dei processi decisionali dell'impresa secondo un quadro coerente di concetti, definizioni e percorsi di compilazione (Paragrafo 3.2).
- c) interazione con il top management aziendale sulla base di rilevatori esperti di tematiche aziendali selezionati sulla base di una gara internazionale che ha visto la partecipazione delle principali società nazionali ed internazionali nel settore della consulenza aziendale e della revisione e analisi dei bilanci (Paragrafo 3.4).

La progettazione della rilevazione, sia da un punto di vista concettuale, metodologico che organizzativo ha richiesto un periodo molto lungo ma è stata comunque completata nei tempi programmati. Le attività di raccolta dati sul campo sono invece iniziate con un forte ritardo causa il protrarsi dei tempi di espletamento della gara internazionale per la selezione del team di esperti aziendali incaricati dall'Istat di svolgere le interviste al top management delle unità complesse incluse nella rilevazione. Al fine assicurare un elevato grado di aggiornamento delle informazioni raccolte il periodo di riferimento dei dati è stato posticipato al 2013 con alcuni quesiti che considerano la serie storica per il triennio 2011-2013.

3.2 Il questionario

Il questionario fa riferimento ad un insieme di definizioni di base che costituiscono il principale riferimento per individuare l'unità di indagine, definire la struttura delle domande e individuare i percorsi e le modalità di compilazione.

L'unità complessa costituisce l'unità di analisi per tutti i quesiti del questionario. Essa comprende l'impresa intervistata e, se presenti, tutte le sue controllate, dirette o indirette, residenti in Italia o anche all'estero. L'insieme di queste imprese costituisce il perimetro dell'unità complessa. Per

controllo si intende la capacità di determinare l'attività generale dell'impresa, anche scegliendo gli amministratori più idonei. La quota di proprietà del capitale sociale di un'impresa viene di solito impiegata come *proxy* per misurare la presenza o meno di controllo. In particolare, un'impresa è controllata quando un'altra impresa detiene direttamente o indirettamente il cinquanta per cento più 1 delle sue quote o azioni con diritto di voto.

Ad ogni unità complessa corrisponde una ed una sola unità di rilevazione che è tenuta a fornire risposte al questionario considerando sempre come riferimento l'intero perimetro. È una impresa residente in Italia selezionata sulla base di informazioni integrate provenienti dagli archivi e dalle indagini dell'Istat ed è costituita di solito dalla capo gruppo, la holding operativa o l'impresa, a seconda dell'articolazione ed organizzazione dell'unità complessa in una o più unità legali.

Al fine di rilevare informazioni coerenti e complete sulla struttura organizzativa interna dell'unità complessa e delle sue relazioni esterne con clienti e fornitori il questionario fa riferimento alle seguenti due definizioni che costituiscono di fatto le principali variabili di classificazione dell'indagine.

Linea di attività (*business line*): è il settore operativo in cui si articola l'attività di base dell'unità complessa che può comprendere uno o più linee di attività. Al fine di definire una linea di attività è necessario che l'unità complessa intraprenda attività imprenditoriali generatrici di ricavi e di costi (compresi i ricavi e i costi riguardanti operazioni con altre componenti della medesima entità); che i ricavi e costi siano individuabili e misurabili separatamente per ogni linea di attività e che i risultati operativi siano rivisti periodicamente dal più alto livello decisionale operativo ai fini dell'adozione di decisioni in merito alle risorse da allocare al settore e della valutazione dei risultati.

Nell'ambito di ciascuna linea di attività è possibile decomporre il processo aziendale nelle seguenti tipologie di attività:

Produzione di beni o servizi (*core activity*): costituisce l'attività economica principale dell'unità complessa nell'ambito della linea di attività ed è finalizzata alla produzione di beni o servizi.

Funzione aziendale (*business function*): nell'ambito del processo aziendale sono le funzioni a supporto della produzione di beni o servizi. Queste includono la progettazione e altri servizi tecnici per la produzione (incluso *engineering* e *design*); la ricerca e sviluppo; la vendita, marketing e servizi commerciali; la distribuzione e logistica; l'assistenza e servizi post vendita di natura tecnica per la commercializzazione; l'attività di supporto contabile, finanziario e manageriale ed i servizi di tecnologie dell'informazione e della comunicazione (ICT/Tic).

Mentre la prima tipologia di variabile di classificazione - la linea di attività - è di tipo "aperto" ossia è definita in modo autonomo dall'impresa rispondente nel pieno rispetto delle definizioni e dei criteri di classificazione forniti dall'Istat, la seconda - le fasi del processo aziendale - è invece di tipo "chiuso" ossia viene fornita nel questionario una lista predefinita di modalità che devono essere impiegate in tutte le domande.

