

Validazione dei dati della rilevazione della popolazione residente da microdato anagrafico effettuata dalla Regione Emilia-Romagna in collaborazione con le Province e la Città Metropolitana

I microdati della popolazione residente sono forniti, al 1° gennaio di ogni anno, dai Comuni per ottemperare a quanto previsto dalla scheda PBO-0004 del PSN dal 2015, e vengono caricati sul portale web al link <https://sasweb.regione.emilia-romagna.it/statistica/restricted/Rilevazione.do>

I dati possono essere caricati una volta sola dal referente individuato dalla Provincia/Città Metropolitana o dal Comune. Tale referente sarà dotato di credenziali personali. Nel caso si rendesse necessario ricaricare un dato più aggiornato, viene fatta richiesta alla Regione, all'indirizzo di posta elettronica statdatianagrafici@regione.emilia-romagna.it, che provvederà a rendere possibile il nuovo upload e poi solo successivamente il Comune o la Provincia/Città Metropolitana potranno riprovare l'operazione.

Il fatto che il caricamento sia effettuato dalla Provincia/Città Metropolitana o dal Comune dipende dalla modalità scelta dalle Province/Città Metropolitana e sottoscritta in una specifica convenzione annuale con la Regione.

La schermata del portale web per il caricamento dei dati che si presenta alla Regione è la seguente:

Rilevazione popolazione - Regione Emilia-Romagna - Internet Explorer

https://sasweb.regione.emilia-romagna.it/statistica/restricted/Rilevazione.do

ER Statistica

Primo Piano Entra in Regione

Rilevazione Popolazione

Salve Volonino Giuseppe!

Report disponibili

Anno di riferimento: 2016

- Distribuzione della popolazione complessiva per comune, sesso e singolo anno di età
- Distribuzione della popolazione straniera per comune, sesso e singolo anno di età
- Distribuzione della popolazione italiana per comune, sesso e singolo anno di età
- Distribuzione delle famiglie per comune e numero di componenti
- Distribuzione dei residenti nelle famiglie per comune e numero di componenti
- Distribuzione per comune e numero di componenti dei residenti in convivenza e numero di convivenze
- File csv della popolazione straniera per comune, sesso, singolo anno di età e cittadinanza
- Residenti nelle famiglie unipersonali per comune, classe di età e sesso
- Numero famiglie per comune, numero di componenti e componenti stranieri
- Numero famiglie per comune, numero componenti minorenni e componenti stranieri
- Numero di famiglie per comune, numero componenti anziani e numero componenti

Situazione caricamenti

- Caricamenti dei comuni
- Validazione dati
- Esporta dataset di lavoro

Caricamento dati

Sfoggia...

Numero di record attesi *

Carica

Per controllare la situazione del caricamento dei Comuni basta selezionare “Caricamenti dei comuni” e si visualizzerà la seguente videata:

The screenshot shows a web browser window displaying the 'Rilevazione Popolazione' page. The page has a header with the 'ER Statistica' logo and navigation buttons for 'Primo Piano' and 'Entra in Regione'. The main content area is titled 'Rilevazione Popolazione' and contains a table titled 'Situazione caricamenti dei comuni'. The table has four columns: 'Comune', 'Bloccato', 'Completo', and 'Validato'. All 16 municipalities listed have 'SI' in all three status columns. A 'Torna alla home' button is located at the bottom of the table area.

Comune	Bloccato	Completo	Validato
Agazzano	SI	SI	SI
Alseno	SI	SI	SI
Besenzone	SI	SI	SI
Bettola	SI	SI	SI
Bobbio	SI	SI	SI
Borgonovo Val Tidone	SI	SI	SI
Cadeo	SI	SI	SI
Calendasco	SI	SI	SI
Caminata	SI	SI	SI
Caorso	SI	SI	SI
Carpaneto Piacentino	SI	SI	SI
Castell'Arquato	SI	SI	SI
Castel San Giovanni	SI	SI	SI
Castelvetro Piacentino	SI	SI	SI
Cerignale	SI	SI	SI
Coli	SI	SI	SI

“Bloccato”=Sì vuol dire che è già stato caricato il file del Comune e che per poterlo caricare nuovamente bisogna che venga sbloccato (bloccato ridiventerebbe così =No).

