

I Tartufi n. 22
Aprile 2006

**GIOVE: UN DATABASE STATISTICO SUL MERCATO
DEL LAVORO VENETO**
Costruzione e indicazioni per l'utilizzo

di Danilo Maurizio



VENETO LAVORO

www.venetolavoro.it

Via Ca' Marcello, 67 - 30172 Venezia Mestre VE
tel. +39.041.29.19.311, fax +39.041.29.19.312
osservatorio.mdl@venetolavoro.it

Sommario

1.	Introduzione	5
2.	Storia e caratteristiche del Silvr	7
2.1.	I flussi informativi	7
2.2.	L'architettura di Netlabor 4.1	9
2.3.	Una tabella chiave: i rapporti di lavoro	10
3.	Giove: il percorso di sviluppo.....	11
3.1.	Stage 1: l'estrazione dei dati.....	11
3.2.	Stage 2: le correzioni	12
3.3.	Ricostruzione delle storie lavorative	14
4.	Authoring di Giove	16
4.1.	L'anagrafica dei lavoratori (DBS_LAVORATORI)	17
4.2.	L'anagrafica delle aziende (DBS_AZIENDE)	17
4.3.	L'anagrafica delle agenzie (DBS_AGENZIE)	17
4.4.	La tabella dei rapporti di lavoro (DBS_RAPLAV)	18
4.5.	Le tabelle di appoggio (DBS_TB)	18
5.	Esempi di utilizzo	20
5.1.	Strumenti di esportazione dei dati.	20
5.2.	Determinare il tasso di conferma per un gruppo di lavoratori interinali.....	21
5.3.	Determinare lo stock di lavoratori da netlabor.	23
5.4.	Un'analisi dei flussi.....	25
Appendice A		
	Il tracciato record delle strutture sorgenti: le tabelle di Netlabor	27
Appendice B		
	Il tracciato record delle strutture di destinazione: le tabelle di Giove	33

1. Introduzione

A partire dal 1996, l'*Agenzia del lavoro del Veneto* e, dopo il passaggio delle competenze alle Regioni, l'area Osservatorio & Ricerche di *Veneto lavoro* hanno dedicato ampia e continua attenzione all'elaborazione e all'analisi dei microdati raccolti dai Centri per l'impiego nell'ambito della loro attività e informatizzati mediante le varie versioni del software Netlabor. Questi microdati sono alla base dell'attività di ricerca sviluppata dall'Osservatorio e dalla struttura pre-esistente nel corso di ormai 10 anni nell'ambito del Progetto "Val.net"¹, inteso a valorizzare a fini conoscitivi le informazioni amministrative informatizzate con Netlabor: numerosi sono i temi trattati e si può dire completo il saggio offerto delle potenzialità insite nell'uso della fonte amministrativa in oggetto. Non è questo il luogo per passare in rassegna pregi e limiti delle principali fonti di dati sul mercato del lavoro presenti nel nostro Paese, ma va comunque sottolineato che grazie all'esperienza maturata in Veneto, e non solo, il "sistema Netlabor" è riuscito a ritagliarsi uno spazio proprio, per originalità e complementarità, indicando numerose direzioni fertili per l'indagine, la ricerca e l'informazione statistica sul mercato del lavoro.

L'Osservatorio di *Veneto lavoro* intende non essere l'unico e perciò privilegiato utilizzatore dei dati Netlabor. Intende anzi favorire l'apertura e l'utilizzo della gran mole di dati amministrativi che affluiscono ai Centri per l'impiego (Cpi), raccogliendo in tal modo un'istanza che il mondo della ricerca manifesta con particolare intensità negli ultimi anni, in particolare da quando pone maggior attenzione alla valutazione delle politiche pubbliche (benchmarking) e ha la necessità di disporre di informazioni adeguate e facilmente accessibili. L'azione di *Veneto lavoro* è favorita dalla recente disponibilità (dal 1 ottobre 2004) di un nuovo testo unico sulla *privacy* che regola in maniera specifica l'accesso, per motivi di ricerca, ad informazioni registrate da enti pubblici per gestire la propria attività amministrativa. Si può dire che il progetto di *Veneto lavoro* è, in definitiva, un progetto di "democratizzazione del dato" – che comincia ad allinearci ad altre avanzate esperienze europee (inglesi, svedesi e danesi) – guidato da questi obiettivi:

- rendere il più possibile trasparente e facile, per scopi scientifici, l'accesso alle informazioni contenute negli archivi dei Cpi;
- diminuire l'autoreferenzialità del Sistema informativo del lavoro (e delle basi dati amministrative in genere), rompendo il monopolio nell'uso delle informazioni attualmente detenuto dalla Pubblica amministrazione;
- valorizzare, in tal modo, il contenuto distintivo che il dato amministrativo porta con sé;
- aumentare il numero di reimpieghi delle informazioni secondo esigenze e linee di ricerca che il sistema dei servizi pubblici non riesce ad intercettare o non è in grado, per diversi motivi, di portare ad un giusto livello di attenzione.

1. Una bibliografia completa delle ricerche riconducibili al "Progetto Val.net" (Valorizzazione Netlabor) è disponibile nel sito di Veneto Lavoro (www.venetolavoro.it).

Il Db statistico che abbiamo battezzato Giove² è il risultato di un'attività di correzione/integrazione dei dati amministrativi estratti dalle banche dati in gestione presso i Cpi del Veneto. Giove è utilizzato dai ricercatori dell'Osservatorio ed è disponibile ai ricercatori esterni – come Public Use File (PUF) – secondo le procedure e le modalità indicate nel sito di *Veneto lavoro*.³

I dati raccolti dai Cpi sono in massima parte generati dal flusso di comunicazioni cui le imprese sono tenute per obbligo di legge (assunzioni, cessazioni e trasformazioni) e in misura minore dal flusso di informazioni raccolte quando il lavoratore in cerca di occupazione spontaneamente si presenta agli sportelli dei servizi per l'impiego.

L'area del mercato del lavoro esplorabile usando le versioni di Giove finora rilasciate (prima edizione: 2004; seconda edizione: 2005; è in preparazione la terza edizione, quella del 2006, prevista per giugno/luglio) è circoscritta dalle informazioni derivanti dalle comunicazioni delle aziende. Per diversi ordini di motivi è stato più semplice, nella prima fase, “aprire” queste informazioni, per le quali sono state già saggiate tutte le potenzialità e tutti i limiti (si veda la letteratura disponibile): in sostanza possono essere analizzate, con diversi livelli di dettaglio, tutte le informazioni relative ai rapporti di lavoro che le aziende venete hanno instaurato nel corso dell'ultimo decennio. Nei prossimi rilasci l'offerta di informazioni sarà ampliata anche alla disoccupazione amministrativa e ad altri temi importanti per l'analisi del mercato del lavoro quali l'accesso agli ammortizzatori sociali (liste di mobilità), la condizione di disabilità etc..

Scopo del presente documento è:

- descrivere i flussi informativi che alimentano gli archivi amministrativi che stanno alla base dell'esperienza che qui resocontiamo;
- documentare quali problemi sono stati affrontati nella realizzazione di Giove;
- fornire indicazioni su come utilizzare il database Giove, grazie anche ad una serie di esempi.

Pertanto ci soffermeremo brevemente sulla storia e sulle caratteristiche del Sistema Informativo Lavoro Regionale del Veneto (Silrv) nel paragrafo 2; nel paragrafo 3, passeremo a descrivere le principali attività di editing; quindi nel paragrafo 4 descriveremo la struttura e le caratteristiche del Public Use File Giove; infine nel paragrafo 5 verranno presentati alcuni esempi di utilizzo.

Si è cercato di mantenere uno stile espositivo il più possibile piano e comprensibile anche a quanti non possiedono conoscenze sulla gestione dei database e la manipolazione dei dati, rinviando alle appendici per una serie di informazioni di carattere più tecnico. L'appendice A riporta il tracciato record delle strutture sorgenti, vale a dire le tabelle di Netlabor; l'appendice B il tracciato record delle strutture di destinazione, vale a dire le tabelle di Giove. La documentazione rilasciata, di natura piuttosto tecnica, ha come prerequisito una buona conoscenza del linguaggio SQL, in particolare nella versione/dialetto implementata dal database opensource MySQL⁴. Per

2. Il primo database statistico costruito nel 2002 con i dati di quattro province è stato sviluppato a partire dagli archivi del Netlabor 1 e perciò denominato Nettuno2002; è sembrato di buon auspicio continuare a denominare i database frutto della nostra attività con il nome di un altro pianeta, più vicino alla terra.

3. Vedi la sezione appositamente dedicata all'accesso al Public Use File in www.venetolavoro.it.

4. www.mysql.com.

chi non volesse utilizzare questo strumento per interrogare/elaborare i dati del PUF verrà brevemente illustrato come esportare i dati in formato ASCII.

2. Storia e caratteristiche del Silvr

L'informatizzazione dei Cpi è cominciata verso l'inizio degli anni novanta, ai tempi del "vecchio" collocamento, prima del trasferimento delle competenze dal Ministero del Lavoro alle Province. È stato un processo incrementale e per lo più volontaristico cominciato grazie all'impegno, alla volontà e alla professionalità di un nucleo molto motivato di dipendenti delle strutture ministeriali periferiche del Veneto. La prima versione del software (Netlabor1), sviluppato in clipper, è rimasta in esercizio fino a tutto il 2000; dall'inizio 2001 è cominciata la fase di migrazione degli archivi e degli applicativi alla nuova versione sviluppata da Insiel in ambiente client-server. La fase di esportazione/importazione dei dati tra versioni successive è stata spesso trascurata ed ha aumentato il numero di sovrapposizioni, incongruenze e duplicazioni che in Giove, nel limite del possibile, abbiamo cercato di superare. Il nuovo software di gestione del Silvr (Netlabor4) è entrato a regime verso la fine del 2002 ed è stato successivamente interessato da numerose modifiche per rispondere ai mutamenti intervenuti nella normativa in vigore⁵.

2.1 I flussi informativi

Netlabor è il software per la gestione delle procedure previste dalla legislazione italiana in materia di mercato del lavoro. Esso permette di registrare in appositi archivi le informazioni individuali relative ai lavoratori e alle aziende che derivano da:

- a. le comunicazioni che le aziende sono obbligate a rendere in base alla normativa vigente ogni volta che un rapporto di lavoro viene avviato, cessato o trasformato;
- b. le autocandidature dei soggetti che dichiarano la propria disponibilità al lavoro.

L'universo di riferimento dell'archivio è rappresentato perciò, limitatamente al periodo coperto dall'informatizzazione, da:

1. i lavoratori, oggetto di almeno una comunicazione obbligatoria da parte delle aziende venete presso le quali lavorano o hanno lavorato, e i soggetti che rendono agli Spi la loro dichiarazione di disponibilità (gli ex iscritti al collocamento), siano essi persone in cerca di occupazione (distinti sulla base di criteri amministrativi secondo diversi "stati" di disoccupazione) o occupati in cerca di ricollocazione professionale;

5. Nel corso degli anni la normativa ha reso disponibili nuovi istituti contrattuali (es. interinale, lavoro in somministrazione, contratto di inserimento) e ha rivisto quasi completamente la definizione della disoccupazione amministrativa.

2. le unità produttive del settore privato che danno luogo a modifiche di posizioni di lavoro dipendente⁶: il dominio di osservazione è rappresentato principalmente dalle attività con sede in comuni afferenti ai Cpi del Veneto e, secondariamente, dalle attività localizzate fuori dai confini del Veneto ma che movimentano lavoratori con domicilio/residenza in Veneto;
3. le caratteristiche, le modalità e i tempi della disoccupazione osservabile secondo le modalità amministrative previste dalla normativa (e perciò stesso mutabili nel tempo);
4. i rapporti di lavoro, descritti secondo le loro principali caratteristiche (la qualifica professionale, il ccnl, la tipologia contrattuale e d'orario...).

