



Osservatorio Mercato del Lavoro

PUF 4.0 – GUIDA A MERCURIO  
Storia, contenuto e specifiche

Luglio 2023

## 1. Storia del Puf (Public use file): da Giove a Planet, da Planet a Mercurio

**Mercurio** è il database statistico (la quarta versione), costruito con i dati che l'Osservatorio di Veneto Lavoro mette a disposizione dei ricercatori – come Puf (**Public use file**) - per dare loro la possibilità di sfruttare a pieno il bagaglio di informazioni derivanti dal sistema gestionale amministrativo dei Centri per l'impiego, sistema che organizza i flussi delle comunicazioni obbligatorie (CO) e delle dichiarazioni di disponibilità al lavoro rilasciate dai lavoratori. Dopo il primo trimestre di ogni anno viene rilasciato il database con i dati aggiornati alla fine dell'anno precedente.

Nella sua **prima versione** il **Puf**, denominato “**Giove**”, ha rappresentato un'esperienza innovativa di accesso alle basi dati amministrative, tra le prime a livello nazionale a mettere a disposizione del mondo della ricerca informazioni a livello elementare nel rispetto della legge sulla privacy e del codice deontologico.

La principale fonte di alimentazione di “Giove” era il sistema **NetLabor**, in uso presso i Centri per l'Impiego del Veneto. Le caratteristiche di quel sistema informativo richiedevano una consistente lavorazione del dato, a causa di modalità di interrogazione e tracciati record condizionati dall'organizzazione “amministrativa” delle informazioni. In Giove non erano disponibili informazioni sulla disoccupazione.

La prima release di Giove risale al 2004, l'ultima al febbraio 2008.

La svolta che ha apportato un cambiamento radicale nell'organizzazione delle informazioni che pervengono ai Centri per l'impiego in materia di rapporti di lavoro è stata segnata dall'avvento a livello nazionale del sistema di **Comunicazioni Obbligatorie (CO)** e, per il Veneto, dall'adozione del **Sistema Informativo del Lavoro (Sil Veneto)**.

Queste modifiche di contesto hanno portato l'Osservatorio di Veneto Lavoro a rivedere le modalità di costruzione del Puf, dando luogo alla versione 2, denominata **Planet**. Rispetto a Giove, Planet ha comportato un affinamento delle modalità di elaborazione, correzione e distribuzione del dato statistico. Ha visto il rilascio di versioni successive, che hanno mantenuto la medesima struttura, pur aggiornandosi nei contenuti e riportando correzioni sempre più affidabili (e di assoluta trasparenza per il ricercatore-utilizzatore).

Le differenze con Giove sono numerose. Il tracciato record di Planet e le modalità di interrogazione sono completamente cambiate; inoltre la versione 2 inglobava la disoccupazione in coerenza con i rapporti di lavoro e mettendo a disposizione una tabella per seguire i cambiamenti di status di tutti i lavoratori presenti nella base dati.

La prima release di Planet risale al maggio 2008, l'ultima al luglio 2015.

Il consolidamento del sistema Silv ed alcune modifiche introdotte nelle CO, assieme alla volontà di semplificare l'interrogazione del database statistico, hanno portato l'Osservatorio a realizzare la terza versione del Puf (evoluto poi nella quarta), denominata **Mercurio**. I nuovi Puf 3.0 e 4.0 condividono con Planet la fonte dell'informazione (Silv e CO), ma prevedono una diversa organizzazione nelle tabelle dei rapporti di lavoro e disoccupazione.

Dal mese di aprile 2019, a seguito delle recenti modifiche regolamentari in materia di trattamento dei dati individuali, la modalità di distribuzione di Mercurio si è ridotta alla sola versione anonimizzata.

Il rilascio del Puf è subordinato alla presentazione, oltre che di un adeguato progetto di ricerca, della documentazione prevista dalla normativa in materia di protezione dei dati personali, il tutto sottoposto alla valutazione e approvazione da parte di una Commissione appositamente istituita presso Veneto Lavoro.

## 2. Il contenuto di Mercurio

### ***Tabelle dei fatti***

- ft\_datore\_lavoro: dataset dei datori di lavoro;
- ft\_lavoratore: anagrafiche minime dei lavoratori;
- ft\_rapporti\_lavoro: ciascun record contiene le informazioni relative agli eventi (assunzione, trasformazione contrattuale, proroga e cessazione) che compongono un singolo rapporto di lavoro;
- ft\_disoccupazione: eventi di disoccupazione (per calcolo di flussi e stock); sono indicati il motivo di ingresso e quello di uscita dall'episodio di disoccupazione e sono esclusi i periodi di lavoro in regime di disoccupazione amministrativa.