“Completo” mostra se siano presenti alcuni errori riscontrati dal sistema come più avanti specificato.

“Validato”: la Regione valuterà e selezionerà i comuni da validare al termine di tutti i controlli e solo se “Completo” =Sì.

Controlli automatici sui microdati durante il caricamento dei dati sul portale

I dati quando vengono caricati dal sistema, vengono scritti sia in una directory in cui viene conservato il file originario, sia, dopo una serie di controlli automatici, in un file che divide i record senza errori da quelli con qualche errore aggiungendo dei flag di errore specifici.

Vi sono alcuni controlli che avvengono durante l'upload che determinano il **blocco del caricamento** stesso:

- quando il numero dei record riscontrati non coincida con il numero dei residenti inserito/dichiarato;
- quando i codici della provincia e/o del comune non siano validi;
- quando il codice individuale contemporaneamente alla data di nascita risultino duplicati;
- quando viene caricato un file Excel o non csv delimitato da “;”, in questo caso viene fuori il messaggio di errore e si vede il box: Ovviamente non è possibile caricare file che non siano csv ben formattati

Controlli sul **tracciato** che determinano creazione di specifiche variabili di errore (contenenti flag 1 se errato, 0 se corretto) ma che **non bloccano** il caricamento dei dati:

- presenza di tutte le variabili previste con nome corretto;
- formato delle variabili uguale a quello previsto dal tracciato (tranne alcune variabili testo che sono state trattate dall'utente come numero e a cui manca lo zero davanti; infatti tali variabili verranno trasformate dal sistema in variabili alfa-numeriche aggiungendo lo zero davanti senza considerarlo un errore: esempi sono il codice del comune, della provincia o dello stato civile);
- se per caso venissero fornite le variabili relative al codice fiscale, nome e cognome, esse verranno individuate e sbianchettate (rimangono, quindi, senza contenuti).
- In caso di data malformata viene segnalato un errore di tracciato.

Altri controlli **sui contenuti** che determinano creazione di specifiche variabili di errore (contenenti flag 1 se errato, 0 se corretto) ma che **non bloccano** il caricamento:

- codice di cittadinanza esistente, cioè che il codice di cittadinanza sia compreso nell'elenco delle cittadinanze ISTAT con l'aggiunta del codice 224; 100 è il codice di cittadinanza italiana;
- anno di nascita non successivo all'anno precedente la rilevazione e non antecedente al 1900;
- il problema sulle date di nascita 00/00/xxxx è stato risolto assegnando un fittizio 02/01/xxxx in questi casi.
Il controllo è stato esteso anche alla data di iscrizione.
- Se la data di iscrizione è missing viene modificata in 31/12/ anno precedente la rilevazione (questo perché risulta troppo dispendioso richiederla al comune e questa data non crea problemi nel confronto con la data di nascita);
- se nella data di nascita l'anno è 0 o la data è missing allora viene segnalato l'errore;
- se il tipo di residenza è 1 allora il codice della famiglia deve contenere un valore maggiore di zero e il codice di convivenza deve essere vuoto o missing o 0, altrimenti si va in errore;

- se il tipo di residenza è 2 sia il codice di famiglia che quello di convivenza possono contenere un valore diverso da 0 o missing; il codice di convivenza deve essere maggiore di 0 altrimenti si va in errore;
- la variabile numero di componenti verrà ignorata e ricalcolata alla fine della validazione finale del dato;
- la codifica della variabile relazione di parentela cambia in base al contenuto del tipo di residenza
 tipo di residenza=1 relazione di parentela varia da 01 a 12 (da 1 a 12)
 tipo di residenza=2 relazione di parentela varia da 01 a 04 e 11 (da 1 a 4 e 11)
- provincia e comune di nascita devono essere nulli se è valorizzato il codice del paese di nascita estero

Altre varie:

- I record caricati per comune vengono contati per comune e inseriti nella tabella riassuntiva.
- Caricamento: viene inviato un messaggio via e-mail a chi effettua l'upload quando questo è terminato.
- Viene segnalato errore di sistema e mostrato messaggio per riprovare successivamente se il file non sia stato caricato per problemi di sovraccarico del sistema.
- Calcolo dell'età. Es. per una persona nata nel 1940: al 31.12.2013 quanti anni avrà?
 per l'età bisogna fare 2013-anno di nascita. Quindi se uno è nato nel 2013, al 31.12.2013 avrà 0 anni. Nell'esempio sarà $2013-1940=73$ anni. Bisogna stare attenti che non ci siano date di nascita del 2014 se la rilevazione è relativa al 31.12.2013.

Solo se tutti gli errori segnalati automaticamente dal sistema saranno stati corretti, sarà tecnicamente possibile validare i dati, anche se, in realtà, vi sono molti altri controlli che vengono svolti successivamente prima della loro validazione.

I flag di errore sono:

err_tracciato: se c'è la data di nascita missing o se qualche variabile ha il nome sbagliato o la variabile filler non è 1

err_cittadinanza: se un codice di cittadinanza è assente o fuori range (fuori dall'elenco delle cittadinanze ISTAT con l'aggiunta di 224 e 100 per quella italiana)

err_anno_nascita: se prima del 1900 e successivo all'anno di rilevazione

err_famiglia_1: se tipo di residenza=1 e codice di convivenza>0 o codice di famiglia =missing o zero oppure se tipo di residenza =2 e codice di convivenza missing o zero.

err_relazione_parentela: se fuori range a seconda del tipo di residenza;

err_tipores: tipo di residenza diverso da 1 e 2

err_paese_estero: se è presente lo stato estero di nascita e ci sono valorizzati la provincia e/o il comune di nascita;

completo riassume tutti i flag di errore: completo=0 se almeno un flag=1 (cioè in errore) su almeno un record nel comune, completo=1 se tutti i flag sono=0 non solo nel record, ma su tutti i record del comune stesso;

Altri controlli e correzioni/integrazioni

Successivamente ai controlli automatici sui microdati in fase di caricamento del dato sul portale, viene effettuata un'ulteriore serie di verifiche che completano i controlli.

Sono praticati con una serie di programmi e macro SAS sviluppati, talvolta anche ad hoc, in locale. Per prima cosa viene verificata la coerenza dell'anno di rilevazione.

Successivamente si accerta la completezza dei dati (viene controllato anche l'ammontare dei residenti in famiglia e in convivenza confrontandolo con i valori presenti nel bilancio della popolazione annuale Istat) e se tutti i dati siano stati effettivamente caricati nella banca dati di lavoro.

A livello provinciale si controlleranno **tutte** le variabili in esame: **cittadinanza, sesso, data di nascita e di iscrizione anagrafica, stato civile, relazione di parentela, tipo di residenza, codice di famiglia o di convivenza, provincia, comune e stato estero di nascita e il filler**(se i valori siano fuori dal range o con formati non propri, o mancanti o incoerenti o slittati in altre variabili ecc...) e tutti i record con errori segnalati dai flag di errore dei controlli automatici.

Si cercherà, di volta in volta, una possibile **correzione** agli errori riscontrati cercandola nei dati dell'anno precedente o, **eventualmente**, chiedendo alle Province/Città Metropolitana di **contattare i referenti comunali** per poter correggere o integrare (fino anche a ricaricare) i dati.

Si produrrà quindi un file csv contenente i dati corretti dei soli comuni con problemi che verranno dunque ricaricati sul portale facendo ripartire l'iter dei controlli sui microdati.

I controlli sui macrodati vengono svolti successivamente, una volta che non ci siano più errori a livello del dato individuale. Tali verifiche sono relative ai dati aggregati a livello comunale per sesso e singolo anno di età, per sesso e cittadinanza, per cittadinanza e singolo anno di età confrontandoli con i valori analoghi ma relativi all'anno precedente. Praticamente viene individuato un intervallo di accettabilità di variazione del dato e se il dato riscontrato esce da tale intervallo avviene la segnalazione di possibile errore che sarà da valutare di volta in volta sulla base dei flussi migratori o di nascita, morte o acquisizione di cittadinanza (che viene calcolata in base al record linkage con i microdati dell'anno precedente) del periodo considerato.