2.1.1 I soggetti dell'offerta di lavoro

Per quanto riguarda i soggetti dell'offerta di lavoro la modalità principale e più ricca di raccolta delle informazioni è quella connessa alla dichiarazione di disponibilità (ex iscrizione alle liste del collocamento) che avviene tramite un'intervista faccia a faccia condotta dagli operatori dei centri, opportunamente formati. Le informazioni richieste, oltre che di tipo socio-anagrafico, riguardano il possesso di alcuni requisiti documentati (titoli di studio, qualifiche professionali acquisite, eventuali esperienze lavorative condotte) o dichiarati (conoscenze linguistiche e informatiche, con indicazione del relativo grado di padronanza e modalità di acquisizione), l'esplicitazione della propria disponibilità e l'indicazioni delle personali propensioni rispetto al lavoro.

Un secondo canale rilevante di raccolta delle informazioni è garantito dalle comunicazioni delle aziende. In effetti, fino a tutti gli anni '90 e coerentemente con la logica che informava l'"avviamento" al lavoro da parte degli uffici pubblici, prima dell'assunzione presso un'impresa era d'obbligo l'iscrizione al collocamento, un passaggio burocratico (di recente abolito in una prospettiva di semplificazione) che garantiva comunque una modalità "ricca" di raccolta delle informazioni individuali. Oggi, così, capita sempre più di frequente che l'ingresso di un individuo nell'anagrafica del sistema possa essere guidato sempre e solo dalle comunicazioni aziendali.⁷ Le informazioni raccolte sono di norma registrate nel sistema gestionale del Cpi dove ha sede l'azienda, se il lavoratore risiede o è domiciliato in un comune di pertinenza di un altro Cpi, la comunicazione viene trasmessa ed archiviata anche nel sistema di quest'ultimo (dando così luogo ad una cosiddetta "migrazione").⁸

2.1.2 Le aziende

Per quanto riguarda le aziende, classificate sulla base dell'attività economica (la classificazione utilizzata è Ateco91), l'organizzazione del database prevede la costituzione di un unico archivio in

6. Sono quindi esclusi il lavoro autonomo e quello nella Pubblica Amministrazione (ad eccezione delle mansioni di basso livello per le quali è previsto il passaggio attraverso le graduatorie predisposte dai Cpi).

7. In passato ciò riguardava solo quei lavoratori non iscritti e/o non residenti, "movimentati" comunque da aziende che ricadevano nel territorio di riferimento di un Cpi.

8. Queste, come vedremo meglio oltre, costituiscono fattori di ridondanza del database regionale che possono tuttavia essere sfruttati per le attività di editing e controllo della qualità dei dati.

cui sono comprese sia le imprese (a livello di singola unità locale) localizzate nel territorio di pertinenza di ogni Cpi, sia quelle “esterne” (ricadenti in altri Cpi) purché abbiano movimentato lavoratori afferenti a quel Cpi. Larga parte delle informazioni relative alle imprese registrate negli archivi di Netlabor sono originate dalle comunicazioni obbligatorie; non è quindi garantita la copertura dell'intero universo delle unità attive presenti, in quanto risulteranno sempre escluse dall'archivio le imprese senza addetti e quelle che, a partire dalla fase di avvio del sistema informatizzato, non abbiano segnalato alcuna modifica delle posizioni di lavoro dei propri dipendenti. In assenza di flussi di informazioni che diano vita ad una vera e propria gestione anagrafica, non è possibile disporre di un quadro sulla nati/mortalità delle imprese.

2.2 L'architettura di Netlabor 4.1

La professionalità e la costante assistenza di un gruppo di operatori dei Cpi ha sostenuto l'attività dell'Osservatorio di *Veneto Lavoro* nel processo di accumulazione delle competenze necessarie all'analisi dei dati Netlabor. In un contesto in cui la maggior parte della documentazione prodotta non era sufficientemente analitica da un punto di vista tecnico e certamente non utile ad un approccio statistico, è stato necessario auto-documentare strutture e regole del database. Non è certo questo il luogo per ricostruire nel dettaglio le caratteristiche architetturali del software in gestione presso i Cpi del Veneto; ci si limiterà quindi a considerare soltanto quegli aspetti che maggiormente hanno avuto impatto nella fase di analisi e sviluppo di Giove.

Netlabor è un insieme di basi dati distribuite⁹ e allineate tramite un sistema di replica¹⁰ (cooperazione applicativa). Il meccanismo di replica viene scatenato ogni qual volta un'informazione deve raggiungere, per competenza, anche uffici diversi da quello che l'ha raccolta. Per garantire che ogni singola transazione oggetto di replica raggiunga effettivamente i destinatari è stato scelto un meccanismo di “*queued updating*”. In particolare il meccanismo di replica deve garantire che informazioni anagrafiche e informazioni di flusso siano presenti sia nella base dati del Cpi che riceve la comunicazione dell'azienda, sia nella base dati del Cpi dove il lavoratore ha eletto il suo domicilio.¹¹ Quindi è ragionevole aspettarsi che i dati anagrafici di un lavoratore siano presenti in tutte le basi dati dei Cpi dove è stato chiamato a lavorare¹², così come un'azienda che assume lavoratori domiciliati in territori afferenti a diversi Cpi sarà presente in tutte le loro basi dati.¹³

9. L'architettura originale prevedeva un'istanza del gestionale per ognuno dei 40 Cpi del Veneto più un database regionale come unione delle singole basi dati. Nel 2004/2005 le province di Padova e Verona hanno deciso di eliminare i livelli circoscrizionali e costituire un unico database provinciale. Il database regionale ha svolto correttamente il compito di tenere allineate le informazioni anagrafiche di lavoratori e imprese, ma non si è rivelato capace di mantenere consistenti le informazioni relative ai movimenti dei lavoratori, pregiudicandone l'utilizzo a scopo di monitoraggio e ricerca del mercato del lavoro.

10. Ogni singola transazione che avviene in un database (*transactional replication*) comporta la copia di dati secondo il modello “*publisher to the subscriber(s)*”.

11. Ovviamente se c'è identità tra Cpi di azienda e lavoratore le informazioni saranno registrate una sola volta.

12. Il caso limite del lavoratore iper-mobile è la presenza nel database di tutti e 40 i Cpi, in realtà il numero massimo riscontrato è 19 Cpi.

13. Il caso limite, la presenza in tutte e 40 le basi dati, è la norma per alcune grosse aziende e per le agenzie di lavoro interinale.

Il meccanismo non sempre ha funzionato correttamente; non riuscendo a smaltire le code di transazione accumulate ha ritardato di molto l'allineamento dei dati tra i vari Cpi, con conseguente perdita di tempestività nell'aggiornamento di informazioni molto importanti quali ad esempio lo stato del lavoratore.

La logica che muove il tutto, al netto dell'*upgrade* tecnologico, è di fatto rimasta inalterata rispetto alla vecchia versione di Netlabor¹⁴, quando l'allineamento delle informazioni era garantito dallo scambio della documentazione cartacea tra i vari Cpi e quindi da una doppia imputazione.¹⁵

Il database di Netlabor 4 è un database formato da oltre 600 tabelle tra dati, dizionari dati (tabelle di appoggio o decodifica) e tabelle di sistema (configurazioni, repliche etc.). Il nucleo principale è formato da circa 20 tabelle, quelle della gerarchia "UCL_*"; le principali tabelle anagrafiche sono quella dei lavoratori (UCL_LAVORATORI) e quella delle aziende (UCL_AZIUNPROD) mentre la tabella che fa da pivot per la gestione di tutte le logiche applicative è invece la tabella dei "movimenti" (UCL_MOVIMENTI), che in gerarchia governa la tabella delle assunzioni (UCL_AVVIAMENTI), la tabella della mobilità (UCL_MOBILITA'), la tabella delle dichiarazioni di disponibilità (UCL_PROVIDIMENTI) etc..

2.3 Una tabella chiave: i rapporti di lavoro

Nel percorso di sviluppo di Giove il primo step da assicurare è stato quello della ricostruzione delle storie lavorative in quanto informazione fondamentale per incastrare episodi di disoccupazione e mobilità (episodi questi che, come abbiamo già detto, saranno oggetto di integrazione nelle prossime versioni di Giove). La ricostruzione delle carriere passa quindi per l'analisi della tabella UCL_AVVIAMENTI che di fatto registra tutti i movimenti di assunzione, cessazione, proroga e trasformazione dei rapporti di lavoro.

La tabella mutua per buona parte la struttura delle vecchie versioni di Netlabor1. Infatti viene scritto un record per ogni comunicazione (le cessazioni sono invece registrate sullo stesso record a chiusura del segmento) e pertanto, nel caso in cui uno stesso rapporto venga più volte prorogato e/o trasformato, esso viene articolato in più record fisici, ciascuno dei quali costituisce un segmento del medesimo rapporto di lavoro. *Il fatto che il sistema gestionale ragioni in termini di comunicazioni e non in termini di "rapporto di lavoro" è una delle maggiori criticità perché impone di ricostruire ogni singolo rapporto¹⁶ a partire da una serie di record per i quali non sempre è garantita coerenza logico/temporale.* Così com'è disegnata la struttura del sistema non garantisce una facile

14. Ci si riferisce alla versione clipper/dbf, versione con la quale il nuovo gestionale mantiene parecchi punti in comune. Le differenze più rilevanti sono legate alla gestione dei rapporti interinali e alla scelta di diversi criteri di classificazione per alcune dimensioni di analisi quali il titolo di studio. È differente anche la gestione delle dichiarazioni di disponibilità e quindi della disoccupazione, informazioni che però le prime versioni di Giove non rendono ancora disponibili.

15. Alcune delle soluzioni che abbiamo scelto nell'attività di sviluppo di Giove possono essere applicate anche agli archivi del vecchio Netlabor: speriamo quindi che possano essere di aiuto per chi con quelle basi dati si dovesse ancora confrontare.

16. A partire da una delle ultime versioni di Netlabor4, Insiel ha cercato di porre rimedio al problema aggiungendo una serie di campi utili alla ricostruzione dei rapporti di lavoro a partire dalla catena di comunicazioni, ma non sufficiente a coprire la totalità dei casi da risolvere.

interrogabilità: non si può ad esempio conoscere la durata di un rapporto di lavoro ed è difficile verificare la coerenza temporale tra diversi rapporti di lavoro.

Altra criticità che presenta la tabella delle assunzioni è quella legata alla gestione dei rapporti di lavoro interinale, dove la triangolazione lavoratore – agenzia di somministrazione – azienda utilizzatrice è gestita duplicando ogni singola comunicazione e registrando i dati dell’agenzia in un record e i dati dell’azienda utilizzatrice in un altro. Anche qui il problema sta nell’interrogabilità di una struttura così disegnata, senza contare il fatto che per diversi ordini di motivi sono moltissimi i casi di coppie di record “orfane” o dell’azienda utilizzatrice o dell’agenzia di lavoro interinale oppure vittime di inversioni logiche tra le due.

3. Giove: il percorso di sviluppo

3.1 Stage 1: l'estrazione dei dati

Dal momento che le tabelle principali oggetto del processo di estrazione/elaborazione non contengono nessuna colonna che permetta di identificare quali record siano variati tra un'estrazione e quella precedente, ogni volta si deve procedere ricostruendo la base dati da zero (*applicazione in rimpiazzo*). Lo sforzo connesso all'estrazione dei dati di ogni singolo Cpi, il trasporto dei dati e la ricostruzione di un'istanza regionale come sommatoria dei singoli database¹⁷ rende onerosa l'attività di estrazione e quindi limita la frequenza¹⁸ di aggiornamento dei dati. L'attività di estrazione dei dati, rigorosamente in formato testo, ha fine quando i dati, opportunamente compressi, vengono trasferiti sul server di elaborazione. Il database risultato della sommatoria¹⁹ dei singoli database è la base dalla quale si separano due linee di analisi: una per il trattamento del dato amministrativo e l'altra per il trattamento del dato statistico.

L'utilizzo del dato amministrativo si basa sull'assunzione (forte) che i dati raccolti dal gestionale siano corretti da un punto di vista qualitativo; ha come limite il fatto che ci si deve accontentare del “modello dati” scelto da chi sviluppa il software e della reale qualità dei dati.

Il forte numero di incongruenze, duplicazioni e sovrapposizioni impone una fase di editing sui dati sia per migliorarne la qualità sia per giungere ad un “modello dati” che risponda alle istanze dell'attività di ricerca e che sia in grado di servire gli output di cui la ricerca ha bisogno. Un database statistico come Giove ha bisogno di lunghi tempi di sviluppo, necessita di specifiche professionalità e di molta attività di coordinamento tra i soggetti che, a diverso titolo, sono coinvolti nella conduzione di un sistema complesso come il Sistema Informativo Lavoro.

