

## Il glossario d'Istituto on-line

### NOTA METODOLOGICA

Il glossario diffuso on-line contiene i termini specialistici utilizzati dall'Istituto nazionale di statistica nei diversi prodotti editoriali (annuari, rapporti, comunicati stampa, banche dati ecc.). I termini pubblicati – passati rispetto al primo rilascio del 2015 da circa 700 a quasi 1.400 – si riferiscono a unità di analisi, variabili e classificazioni in uso nelle rilevazioni nonché alla terminologia tecnico-scientifica di riferimento.

Molte pubblicazioni dell'Istat sono corredate da un proprio glossario che, pur rappresentando un supporto utilissimo per l'utente finale e completo nella sua unicità, talvolta presenta termini<sup>1</sup> incoerenti con quelli risultanti in altri glossari; i diversi glossari, a loro volta, spesso adottano criteri nella formulazione dei lemmi e delle definizioni non coordinati fra loro.

Per sanare e risolvere queste difformità, che nella fattispecie si rilevano nella fase di diffusione del processo produttivo statistico<sup>2</sup>, ma possono verificarsi anche in altre fasi, si è avviato un processo finalizzato alla definizione di un glossario unico d'Istituto (d'ora in poi GU) **integrato, armonizzato** nei contenuti e **standardizzato** nella formulazione. L'attività si articola in diverse fasi – acquisizione, armonizzazione, standardizzazione, monitoraggio e aggiornamento – di seguito descritte. I termini finora documentati nel GU non sono ancora tutti quelli correntemente utilizzati in Istituto, ma certamente i più importanti; quelli pubblicati on-line costituiscono il sottoinsieme dei termini del GU utilizzati esclusivamente nei prodotti editoriali e fruiti correntemente dagli utenti che accedono al portale dell'Istat.

### FASI DEL PROCESSO DI DEFINIZIONE DEL GLOSSARIO UNICO

La prima fonte di **acquisizione** dei termini utilizzata nel 2015 è stato il glossario on-line, costituito da un elenco dei termini presenti nell'Annuario Statistico Italiano (ASI) del 1999. Questo è stato

---

<sup>1</sup> Il termine è definito dalla coppia lemma-definizione.

<sup>2</sup> Il riferimento è alla modellazione GSBPM, che descrive e definisce le fasi del processo di produzione delle statistiche.  
<http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>.

arricchito con i termini dell'edizione 2014 dell'ASI, successivamente aggiornati mediante la rete dei referenti tematici dell'ASI del 2015 e del 2016.

Nell'ambito delle pubblicazioni a carattere generale, in aggiunta alle edizioni successive dell'ASI, un'altra fonte è rappresentata dalle edizioni del Rapporto Annuale 2017, 2018, 2019 e 2020.

Relativamente agli ambiti tematici, attualmente sono stati acquisiti per "istruzione" e "lavoro" i glossari da rapporti tematici, pubblicazioni in collaborazione con altri enti e dai seguenti format dei comunicati stampa: Statistiche report, Statistiche focus e Statistiche Flash, pubblicati negli ultimi anni. Sulla base delle fonti utilizzate, sono stati pertanto documentati nel glossario i processi attualmente attivi e/o recentemente dismessi.

Per quanto riguarda i registri, è stata acquisita la terminologia di quelli dei luoghi (Situas e RSBL) ed è in fase di completamento quella del Registro degli individui (RBI). L'attività sul Registro del lavoro (RTL) è invece solo all'inizio.

Altra fonte di acquisizione sono le Classificazioni e code list ufficiali, il glossario riporta i termini di: Titolo di studio e Relazioni di parentela; al momento è in corso il consolidamento di Tipologia familiare, Tipo nucleo e Stato civile.

L'acquisizione dei glossari provenienti dalle fonti sopracitate sta consentendo di documentare in modo sempre più approfondito l'uso dei termini nella fase di diffusione. Questa attività gestisce dunque l'inserimento di nuovi termini, in precedenza non utilizzati nei glossari delle fonti fino a quel momento considerate.

Il processo immediatamente successivo a quello di acquisizione è l'**armonizzazione** che comporta un'attenta analisi dei lemmi e delle definizioni ed un eventuale intervento finalizzato all'integrazione dei contenuti. I glossari acquisiti presentano spesso gli stessi lemmi, ma con definizioni diverse (polisemie) a seconda della tipologia del processo di produzione dei dati (ad esempio censimento o indagine). In alcuni casi nei glossari si presentano anche lemmi che seppur diversi, assumono lo stesso significato (sinonimie). Il processo di armonizzazione si esplica, pertanto, nell'individuazione di concetti impropriamente duplicati oppure nel trattamento di tutte le eventuali incongruenze.