La combinazione di queste due variabili di classificazione, che è presente in alcuni dei principali quesiti della rilevazione, consente di realizzare una mappatura accurata e completa dell'organizzazione interna all'unità complessa nonché di analizzare la struttura dei rapporti esterni (clienti e fornitori) distinguendo anche quali attività o fasi sono svolte in Italia o all'estero.

Il questionario include settanta domande e si articola in quattro principali sezioni. Ognuna di queste risponde a specifici obiettivi conoscitivi in un quadro coerente di definizioni e classificazioni. Ciascuna sezione può essere suddivisa in aree tematiche. Ad ogni area tematica corrisponde un numero di quesiti che realizzano un progressivo approfondimento sul fenomeno oggetto di interesse. In generale, le informazioni raccolte dal questionario - secondo criteri di identificazione delle unità e di classificazione innovativi - non sono rilevabili da altre rilevazioni o da fonti fiscali/amministrative già disponibili all'Istat. Le informazioni più rilevanti ed innovative per l'analisi

della competitività delle unità economiche complesse riguardano: il posizionamento competitivo per linee di attività e il relativo grado di specializzazione; le caratteristiche e i profili professionali del management; l'analisi dei processi decisionali; la capacità del management di pianificare e gestire processi complessi come l'internazionalizzazione e la creazione di nuove conoscenze; la pianificazione e la gestione strategica di risorse interne ed esterne. Viene riportata a seguire una descrizione sintetica dei contenuti informativi del questionario.

Sezione A - Principali caratteristiche dell'unità economica complessa

La Sezione A include solo tre domande che predispongono alla compilazione del resto del questionario a partire da una corretta identificazione delle principali caratteristiche dell'unità complessa.

Sezione B - Organizzazione e sistemi informativi aziendali

La Sezione B si compone di due aree tematiche per un totale di diciotto quesiti. Ha come obiettivo quello di mappare l'organizzazione dell'unità complessa e i relativi flussi informativi aziendali. In particolare:

L'Area tematica 1 - Organizzazione dell'Uec – include dieci domande che analizzano l'organizzazione interna e le relazioni esterne dell'unità complessa e ne identificano le linee di attività. Per ciascuna linea di attività individuata, vengono richieste informazioni su:

- la destinazione dell'output e il grado di autonomia finanziaria e organizzativa;
- la rilevanza economica e la localizzazione della produzione in Italia o all'estero;
- le principali unità legali residenti in Italia che realizzano due o più linee di attività;
- la presenza di imprese o unità legali dedicate a specifiche funzioni operative o finanziarie in Italia e all'estero;
- le funzioni del processo aziendale che sono realizzate all'interno dell'unità complessa in Italia e all'estero;
- le funzioni del processo aziendale che sono realizzate da imprese all'esterno dell'unità complessa in Italia e all'estero.

L'Area tematica 2 - Organizzazione dei flussi informativi all'interno e all'esterno dell'Uec - comprende otto domande che approfondiscono le principali caratteristiche tecnologiche e funzionali dei sistemi informativi aziendali. Queste domande raccolgono informazioni su:

- la presenza di strutture organizzative dedicate a pianificazione e controllo;
- la presenza di sistemi di reportistica aziendale dedicati;
- l'organizzazione dei flussi informativi a livello di singola linea di attività in Italia e/o all'estero;
- il livello di impiego di tecnologie ICT;
- la tipologia di informazione prodotta a fini amministrativi e fiscali;
- il livello di integrazione della documentazione di carattere economico con i sistemi di reportistica aziendale;
- gli standard tecnologici e contabili utilizzati per la predisposizione e la trasmissione dei bilanci;
- la disponibilità di macro informazioni, per linea di attività, in relazione alle principali voci di bilancio.

Sezione C - Determinanti della performance dell'unità complessa

La Sezione C si articola in sei aree tematiche per un totale di quarantadue domande. Ciascuna area tematica ha l'obiettivo di approfondire, da una specifica prospettiva di analisi, l'articolazione degli assetti organizzativi e dei processi decisionali nell'ambito dell'unità complessa.

L'Area tematica 1 - Fattori di competitività e pianificazione di nuove strategie - comprende quattro quesiti che analizzano, per ciascuna linea di attività, il posizionamento competitivo dell'unità complessa con particolare riguardo al mercato di riferimento, ai fattori di competitività e ai punti di forza dei principali concorrenti. Sono inoltre raccolte informazioni sulle strategie di sviluppo che l'unità complessa intende adottare nel periodo 2013-2015, distinguendo tra Italia ed estero.