17. In totale vengono trasferiti dati per circa 10Gb non compressi.

18. Per le necessità della ricerca abbiamo ritenuto sufficienti due estrazioni annuali, una in primavera ed una in autunno.

19. La tabella con la cardinalità più alta è la UCL_MOVIMENTI con oltre 15 milioni di record, la tabella più “grossa” è la tabella UCL_AVVIAMENTI che occupa, una volta indicizzata, circa 3GB di spazio su disco.

3.2 Stage 2: le correzioni

L'attività di costruzione di Giove passa attraverso un processo circolare tra le domande di ricerca e i possibili output. Partendo dalla definizione dei requisiti, attraverso l'individuazione delle principali criticità si giunge all'analisi e alla definizione del "modello dati". *In Giove il rapporto di lavoro, e non la comunicazione, è il vertice dei processi di elaborazione.* Sulla base di questo flusso (finito? infinito?) si giunge allo sviluppo delle routine di correzione secondo le seguenti linee di intervento:

1. predisposizione di una lista di codici fiscali univoca;
2. soluzione del problema della duplicazione dei dati;
3. ricostruzione dei rapporti di lavoro;
4. gestione della triangolazione nei rapporti di lavoro interinale;
5. ricostruzione delle storie lavorative e gestione dell'overlapping tra rapporti di lavoro.

3.2.1 Predisposizione di una lista di codici fiscali univoca

Il primo dei problemi affrontato è stato quello di individuare una lista univoca di soggetti che abbiano avuto almeno un rapporto di lavoro dall'informatizzazione ad oggi. Esistono, infatti, individui presenti nel sistema con due o più codici fiscali; la maggior parte di queste duplicazioni è dovuta:

- a problemi legati al rilascio di nuovi codici da parte del Ministero delle Finanze;²⁰
- alla mancata manutenzione della tabella dei codici comunali (UCL_TBCOMUNI) che sta alla base dell'algoritmo che genera i codici fiscali;²¹
- ad errori di imputazione da parte degli operatori.

L'approccio scelto risolve solo i casi in cui esiste identità tra i primi 11 caratteri del codice fiscale, e non tutti quei casi in cui ad essere sbagliata è un'altra porzione dell'identificativo²²; questa è infatti l'unica correzione che si può applicare senza incorrere nel rischio di accorpare soggetti effettivamente diversi.

Quando è risultato chiaramente che le diverse codifiche appartenevano al medesimo lavoratore abbiamo proceduto aggregando i due o più identificativi in quello corretto, altrimenti abbiamo solo segnalato il problema mantenendo distinti gli individui.

3.2.2 Soluzione del problema della duplicazione dei dati

Risolto il problema dell'identificazione dei lavoratori, processando un lavoratore alla volta abbiamo proceduto creando una struttura definita "magma" (l'idea è quella di farlo raffreddare lentamente favorendo la formazione di cristalli). Questa struttura intermedia raccoglie tutte le informazioni

20. Il caso numericamente più rilevante coinvolge i cittadini di Mestre-Venezia che hanno visto modificare gli ultimi 5 caratteri (codice del comune da F159 a L736 e il check-digit di conseguenza) del loro codice fiscale.

21. I casi più significativi sono quelli legati al codice comune di nascita generato per i cittadini della ex Jugoslavia.

22. I primi sei caratteri, generati dal nome e cognome del lavoratore (in particolare se oggetto di translitterazione da lingue in alfabeto non latino o in presenza di nomi composti), o le posizioni tra il 7° e l'11° carattere per quanto riguarda la data di nascita.

utili alla ricostruzione della catena di rapporti di lavoro che un soggetto ha avuto, siano esse relative alla anagrafica del soggetto o a quella dell'azienda che lo ha assunto o ancora al rapporto di lavoro che è stato instaurato.

Essendo quindi il risultato della sommatoria di tutte le informazioni registrate nel sistema per un dato soggetto, sono possibili tre tipi di duplicazioni²³ (moltiplicazioni):

1. duplicazioni di natura applicativa (gestione del lavoro interinale);
2. duplicazioni dovute all'architettura del sistema (migrazioni/repliche);
3. duplicazioni legate all'attività di *data entry* (doppie imputazioni).

Netlabor 4.1 per ogni comunicazione di assunzione/proroga/trasformazione di un rapporto interinale scrive due record, uno con le informazioni dell'azienda utilizzatrice ed uno con i dati dell'agenzia di somministrazione. Come già evidenziato, una struttura del genere non è di facile interrogazione e quindi abbiamo scelto di riportare l'identificativo dell'agenzia di lavoro interinale in coda al record con le informazioni dell'azienda utilizzatrice, mutuando di fatto la stessa struttura già in uso nelle varie declinazioni del vecchio Netlabor. Quest'attività è stata complicata da due ordini di problemi: i. l'inversione logica delle informazioni tra utilizzatrice e agenzia; ii. la presenza di rapporti "orfani" di uno dei due soggetti della triangolazione lavoratore/agenzia/impresa utilizzatrice.

L'accoppiamento dei record è stato fatto a partire dal campo *mov_indcanc* della tabella dei movimenti (UCL_MOVIMENTI) che assume i valori²⁴ "0" quando registra i dati dell'agenzia e "3" quando registra i dati dell'impresa utilizzatrice.

Nel caso di parziale abbinamento abbiamo usato le informazioni di un campo testuale, il campo *avv_luogolav* della tabella delle assunzioni, retaggio di un vecchio campo del precedente gestionale che tiene memoria dell'identificativo dell'azienda utilizzatrice, che in alcuni casi ci ha permesso di risolvere il problema dei rapporti di lavoro di somministrazione "orfani". Abbiamo tenuto in sospenso i record non abbinati fino alla fine del processo, quando un'opportuna procedura decide se "promuovere" i casi dubbi a rapporti di lavoro interinali o "retrocedere" gli stessi a rapporti di lavoro normali.²⁵

Le duplicazioni di natura architettureale o legate all'attività di *data-entry* sono state risolte attraverso l'utilizzo di funzioni ricorsive, cancellando i record duplicati con diversi livelli di filtro. Si è sempre e comunque cercato di mantenere il record con le informazioni più complete e di recuperare dal record cancellato eventuali informazioni importanti, quali ad esempio la data di cessazione del rapporto di lavoro.

Il filtro fondamentale è l'individuazione dei record interni o esterni. Interno è il record registrato nel Cpi che riceve la comunicazione da parte dell'azienda; esterno è invece il record generato dal meccanismo di replica sulla base della proprietà/competenza del lavoratore. Abbiamo sempre e

23. I tre casi si possono ovviamente combinare.

24. Il campo *mov_indcanc* oltre a gestire le logiche di registrazione dei rapporti di lavoro interinale è anche l'indicatore di cancellazione logica per tutti i record dei movimenti di Netlabor e assume anche i valori 1 e 2 che di fatto servono ad annullare il record così marcato. La scelta di implementare due logiche così cruciali in un solo campo certamente non può essere definita una soluzione "elegante".

25. L'uso di "promuovere" e "retrocedere" non nasconde nessun giudizio di valore.

comunque privilegiato i record interni eccetto in quei casi²⁶ in cui un rapporto di lavoro era registrato solo nel database del Cpi del lavoratore e quindi non duplicato.

3.2.3 Ricostruzione dei rapporti di lavoro

Eliminato il problema delle duplicazioni,²⁷ “raffreddato” in parte il magma iniziale, si passa ad una serie di correzioni minori sugli estremi temporali dei singoli segmenti di un rapporto di lavoro:

1. se un rapporto di lavoro a tempo definito ha indicata una durata prevista ma risulta ancora aperto alla data di estrazione esso viene chiuso sulla base della durata prevista;
2. se un lavoratore alla data di estrazione ha più di un rapporto aperto (chiaramente il ragionamento non vale per chi compone più part-time), il rapporto con data di assunzione più vecchia viene chiuso in modo da rispettare la successione degli eventi (successione che viene ricontrollata in seguito);
3. i rapporti di lavoro con data di assunzione fittizia²⁸ dove possibile sono stati corretti.

3.3 Ricostruzione delle storie lavorative

Come abbiamo già indicato, avendo tra i suoi obiettivi quello di assicurare la computabilità degli stock e il calcolo delle durate, il “modello dati” di Giove ha come vertice del processo di elaborazione il rapporto di lavoro. La ricostruzione delle storie lavorative, quindi, passa attraverso la corretta individuazione dei singoli rapporti di lavoro. Insiel ha messo a punto un meccanismo per ricostruire interamente il rapporto di lavoro a partire dalla catena di eventi/comunicazioni correlati ma questo meccanismo non copre la totalità dei casi da risolvere e risulta negativamente condizionato dai meccanismi della cooperazione applicativa. Vanno identificate correttamente la data di assunzione vera e propria del rapporto di lavoro e la data di cessazione/licenziamento.

La scelta fatta è stata quella di riportare le informazioni essenziali, quali la data di avviamento, il tipo di contratto alla nascita e la tipologia di assunzione del record di origine (padre) in tutti i record successivi (figli) fino all'ultimo dei record che formano un unico rapporto di lavoro; una serie di puntatori garantisce la possibilità di risalire ai vari stati intermedi.

26. Molto probabilmente questi rapporti di lavoro non sono stati generati da una comunicazione obbligatoria ma sulla base di documentazione presentata dal lavoratore durante i colloqui legati alla dichiarazione di disponibilità.

27. Il gestionale nel suo complesso, in virtù del suo disegno decentrato, oggi (al momento in cui è stato realizzato Giove 2005) gestisce un totale di 8.200.000 record di comunicazioni a fronte di 6.800.000 con un vero contenuto informativo. Il risultato della nostra attività di correzione ha infatti ridotto di circa 1.500.000 record il peso della tabella delle assunzioni, quasi il 18% del totale; è una massa notevole di dati che, se pur inutili, devono essere gestiti determinando la crescita di complessità di tutto il sistema.

28. Soprattutto in passato è stata una consuetudine comune tra gli operatori quella di imputare una data di assunzione arbitraria in sostituzione del valore mancante (per esempio laddove pervenivano comunicazioni di cessazione di rapporti di lavoro non presenti in archivio). Chiaramente queste date creano spesso problemi di varia natura e il loro effetto in Giove è stato quasi completamente risolto. Occorre comunque assicurarsi di trattare con la dovuta cautela le informazioni di durata per i rapporti di lavoro che hanno una data di inizio del tipo “01-01-[1949,1950,1980]”.

I record di una catena vengono numerati da 0 a n (dove n è il numero di record -1 che formano ogni singolo episodio di occupazione) e nei record figli è sempre riportato l'identificativo univoco del record padre della catena. Il record con il maggiore contenuto informativo è l'ultimo, che ricostruisce l'intero rapporto ed è identificato da un apposito flag (rap_flu) che assume il valore 1. Per i rapporti di lavoro che non sono stati oggetto né di trasformazioni né di proroghe è chiaramente presente un solo record, con il flag rap_flu posto uguale a 1.

3.3.1 La soluzione dei problemi di overlapping tra rapporti di lavoro

Altro passaggio chiave per arrivare ad un database statistico consistente è la soluzione del problema dell'*overlapping* tra i diversi rapporti di lavoro. Esistono, infatti, moltissimi soggetti (quasi il 10% del totale) che hanno rapporti di lavoro che si sovrappongono. I fattori che generano questo tipo di incongruenze sono legati a:

- la qualità delle comunicazioni obbligatorie;
- le operazioni di migrazione dati tra le varie versioni del gestionale;
- la qualità del data-entry;
- la carenza di controlli messi a disposizione dai gestionali e la particolare struttura scelta per la tabella dei rapporti di lavoro.