Una complicazione ai controlli dei macro dati riguarda le continue nuove fusioni di comuni delle quali bisogna tener conto nell'appaiamento dei dati rispetto all'anno precedente e nella pubblicazione dei dati sul sito della statistica, visto che il nuovo formato dei comuni sarà da gestire in tutte le tabelle pubblicate.

Una volta che tutti i record siano stati corretti, avviene la validazione dei dati spuntando tutti i comuni da validare nell'apposita schermata.

A questo punto viene creato il dataset con i dati validati e dal quale vengono estratti i report visibili da parte delle Province/Città metropolitana e della Regione, ciascuna per il proprio territorio, sul portale web di caricamento. Seguirà poi la pubblicazione dei dati sul sito della statistica con relativi controlli post-pubblicazione nell'area di test.

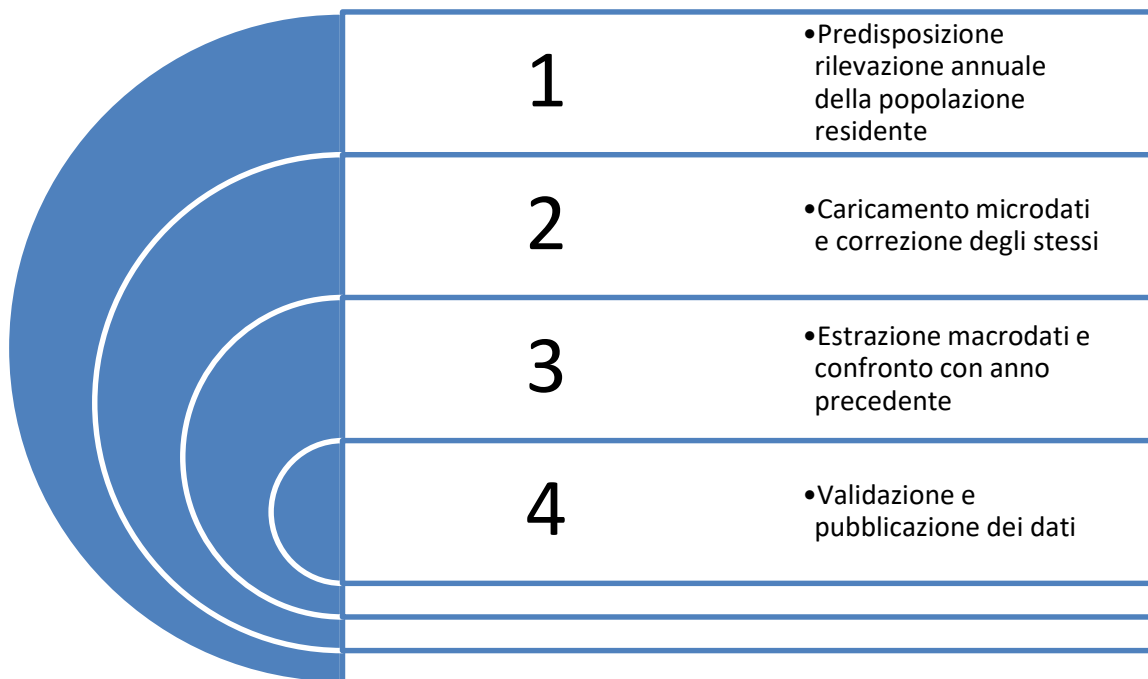
Processo di validazione: sintesi

Dal 1.1 2015 la rilevazione, basata su elenchi di microdati, è stata completata annualmente su tutti i comuni della regione

