



Istituto Nazionale
di Statistica

MANUALE PER LA REPORTISTICA DI QUALITÀ NEL SISTAN





MANUALE PER LA REPORTISTICA DI QUALITÀ NEL SISTAN

A cura di Andrea Bruni.

Revisione redazionale: Lara Parisella.

Attività editoriali: Claudio Bava e Manuela Marrone.

Grafica: Sofia Barletta.

ISBN 978-88-458-2181-3

© 2025

Istituto Nazionale di Statistica
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti
pubblicati sono soggetti alla licenza
Creative Commons - Attribuzione - versione 4.0.
<https://creativecommons.org/licenses/by/4.0/deed.it>

È dunque possibile riprodurre, distribuire,
trasmettere e adattare liberamente dati e analisi
dell'Istituto nazionale di statistica, anche a scopi
commerciali, a condizione che venga citata la fonte.

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

INDICE

	Pag.
Presentazione	5
Introduzione	7
1. Definizione del <i>quality report</i> italiano	9
1.1 Percorso e strumenti del <i>framework</i> italiano per la qualità	9
1.2 Standard europei per il <i>quality reporting</i>	11
1.3 Progettazione del <i>quality report</i> per il Sistan	13
1.4 Standard SDMX per la gestione e la diffusione dei metadati	16
1.5 Riutilizzo dei metadati del Programma statistico nazionale	18
1.6 Prossimi passi per la qualità nel Sistan	19
2. Guida alla compilazione	21
2.1 Contatti	21
2.2 Aggiornamento dei metadati	22
2.3 Descrizione dei dati	23
2.3.1 <i>Caratteristiche dei dati</i>	23
2.3.2 <i>Classificazioni</i>	24
2.3.3 <i>Variabili statistiche</i>	25
2.3.4 <i>Unità statistica</i>	26
2.3.5 <i>Popolazione statistica</i>	27
2.3.6 <i>Territorio di riferimento</i>	28
2.3.7 <i>Copertura temporale</i>	29
2.3.8 <i>Unità di misura</i>	30
2.3.9 <i>Periodo temporale di riferimento</i>	30
2.4 Descrizione del processo statistico	31
2.4.1 <i>Fonte dei dati</i>	31
2.4.2 <i>Periodicità della raccolta dei dati</i>	37
2.4.3 <i>Periodicità della diffusione</i>	37
2.4.4 <i>Raccolta dei dati</i>	38
2.4.5 <i>Trattamento dei dati</i>	42
2.5 Mandato istituzionale	44
2.6 Politica della riservatezza	45

	Pag.
2.7 Gestione della qualità	48
2.8 Pertinenza	50
2.9 Accuratezza	53
2.9.1 <i>Indagini con campionamento probabilistico</i>	54
2.9.2 <i>Indagini con campionamento non probabilistico</i>	57
2.9.3 <i>Indagini censuarie</i>	57
2.9.4 <i>Statistiche basate su dati amministrativi</i>	57
2.9.5 <i>Processi multifonte</i>	58
2.10 Attendibilità - Revisione dei dati	62
2.11 Tempestività e puntualità	64
2.12 Comparabilità nel tempo	65
2.13 Accessibilità e chiarezza	66
2.13.1 <i>Metodi e formati di diffusione</i>	67
2.13.2 <i>Documentazione metodologica</i>	71
2.14 Costi e onere per i rispondenti	72
2.15 Commenti	74
Appendice 1 - Esempio di <i>quality report</i> precompilato	75
Appendice 2 - Generazione SDMX del <i>quality report</i>	77
Riferimenti bibliografici	79

PRESENTAZIONE

Il passaggio dalla produzione di dati basata principalmente su indagini dirette a un nuovo paradigma che prevede anche lo sfruttamento di fonti amministrative, lavorate e trasformate in registri, sta caratterizzando, e lo farà sempre più, il lavoro degli statistici ufficiali.

Tra le conseguenze di tale passaggio si evidenziano, dal lato delle relazioni interistituzionali, l'inevitabile necessità di "dialogare" con un linguaggio comune e, dal lato della produzione, la possibilità di avere dati granulari. Ciò arricchisce l'ambito in cui si muove la statistica ufficiale sia nei confronti degli utenti, che possono sfruttare misure territoriali fini e tempestive, sia nei rapporti tra istituzioni, che devono necessariamente sviluppare modelli sempre più efficienti di interoperabilità.

In tale contesto, quindi, la fruibilità dei dati statistici è di fondamentale importanza per assicurare che quanto raccolto, analizzato e diffuso possa essere effettivamente utilizzato e valorizzato da una vasta gamma di utenti: non solo ricercatori e statistici, ma anche i responsabili delle politiche (secondo il principio del "comprendere per decidere"), le istituzioni educative a vari livelli, le imprese e i cittadini. Ciò implica garantire l'accesso simultaneo sia ai dati sia ai metadati con la massima facilità per tutti gli attori in scena.

La disponibilità di informazioni di dettaglio sulle fonti dei dati, cioè i metadati (i concetti e le definizioni, le classificazioni, le metodologie e le procedure di calcolo), è essenziale per assicurare che i dati stessi possano essere interpretati correttamente e utilizzati efficacemente, migliorando così la qualità delle analisi e delle decisioni basate sulle evidenze qualitative e quantitative.

A tale proposito, il rispetto delle raccomandazioni internazionali, come quelle emerse durante l'ultimo *round* di *peer review* dell'Istat a fine 2022, sottolinea l'importanza di affinare gli strumenti per il coordinamento del Sistema statistico nazionale (Sistan), compresi l'attuazione e il monitoraggio degli standard europei e la fornitura di assistenza metodologica e metadati accurati, contribuendo a una maggiore coerenza e comparabilità delle statistiche nello spazio e nel tempo.

Sebbene tale azione di miglioramento sia rivolta principalmente alle autorità statistiche nazionali nel loro perseguitamento della conformità al Codice delle statistiche europee (*European Statistics Code of Practice*), è essenziale notare che l'approccio dell'Istat, coerente con la politica per la qualità della produzione statistica dell'Istat, va oltre la semplice conformità. L'Istat, infatti, interpreta gli input dal contesto europeo in un senso ampio, abbracciando un'applicazione completa delle raccomandazioni, per quanto possibile, nel Sistan tutto. In tale modo, l'Istat mira a promuovere una cultura dell'eccellenza e dell'integrità nella pratica statistica a tutti i livelli e in tutti i domini di interesse, migliorando così l'affidabilità, la credibilità e la rilevanza delle informazioni prodotte nell'intero panorama statistico italiano.

Garantire la qualità dei dati prodotti rappresenta, infatti, un pilastro fondamentale per favorire la fiducia nelle statistiche pubbliche. Per questo motivo, un sistema di governance efficace deve stabilire requisiti standardizzati e riconosciuti, facilitando l'identificazione dei meccanismi per controllare i processi statistici, misurare la qualità e comunicarla in modo trasparente agli utenti esterni.

Il ricorso agli standard dell'Unione europea costituisce la soluzione naturale a questa necessità, come peraltro suggerito dagli organi di governo e vigilanza del Sistan, il Comitato di indirizzo e coordinamento dell'informazione statistica (Comstat) e la Commissione per la garanzia della qualità dell'informazione statistica (Cogis).

Il presente documento porta a compimento proprio l'allineamento al quadro concettuale per la qualità sviluppato nell'ambito del Sistema statistico europeo (*European Statistical System - ESS*), con la definizione della struttura di *quality reporting* derivata dallo standard *SIMS (Single Integrated Metadata Structure)* e, contestualmente, delle istruzioni per una corretta compilazione della reportistica, calibrando sulla realtà nazionale l'*European Statistical System Handbook for Quality and Metadata Reports*.

Matteo Mazziotta
Istat, Direttore centrale Sistan e territorio

INTRODUZIONE¹

Il Sistan rappresenta un elemento fondamentale per lo sviluppo della società italiana, poiché opera, ai vari livelli di competenza, per assicurare informazioni adeguate e tempestive a tutti gli operatori pubblici e privati, idonee a soddisfare le esigenze espresse per fini scientifici, per l'efficiente gestione e il controllo dell'operato della Pubblica amministrazione. In questo sistema, all'Istat viene affidato il ruolo di coordinamento, indirizzo e controllo, per assicurare la completezza, la qualità e la rispondenza delle statistiche ai bisogni di conoscenza espressi dal Paese.

Per migliorare l'informazione statistica è necessario dotarsi di strumenti adeguati alla valutazione degli assetti istituzionali, dei processi e dei prodotti; occorre, cioè, potenziare il sistema di gestione della qualità della statistica pubblica. Nel novero delle migliori pratiche riconosciute a livello europeo (e internazionale), la documentazione della qualità in ambito statistico è fondamentale: essa fornisce, in maniera standardizzata attraverso *template* predefiniti, le evidenze necessarie per dimostrare che i dati sono stati raccolti, analizzati e diffusi secondo procedure rigorose e disciplinate; fornisce, inoltre, i metadati necessari a supportare i dati, facilitando così un utilizzo corretto delle informazioni (Istat 2021).

Il presente ebook ha, dunque, come obiettivo la definizione di un *template* per la documentazione della qualità delle statistiche ufficiali prodotte dai soggetti del Sistan. Viene messo a disposizione dei soggetti del Sistema non solo come strumento concettuale, ma anche, in termini concreti, all'interno di un applicativo che ne faciliti la compilazione e la messa in produzione.

I principali benefici di questa pratica sono di seguito elencati.

- **Trasparenza e responsabilità:** quando le metodologie, le fonti di dati e i prodotti statistici sono chiaramente descritti, gli utenti (decisori politici, ricercatori, aziende e cittadini) possono comprendere meglio come sono stati prodotti i dati e quali sono le limitazioni. Ciò aiuta a mantenere le istituzioni statistiche responsabili e favorisce una maggiore fiducia pubblica.
- **Valutazione e miglioramento continuo:** è possibile identificare punti di forza e di debolezza nei processi statistici attraverso un'analisi sistematica, che valuta l'efficacia delle metodologie adottate, allo scopo di identificare spazi di ottimizzazione. Questo approccio favorisce un ciclo di miglioramento continuo, contribuendo a una statistica più accurata e pertinente.
- **Comparabilità e coerenza:** in un contesto globale, è importante che i dati siano confrontabili. La documentazione aiuta a garantire che le stesse definizioni e metodologie siano utilizzate nel tempo e tra diversi contesti: ciò facilita il confronto tra dati provenienti da diverse fonti o enti, rendendo più facile l'analisi e l'interpretazione delle informazioni.
- **Supporto alle decisioni politiche:** le politiche pubbliche si basano su dati statistici. Una buona documentazione della qualità permette ai decisori di comprendere le caratteristiche dei dati su cui si fondano le loro decisioni, aspetto particolarmente importante in situazioni di crisi o di cambiamento rapido, dove le decisioni devono essere basate su informazioni solide e affidabili.
- **Formazione e consapevolezza:** documentare la qualità dei dati contribuisce anche alla formazione degli operatori statistici, aiuta i professionisti a comprendere l'importanza

¹ L'Introduzione è stata redatta da Andrea Bruni.

della qualità e a sviluppare competenze e, inoltre, aumenta la consapevolezza degli utenti sui criteri di qualità, consentendo una valutazione critica delle informazioni utilizzate.

Garantire la qualità delle statistiche è una pietra miliare nel percorso di promozione della fiducia nelle statistiche pubbliche: un sistema di governance solido ed efficace per le statistiche ufficiali deve, infatti, stabilire requisiti standardizzati e riconosciuti, facilitando l'identificazione di meccanismi per controllare i processi statistici, misurare la qualità e comunicarla agli utenti esterni.

Il modello di riferimento è il quadro concettuale sviluppato per il Sistema statistico europeo, con le sue linee di indirizzo, definizioni e raccomandazioni. In sintesi, seguire tali standard è fondamentale per definire valori comuni e parametri condivisi in un sistema statistico nazionale, perché non solo garantisce la qualità e l'affidabilità dei dati, ma ne promuove anche la trasparenza e stimola l'innovazione. Investire nella conformità a questi standard rappresenta un passo decisivo per migliorare la gestione della cosa pubblica, contribuendo così al benessere sociale ed economico del Paese, in quanto un sistema statistico robusto e ben integrato a livello europeo è una risorsa preziosa per affrontare le sfide contemporanee e costruire un futuro sostenibile.

Il presente ebook è strutturato come segue: il Capitolo 1 illustra il percorso di definizione del *template* e, in particolare, vengono introdotte le relazioni con il quadro di riferimento per la qualità elaborato a livello italiano e con le linee di intervento per il Sistan elaborate nell'ambito della politica per la qualità dell'Istat; successivamente si passa a descrivere la procedura adottata per predisporre la versione finale del *template* italiano, l'implementazione tramite procedure fondate sullo standard *Statistical Data and Metadata eXchange* (SDMX) e le operazioni volte a minimizzare l'onere per la compilazione del *quality report*. Nel Capitolo 2 sono riportati i concetti selezionati, con annessa definizione e linee guida per la compilazione. Per facilitare l'operatività, sono altresì stati inseriti esempi tratti dalle effettive pratiche dei soggetti del Sistan in materia di *quality reporting*.

1. DEFINIZIONE DEL *QUALITY REPORT ITALIANO*¹

1.1 Percorso e strumenti del *framework* italiano per la qualità

La legislazione, a mezzo del decreto legislativo 6 settembre 1989, n. 322, sancisce tra i compiti dell'Istat la funzione di indirizzo e coordinamento in ambito Sistan, la rete di soggetti pubblici e privati che fornisce al Paese e agli organismi internazionali l'informazione statistica ufficiale.

L'Istituto Nazionale di Statistica è incaricato di coordinare il Sistan, composto da oltre tremila uffici, sotto la supervisione dell'organo di governo del sistema, il Comitato di indirizzo e coordinamento per l'informazione statistica (Comstat), presieduto dal Presidente dell'Istat. Inoltre, la Commissione per la garanzia della qualità dell'informazione statistica (Cogis), in quanto organo collegiale indipendente, vigila sulle attività all'interno del Sistan.

Data la sua posizione centrale, l'Istat assume un ruolo di guida indispensabile, in particolare nel fornire assistenza tecnica e supervisione per garantire la conformità delle statistiche prodotte e diffuse agli standard di qualità, in linea con le indicazioni di Comstat e Cogis, e con il quadro di riferimento europeo sulla qualità, fondato sui principi e sugli indicatori del Codice delle statistiche europee.

Adottato dal Comitato del Sistema statistico europeo nel novembre 2017 (*European Statistical System 2017*), il Codice delle statistiche europee rappresenta uno strumento di autoregolamentazione condiviso dalle autorità statistiche dei paesi europei, che, periodicamente, valutano il livello di aderenza di ciascun sistema statistico nazionale ai principi del Codice stesso, attraverso il meccanismo delle *peer review*.

L'adozione del Codice è prevista dalla cosiddetta legge statistica europea, il regolamento (CE) n. 223/2009, recentemente modificata dal regolamento (UE) 2024/3018 del Parlamento europeo e del Consiglio, del 27 novembre 2024, entrato in vigore il 26 dicembre 2024, per tenere conto dell'evoluzione nella produzione delle statistiche ufficiali.

Non solo l'Istat, ma tutte le autorità statistiche nazionali responsabili dello sviluppo, della produzione e diffusione di statistiche europee – le cosiddette ONA (*Other National Authorities*), attualmente tredici tra Ministeri e altri enti, membri del Sistan – seguono il quadro di riferimento europeo per la garanzia della qualità. Tale quadro è costituito non solo dal Codice delle statistiche europee, ma anche dal *Quality Assurance Framework* (QAF) del Sistema statistico europeo, approvato dal Comitato del Sistema statistico europeo nel maggio 2019 (*European Statistical System 2019*): per ogni principio e indicatore del Codice, viene riportata una serie di metodi utili a guidare l'implementazione del principio e allo stesso tempo a valutarne l'aderenza.

Fin dalla prima emanazione del Codice delle statistiche europee, nel 2005, l'Istat ha dedicato un impegno consistente alla costruzione di un linguaggio comune sulla qualità, prendendo a riferimento quanto sviluppato nell'ambito del Sistema statistico europeo, trasferendo su scala nazionale contenuti e modalità di applicazione dei dettami europei.

¹ Il Capitolo è stato redatto da: Paola Giordano (paragrafi 1.1, 1.3 e 1.5), Giorgia Simeoni (paragrafo 1.1), Gabriele Ascani (paragrafo 1.2), Loredana Amilcare (paragrafi 1.3 e 1.5), Simone Coccia (paragrafo 1.4) e Andrea Bruni (paragrafo 1.6).

Nel 2010 viene approvata, infatti, la prima versione del Codice italiano per la statistica ufficiale, che rispecchia i principi del Codice europeo ma adattandoli alla realtà del Sistan, e viene avviata una rilevazione esplorativa sulla qualità presso i soggetti del sistema per verificare l'applicazione dei principi di qualità nella produzione delle statistiche.

Il percorso per la definizione di buone pratiche continua con lo sviluppo di ulteriori strumenti di indirizzo e di coordinamento: ad esempio, il Comstat emana le *Linee guida per il miglioramento della qualità della diffusione delle statistiche ufficiali da parte dei soggetti del Sistema statistico nazionale* (Comstat 2011), con l'obiettivo di fornire concrete indicazioni operative agli uffici di statistica, per promuovere e facilitare l'effettiva attuazione del Codice italiano per la statistica ufficiale. Un altro documento chiave del Comstat è costituito dall'*Atto di indirizzo n. 3* (Comstat 2018), che definisce i requisiti di qualità, le forme e le modalità di diffusione dei dati che gli enti e gli uffici del Sistan raccolgono e trasmettono all'Istituto Nazionale di Statistica per l'esecuzione dei lavori del Programma statistico nazionale (PSN). Di particolare rilevanza è l'articolo 4, che sottolinea l'importanza di una diffusione in forma chiara e comprensibile dei dati statistici e della presenza di opportuni metadati a corredo.

Nello stesso anno, l'Istat sviluppa anche le *Linee guida per la qualità delle statistiche prodotte dal Sistema statistico nazionale* (Istat 2018), un ebook contenente i principi e le buone pratiche da applicare nei processi statistici basati sulle indagini o sui dati di fonti amministrative. Le linee guida vengono adottate, tra il 2018 e il 2021, come riferimento per la conduzione da parte dell'Istat di un programma di valutazione, attraverso procedure di *audit* e autovalutazione, sui processi statistici alla base delle statistiche europee prodotte dalle ONA. Tale programma contribuisce a instaurare un clima di scambio e collaborazione tra l'Istat e le ONA per l'obiettivo comune del miglioramento della qualità delle statistiche europee.

Nel 2021 l'Istat rafforza il suo impegno per la qualità con l'approvazione della *Nuova politica per la qualità della produzione statistica* (Istat 2021), in cui dedica particolare attenzione al ruolo di coordinamento del Sistan, esplicitando la strategia che comporta l'allestimento di una serie di strumenti specificamente calibrati al contesto nazionale, concepiti per funzionare in parallelo e in modo complementare rispetto a quelli europei, e pianifica la produzione di appositi documenti che realizzano in concreto la strategia.

Nel frattempo, nel 2017, a livello comunitario, il Codice europeo viene rivisto e rafforzato proprio nella componente del coordinamento del Sistema statistico nazionale, attraverso l'introduzione del principio 1bis "Coordinamento e Cooperazione"; successivamente, nel 2019, viene completata la versione 2.0 del QAF che include metodi per ogni principio e indicatore, contrariamente alla versione precedente che non aveva carattere di esaustività.

Il percorso di adeguamento del quadro di riferimento italiano al *framework* europeo conduce alla pubblicazione della seconda versione del Codice italiano per la qualità delle statistiche ufficiali, attraverso l'emanazione della Direttiva n. 12/Comstat pubblicata nella Gazzetta Ufficiale Serie generale n. 23 del 29 gennaio 2022, e alla produzione della *Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali* (Istat 2023), che rappresenta la versione italiana speculare al QAF europeo.

L'adozione di standard coerenti con quelli europei consente di promuovere al massimo grado l'indipendenza, l'integrità e la responsabilità dei produttori di dati pubblici: la conformità a enunciati uniformi e criteri rigorosi garantisce la valorizzazione della funzione statistica all'interno del Sistema statistico nazionale. Tale iniziativa di trasparenza incrementa la visibilità del Sistan e rafforza la fiducia nelle istituzioni e la credibilità delle statistiche prodotte.

1. Definizione del *quality report* italiano

In questo processo di convergenza, un ultimo passo si rendeva necessario: prendere le mosse dagli standard europei in materia di *quality reporting* per definire un *template* italiano per la reportistica sulla qualità, con annesse linee guida per la compilazione. Tale esigenza era stata altresì sollevata dalla Cogis, dal Comstat e dal Ministro della Pubblica amministrazione durante l'iter di approvazione del Codice italiano per la qualità delle statistiche ufficiali.

A livello europeo, lo standard di riferimento per la compilazione e lo scambio di metadati e di informazioni sulla qualità è rappresentato dal *Single Integrated Metadata Structure* (SIMS) v2.0 (Eurostat 2015), che è ulteriormente dettagliato nello *European Statistical System Handbook for Quality and Metadata Reports* (Eurostat 2021). La conformità con il *framework* europeo potrà dunque dirsi pienamente raggiunta con la predisposizione del presente ebook, che si pone quale documento operativo di riferimento, in ambito Sistan, per l'appropriata raccolta del *set* di metadati essenziale per un'idonea interpretazione dei dati diffusi. In particolare, ispirandosi allo *ESS Handbook for Quality and Metadata Reports*, gli obiettivi di questo documento italiano sono tre:

- armonizzare la documentazione prodotta per diversi processi statistici e da diversi enti del Sistan;
- facilitare l'uso e il confronto sia per i processi statistici sia per gli output;
- garantire una diffusione chiara e standardizzata dei dati, consentendone un corretto utilizzo e interpretazione.

Allineando le pratiche di documentazione e garantendo la coerenza dei *quality report*, questo manuale ha l'obiettivo di migliorare la qualità e la fruibilità dell'informazione statistica prodotta nell'ambito del Sistan, promuovendo così una maggiore aderenza al Codice italiano per la qualità delle statistiche ufficiali. In questa prospettiva, la sua realizzazione dà anche attuazione a quanto previsto nella politica per la qualità dell'Istat e rappresenta un utile strumento per definire i prossimi passi nel percorso di miglioramento continuo della qualità delle statistiche prodotte dal Sistan.