### ***Tabelle di decodifica (o di appoggio)***

- dt\_ateco: classificazione dei settori di attività;
- dt\_contratto: classificazione dei contratti;
- dt\_dove: classificazione della geografia;
- dt\_evento\_uscita\_disoccupazione: classificazione degli eventi che determinano l'uscita dalla disoccupazione;
- dtlivello\_studio: classificazione dei livelli di istruzione;
- dt\_motivo\_cessazione: classificazione dei motivi di cessazione;
- dt\_nazione\_cittadinanza: classificazione delle cittadinanze;
- dt\_orario: classificazione degli orari di lavoro;
- dt\_qualifica: classificazione delle qualifiche/figure professionali;
- dt\_stato\_entrata\_disoccupazione: classificazione degli status occupazionali all'inizio del periodo di disoccupazione;
- dt\_trasformazione: classificazione delle tipologie di trasformazioni contrattuali;

Alcune tabelle di appoggio, oltre alle codifiche “ministeriali”, propongono delle aggregazioni così come definite e utilizzate dall'Osservatorio di Veneto Lavoro.

### ***Livello di aggiornamento delle anagrafiche***

Le CO descrivono contenuti e situazioni di natura mutevole nel tempo: si pensi infatti alle caratteristiche dei lavoratori (quali il domicilio, la residenza, il livello d'istruzione, la cittadinanza) oppure a quelle del rapporto di lavoro (la qualifica, l'orario di lavoro, il settore di attività e l'ubicazione). E' probabile che nel periodo di vita del sistema informativo questi attributi siano variati almeno una volta. Per quanto riguarda le anagrafiche dei lavoratori, la nostra scelta è stata quella di riportare nel Puf l'ultima informazione comunicata. Per quanto concerne le informazioni riguardanti il settore di attività, l'orario e la qualifica del rapporto di lavoro, è stato inserito il dato contenuto nell'ultima CO inviata; per la localizzazione, invece, oltre all'ultima informazione comunicata, è presente anche quella relativa al momento dell'assunzione.

### ***Tracciato Record***

Nel file Excel allegato è riportato il tracciato record delle tabelle dei fatti, con i riferimenti alle tabelle di appoggio.

### 3. Data quality sul rapporto di lavoro

I gestionali di Silv e CO - le due fonti di alimentazione di **Mercurio** – sono oggetto di continue modificazioni e correzioni, sia dal punto di vista logico, sia per quanto riguarda il sistema di validazione (oltre che per tenere conto delle novità normative).

Nonostante alcuni dei problemi riscontrati negli anni scorsi siano stati risolti in fase di validazione delle Comunicazioni Obbligatorie, permangono parziali incongruenze, determinate soprattutto:

- dalla definizione di algoritmi deterministici, a volte un po' rigidi, rispondenti all'esigenza di ridurre al minimo la quantità di comunicazioni “da validare manualmente”;
- dalla bassa qualità delle informazioni ereditate dal precedente gestionale (Netlabor).

Nelle diverse versioni del Puf abbiamo sviluppato degli algoritmi in grado di correggere, dove possibile, situazioni di palese incertezza, cercando – nei casi dubbi - di mantenere un atteggiamento “conservativo” per evitare di modificare in modo eccessivo l'informazione originale.

Le correzioni riguardano diversi aspetti e possono introdurre delle modifiche nei dati o, nel caso in cui risulta impossibile ricostruire correttamente l'informazione, portare a cancellare il rapporto di lavoro.

#### ***Verifica della coerenza delle date***

\* Assunzioni e cessazioni

In caso di data inizio rapporto mancante o di data di cessazione antecedente all'inizio, il rapporto di lavoro è stato cancellato.

\* Trasformazioni contrattuali

Nel sistema sono presenti diverse trasformazioni contrattuali effettuate in date non coerenti con il periodo di vita del rapporto di lavoro. In questi casi lo schema di comportamento è il seguente:

- se la trasformazione avviene nell'intervallo compreso tra la data inizio del rapporto e il giorno successivo alla data di cessazione è valida ed in particolare:
  - se avviene nella data di cessazione o nel giorno successivo viene eliminata la cessazione
  - se avviene successivamente a questo termine (o antecedentemente alla data di inizio del rapporto) non viene considerata; il rapporto di lavoro quindi si conclude alla data originalmente.
- in tutti gli altri casi la trasformazione viene considerata nulla ed eliminata.

#### ***Verifica del tipo di trasformazione comunicato***

Per valutare la correttezza delle trasformazioni contrattuali rispetto al rapporto di lavoro cui sono riferite è necessario conoscere:

- il contratto di assunzione
- il tipo di trasformazione comunicato e la sua compatibilità con il contratto di assunzione
- il contratto al momento dell'eventuale cessazione.