Chiaramente è legittimo che si sovrappongano solo i rapporti a part-time, i quali non sono quindi stati fatto oggetto di pulizia. La casistica osservata è quanto mai ampia ed una rassegna non può che limitarsi ad esporre le situazioni più frequenti:²⁹

- tutti quei rapporti che si sovrappongono per un solo giorno sono stati risolti "banalmente", riportando indietro di un giorno la data di cessazione del primo;
- i vecchi rapporti di lavoro rimasti aperti (ovvero quelli per i quali non è stata comunicata/imputata la data di cessazione), che attraversano tutto il periodo e vanno a sovrapporsi con tutte le eventuali successive esperienze di impiego, sono stati chiusi secondo due criteri alternativi: o la mediana della durata osservata nella distribuzione per quella tipologia contrattuale o, se minore, la distanza dal successivo rapporto di lavoro;
- i rapporti di lavoro con durate negative³⁰ che si sovrappongono con altri rapporti di lavoro sono stati eliminati;
- nel caso di rapporti che uno stesso soggetto risulta aver avuto con la medesima azienda e che si sovrappongono per oltre il 60% del tempo è stato scelto di eliminare uno dei due rapporti;
- nel caso di rapporti di lavoro di uno stesso soggetto con la medesima impresa che si sovrappongono per meno del 60% ma che terminano o iniziano lo stesso giorno, è stato scelto di eliminare quello più breve dei due rapporti;

29. Alcune delle scelte fatte in questo come in altri casi sono il frutto di convenzioni assunte come ragionevoli, non necessariamente condivisibili ma sempre guidate dal buon senso.

30. Se la data di cessazione è antecedente la data di assunzione, spesso per effetto di uno scambio tra record/segmenti afferenti al medesimo rapporto, la durata calcolata è ovviamente negativa.

- nel caso di rapporti di lavoro con data di avviamento fittizia e data di cessazione in contraddizione con un successivo rapporto, si è proceduto con l'eliminazione;
- se tutti questi criteri falliscono, ovvero sono ancora presenti dei rapporti di lavoro che si sovrappongono, viene cancellato un record secondo criteri meno stringenti;
- se un singolo rapporto di lavoro, per risolvere problemi di coerenza, è stato oggetto di due o più correzioni si è proceduto alla sua eliminazione.³¹

3.3.2. Una questione specifica: i rapporti di lavoro interinale “dubbi”

Ultimo passaggio per la messa in qualità dei dati di Netlabor è quello legato ai rapporti di lavoro interinali dubbi, ovvero orfani di uno dei soggetti della triangolazione o mal codificati all'atto dell'inserimento della comunicazione di assunzione. Sono state prese in considerazione 4 variabili di controllo:

- il tipo di avviamento codificato nel campo *avv_codavv* della tabella UCL_AVVIAMENTI;
- il tipo di contratto codificato nel campo *avv_codrap* della tabella UCL_AVVIAMENTI;
- la corretta identificazione di una agenzia di somministrazione;
- la presenza di un'impresa utilizzatrice.

Se almeno tre delle informazioni suddette sono coerenti allora il rapporto di lavoro è stato considerato interinale; se c'è coerenza per quanto riguarda il tipo di avviamento ed è identificabile un'agenzia si è provveduto a promuovere a rapporti di lavoro interinale i contratti a tempo determinato e a retrocedere a rapporti di lavoro “normale” tutti gli altri.

4. **Authoring di Giove**

Terminata la fase di pulizia si possono ritenere raggiunti i seguenti obiettivi:

- consolidamento di una lista di soggetti attivi nel mercato del lavoro veneto nell'ultimo decennio;
- eliminazione delle duplicazioni di qualsiasi natura;
- ricostruzione dei rapporti di lavoro e quindi delle carriere osservabili da Netlabor;
- risoluzione del problema dell'overlapping tra i rapporti di lavoro;
- risoluzione dei problemi legati alla codifica dei rapporti di lavoro interinale.

Il semilavorato fin qui ottenuto deve essere rifinito e preparato per essere mandato in produzione, ovvero per essere usato ai fini dell'attività di ricerca e per essere anonimizzato in vista dei rilasci all'esterno su richiesta e convenzione.

La fase di *authoring* prevede l'alimentazione delle tre anagrafiche di riferimento (lavoratori, aziende e agenzie), della tabella dei rapporti di lavoro (la tabella principale) e di tutta una serie di tabelle di appoggio.

31. I rapporti di lavoro eliminati sono stati in totale 84.000, quelli corretti e resi coerenti oltre 100.000.

La nomenclatura scelta per le strutture di Giove è del tutto simile a quella scelta da Insiel per il Netlabor: questo è un vantaggio per chi ha già un po' di familiarità con le strutture di questo gestionale; rimane comunque sufficientemente "parlante" e quindi utile anche a chi si avvicina per la prima volta a questo set di informazioni.

4.1 *L'anagrafica dei lavoratori (DBS_LAVORATORI)*

I lavoratori censiti in Giove edizione 2005 sono 1.972.215. Laddove sono stati riscontrati problemi di duplicazione³² è stato valorizzato il campo *lav_gemelli* che ha la sua codifica nella tabella *DBS_TBGMEMELLI*. Nel caso il lavoratore fosse presente in più anagrafiche, ovvero nelle anagrafiche di più Cpi, sono state considerate in prima battuta più attendibili le informazioni registrate nel db del Cpi di domicilio/residenza del lavoratore; nei casi in cui non è stato possibile risalire al Cpi di residenza, in alternativa è stato scelto il record di anagrafica più recente (*lav_dtultagg*). Oltre al Cpi relativo al comune di domicilio viene indicato anche il Cpi da cui sono state estratte le informazioni. Sono ovviamente state riportate anche le informazioni di tipo socio-demografico, compreso il titolo di studio;³³ alcune informazioni mancanti o palesemente errate sono state desunte dal codice fiscale.

4.2 *L'anagrafica delle aziende (DBS_AZIENDE)*

Le unità locali censite sono 405.748, siano esse unità locali di imprese venete o unità locali di imprese esterne che hanno assunto lavoratori domiciliati in Veneto. Nel caso l'azienda/unità locale risulti presente nelle anagrafiche di più Cpi abbiamo considerato più attendibili le informazioni presenti nel database del Cpi di localizzazione dell'azienda; per le imprese localizzate fuori dal Veneto abbiamo preferito il record di anagrafica più recente. Oltre al Cpi di localizzazione dell'impresa è stato indicato anche il Cpi da cui le informazioni di anagrafica sono state estratte. Tutte le informazioni descrittive, quali indirizzo, codice di attività, contratto nazionale applicato etc., quando valorizzate, sono state riportate. Il problema delle duplicazioni/moltiplicazioni delle unità locali è stato affrontato solo in questa fase, e non all'inizio del processo di elaborazione come per i lavoratori gemelli, accorpando unità locali apparentemente diverse sulla base del confronto dell'indirizzo di localizzazione.

4.3 *L'anagrafica delle agenzie (DBS_AGENZIE)*

La tabella delle agenzie, 1.765 in totale tra regionali ed extraregionali, è stata alimentata da un processo del tutto analogo a quello seguito per le aziende. Vengono riportate le stesse informazioni, ad eccezione ovviamente del codice di attività.

32. Si consiglia di verificare per tutti i lavoratori oggetto di estrazione che valori assume il flag relativo e di trattare il dataset secondo le regole che si ritengono opportune.

33. Mancando un'indicazione della data di conseguimento del titolo è stato selezionato il titolo di studio più alto presente per ogni lavoratore.

4.4 *La tabella dei rapporti di lavoro (DBS_RAPLAV)*

La tabella dei rapporti di lavoro è quella nella quale sono registrate le storie lavorative dei soggetti. La tabella contiene 6.814.293 record, pari a 5.749.208 rapporti di lavoro, ognuno dei quali riporta gli identificativi di lavoratore e azienda; nel caso di lavoro interinale (identificabili dal flag rap_inter = 1) è prevista anche la valorizzazione dell'identificativo dell'agenzia di somministrazione.

Un progressivo numerico rap_seqnum identifica ogni singolo record; nel caso un rapporto³⁴ di lavoro risulti formato da più record sono stati valorizzati alcuni campi nel seguente modo:

rap_pat: indica, se valorizzato, a quale record fare riferimento per identificare il record padre;

rap_chain: numera la catena di record che formano un rapporto di lavoro da 0 (il padre) a n-1, dove n è il numero di record;

rap_flu: assume i valori 1 e 0, vale uno se quel record è l'ultimo o l'unico e 0 altrimenti.

In ogni record sono comunque riportate le seguenti informazioni del record padre, quindi relative alla nascita del rapporto di lavoro:

rap_pdtavv: la data di avviamento del rapporto di lavoro;

rap_pcontratto: tipologia contrattuale alla nascita;

rap_ptempo: orario di lavoro alla nascita.

L'ultimo record di ogni catena, o l'unico se il rapporto non è mai stato trasformato/prorogato, ha le seguenti caratteristiche distintive che ne fanno il record con il maggior contenuto informativo: è identificabile dal fatto che il flag rap_flu = 1, riporta le effettive date di inizio e, se valorizzata³⁵, di fine del rapporto, riporta anche sia la tipologia contrattuale che l'orario di lavoro di inizio. Il contratto, l'orario, la qualifica, il ccnl applicato, il livello e la durata prevista sono riportate (se valorizzate) in ogni record e valgono per il periodo/segmento di rapporto compreso tra rap_dtavv e rap_dtcess. Se infine il rapporto di lavoro è stato estratto dal Cpi dove ha sede l'azienda, il campo rap_rec assume il valore "I" (record interno); se invece è stato estratto dal database del Cpi dove ha domicilio il lavoratore assume il valore "E" (record esterno).

4.5 *Le tabelle di appoggio (DBS_TB)*

Le tabelle di appoggio necessarie sono state incluse e sono riconoscibili dal prefisso DBS_TB; nella maggior parte dei casi sono state rispettate le stesse classificazioni in uso nel gestionale, con alcune modifiche a carico della tabella delle cittadinanze oltre ad altri interventi minori.

DBS_TBATECO: classificazione delle attività economiche ATECO91, non del tutto sovrapponibile con la versione Istat perché tronca nel descrittivo e con alcune codifiche anomale.

DBS_TBBENEFICIO: benefici goduti dal datore di lavoro, solo parzialmente utile perché l'informazione è lacunosa, incompleta e distribuita anche su altri campi.

34. I rapporti di lavoro che sono stati oggetto di almeno una trasformazione o proroga sono quasi 800.000.

35. Se il rapporto di lavoro alla data di estrazione risulta ancora aperto, la data di cessazione assume il valore 0000-00-00.

DBS_TBCCNL: contratto nazionale applicato, ha alcune voci ripetute con codifiche/descrizioni leggermente diverse (es: A05C = "ACQUE E BEVANDE GASSATE - Cooperative" e B02C = "BEVANDE E ACQUE GASSATE - Cooperative").

DBS_TBCESS: codice del motivo di cessazione, indica il motivo per cui viene concluso un rapporto di lavoro o un suo segmento.

DBS_TBCITTADINANZE: tabella delle cittadinanze secondo la classificazione del Ministero del Lavoro; è stata modificata rispetto a quella originale per correggere l'indicazione degli Stati appartenenti all'Unione Europea e per arricchirla con alcune informazioni desunte dalla tabella delle cittadinanze ISTAT.

DBS_TBCODAVV: tabella dei codici di assunzione; qualifica la tipologia di assunzione, in particolare per i rapporti di lavoro interinale; contiene anche informazioni circa l'apertura di nuovi segmenti di rapporti in essere. Alcuni rapporti di lavoro di Giove hanno codifiche di questo campo non legate a novazioni contrattuali quando la prima informazione disponibile per un determinato rapporto di lavoro è una trasformazione o una proroga.

DBS_TBCOMUNI: tabella dei codici dei comuni d'Italia secondo la classificazione del Ministero delle Finanze. Viene riportata anche la provincia di appartenenza e il Cpi di appartenenza (nel caso di comuni non Veneti è valorizzato con "EXV").

DBS_TBCONTRATTO: tabella dei contratti, indica la tipologia contrattuale, eccetto nel caso del lavoro interinale che viene gestito con un apposito flag, rap_inter, della tabella dei rapporti di lavoro.

DBS_TBCPI: tabella dei Cpi; ogni Cpi del Veneto ha una label di tre caratteri che lo identifica, nella tabella è riportata anche la provincia. Il campo *Cpi_aggiornato* indica se alla data dell'estrazione³⁶ erano state imputate tutte le comunicazioni obbligatorie giunte dalle aziende.