1.2 Standard europei per il *quality reporting*

Comprendere come un dato statistico è stato prodotto, quali sono i suoi possibili usi e quale è il suo livello di qualità rappresenta un aspetto di fondamentale importanza per molti utenti della statistica ufficiale. Tali informazioni costituiscono i cosiddetti metadati referenziali, ovvero metadati che accompagnano il dato statistico e ne chiariscono, con diversi dettagli, i contenuti informativi, le metodologie impiegate per la loro produzione e gli attributi di qualità. La caratteristica distintiva dei metadati referenziali è che possono essere letti separatamente dal prodotto statistico cui si riferiscono; questo tipo di informazione, infatti, è in genere raccolto in appositi documenti. Nell'ambito del Sistema statistico europeo tali documenti sono denominati "relazioni sulla qualità" o *quality report*.

I *quality report* rappresentano un elemento essenziale per garantire la trasparenza e la comparabilità dei dati statistici, offrendo una valutazione critica degli aspetti di accuratezza, coerenza, tempestività e accessibilità dei dati, permettendo così di apprezzare pienamente il valore e l'affidabilità delle statistiche prodotte. In questo contesto, il mantenimento di standard elevati nella documentazione è cruciale per sostenere la fiducia e l'uso efficace delle informazioni statistiche: si è quindi rapidamente avvertita l'esigenza di armonizzare sia la struttura sia i contenuti della documentazione contenente i metadati referenziali e le informazioni sulla qualità dell'output.

Parallelamente, si è ritenuto opportuno distinguere le informazioni destinate agli utenti finali da quelle che i produttori delle statistiche sono in genere vincolati a fornire in ottemperanza ai regolamenti e alle normative vigenti. A tale scopo sono state inizialmente definite due strutture di *report* distinte, integrate poi in un unico *superset* per garantire una maggiore coerenza e facilità di utilizzo. Di conseguenza, come accennato nel paragrafo precedente, nel 2015 è entrato a regime il *template* SIMS che, nella versione 2.0, rappresenta l'attuale sistema di riferimento per tutte le attività di *quality reporting* all'interno del Sistema statistico europeo, come tra l'altro confermato dalla raccomandazione (UE) 397/2023, nella quale si invitano gli Stati membri a garantire che le rispettive autorità statistiche nazionali applichino i concetti statistici elencati nell'ultima versione della struttura SIMS per la compilazione dei metadati referenziali e dei *quality report*.

L'impianto riprende tutte le informazioni contenute nelle precedenti strutture, organizzandole gerarchicamente in una serie di concetti e sottoconcetti. La versione attuale ne include rispettivamente 19 e 80 (Prospetto 1.1); ciascuna di queste voci è descritta nel dettaglio nello *ESS Handbook for Quality and Metadata Reports*, la cui più recente edizione è del febbraio 2021.

Prospetto 1.1 - La struttura di SIMS

Section	Concept name	Section	Concept name	Section	Concept name
S.1	Contact	S.8.2	<i>Release calendar access</i>	S.13.3.3.2	A5. <i>Item non-response</i>
S.1.1	<i>Contact organisation</i>	S.8.3	<i>User access</i>	S.13.3.4	<i>Processing error</i>
S.1.2	<i>Contact organisation unit</i>	S.9	Frequency of dissemination	S.13.3.5	<i>Model assumption error</i>
S.1.3	<i>Contact name</i>	S.10	Accessibility and clarity	S.14	Timeliness and punctuality
S.1.4	<i>Contact person function</i>	S.10.1	<i>News release</i>	S.14.1	<i>Timeliness</i>
S.1.5	<i>Contact mail address</i>	S.10.2	<i>Publications</i>	S.14.1.1	TP1. <i>Time lag - first results</i>
S.1.6	<i>Contact email address</i>	S.10.3	<i>Online database</i>	S.14.1.2	TP2. <i>Time lag - final results</i>
S.1.7	<i>Contact phone number</i>	S.10.3.1	AC1. <i>Data tables - consultations</i>	S.14.2	<i>Punctuality</i>
S.1.8	<i>Contact fax number</i>	S.10.4	<i>Micro-data access</i>	S.14.2.1	TP3. <i>Punctuality - delivery and publication</i>
S.2	Metadata update	S.10.5	<i>Other</i>	S.15	Coherence and comparability
S.2.1	<i>Metadata last certified</i>	S.10.5.1	AC 2. <i>Metadata - consultations</i>	S.15.1	<i>Comparability - geographical</i>
S.2.2	<i>Metadata last posted</i>	S.10.6	<i>Documentation on methodology</i>	S.15.1.1	CC1. <i>Asymmetry for mirror flows statistics - coefficient</i>
S.2.3	<i>Metadata last update</i>	S.10.6.1	AC 3. <i>Metadata completeness - rate</i>	S.15.2	<i>Comparability - over time</i>
S.3	Statistical presentation	S.10.7	<i>Quality documentation</i>	S.15.2.1	CC2. <i>Length of comparable time serie</i>
S.3.1	<i>Data description</i>	S.11	Quality management	S.15.3	<i>Coherence- cross domain</i>
S.3.2	<i>Classification system</i>	S.11.1	<i>Quality assurance</i>	S.15.3.1	<i>Coherence - sub annual and annual statistics</i>
S.3.3	<i>Sector coverage</i>	S.11.2	<i>Quality assessment</i>	S.15.3.2	<i>Coherence- National Accounts</i>
S.3.4	<i>Statistical concepts and definitions</i>	S.12	Relevance	S.15.4	<i>Coherence - internal</i>
S.3.5	<i>Statistical unit</i>	S.12.1	<i>User needs</i>	S.16	Cost and burden
S.3.6	<i>Statistical population</i>	S.12.2	<i>User satisfaction</i>	S.17	Data revision
S.3.7	<i>Reference area</i>	S.12.3	<i>Completeness and R1. Data completeness</i>	S.17.1	<i>Data revision - policy</i>
S.3.8	<i>Time coverage</i>	S.12.3.1	R1. <i>Data completeness</i>	S.17.2	<i>Data revision - practice</i>
S.3.9	<i>Base period</i>	S.13	Accuracy and reliability	S.17.2.1	A6. <i>Data revision - average size</i>
S.4	Unit of measure	S.13.1	<i>Overall accuracy</i>	S.18	Statistical processing
S.5	Reference period	S.13.2	<i>Sampling error</i>	S.18.1	<i>Source data</i>
S.6	Institutional mandate	S.13.2.1	A1. <i>Sampling errors</i>	S.18.2	<i>Frequency of data collection</i>
S.6.1	<i>Legal acts and other agreements</i>	S.13.3	<i>Non-sampling error</i>	S.18.3	<i>Data collection</i>
S.6.2	<i>Data sharing</i>	S.13.3.1	<i>Coverage error</i>	S.18.4	<i>Data validation</i>
S.7	Confidentiality	S.13.3.1.1	A2. <i>Over-coverage - rate</i>	S.18.5	<i>Data compilation</i>
S.7.1	<i>Confidentiality - policy</i>	S.13.3.1.2	A3. <i>Common units - proportion</i>	S.18.5.1	A7. <i>Imputation - rate</i>
S.7.2	<i>Confidentiality - data treatment</i>	S.13.3.2	<i>Measurement error</i>	S.18.6	<i>Adjustment</i>
S.8	Release policy	S.13.3.3	<i>Non response error</i>	S.18.6.1	<i>Seasonal adjustment</i>
S.8.1	<i>Release calendar</i>	S.13.3.3.1	A4. <i>Unit non-response</i>	S.19	Comment

1. Definizione del *quality report* italiano

Come si può osservare, SIMS si apre con alcune informazioni di base sull'ente statistico produttore del dato, per poi offrire una panoramica esaustiva delle caratteristiche intrinseche del prodotto statistico, includendo dettagli sui sistemi di classificazione utilizzati, le unità statistiche, la popolazione di riferimento, le unità di misura utilizzate e così via. Queste prime sezioni coprono i primi cinque concetti del SIMS; i concetti dal sesto all'ottavo si concentrano su aspetti istituzionali e normativi che regolano il processo statistico, includendo le politiche di riservatezza e le pratiche di rilascio dei dati.

Le cosiddette dimensioni della qualità, ovvero i criteri di qualità che un prodotto statistico deve soddisfare secondo quanto stabilito dal regolamento (CE) n. 223/2009, sono introdotte a partire dal decimo concetto (*Accessibilità e chiarezza*). Dopo un concetto dedicato alla gestione della qualità, le altre dimensioni di qualità si presentano a partire dalla dodicesima voce del *template*; in particolare, l'*Accuratezza*, comprensiva dei suoi numerosi sottoconcetti, è esplorata in profondità nella sezione tredici. La parte conclusiva del *template* è dedicata ad aspetti complementari ma non meno importanti per la valutazione complessiva della qualità del prodotto statistico: questo segmento include i costi associati al processo e il cosiddetto “fastidio statistico” imposto sui rispondenti, che rappresentano dei vincoli essenziali per la qualità. Sono inoltre incluse la politica di revisione, la descrizione del processo che ha generato i dati, dalle fonti della raccolta dati fino alle metodologie utilizzate per la validazione dei dati, nonché le metodologie relative alla produzione degli output statistici, incluse le procedure per gli aggiustamenti da effetti stagionali. Chiude il *template* uno spazio libero riservato ai commenti da parte del compilatore, offrendo un'opportunità per ulteriori osservazioni o chiarimenti.

1.3 Progettazione del *quality report* per il Sistan

Il presente documento intende guidare i soggetti del Sistan verso la produzione di *quality report* standardizzati dei processi e dei prodotti statistici. Nel concreto il manuale, che trova fondamento negli standard europei in materia, intende promuovere nel Sistan un'armonizzazione della reportistica tra processi statistici e tra soggetti del Sistan, facilitando così i confronti relativi sia ai processi sia ai prodotti. Considerata l'eterogeneità delle realtà statistiche nel Sistema statistico nazionale, si è ritenuto opportuno non adottare in maniera rigida e automatica gli standard europei, bensì procedere a un esame capillare delle pratiche relative alla diffusione dei metadati presso i vari enti, per derivare il sottoinsieme di metadati più utile e adatto per gli specifici scopi nazionali.

L'attività si è svolta secondo le seguenti direttive:

- analisi delle attuali pratiche degli enti del Sistan in materia di *quality reporting*, con particolare attenzione ai metadati strutturali e referenziali;
- mappatura delle metainformazioni disponibili presso gli enti del Sistan in relazione ai concetti facenti parte dello standard europeo SIMS;
- selezione dei contenuti da inserire nel modello italiano di *quality reporting* standardizzato.

A valle dell'individuazione del sottoinsieme utile ai fini Sistan, in analogia con l'*ESS Handbook for Quality and Metadata Reports*, sono state predisposte le linee guida operative per la compilazione del *quality report* secondo il formato standardizzato, descritte nel Capitolo 2 del presente ebook.

L'analisi si è concentrata sui metadati presenti nel Programma statistico nazionale (PSN), nel quale i lavori statistici sono classificati in differenti tipologie: Statistiche (Sta), Studi progettuali (Stu) e Sistemi informativi statistici (Sis).

Nello specifico, il PSN 2020-2022 aggiornamento 2022 comprende 805 lavori:

- 714 Sta (274 con titolarità Istat e 440 con titolarità di altri enti del Sistan);
- 62 Stu (36 con titolarità Istat e 26 con titolarità di altri enti Sistan);
- 29 Sis (10 con titolarità Istat e 19 con titolarità di altri enti Sistan).

Occorre prestare attenzione al fatto che alcuni metadati, così come alcune definizioni dei concetti presenti nel PSN utilizzato come riferimento per l'individuazione dei concetti da includere nel *template*, sono stati oggetto di modifiche nei PSN successivi.

Al momento della presente analisi, si è posta particolare attenzione alle statistiche non Istat di maggiore rilevanza, conducendo un approfondimento dettagliato su 379 Sta. Nella pratica, sono stati esaminati i metadati relativi ai lavori pubblicati sui siti ufficiali degli enti del Sistan titolari dei lavori medesimi. L'obiettivo di questa fase era raccogliere tutti i metadati disponibili e, al contempo, identificare quali fossero i metadati più rilevanti, e quindi necessari per la pubblicazione, da parte dagli uffici di statistica.

Successivamente, per ogni lavoro statistico, i metadati sono stati mappati con le 99 voci che compongono la struttura del SIMS 2.0, al fine di identificare gli *item* dello standard europeo più ricorrenti nelle pratiche nazionali. Nell'analisi delle 379 Sta, gli esperti dell'Istat sono riusciti a reperire i metadati di 213 di esse (56,2 per cento) sui siti ufficiali esaminati.

L'attività di mappatura con i concetti SIMS è dettagliata nella tavola seguente, nella quale, per comodità di analisi, sono riportate solo le voci con tre o più occorrenze.

Tavola 1.1 - Concetti SIMS nei metadati diffusi dagli enti Sistan non Istat. Anno 2024 (valori assoluti)

Codice	Nome del concetto	Occorrenze
S.1.1	<i>Contact organisation</i>	40
S.1.2	<i>Contact organisation unit</i>	34
S.1.3	<i>Contact name</i>	26
S.1.4	<i>Contact person function</i>	17
S.1.5	<i>Contact mail address</i>	17
S.1.6	<i>Contact email address</i>	28
S.1.7	<i>Contact phone number</i>	21
S.2.3	<i>Metadata last update</i>	3
S.3.1	<i>Data description</i>	202
S.3.2	<i>Classification system</i>	92
S.3.3	<i>Sector coverage</i>	39
S.3.4	<i>Statistical concepts and definitions</i>	148
S.3.5	<i>Statistical unit</i>	152
S.3.6	<i>Statistical population</i>	119
S.3.7	<i>Reference area</i>	116
S.3.8	<i>Time coverage</i>	113
S.3.9	<i>Base period</i>	7
S.4	<i>Unit of measure</i>	48
S.5	<i>Reference period</i>	46
S.6.1	<i>Legal acts and other agreements</i>	92
S.6.2	<i>Data sharing</i>	4
S.7.1	<i>Confidentiality - policy</i>	11
S.7.2	<i>Confidentiality - data treatment</i>	5
S.8.1	<i>Release calendar</i>	40
S.8.2	<i>Release calendar access</i>	8
S.8.3	<i>User access</i>	3

Fonte: Istat

1. Definizione del *quality report* italiano

Tavola 1.1 segue - Concetti SIMS nei metadati diffusi dagli enti Sistan non Istat. Anno 2024 (valori assoluti)

Codice	Nome del concetto	Occorrenze
S.9	<i>Frequency of dissemination</i>	52
S.10.1	<i>News release</i>	71
S.10.2	<i>Publications</i>	69
S.10.3	<i>Online database</i>	102
S.10.4	<i>Microdata access</i>	10
S.10.5	<i>Other</i>	8
S.10.6	<i>Documentation on methodology</i>	29
S.10.7	<i>Quality documentation</i>	4
S.11.1	<i>Quality assurance</i>	10
S.11.2	<i>Quality assessment</i>	12
S.13.1	<i>Overall accuracy</i>	11
S.13.3.3	<i>Non response error</i>	4
S.13.3.3.1	<i>A4. Unit non-response - rate</i>	3
S.14.1	<i>Timeliness and TP2. Time lag - final results</i>	42
S.15.1	<i>Comparability - geographical</i>	3
S.18.1	<i>Source data</i>	132
S.18.2	<i>Frequency of data collection</i>	58
S.18.3	<i>Data collection</i>	162
S.18.4	<i>Data validation</i>	56
S.18.5	<i>Data compilation</i>	35
S.19	<i>Comment</i>	8

Fonte: Istat

I risultati di questo prospetto sono serviti come fondamento per la definizione del *template* italiano, che operativamente si traduce, come si vedrà nel successivo paragrafo, nella creazione dell'artefatto SDMX denominato *Metadata Structure Definition* (MSD).

Dopo un'attenta valutazione è stato, infatti, deciso di focalizzarsi su un insieme ridotto di 37 concetti dei 99 totali presenti nel SIMS 2.0, con l'obiettivo rendere più efficace e sostenibile il processo di redazione del *quality report* italiano. Questa selezione segue una duplice logica: da un lato, sono stati mantenuti i concetti ritenuti chiave a livello europeo, in quanto fondamentali per garantire l'allineamento con lo standard internazionale; dall'altro, si è tenuto conto della frequente ricorrenza di tali concetti nei metadati già in uso dai soggetti del Sistan, valorizzando dunque elementi informativi già consolidati nella prassi nazionale.

Questa strategia ha permesso di raggiungere un equilibrio tra adesione agli standard internazionali, pertinenza operativa per i produttori di statistiche e sostenibilità del processo di compilazione, consentendo di mantenere un elevato livello di dettaglio informativo e, al contempo, ridurre la complessità e il carico di lavoro per gli operatori.

Il seguente Prospetto 1.2 presenta il *template* italiano, del quale si rende contestualmente disponibile anche la versione in lingua inglese. Inoltre, è presente una mappatura con i concetti del SIMS, che permette di cogliere la corrispondenza tra i concetti domestici e i concetti dello standard ESS.

Prospetto 1.2 - *Template* italiano e mappatura con i concetti di SIMS

Versione inglese		Versione italiana		SIMS
First level	Second level	Primo livello	Secondo livello	
Contact	Contact organisation	Contatti	Ente	1.1
	Contact organisation unit		Unità organizzativa	1.2
	Contact name		Responsabile	1.3
	Contact person function		Funzione del responsabile	1.4
	Contact email address		E-mail	1.6
	Contact phone number		Telefono	1.7
Metadata last update		Ultimo aggiornamento		2.3
Statistical presentation	Data description	Descrizione dei dati	Caratteristiche dei dati	3.1
	Classification system		Classificazioni	3.2
	Statistical concepts and definitions		Variabili statistiche	3.4
	Statistical unit		Unità statistica	3.5
	Statistical population		Popolazione statistica	3.6
	Reference area		Territorio di riferimento	3.7
	Time coverage		Copertura temporale	3.8
	Unit of measure		Unità di misura	4
	Reference period		Periodo temporale di riferimento	5
Statistical processing	Source data	Descrizione del processo statistico	Fonte dei dati	18.1
	Frequency of data collection		Periodicità della raccolta dei dati	18.2
	Frequency of dissemination		Periodicità della diffusione	9
	Data collection		Raccolta dei dati	18.3
	Data processing		Trattamento dei dati	18.(4+5)
Institutional mandate		Mandato istituzionale		6
Confidentiality - policy		Politica della riservatezza		7.1
Quality management		Gestione della qualità		11
Relevance		Pertinenza		12
Accuracy		Accuratezza		13
Reliability - Data revision		Attendibilità - Revisione dei dati		17
Timeliness and punctuality		Tempestività e puntualità		14
Comparability - over time		Comparabilità nel tempo		15.2
Accessibility and clarity	Release calendar	Accessibilità e chiarezza	Calendario di diffusione	8.1
	News release		Comunicati stampa	10.1
	Publications		Pubblicazioni	10.2
	Online database		Portale online	10.3
	Other		Altro	10.5
	Documentation on methodology		Documentazione metodologica	10.6
Cost and burden		Costi e onere sui rispondenti		16
Comment		Commenti		19

Fonte: Istat

1.4 Standard SDMX per la gestione e la diffusione dei metadati

SDMX è uno standard internazionale progettato per facilitare lo scambio, la gestione e l'integrazione di dati e di metadati statistici attraverso l'uso di tecnologie informatiche avanzate. Questo standard è stato sviluppato con l'obiettivo di migliorare l'efficienza e la qualità nella produzione, nell'archiviazione e nella diffusione delle informazioni statistiche attraverso un modello informativo composto da diversi artefatti concepiti per descrivere con precisione sia i dati sia i metadati relativi ai vari domini statistici. Questo approccio permette di definire schemi chiari e condivisi per la rappresentazione e il trasferimento delle informazioni, assicurando la loro comprensibilità e fruibilità da parte di diversi attori all'interno della comunità statistica (Cammarota 2023).

Uno degli aspetti fondamentali di SDMX è la sua capacità di standardizzare e industrializzare i processi di diffusione dei dati e metadati, promuovendo l'interoperabilità tra istituzioni nazionali e internazionali, enti statistici, organizzazioni governative e organismi di ricerca.

1. Definizione del *quality report* italiano

L'adozione di questo standard consente di superare le barriere legate all'eterogeneità dei sistemi informativi, garantendo un flusso di dati e metadati più efficiente e conforme ai principi di trasparenza e comparabilità.

A tale proposito, l'Istat ha progettato e realizzato il *framework* “SDMX Istat StatKit”, un insieme di moduli software mediante i quali è possibile gestire l'intero flusso di attività dalla modellazione dei dati fino alla pubblicazione dei metadati associati. Questo comporta una maggiore trasparenza, una migliore qualità di dati e metadati e una riduzione dei costi operativi, rendendo SDMX uno strumento fondamentale nella gestione delle attività statistiche. Nel dettaglio, la piattaforma comprende:

- il *MetaDataManager*, per la modellazione dei dati, l'*editing*, la pubblicazione di metadati descrittivi e dei cataloghi DCAT-AP_IT di dati aperti e glossari;
- il *Data browser*, applicazione web per la navigazione, la visualizzazione e lo scarico di dati e metadati da una o più sorgenti;
- il *Reference metadata tool*, applicazione *desktop* per il supporto alla gestione dei metadati referenziali.

Attraverso quest'ultima funzionalità, si può associare al particolare insieme di dati diffusi (il *dataflow*) un *report* di metadati referenziali, il *quality report*. In una situazione ottimale, si dovrà arrivare ad associare un *report* a ogni *dataflow* esposto. Ovviamente, lo stesso *quality report* potrà essere utilizzato per i diversi *dataflow* che afferiscono al medesimo processo statistico.

Per giungere alla pubblicazione del *quality report*, il flusso di lavoro necessario nello SDMX Istat Toolkit può essere concettualmente suddiviso in tre fasi:

- creazione di un *template*, seguendo la struttura definita nel paragrafo precedente;
- definizione degli artefatti SDMX, specifici per i metadati: creazione della MSD;
- generazione e pubblicazione del *quality report*.