L'Area tematica 2 - Caratteristiche del management e *governance* dei processi decisionali – include sette quesiti che raccolgono informazioni su:

- la tipologia della proprietà, la modalità di controllo proprietario e gli organi di governo dell'unità complessa;
- il grado di autonomia decisionale tra livello centralizzato e linee di attività in relazione ad alcune scelte strategiche;
- i profili anagrafici, formativi e professionali del top e del middle management;
- l'esperienza professionale precedente e l'anzianità aziendale media del top management;
- i percorsi di formazione e l'esperienza professionale del top management che ha legami familiari con la proprietà;
- la localizzazione della responsabilità (Italia e/o estero) del top e del middle management e la provenienza geografica di quello stabilmente impiegato all'estero.

L'Area tematica 3 - Gestione delle risorse umane – comprende sei domande che analizzano i seguenti aspetti:

- il titolo di studio dei dipendenti con funzioni non manageriali;
- le modalità di selezione e l'esistenza di percorsi formativi per le nuove risorse umane;
- le strategie aziendali relative alla formazione, alla gestione e alla crescita professionale dei dipendenti;
- la struttura e l'autonomia decisionale dei gruppi di lavoro formalizzati;
- l'utilizzo di strumenti di comunicazione online per la gestione corrente dei rapporti di lavoro.

L'Area tematica 4 - Rapporti con fornitori e clienti esterni – comprende dodici domande, finalizzate a raccogliere informazioni su:

- la struttura degli acquisti di beni e servizi da imprese esterne al perimetro dell'unità complessa per linea di attività;
- le principali strategie adottate nei rapporti con i fornitori di beni e servizi;
- le scelte strategiche rilevanti nella gestione corrente dei rapporti con i fornitori esterni di beni e servizi;
- la valutazione dei principali aspetti dell'organizzazione dedicata agli acquisti di beni e servizi;
- le strategie di approvvigionamento adottate per l'approvvigionamento di merci e servizi sui mercati internazionali;
- la distribuzione percentuale dei destinatari delle vendite e la destinazione geografica del fatturato per linee di attività;
- l'orientamento prevalente, per linea di attività, delle strategie di vendita e dei criteri di fissazione del prezzo;

- le strategie adottate per la distribuzione e il marketing per linea di attività;
- il grado di centralizzazione nella gestione dei rapporti con i clienti e i criteri adottati per individuare i clienti;
- la valutazione dei principali aspetti dell'organizzazione dedicata alla gestione dei rapporti con i clienti;
- le tipologie di attività di supporto commerciale e logistico per linea di attività;
- le strategie e modalità di vendita di beni e servizi sui mercati internazionali.

L'Area tematica 5 – Internazionalizzazione dei processi produttivi e tecnologici dell'unità complessa – include quattro quesiti orientati a raccogliere informazioni su:

- le modalità di internazionalizzazione dei processi produttivi e tecnologici (in attività di *engineering* o R&S);
- le motivazioni sull'eventuale assenza di internazionalizzazione dei processi produttivi e tecnologici;
- i principali vantaggi connessi con l'internazionalizzazione dei processi produttivi per linea di attività;
- i fattori determinanti nella scelta della localizzazione, in Italia o all'estero, delle attività di R&S, progettazione ed *engineering*.

L'Area tematica 6 - Gestione della conoscenza (*knowledge management*) e processi di innovazione - comprende nove domande che esplorano le principali caratteristiche dei processi di creazione di conoscenza che interessano l'unità complessa nonché sui processi decisionali che ne orientano la gestione all'interno dell'organizzazione aziendale. In particolare, le informazioni raccolte riguardano:

- le principali tipologie e aree di applicazione della conoscenza disponibili nell'unità complessa per la progettazione, produzione e vendita di beni e servizi, suddivise per fonte di provenienza;
- la valutazione delle principali tipologie di conoscenza necessarie per sostenere la competitività e di quelle disponibili rispetto al fabbisogno;
- le azioni programmate per ridurre il divario tra le conoscenze disponibili e quelle necessarie a sostenere la competitività nel periodo 2014-2015;
- la distribuzione per tipologia di lavoratori della conoscenza in Italia e all'estero;
- il grado di mobilità dei lavoratori della conoscenza tra le sedi di lavoro;
- il grado di utilizzo degli strumenti telematici da parte dei lavoratori della conoscenza;
- la presenza di figure professionali o modalità organizzative esplicitamente dedicate alla gestione o alla crescita professionale dei lavoratori della conoscenza;
- le modalità prevalenti delle caratteristiche dei processi innovativi realizzati a livello *corporate* o all'interno di ogni linea di attività;
- i meccanismi e gli strumenti utilizzati per il trasferimento e l'integrazione delle conoscenze a livello internazionale.