A ciò si aggiunga che la mancanza di omogeneità nella stesura di lemmi e definizioni presenti nei vari glossari rende necessaria la loro **standardizzazione**.

Il lemma, rappresentato da una parola o da un'espressione complessa, va uniformato in base ai seguenti standard concernenti:

- *forma base*: formulazione dei sostantivi al singolare tranne nel caso in cui il loro utilizzo sia riferito ad una collettività;

- *dominio di riferimento*: è indispensabile esplicitarlo per evitare ambiguità nell'uso, ad esempio: Clima di fiducia *delle famiglie* e Clima di fiducia *delle imprese italiane*;
- *nomi delle classificazioni*: espresse in italiano anche se internazionali;
- *tipologia del processo di produzione dei dati*: da esplicitare nel caso di lemmi uguali, ma con definizioni diverse ad es.: Abitazione (censimento della popolazione e delle abitazioni 2011) e Abitazione (rilevazione dei permessi di costruire);
- *locuzioni tecniche ed espressioni complesse*: rimangono invariate nel loro ordine naturale ad es.: *tasso di natalità* e *non natalità tasso di*;
- *acronimi, sigle e abbreviazioni*: non rappresentano dei lemmi, ma un loro attributo.

La definizione, che ha la funzione di descrivere il concetto designato dal lemma, va formulata in modo che risulti immediatamente chiaro il dominio di appartenenza e uso del termine. Eventuali spiegazioni vanno collocate successivamente, così come anche eventuali specifiche o usi peculiari. La formulazione non deve ripetere al principio della frase il termine medesimo e deve essere:

- sintetica, ma completa;
- con termini consolidati e non ambigui;
- priva di abbreviazioni e acronimi;
- espressa in senso positivo;
- non circolare.

Ancora, altri standard all'interno della definizione possono riguardare:

- le citazioni degli atti normativi o di parte di essi: si è preso come riferimento la "Guida alla redazione degli atti amministrativi", a cura del gruppo di lavoro promosso da Istituto di teoria e tecniche dell'informazione giuridica e Accademia della Crusca;
- la coerenza con la legislazione vigente: soprattutto per quei lemmi appartenenti ad ambiti oggetto di frequenti interventi legislativi (ad es. il lavoro);
- le citazioni di parole in lingue diverse dall'italiano: in carattere corsivo;
- gli acronimi: solo l'iniziale in lettera maiuscola e senza punti.

I termini del GU possono nel tempo modificarsi, estinguersi oppure essere di nuova introduzione. L'attività di **monitoraggio** ha l'obiettivo di registrare queste variazioni e riportarle nel database di gestione che conserva anche alcune metainformazioni quali i referenti, la fonte, la data di validazione, il processo di riferimento ed altre.

L'**aggiornamento** del GU permette di gestire:

- le variazioni delle definizioni dei termini, ad es. per una maggior conformità alla legislazione vigente oppure per una variazione intervenuta nei processi di produzione;
- l'acquisizione di nuovi termini derivanti ad es. da nuovi processi;

- le cancellazioni dei termini divenuti obsoleti perché non più in uso nei processi di produzione statistica.

## PROSPETTIVE FUTURE

Attualmente è prioritario completare il glossario on-line con la documentazione di tutti i termini relativi alla diffusione, coerentemente con l'esigenza di fornire un supporto esaustivo ed efficace per l'utenza esterna che accede all'informazione statistica mediante il portale dell'Istituto.

Nel prossimo futuro il GU sarà arricchito anche con termini derivanti da altre fonti con l'obiettivo di dotare l'Istituto di una terminologia specialistica il più possibile completa e sarà necessario curare l'uso di questi termini nelle altre fasi del processo di produzione statistica.

È prevista la pubblicazione di una versione del glossario on-line in lingua inglese.

La progettazione del nuovo sistema di metadati infine prevede l'acquisizione dei termini della raccolta e il loro collegamento ai diversi oggetti che caratterizzano il processo di produzione statistica (variabili, unità di analisi e code list), in modo da consentire l'attivazione di sistemi metadata driven.