Le prime due fasi vengono svolte un'unica volta e garantiscono l'infrastruttura necessaria all'esecuzione della terza fase, che può esser replicata per tutte le occorrenze necessarie.

I passaggi operativi sono illustrati nel dettaglio nell'Appendice 2.

1.5 Riutilizzo dei metadati del Programma statistico nazionale

Una volta giunti alla definizione finale del *template*, così come illustrato nel paragrafo 1.4, si è proceduto a utilizzare le informazioni riferite ai lavori del PSN già raccolte dall'Istat all'interno del *PsnPlus*, sistema informativo cui i titolari di lavori statistici accedono per confermare, aggiornare, eliminare lavori correnti o per inserire nuovi lavori che siano stati approvati dal circolo di qualità competente per materia.

Per ogni lavoro contenuto nel PSN, l'ente titolare è tenuto a compilare un'apposita scheda, contenente le principali informazioni, standardizzate. Tale scheda (Figura 1.1) funge da *repository* di metadati, in cui i responsabili degli uffici di statistica del Sistan forniscono quelli fondamentali per la definizione del PSN.

Figura 1.1 - Parziale della Sezione 1 della scheda PSN (Sta)

Sezione 1 - Contesto istituzionale e informazioni di base	
1 Titolare	
Dipartimento/Direzione	
Ufficio incaricato	
Tipologia dell'ufficio	
Incarico istituzionale del responsabile del lavoro statistico	
Referente per i contenuti della scheda del lavoro	
Cognome e Nome del referente	
Email del referente	
Recapito telefonico del referente	

Fonte: piattaforma *PsnPlus*

Si è proceduto quindi a mappare la scheda PSN con i 99 *item* di Sims e, quindi, per estensione, con i 37 concetti del *quality report* italiano. Ne è risultato, come illustrato nel dettaglio nel successivo Prospetto 1.3, che ben 25 (il 68 per cento) dei 37 concetti che compongono il *quality report* italiano sono già presenti nella scheda PSN. La presenza di una siffatta quota di concetti comuni tra PSN e SIMS, già standardizzati, ha portato all'esigenza di sfruttare il patrimonio informativo del *PsnPlus* in una duplice prospettiva: fornire nel *quality report* campi precompilati, alimentati dai contenuti della scheda PSN, consente sia l'armonizzazione dei metadati sia la riduzione dell'onere di documentazione per il compilatore del *quality report*.

Si lavorerà quindi, dal punto di vista tecnico, per fornire un *template* in gran parte già valorizzato, richiedendo solo una conferma da parte dei compilatori. Questo approccio semplificato non solo migliora l'efficienza, ma contribuisce anche alla sostenibilità del progetto Sistan per il *quality reporting* nel lungo periodo.

Un esercizio di *quality report* precompilato è disponibile a titolo di esempio in Appendice 1.

1. Definizione del *quality report* italiano

Prospetto 1.3 - Campi del *quality report* italiano presenti nella scheda PSN (in verde)

Primo livello	Secondo livello
Contatti	Ente
	Unità organizzativa
	Responsabile
	Funzione del responsabile
	E-mail
	Telefono
Ultimo aggiornamento	
Descrizione dei dati	Caratteristiche dei dati
	Classificazioni
	Variabili statistiche
	Unità statistica
	Popolazione statistica
	Territorio di riferimento
Descrizione del processo statistico	Copertura temporale
	Unità di misura
	Periodo temporale di riferimento
	Fonte dei dati
	Periodicità della raccolta dei dati
	Periodicità della diffusione
Mandato istituzionale	Raccolta dei dati
	Trattamento dei dati
	Politica della riservatezza
	Gestione della qualità
	Pertinenza
	Accuratezza
Attendibilità - Revisione dei dati	
Tempestività e puntualità	
Comparabilità nel tempo	
Accessibilità e chiarezza	Calendario di diffusione
	Comunicati stampa
	Pubblicazioni
	Portale on line
	Altro
	Documentazione metodologica
Costi e onere sui rispondenti	
Commenti	

Fonte: Istat

1.6 Prossimi passi per la qualità nel Sistan

Il pieno sfruttamento del sistema informativo pubblico consentirebbe, se adeguatamente utilizzato, di rivelare un prezioso patrimonio di informazioni su innumerevoli aspetti della vita sociale ed economica del Paese. In tale prospettiva, è compito dell'Istat, in qualità di ente responsabile per l'indirizzo, il coordinamento, l'assistenza tecnica e la formazione all'interno del Sistan, definire e promuovere un quadro di azione che costruisca e garantisca la qualità della statistica ufficiale, al fine di rendere disponibili statistiche fruibili per i diversi segmenti di utenza.

La documentazione della qualità in ambito statistico è di primaria importanza per garantire che i dati siano affidabili, trasparenti e utili: questa attività promuove la fiducia degli

utenti, supporta il miglioramento continuo, facilita la comparabilità e contribuisce a decisioni informate e responsabili. Investire nella documentazione della qualità è, quindi, un passo essenziale per qualsiasi sistema statistico che ambisca a essere rilevante e utile nel contesto attuale. Occorre impostare un progetto, sicuramente complesso sotto il profilo tecnico-organizzativo e impegnativo in quanto a risorse umane e finanziarie, ma che abbia la capacità di imprimere le dinamiche necessarie a dare un nuovo impulso al Sistan e a tutti i soggetti che ne fanno parte. In particolare, si deve:

- promuovere e supportare un più intenso impegno sulla qualità dei processi e dei prodotti della statistica pubblica;
- rafforzare la fiducia nella statistica ufficiale e la visibilità del Sistema statistico nazionale.

Per il raggiungimento di tali obiettivi è indispensabile formulare un piano di lavoro che si sviluppi a breve, medio e lungo termine, articolato in azioni specifiche e attività complesse, in cui non solo l'Istat, nel ruolo di coordinamento istituzionale, ma anche tutti i soggetti del Sistema statistico nazionale, nell'ambito delle rispettive competenze, devono necessariamente collaborare.

La nuova edizione del Codice italiano e l'associata Guida per l'implementazione, le linee guida del 2018, il presente manuale per il *quality reporting* e il progetto *Hub della statistica pubblica* sono tutte iniziative dell'Istat che operano nella medesima direzione: i nuovi strumenti mirano a sensibilizzare gli uffici di statistica sull'importanza della qualità e mirano a creare un clima favorevole e collaborativo allo sviluppo operativo di attività volte al miglioramento delle statistiche prodotte e diffuse. Si possono identificare in tale senso alcune linee di intervento:

- diffusione della conoscenza degli strumenti per la qualità predisposti dall'Istat attraverso iniziative di comunicazione e di sensibilizzazione presso gli operatori, le autorità politiche, i media, il pubblico e ogni soggetto interessato, anche attraverso la messa a disposizione di documentazione sulle esperienze maturate;
- interventi strutturati e regolari di formazione indirizzati oltre che ai responsabili degli uffici di statistica, anche al personale dell'amministrazione che si occupa di produzione, incentrati principalmente sui principi del Codice e su aspetti cruciali per la qualità della statistica, quali le metodologie e le classificazioni, gli aspetti legati alla *privacy*, i processi di produzione del dato, eccetera;
- monitoraggio sistematico degli enti del Sistan sul rispetto e l'applicazione dei principi, criteri, e metodi del *quality framework* italiano;
- iniziative per il miglioramento della qualità e per favorire la circolazione delle migliori pratiche tra gli enti del Sistan.

Per quanto lo sforzo attuale sia già sostanziale e importante, è necessario investire in maniera incessante e incisiva, andando oltre la semplice adozione formale degli strumenti, per promuovere concretamente il miglioramento continuo della qualità di processi e prodotti, a servizio della credibilità della statistica pubblica nel Paese.

2. GUIDA ALLA COMPILAZIONE¹

2.1 Contatti

In questa sezione vanno riportate le persone di riferimento dell'ufficio statistico dell'ente che hanno responsabilità del processo e degli output (sia i dati sia i metadati). Per ciascun riferimento deve essere indicato nome, cognome, organizzazione, unità organizzativa, area tecnica di appartenenza, contatto telefonico e indirizzo di posta elettronica.

Codice	Nome del concetto	Definizione	Linee guida
1.1	Ente	Organizzazione di riferimento	Fornire il nome completo dell'ente responsabile del processo e degli output (dati e metadati) oggetto del <i>report</i> .
1.2	Unità organizzativa	Unità organizzativa di riferimento	Fornire il nome completo dell'unità organizzativa responsabile del processo e degli output a cui ci può rivolgere.
1.3	Responsabile	I riferimenti dei punti di contatto	Fornire il nome e cognome della persona o delle persone indicate come contatto.
1.4	Funzione del responsabile	Il titolo e l'area di responsabilità del contatto	Fornire il titolo (o i titoli) e l'area (o le aree) di responsabilità della persona (o delle persone) indicate come contatto.
1.5	Email	L'indirizzo di posta elettronica del contatto	Fornire l'indirizzo di posta elettronica della persona indicata come contatto.
1.6	Telefono	Il numero di telefono del contatto	Fornire il numero di telefono della persona indicata come contatto.

Nel campo 1.1 occorre fornire il nome completo, non un codice o l'acronimo dell'ente.

Nel campo 1.2 si può aggiungere il numero o il codice della unità organizzativa.

Nel campo 1.3:

- se sono indicate più persone di riferimento, deve essere esplicitato il contatto principale;
- se l'autore del *report* è diverso dalla persona o dalle persone responsabili del processo e dei suoi output, deve essere fornito anche il contatto dell'autore del *report*.

Nel campo 1.4:

- esempi di titolo sono: responsabile dell'ufficio statistico, capo divisione, dirigente, quadro, funzionario, ricercatore *senior*, tecnologo, impiegato;
- esempi di area di responsabilità sono: divisione economia, gestione del database, area metodologica, diffusione.

Nel campo 1.5 l'indirizzo di posta elettronica, appartenente al dominio dell'ente:

- può essere individuale, relativo alla persona o persone di riferimento;
- può riferirsi a un'utenza di servizio, cui la persona (o le persone) di riferimento ha (hanno) accesso.

¹ Il Capitolo è stato redatto da: Maria Antonietta Liguori (paragrafi 2.1, 2.2, 2.3.6 e 2.3.7), Gabriele Ascari (paragrafi 2.3, 2.3.1, 2.3.2 e 2.3.3), Francesca Budano (paragrafi 2.3.4, 2.3.5, 2.7 e 2.13.2), Chiara Capogrossi (paragrafi 2.3.8, 2.5, 2.6 e 2.8), Daniela Fusco (paragrafi 2.3.9, 2.4.2, 2.4.3, 2.11, 2.12), Paola Giordano (paragrafi 2.4, 2.4.1 e 2.4.4), Giorgia Simeoni (paragrafi 2.4.5, 2.9 e 2.10) e Manuela Bartoloni (paragrafi 2.13, 2.13.1, 2.14 e 2.15).

Esempi

Si noti che negli esempi i nominativi e i dettagli di contatto personali sono stati soppressi per preservare la *privacy*.

Anagrafe nazionale degli studenti (MUR-00035)

Ministero dell'Istruzione e del Merito

1.1 Ente: Ministero dell'Istruzione e del Merito.

1.2 Unità organizzativa: Direzione generale per i contratti, gli acquisti e per i sistemi informativi e la statistica Ufficio VI - Gestione patrimonio informativo e statistica.

1.3 Responsabile: xxxx.

1.4 Funzione del responsabile: Referente ufficio di statistica.

1.5 Email: xxxx@istruzione.it.

1.6 Telefono: xx

Registro nazionale degli ipotiroidei congeniti (ISS-00007)

ISS - Istituto superiore di sanità

1.1 Ente: Istituto superiore di sanità.

1.2 Unità organizzativa: Dipartimento di Malattie Cardiovascolari, Endocrino-Metaboliche e Invecchiamento - ISS.

1.3 Responsabile: xxxx.

1.4 Funzione del responsabile: Responsabile scientifico del dipartimento

1.5 Email: xxxx@iss.it.

1.6 Telefono: xxxx.

2.2 Aggiornamento dei metadati

Codice	Nome del concetto	Definizione	Linea guida
2	Ultimo aggiornamento dei metadati	La data dell'ultimo aggiornamento del contenuto dei metadati diffusi.	Deve essere fornita la data dell'inserimento o dell'ultimo aggiornamento dei metadati che vengono messi a disposizione degli utenti.

Si intende l'ultima modifica apportata ai metadati, avvenuta manualmente o in via automatica, all'interno del proprio sistema dei metadati. Nel caso di certificazione degli stessi, documentare la data di certificazione, anche nel caso in cui tale operazione non abbia apportato alcuna modifica ai metadati.

Esempio

Sistema epidemiologico integrato dell'epatite virale acuta (SEIEVA) (ISS-00006)

ISS - Istituto superiore di sanità

Ultimo aggiornamento dei metadati:

Data di ultimo aggiornamento: 30 novembre 2023.

2.3 Descrizione dei dati

La presente sezione si propone di fornire una dettagliata descrizione dei dati diffusi, i quali possono essere visualizzati dagli utenti attraverso differenti modalità di rappresentazione, quali tabelle, grafici o mappe. L'obiettivo è garantire una comprensione chiara e precisa delle informazioni statistiche presentate.

Questo concetto introduce l'illustrazione dettagliata del prodotto statistico al quale il *quality report* fa riferimento. Vengono tracciate le caratteristiche principali dei dati, con particolare riguardo alle unità statistiche considerate, alle principali variabili diffuse e alle classificazioni utilizzate. Inoltre, per consentire all'utente di comprendere pienamente il contesto dei dati, vengono fornite informazioni sulla copertura geografica e temporale, nonché sulla popolazione di riferimento e altre caratteristiche riportate in specifiche sottovoci. Nello specifico, la sezione 3 “Descrizione dei dati” risulta articolata nelle seguenti voci di dettaglio:

- 3.1 Caratteristiche dei dati.
- 3.2 Classificazioni.
- 3.3 Variabili statistiche.
- 3.4 Unità statistica.
- 3.5 Popolazione statistica.
- 3.6 Territorio di riferimento.
- 3.7 Copertura temporale.
- 3.8 Unità di misura.
- 3.9 Periodo temporale di riferimento.

2.3.1 Caratteristiche dei dati

Codice	Nome del concetto	Definizione	Linea guida
3.1	Caratteristiche dei dati	Principali caratteristiche dei dati diffusi.	Descrivere brevemente le principali caratteristiche dei dati diffusi, evidenziando il fenomeno che viene descritto. Si può fare riferimento alle unità e alle variabili diffuse, che andranno poi approfondite nei sottoconcetti successivi.

Data la grande varietà di temi e di fenomeni trattati all'interno del Sistan, questa sezione funge da introduzione tematica al prodotto statistico, consentendo di descrivere le principali caratteristiche del *dataset* messo a disposizione degli utenti e di illustrare gli obiettivi del processo statistico che hanno portato alla creazione di tale *dataset* e alla descrizione dei fenomeni chiave in esso rappresentati. In altre parole, questo sottoconcetto rappresenta una premessa introduttiva con l'obiettivo di offrire agli utenti una panoramica generale, che sarà poi dettagliata nelle voci successive di questa sezione e nelle parti successive del documento.

Esempi

Indagine sui comportamenti formativi degli adulti (Indaco Adulti) (IAP-00003)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

L'indagine sui comportamenti formativi degli adulti (Indaco Adulti) è una rilevazione campionaria che ha l'obiettivo di ricostruire il quadro della diffusione dell'apprendimento permanente degli adulti in Italia. L'indagine rileva la partecipazione alle attività di istruzione e formazione.

Oggetto di indagine sono la partecipazione formativa formale, non formale e informale degli adulti, i divari nell'accesso di tipo sociodemografico, culturale e professionale e i benefici percepiti, le competenze, gli ostacoli all'accesso, l'apprendimento intergenerazionale, la transizione pandemica e lo *smart working*.

Monitoraggio del consumo del suolo e del *soil sealing* (APA-00046)

ISPRA - Istituto superiore per la protezione e la ricerca ambientale

Il consumo di suolo è monitorato dal Sistema nazionale per la protezione dell'ambiente che ogni anno realizza il rapporto nazionale “Consumo di suolo, dinamiche territoriali e servizi ecosistemici”.

È un fenomeno associato alla perdita di una risorsa ambientale fondamentale, dovuta all'occupazione di superficie originariamente agricola, naturale o seminaturale. Il fenomeno si riferisce, quindi, a un incremento della copertura artificiale di terreno, legato alle dinamiche insediative. Un processo prevalentemente dovuto alla costruzione di nuovi edifici e infrastrutture, all'espansione delle città, alla densificazione o alla conversione di terreno entro un'area urbana, all'infrastrutturazione del territorio.

Procedimenti penali per delitti commessi con finalità di terrorismo ed eversione dell'ordine democratico (art. 51 comma 3 *quater* c.p.p.) (MGG-00010)

Ministero della Giustizia

L'indagine, che ha subito nel corso del tempo diverse modifiche strutturali, anche a seguito di importanti modifiche normative, fa parte del PSN e rileva il numero dei procedimenti penali iscritti, pendenti e definiti presso le procure e gli uffici giudicanti competenti per la materia in oggetto, insieme al relativo numero di persone coinvolte e alla qualificazione giuridica del fatto contestata.

2.3.2 Classificazioni

Codice	Nome del concetto	Definizione	Linea guida
3.2	Classificazioni	Una classificazione è una suddivisione di oggetti secondo gruppi in base alle caratteristiche omogenee che essi condividono.	Elencare le classificazioni utilizzate nel <i>dataset</i> , fornendo dei link di riferimento pubblici (se disponibili pubblicamente).

In questa voce vanno elencate in maniera dettagliata le classificazioni utilizzate all'interno del *dataset* pubblicato. Per ogni classificazione è necessario riportare il nome completo e, ove disponibili, fornire link di riferimento pubblici, al fine di garantire trasparenza e verificabilità.

È fondamentale che gli enti del Sistan utilizzino, nella misura più ampia possibile, le classificazioni predisposte dall'Istat, nonché quelle definite a livello internazionale. Eventuali deviazioni da tali standard vanno esplicate e adeguatamente motivate, fornendo le ragioni che hanno portato all'adozione di classificazioni alternative.

Esempi

La mortalità per causa in Italia (ISS-00013)

ISS - Istituto superiore di sanità

Le cause di morte vengono classificate a livello internazionale secondo un sistema di codifica denominato ICD (*International Classification of Diseases*) elaborato dall'Orga-

2. Guida alla compilazione

nizzazione mondiale della sanità, che viene periodicamente aggiornata. A partire dall'anno 2003 è stata adottata anche in Italia la classificazione tuttora vigente, la decima revisione (ICD-10).

Sviluppo e potenziamento del Sistema informativo dell'anagrafe delle biblioteche italiane (MBE-00014)

Ministero della Cultura

Il codice ISIL (*International Standard Identifier for Libraries and related organizations*), è l'identificativo standard internazionale conforme alla norma ISO-15511 per le biblioteche e le organizzazioni collegate come archivi e musei ed è ovviamente utilizzato nella base dati Anagrafe. Il codice ISIL è assegnato dall'Istituto centrale per il catalogo unico (ICCU) a tutte le biblioteche registrate in Anagrafe e sarà assegnato, sempre dall'ICCU, alle altre istituzioni che lo richiederanno.

Sistema informativo per l'occupazione e la formazione *Excelsior* (UCC-00007)

Unioncamere

Il sistema informativo *Excelsior* adotta due principali tipologie di classificazioni:

- classificazioni relative alle caratteristiche dell'impresa, tra cui assume rilievo quella riferita all'attività economica;
- classificazioni relative alle caratteristiche della figura professionale. Per quest'ultima sono rilevanti in particolare le classificazioni relative a: tipo di contratto; livello di istruzione e titoli di studio; professioni e figure elementari.

Corrispondenza tra Ateco 2007 e settori *Excelsior*:

https://excelsior.unioncamere.net/sites/default/files/documenti/ATECO_vers_2017.pdf.

Classificazione dei titoli di studio:

<https://excelsior.unioncamere.net/sites/default/files/documenti/Livellistruzione-Indirizzi-Titoli.pdf>.

Figure professionali e collocazione nella CP 2021:

https://excelsior.unioncamere.net/sites/default/files/documenti/Professioni_ISTAT.pdf.

2.3.3 Variabili statistiche

Codice	Nome del concetto	Definizione	Linea guida
3.3	Variabili statistiche	Caratteristiche delle osservazioni statistiche.	Definire e descrivere brevemente le principali variabili statistiche osservate o derivate.

Ogni prodotto statistico include una o più variabili, spesso definite da normative nazionali; descrivono le caratteristiche delle unità osservate e possono essere di diverso tipo, in particolare qualitative oppure quantitative. È opportuno in questa sezione elencare almeno le principali variabili presenti nel *dataset*, fornirne tipologia, definizione e metodo di calcolo. Inoltre, se per alcune variabili si adottano definizioni diverse da quelle ufficiali, è utile riportare tali discrepanze (può dunque essere utile riportare link a eventuale documentazione che contenga le loro definizioni).

Esempi

Statistiche del registro e delle successioni (FIN-00011)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

https://www1.finanze.gov.it/finanze/registro/public/v_2_0_0/contenuti/it/definizione_variabili.pdf.

Stima della incidenza e della prevalenza dei principali tumori in Italia e nelle regioni italiane (ISS-00017)

ISS - Istituto superiore di sanità

La frequenza dei tumori viene descritta come incidenza, cioè il numero di nuove diagnosi in un intervallo di tempo – in genere un anno – e come prevalenza, cioè il numero di casi in vita a una certa data. L'esito della malattia viene studiato in termini di mortalità (numero di decessi annui per tumore) e di sopravvivenza (probabilità di sopravvivere al crescere della distanza temporale dalla diagnosi).

Statistica annuale della produzione e del consumo di energia elettrica in Italia (TER-00001)

Terna S.p.A.