Sezione D - Ottimizzazione della raccolta dati e nuovi servizi offerti dall'Istat

La Sezione D comprende sette quesiti e si pone obiettivi di carattere relazionale ed informativo che si inquadrano nella strategia più generale dell'Istat tesa a ridefinire i rapporti con le grandi imprese o comunque con le unità più complesse sia da un punto di vista tecnologico (Portale delle imprese e questionari elettronici) che organizzativo. In particolare, questa sezione raccoglie informazioni su:

- le principali difficoltà incontrate dalle imprese incluse nel perimetro dell'unità complessa nel rispondere alle indagini Istat (barriere tecnologiche, complessità questionari, difficoltà nel reperire informazioni data l'articolazione delle attività aziendali);
- la valutazione sulle modalità organizzative meno onerose per l'impresa e più efficaci per la raccolta delle informazioni per tipologia di indagine;
- in relazione all'esigenza precedentemente segnalata (ovvero di inviare i questionari a specifiche strutture organizzative), si richiede di fornire informazioni dettagliate relative alle singole strutture;
- eventuale indicazione dei referenti interni all'impresa o all'unità complessa più idonei a rispondere alle diverse tipologie di indagine;
- le tecniche di raccolta più idonee per interagire con l'Istat nella raccolta e trasmissione dei dati;
- i servizi/strumenti adottati dall'Istat ritenuti più utili per agevolare la compilazione dei questionari;
- il tipo di informazione statistica di maggiore interesse per l'unità complessa nell'ambito del sistema di "restituzione" di informazioni statistiche.

Per quanto riguarda la natura delle informazioni richieste, i quesiti sono di carattere prevalentemente qualitativo. In particolare, è richiesto di fornire informazioni o di esprimere valutazioni del tipo "sì/no" oppure "presente/non presente" oppure "rilevante/non rilevante" rispetto ad una lista di item predefiniti dall'Istat secondo opzione a risposta multipla (quadrato riportato nella domanda) o una sola risposta per ogni domanda o sezione della domanda (pallino riportato nella domanda).

Al fine di consentire ai rilevatori selezionati come esperti aziendali di operare sul campo in modo autonomo e pienamente coerente (rilevazione CAPI), il questionario è stato implementato come sistema informativo autonomo nell'ambito del Portale Istat delle imprese. In particolare, la piena implementazione del complesso piano di check nel questionario elettronico ha consentito percorsi di compilazione sempre coerenti per quanto molto articolati. Inoltre, ogni domanda è stata corredata di uno specifico insieme di definizioni, note esplicative e sistemi di segnalazione (*alert* e *pop-up*) finalizzate a supportare il rilevatore durante la compilazione di tutto il questionario.

3.3 La popolazione di interesse e i domini di studio

La popolazione di interesse è stata individuata secondo un approccio di tipo *register-based* a partire dalla definizione di unità complessa riportata nel paragrafo precedente.

Le fonti informative utilizzate sono state l'archivio statistico sui gruppi di imprese, integrato con dati di fonte amministrativa (bilanci di esercizio e consolidati) e con i risultati delle rilevazioni sulle imprese a controllo estero residenti in Italia (*Inward FATS*) e sulle attività all'estero delle imprese a controllo nazionale (*Outward FATS*). In particolare, l'archivio statistico sui gruppi di imprese ha consentito, sulla base della gerarchia delle relazioni di controllo proprietario, di individuare sia il perimetro dell'unità complessa che le imprese meglio qualificate a rispondere alla rilevazione (di solito in posizione di vertice con funzione di capo gruppo o di holding operativa). I principali risultati delle rilevazioni FATS hanno consentito, da un lato, di estendere il perimetro e la disponibilità di variabili economiche anche alle imprese a controllo nazionale residenti all'estero ed incluse nel perimetro dell'unità complessa (*Outward FATS*) e, dall'altro, di avere un maggiore controllo sulle imprese o gruppi di impresa che appartengono a gruppi multinazionali esteri operanti in Italia

(*Inward FATS*), le quali presentano, come spiegato a seguire, alcuni rilevanti problemi di rilevazione.

Dato che il grado di complessità organizzativa e dei processi decisionali delle unità economiche - si ricorda che l'individuazione delle unità complesse è il principale oggetto di rilevazione - non è osservabile *a priori*, sono stati adottati dei criteri basati su variabili *proxy* lasciando poi alla raccolta dati il ruolo di rilevare e misurare l'effettivo grado di complessità delle unità economiche incluse nella lista di rilevazione.