Sulla base di queste tre informazioni si determina la situazione in cui ricade la trasformazione esaminata e, di conseguenza, si definiscono le misure da intraprendere per correggere eventuali incoerenze. La gestione di queste situazioni avviene secondo lo schema di seguito riportato. Ne consegue che in alcuni casi di incompatibilità l'evento di trasformazione viene modificato in modo da risultare coerente con il rapporto di lavoro, in altri casi l'informazione portata dalla comunicazione di trasformazione permette di correggere gli altri eventi registrati (assunzione o cessazione) afferenti al medesimo rapporto di lavoro.

Contratto di assunzione	Trasformazione "concorde" con l'assunzione	Cessazione con contratto a tempo indet.	Soluzione
Cap, Ctd	Si	Si	Situazione corretta.
Cap, Ctd	Si	No	Correzione del contratto di cessazione (tempo indeterminato).
Cap, Ctd	No	Si	Correzione del tipo di trasformazione in base al contratto di assunzione.
Cap, Ctd	No	No	Correzione del contratto di cessazione (tempo indeterminato). Correzione del tipo di trasformazione in base al contratto di assunzione.
Som, Int, Dom, Par, Esp	No	-	Il contratto di cessazione viene sovrascritto dal contratto di assunzione. Cancellazione dell'evento di trasformazione.
Cti	No	-	Modificazione del contratto di assunzione in base al tipo di trasformazione.

### ***Trasformazioni contrattuali implicite***

In alcuni casi, il rapporto di lavoro può essere avviato con un determinato contratto (es. tempo determinato) e concludersi con un contratto diverso (es. tempo indeterminato), pur senza che sia stata comunicata una trasformazione contrattuale. In altre situazioni, alcuni rapporti di lavoro a termine (es. tempo determinato) non hanno una data di fine prevista valorizzata.

In generale, laddove l'evoluzione contrattuale risulta coerente con una possibile trasformazione "mancante", questa viene generata d'ufficio.

### ***Verifica delle proroghe***

Le comunicazioni di proroga devono attenersi ad alcune regole, per essere coerenti con il rapporto di lavoro a cui afferiscono, in particolare:

- le proroghe possono riguardare solo rapporti di lavoro a termine: tutte le proroghe applicate a contratti "non compatibili" (in generale rapporti a tempo indeterminato, anche se non strettamente dipendenti) verranno annullate.
- l'inizio della prima proroga deve avvenire contestualmente alla data comunicata di fine prevista. La fine dell'ultima proroga comunicata non deve superare il termine ultimo del rapporto (data di cessazione) o l'eventuale data di trasformazione contrattuale (verso il tempo indeterminato). Le comunicazioni di proroga che non rispettano questi vincoli verranno annullate.

### ***Ricostruzione della data di cessazione***

Nel caso in cui la mancanza dell'evento di cessazione non sia giustificabile, viene generata d'ufficio una chiusura del rapporto di lavoro e la durata è determinata ponendola pari alla moda delle durate per la medesima tipologia contrattuale. Questa situazione si verifica in un numero molto limitato di casi.

## **4. Data quality sulle storie lavorative**

Il database statistico ha l'obiettivo di garantire, dove possibile, la coerenza dei rapporti di lavoro intesi come sequenza degli eventi che nel contesto informatico sono identificati dalle Co.

La messa in qualità del singolo rapporto di lavoro può essere estesa anche alla storia lavorativa di un singolo individuo, intesa come sequenza di rapporti di lavoro.

In questo modo, così come all'interno del singolo rapporto di lavoro non possono coesistere ad esempio più eventi di assunzione o di cessazione, nella carriera lavorativa di un soggetto non dovrebbero essere presenti rapporti di lavoro erroneamente sovrapposti o addirittura duplicati, se non in casi prestabiliti.

### ***Rapporti di lavoro duplicati***

Nonostante i controlli operati a monte dal Sil è possibile che nella storia lavorativa di un soggetto esistano due o più rapporti di lavoro apparentemente duplicati. La definizione di univocità di un rapporto di lavoro è legata ad alcuni aspetti chiave ossia: il lavoratore coinvolto, il datore di lavoro, la data di inizio del rapporto ed eventualmente la sede di lavoro ed il contratto. Consideriamo un rapporto come doppio, e quindi da eliminare, quando per la stessa coppia lavoratore/datore esiste un altro rapporto di lavoro avviato lo stesso giorno, con lo stesso contratto e nella stessa sede lavorativa.