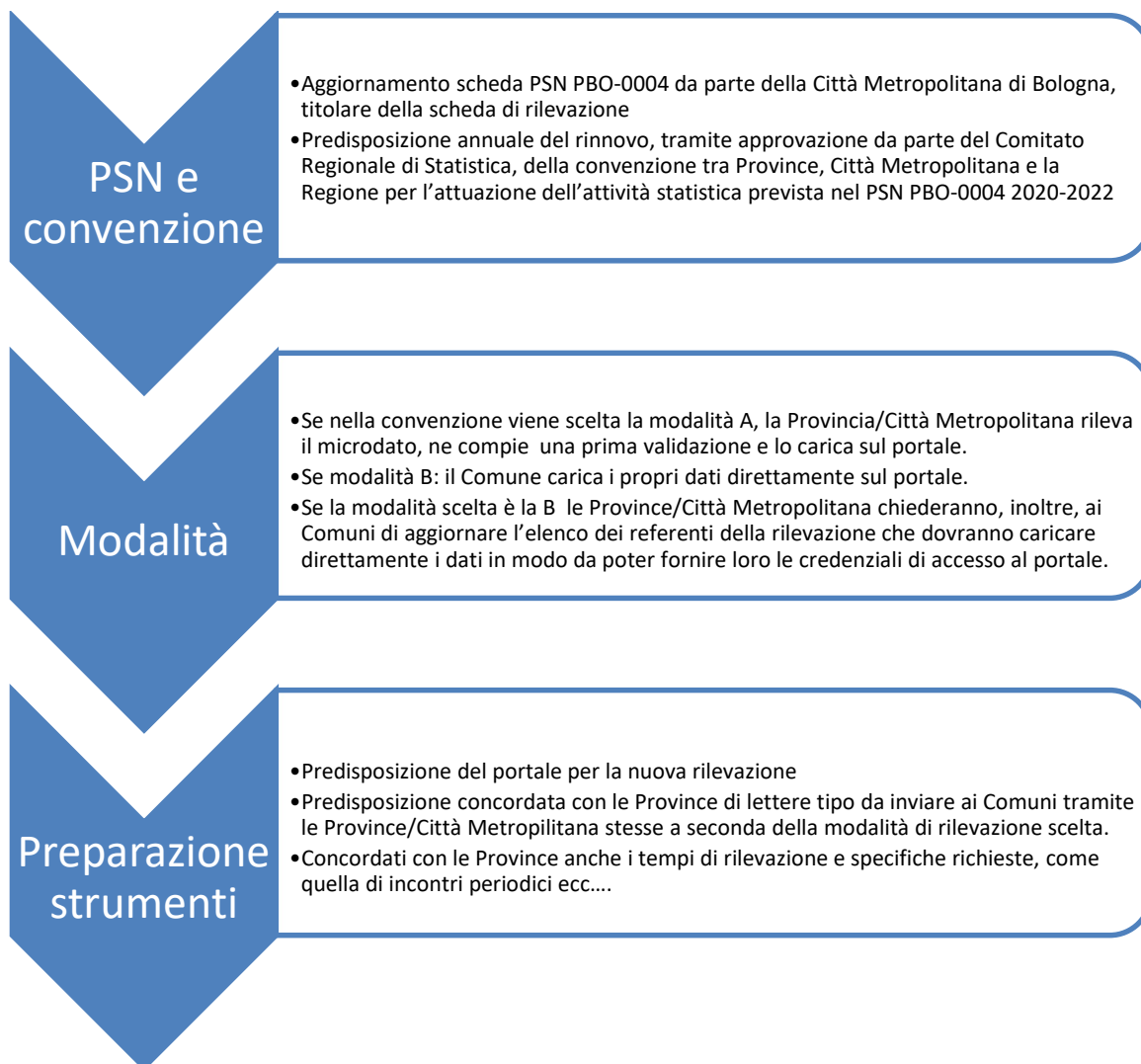
Le Province/Città Metropolitana che effettuano direttamente il caricamento, svolgono anche un primo controllo/correzione dei dati.

Tutti i processi di correzione/validazione effettuati successivamente dalla Regione sono stati concordati tra Regione e Province/Città Metropolitana.