Elenco delle variabili incluse nella raccolta dati:

- producibilità media annua degli impianti idroelettrici; potenza efficiente (londa e netta) e potenza nominale (motori primi e generatori) degli impianti di generazione al 31 dicembre dell'anno di riferimento;
- produzione londa e netta di energia elettrica, energia elettrica immessa in rete ed energia elettrica autoconsumata (per attività economica) degli impianti di generazione nel mese di riferimento;
- consumo mensile di combustibile degli impianti di generazione;
- consumo annuale di energia elettrica per attività economica;
- potere calorifico inferiore e consumo specifico elettrico;
- tensione.

Le suddette variabili vengono raccolte per ogni sezione di tutti gli impianti di generazione (produttori e autoproduttori); i dati sul consumo di energia elettrica vengono raccolti anche per ogni distributore.

2.3.4 Unità statistica

Codice	Nome del concetto	Definizione	Linee guida
3.4	Unità statistica	Entità alla quale si riferiscono le informazioni osservate/di interesse in un processo statistico.	Definire il tipo di unità statistica su cui vengono raccolti i dati, ad esempio impresa, unità locale, nucleo familiare, abitazione privata, persona. Definire ogni tipo di unità, se esistono più tipi.

Si tratta dell'entità oggetto di osservazione del processo. Le unità statistiche possono essere sia direttamente osservabili (ad esempio famiglie, componenti delle famiglie, imprese, lavoratori dipendenti) sia eventi (ad esempio vacanze, ricoveri ospedalieri).

Esempi

Movimento dei procedimenti civili presso le Corti di appello e i Tribunali ordinari (MGG-00124)

2. Guida alla compilazione

Ministero della Giustizia

Oggetto di rilevazione sono i procedimenti pendenti dinanzi a tutti gli uffici giudiziari italiani (Corte di cassazione, Corte di appello, Tribunale ordinario, Giudice di pace e Tribunale per i minorenni), distinti per area SICID (affari contenziosi, controversie in materia di lavoro e previdenza, procedimenti speciali e sommari e volontaria giurisdizione) e SIECIC (procedure esecutive e concorsuali). Sono esclusi dal calcolo i procedimenti del giudice tutelare, quelli di accertamento tecnico preventivo in materia previdenziale e l'attività di "ricevimento e verbalizzazione di dichiarazione giurata".

Disabilità, DSA e accesso alla Formazione universitaria (ANV-00002)

ANVUR - Agenzia nazionale per la valutazione del sistema universitario e della ricerca

Studenti con disabilità o con disturbi specifici dell'apprendimento (DSA) iscritti alle università statali, non statali e telematiche. Si considerano le categorie amministrative degli studenti con invalidità non inferiore al 66 per cento e quelli con invalidità inferiore al 66 per cento, gli studenti con disabilità certificata ai sensi dell'art. 3 comma 1 e comma 3 della legge 104/1992 e gli studenti con DSA certificati in base a quanto disposto della legge 170/2010.

Pensioni erogate dall'INPS (escluse le gestioni dipendenti pubblici e ex-Enpals) - vigenti e liquidate (IPS-00001)

INPS - Istituto nazionale della previdenza sociale

L'unità statistica è rappresentata da:

- le pensioni vigenti al primo gennaio dell'anno e cioè quelle che hanno ottenuto il rinnovo del mandato di pagamento al 1° gennaio con la relativa perequazione dell'importo della pensione. Esse costituiscono quindi un dato di *stock*;
- le pensioni liquidate e cioè quelle che hanno dato luogo al primo pagamento nel corso dell'anno precedente. Esse costituiscono un dato di *flusso*.

2.3.5 Popolazione statistica

Codice	Nome del concetto	Definizione	Linea guida
3.5	Popolazione statistica	L'insieme esaustivo di entità, detto anche universo, oggetto di un'analisi statistica.	Definire la popolazione obiettivo (<i>target population</i>) delle unità statistiche per cui si desidera ottenere le informazioni.

La popolazione statistica rappresenta l'insieme completo di tutte le unità statistiche per le quali si intende ottenere l'informazione, in base agli obiettivi specifici della ricerca. Nella pratica, tuttavia, non sempre è possibile raccogliere informazioni su tutte le unità della popolazione obiettivo. La popolazione di indagine (*survey population*), cioè l'insieme delle unità statistiche effettivamente osservate, rappresenta l'approssimazione della popolazione obiettivo utilizzata in via pratica. Essa dipende dalla popolazione presente nella lista di riferimento (*frame*) e deve essere descritta nel campo "Fondate dei dati". Eventuali discrepanze tra popolazione obiettivo e popolazione di indagine costituiscono un problema di copertura che va discusso nella sezione "Accuratezza".

Esempi

Condizioni di vita delle famiglie trentine (PAT-00007)

Provincia autonoma di Trento
Famiglie trentine.

Disoccupazione nei settori non agricoli (IPS-00006)
INPS - Istituto nazionale della previdenza sociale
Lavoratori disoccupati nei settori non agricoli.

Società, praticanti tesserati e operatori delle Federazioni sportive nazionali e delle Discipline sportive associate (CON-00002)
CONI - Comitato olimpico nazionale italiano
Entità associative affiliate e tutti i soggetti tesserati a vario titolo alle Federazioni sportive nazionali e alle Discipline sportive associate.

Registro nazionale degli ipotiroidei congeniti (ISS-00007)
ISS - Istituto superiore di sanità
Bambini affetti da ipotiroidismo congenito identificati su tutto il territorio nazionale.

2.3.6 Territorio di riferimento

Codice	Nome del concetto	Definizione	Linea guida
3.6	Territorio di riferimento	Il Paese o l'area geografica a cui si riferisce il fenomeno statistico.	Indicare il Paese, le regioni o gli altri aggregati territoriali a cui si riferiscono i dati.

Va segnalata la eventualità di territori esclusi dai dati diffusi. È opportuno che le motivazioni di tale scelta vengano documentate.

Esempi

Statistiche sugli indici sintetici di affidabilità (FIN-00001)
Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze
Classificazione territoriale: Regione e Regione/provincia.

L'indagine sul dolore cronico in Italia a partire dai dati dell'EHIS 2019 (ISS-00021)
ISS - Istituto superiore di sanità

Per soddisfare i bisogni informativi a livello territoriale e consentire stime regionali utili alla programmazione sanitaria locale, nonché, ove possibile, stime subregionali, i domini di studio, ossia gli ambiti territoriali ai quali sono riferiti i parametri di popolazione oggetto di stima, sono: le cinque ripartizioni geografiche (Nord-ovest, Nord-est, Centro, Sud, Isole); le regioni (a eccezione del Trentino-Alto Adige/Südtirol, le cui stime sono prodotte distintamente per le Province autonome di Bolzano/Bozen e di Trento).

Proiezioni della popolazione e delle famiglie (MIL-00033)
Comune di Milano

Il documento *Previsioni della popolazione e delle famiglie a Milano 2023-2039* presenta le previsioni della popolazione residente (scenario centrale) per l'intero territorio comunale.

La domanda di lavoro per bacino dei centri per l'impiego (ILA-00002)
Sviluppo Lavoro Italia S.p.A (già ANPAL Servizi S.p.A)

2. Guida alla compilazione

La costruzione dell'aggregato dei confini amministrativi dei bacini di competenza degli uffici pubblici del lavoro è determinata dai dati attinenti all'anagrafica e ai comuni associati ai Centri per l'impiego (CPI) contenuti nel file della "Classificazione standard" di fonte ANPAL. La realizzazione dell'aggregato si basa sull'universo dei comuni italiani (7.904, dato Istat 2022) raggruppati per bacino amministrativo dei CPI (540, dato ANPAL aprile 2022). Il bacino amministrativo di riferimento dei CPI è ottenuto dall'insieme dei comuni limitrofi di competenza territoriale di ogni singolo Centro per l'impiego, tranne che per le aree metropolitane di Genova, Napoli e Roma. Inoltre, i CPI di Ales, Assemini, Quartu Sant'Elena e Lanusei, avendo competenze su comuni appartenenti a più province, sono stati ricostruiti nel seguente modo:

- tutti i comuni associati al bacino territoriale del CPI Ales sono stati considerati come appartenenti alla provincia di Oristano;
- tutti i comuni associati ai bacini territoriali del CPI Assemini e del CPI Quartu Sant'Elena sono stati considerati come appartenenti alla provincia di Cagliari;
- tutti i comuni associati al bacino territoriale del CPI Lanusei sono stati considerati come appartenenti alla provincia di Nuoro.

Il numero totale di aree territoriali di competenza dei CPI oggetto di analisi è, dunque, pari a 529.

Rilevazione delle compravendite immobiliari (FIN-00009)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

La "Base dati mutui ipotecari" comprende tutte le unità immobiliari urbane ipotecate situate nel territorio nazionale a eccezione di quelle nei comuni delle Province autonome di Trento (n. 177 comuni) e Bolzano/Bozen (n. 116 comuni), del Veneto (n. 2 comuni), della Lombardia (n. 2 comuni) e del Friuli-Venezia Giulia (n. 44 comuni) dove vige il sistema dei libri fondiari (o sistema tavolare) per la pubblicità immobiliare.

2.3.7 Copertura temporale

Codice	Nome del concetto	Definizione	Linea guida
3.7	Copertura temporale	Il periodo temporale di disponibilità dei dati.	Indicare il periodo o i periodi di tempo di copertura temporale dei dati diffusi (ad esempio, il primo trimestre 2018 o i trimestri 2015-2018 o l'anno 2018 o gli anni 1985-2018).

Si noti che eventuali problematiche relative alla comparabilità nel tempo vanno documentate nella voce dedicata (2.12 "Comparabilità nel tempo").

Esempi

Procedimenti penali per delitti commessi con finalità di terrorismo ed eversione dell'ordine democratico (MGG-00010)

Ministero della Giustizia

L'indagine conoscitiva è iniziata nel 1984, in osservanza di una circolare del Direttore generale degli Affari penali.

Certificato di assistenza al parto (SAL-00038)

Ministero della Salute

L'attuale rilevazione dati del Certificato di assistenza al parto (CeDAP) ha avuto inizio il 1° gennaio 2002, in attuazione del decreto del Ministro della Sanità 16 luglio 2001, n.349.

Produzione, recupero, trattamento e smaltimento di rifiuti urbani, speciali e pericolosi (APA-00001)

ISPRA - Istituto superiore per la protezione e la ricerca ambientale

Dati fino al dettaglio comunale: 2010-2022; dati fino al dettaglio provinciale: 2001-2022.

2.3.8 Unità di misura

Codice	Nome del concetto	Definizione	Linea guida
3.8	Unità di misura	L'unità nella quale sono misurati i dati.	I dati di solito coinvolgono diverse unità di misura a seconda delle variabili. Esempi sono euro, valuta nazionale, numero di persone e tasso per 100 mila abitanti.

Si faccia attenzione a includere, in questo concetto, anche l'informazione relativa all'ordine di grandezza (ad esempio: migliaia, milioni) delle unità numeriche.

Esempi

Panel degli acquisti agroalimentari delle famiglie italiane (ISM-00001)

ISMEA - Istituto di servizi per il mercato agricolo alimentare

Prezzi internazionali bovini: 100 euro al chilogrammo.

Principali input produttivi:

- Vitelli da ristallo, baliotti: euro/chilogrammo/peso vivo;
- Mais, foraggi, mangimi: euro/tonnellata.

Il patrimonio dello stato - informazioni e statistiche (TES-00022)

Ministero dell'Economia e delle Finanze - Settore ex Tesoro

Attività finanziarie e non finanziarie, passività finanziarie: migliaia di euro.

Produzione, recupero, trattamento e smaltimento di rifiuti urbani, speciali e pericolosi (APA-00001)

ISPRA - Istituto superiore per la protezione e la ricerca ambientale

Produzione e raccolta rifiuti urbani:

- raccolta differenziata, rifiuti urbani: tonnellate;
- raccolta differenziata pro capite, rifiuti urbani pro capite: chilogrammi per abitante.

2.3.9 Periodo temporale di riferimento

Codice	Nome del concetto	Definizione	Linea guida
3.9	Periodo temporale di riferimento	Il periodo di tempo o il momento nel tempo a cui fanno riferimento le osservazioni misurate.	Il valore di una variabile si riferisce a un periodo di tempo specifico (ad esempio: l'ultima settimana di un mese, un mese, un anno fiscale, un anno solare o diversi anni solari) o a un momento preciso nel tempo (ad esempio: un giorno specifico o l'ultimo giorno di un mese).

2. Guida alla compilazione

Le variabili in un *dataset* possono riferirsi a più periodi di riferimento: è necessario indicarli chiaramente tutti. Si noti che l'eventuale differenza tra il/i periodo/i di riferimento obiettivo e il/i periodo/i di riferimento effettivo è un problema di accuratezza e dovrebbe essere discusso nella apposita sezione. Si noti che, se la popolazione di indagine non include tutte le unità della popolazione obiettivo per il periodo di riferimento specificato, si tratta di un problema di copertura e dovrebbe essere discusso sempre nella sezione “Accuratezza”.

Esempi

Analisi statistiche dei dati derivanti dal processo tributario (FIN-00004)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

Il monitoraggio dei ricorsi tributari pendenti fa riferimento al 31° dicembre dell'anno precedente alla pubblicazione.

2.4 Descrizione del processo statistico

Le pratiche di documentazione del processo statistico nel Sistan differiscono ampiamente.

Alcuni enti producono rapporti tecnici che descrivono in dettaglio metodi statistici e procedure operative, ad esempio includendo procedure di campionamento, formule di stima, metodi di misurazione e regole di *editing*. Quando tale documentazione tecnica esiste ed è facilmente accessibile, il *quality report* dovrebbe semplicemente riassumerla e fornire un collegamento ipertestuale a essa. Tuttavia, quando tale documentazione non è disponibile, le informazioni sui metodi e sulle procedure dovrebbero essere incluse nel *quality report* stesso. La presente sezione è il luogo in cui inserire tali elementi, che riguardano le fasi di progettazione, raccolta, elaborazione e analisi di un processo statistico, insieme alla periodicità della diffusione.

2.4.1 Fonte dei dati

Codice	Nome del concetto	Definizione	Linee guida
4.1	Fonte dei dati	Caratteristiche e componenti dei dati statistici grezzi utilizzati per la compilazione degli aggregati statistici.	Indicare se i dati sono basati su un'indagine, fonti amministrative, più fonti di dati o macroaggregati. In caso di processi multifonte o macroaggregati, descrivere ogni fonte. Per ogni fonte di indagine: riportare la popolazione dell'indagine, facendo riferimento alla descrizione della popolazione obiettivo, e riassumere il disegno campionario. Per ogni fonte amministrativa: riassumere la fonte, il suo scopo primario e i principali dati acquisiti.

Di seguito ulteriori linee guida, comuni a tutti i tipi di processo statistico o specifiche per le diverse tipologie di processo statistico:

Tutti i tipi di processo statistico

Descrivere le attività di progettazione e sviluppo associate al processo, inclusi tutti i lavori di ricerca necessari per definire output statistici, concetti, metodi, strumenti di raccolta e procedure operative.

Descrivere in quale misura la progettazione ha fatto uso di definizioni, classificazioni e metodi standard internazionali e nazionali.

Indagine probabilistica o censuaria

Il disegno dell'indagine può prevedere l'osservazione di un sottoinsieme di elementi della popolazione oppure essere esaustivo. Nella misura in cui non sono trattati altrove:

- descrivere i metodi con cui è stata ottenuta o creata la lista di riferimento o archivio di estrazione delle unità dell'indagine, incluse le azioni per prevenire sovraccopertura, sottocopertura e inaccuratezza nella classificazione delle unità e dei dettagli di contatto;
- indicare il numero di unità presenti nella lista di indagine, in totale e per le disaggregazioni di livello elevato.

Descrivere il disegno di campionamento, compresi l'unità statistica, la popolazione obiettivo, la popolazione di indagine e la lista di riferimento dell'indagine.

Descrivere le procedure di stratificazione, allocazione e selezione del campione.

Indicare il numero di unità nella lista di riferimento dell'indagine e nel campione, in totale e (per le indagini stratificate) negli strati più significativi.

Indagine non probabilistica

Descrivere le motivazioni che hanno portato a non utilizzare il campionamento probabilistico e qualsiasi valutazione fatta sulla procedura di campionamento che possa supportare le inferenze sulla popolazione a partire dal campione.

Descrivere l'approccio effettivo per la selezione delle unità di campionamento.

Per le indagini con campionamento per quote, indicare le variabili utilizzate per stabilire le quote e descrivere come gli intervistatori sono stati formati per la selezione delle unità campionarie finali.

Per le indagini con campionamento *cut-off* con meccanismo soggettivo di selezione del campione, indicare quali variabili e criteri sono stati utilizzati per identificare le unità di campionamento più grandi o più importanti e descrivere le procedure utilizzate.

Per le indagini su base volontaria, descrivere il meccanismo di ingaggio dei rispondenti e fornire una descrizione del campione finale in termini di variabili potenzialmente importanti, come la posizione geografica, il sesso, l'età, le variabili di status sociale.

Processo basato su dati amministrativi

Descrivere la fonte dei dati, il processo con cui sono stati prodotti e il processo con cui sono stati acquisiti dall'ente.

Descrivere le unità e la popolazione a cui si riferiscono i dati, il contenuto dei dati (variabili e caratteristiche) e le loro relazioni con i concetti statistici obiettivo.

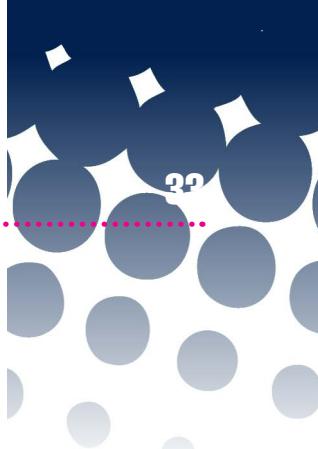
Processo multifonte

Descrivere come è composto il processo in termini di tutti i suoi componenti.

Descrivere ogni componente in modo sufficientemente dettagliato, tenendo conto della sua importanza complessiva per gli output.

Processo di compilazione dei macroaggregati

Descrivere la fonte di ciascun *set* di dati di input e il processo con cui è stato prodotto e acquisito. Descrivere il contenuto dei dati (variabili e caratteristiche) e le loro relazioni con i concetti statistici obiettivo.



Esempi

Audit dei fabbisogni professionali (IAP-00011)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

(L'esempio rientra nella casistica "Indagine probabilistica").

Il campione di riferimento per l'indagine è costituito dalle imprese private (35 mila) con almeno un dipendente di tutti i settori economici, con esclusione dunque della Pubblica amministrazione e prevede una quota longitudinale. Le imprese appartenenti al campione sono state selezionate casualmente dai principali archivi statistici del Paese, in particolare da ASIA (Archivio statistico imprese attive) gestito dall'Istat. La stratificazione del campione di imprese rappresentativo dell'intera economia è stata articolata, in particolare, sulla base di 13 settori economici (individuati a partire dalla classificazione Ateco sulle attività economiche), quattro dimensioni di impresa (piccolissime imprese da uno a nove dipendenti; imprese piccole da 10 a 49 dipendenti, imprese medie da 50 a 249 dipendenti e imprese grandi con oltre 250 dipendenti) e quattro ripartizioni geografiche (Nord-ovest, Nord-est, Centro, Mezzogiorno). Le liste anagrafiche delle imprese da intervistare hanno previsto l'individuazione di imprese "titolari" e di un congruo numero di "riserve" eventualmente da utilizzare in corso di opera.

Il quadro concettuale di riferimento per l'indagine e le tassonomie di variabili utilizzate sono per lo più mutuate dal modello statunitense dell'*Occupational Information Network* già utilizzato in Italia nell'ambito dell'Indagine campionaria sulle professioni, realizzata congiuntamente da INAPP e Istat.

Panel Industria Alimentare (ISM-00003)

ISMEA - Istituto di servizi per il mercato agricolo alimentare

(L'esempio rientra nella casistica "Indagine probabilistica").

Per lo svolgimento dell'indagine, ISMEA si avvale di un *panel* di 1.550 operatori dell'industria alimentare italiana, individuati applicando uno schema di campionamento stratificato casuale. Fanno parte integrante del campione 215 imprese assunte come unità autorappresentative, individuate tra le realtà di eccellenza dell'industria alimentare italiana e incluse a priori, in quanto sono ritenute di particolare rilevanza ai fini delle analisi del settore. Per l'universo di riferimento si è fatto ricorso alle liste dell'Archivio statistico delle imprese attive (ASIA) dell'Istat; per la stratificazione del campione sono stati presi in considerazione 24 settori del manifatturiero alimentare definiti a partire dalla classificazione Ateco 2007 (accorpati, poi, in 18 ai fini della divulgazione esterna dei risultati dell'indagine), le quattro macroaree geografiche adottate dall'Istat per le indagini strutturali e congiunturali, le quattro classi di addetti definite in base al decreto del Ministro delle attività produttive del 18 aprile 2005. Poi, il numero delle imprese da allocare nei diversi strati del campione è stato definito in base alla quota di fatturato conseguita da ogni settore nelle diverse aree geografiche, rispetto al fatturato complessivo dell'universo delle industrie alimentari considerate per la definizione del campione. Per la costruzione dell'indice del clima di fiducia, ISMEA, come suggerito dalle direttive dell'Organizzazione per la cooperazione e lo sviluppo economico del 2003 e dell'Unione europea del 2007, ha predisposto un algoritmo che sintetizza i giudizi degli operatori sul livello generale degli ordini, sul livello delle giacenze di prodotti finiti e sulle aspettative circa la tendenza della produzione nell'immediato futuro.

Osservatorio nazionale per il monitoraggio della iodoprofilassi in Italia (OSNAMI) (ISS-00063)

ISS - Istituto superiore di sanità

(L'esempio rientra nella casistica "Indagine probabilistica").

L'OSNAMI, grazie alla collaborazione con gli Osservatori regionali per la prevenzione del gozzo, tra il 2015 e il 2019 ha condotto una seconda sorveglianza con l'obiettivo di valutare lo stato nutrizionale iodico della popolazione italiana a 14 anni dall'approvazione della legge.

All'indagine hanno partecipato nove Regioni (Liguria, Veneto, Emilia-Romagna, Toscana, Umbria, Marche, Lazio, Calabria, Sicilia) e sono stati reclutati 3.976 scolari (52,3 per cento di sesso maschile; 47,7 per cento di sesso femminile) di età compresa tra gli 11 e i 13 anni, e residenti sia in aree rurali (48 per cento, aree sentinella) sia in aree urbane (52 per cento, aree di riferimento).