I criteri di selezione adottati per restringere il campo di osservazione da tutte le imprese o gruppi di imprese residenti in Italia alle sole imprese caratterizzate da una elevata probabilità di presentare un elevato grado di complessità organizzativa e decisionale, sono stati di tipo ragionato. In particolare, sono state individuate delle soglie di esclusione che bilanciano, secondo uno schema a griglia, le classi di addetti con differenti classi di valore del fatturato, relative all'intero perimetro dell'unità complessa. Soglie più elevate sono state utilizzate per unità complesse caratterizzate da particolari forme giuridiche (società cooperative) o attive in settori di attività economica per cui è nota la minor presenza, a parità di dimensione economica, di strutture organizzative e processi decisionali complessi.

La lista delle unità selezionate come risultato di questo processo (circa 3.000 nell'ipotesi più restrittiva e 3.500 nella versione più ampia) è stata sottoposta ad una intensa attività di *profiling* al fine di verificare la corretta identificazione del perimetro dell'unità complessa e confermare l'idoneità dell'impresa selezionata come unità di rilevazione a rispondere al questionario. In particolare, sono state esaminate in modo accurato le seguenti quattro tipologie di unità complesse: imprese indipendenti, gruppi con una sola impresa attiva, gruppi con vertici eleggibili e gruppi con vertici non eleggibili. Questi ultimi riguardano i casi in cui al vertice del gruppo sia presente holding finanziaria, persona fisica, ente pubblico o multinazionale estera.

Nel controllare l'idoneità del rispondente selezionato sono state analizzate le strutture di controllo del gruppo seguendo la logica del controllo dall'alto in basso (*top-down*) poiché un livello inferiore della catena di controllo non è sempre idoneo a rispondere per la sua controllante. Nel caso delle multinazionali estere non è possibile seguire questa logica in quanto le affiliate possono rispondere solo per la parte di gruppo che controllano in Italia. Nel caso della frammentazione dei gruppi è stato verificato se la frammentazione proposta in sottogruppi seguiva la logica del macro settore; nei casi in cui non è stato possibile seguire questa logica, e non è stato possibile riunire il gruppo in quanto il vertice era ineleggibile, tale gruppo è stato sostituito o eliminato dalla lista. Inoltre, sempre nel caso dei gruppi frammentati, è stata fatta un'analisi *ad hoc* sui vertici al fine di individuarne la tipologia ed, eventualmente, verificare la possibilità di riunire il gruppo. Per esempio, per quanto riguarda i gruppi di imprese a controllo estero è stata verificata l'idoneità a rispondere dell'impresa selezionata come unità di rilevazione in funzione delle linee di attività del gruppo, inoltre si è cercato di non spezzare i gruppi per i quali le linee di attività erano trasversali o differenziate. Infine, per quanto riguarda i gruppi italiani che presentano la struttura di conglomerati differenziati per macro-settori di attività economica (industriale, utilities, distribuzione, servizi avanzati e intermediazione finanziaria) con al vertice una società di tipo finanziario è stata verificata la correttezza della frammentazione del gruppo in funzione delle linee di attività al fine di consentire alle imprese che hanno un ruolo di holding operative di fornire risposte più coerenti e puntuali.

Per quanto riguarda i domini di studio, questi sono di tipo non standard e riflettono le informazioni raccolte nell'ambito della rilevazione, con particolare riguardo alle linee di attività e la relativa decomposizione per fasi del processo aziendale definite nel paragrafo precedente. Altre informazioni rilevate dal questionario, quali la tipologia di unità complessa (impresa indipendente,

gruppo di imprese con unità residenti solo in Italia, gruppo di imprese con unità residenti anche all'estero), l'attività economica prevalente dell'unità complessa nonché la sua dimensione economica (fatturato consolidato interno ed esterno al gruppo e numero complessivo di addetti in Italia e all'estero) saranno utilizzate, opportunamente riclassificate per rispettare la tutela della riservatezza anche in considerazione della dimensione piuttosto limitata del campione, per classificare le informazioni che verranno pubblicate a livello aggregato.

3.4 L'organizzazione della rilevazione

La rilevazione sul campo, di tipo censuario, ha interessato 3.500 imprese selezionate dall'Istat a rappresentare altrettante unità più ampie (unità complesse) potenzialmente caratterizzate da strutture organizzative e processi decisionali complessi.

La peculiare natura della rilevazione, sia da un punto di vista concettuale, metodologico e di modalità di interazione con le imprese (interviste dirette approfondite al top management al fine di garantire la qualità dell'informazione raccolta) ha richiesto la progettazione e implementazione di un impianto organizzativo autonomo rispetto al resto delle rilevazioni campionarie del Cis2011¹²².