DBS_TBGEMELLI: indica per ogni lavoratore della tabella dei lavoratori se sono state fatte operazioni di accorpamento o meno.

DBS_TBNORM: completa in parte la tabella dei benefici in quanto indica sulla base di quale norma è stato dato seguito ad un'assunzione. È comunque non completa.

DBS_TBQUALIFICHE: tabella delle qualifiche secondo la classificazione ministeriale.

DBS_TBSESSO: sesso del lavoratore.

DBS_TBTEMPO: indica l'articolazione dell'orario di lavoro applicato. Nel caso di rapporti a part-time si consiglia di verificare com'è valorizzato il campo rap_oreavv della tabella dei rapporti di lavoro.

DBS_TBTITSTUDI: tabella di decodifica per i titoli di studio secondo la classificazione ministeriale.

36. Febbraio 2005.

5. Esempi di utilizzo

Gli esempi di seguito riportati rispondono ad un intento puramente “didattico” e non sono tratti dall’attività concreta di ricerca sul mercato del lavoro; più che il loro contenuto ciò che interessa è indicare alcuni percorsi di trattamento della base dati.

Il primo “how-to” si rivolge principalmente a coloro che per la fase di estrazione e trattamento dei dati non vogliono usare un database server (MySQL in particolare) il secondo ci porterà alla determinazione dei tassi di conferma³⁷ per un gruppo di lavoratori interinali; nel terzo verrà affrontato il tema della generazione dello stock di lavoratori ad una certa data, mentre nell’ultimo calcoleremo il flusso di assunzioni registrato in un determinato sottoinsieme spazio-temporale.

5.1 Strumenti di esportazione dei dati

Fermo restando che il consiglio che avanziamo è quello di usare MySQL in tutte le fasi di lavoro, fino alla determinazione del dataset da analizzare, è utile comunque indicare quali sono le funzioni per l’export dei dati messe a disposizione da MySQL o quali strumenti usare per l’interfacciamento.

Il formato di rilascio di Giove è un dump “FULL SQL”; quindi un unico file contiene tutti gli statement per la creazione e l’alimentazione delle strutture del database. Per il ripristino del database si rimanda alla nutrita sezione di documentazione del sito di MySQL, in particolare <http://dev.mysql.com/doc/>. La maggior parte dei tools³⁸ di amministrazione presenti sul mercato hanno specifiche funzioni per il ripristino dei dati così come da noi rilasciati.

Una volta ripristinato il database si può interagire con Giove attraverso semplici istruzioni in linguaggio SQL. In particolare per estrarre dati in formato testo si fa uso dell’operatore “into outfile” dello statement select come nell’esempio che segue, dove a partire dalla tabelle dei lavoratori – `db_s_lavoratori` – verrà generato un file CSV (comma separated values):

```
SELECT * INTO OUTFILE 'C:\\percorso_al_file\\db_s_lavoratori.csv'  
FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY ''''  
LINES TERMINATED BY '\n'  
FROM db_s_lavoratori;
```

Questo statement può essere ripetuto per ognuna delle tabelle di Giove in modo da renderle disponibili per l’import a qualsiasi altro strumento di trattamento e analisi dei dati.

Ci si può, ovviamente, interfacciare a MySQL anche tramite un apposito driver ODBC, scaricabile dal sito di MySQL, che permette di interagire direttamente con il database per l’import dei dati.

37. Per tasso di conferma intendiamo la percentuale di soggetti assunti direttamente da un’impresa utilizzatrice al termine della missione sul totale delle missioni realizzate.

38. Compresi quelli messi a disposizione da MySQL stesso.

Tutti i gli strumenti di produttività statistica³⁹ mettono a disposizione delle componenti per l'interfacciamento a basi di dati tramite ODBC; nel caso usiate R⁴⁰ fate riferimento ai moduli RODBC e RMySQL per la configurazione, la sintassi ed esempi di utilizzo.

5.2 *Determinare il tasso di conferma per un gruppo di lavoratori interinali*

I passi da seguire per rispondere ad una domanda sui tassi di conferma sono i seguenti:

1. identificare le missioni terminate in un determinato ambito spazio-temporale (nell'esempio useremo i primi sei mesi del 2001, per missioni attivate nel territorio del comune di Padova);
2. identificare i vertici della triangolazione coinvolti ed eventualmente reperire loro caratteristiche nelle tabelle correlate;
3. verificare quali lavoratori hanno avuto ulteriori rapporti di lavoro dopo il termine della missione e per essi verificare con che azienda, contratto, etc..

1. Per identificare le missioni terminate nel primo semestre del 2001 nel comune di Padova bisogna porre una condizione temporale sulla data di cessazione e una geografica a partire dal comune dell'impresa utilizzatrice. In particolare lo statement SQL da passare al server è il seguente:

```
select
rap_azi, rap_lav, rap_pdtavv, rap_dtcess, b.azi_codatt
from dbs_raplav a inner join dbs_aziende b
on a.rap_azi = b.azi_id
where rap_flu = 1
and rap_inter = '1'
and rap_dtcess between "2001-01-01" and "2001-06-30"
and b.azi_comupr = 'G224';
```

La query deve mettere in join la tabella dei rapporti di lavoro e la tabella delle aziende, estrae una parte delle informazioni e non tutte (identificativi di lavoratore e impresa e di questa anche il codice di attività economica oltre agli estremi temporali della missione) e ha le seguenti condizioni di selezione:

- il flag rap_flu deve essere posto a "1" in modo da selezionare il record con il massimo del contenuto informativo (in questo caso il focus è infatti sulla data di cessazione);
- il rap_inter deve essere posto a "1" in modo da identificare solo i rapporti di lavoro interinale;
- la data di cessazione deve essere compresa tra il "2001-01-01" e il "2001-06-30", estremi inclusi;
- il comune di localizzazione dell'impresa, da cui la join per il reperimento del campo azi_comupr, deve essere G224.

39. In alcuni casi tramite moduli acquistabili a parte.

40. Software open-source di analisi statistica reperibile al sito www.r-project.org.

Anteponendo allo statement precedente l'istruzione `create table temp_inter`, si genererà un semilavorato utile al proseguimento dell'esercizio.

2. Dopo aver opportunamente indicizzato il campo `lav_id` del semilavorato `temp_inter` si procede con un'ulteriore join verso la tabella dei lavoratori per prelevare alcune informazioni socio-demografiche potenzialmente utili, quindi:

```
create index rap_lav on temp_inter(rap_lav);
create table temp_inter2
select a.*,lav_sesso, lav_dtnasc
from temp_inter a inner join dbs_lavoratori b
on a.rap_lav = b.lav_id;
```

Anche in questo caso passo attraverso un successivo semilavorato `temp_inter2`.

3. Per verificare quali delle missioni del dataset appena generato sono state seguite da ulteriori rapporti di lavoro si possono seguire diverse strade, una delle quali, di immediata semplicità, è ricorrere all'uso delle subquery⁴¹ (preventivamente vanno opportunamente indicizzati alcuni campi coinvolti al fine di migliorare le prestazioni della query).

```
create index rap_lav on temp_inter2(rap_lav);
create table temp_inter3
select
a.*,
(
select
    rap_seqnum
    from dbs_raplav b
    where b.rap_flu = 1
    and a.rap_lav=b.rap_lav
    and b.rap_pdtavv>a.rap_dtcess
    order by b.rap_pdtavv asc limit 1
)
from temp_inter2 a;
```

La query precedente permette di risalire all'identificativo univoco del rapporto di lavoro immediatamente successivo al termine della missione, se presente.

La seguente semplice query restituisce il numero di missioni non seguite da nessun altro rapporto di lavoro:

```
select count(*) from temp_inter3 where rap_seqnum is null;
```

41. Quest'uso delle subquery ha il solo limite di imporre il ritorno di un solo campo e quindi si è scelto l'identificativo univoco del rapporto di lavoro. Torna il valore NULL quando non riesce a recuperare l'informazione.

Una volta indicizzato il campo rap_seqnum appena reperito si può tornare alla tabella dei rapporti di lavoro e recuperare l'id dell'azienda, il contratto di lavoro alla nascita e/o altre informazioni di interesse.

Quindi l'ultimo dei semilavorati temp_inter4 viene generato dal seguente statement SQL:

```
create index rap_seqnum on temp_inter3(rap_seqnum);
create table temp_inter4
select
a.*,b.rap_pdtavv brap_pdtavv, b.rap_pcontratto brap_pcontratto, b.rap_azi brap_azi,
b.rap_inter brap_inter
from temp_inter3 a
left join dbs_raplav b on a.rap_seqnum = b.rap_seqnum;
```

Il record set è generato e può essere estratto per l'analisi dei dati, una semplice query risponde almeno in parte alla nostra curiosità:

```
select count(*) from temp_inter4 where rap_azi = brap_azi and brap_inter !=1;
```

La query torna il numero di missioni seguite da un rapporto di lavoro con contratto non interinale e attivato dalla medesima impresa utilizzatrice. Una curiosità che lasciamo al lettore è quella di studiare la distribuzione dei tempi di attesa tra la fine della missione e l'inizio del rapporto di lavoro.

5.3 *Determinare lo stock di lavoratori da netlabor*

Nel seguente esempio verrà determinato lo stock di lavoratori attivi al '2003-12-31' nel settore turistico (classe ateco 55) della provincia di Belluno.

I passi da seguire per determinare lo stock appena descritto possono essere i seguenti:

1. identificare tutte le aziende dell'anagrafica di Giove con il vincolo sulla provincia di localizzazione e sul settore economico scelto;
2. identificare i rapporti di lavoro che alla data scelta risultano attivati in una delle aziende appena selezionate;
3. ricavare alcune informazioni del rapporto di lavoro in modo da caratterizzare lo stock estratto.

1. Per identificare le aziende bellunesi del turismo procediamo con la seguente query:

```
create table temp_stock
select
a.azi_id,azi_codatt
from
dbs_aziende a inner join dbs_tbcomuni b
on a.azi_comupr = b.com_cod
where azi_codatt like '5.5%'
and b.com_prov = 'BL';
```

2. Una volta determinato questo set di aziende si identificano i rapporti in essere alla data di stock:

```
create index azi_id on temp_stock(azi_id);
create table temp_stock2
select * from dbs_raplav a
inner join temp_stock b
on a.rap_azi = b.azi_id
where
rap_flu = 1
and '2003-12-31' between rap_pdtavv and if(rap_dtcess = 0,'2005-02-28',rap_dtcess);
```

In particolare lo statement precedente pone come condizione che la data di stock sia compresa tra la data di avviamento e la data di cessazione o la data di chiusura degli archivi se il rapporto non è ancora cessato (quindi con censura a destra).

3. Vanno ora estratte solo una parte di informazioni a partire dal precedente dataste, in particolare il rapporto di lavoro alla nascita, alla cessazione (se presente) e alla vera data di stock. Determinare la tipologia contrattuale alla nascita e alla fine è semplice perché sono informazioni già registrate nel record con il rap_flu = 1; un po' più complicato è recuperare questa informazione se ci si limita puntualmente alla data di stock.

```
create index rap_seqnum on temp_stock2(rap_seqnum);
create index rap_pat on temp_stock2(rap_pat);
create table temp_stock3
select
rap_lav,rap_pcontratto,rap_contratto,rap_chain,rap_dtavv,
case
when rap_chain = 0 then rap_contratto
when rap_chain = 1 and rap_dtavv>'2003-12-31' then rap_pcontratto
when rap_chain = 1 and rap_dtavv<='2003-12-31' then rap_contratto
when rap_chain > 1 then
(
select
rap_contratto
from dbs_raplav b
where a.rap_lav = b.rap_lav
and a.rap_pat = b.rap_pat
and '2003-12-31'
between
b.rap_dtavv
and
if(b.rap_dtcess = 0,"2005-02-28",rap_dtcess)
limit 1
)
end rap_scontratto
from temp_stock2 a;
```


In temp_stock3 si avranno per ogni rapporto di lavoro aperto alla data del “2003-12-31” nel settore turistico della provincia di Belluno le seguenti informazioni sul contratto:

- rap_pcontratto: il contratto di lavoro all’atto dell’assunzione;
- rap_contratto: il contratto al momento della cessazione del rapporto di lavoro o nel caso sia ancora aperto alla data di estrazione dei dati (quindi l’ultimo rilevato);
- rap_scontratto: il contratto di lavoro alla data di stock.