### ***Overlapping***

In presenza di sovrapposizione tra rapporti di lavoro per il medesimo lavoratore, prima di apportare delle modifiche è necessario accertarsi se l'overlapping sia legittimo o meno. Consideriamo la sovrapposizione "legittima" se:

- almeno uno dei due rapporti di lavoro è non subordinato/dipendente. Sono quindi ammissibili i casi in cui rapporti con contratto Intermittente, Parasubordinato, Domestico o Work experience si sovrappongono ad altri, oppure
- entrambi i rapporti di lavoro rientrano tra i contratti dipendenti (tempo indeterminato, apprendistato, tempo determinato, somministrato) e sono part-time.

Sono invece considerate non legittime, quindi da correggere, le situazioni in cui:

- entrambi i rapporti di lavoro sono subordinati (tempo indeterminato, apprendistato, tempo determinato, somministrato) e
- almeno uno dei due è a full-time.

Ad esempio:

- la sovrapposizione di un tempo indeterminato full-time con un rapporto parasubordinato è ammissibile,
- la sovrapposizione di due rapporti di lavoro a tempo determinato part-time è ammissibile,
- la sovrapposizione di un rapporto di lavoro a tempo indeterminato full-time con un altro a tempo determinato/apprendistato/somministrato non è ammissibile.

In linea generale, si procede per iterazioni successive individuando il rapporto di lavoro che causa il maggior numero di sovrapposizioni, all'interno della carriera di un lavoratore.

La data di inizio del rapporto non viene mai modificata, la procedura agisce solamente:

- "arretrando" la data di cessazione, nei casi meets, overlaps, finished by e starts (e reciproci) oppure
- cancellando il rapporto di lavoro, nei casi contains ed equals.

La procedura prosegue fino alla risoluzione di tutte le sovrapposizioni non consentite nella carriera lavorativa.

Ci sono poi una serie di eccezioni e particolarità dettate dalle caratteristiche del rapporto di lavoro (contratti, orari, ecc.).

## 5. L'acquisizione di Mercurio

I dataset vengono consegnati sottoforma di dump formato MySQL/MariaDB<sup>1</sup> in un singolo file compresso.

Per creare le strutture e popolarle con i dati è necessario installare una versione server del DBMS MySQL/MariaDB; se non presente

- scaricare l'archivio del Puf e decomprimerlo per ottenere il file con estensione .sql
- creare un database vuoto nel proprio server MySQL/MariaDB
- eseguire il file .sql nell'ambiente di sviluppo MySQL/MariaDB<sup>2</sup> per creare le tabelle e caricare i dati nel database creato.

Al fine di ridurre il volume di dati da trasferire, si è deciso di non indicizzare i campi chiave nelle tabelle fornite nel PUF. È consigliabile creare degli indici nei campi di join utilizzati<sup>3</sup>.

Da dicembre 2020 si è scelto di mettere a disposizione, oltre al formato SQL, i dati in formato CSV. La prima riga di ciascun file contiene il nome delle colonne e il separatore utilizzato è la virgola “,” . Tutti i campi sono racchiusi tra doppi apici “ ” “ ”.

---

<sup>1</sup> Come per le versioni precedenti del Puf, abbiamo ritenuto opportuno affidare la modalità di distribuzione ad un database open source. Si tratta in questo caso di MariaDB, controparte open source e non commerciale di MySQL. Per maggiori informazioni si veda <https://mariadb.org/> e <http://www.mysql.com/>.

<sup>2</sup> Per maggiori informazioni sul caricamento di un dump sql, si veda <https://mariadb.com/kb/en/mariadb/backup-and-restore-overview/>.

<sup>3</sup> Per maggiori informazioni sulla creazione e modifica di indici, si veda <https://mariadb.com/kb/en/mariadb/optimization-and-indexes/>

## Riferimenti bibliografici

Prima dell'utilizzo del PUF, si consiglia vivamente di prendere visione dei seguenti documenti:

- Maurizio D., *Giove: un database statistico sul mercato del lavoro veneto*, collana “I tartufi”, n. 22, 2006, <https://www.venetolavoro.it/tartufi1>
- Anastasia B., Disarò M., Gambuzza M. e Rasera M., *Comunicazioni obbligatorie e analisi congiunturale del mercato del lavoro: evoluzione, problemi metodologici, risultati*, collana “I tartufi”, n. 35, 2009, <https://www.venetolavoro.it/tartufi1>
- Anastasia B., Disarò M., Emireni G., Gambuzza M. e Rasera M., *Guida all'uso delle Comunicazioni Obbligatorie nel monitoraggio del mercato del lavoro. Seconda versione: dicembre 2010*, collana “I tartufi”, n. 36, 2010, <https://www.venetolavoro.it/tartufi1>
- Anastasia B., Emireni G., Gambuzza M., Maschio S., Rasera M., *Grammatica delle comunicazioni obbligatorie  $\beta$  – Guida alle elaborazioni a partire dai dati di flusso*, <https://www.venetolavoro.it/tempi-metodi>