La selezione di un fornitore esterno capace di garantire una rete di rilevazione sul territorio formata da intervistatori altamente qualificati sui temi di interesse dell'indagine ha costituito l'elemento qualificante della strategia di rilevazione.

La selezione della società incaricata della raccolta dei dati è stata realizzata sulla base di una gara internazionale che ha richiesto la progettazione di un capitolato tecnico orientato a selezionare un fornitore in grado di operare sull'intero territorio nazionale sulla base di una rete di intervistatori particolarmente qualificati nella raccolta e analisi dei dati relativi alle caratteristiche organizzative e manageriali delle imprese più complesse. Data la natura altamente specialistica dei contenuti analitici proposti, il *target* del questionario in termini di rispondenti è chiaramente rivolto ai manager con responsabilità o funzioni apicali. Infatti, è cruciale individuare degli interlocutori in grado di interagire e supportare la compilazione del questionario con tecnica CAPI.

Al fine di individuare gli intervistatori più idonei a conseguire questo obiettivo la lista di indagine è stata segmentata di tre diversi sotto-insiemi di imprese da intervistare:

- unità complessa di tipo A: principali imprese o gruppi di imprese dell'industria e dei servizi operanti in Italia che, per volume di fatturato e/o numero di addetti, si qualificano tra i principali attori dell'economia nazionale e richiedono la conduzione delle interviste da parte di esperti con una notevole e comprovata competenza professionale nelle tematiche oggetto di indagine. Questo sotto-insieme includeva nel capitolato 75 unità;
- unità complessa di tipo B: imprese o gruppi di imprese che, per volume di fatturato e/o numero di addetti, si qualificano tra i principali attori di uno o più settori industriali o dei servizi a livello nazionale e richiedono la conduzione delle interviste da parte di esperti senior con una significativa e comprovata esperienza professionale nelle tematiche oggetto di indagine. Questo sotto-insieme includeva nel capitolato 325 unità;
- unità complessa di tipo C: imprese o gruppi di imprese che presentano volumi di fatturato e/o un numero di addetti rilevanti ai fini dell'indagine e richiedono la conduzione delle interviste da parte di esperti anche *junior* ma con una significativa e comprovata esperienza professionale e/o di studio specialistico nelle tematiche oggetto di indagine. Questo sotto-insieme includeva nel capitolato le restanti unità.

¹²² Cfr. Capitolo 2 – La Rilevazione sulle unità locali delle imprese e Capitolo 3 – La Rilevazione multiscopo sulle imprese.

Le attività preliminari al lancio della rilevazione hanno riguardato il controllo e la validazione della lista di indagine fornita dall'Istat da parte della società vincitrice dalla gara internazionale (di seguito Rti, acronimo di raggruppamento temporaneo di imprese) e l'attiva partecipazione degli intervistatori all'evento di formazione (tre giorni full time) sugli aspetti tematici (descrizione analitica dei quesiti) e tecnici (utilizzo questionario elettronico) della rilevazione.

I rapporti tra l'Istat e l'Rti sono disciplinati oltre che dal capitolato tecnico incluso nella gara di selezione internazionale anche da un manuale per la conduzione dell'indagine, previsto nel capitolato stesso, che descrive in modo analitico tutte le fasi del processo di conduzione della rilevazione ed il ruolo svolto dall'Istat e dall'Rti per la raccolta dei dati. L'Rti ha inoltre realizzato gli strumenti di lavoro richiesti dal capitolato tecnico della gara ossia l'agenda condivisa ed ha attivato il servizio di segreteria richiesto.

Le attività di raccolta dati sul campo sono iniziate come fase di test a settembre 2014 mentre sono state ufficialmente attivate a partire dagli inizi di novembre 2014. Sono terminate a giugno 2015, termine ultimo per la raccolta dei dati secondo l'informativa trasmessa dall'Istat. Le operazioni sul campo sono proseguite nel mese di luglio ma solo per completare i questionari già inviati all'Istat che sono risultati ancora incompleti. Ad inizio settembre 2015 è stata attivata la procedura sanzionatoria per le imprese non rispondenti. La natura fortemente innovativa della rilevazione ha richiesto numerosi incontri di coordinamento tra l'Istat e l'Rti al fine di gestire le criticità emergenti e monitorare puntualmente la corretta conduzione della rilevazione. In particolare, la complessità del questionario e la difficoltà per l'Rti di organizzare interviste al top management hanno reso necessari numerosi incontri di *debriefing* per assicurare il completamento della rilevazione secondo elevati standard di qualità.