Il costrutto case e la subquery servono proprio a reperire la tipologia contrattuale puntuale per la data di stock. Lasciamo al lettore come esercizio la verifica della variazione dello stock di occupati rispetto al precedente anno.

5.4 Un’analisi dei flussi

Nel seguente esempio ci limiteremo a scegliere alcune dimensioni per l’analisi dei flussi di assunzione attivati da aziende della provincia di Rovigo nel 2002 e nel 2003.

I passi da seguire per arrivare alla produzione del dataset in oggetto sono i seguenti:

1. identificare tutte le aziende dell’anagrafica di Giove con il vincolo sulla provincia di localizzazione scelta;
2. identificare tutti i rapporti di lavoro attivate dalle aziende come precedentemente selezionate e con il vincolo temporale scelto;
3. recuperare alcune informazioni socio-demografiche del lavoratore;
4. recuperare i descrittori per le dimensioni di analisi scelte.

1. Analogamente a quanto visto in precedenza per le imprese bellunesi del turismo procediamo con la seguente query:

```
create table temp_flu
select
a.azi_id,azi_codatt
from
dbs_aziende a inner join dbs_tbcomuni b
on a.azi_comupr = b.com_cod
where b.com_prov = 'RO';
```

2. A partire da queste aziende si recuperano tutti i rapporti di lavoro nati nel periodo preso in esame:

```
create index azi_id on temp_flu(azi_id);
create table temp_flu2
select
b.rap_lav, a.azi_codatt, b.rap_pcontratto,
year(b.rap_pdtavv) anno
from temp_flu a
```

```

inner join dbs_raplav b
on a.azi_id = b.rap_azi
where
rap_flu = 1
and year(rap_pdtavv) between 2002 and 2003;

```

3. Si aggiungono quindi alcune informazioni dall'anagrafica dei lavoratori:

```

create index rap_lav on temp_flu2(rap_lav);
create table temp_flu3
select a.*,b.lav_sesso,b.lav_dtnasc from temp_flu2 a
inner join dbs_lavoratori b
on a.rap_lav = b.lav_id;

```

4. Si indicizzano opportunamente i campi coinvolti in tabelle correlate per il reperimento dei descrittivi, e si arricchisce così il dataset da esportare:

```

create index azi_codatt on temp_flu3(azi_codatt);
create index rap_pcontratto on temp_flu3(rap_pcontratto);
create table temp_flu4
select
a.anno,a.lav_sesso,b.ateco_fulldescriz,c.con_descriz,
case
when a.anno - year(a.lav_dtnasc) <= 30 then 'fino a 30'
when a.anno - year(a.lav_dtnasc) > 30 then 'oltre i 30'
end eta
from temp_flu3 a
inner join dbs_tbateco b
on a.azi_codatt = b.ateco_cod
inner join dbs_tbcontratto c
on a.rap_pcontratto = c.con_cod;

```

Il dataset così ottenuto rappresenta la lista di assunzioni effettuate nella provincia di Rovigo nel 2002 e nel 2003 secondo alcune caratteristiche; per esercizio si consideri l'import del dataset su excel per la produzione di una tabella pivot.

Appendice A

**IL TRACCIATO RECORD DELLE STRUTTURE SORGENTI:
LE TABELLE DI NETLABOR**

Tab. A1 - UCL_AVVIAMENTI

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	AVV_CODFISC	Key Mov - Codice fiscale del lavoratore	CHAR	16	UCL_MOVIMENTI
2	AVV_ANNO	Key Mov - Anno del movimento	NUMERIC	4	UCL_MOVIMENTI
3	AVV_NUMERO	Key Mov - Progressivo del movimento	NUMERIC	6	UCL_MOVIMENTI
4	AVV_PROPRIE_CODCIRC	Key Mov - Codice del Centro per l'Impiego - Proprietà	CHAR	5	UCL_MOVIMENTI
5	AVV_CODCIRC	Key Mov - Codice del Centro per l'Impiego	CHAR	14	UCL_MOVIMENTI
6	AVV_DTAVV	Data di assunzione	DATETIME	0	
7	AVV_NUMPROT1	Numero di protocollo dell'avviamento	NUMERIC	9	
8	AVV_PROPRIE_CODAVV	Codice dell'assunzione - Proprietà	CHAR	5	UCL_TBTIPAVV
9	AVV_CODAVV	Codice dell'assunzione	CHAR	3	UCL_TBTIPAVV
10	AVV_PROPRIE_CODRIS	Codice riservatario - Proprietà	CHAR	5	UCL_TBRISERVA
11	AVV_CODRIS	Codice riservatario	CHAR	1	UCL_TBRISERVA
12	AVV_PROPRIE_CODRAP	Codice del tipo di rapporto - Proprietà	CHAR	5	UCL_TBRAPLAV
13	AVV_CODRAP	Codice del tipo di rapporto	CHAR	2	UCL_TBRAPLAV
14	AVV_PROPRIE_CODNORM	Codice del tipo di normativa - Proprietà	CHAR	5	UCL_TBNORMAVV
15	AVV_CODNORM	Codice del tipo di normativa	NUMERIC	2	UCL_TBNORMAVV
16	AVV_PROPRIE_CODQUA	Key AIL/CFL -Codice della qualifica - Proprietà	CHAR	5	UCL_TBQUALIFICHE
17	AVV_CODQUA	Key AIL/CFL -Codice della qualifica	CHAR	10	UCL_TBQUALIFICHE
18	AVV_PROG_QUALIFICA	Key AIL/CFL -Progressivo della qualifica	NUMERIC	3	UCL_AZICFL_QUALIFICHE
19	AVV_PROPRIE_CODGR	Codice del grado di qualifica - Proprietà	CHAR	5	UCL_TBGRQUAL
20	AVV_CODGR	Codice del grado di qualifica	NUMERIC	2	UCL_TBGRQUAL
21	AVV_PROPRIE_CODCCN	Codice del CCNL - Proprietà	CHAR	5	UCL_TBCCNL
22	AVV_CODCCN	Codice del CCNL	CHAR	4	UCL_TBCCNL
23	AVV_PROPRIE_CODCFL	Codice contratto formazione lavoro - Proprietà	CHAR	5	UCL_TBCFL
24	AVV_CODCFL	Codice contratto formazione lavoro	CHAR	2	UCL_TBCFL
25	AVV_PROPRIE_CODCES	Codice del tipo di cessazione - Proprietà	CHAR	5	UCL_TBCESS
26	AVV_CODCES	Codice del tipo di cessazione	CHAR	2	UCL_TBCESS
27	AVV_PROPRIE_CODPRIO	Codice della domanda di priorità (Art. 9 bis) - Proprietà	CHAR	5	UCL_TBPREC
28	AVV_CODPRIO	Codice della domanda di priorità (Art. 9 bis)	CHAR	2	UCL_TBPREC
29	AVV_PROPRIE_CODNOT	Codice delle note degli avviamenti - Proprietà	CHAR	5	UCL_TBNAVV
30	AVV_CODNOT	Codice delle note degli avviamenti	CHAR	1	UCL_TBNAVV
31	AVV_PROPRIE_CODBEN	Codice del beneficio richiesto - Proprietà	CHAR	5	UCL_TBBEN
32	AVV_CODBEN	Codice del beneficio richiesto	CHAR	2	UCL_TBBEN
33	AVV_PROPRIE_CODAREA	Codice dell'area di lavoro - Proprietà	CHAR	5	UCL_TBAREA
34	AVV_CODAREA	Codice dell'area di lavoro	CHAR	2	UCL_TBAREA
35	AVV_DTAPPO	Data di appoggio	DATETIME	0	
36	AVV_PROPRIE_CFLNOTE	Codice note o ass. relativa al C.F.L. - proprietà	CHAR	5	UCL_TBALBO
37	AVV_CFLNOTE	Codice note o ass. relativa al C.F.L.	CHAR	5	UCL_TBALBO
38	AVV_APPRA3	-- Non utilizzato --	CHAR	1	
39	AVV_DTCESS	Data di cessazione dell'avviamento	DATETIME	0	
40	AVV_DTPRECED	Data di presentazione domanda diritto di precedenza	DATETIME	0	
41	AVV_NUMPRECED	Numero domanda diritto di precedenza	NUMERIC	6	
42	AVV_DTPROT1	Data di protocollo dell'avviamento	DATETIME	0	
43	AVV_NUMPROT2	Numero di protocollo della cessazione	NUMERIC	9	
44	AVV_DTPROT2	Data di protocollo della cessazione	DATETIME	0	
45	AVV_GGAVV	Numero giorni (T.D. & apprendistato)	NUMERIC	4	
46	AVV_MMAVV	Numero mesi (T.D. & apprendistato)	NUMERIC	2	
47	AVV_AA AVV	Numero anni (T.D. & apprendistato)	NUMERIC	2	
48	AVV_ OREAVV	Numero ore (T.D. & apprendistato)	NUMERIC	3	
49	AVV_LIVELLO	Livello	CHAR	5	
50	AVV_AZIFISC	Unità produttiva - Codice fiscale	CHAR	16	UCL_AZIUNPROD
51	AVV_AZIPIVA	Unità produttiva - Partita IVA	CHAR	11	UCL_AZIUNPROD
52	AVV_AZIVER	Unità produttiva - Versione	NUMERIC	3	UCL_AZIUNPROD
53	AVV_AZIPROGR	Unità produttiva - Progressivo	NUMERIC	4	UCL_AZIUNPROD
54	AVV_MATR	Matricola del lavoratore nell'unità produttiva	CHAR	5	
55	AVV_PROPRIE_AILCFLCIRC	Key AIL/CFL - Codice del Centro per l'Impiego - Proprietà	CHAR	5	UCL_AZICFL_QUALIFICHE
56	AVV_AILCFLCIRC	Key AIL/CFL - Codice del Centro per l'Impiego	CHAR	14	UCL_AZICFL_QUALIFICHE
57	AVV_ANNOAIL	Key AIL - Anno	NUMERIC	4	UCL_AZIAIL_QUALIFICHE
58	AVV_NUMAIL	Key AIL - Numero	NUMERIC	6	UCL_AZIAIL_QUALIFICHE
59	AVV_ANNOCFL	Key CFL - Anno	NUMERIC	4	UCL_AZICFL_QUALIFICHE
60	AVV_NUMCFL	Key CFL - Numero	NUMERIC	6	UCL_AZICFL_QUALIFICHE
61	AVV_DTCOMOCE	Data di comodo per la cessazione	DATETIME	0	
62	AVV_DTSPE1	Data di spedizione (Protocollo 1)	DATETIME	0	
63	AVV_DTSPE2	Data di spedizione (Protocollo 2)	DATETIME	0	
64	AVV_DTSPEPREC	Data di presentazione domanda di precedenza (Art. 9 bis)	DATETIME	0	
65	AVV_LUOGOLAV	Descrizione del luogo di lavoro	CHAR	255	
66	AVV_DTFINEPREC	Data fine validità domanda di precedenza (Art. 9 bis)	DATETIME	0	
67	AVV_PROPRIE_NAZESTERO	Codice della nazione per lavoro all'estero - Proprietà	CHAR	5	UCL_TBNAZIONI
68	AVV_NAZESTERO	Codice della nazione per lavoro all'estero	CHAR	4	UCL_TBNAZIONI
69	AVV_DURATAAPPR	Durata prevista del contratto di apprendistato	NUMERIC	2	
70	AVV_MANSIONE	Mansione	VARCHAR	255	
71	AVV_RIF_ANNO	Key Mov originale - Anno	NUMERIC	4	
72	AVV_RIF_NUMERO	Key Mov originale - Numero	NUMERIC	6	
73	AVV_RIF_PROPRIE_CODCIRC	Key Mov originale - Proprietà codice ufficio	CHAR	5	
74	AVV_RIF_CODCIRC	Key Mov originale - Codice ufficio	CHAR	14	
75	AVV_RIF_PROPRIE_TIPO	Proprietà codice tipo trasformazione	CHAR	5	
76	AVV_RIF_TIPO	Codice tipo trasformazione	CHAR	3	
77	AVV_RIFS_ANNO	Key Mov precedente - Anno	NUMERIC	4	
78	AVV_RIFS_NUMERO	Key Mov precedente - Numero	NUMERIC	6	
79	AVV_RIFS_PR_CODCIRC	Key Mov precedente - Proprietà codice ufficio	CHAR	5	
80	AVV_RIFS_CODCIRC	Key Mov precedente - Codice ufficio	CHAR	14	