Su tutti i bambini reclutati è stata effettuata la visita medica con somministrazione di un questionario ad hoc per la verifica dell'uso di sale iodato ed è stato raccolto un campione estemporaneo di urine (*spot*) per la determinazione della UIC mediante spettrometria di massa al plasma accoppiato induttivamente (*Inductively Coupled Plasma-Mass Spectrometry*).

Le determinazioni sono state effettuate dal Dipartimento di medicina clinica e sperimentale, Sezione di endocrinologia dell'università di Pisa nei primi due anni di reclutamento, e dal Dipartimento di sicurezza alimentare, nutrizione e sanità pubblica veterinaria dell'Istituto superiore di sanità nel restante periodo di reclutamento.

PLUS: Partecipation Labour Unemployment Survey - Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro (IAP-00004)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

(L'esempio rientra nella casistica "Indagine non probabilistica").

L'indagine PLUS è una rilevazione statistica periodica condotta su un campione di 45 mila individui dai 18 ai 74 anni intervistati telefonicamente attraverso un sistema CatI (*Computer-Assisted Telephone Interviewing*) e in assenza di rispondenti *proxy*. I nominativi da intervistare vengono estratti dagli elenchi telefonici. La pianificazione delle interviste da effettuare è stata realizzata sulla base di un campionamento per quote stratificato con definizione di domini di studio parzialmente sovrapposti. La scelta del campionamento per quote è stata motivata dall'esigenza di ridurre notevolmente la numerosità campionaria necessaria alla produzione di stime statisticamente significative per piccole sottopopolazioni di interesse. Come scelta alternativa si sarebbero potute utilizzare strategie classiche di campionamento a due stadi (per esempio, comuni e famiglie) che però, oltre a richiedere rilevazioni sul campo molto più onerose, avrebbero comportato la rinuncia a una delle principali caratteristiche dell'indagine PLUS, ovvero quella dell'assenza di rispondenti *proxy*.

Rete di informazione contabile agricola (RICA) (INE-00001)

CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

(L'esempio rientra nella casistica "Indagine non probabilistica").

La RICA è una indagine campionaria annuale istituita dalla Commissione economica europea nel 1965, con il regolamento (CE) n. 79/56 e aggiornata con il regolamento (CE) n. 1217/2009 e s.m.i. Essa viene svolta, in Italia a partire dal 1968, con un'impostazione

2. Guida alla compilazione

analoga in tutti i paesi membri dell'Unione europea e rappresenta l'unica fonte armonizzata di dati microeconomici sull'evoluzione dei redditi e sulle dinamiche economico-strutturali delle aziende agricole.

L'indagine RICA non rappresenta tutto l'universo delle aziende agricole censite in un determinato territorio, ma solo quelle che, per la loro dimensione economica, possono essere considerate professionali e orientate al mercato. La metodologia adottata mira a fornire dati rappresentativi su tre dimensioni: regione, dimensione economica e ordinamento tecnico economico.

Le oltre 86 mila aziende della RICA comunitaria rappresentano quasi cinque milioni di aziende unionali, il 90 per cento della superficie agricole e il 90 per cento della produzione standard.

Attualmente (dati medi 2014-2019), il campione RICA italiano si basa su un campione ragionato di circa 11 mila aziende, strutturato in modo da rappresentare le diverse tipologie produttive e dimensionali presenti sul territorio nazionale. Esso consente una copertura media a livello nazionale del 95 per cento della superficie agricola utilizzata, del 97 per cento del valore della produzione standard, del 92 per cento delle unità di lavoro e del 91 per cento delle unità di bestiame.

Pensioni erogate dall'INPS (escluse le gestioni dipendenti pubblici e ex-Enpals) - vi- genti e liquidate (IPS-00001)

INPS - Istituto nazionale della previdenza sociale

(L'esempio rientra nella casistica "Processo basato su dati amministrativi - Statistiche da fonti amministrative organizzate").

La fonte dei dati dell'Osservatorio statistico sulle pensioni erogate da INPS è l'archivio gestionale alimentato dalle procedure amministrativo-contabili per la liquidazione e gestione delle pensioni. Dall'archivio, si possono desumere informazioni anagrafiche, contabili, reddituali e contributive, quali ad esempio, dati di carattere generale della pensione, dati anagrafici del titolare e dei contitolari della pensione, oppure dei familiari o del delegato alla riscossione, dati sul dante causa della pensione ai superstiti, dati di calcolo della pensione e delle trattenute, dati sui conguagli per rinnovo o sulle certificazioni fiscali; dati contabili dell'anno in corso; dati contabili analitici storici; dati dei mandati di pagamento.

Lavoratori domestici (IPS-00015)

INPS - Istituto nazionale della previdenza sociale

(L'esempio rientra nella casistica "Processo basato su dati amministrativi - Statistiche da fonti amministrative organizzate").

La fonte dei dati utilizzati per la costruzione dell'Osservatorio sui lavoratori domestici è rappresentata dagli archivi amministrativi generati dall'acquisizione delle informazioni contenute:

- nelle comunicazioni obbligatorie di assunzione, trasformazione, proroga e cessazione del rapporto di lavoro domestico effettuate dai datori di lavoro (modelli cartacei e comunicazioni telematiche);
- nei versamenti (bollettini di conto corrente postale o online) effettuati dai datori di lavoro per il pagamento dei contributi previdenziali.

Nelle comunicazioni obbligatorie di assunzione, trasformazione, proroga e cessazione del rapporto di lavoro domestico sono disponibili una serie di informazioni sul lavoratore, tra cui il sesso, l'età, la nazionalità, eccetera. Nei versamenti possono desu-

mersi informazioni sui periodi coperti da contribuzione, sulle ore settimanali retribuite, sulla retribuzione oraria, eccetera. Detti archivi, come ogni archivio amministrativo, sono soggetti a aggiornamenti dovuti a nuove acquisizioni e a correzioni delle informazioni già acquisite. In generale tali aggiornamenti riguardano maggiormente l'anno più recente e in maniera via via minore gli altri anni.

Spesa pubblica in agricoltura (INE-00018)

CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

(L'esempio rientra nella casistica "Processo basato su dati amministrativi").

L'indagine nasce nei primi anni Novanta con l'obiettivo di approfondire il tema dell'intervento pubblico nel settore agricolo attraverso l'analisi e la classificazione della spesa erogata a diversi livelli dall'amministrazione centrale e regionale.

La spesa pubblica in agricoltura viene analizzata sulla base dei dati contenuti nei bilanci preventivi e consuntivi delle amministrazioni regionali e centrali, attraverso una metodologia omogenea che ha consentito di quantificare e qualificare le voci che compongono, in maniera diretta o indiretta, l'ammontare del sostegno pubblico al settore e di implementare un'apposita banca dati.

Lo scopo finale del progetto è quello di fornire ai *policy maker*, ma anche ai ricercatori e agli operatori del settore uno strumento per l'analisi e la valutazione delle politiche pubbliche attuate in agricoltura.

Il progetto si avvale del supporto di una rete di monitoraggio per la raccolta sistematica dei dati a livello centrale e regionale. L'informazione raccolta è mediata e interpretata con l'aiuto di esperti e ricercatori del Centro di ricerca politiche e bioeconomia e validata attraverso il coinvolgimento delle amministrazioni competenti.

La Banca Dati SoPiA – Sostegno pubblico in agricoltura contiene informazioni sui trasferimenti e le agevolazioni fiscali beneficate dal settore agricolo nel periodo 2000-2019.

La banca dati è aggiornata con cadenza annuale ed è alimentata da un flusso di informazioni derivanti da fonti ufficiali dei soggetti coinvolti nell'attuazione delle politiche: regioni, ministeri competenti, INPS, Agenzia delle Entrate, eccetera.

I dati raccolti sono di tipo finanziario, anche se non mancano collegamenti con i dispositivi legislativi che ne hanno generato i flussi di spesa e informazioni di tipo qualitativo, mediante un sistema di codifica della spesa appositamente elaborato (tipo di politica, beneficiari, comparto produttivo, eccetera).

Sistema Informativo Nazionale per la Sorveglianza delle Esposizioni Pericolose e delle Intossicazioni (SIN-SEPI) (ISS-00040)

ISS - Istituto superiore di sanità

(L'esempio rientra nella casistica "Processo basato su dati amministrativi").

Le fonti di dati sono i database dei Centri antiveneni (strutture del Servizio sanitario nazionale con competenze tossicologiche, che operano per una corretta diagnosi e gestione delle intossicazioni) dove vengono raccolte le informazioni riguardanti le consulenze effettuate.

Codice	Nome del concetto	Definizione	Linea guida
4.2	Periodicità della raccolta dei dati	Frequenza con cui vengono raccolti i dati.	Indicare la frequenza della raccolta dati (per esempio: mensile, quadrimestrale, annuale o continua).

2. Guida alla compilazione

2.4.2 Periodicità della raccolta dei dati

Esempi

Indagine PIAAC (Programme for the International Assessment of Adult Competencies) (IAP-00021)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

PIAAC è concepito come studio a cicli ripetuti, con cadenza decennale.

Rapporto sulle ipoteche immobiliari iscritte a garanzia di mutui (FIN-00015)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

L'intervallo temporale base di osservazione è l'anno, periodo al quale sono associati gli atti di mutuo con riferimento alla data di iscrizione nei registri della pubblicità immobiliare.

Misure di prevenzione personali e patrimoniali emesse ex l.646/82 (stampo mafioso) (MGG-00004)

Ministero della Giustizia

I prospetti di rilevazione, debitamente compilati dagli uffici interessati, dovranno essere restituiti a questo ministero ogni sei mesi (cioè, al 31 dicembre e al 30 giugno di ciascun anno) entro i 30 giorni successivi alla scadenza semestrale.

Codice	Nome del concetto	Definizione	Linea guida
4.3	Periodicità della diffusione	L'intervallo di tempo in cui le statistiche vengono diffuse in un determinato periodo di tempo.	Indicare la frequenza con cui i dati vengono diffusi, per esempio mensile, trimestrale, annuale.

2.4.3 Periodicità della diffusione

Esempi

Artigiani e commercianti (IPS-00017)

INPS - Istituto nazionale di previdenza sociale

L'aggiornamento dell'Osservatorio ha cadenza annuale e viene effettuato, attraverso la rielaborazione della serie storica decennale, con l'eliminazione dell'anno più vecchio e l'inserimento di quello nuovo.

Movimento dei procedimenti civili presso gli uffici del giudice di pace (MGG-00097)

Ministero della Giustizia

Il monitoraggio fornisce ogni trimestre informazioni a livello nazionale sull'andamento dei procedimenti pendenti civili e penali e dell'arretrato civile.

Diritto allo studio (MUR-00008)

Ministero dell'Università e della Ricerca

La raccolta e l'elaborazione delle informazioni di interesse avviene ad anno accademico concluso; pertanto, nell'anno ($t+1$) sono rilasciati i dati relativi all'anno accademico ($t-1$) - (t) e/o all'anno solare ($t-1$). Ad esempio, nell'anno 2015 sono rilasciati i dati relativi all'anno accademico 2013-2014 e/o all'anno solare 2013.

2.4.4 Raccolta dei dati

Codice	Nome del concetto	Definizione	Linee guida
4.4	Raccolta dei dati	Processo sistematico di raccolta dei dati per le statistiche ufficiali.	<p>Per ciascuna fonte di dati da indagine: descrivere i metodi utilizzati per raccogliere i dati presso i rispondenti; allegare o collegare tramite collegamento ipertestuale i questionari.</p> <p>Per ciascuna fonte di dati amministrativi: descrivere il processo di acquisizione dei dati e come è stato testato.</p> <p>Per tutte le fonti: descrivere le tipologie di controlli applicate durante l'immissione dati.</p>

Per un processo basato su dati da indagine, descrivere:

- il test sulle procedure e sui sistemi di raccolta dei dati e la formazione del personale deputato alla raccolta dei dati;
- le procedure e il programma di gestione delle mancate risposte e indicare in quale misura il programma è stato rispettato;
- le procedure per l'inserimento dei dati raccolti nel database di input per ulteriori trattamenti, nonché le modalità di controllo di tali dati in fase di inserimento.

Per un processo basato su dati amministrativi, descrivere le procedure per il controllo dei dati prima e durante l'inserimento nel database.

Esempi

Panel degli acquisti agroalimentari delle famiglie italiane (ISM-00001)

ISMEA - Istituto di servizi per il mercato agricolo alimentare

(L'esempio rientra nella casistica "rilevazione campionaria").

Per la valutazione quali-quantitativa e il monitoraggio degli acquisti di prodotti agroalimentari delle famiglie italiane l'ISMEA si avvale di un *panel* famiglie gestito in collaborazione con la AC Nielsen. Il processo di rilevazione avviene attraverso la tecnologia *home scanning*. A ogni atto di acquisto la famiglia rileva, attraverso un minicomputer dotato di lettore ottico, tutti i fatti che ne definiscono il comportamento. L'atto di acquisto deve essere rilevato dal soggetto che lo compie e nel momento stesso in cui si completa ("il prodotto entra in casa"). La scansione del prodotto rende disponibile l'informazione a livello di singolo codice, mentre una serie di domande inviate dal minicomputer guidano la famiglia e permettono la precisa osservazione del comportamento di acquisto. A ogni atto di acquisto, infatti, la famiglia rileva tutti i fatti che ne descrivono il comportamento, indicando in modo puntuale: data di acquisto; fonte di acquisto (negozi visitati); individuo acquirente; quantità acquistata per ciascun prodotto; prezzo; promozioni.

Per mantenere uno standard analogo nella registrazione dei prodotti senza codice a barre, viene fornita una guida referenze che permette alla famiglia di individuare i dettagli di questi prodotti. In tale modo vengono raccolte tutte le informazioni relative non solo al *grocery* confezionato, ma anche ai prodotti freschi, sfusi e autoprodotti. La famiglia trasmette direttamente, in via telematica, le informazioni rilevate con cadenza settimanale così da garantirne una lettura veloce e precisa.

Indagine Almalaurea sulla condizione occupazionale dei laureati a dieci anni dal titolo (ALM-00003)

Consorzio interuniversitario AlmaLaurea

2. Guida alla compilazione

(L'esempio rientra nella casistica "rilevazione campionaria").

La rilevazione AlmaLaurea sulla condizione occupazionale è organizzata in modo da realizzare le interviste sostanzialmente alla medesima distanza temporale dal conseguimento del titolo. La metodologia di rilevazione prevede un primo contatto via email; successivamente, tutti coloro che non hanno compilato il questionario online vengono contattati anche telefonicamente. La rilevazione telefonica è di anno in anno affidata a una società esterna, che risulta aggiudicataria di una procedura di gara europea. Per garantire a tutti gli atenei il raggiungimento di un elevato numero di interviste, il contatto telefonico avviene dal lunedì al venerdì dalle 9:30 alle 20:30 e il sabato dalle 10:30 alle 18:30 (salvo diversi accordi presi con i laureati). Per qualsiasi altra esigenza non legata all'indagine telefonica, AlmaLaurea mette a disposizione dei laureati e degli studenti un servizio dedicato di assistenza, accessibile attraverso diversi canali. Tutte le informazioni sono raccolte nel rispetto della normativa vigente in materia di protezione dei dati personali e utilizzate solo a fini statistici.

Rilevazione statistica sugli investimenti intangibili (*Intangible Assets Survey*) (IAP-00012)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

(L'esempio rientra nella casistica "rilevazione campionaria").

Il primo contatto con le imprese del campione è avvenuto tramite Pec in coerenza con le recenti disposizioni di Agenda digitale. È stato, quindi, utilizzato l'invio delle email tramite Pec a circa 25 mila imprese estratte dall'archivio ASIA di Istat e presenti nell'anagrafe Pec delle Camere di commercio. Invece, per le imprese che ancora non si erano adeguate alla nuova normativa, si è fatto ricorso al contatto telefonico. Per la raccolta dati è stata adottata una metodologia mista, con tecniche Cati e Cawi (*Computer-Assisted Web Interviewing*), attraverso un questionario strutturato in otto sezioni. Il questionario è stato informatizzato attraverso lo sviluppo di questi due sistemi di rilevazione per consentire, se richiesto dall'impresa, la raccolta dati mediante intervista telefonica o la registrazione di questionari cartacei inviati per email o fax. La tecnica Cati e Cawi sono state sviluppate tenendo conto dei filtri relativi al questionario e implementando dei controlli sia *hard check* (con blocco della compilazione), sia *soft check*, con la comparsa di specifici messaggi di errore o di messaggi di *warning*. L'utilizzo di una tecnica mista ha consentito di massimizzare l'efficacia della rilevazione, sfruttando al massimo le potenzialità delle due metodologie e garantendo così un alto tasso di risposta del target oggetto di studio. Le imprese che non hanno compilato il questionario, entro la data indicata nella email iniziale, sono state sollecitate tramite diversi *recall* inviati via email e successivamente sollecitate telefonicamente. Al termine della raccolta dati sono pervenuti 10.631 questionari: di questi la quasi totalità è stata compilata con la tecnica Cawi (85 per cento circa), solo una piccola percentuale attraverso Cati (meno del cinque per cento) mentre il 10 per cento circa ha preferito compilare un questionario cartaceo.

Durante la fase di rilevazione sul campo, la società incaricata ha attivato un numero verde per fornire un supporto ai partecipanti; gli intervistati potevano chiamare questo numero di telefono per avere chiarimenti e richiedere assistenza sulla rilevazione. L'Isfol², invece, ha attivato una email dedicata, per tutto il periodo della rilevazione, per fornire assistenza a tutte le imprese che chiedevano informazioni sia su aspetti formali dell'indagine (obbligo di risposta, titolarità dell'indagine, segreto statistico e tutela della riservatezza, eccetera) sia per specifiche problematiche relative ai diversi *asset*; in particolare, sono state ricontattate telefonicamente tutte quelle imprese che ponevano problemi di interpretazione delle definizioni o quesiti molto peculiari.

2 L'Istituto per lo sviluppo della formazione professionale dei lavoratori (Isfol) dal 1° dicembre 2016 ha assunto la denominazione di INAPP.

È stato, inoltre, predisposto un sistema per monitorare l'andamento della rilevazione, con una periodicità giornaliera (numero di accessi al sito, numero di registrazioni, numero di questionari compilati parzialmente e numero di questionari completi inviati) e settimanale (analisi delle variabili di interesse) per tenere sotto controllo la tempistica e il rispetto del disegno campionario. Sulla base delle informazioni ricevute sono stati calcolati alcuni indicatori che hanno consentito di mettere in atto adeguati interventi (*recall* inviati via email alle imprese registrate, solleciti telefonici dopo un mese dall'invio ufficiale, eccetera).

Questionario: <https://www.inapp.gov.it/wp-content/uploads/Non-organizzati/Questionario-IAS-2020.pdf>.

Indagine sulla consistenza e dislocazione territoriale degli appartenenti alle popolazioni di lingua ladina, mòchena e cimbra (Rcdt) (PAT-00044)

Provincia autonoma di Trento

(L'esempio rientra nella casistica “rilevazione censuaria” delle famiglie residenti nella Provincia autonoma di Trento).

L'Istituto di statistica della Provincia di Trento (ISPAT) si è fatto carico della predisposizione dell'applicativo informatico dedicato alla rilevazione, progettato in modo da garantire fin dall'origine che la compilazione del questionario avvenisse in modo anonimo nel rispetto delle norme sul trattamento dei dati personali e del segreto statistico, nonché del monitoraggio della rilevazione, del coinvolgimento dei comuni per la raccolta di eventuali questionari cartacei, della traduzione in lingua di minoranza del questionario e dell'altra documentazione necessaria al buon andamento della rilevazione. L'Istituto ha inoltre provveduto ad attivare un numero verde e un indirizzo email dedicato per i cittadini; inoltre ha realizzato una pagina web dedicata, contenente le informazioni principali relative alla rilevazione. Gli sportelli linguistici hanno provveduto a redigere le traduzioni della lettera, dei questionari e delle istruzioni di compilazione nelle tre lingue di minoranza.

In Trentino-Alto Adige/Südtirol risultavano residenti, al 1° gennaio 2021, 239.412 famiglie, le quali sono state preventivamente informate sulle modalità della rilevazione con un'apposita lettera con le istruzioni per la compilazione del questionario online e le credenziali per accedere all'applicativo. A metà del periodo di rilevazione è stata inviata una nuova lettera alle famiglie per ricordare l'approssimarsi della scadenza per la compilazione del questionario con le credenziali per l'accesso all'applicativo per chi le avesse smarrite.

Per quanto riguarda la rilevazione nelle convivenze anagrafiche, i responsabili di ciascuna convivenza anagrafica sono stati avvisati preventivamente attraverso email della rilevazione e a ciascuna convivenza anagrafica è stato inviato un plico contenente i questionari cartacei, le istruzioni e le buste anonime affinché ciascun residente in queste strutture interessato alla compilazione del questionario potesse partecipare alla rilevazione. È stata predisposta una specifica formazione per gli ufficiali dell'anagrafe di tutti i comuni della provincia autonoma di Trento e consegnato del materiale informativo da esporre sul territorio comunale. L'ISPAT ha assicurato ai comuni assistenza telefonica e via email in modo da agevolarli nei loro compiti.

Nelle località vociate ladine, mòchene e cimbre, nel mese di dicembre si è provveduto a recuperare le mancate risposte tramite i rilevatori che hanno consegnato e poi ritirato presso le famiglie il questionario cartaceo, che riportava sullo stesso modello sia la lingua italiana sia la traduzione nella lingua della minoranza locale. Durante la rilevazione importante è stata la collaborazione degli ufficiali dell'anagrafe che hanno consegnato a

2. Guida alla compilazione

coloro che non disponevano del supporto informatico, o non erano in grado di utilizzarlo, il questionario cartaceo e ne hanno curato la raccolta in forma anonima e la successiva consegna all'ISPAT.

Rilevazione sull'attività estrattiva (TOS-00017)

Regione Toscana

(L'esempio rientra nella casistica “rilevazione censuaria”).