Data la peculiare natura delle informazioni raccolte non è stato possibile adottare tecniche statistiche e strumenti di controllo e correzione. La conduzione delle interviste da parte di intervistatori altamente specializzati sulla base di un questionario elettronico molto avanzato - in grado di garantire la completezza e coerenza formale dei quesiti pur in presenza di percorsi di compilazione molto articolati - ha consentito di raccogliere informazioni di qualità.

L'Istat ha inoltre realizzato un'approfondita valutazione puntuale dei dati raccolti tramite attività di *profiling* dei questionari anche sulla base di informazioni disponibili presso siti aziendali o desumibili dalle relazioni tecniche di bilancio (bilanci di esercizio e consolidati). Come previsto dal capitolato tecnico, l'Rti ha svolto, a fronte delle segnalazioni dell'Istat, i controlli e le verifiche necessarie con la massima collaborazione consentendo di conseguire elevati standard di qualità dei dati.

3.5 In sintesi

La presenza di una struttura organizzativa adeguata a garantire l'efficienza dei processi e a supportare la crescita equilibrata dell'impresa - nonché la possibilità di far leva su una funzione manageriale specializzata e qualificata per progettare e gestire con successo strategie aziendali complesse in termini di innovazione, internazionalizzazione e valorizzazione del capitale umano - rappresentano indubbiamente due dei principali fattori di competitività del segmento più evoluto del sistema produttivo italiano.

La misurazione e l'analisi di questi fenomeni, a partire dall'individuazione della dimensione organizzativa e decisionale rilevante per le imprese, costituiscono un sfida per la statistica ufficiale. L'Istat ha voluto cogliere questa sfida progettando e coordinando la Rilevazione multiscopo sulle unità complesse a partire da un sostanziale cambiamento di approccio rispetto alle tradizionali metodologie di impianto delle rilevazioni statistiche sulle imprese, che risulta poco coerente ed

efficiente nel cogliere tali fenomeni nella loro complessità. L'impresa come unità legale autonoma non individua sempre in modo coerente ed esaustivo i confini dell'impresa, mentre la singola impresa nell'ambito di un gruppo di imprese più complesso non è necessariamente il centro decisionale delle strategie aziendali. Inoltre, i tradizionali schemi di classificazione, quali il settore di attività economica e la dimensione di impresa, forniscono una rappresentazione oramai parziale della complessità aziendale per un segmento rilevante del tessuto produttivo italiano.

Anche sulla base del dibattito in corso a livello europeo sulla definizione delle nuove unità statistiche, l'indagine multiscopo sulle unità complesse è stata attentamente e approfonditamente progettata dall'Istat con l'obiettivo di superare i limiti conoscitivi dei tradizionali schemi di rilevazione.

I principali punti di forza dell'indagine sono costituiti dall'impiego di un approccio più vicino, come concetti e definizioni, all'operatività reale delle organizzazioni complesse e quindi più efficace nella raccolta delle informazioni. In particolare, l'impiego di "griglie" o matrici di classificazione più efficaci e coerenti con la realtà operativa delle imprese - linee di attività e fasi del processo aziendale - ha consentito di mappare in modo approfondito la struttura organizzativa interna e l'insieme di relazioni esterne dell'unità complessa. Inoltre, gli approfondimenti realizzati nell'ambito delle singole sezioni tematiche hanno consentito di analizzare e classificare aspetti particolarmente complessi dell'organizzazione aziendale, attraverso richiesta di informazioni più dirette e focalizzate. Infine l'utilizzo di un questionario elettronico avanzato ha consentito di raccogliere informazioni in modo sempre coerente, agevolando il contributo essenziale fornito dai rilevatori esperti nell'analisi di questi fenomeni e impiegati sul campo nell'intervista diretta al top management dell'unità complessa.