Tab. A2 - UCL_AZIUNPROD

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	UNP_CODFISC	Unità produttiva - Codice fiscale	CHAR	16	
2	UNP_PIVA	Unità produttiva - Partita IVA	CHAR	11	
3	UNP_VER	Unità produttiva - Versione	NUMERIC	3	
4	UNP_PROGR	Unità produttiva - Progressivo	NUMERIC	4	
5	UNP_RAGSOC	Ragione sociale	VARCHAR	255	
6	UNP_PROPRIE_COMUPR	Codice del comune sede dell'unità produttiva - Proprietà	CHAR	5	UCL_TBCOMUNI
7	UNP_COMUPR	Codice del comune sede dell'unità produttiva	CHAR	4	UCL_TBCOMUNI
8	UNP_INDIRUPR	Indirizzo dell'unità produttiva	VARCHAR	40	
9	UNP_CAPUPR	CAP dell'unità produttiva	CHAR	6	
10	UNP_FRAZUPR	Località dell'unità produttiva	VARCHAR	30	
11	UNP_NTEL	Numero telefonico	VARCHAR	20	
12	UNP_NFAX	Numero di FAX	VARCHAR	20	
13	UNP_MAIL	Indirizzo di e-mail	VARCHAR	60	
14	UNP_PROPRIE_CODATT	Codice del tipo di attività - proprietà	CHAR	5	UCL_TBATE91
15	UNP_CODATT	Codice del tipo di attività	CHAR	9	UCL_TBATE91
16	UNP_NUMUNP	Numero interno dell'unità produttiva	VARCHAR	20	
17	UNP_INPS	Numero I.N.P.S.	VARCHAR	20	
18	UNP_INAIL	Numero I.N.A.I.L.	VARCHAR	20	
19	UNP_ASL	Numero A.S.L.	VARCHAR	20	
20	UNP_ANLT	Numero Albo Nazionale Lavoro Interinale	VARCHAR	20	
21	UNP_ANSI	Numero Albo Nazionale Società Mediazione	VARCHAR	20	
22	UNP_PROPRIE_STATO	Codice dello stato dell'unità produttiva - proprietà	CHAR	5	UCL_TBSTATAZ
23	UNP_STATO	Codice dello stato dell'unità produttiva	NUMERIC	2	UCL_TBSTATAZ
24	UNP_DTSTAT	Data validità dello stato	DATETIME	0	
25	UNP_DTCREAZ	Data di creazione dell'unità produttiva	DATETIME	0	
26	UNP_NDIPDICH	Numero di dipendenti dichiarati	NUMERIC	6	
27	UNP_NOTE	Note	VARCHAR	60	
28	UNP_PROPRIE_CODCIRC	Codice del Centro per l'Impiego - proprietà	CHAR	5	UCL_TBCIRC
29	UNP_CODCIRC	Codice del Centro per l'Impiego	CHAR	14	UCL_TBCIRC
30	UNP_REAPROV	REA - Provincia	CHAR	2	
31	UNP_REANUM	REA - Numero	NUMERIC	9	
32	UNP_REAPROGR	REA - Progressivo	NUMERIC	4	
33	UNP_INSEGNA	Insegna	VARCHAR	50	
34	UNP_PROPRIE_CCNL	Codice del CCNL principale applicato - Proprietà	CHAR	5	UCL_TBCCNL
35	UNP_CCNL	Codice del CCNL principale applicato	CHAR	4	UCL_TBCCNL
36	UNP_CL9	Unità produttiva obbligata a presentare il CL/9	CHAR	1	
37	UNP_DTULTAGG	Data ultima modifica sull'unità produttiva	DATETIME	0	

Tab. A3 - UCL_LAVORATORI

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	LAV_CODFISC	Codice fiscale del lavoratore	CHAR	16	UCL_LAVORATORI
2	LAV_COGNOME	Cognome del lavoratore	VARCHAR	255	UCL_LAVORATORI
3	LAV_NOME	Nome del lavoratore	VARCHAR	255	
4	LAV_SESSO	Sesso	CHAR	1	
5	LAV_DTNASC	Data nascita del lavoratore	DATETIME	0	
6	LAV_PROPRIE_COMNASC	Codice ministeriale del comune di nascita - Proprietà	CHAR	5	UCL_TBCOMUNI
7	LAV_COMNASC	Codice ministeriale del comune di nascita	CHAR	4	UCL_TBCOMUNI
8	LAV_PROPRIE_CODCIT	Codice della cittadinanza - Proprietà	CHAR	5	UCL_TBCITTA
9	LAV_CODCIT	Codice della cittadinanza	CHAR	3	UCL_TBCITTA
10	LAV_PROPRIE_CODCIV	Codice dello stato civile - Proprietà	CHAR	5	UCL_TBSTACIV
11	LAV_CODCIV	Codice dello stato civile	CHAR	2	UCL_TBSTACIV
12	LAV_PROPRIE_RESCOM	Codice del comune di residenza - Proprietà	CHAR	5	UCL_TBCOMUNI
13	LAV_RESCOM	Codice del comune di residenza	CHAR	4	UCL_TBCOMUNI
14	LAV_RESINDIR	Indirizzo di residenza	VARCHAR	60	
15	LAV_RESCAP	Cap della residenza	CHAR	6	
16	LAV_RESFRAZ	Località della residenza	VARCHAR	60	
17	LAV_RESNTEL	Numero di telefono della residenza	VARCHAR	255	
18	LAV_PROPRIE_RESSTATO	Codice dello stato di residenza - Proprietà	CHAR	5	UCL_TBNAZIONI
19	LAV_RESSTATO	Codice dello stato di residenza	VARCHAR	4	UCL_TBNAZIONI
20	LAV_PROPRIE_DOMCOM	Proprietà del codice del comune di domicilio	CHAR	5	UCL_TBCOMUNI
21	LAV_DOMCOM	Comune del domicilio	CHAR	4	UCL_TBCOMUNI
22	LAV_DOMINDIR	Indirizzo del domicilio	VARCHAR	60	
23	LAV_DOMCAP	Cap del domicilio	CHAR	6	
24	LAV_DOMFRAZ	Località del domicilio	VARCHAR	60	
25	LAV_DOMNTEL	Numero telefonico	VARCHAR	255	
26	LAV_PROPRIE_DOMSTATO	Codice dello stato del domicilio - Proprietà	CHAR	5	UCL_TBNAZIONI
27	LAV_DOMSTATO	Codice dello stato del domicilio	VARCHAR	4	UCL_TBNAZIONI
28	LAV_PROPRIE_CODCLA	Codice della classe - Proprietà	CHAR	5	UCL_TBCLASSE
29	LAV_CODCLA	Codice della classe	CHAR	2	UCL_TBCLASSE
30	LAV_NARCHI	Numero o sigla archiviazione	VARCHAR	9	
31	LAV_PROPRIE_CODSTA	Codice dello stato del lavoratore - Proprietà	CHAR	5	UCL_TBSTATO
32	LAV_CODSTA	Codice dello stato del lavoratore	NUMERIC	2	UCL_TBSTATO
33	LAV_PROPRIE_CODCIRC	Codice del Centro per l'Impiego attuale - Proprietà	CHAR	5	UCL_TBCIRC
34	LAV_CODCIRC	Codice del Centro per l'Impiego attuale	CHAR	14	UCL_TBCIRC
35	LAV_L407	Lavoratore inserito nella lista appartenente alla L.407	NUMERIC	1	
36	LAV_VALISCR	Indicatore di validità del codice fiscale	NUMERIC	1	
37	LAV_PROPRIE_CIRCIN	Codice del Centro per l'Impiego di provenienza - Proprietà	CHAR	5	UCL_TBCIRC
38	LAV_CIRCIN	Codice del Centro per l'Impiego di provenienza	CHAR	14	UCL_TBCIRC
39	LAV_DTACCERT	Data accertamento sanitario	DATETIME	0	
40	LAV_LIBDATA	Data rilascio libretto di lavoro	DATETIME	0	
41	LAV_LIBNUM	Numero libretto di lavoro	NUMERIC	6	
42	LAV_LIBCOPIA	Indicatore di libretto originale o copia	NUMERIC	1	
43	LAV_LIBUFF	Tipo ufficio che ha rilasciato il libretto	CHAR	1	
44	LAV_PROPRIE_LIBCODCIRC	Codice ufficio che ha rilasciato il libretto - Proprietà	CHAR	5	UCL_TBCIRC
45	LAV_LIBCODCIRC	Codice ufficio che ha rilasciato il libretto	VARCHAR	14	UCL_TBCIRC
46	LAV_PROPRIE_LIBCODCOM	Codice del comune che ha rilasciato il libretto - Proprietà	CHAR	5	UCL_TBCOMUNI
47	LAV_LIBCODCOM	Codice del comune che ha rilasciato il libretto	VARCHAR	4	UCL_TBCOMUNI
48	LAV_DATSCA	Data scadenza permesso di soggiorno	DATETIME	0	
49	LAV_PROPRIE_INPS	Codice I.N.P.S. - Proprietà	CHAR	5	UCL_TBINPS
50	LAV_INPS	Codice I.N.P.S.	CHAR	6	UCL_TBINPS
51	LAV_USL	Codice A.S.L.	CHAR	2	
52	LAV_PRESSO	Indicazione presso cui abita il lavoratore	VARCHAR	255	
53	LAV_RISERVAT	Indicat.riservatezza dati (0,1)	NUMERIC	1	
54	LAV_FVARST	Ind. variazione stato	NUMERIC	1	
55	LAV_DTULTVAR	Data ultima variazione	DATETIME	0	
56	LAV_TIPOINS	Tipo lav.(0=int.,1=est.caricato da me,2=est. da centrale)	NUMERIC	1	
57	LAV_DTULTMOV	Data ultimo movimento	DATETIME	0	
58	LAV_EMAIL	Indirizzo di e-mail	VARCHAR	60	
59	LAV_DTULTAGG	Data ultimo aggiornamento	DATETIME	0	

Tab. A4 - UCL_MOVIMENTI

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	MOV_CODFISC	Key Mov - Codice fiscale del lavoratore	CHAR	16	UCL_LAVORATORI
2	MOV_ANNO	Key Mov - Anno del movimento	NUMERIC	4	
3	MOV_NUMERO	Key Mov - Numero del movimento	NUMERIC	6	
4	MOV_PROPRIE_CODCIRC	Key Mov - Codice del Centro per l'Impiego - Proprietà	CHAR	5	UCL_TBCIRC
5	MOV_CODCIRC	Key Mov - Codice del Centro per l'Impiego	CHAR	14	UCL_TBCIRC
6	MOV_PROPRIE_TIPO	Codice del tipo di movimento - Proprietà	CHAR	5	UCL_TBTIPOMOV
7	MOV_TIPO	Codice del tipo di movimento	CHAR	2	UCL_TBTIPOMOV
8	MOV_DATAMOV	Data del movimento	DATETIME	0	
9	MOV_INDCANC	Indicatore di cancellazione logica (0 = OK, 1 = Cancellato)	NUMERIC	1	
10	MOV_USER	Identificativo di chi ha aggiornato il mov.	VARCHAR	32	
11	MOV_DTAGG	Data dell'ultimo aggiornamento del movimento	DATETIME	0	
12	MOV_NOTE	Note del movimento	VARCHAR	255	
13	MOV_PROTANNO	Movimento annullato - Anno di procollazione	NUMERIC	4	
14	MOV_PROTNUM	Movimento annullato - Numero di procollazione	NUMERIC	9	
15	MOV_PROTDATA	Movimento annullato - Data di procollazione	DATETIME	0	
16	MOV_PROTNOTE	Movimento annullato - Note di procollazione	VARCHAR	255	
17	MOV_PROPRIE_CIRCMOV	Proprietario del movimento - Codice del Centro per l'Impiego - Proprietà	CHAR	5	UCL_TBCIRC
18	MOV_CIRCMOV	Proprietario del movimento - Codice del Centro per l'Impiego	CHAR	14	UCL_TBCIRC
19	MOV_PROPRIE_CIRCRIF	Movimento di riferimento - Codice del Centro per l'Impiego - Proprietà	CHAR	5	UCL_TBCIRC
20	MOV_CIRCRIF	Movimento di riferimento - Codice del Centro per l'Impiego	CHAR	14	UCL_TBCIRC
21	MOV_ANNORIF	Movimento di riferimento - Anno (-1 = padre), (> 0 = figli Indica anno del padre)	NUMERIC	4	
22	MOV_NUMRIF	Movimento di riferimento - Numero (-1 = padre), (> 0 = figli Indica numero del padre)	NUMERIC	6	
23	MOV_PROTPARZ	Indicatore di protocollazione parziale	NUMERIC	1	