Il titolare dell'autorizzazione è tenuto a fornire al Comune e alla Regione ogni informazione richiesta in ordine all'attività estrattiva; il Comune a sua volta invia entro il mese di marzo di ogni anno le informazioni relative all'andamento delle attività estrattive relative all'anno precedente. Inoltre, i Comuni comunicano alla Regione trimestralmente i controlli effettuati e l'esito degli stessi e, entro il mese di luglio di ogni anno, trasmettono alla Regione la stima relativa all'entità dei contributi di estrazione dell'anno in corso. Il comune risulta essere il primo detentore delle informazioni riguardanti le tipologie e le volumetrie dei materiali estratti in Toscana. Le informazioni del settore estrattivo possono essere comunicate dai comuni tramite procedura telematica che prevede il *login* degli uffici di ogni Comune e la relativa compilazione delle schede di rilevamento.

Relazione al Parlamento sul patrocinio a spese dello Stato nel processo penale (ai sensi dell'art. 294 del DPR n.115/02) (MGG-00048)

Ministero della Giustizia

(L'esempio rientra nella casistica “processo basato su dati amministrativi”).

I prospetti di rilevazione predisposti per gli uffici giudiziari sono due, uno per gli uffici giudicanti ordinari (ossia per i maggiorenni) e l'altro per gli uffici giudicanti per i minorenni.

Le percentuali degli uffici che hanno risposto, seppure non pari al 100 per cento, sono risultate comunque significative ai fini della corretta analisi e valutazione del fenomeno, anche grazie a una attenta stima dei dati mancanti. A causa della persistente difficoltà di acquisizione dei dati presso alcuni uffici giudiziari, problema che di fatto comporta l'assegnazione di un valore pari a zero a tutte le variabili dell'ufficio inadempiente, si è ritenuto opportuno effettuare, per gli anni 2005-2020, una stima dei dati mancanti, procedura che ha consentito di mitigare in modo apprezzabile tale carenza.

Certificato di assistenza al parto (SAL-00038)

Ministero della Salute

(L'esempio rientra nella casistica “processo basato su dati amministrativi”).

L'originale del certificato (CeDAP) viene conservato presso la direzione sanitaria degli istituti di cura pubblici e privati in cui è avvenuto il parto. Nei casi di nascita avvenuta a domicilio o in struttura diversa da istituto di cura pubblico o privato, il certificato deve essere consegnato dall'ostetrica/o o dal medico che ha assistito il parto alla azienda sanitaria di evento, non oltre il decimo giorno dall'evento. Le aziende sanitarie ricevono anche i dati relativi agli eventi di nascita che hanno luogo presso gli ospedali a gestione diretta, classificati e gli istituti di ricovero privati.

I certificati relativi agli eventi di nascita che hanno luogo presso le aziende ospedaliere, gli istituti di ricovero e cura a carattere scientifico e i policlinici universitari, sono invece trasmessi da queste strutture direttamente alle Regioni.

Le Regioni trasmettono i dati semestralmente al Ministero della Salute - Direzione generale del sistema informativo, ufficio di direzione statistica, secondo le seguenti tempistiche:

- entro il 30 settembre di ogni anno, i dati relativi ai parti del primo semestre dell'anno;
- entro il 31 marzo di ogni anno, i dati relativi ai parti del secondo semestre dell'anno precedente ed eventuali correzioni e integrazioni riguardanti il primo semestre.

Dall'anno di riferimento 2016, ai fini del miglioramento continuo della qualità dei dati rilevati dal CeDAP, sono stati messi a disposizione delle regioni e delle province autonome appositi strumenti di monitoraggio online che forniscono:

- *report* dettagliati della qualità di ciascuna delle variabili rilevate, in grado di evidenziare tempestivamente eventuali errori e anomalie riscontrate nel flusso informativo.
- analisi dettagliate per ciascun punto nascita, di alcune informazioni rilevanti del CeDAP, anche attraverso l'integrazione con la fonte informativa della scheda di dimissione ospedaliera.

2.4.5 Trattamento dei dati

Codice	Nome del concetto	Definizione	Linea guida
4.5	Trattamento dei dati	Insieme di operazioni di controllo e trasformazione finalizzate alla produzione di informazioni statistiche a partire dai dati di input.	Descrivere le diverse operazioni applicate ai dati di input per produrre gli output statistici.

In genere, si tratta di operazioni di codifica, integrazione, controllo e correzione, imputazione, derivazione di nuove unità e variabili, calcolo di aggregati semplici o complessi come i numeri indice, applicazione di metodi di stima. Inoltre, vanno riportate le procedure di validazione degli output adottate.

Descrivere le procedure per il controllo e la validazione dei dati di origine.

Descrivere le procedure applicate per garantire una maggiore accuratezza degli output, come le eventuali tecniche di localizzazione e trattamento degli errori e valori anomali e i loro possibili effetti sulle stime, così come le procedure di calibrazione e aggiustamento dei pesi per mancate risposte ed errori di copertura. In caso di applicazione di modelli per la stima, descrivere le eventuali ipotesi di modello applicate quando i dati primari sono mancanti o incompleti.

Rispetto alla validazione degli output, descrivere le procedure per confrontare i dati con i cicli precedenti, con altri dati rilevanti, aspettative generali e altre informazioni di dominio.

Descrivere le procedure di imputazione, i motivi più comuni di imputazione e i tassi di imputazione per le principali variabili. Descrivere le procedure per derivare nuove variabili e per calcolare aggregati e statistiche complesse.

Descrivere le eventuali procedure di aggiustamento dei pesi applicate per compensare gli effetti degli errori di mancata risposta e copertura, inclusa la calibrazione, se utilizzata.

Descrivere le procedure per combinare dati di input provenienti da fonti diverse.

Esempi

Lavoratori agricoli dipendenti (IPS-00025)

INPS - Istituto nazionale della previdenza sociale

L'Osservatorio è caratterizzato da un processo di elaborazione che prevede una serie di controlli allo scopo di individuare e superare incompatibilità, anomalie ed errori sistematici

2. Guida alla compilazione

o casuali che possono manifestarsi negli archivi amministrativi. I controlli statistici vengono effettuati dall'Osservatorio realizzato dal Coordinamento generale statistico attuariale.

Il lavoratore presente in più province è rilevato soltanto nella provincia in cui è stato riscontrato il maggior numero di giornate lavorate. Il numero di lavoratori è ottenuto dalla somma delle unità statistiche (indica le “teste”), quindi nel caso in cui un singolo lavoratore abbia avuto più di un rapporto di lavoro nello stesso mese, viene contato una sola volta.

Il numero dei lavoratori è inferiore a quello della somma dei lavoratori a tempo indeterminato e determinato, in quanto uno stesso lavoratore, può avere assunto, nel corso dell'anno, entrambe le qualifiche. La media annua è calcolata in funzione del numero dei mesi di presenza effettiva; un soggetto presente per soli sei mesi è equivalente a 0,5.

I lavoratori extracomunitari sono identificati dalla cittadinanza ottenuta dall'archivio dei permessi di soggiorno del Ministero dell'Interno. Inoltre, il numero dei lavoratori extracomunitari differisce da quello pubblicato nell'osservatorio sugli stranieri in cui il soggetto è classificato nello stato di lavoratore, pensionato o percettore di prestazioni a sostegno del reddito a seconda della caratteristica prevalente, cioè lo stato nel quale il soggetto ha trascorso la maggior parte del tempo nell'anno.

Movimento dei procedimenti civili presso gli uffici del giudice di pace (MGG-00097)

Ministero della Giustizia

Nei casi di inadempienza, al fine di garantire la completezza e la consistenza della serie storica, i dati vengono elaborati mediante la stima dei valori degli uffici che non hanno fornito i dati, sulla base del loro *trend* storico nonché con opportune tecniche di interpolazione statistica.

Indicatori per il monitoraggio delle acque dolci superficiali ai sensi della Direttiva 2000/60/CE (APA-00049)

ISPRA - Istituto superiore per la protezione e la ricerca ambientale

Descrizione della metodologia di elaborazione per “Acque interne superficiali – elementi di qualità biologica nei fiumi”: l'indice multimetrico finale per il *Macrofauna STAR_ICMi* è ottenuto dalla somma delle metriche normalizzate, ciascuna delle quali è moltiplicata per il proprio peso. Tale peso varia tra le singole metriche mentre assume un unico valore per le tre categorie generali (tolleranza, abbondanza/habitat, ricchezza/diversità). Dopo il calcolo della media ponderata delle metriche, i valori risultanti vengono normalizzati sul valore mediano di *STAR_ICMi* osservato per i siti di riferimento. Per ogni componente biologica monitorata, il risultato degli RQE (Rapporto di Qualità Ecologica, ovvero il rapporto tra valore dell'indice riscontrato nel sito e valore dell'indice relativo alle condizioni di riferimento) viene confrontato con i valori soglia riportati nel decreto del Ministro dell'Ambiente 260/2010 corrispondenti alle cinque classi di stato ecologico previste per ogni componente monitorata. L'indice multimetrico finale per le Diatomee *ICMi* si basa sulla identificazione e conteggio di 400 individui. La formula finale è $ICMi = (EQR_IPS + EQR_TI)/2$.

PLUS: Partecipazione al mercato del lavoro - Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro (IAP-00004)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

Il trattamento statistico si basa sullo studio del sottostante modello di mancata risposta *panel* e di alcuni flussi tra condizioni di interesse primario per l'indagine, quali quello tra disoccupazione e occupazione, da lavoratore “standard” a “non standard”, da part-time a full-time, eccetera.

Attraverso l'utilizzo di algoritmi di classificazione non parametrica *Classification Tree Analysis* si individuano delle sottopopolazioni omogenee rispetto ad alcune caratteristiche individuali, rispetto alle quali specificare i totali noti da imporre come vincoli di calibrazione longitudinale.

Il peso *panel* finale è ottenuto a partire dal peso sezionale corrispondente all'anno di inizio periodo t_0 e applicando un correttore di calibrazione che da una parte tiene conto degli effetti di attrito, dall'altra sintetizza molta dell'informazione utile ai fini della produzione delle principali stime longitudinali. È interessante notare come la selezione endogena (da modello) delle variabili utili ai fini della spiegazione della mancata risposta del *panel* siano rimaste pressoché invariate nel tempo. Per quanto riguarda, invece, lo studio delle transizioni notevoli si è preferito svincolare gli algoritmi di classificazione dai risultati precedenti, garantendo "omogeneità metodologica" piuttosto che una più stringente stabilità nella definizione delle sottopopolazioni di riferimento.

2.5 Mandato istituzionale

Codice	Nome del concetto	Definizione	Linea guida
5	Mandato istituzionale	Leggi, disposizioni normative e regolamentari, altri impegni istituzionali formalizzati che assegnano responsabilità e autorità a un'organizzazione per la raccolta, l'elaborazione e la diffusione di dati statistici.	Disposizioni legislative e regolamentari o altri atti giuridici, che attribuiscono a un ente la responsabilità e l'autorità per la raccolta, l'elaborazione e la diffusione di dati statistici.

Indicare le disposizioni normative e regolamentari o gli altri atti giuridici vincolanti, compresi gli atti giuridici dell'UE e l'attuazione delle direttive dell'UE.

Specificare anche se il processo è presente nel Programma statistico nazionale, atto di programmazione in cui sono stabiliti i lavori statistici di interesse pubblico, e se la fornitura dei dati è obbligatoria ai sensi dell'art.7 del decreto legislativo n. 322 del 1989.

Esempi

L'indagine sul dolore cronico in Italia a partire dai dati dell'EHIS 2019 (ISS-00054)

ISS - Istituto superiore di sanità

L'indagine è prevista nei regolamenti dell'Unione europea che attuano il regolamento (CE) n. 1338/2008 relativo alle statistiche comunitarie in materia di sanità pubblica e di salute e sicurezza sul luogo di lavoro. In particolare, la *wave 2*, condotta in Italia nel 2015, è normata dal regolamento (UE) n. 141/2013; la *wave 3*, effettuata nel 2019, è prevista nel regolamento (UE) n. 255/2018, mentre è in corso di approvazione il regolamento (UE) per la *wave 4* che sarà realizzata nel 2025.

Panorama Spettacolo (MBE-00018)

Ministero della Cultura

L'articolo 5 della legge n. 163 del 30 aprile 1985 "Nuova disciplina degli interventi dello Stato a favore dello spettacolo" istituisce, nell'ambito dell'ufficio studi e programmazione dell'allora Ministero del turismo e dello spettacolo, l'Osservatorio dello spettacolo con i compiti di:

2. Guida alla compilazione

- raccogliere e aggiornare dati e notizie relativi all'andamento dello spettacolo, nelle sue diverse forme, in Italia e all'estero;
- acquisire elementi di conoscenza sulla spesa annua complessiva in Italia, ivi compresa quella delle regioni e degli enti locali, e all'estero, destinata al sostegno e alla incentivazione dello spettacolo;
- elaborare documenti di raccolta e analisi di tali dati e notizie, che consentano di individuare le linee di tendenza dello spettacolo nel suo complesso e dei singoli settori di esso sui mercati nazionali e internazionali.

Nell'ambito dei suoi compiti, l'Osservatorio dello spettacolo realizza i documenti di ricerca prodotti nell'ambito del Progetto Panorama Spettacolo (il Progetto Panorama Spettacolo è presente nel PSN).

2.6 Politica della riservatezza

Codice	Nome del concetto	Definizione	Linea guida
6	Politica della riservatezza	Disposizioni normative e regolamentari o altri impegni formali che impediscono la divulgazione non autorizzata di dati che consentono di identificare direttamente o indirettamente una persona o un'entità economica.	Illustrare le disposizioni legislative e regolamentari o gli altri atti riguardanti la riservatezza. Descrivere la normativa applicabile (se presente).

Tutti i dati prodotti nell'ambito dei processi statistici condotti dai soggetti appartenenti al Sistan sono sottoposti alla medesima disciplina in materia di tutela della riservatezza e di accesso ai dati per fini scientifici, in conformità con le normative vigenti. Alla luce di ciò, nella compilazione del campo dedicato alla riservatezza, è opportuno trattare in maniera articolata i seguenti tre aspetti:

- *policy*: indicare se il personale coinvolto nell'indagine è tenuto a sottoscrivere un impegno di riservatezza, in base alle azioni e misure adottate in relazione ai trattamenti;
- linee guida generali e coordinamento: indicare chi può avere accesso ai dati riservati e in quali circostanze, compresi gli obblighi del personale e l'accesso degli utenti ai microdati per scopi di ricerca;
- sicurezza: indicare quali misure di sicurezza, se presenti, sono state introdotte tenendo conto della riservatezza.

Verificare se, come prassi:

- i dati personali, cioè i dati che consentono l'identificazione diretta e indiretta, vengono rimossi il più presto possibile nel processo di produzione statistica e se vengono adottate adeguate tecniche di pseudonimizzazione;
- l'accesso ai dati personali è limitato e giustificato.

Inserire, inoltre, i riferimenti per contattare il responsabile della protezione dei dati.

Aggiungere informazioni sulle regole applicate per il trattamento dei *dataset* al fine di garantire la riservatezza statistica e impedire la divulgazione non autorizzata:

- per output aggregati: fornire le regole che definiscono una cella confidenziale; descrivere le procedure, individuare le celle confidenziali (riservatezza primaria) e controllare la divulgazione residua (derivazione o riservatezza secondaria); descrivere le procedure per ridurre il rischio di divulgazione mediante il trattamento delle celle confidenziali, ad esempio mediante perturbazione, arrotondamento controllato, soppressione delle celle o aggregazione di modalità;

- per output a livello micro: descrivere le procedure adottate per proteggere la riservatezza.

Se gli utenti esterni possono accedere ai microdati per scopi di ricerca, descrivere le disposizioni sulla riservatezza applicate. Descrivere le procedure per garantire la sicurezza dei dati durante la raccolta, l'elaborazione, l'analisi e la diffusione dei dati, prevenendo così l'accesso non autorizzato.

Esempi

Rete di informazione contabile agricola (RICA) (INE-00001)

CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria Riservatezza

I dati raccolti nella RICA sono protetti e tutelati dal segreto statistico e sottoposti alle normative comunitarie (regolamento UE 2016/679) e nazionali (d.lgs. 196/2003) sulla protezione dei dati personali, anche per successivi trattamenti, esclusivamente per finalità statistiche da parte della Commissione europea e dai soggetti afferenti al Sistan e potranno, inoltre, essere comunicati per finalità di ricerca scientifica alle condizioni e secondo le modalità previste dalle disposizioni vigenti per il trattamento di dati personali effettuato nell'ambito del sistema statistico nazionale. I medesimi dati vengono diffusi in forma aggregata, in modo tale che non sia possibile risalire ai soggetti che li hanno forniti o a cui si riferiscono. I regolamenti comunitari ne vietano un uso diverso da quello della valutazione delle politiche agricole e per studi e analisi sul settore. I dati della RICA non possono essere utilizzati per controlli amministrativi o per verifiche fiscali.

Ai sensi delle normative vigenti, titolare del trattamento dei dati personali raccolti con la RICA è il CREA e responsabili del trattamento per il CREA sono, per le diverse fasi del processo di rilevazione e gestione di rispettiva competenza, il direttore *protempore* del centro di politiche e bioeconomia del CREA e il direttore generale del CREA.

PITER-HCV (ISS-00055)

ISS - Istituto superiore di sanità

Confidenzialità e sicurezza dei dati

La procedura di arruolamento e la raccolta dei dati saranno centralizzate attraverso un sistema web appositamente predisposto per il progetto PITER. L'accesso al sistema sarà criptato e possibile solo grazie a *username* e *password* nominale assegnata al medico responsabile dello studio presso il centro clinico. Durante le fasi di registrazione verrà automaticamente generato un codice univoco anonimo, ovvero non desumibile dai dati personali identificativi degli interessati. La lista di corrispondenza tra i dati anagrafici e i codici identificativi dei pazienti dovrà essere conservata presso il centro clinico che ne sarà responsabile.

Durante le operazioni di elaborazione dei dati memorizzati nel database centralizzato, saranno adottate specifiche misure per incrementare il livello di sicurezza dei dati acquisiti, tramite:

- idonei sistemi di autenticazione e di autorizzazione per le esigenze di accesso e trattamento, avendo cura di utilizzare credenziali di validità limitata alla durata dello studio e di disattivarle al termine dello stesso;
- procedure per la verifica periodica della qualità e coerenza delle credenziali di autenticazione e dei profili di autorizzazione assegnati agli incaricati del trattamento;

2. Guida alla compilazione

- sistemi di *audit log* per il controllo degli accessi al database e per il rilevamento di eventuali anomalie.

Il centro potrà accedere ai propri dati attraverso la stessa piattaforma web utilizzata per l'inserimento dei dati. La piattaforma del Progetto PITER è conforme al regolamento (UE) 2016/679. I dati raccolti mediante l'applicazione online Progetto PITER verranno ubicati su infrastruttura ISO 27001.

Aspetti etici

Prima dell'arruolamento, ogni paziente dovrà sottoscrivere un modulo di consenso informato per la partecipazione allo studio PITER. Lo studio verrà condotto in accordo con la dichiarazione di Helsinki e con la normativa italiana vigente sulla ricerca biomedica. Il protocollo dovrà essere approvato dal comitato etico dell'Iss e successivamente dal comitato etico dei centri partecipanti.

Trattamento dei dati personali

Il trattamento dei dati personali raccolti e dei campioni biologici eventualmente conservati avverrà in forma pseudoanonimizzata, mediante assegnazione a ogni paziente, da parte del centro di riferimento, di un codice univoco alfanumerico personale (Codice Studio PITER). Nessuno, a eccezione del medico specialista o specializzando che ha in cura il paziente presso il centro di riferimento, potrà risalire all'identità del paziente. La lista di corrispondenza tra i dati anagrafici e il codice identificativo sarà conservata presso il centro di riferimento, che ne sarà responsabile. Pertanto, il centro di riferimento ha la responsabilità di proporre le informazioni *privacy* e fare sottoscrivere il consenso al trattamento dei dati personali, in cui dovrà indicare il titolare del trattamento dei dati, al fine di rispettare la normativa sulla tutela dei dati personali ai sensi del regolamento (UE) 2016/679. Sarà inoltre conservata presso il centro di riferimento la modulistica relativa al consenso al trattamento dei dati personali, attraverso misure idonee a garantire la sicurezza delle informazioni in essa contenute, e in luogo separato rispetto alla lista di corrispondenza. Il consenso al trattamento dei dati personali è una condizione essenziale ai fini della partecipazione allo studio.

I dati raccolti dal medico curante specialista del centro clinico di riferimento durante la visita di arruolamento del paziente e nel corso delle varie visite di *follow-up* verranno caricati in forma pseudoanonimizzata all'interno di un database centralizzato creato appositamente per il progetto PITER e in dotazione presso l'ISS. La conservazione di tali dati avverrà secondo modalità digitali per tutto il tempo necessario alla conduzione dello studio in oggetto; al termine, i dati verranno cancellati, salvo la possibilità di permettere la conservazione per un tempo illimitato qualora gli stessi vengano adeguatamente anonimizzati. La conservazione del materiale biologico raccolto avverrà nel centro partecipante presso il quale è stato raccolto, e non verrà in alcun modo trasferito all'Iss o ad altri centri partecipanti. Il personale addetto allo studio, i rappresentanti del comitato etico e delle autorità sanitarie italiane, potranno conoscere i dati contenuti anche nella documentazione clinica originale, sempre in conformità con la legislazione vigente per gli studi farmacologici clinici.

Elaborazione sulle biblioteche (LIG-00006)

Regione Liguria



Per le biblioteche che non avessero ancora inviato la richiesta di registrazione ma fossero interessate, nella pagina è presente il modello da utilizzare (in formato PDF) e l'informatica *privacy*³.

2.7 Gestione della qualità

La definizione di un *framework* per la qualità interno all'organizzazione risponde all'esigenza di perseguire i seguenti obiettivi:

- fornire un meccanismo sistematico per l'identificazione continua dei problemi di qualità e delle possibili azioni per la loro risoluzione, dai miglioramenti incrementali alla reingegnerizzazione su larga scala;
- fornire una base per la creazione e il mantenimento di una cultura della qualità;
- sostenere il miglioramento continuo della qualità;
- stimolare l'interazione tra il personale in tutta l'organizzazione;
- accogliere materiale di riferimento utile nella formazione;
- garantire la trasparenza dei processi.