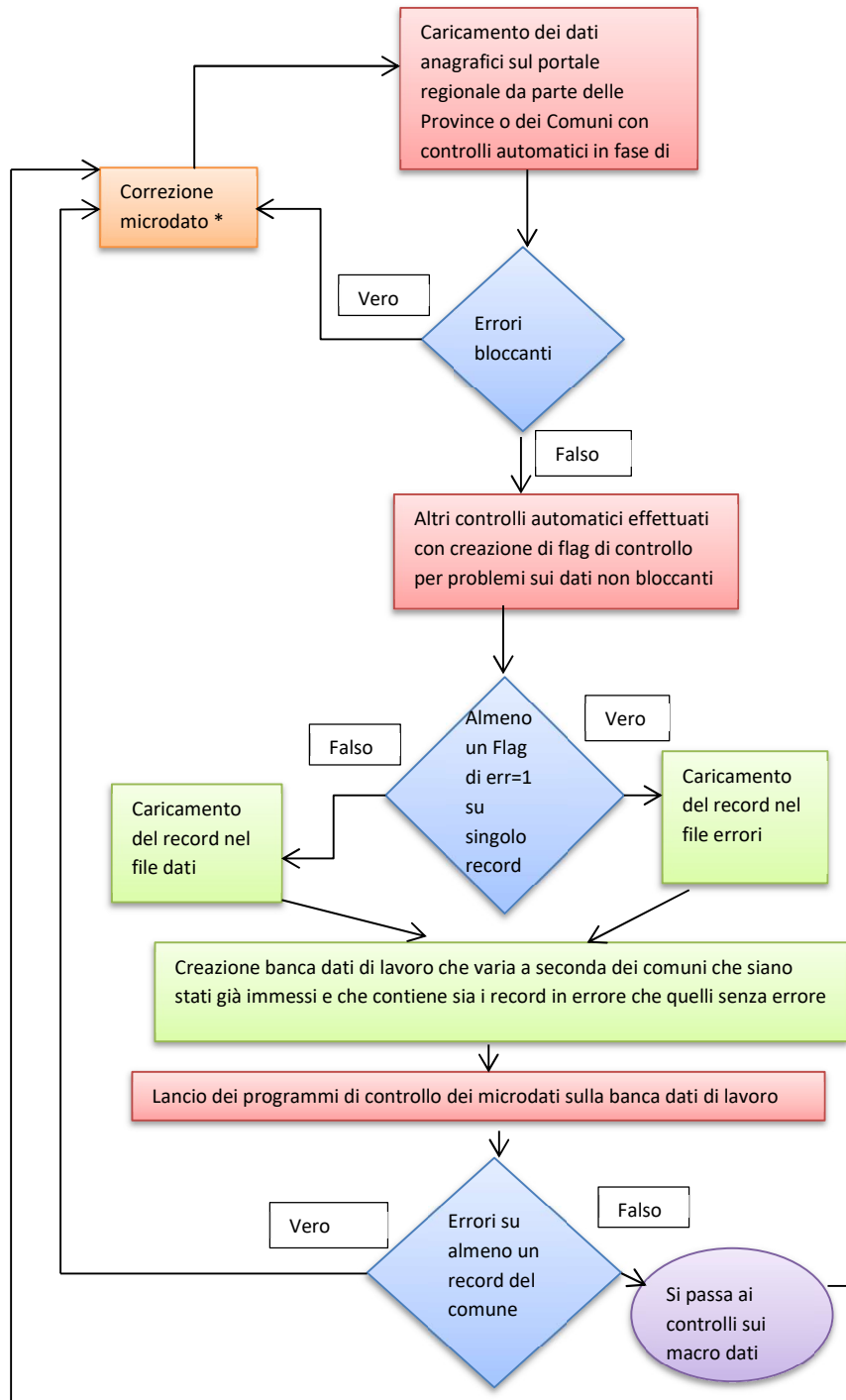
Il seguente schema riassume i processi di validazione dei dati che vengono meglio specificati nei diagrammi di flusso successivi.