Riferimenti bibliografici

- Regolamento del Consiglio (CEE) n° 2186/93 del 22 luglio 1993, relativo al coordinamento comunitario dello sviluppo di registri di imprese utilizzati a fini statistici.
- Regolamento (CE) n. 177/2008 del Parlamento Europeo e del Consiglio del 20 febbraio 2008 che istituisce un quadro comune per i registri di imprese utilizzati a fini statistici e abroga il Regolamento (CEE) n. 2186/93 del Consiglio.
- Regolamento (CE) n. 696/1993 del Consiglio Europeo del 15 marzo 1993, sulle unità statistiche di osservazione e di analisi del sistema produttivo nella Comunità.
- Regolamento (CE) n. 1893/2006 del Parlamento Europeo e del Consiglio del 20 dicembre 2006 che definisce la classificazione statistica delle attività economiche Nace Revisione 2 e modifica il Regolamento (CEE) n. 3037/90 del Consiglio nonché alcuni regolamenti (CE) relativi a settori statistici specifici.
- Eurostat. 2010. *Business Register Recommendations Manual*. Lussemburgo: methodologies and Working Papers. ISBN 978-92-79-14659-6.
<http://ec.europa.eu/eurostat/ramon/statmanuals/files/KS-32-10-216-EN-C-EN.pdf>
- EU. 2013. *Labour market policy statistics LMPS Methodology 2013*. Eurostat Manuals and guidelines.
- Istat. 2009. *Classificazione delle attività economiche - Ateco 2007*. Metodi e norme n. 40. Roma.
- Abbate, C. 1995. *Una metodologia per la definizione ottimale degli attributi*, in “Verso un sistema statistico integrato delle imprese in Europa”, Franco Angeli, Milano.
- Ambroselli, S. 2011. *Metodologia per l'attribuzione del codice Ateco 2007 - Registro Asia*. Roma: Working Papers, n. 5 Istat.
- Ambroselli S. (2015) *I codici identificativi univoci all'interno del SIM (Sistema Integrato di Microdati)*, [Istat Working Papers 5/2015](#), 15 Aprile 2015
- Cerrori F., Morganti E., *La metodologia e il potenziale informativo dell'archivio sui gruppi di impresa: primi risultati*. Contributi, n. 3. Istat, 2003
- Cerroni F., Di Bella G., Galiè L. 2014. *Evaluating administrative data quality as input of the statistical production process*. Rivista di Statistica Ufficiale, n. 1-2, 2014.
- Calabrese, L. Consalvi, M., *A comparative approach to assess the quality of the economic activity code*, 17th International Roundtable on Business Survey Frames, Rome 2003
- Calzaroni M. (2008) *Le fonti amministrative nei processi e nei prodotti della statistica ufficiale*, Nona Conferenza di Statistica, Sessione 4, Istat, dicembre 2008, <http://www.istat.it/it/files/2011/02/Calzaroni.pdf>
- Consalvi, M. Garofalo, G., 2000. *The real and spurious changes in the Economic Activity Code: methodological treatments and stability criterion*, 14th International Roundtable on Business Survey Frames, Auckland 2000

- Consalvi, M. Fazio, N.R., 2012. *The Business Portal*, 23rd Meeting of the Wiesbaden Group on Business Registers - International Roundtable on Business Survey Frames, Washington, D.C. 17 – 20 September 2012
- Davies & Gather(1993): "The identification of outliers", in Journal of the American statistical association 88: 782-792
- ESSnet on consistency of concepts and methods of business-related statistics - 2010 project on statistical units. *Deliverable 5.3 - Proposals for a revision of the statistical units regulation, Part 1 Proposal for a new definition of the statistical units Enterprise.*
- ESSnet on profiling of large and complex multinational enterprises (MNEs). *Deliverables WP_B - Conceptual framework, methodology rules and standards.*
- Frenda A., Garofalo. G, Morganti E., 2001. *Delineation of European Standards for the inclusion of Enterprise Groups in Business Registers.* 2001. Eurostat Official Website CIRCA.
- Garofalo, G. 2002. *To exploit Administrative sources: a framework of concepts.* Presentato alla 16th Roundtable on Business Survey Frames, Ottobre 2002, Lisbona.
- Garofalo G., 2014, *Il progetto Archimede, obiettivi e risultati sperimentali*, Istat Working Papers n. 9/2014 <http://www.istat.it/it/archivio/140232>
- Garofalo G., Morganti E., 2000. *Relazione finale del Gruppo di lavoro per la Progettazione di un archivio statistico sui gruppi d'impresa*, Settembre 2000.
- Garofalo G., Porcu M., Viviano C. 2003. *Lo Stato di Attività delle Imprese. L'impiego delle fonti amministrative per l'adattamento di un modello probabilistico*, in Quaderni di ricerca Istat n. 3/2003, e nel volume "L'integrazione di dati di fonti diverse", n. 365.397, Franco Angeli, 2005
- Kuhnt, S., (2004), "Outlier identification procedures for contingency tables using maximum likelihood and L1 estimates", in Scandinavian Journal of Statistics 31:431-442.
- Martini A., Sisti M. 2009. *Valutare il successo delle politiche pubbliche.* Il Mulino, Bologna.
- Viviano C., 1997. *The determination of the state of activity of a statistical unit: a probabilistic approach*, in 'Proceedings of session 51th', Bulletin of ISI – August 18-19, Istanbul, 1997, pp. 625-626

Elenco allegati disponibili in formato digitale

- Quadro normativo di riferimento e disposizioni tecniche
- Newsletter
- Questionari e materiali per le rilevazioni
- *Frequently Asked Questions*