Codice	Nome del concetto	Definizione	Linee guida
7	Gestione della qualità	L'insieme dei sistemi, delle strutture e delle procedure per gestire la qualità dei prodotti e dei processi statistici all'interno della organizzazione.	Descrivere l'impianto di <i>quality management</i> adottato. In particolare, fare riferimento al Codice italiano per la qualità delle statistiche ufficiali e alla relativa Guida all'implementazione. Se il processo è stato sottoposto a valutazione della qualità, riassumerne i risultati. Descrivere azioni di miglioramento del processo in corso o pianificate. Se possibile, riportare punti di forza e limitazioni dell'output prodotto.

Come descritto nel Capitolo 1, il vigente quadro di riferimento nazionale per la qualità prevede che gli enti Sistan qualificati come Autorità statistiche nazionali aderiscano al Codice delle statistiche europee. Parallelamente, il Codice italiano per la qualità delle statistiche ufficiali estende e calibra il codice europeo all'intera statistica ufficiale prodotta in ambito nazionale. In aggiunta, la *Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali* (Istat 2023) elenca una serie di metodi che offrono indicazioni pratiche per conformarsi ai principi e ai criteri stabiliti dal Codice italiano. In questa sezione si può fare riferimento alle attività in corso nell'ente per soddisfare i principi e i criteri del Codice italiano e i metodi della *Guida per l'implementazione*.

Si può documentare l'esistenza di strutture, come un'unità organizzativa o un comitato interno, incaricati della gestione della qualità. È possibile evidenziare la presenza di programmi di formazione sulla metodologia statistica e sui concetti di qualità, di linee guida sulla qualità per la produzione statistica sviluppate internamente o di fonte esterna, come le *Linee guida per la qualità delle statistiche del Sistema statistico nazionale* (Istat 2018), la disponibilità di documentazione e indicatori di qualità, nonché la produzione di *quality report*, eccetera.

³ Cfr. <https://www.regione.liguria.it/component/publiccompetitions/document/54714:privacy-anagrafe-biblioteche.html?Itemid=13366>.

2. Guida alla compilazione

Descrivere le procedure applicate al processo statistico a tutela della qualità, ad esempio corsi di formazione, monitoraggio del processo, *benchmarking*, valutazioni e utilizzo delle migliori pratiche.

Inoltre, includere descrizioni delle attività di valutazione a cui è stato sottoposto il processo statistico, come *audit* o autovalutazione, e gli eventuali risultati di tale valutazione. Si può fornire una sintesi delle caratteristiche principali della qualità del prodotto, ad esempio in termini delle dimensioni della qualità: si può evidenziare, se presenti, limitazioni e punti di forza.

Si noti che l'ulteriore documentazione metodologica disponibile, in termini di elenco e link, va riportata alla voce 13.6 “Documentazione metodologica”, anche se il contenuto di tale documentazione va descritto in questa sezione.

Esempi

Indicatori per il monitoraggio delle acque dolci superficiali ai sensi della Direttiva 2000/60/CE (APA-00049)

ISPRA - Istituto superiore per la protezione e la ricerca ambientale

Qualità dell'informazione: l'indicatore risponde alle richieste della normativa di settore e si basa su indici consolidati e in uso su tutto il territorio. La comparabilità temporale e spaziale è garantita dall'uso su scala nazionale della stessa metodologia di analisi, campionamento e classificazione e, a livello internazionale, dal processo di intercalibrazione dei metodi con gli altri Stati membri.

Utenza e attività dei servizi pubblici per le tossicodipendenze (SERT) (SAL-00012)

Ministero della Salute

Al fine di condividere gli aspetti principali del Sistema informativo nazionale dipendenze (SIND) è stato attivato un apposito gruppo di lavoro, costituito dai rappresentanti delle regioni, che ha supportato la realizzazione dello studio di fattibilità; sulla base dei risultati dello studio, è stato implementato il sistema informatico a supporto. Il gruppo di lavoro ha contribuito alla redazione dello schema del decreto istitutivo e relativo disciplinare tecnico, alla definizione delle specifiche funzionali e da ultimo alla stesura del manuale operativo. Il SIND è stato messo a disposizione delle regioni a partire dal 2011 e, il gruppo di lavoro partecipa a incontri periodici per garantire la corretta interpretazione dei dati raccolti e per analizzare le eventuali criticità legate all'alimentazione del flusso. Tale modalità di lavoro permette di mantenere il flusso sempre aggiornato rispetto alle modificazioni strutturali e normative che caratterizzano lo specifico fenomeno assistenziale.

A partire dal 2012 il gruppo di lavoro ha demandato a uno specifico sottogruppo l'individuazione di un set di indicatori rappresentativo del fenomeno di assistenza. Gli indicatori, sintetizzati in un documento descrittivo, sono stati discussi nel dettaglio individuandone finalità e specifiche di calcolo.

Disabilità, DSA e accesso alla Formazione universitaria (ANV-00002)

ANVUR - Agenzia nazionale di valutazione del sistema universitario e della ricerca

Ogni università ha fornito un nominativo di una persona responsabile, che ha avuto un ruolo di coordinamento interno agli atenei per il reperimento delle informazioni richie-

ste, con la supervisione del delegato. La struttura del questionario, infatti, richiedeva dati che di norma vengono alimentati e gestiti da uffici diversi, non sempre accessibili su piattaforme locali integrate. Proprio la natura trasversale delle tematiche legate all'inclusione sociale rappresenta una criticità per la raccolta e il monitoraggio dei dati a livello centrale, da considerare nella prospettiva evolutiva di questo progetto. La ricerca ha messo in luce che è necessario avere una unica rilevazione per quanto riguarda il fenomeno degli studenti con disabilità e disturbi specifici dell'apprendimento (DSA). Una simile base dati dovrebbe essere alimentata da rilevazioni periodiche in grado di produrre informazioni e dati attendibili, utili alla programmazione delle politiche per il diritto allo studio rivolte agli studenti con disabilità o con DSA e alla valutazione del loro impatto. A tale proposito essa dovrebbe essere inserita in un più ampio sistema di rilevazione che preveda anche l'utilizzo delle base dati già esistenti che, con gli opportuni adeguamenti, potrebbero essere utilizzate per conoscere altri importanti aspetti della vita universitaria degli studenti con disabilità o con DSA, come accade con l'Anagrafe nazionale degli studenti per la generalità di essi.

2.8 Pertinenza

La pertinenza è la misura e il grado in cui il contenuto di un processo soddisfa le esigenze di diversi utilizzatori. Dipende dal fatto che vengano prodotte tutte le statistiche necessarie e dal grado in cui i concetti utilizzati (definizioni, classificazioni, eccetera) riflettano le esigenze degli utenti.

Le esigenze e la soddisfazione degli utenti sono inoltre correlate ad altre componenti della qualità, come accuratezza, tempestività e comparabilità, che vengono discusse in paragrafi separati. Ma, per scopi pratici, una panoramica completa delle esigenze e della soddisfazione degli utenti è fornita in questa apposita sezione del *report*.

Codice	Nome del concetto	Definizione	Linea guida
8	Pertinenza	Il grado in cui le informazioni statistiche soddisfano le esigenze attuali e potenziali degli utenti.	Descrivere gli utenti e le loro esigenze specifiche rispetto ai dati statistici. Descrivere le misure per valutare la soddisfazione degli utenti.

Esistono esigenze differenti a seconda delle categorie di utenti, ad esempio: governo, aziende, mondo accademico e il pubblico in generale. La sfida dell'ente è quella di saperizzare e bilanciare le esigenze (eventualmente contrastanti) degli utenti attuali e potenziali e di produrre statistiche che soddisfino le esigenze più importanti, o la maggior parte di esse, entro i limiti delle risorse disponibili.

La descrizione delle esigenze degli utenti dovrebbe considerare tutte le componenti della qualità. Queste sono espresse non solo in termini di contenuto dei dati, ma anche in termini di grado di accuratezza richiesto, tempistica, modalità di diffusione, metadati necessari per l'interpretazione e relazione con altri output statistici rilevanti. In altre parole, coprono l'intera gamma delle componenti della qualità dell'output.

Il primo passo è raccogliere informazioni sugli utenti: chi sono, quanti sono e quanto sono importanti individualmente e collettivamente. Sulla base delle informazioni disponibili, ad esempio, da comitati consultivi, elenchi di destinatari paganti e accessi a Internet,

2. Guida alla compilazione

l'approccio usuale è quello di sviluppare una classificazione degli utenti, stimare il numero di ogni tipo e identificare gli utenti chiave, ovvero gli utenti che sono sufficientemente importanti da richiedere una considerazione individuale.

Il secondo passo consiste nel determinare le esigenze di ogni gruppo di utenti in termini di popolazione di riferimento ideale e contenuto dei dati e, nel caso di utenti chiave, le loro esigenze individuali. L'ottenimento di informazioni sulle esigenze degli utenti e sugli usi per cui sono necessari i dati viene in genere realizzato tramite comitati consultivi specifici per dominio, categorie di utenti, *focus group* ad hoc, richieste, reclami, sondaggi e altri *feedback* degli utenti. Dovrebbe essere fatta una valutazione degli output chiave desiderati dalle diverse categorie di utenti e dovrebbero essere menzionate eventuali carenze negli output per importanti singoli utenti e categorie di utenti. Le lacune potrebbero riguardare, ad esempio, la non disponibilità di alcuni dati, definizioni inappropriate, suddivisione insufficiente dei dati in sottodomini, serie temporali troppo brevi o output troppo poco frequenti, ad esempio trimestrali anziché mensili. Non tutte le esigenze degli utenti possono essere soddisfatte, per ragioni di bilancio e/o tecnici.

Il terzo passo consiste nel determinare, in termini generali, le priorità da assegnare agli utenti chiave e alle diverse categorie di utenti nel soddisfare le loro esigenze. Ad esempio, le esigenze dei responsabili delle politiche governative potrebbero essere prioritarie rispetto a quelle dei ricercatori accademici. Alcune di queste sono importanti ma transitorie; alcuni utenti potrebbero anche essere rispondenti e le loro esigenze meritano un'attenzione particolare.

In sintesi, il rapporto dovrebbe contenere una classificazione degli utenti, i nomi completi degli utenti chiave, un'indicazione degli usi per i quali gli utenti desiderano gli output, le priorità nel soddisfare le loro esigenze e un resoconto di come tutte queste informazioni sono state ottenute e utilizzate.

La soddisfazione dell'utente è la priorità assoluta. Il metodo di valutazione più completo è un'indagine sulla soddisfazione degli utenti su vasta scala, condotta secondo le migliori pratiche: estraendo un campione rappresentativo di utenti da un elenco appropriato (se disponibile), progettando e testando un questionario adatto, raccogliendo, elaborando e analizzando i risultati, eccetera.

Condurre un'indagine sulla soddisfazione degli utenti non è sempre conveniente, in particolare per piccoli processi statistici in cui rappresenterebbe una quota significativa del *budget* totale dell'attività. Altri metodi di valutazione comprendono l'analisi delle vendite di pubblicazioni, commenti degli utenti, richieste e reclami ricevuti, accessi al sito web e *feedback* da parte di comitati consultivi e *focus group*. I metodi utilizzati per la valutazione e le misure adottate per migliorare la soddisfazione degli utenti dovrebbero essere descritti. Qualora disponibili e pertinenti, è opportuno presentare i risultati di un'indagine sulla soddisfazione degli utenti, o le analisi condotte, suddivisi per le categorie più rilevanti di utenti, con riferimento di informazioni più complete reperibili altrove.

Esempi

Rete di informazione contabile agricola (RICA) (INE-00001)

CREA - Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria

Compiti istituzionali e utilizzatori

Compito primario della RICA è quello di soddisfare i bisogni informativi della UE per la definizione e la valutazione della Politica agricola comunitaria (PAC). I dati della RICA rappresentano la principale fonte informativa sia per la Commissione europea sia per i paesi membri, per valutare l'impatto delle proposte di modifica della PAC attraverso la simulazione di diversi scenari sulla sostenibilità aziendale (economica, ambientale, sociale e delle innovazioni).

La RICA è utilizzata per la giustificazione degli aiuti pubblici all'agricoltura cofinanziati dall'Unione europea, per la valutazione dell'importanza delle imprese agricole come fornitori di beni comuni. Le informazioni raccolte con la RICA consentono inoltre di rispondere ai fabbisogni della ricerca e dei servizi alla consulenza aziendale, attraverso una serie di variabili e indici sulle caratteristiche tecniche, economiche, patrimoniali e reddituali delle aziende agricole.

Il quadro informativo della RICA italiana, molto più ampio rispetto delle esigenze istituzionali della Commissione europea, consente di realizzare analisi su diversi temi che vanno dalla produttività delle aziende agricole ai costi di produzione, dalla sostenibilità ambientale al ruolo della famiglia agricola.

Il database RICA viene utilizzato anche da utenti esterni alle istituzioni comunitarie: Ministero dell'Agricoltura, Regioni, università e istituti di ricerca, così come valutatori indipendenti, organizzazioni professionali e rappresentanti dei produttori agricoli, che richiedono informazioni utili a definire il contesto entro il quale sono attuate le misure di politica agricola e di sviluppo rurale.

Conto annuale delle spese di personale delle amministrazioni pubbliche (TES-00021)

Ministero dell'Economia e delle Finanze - Settore ex Tesoro

Il Dipartimento della Ragioneria generale dello Stato cura dal 1993, quale obbligo istituzionale, la rilevazione del "Conto annuale", riguardante la consistenza del personale e le relative spese sostenute dalle Amministrazioni pubbliche.

Tali informazioni costituiscono la base informativa ufficiale per le determinazioni del Governo e del Parlamento in materia di personale pubblico nonché, per espressa disposizione di legge, per la quantificazione dei costi dei rinnovi contrattuali.

La rilevazione consente in particolare:

- alla Corte dei conti di redigere il referto annuale al Parlamento sulla gestione delle risorse finanziarie impiegate dalle Amministrazioni pubbliche per le prestazioni lavorative del personale ai sensi dell'art. 60, comma 4 del d.lgs. 165/2001 e di attuare gli specifici compiti di controllo in materia di contrattazione integrativa previsti dall'articolo 40bis del medesimo decreto legislativo;
- al Ministero dell'Economia e delle Finanze - Dipartimento della Ragioneria generale dello Stato di attuare, con riferimento al mondo del lavoro pubblico, i propri compiti di controllo e monitoraggio degli andamenti di finanza pubblica;
- al Ministero dell'Interno - Dipartimento degli Affari interni e territoriali di effettuare il censimento del personale degli enti locali (CePEL), previsto dall'art. 95 del d.lgs. 18 agosto 2000, n. 267;
- ai soggetti istituzionalmente destinatari ai sensi del titolo V del d.lgs. n.165 del 2001 (Corte dei conti, Parlamento, Dipartimento della Funzione pubblica), di effettuare, per lo svolgimento delle funzioni di rispettiva competenza, analisi delle dinamiche occupazionali e di spesa del personale delle pubbliche amministrazioni.

2. Guida alla compilazione

Tutte le informazioni di organico sono presentate con il dettaglio uomini e donne, realizzando per il pubblico impiego una statistica di genere.

Indice dei prezzi dei mezzi correnti di produzione dell'agricoltura (ISM-00002)

ISMEA - Istituto di servizi per il mercato agricolo alimentare

Il monitoraggio dei livelli e dell'andamento dei prezzi dei fattori di produzione impiegati in agricoltura e dei prezzi dei prodotti agricoli venduti consente di ottenere un valido indicatore della possibile evoluzione della redditività delle produzioni agricole e rappresenta uno strumento importante per valutare *ex ante* ed *ex post* l'impatto delle politiche economiche e agricole.

Con il progressivo smantellamento del sistema di sostegno dei prezzi dei prodotti agricoli creato dalla PAC, è aumentata anche la domanda di informazione da parte degli agricoltori sull'evoluzione dei prezzi, sia dal lato delle vendite delle proprie produzioni sia dal lato degli acquisti dei mezzi di produzione. Infatti, l'ampliamento delle scelte decisionali degli imprenditori conseguente alla riforma della PAC del 2003 e degli anni successivi, che hanno introdotto il pagamento unico aziendale disaccoppiato dalle produzioni, richiede corrispondentemente la disponibilità di dati costantemente aggiornati, affidabili, fruibili e rapidamente consultabili sui prezzi e sui costi delle diverse produzioni agricole, al fine di orientare le decisioni produttive.

2.9 Accuratezza

Codice	Nome del concetto	Definizione	Linee guida
9	Accuratezza	Vicinanza tra le stime e i valori reali non noti	Descrivere le principali fonti di errori casuali e sistematici negli output statistici e fornire una valutazione riassuntiva di tutti gli errori tenendo conto del possibile impatto sulle stime. La valutazione dell'eventuale distorsione nelle stime può essere espressa in termini quantitativi e/o qualitativi. Riportare i valori di rilevanti indicatori di qualità, come i coefficienti di variazione o il tasso di risposta.

Il concetto di accuratezza esprime la relazione tra una stima e il suo valore reale: più la stima si avvicina al valore vero, maggiore è la sua accuratezza. La differenza tra la stima e il valore reale è definita "errore della stima". L'errore rappresenta quindi il grado di mancanza di accuratezza. Tuttavia, poiché il valore reale è raramente disponibile, l'accuratezza viene valutata attraverso stime parziali o indicatori indiretti, sia quantitativi sia qualitativi.

L'errore totale può generalmente essere scomposto in due componenti: una casuale (varianza) e una sistematica (distorsione o *bias*). Queste derivano dagli errori che si verificano lungo il processo produttivo e che si distinguono, in genere, in errore campionario, dovuto all'osservazione di un sottosinsieme della popolazione, ed errore non campionario, che include tutte le altre fonti di errore.

Errori campionari e non campionari possono avere impatto sulle stime finali in termini di distorsione o variabilità. L'errore campionario impatta principalmente sulla variabilità della stima, soprattutto se si adotta un campionamento probabilistico.

I principali errori non campionari sono:

- errori di copertura: errori dovuti alla divergenza tra la popolazione oggetto dell'indagine e la popolazione obiettivo;
- errori di misurazione: errori che si verificano durante la raccolta dei dati;
- errori di mancata risposta: errori che derivano dalla mancata risposta ad alcune, o a tutte, le domande;
- errore di trattamento: errori nell'elaborazione dei dati dopo la raccolta, ad esempio durante la registrazione, la codifica, l'integrazione;
- errori di assunzione del modello: errori dovuti ai modelli specifici del dominio necessari per definire l'obiettivo della stima.

Sebbene alcuni errori di misurazione possano avere componenti casuali che contribuiscono alla varianza delle stime, la distorsione rappresenta l'effetto principale di tutte le fonti di errore non campionario. La sua misurazione, tuttavia, risulta spesso complessa e/o costosa. In genere, è possibile avere un'idea informata sulla direzione della distorsione, ma raramente è possibile stimarne la dimensione. Per questo motivo, nel *quality report* è richiesta almeno una valutazione qualitativa.

Alcuni indicatori di qualità, come il tasso di mancata risposta, devono essere documentati, poiché rappresentano segnali di allarme per la possibile presenza di distorsione.

In generale, in qualsiasi processo statistico la sezione dedicata all'accuratezza dovrebbe fornire un quadro sintetico che identifichi le principali fonti di errore. Ogni componente di errore andrebbe trattata in base alla sua importanza relativa: se un tipo di errore prevale sugli altri, se ne dovrebbe fornire un'indicazione della sua dimensione, ove possibile.

2.9.1 Indagini con campionamento probabilistico

La classificazione degli errori sopra riportata si applica in modo particolarmente appropriato alle indagini basate su campionamento probabilistico. In questi casi, l'approccio più diretto per valutare l'accuratezza consiste nell'elencare le principali fonti di errore e identificare quelle più significative.

- Errore campionario

La teoria del campionamento fornisce tecniche per la stima della varianza, permettendo di quantificare e controllare la variabilità dovuta al campionamento già in fase di progettazione. La distorsione da campionamento è, inoltre, generalmente trascurabile grazie alla proprietà di correttezza asintotica degli stimatori.

La variabilità di uno stimatore attorno al suo valore atteso può essere espressa dalla sua varianza, dall'errore standard (radice quadrata della varianza dello stimatore), dal coefficiente di variazione (rapporto tra errore standard e valore atteso dello stimatore) e dall'intervallo di confidenza (intervallo entro cui il valore vero si trova con una certa probabilità).

Di solito, l'errore standard non è adatto per la reportistica poiché può risultare di difficile interpretazione per l'utente medio. Si raccomanda di riportare nel *quality report* il *range* assunto dai coefficienti di variazione delle variabili prodotte o il valore del coefficiente di variazione per le principali variabili.

- Errore di copertura

L'errore di copertura, o errore di lista, deriva da discrepanze tra la popolazione inclusa nella lista da cui viene estratto il campione e la popolazione obiettivo. Si distinguono:

- sottocopertura: alcune unità della popolazione obiettivo non sono presenti nella lista;

2. Guida alla compilazione

- sovracopertura: la lista contiene unità che non appartengono alla popolazione obiettivo (ad esempio, imprese cessate ancora presenti nel registro delle imprese);
- duplicazioni: alcune unità della popolazione obiettivo compaiono più volte nella lista.

La sottocopertura è spesso il problema più critico, poiché è difficile da quantificare; di conseguenza, nel *quality report* viene solitamente indicata solo in termini qualitativi. La sovracopertura, invece, è più facilmente rilevabile, spesso durante la fase di raccolta dati, quando si individuano unità che non appartengono alla popolazione obiettivo, e può quindi essere valutata in termini quantitativi, così come gli errori di mancata risposta.

La lista dovrebbe includere esclusivamente le unità appartenenti alla popolazione obiettivo, insieme alle informazioni necessarie per il loro contatto, come indirizzo, numero di telefono e indirizzo email. Se tali informazioni risultano errate, si parla di errori di lista che possono generare mancata risposta. La lista, inoltre, contiene informazioni ausiliarie utili a migliorare l'efficienza del campione, come variabili impiegate per la stratificazione. Errori in queste variabili possono determinare un aumento dell'errore campionario, riducendo la precisione delle stime.