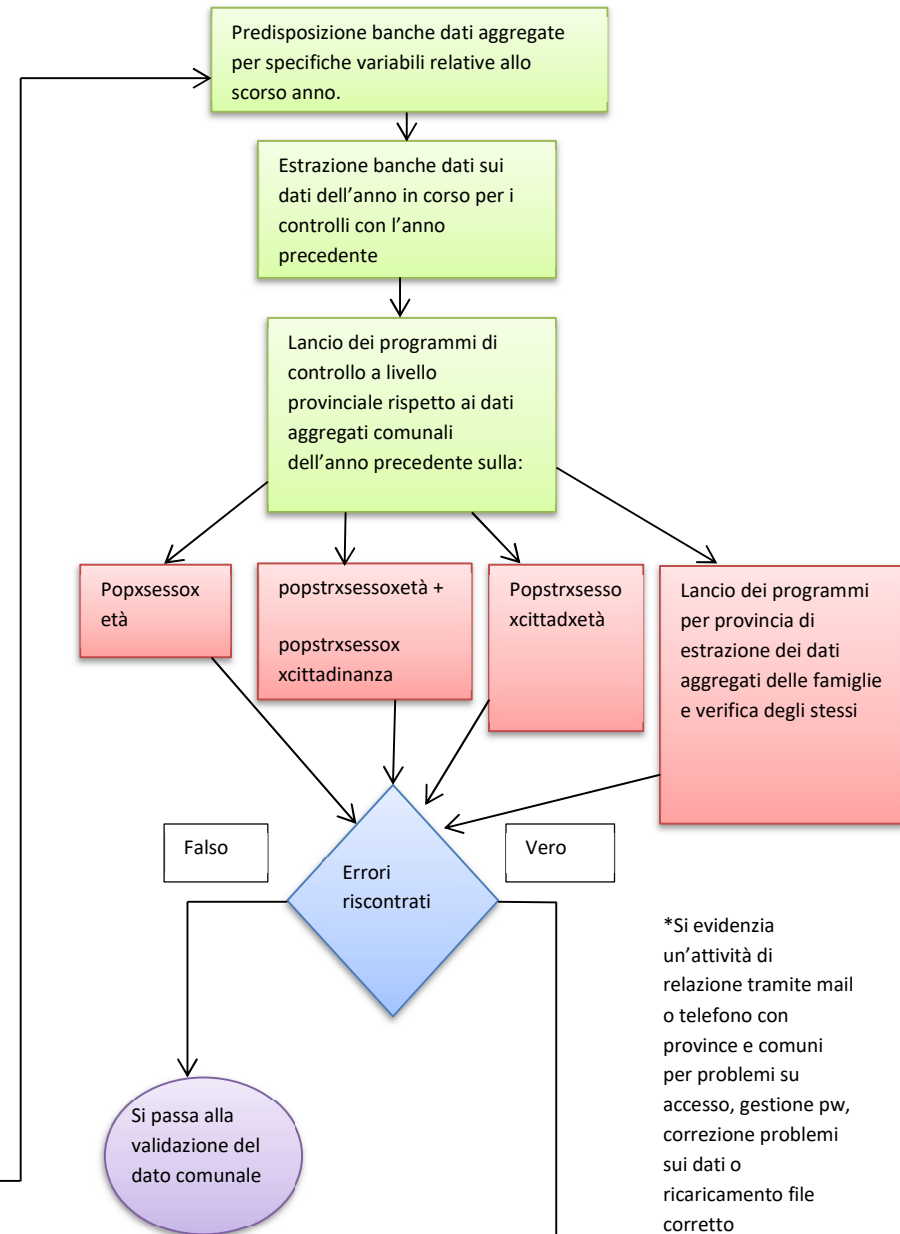
1. Predisposizione rilevazione annuale della popolazione residente



2. Caricamento microdati e controlli



3. Controllo sui macro dati



*Si evidenzia un'attività di relazione tramite mail o telefono con province e comuni per problemi su accesso, gestione pw, correzione problemi sui dati o ricaricamento file corretto

4. Validazione e pubblicazione dei dati