Appendice B

**IL TRACCIATO RECORD DELLE STRUTTURE DI DESTINAZIONE:
LE TABELLE DI GIOVE**

Tab. B1 - DBS_AGENZIE

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	age_id	id progressivo agenzie	int	23	
2	age_codfisc*	Codice fiscale	varchar	16	
3	age_piva*	Partita Iva	varchar	11	
4	age_ver*	Versione - utile al link con Netlabor	int	3	
5	age_progr*	Progressivo - utile al link con Netlabor	int	4	
6	age_ragsoc*	Ragione sociale	varchar	255	
7	age_indirupr*	Indirizzo	varchar	40	
8	age_comupr	Comune	varchar	4	db_s_tbc Comuni
9	age_cir_info	CPI da cui sono prelevati i dati	varchar	3	db_s_tbc CPI
10	age_cir	CPI di localizzazione	varchar	3	db_s_tbc CPI
11	age_dtultag	Data ultimo agg. record di anagrafica	date		

TAB. B2 - DBS_AZIENDE

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	azi_id	id progressivo aziende	int	23	
2	azi_codfisc*	Codice Fiscale	varchar	16	
3	azi_piva*	Partita Iva	varchar	11	
4	azi_ver*	Versione - utile al link con Netlabor	int	3	
5	azi_progr*	Progressivo - utile al link con Netlabor	int	4	
6	azi_ragsoc*	Ragione sociale azienda	varchar	255	
7	azi_indirupr*	Indirizzo	varchar	40	
8	azi_comupr	Comune	varchar	4	db_s_tbc Comuni
9	azi_codatt	Codice attività economica	varchar	9	db_s_tbc ATECO
10	azi_ccnl	Contratto nazionale applicato	varchar	4	db_s_tbc CCNL
11	azi_cir_info	CPI da cui sono prelevati i dati	varchar	3	db_s_tbc CPI
12	azi_cir	CPI di localizzazione	varchar	3	db_s_tbc CPI
13	azi_dtultag	Data ultimo agg. record di anagrafica	date		

TAB. B3 - DBS_LAVORATORI

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	lav_id	id progressivo lavoratore	int	23	
2	lav_codfisc*	Codice Fiscale	varchar	16	
3	lav_gemelli	Indica la presenza di eventuali problemi di duplicazione	int	1	db_s_tbc Gemelli
4	lav_gemelli_rif	In caso di duplicazioni non risolte indica l'eventuale "gemello"	int	11	
5	lav_cpi	CPI da cui sono prelevati i dati	varchar	3	db_s_tbc CPI
6	lav_cpi_info	CPI di localizzazione	varchar	3	db_s_tbc CPI
7	lav_sesso	Sesso	varchar	1	db_s_tbc Sesso
8	lav_coddit	Cittadinanza	varchar	3	db_s_tbc Cittadinanze
9	lav_dtnasc	Data di nascita	date		
10	lav_comnasc	Comune di nascita	varchar	4	db_s_tbc Comuni
11	lav_domcom	Comune di domicilio	varchar	4	db_s_tbc Comuni
12	lav_rescom	Comune di residenza	varchar	4	db_s_tbc Comuni
13	lav_dtultag	Data ultimo agg. record di anagrafica	date		
14	lav_overlap	Segnala problemi di overlapping tra i rapporti di lavoro	int	1	
15	lav_codstu	Codice del titolo di studio	varchar	6	db_s_tbc Titoli di studio

* Disponibili solo per la versione di Giove con gli identificativi in chiaro

TAB. B4 - DBS_RAPLAV

Campo	Nome campo	Descrizione campo	Tipo	Dimensione	Tabella di rif.
1	rap_lav	Identificativo lavoratore	int	23	
2	rap_azi	Identificativo azienda	int	23	
3	rap_aqe	Identificativo agenzia	int	23	
4	rap_flu	Assume i valori 0 e 1, quando posto ad 1 identifica il record con il maggiore contenuto informativo	int	11	
5	rap_seqnum	Progressivo numerico dei record che formano i rapporti di lavoro	int	20	
6	rap_pat	Indica il record padre in rapporti di lavoro scritti su più record	int	11	
7	rap_chain	Numera da 0 a n-1, i record di una catena che formano un rapporti di lavoro	int	11	
8	rap_rec	E = record esterno, I = record interno	varchar	1	
9	rap_cirinfo	CPI da cui sono prelevati i dati	varchar	3	db_s_tbcpi
10	rap_dtavv	Data di inizio di un segmento di un rapporto di lavoro	date		
11	rap_codavv	Codice di apertura di un segmento di un rapporto di lavoro	varchar	3	db_s_tbcodavv
12	rap_contratto	Contratto	varchar	1	db_s_tbcontratto
13	rap_tempo	Orario di lavoro	varchar	1	db_s_tbttempo
14	rap_codnorm	Norma di riferimento per l'assunzione	int	2	db_s_tbnorm
15	rap_codqua	Codice della qualifica	varchar	10	db_s_tbqualifiche
16	rap_codqr	Grado dell'avviamento	int	2	db_s_tbqr
17	rap_codccn	Contratto nazionale applicato	varchar	4	db_s_tbccnl
18	rap_codces	Codice tipologia chiusura di un segmento o del rapporto di lavoro se rap flu = 1	varchar	2	db_s_tbcess
19	rap_codben	Benefici di legge goduti	varchar	2	db_s_tbbeneficio
20	rap_dtcess	Data di chiusura di un segmento di un rapporto di lavoro	date		
21	rap_ggavv	Durata prevista in giorni	int	4	
22	rap_mmavv	Durata prevista in mesi	int	2	
23	rap_oreavv	Orario di lavoro settimanale se part-time	int	3	
24	rap_livello	Livello di inquadramento	varchar	5	
25	rap_durataappr	Durata prevista in mesi se Apprendista	int	2	
26	rap_inter	Indica se il rapporto è interinale o meno	varchar	1	
27	rap_pdtavv	Data di assunzione vera e propria	date		
28	rap_pcontratto	Contratto alla nascita	varchar	1	db_s_tbcontratto
29	rap_ptempo	Orario di lavoro alla nascita	varchar	1	db_s_tbttempo
30	rap_pcodavv	Codice di apertura di un rapporto di lavoro	varchar	3	

I Tartufi già pubblicati:

- 1/2000 **EXTRACOMUNITARI AL LAVORO IN VENETO: LO "STATUS QUAESTIONIS"**
Canali d'ingresso, rischio disoccupazione, livello e caratteristiche della domanda di lavoro
di Bruno Anastasia, Maurizio Gambuzza e Maurizio Rasera
- 2/2000 **LA DISOCCUPAZIONE "AMMINISTRATIVA": UN'APPROSSIMAZIONE (O UNA FINZIONE) IRRINUNCIABILE?**
di Bruno Anastasia, Maurizio Gambuzza e Maurizio Rasera
- 3/2001 **COME CAMBIA L'OFFERTA E LA DOMANDA DI LAVORO FEMMINILE NEL VENETO**
di Anna de Angelini e Luciana Positello
- 4/2001 **OPPORTUNITÀ E DISPARITÀ. L'OCCUPAZIONE FEMMINILE NELLE GRANDI IMPRESE DEL VENETO**
Considerazioni sul monitoraggio dell'art. 9 della legge 125/91
di Maurizio Gambuzza e Maurizio Rasera
- 5/2001 **GIOVANI PERITI AL LAVORO. I PERCORSI DEI DIPLOMATI NEGLI ANNI '90 ALL'ITIS "L. DA VINCI" DI PORTOGRUARO**
di Bruno Anastasia e Maurizio Gambuzza
- 6/2001 **ANNUALE DI DATI E STATISTICHE. DATI 1996-1999**
- 7/2002 **IL NUOVO APPRENDISTATO NEL VENETO**
di Anna de Angelini
- 8/2002 **PERCORSI DI STUDIO E DI LAVORO DI TRE LEVE DI DIPLOMATI DELL'ITIS "A. PACINOTTI" DI MESTRE**
di Marina Camonico
- 9/2002 **SCENARI DI EVOLUZIONE DEMOGRAFICA IN VENETO**
Le conseguenze di ipotetici scenari di flussi migratori prossimi venturi
di Enzo Migliorini
- 10/2002 **DOPO L'APPRENDISTATO**
I percorsi lavorativi degli ex apprendisti a confronto con quelli di altri gruppi di lavoratori
di Anna de Angelini e Alessandra Boldrin
- 11/2002 **MISURE DELL'OCCUPAZIONE TEMPORANEA: CONSISTENZA, DINAMICA E CARATTERISTICHE DI UNO STOCK ETEROGENEO**
di Bruno Anastasia e Danilo Maurizio
- 12/2002 **AL LAVORO. INDAGINE SULLA TRANSIZIONE SCUOLA-LAVORO DEI DIPLOMATI ALL'ISTITUTO PROFESSIONALE "A. PAROLINI" DI BASSANO DEL GRAPPA**
di Marina Camonico
- 13/2003 **DINAMICA DELL'OCCUPAZIONE E MUTAMENTI NELLA STRUTTURA DEL LAVORO DIPENDENTE**
di Maurizio Gambuzza, Danilo Maurizio e Maurizio Rasera
(in corso di pubblicazione)
- 14/2003 **L'APPETITO VIEN MANGIANDO. LA TRANSIZIONE SCUOLA - LAVORO DEI DIPLOMATI PRESSO GLI ISTITUTI PROFESSIONALI PER I SERVIZI ALBERGHIERI E DELLA RISTORAZIONE DEL VENETO**
di Marina Camonico
- 15/2003 **"COMMERCIO E SERVIZI". LA TRANSIZIONE SCUOLA - LAVORO DEI DIPLOMATI PRESSO L'ISTITUTO PROFESSIONALE PER I SERVIZI COMMERCIALI E TURISTICI "MARCO POLO" E L'ISTITUTO TECNICO COMMERCIALE "E. DE AMICIS" DI ROVIGO**
di Marina Camonico
- 16/2004 **OCCUPATI STABILI, MOBILI, TEMPORANEI IN VENETO: MISURE DI CONSISTENZA E DI "LOCK IN"**
di Bruno Anastasia, Massimo Disarò e Danilo Maurizio
- 17/2004 **LA MOBILITÀ DEI LAVORATORI IN VENETO DAL 1993 AL 2003. UN CONFRONTO FRA LE MISURE EFFETTUATE SU DATI RTFL E LE MISURE EFFETTUATE SU DATI NETLABOR**
di Anna de Angelini
- 18/2005 **VENEZIA E TERRAFERMA. LE ESPERIENZE DI STUDIO E DI LAVORO DEI DIPLOMATI**
di Marina Camonico
- 19/2005 **I LAVORATORI ANZIANI NEL MERCATO DEL LAVORO REGIONALE: CONDIZIONAMENTI DEL PASSATO E NUOVI COMPORAMENTI**
di Anna de Angelini
- 20/2005 **I LAVORATORI DIPENDENTI IN VENETO 1998-2003: PROFILI E PERCORSI**
Statistiche sistematiche da "Giove 2005"
- 21/2006 **LA DOMANDA ESPLICITATA DI NUOVI LAVORATORI EXTRACOMUNITARI NELLA PROVINCIA DI VENEZIA**
di Bruno Anastasia, Stefania Bragato e Maurizio Rasera
- 22/2006 **GIOVE: UN DATABASE STATISTICO SUL MERCATO DEL LAVORO VENETO**
Costruzione e indicazioni per l'utilizzo
di Danilo Maurizio