- Errore di misurazione

L'errore di misurazione si verifica durante la raccolta dei dati e si manifesta quando il valore rilevato per una variabile non corrisponde al valore reale. In questo contesto, il termine "misurazione" si riferisce alla rilevazione a livello di unità, ad esempio il reddito mensile di una persona o il fatturato annuo di un'azienda. Il risultato di una misurazione può essere visto come la somma del valore vero e di un termine di errore, che è pari a zero solo in caso di misurazione perfetta.

Le cause possono essere diverse e riguardano vari aspetti del processo di raccolta dati:

- strumento di indagine: il modulo, il questionario o il dispositivo di misurazione utilizzato possono portare alla registrazione di valori errati. Questo può dipendere dalla formulazione delle domande, dall'ordine in cui vengono presentate o dal contesto in cui vengono poste;
- rispondenti: i rispondenti possono, consapevolmente o inconsapevolmente, fornire dati errati a causa di confusione, ignoranza, negligenza o disonestà;
- intervistatore: una formazione inadeguata o una supervisione insufficiente possono portare l'intervistatore a influenzare in modo involontario o improprio le risposte fornite dai rispondenti.

Gli errori di misurazione possono essere di natura sistematica o casuale. Gli errori casuali sono spesso legati all'idea di replicazione del processo: se la misurazione viene ripetuta più volte sulla stessa unità in condizioni costanti, i valori di misurazione registrati varieranno in modo casuale, mentre l'errore sistematico rimarrà costante. Di conseguenza, gli errori di misurazione possono introdurre sia distorsioni sia variabilità aggiuntiva degli output statistici, con la distorsione che rappresenta generalmente il problema principale. La distorsione e la variabilità causate da errori di misurazione possono essere valutate, ad esempio, attraverso reinterviste su un sottocampione di rispondenti.

È possibile identificare errori di misurazione attraverso il *data editing*, quando l'errore genera incongruenze nelle informazioni raccolte. Gli errori identificati possono quindi essere corretti in vari modi, come ricontattando il rispondente o utilizzando tecniche di imputazione (il tasso di imputazione può infatti rappresentare un indicatore indiretto dell'entità dell'errore di misurazione).

Quando gli errori di misurazione sono significativi, le misure adottate per ridurli dovrebbero essere riportate nel *quality report*. Ad esempio, gli sforzi compiuti nella progettazione e nel test del questionario andrebbero brevemente descritti.

- Errore di mancata risposta

Possono verificarsi due tipi di mancata risposta:

- mancata risposta totale, quando non viene ottenuta alcuna informazione da un'unità nel campione;
- mancata risposta parziale, quando i valori di una o più variabili, ma non tutte, non vengono ottenuti per un'unità.

La mancata risposta viene misurata mediante due tipologie di tassi, definiti come segue:

- tasso di mancata risposta totale: complemento a uno (o a 100, se il tasso è espresso in percentuale) del rapporto tra il numero di rispondenti (unità per le quali sono stati raccolti dati per alcuni o tutti i quesiti oggetto di indagine) e il numero totale di unità eleggibili nel campione;
- tasso di mancata risposta parziale: complemento a uno (o a 100, se il tasso è espresso in percentuale) del rapporto tra il numero di unità che hanno fornito dati per una specifica variabile e il numero di unità che avrebbero dovuto rispondere a tale variabile.

I tassi di mancata risposta sono molto importanti poiché rappresentano un campanello di allarme sulla potenziale distorsione, che ha luogo quando i rispondenti differiscono dai non rispondenti rispetto alle variabili di interesse dell'indagine. Inoltre, se il tasso di risposta è molto basso, anche piccole differenze possono generare una distorsione significativa. Per questo motivo, i tassi di mancata risposta totale andrebbero sempre riportati nel *quality report*. Qualora siano rilevanti, è importante includere anche i tassi di mancata risposta parziale per le variabili principali.

- Errore di trattamento

Tra la raccolta dei dati e l'analisi statistica, i microdati sono sottoposti a una serie di operazioni di elaborazione, che includono la registrazione dei dati, l'integrazione, la codifica, il controllo, la correzione, l'imputazione, la derivazione di unità o variabili, l'aggregazione e la ponderazione. Gli errori introdotti in queste fasi sono collettivamente noti come errori di trattamento. Sia gli errori di misurazione sia gli errori di trattamento riguardano i microdati, e le tecniche di valutazione per entrambi sono molto simili.

Un tipo di errore di trattamento che merita particolare attenzione è quello legato alla codifica delle risposte fornite in formato di testo libero. Queste risposte, che generalmente si riferiscono a variabili testuali con classificazioni ufficiali (come attività economica, professione o titolo di studio), devono essere codificate assegnando a ciascun testo un codice univoco secondo uno schema classificatorio predefinito. Gli errori di trattamento influenzano le singole osservazioni e, come gli errori di misurazione, possono causare distorsione e aumentare la variabilità delle statistiche risultanti. L'importanza degli errori di trattamento può variare significativamente tra i diversi processi statistici, e la loro trattazione all'interno di un *quality report* deve riflettere l'importanza che rivestono nel contesto specifico.

- Errore di assunzione del modello

I modelli sono ampiamente utilizzati in statistica. In alcuni casi, l'obiettivo della stima si basa su un modello definito in ambito tematico; In altri, come ad esempio per la destagionalizzazione, il modello è di natura puramente matematico-statistica. Talvolta, un modello viene impiegato per migliorare la precisione della stima o per correggere la mancata risposta. Quando i modelli rivestono un ruolo significativo nell'interpretazione dei risultati, è fondamentale che vengano descritti in dettaglio e che si valuti l'impatto che possono avere sull'accuratezza degli output.

2.9.2 Indagini con campionamento non probabilistico

Le indagini con campionamento non probabilistico sono anch'esse influenzate dall'errore di campionamento, sebbene non vi sia una teoria consolidata che possa essere utilizzata per la stima dell'errore. La questione fondamentale riguarda la verifica della rappresentatività del campione ottenuto. Con un campione probabilistico, la rappresentatività è garantita dal disegno di campionamento (sebbene soggetto anch'esso a distorsione da errori di copertura e di mancata risposta), mentre in un campione non probabilistico è necessario ricorrere a un altro meccanismo che possa assicurare questa rappresentatività. Un esempio di tale meccanismo può essere un modello statistico appropriato al campo di studio. Ad esempio, nelle indagini per quote, la rappresentatività dipende in modo critico dall'efficacia delle variabili utilizzate per la costruzione delle quote, che devono catturare le variazioni nella popolazione obiettivo. Il meccanismo di selezione nelle indagini per quote (che dipende sia dalla selezione effettuata dall'intervistatore sia dall'accettazione del rispondente a partecipare) non deve essere correlato con le variabili target dell'indagine, affinché non ci sia distorsione. Per quanto riguarda gli errori non campionari, nelle indagini con campionamento non probabilistico in genere la mancata risposta non è un concetto significativo, poiché l'insieme degli intervistati non è predeterminato dal disegno del campionamento. Gli errori di misurazione e trattamento sono della stessa natura delle indagini probabilistiche.

2.9.3 Indagini censuarie

L'indagine censuaria raccoglie dati per tutte le unità della popolazione di riferimento e, per definizione, non presenta errori di campionamento. Tuttavia, gli altri tipi di errore sono presenti anche in questo caso, come nelle indagini con campione probabilistico. Il profilo di errore di un'indagine censuaria, però, può differire notevolmente da quello di un'indagine campionaria e variare in base al tipo di censimento e all'approccio adottato. Questo influenza sull'enfasi da attribuire nel *quality report* nel trattare le diverse fonti di errore.

Poiché uno degli obiettivi principali dei censimenti è enumerare con precisione tutte e sole le unità della popolazione, un errore particolarmente rilevante è il cosiddetto sovra o sottoconteggio, che corrisponde all'errore di sovra o sottocopertura. Per valutarlo, si conduce a posteriori un'indagine di copertura, un'analisi campionaria finalizzata a verificare quante unità siano state erroneamente escluse o, al contrario, incluse nel conteggio. Nel *quality report* deve essere fornita una valutazione dell'errore riscontrato.

Gli errori di misurazione e di mancata risposta potrebbero essere rilevanti, così come gli errori di trattamento sotto forma di errori di registrazione dati o di codifica possono essere di grande importanza. Anche tutti questi aspetti dovrebbero essere discussi nel *quality report*.

2.9.4 Statistiche basate su dati amministrativi

La classificazione degli errori riportata precedentemente, come già sottolineato, è particolarmente adatta alle indagini dirette, ma meno ad altre tipologie di processo. Tuttavia, è possibile adattarla anche a quelle basate su dati amministrativi, ad esempio sostituendo il concetto di "mancata risposta" con quello di "valori mancanti" nei dati amministrativi raccolti. Alcuni tipi di errori possono essere utilizzati, ma è necessario adattarne la definizione,

mantenendo comunque la chiarezza. Nelle statistiche basate su dati amministrativi, l'errore di campionamento di norma non si applica. Sono frequenti discrepanze rispetto alla popolazione obiettivo, che possono essere interpretate come errori di copertura (sovra o sottocopertura). Per alcune variabili, inoltre, potrebbero esserci dati mancanti, trattabili come la mancata risposta parziale. Possono verificarsi anche errori di misurazione e di trattamento, che sono della stessa natura delle indagini probabilistiche.

2.9.5 Processi multifonte

Quando l'output statistico deriva da un processo che utilizza più fonti di dati, è fondamentale che la valutazione complessiva consideri tutte le possibili componenti di errore, nonché eventuali problemi di coerenza nella combinazione dei dati provenienti dalle diverse fonti. In particolare, gli errori di integrazione delle fonti, che rientrano tra gli errori di trattamento, sono di grande rilevanza.

Nella maggior parte dei casi, una panoramica sull'accuratezza complessiva rappresenta una parte essenziale del *quality report*. Tuttavia, valutare l'accuratezza è complesso a causa dei numerosi errori che possono emergere dalle diverse fonti di dati ed è dunque importante identificarne i più significativi e fornirne una spiegazione dettagliata. Quando possibile, è utile descrivere le principali fonti di errore in termini quantitativi, per fornire una visione più precisa delle loro implicazioni.

Esempi

L'indagine sul dolore cronico in Italia a partire dai dati dell'EHIS 2019 (ISS-00054)

ISS - Istituto superiore di sanità

Il tasso di risposta, al netto delle famiglie non eleggibili (927 famiglie; quelle, ad esempio, che si sono trasferite in un'altra città) è stato del 78 per cento.

Livello di precisione delle stime e presentazione sintetica degli errori campionari

Le principali statistiche di interesse per valutare la variabilità campionaria delle stime prodotte da un'indagine sono l'errore di campionamento assoluto e l'errore di campionamento relativo.

A ogni stima corrisponde un errore di campionamento relativo; ciò significa che per consentire una lettura corretta delle tabelle pubblicate sarebbe necessario presentare per ogni stima pubblicata il corrispondente errore di campionamento relativo. Ciò, tuttavia, non è possibile sia per limiti di tempo e di costi di elaborazione, sia perché le tavole della pubblicazione risulterebbero appesantite e di non facile consultazione per l'utente finale. Inoltre, non sarebbero comunque disponibili gli errori delle stime non pubblicate, che l'utente può ricavare in modo autonomo. Per le motivazioni sopra esposte, si ricorre frequentemente a una presentazione sintetica degli errori relativi, basata sul metodo dei modelli regressivi. Tale metodo si basa sulla determinazione di una funzione matematica che mette in relazione ciascuna stima con il proprio errore relativo. Nella presente indagine, il modello utilizzato per le stime di frequenze assolute e relative è del tipo seguente:

$$\log(\hat{\varepsilon}^2(\hat{Y}_d)) = a + b \log(\hat{Y}_d) \quad (1)$$

dove i parametri a e b vengono stimati utilizzando il metodo dei minimi quadrati. Nella Tavola 2.1 sono riportati i valori dei coefficienti a e b e dell'indice di determinazione R^2 del modello utilizzato per l'interpolazione degli errori campionari di stime di frequenze assolute e relative, per totale Italia, ripartizione geografica e Regione. Sulla base delle informazioni

2. Guida alla compilazione

contenute in tale prospetto, è possibile calcolare la stima dell'errore di campionamento relativo di una determinata stima di frequenza assoluta \widehat{Y}_d mediante la formula:

$$\hat{\varepsilon}(\widehat{Y}_d) = \sqrt{\exp(a + b \log(\widehat{Y}_d))} \quad (2)$$

che si ricava facilmente dalla (1).

Se, per esempio, la stima \widehat{Y}_d si riferisce agli individui dell'Italia Nord-ovest, l'errore relativo corrispondente si ottiene introducendo nella (2) i valori dei parametri a e b riportati nella prima riga della Tavola 2.1. Le Tavole 2.2 e 2.3, presentate in aggiunta con riferimento agli individui, consentono di rendere più agevole il calcolo degli errori campionari. Esse contengono gli errori di campionamento relativo, per ciascun dominio territoriale di interesse, calcolati mediante la formula (2), corrispondenti alle stime di frequenze assolute. Le informazioni contenute permettono di calcolare l'errore relativo di una generica stima di frequenza assoluta (o relativa) mediante due procedimenti che risultano di facile applicazione, anche se conducono a risultati meno precisi di quelli ottenibili mediante l'espressione (2). Il primo metodo consiste nell'individuare il livello di stima (riportato in colonna) che più si avvicina alla stima di interesse e nel considerare come errore relativo il valore che si trova sulla riga corrispondente al dominio territoriale di riferimento. A partire dagli errori assoluti e relativi è possibile costruire gli intervalli di confidenza che, con un prefissato livello di fiducia, contengono al loro interno il valore "vero", ma ignoto, del parametro oggetto di stima.

Tavola 2.1 - Valori dei coefficienti a , b e dell'indice di determinazione R^2 delle funzioni utilizzate per le interpolazioni degli errori campionari delle stime a livello nazionale per ripartizione geografica e regione. Anno 2019

RIPARTIZIONI GEOGRAFICHE E REGIONI	a	b	R^2 (%)
Italia	8,268	-1,072	98,1
Nord-ovest	8,290	-1,076	97,9
Nord-est	8,492	-1,109	97,3
Centro	7,995	-1,061	97,3
Sud	7,275	-1,006	97,4
Isole	7,466	-1,024	96,5
Valle d'Aosta/Vallée d'Aoste	5,22	-1,09	95,5
Lombardia	8,51	-1,08	97,8
Bolzano/Bozen	5,98	-1,01	95,7
Trento	6,01	-1,02	95,1
Veneto	8,46	-1,11	97,7
Friuli-Venezia Giulia	6,96	-1,05	96,7
Liguria	7,06	-1,05	96,8
Emilia-Romagna	8,36	-1,09	96,6
Toscana	7,51	-1,04	97,4
Umbria	7,03	-1,06	95,2
Marche	6,65	-1,00	96,0
Lazio	8,20	-1,06	96,9
Abruzzo	7,30	-1,07	96,1
Molise	5,42	-0,98	93,8
Campania	7,22	-0,98	96,6
Puglia	7,40	-1,02	97,2
Basilicata	6,53	-1,06	96,3
Calabria	6,72	-0,99	96,7
Sicilia	7,72	-1,03	96,5
Sardegna	6,88	-1,02	97,1

Fonte: Eurostat, European Health Interview Survey (EHIS)

Tavola 2.2 - Valori interpolati degli errori campionari delle stime per ripartizione geografica. Anno 2019 (frequenze assolute)

RIPARTIZIONI GEOGRAFICHE	VALORI DELLE STIME										
	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000	2.500.000	5.000.000	
Italia	27,4	18,9	15,2	13,0	8,0	5,5	4,4	3,8	2,3	1,6	
Nord-ovest	27,2	18,7	15,1	12,9	7,9	5,4	4,4	3,7	2,3	1,6	
Nord-est	25,5	17,3	13,8	11,8	7,1	4,8	3,9	3,3	2,0	1,4	
Centro	25,3	17,5	14,1	12,1	7,5	5,2	4,2	3,6	2,2	1,5	
Sud	23,3	16,4	13,4	11,6	7,3	5,2	4,2	3,6	2,3	1,6	
Isole	23,5	16,4	13,4	11,5	7,2	5,1	4,1	3,6	2,2	1,6	

Fonte: Eurostat, European Health Interview Survey (EHIS)

Tavola 2.3 - Valori interpolati degli errori campionari delle stime per regione. Anno 2019 (frequenze assolute)

REGIONI	VALORI DELLE STIME										
	10.000	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000	2.500.000	
Piemonte	37,1	22,6	15,5	12,4	10,6	6,5	4,4	3,6	3,1	1,9	
Valle d'Aosta/Vallée d'Aoste	8,8	5,3	3,7	2,9	2,5	1,5	1,0	0,8	0,7	0,4	
Lombardia	48,7	29,7	20,4	16,4	14,0	8,6	5,9	4,7	4,0	2,5	
Bolzano/Bozen	19,3	12,2	8,6	7,0	6,1	3,8	2,7	2,2	1,9	1,2	
Trento	18,3	11,4	8,0	6,5	5,6	3,5	2,5	2,0	1,7	1,1	
Veneto	41,9	25,3	17,2	13,7	11,7	7,1	4,8	3,8	3,3	2,0	
Friuli-Venezia Giulia	25,4	15,7	10,9	8,8	7,6	4,7	3,2	2,6	2,3	1,4	
Liguria	27,2	16,8	11,7	9,5	8,1	5,0	3,5	2,8	2,4	1,5	
Emilia-Romagna	42,7	25,9	17,7	14,2	12,1	7,4	5,0	4,0	3,5	2,1	
Toscana	34,9	21,6	15,1	12,2	10,5	6,5	4,5	3,7	3,2	2,0	
Umbria	25,5	15,7	10,9	8,8	7,5	4,6	3,2	2,6	2,2	1,4	
Marche	27,3	17,3	12,2	9,9	8,6	5,4	3,8	3,1	2,7	1,7	
Lazio	46,1	28,4	19,7	15,9	13,6	8,4	5,8	4,7	4,0	2,5	
Abruzzo	27,9	17,1	11,8	9,5	8,1	5,0	3,4	2,8	2,4	1,5	
Molise	16,7	10,7	7,6	6,2	5,4	3,5	2,5	2,0	1,8	1,1	
Campania	41,2	26,3	18,8	15,4	13,4	8,6	6,1	5,0	4,3	2,8	
Puglia	37,7	23,7	16,7	13,6	11,7	7,4	5,2	4,2	3,7	2,3	
Basilicata	20,0	12,3	8,6	6,9	5,9	3,7	2,5	2,0	1,8	1,1	
Calabria	30,3	19,2	13,7	11,2	9,7	6,2	4,4	3,6	3,1	2,0	
Sicilia	40,9	25,5	17,8	14,5	12,5	7,8	5,4	4,4	3,8	2,4	
Sardegna	28,2	17,7	12,4	10,1	8,7	5,5	3,8	3,1	2,7	1,7	

Fonte: Eurostat, European Health Interview Survey (EHIS)

Disabilità, DSA e accesso alla Formazione universitaria (ANV-00002)

ANVUR - Agenzia nazionale di valutazione del sistema universitario e della ricerca

Il questionario è stato preventivamente testato, in presenza, nei tre principali atenei romani, cui si deve un ringraziamento particolare per la disponibilità e per il contributo che hanno apportato alla definizione dello strumento di indagine. La rilevazione si è conclusa nell'ottobre del 2020 con i risultati riportati nella seguente Tavola 2.4.

Tavola 2.4 - I dati sulla rilevazione dell'ANVUR. Anno 2020 (valori assoluti)

PARTECIPAZIONE ATENELI	Statali	Statali-Scuole	Non statali	Telematici	Totale
Accreditati in piattaforma	59	6	9	7	91
Non accreditati	0	0	1	4	5
Totale atenei	61	6	20	11	98
<i>Hanno risposto</i>	59	6	19	6	90
<i>Non hanno risposto</i>	2	0	1	5	8

Fonte: Agenzia nazionale di valutazione del sistema universitario e della ricerca (ANVUR)

2. Guida alla compilazione

Sistema informativo sulle statistiche culturali (MBE-00012)

Ministero della Cultura

Al termine del periodo di raccolta dei dati tramite questionario Istat, delle 5.288 unità presenti nella lista iniziale, sono risultati non raggiungibili, cioè “inevase”, 475 musei e istituzioni simili, pari al nove per cento della lista iniziale. Considerando il totale di unità “risolte”, cioè le unità rispondenti più le unità non eleggibili, si è raggiunto un tasso di risposta complessivo pari al 91,5 per cento. L’analisi delle mancate risposte parziali evidenzia una percentuale media di circa il sette per cento, che si annulla per le domande chiave (ad esempio: numero visitatori paganti e non paganti, tipologia di apertura, supporti alla visita, servizi erogati, eccetera) a cui il responsabile del museo doveva obbligatoriamente rispondere per potere proseguire nel questionario online.

Audimob - Indagine su stili e comportamenti di mobilità dei residenti in Italia (IFT-00001)

ISFORT - Istituto superiore di formazione e ricerca per i trasporti

La base dati dell’Osservatorio è alimentata da un numero variabile di interviste/anno con un errore campionario sul dato nazionale, comunque, sempre inferiore all’uno per cento.

Relazione al Parlamento sul Patrocinio a spese dello Stato nel processo penale ai sensi dell’art. 294 del DPR n.115/02 (MGG-00048)

Ministero della Giustizia

A causa della persistente difficoltà di acquisizione dei dati presso alcuni uffici giudiziari, problema che di fatto comporta l’assegnazione di un valore pari a zero a tutte le variabili dell’ufficio inadempiente, si è ritenuto opportuno effettuare, per gli anni 2005-2020, una stima dei dati mancanti, procedura che ha consentito di mitigare in modo apprezzabile tale carenza.

Sistema di sorveglianza sentinella dell’influenza - INFLUNET (ISS-00014)

ISS - Istituto superiore di sanità

Per ottenere una stima solida dei casi di sindrome simil influenzale in Italia, occorre mantenere la rete di medici di medicina generale e di pediatri di libera scelta sul territorio regionale ed è necessario arruolare un campione di medici sentinella sufficiente a garantire la sorveglianza di almeno il quattro per cento della popolazione nazionale.

Sorveglianza delle malattie batteriche invasive (ISS-00028)

ISS - Istituto superiore di sanità

Periodicamente, per alcune Regioni, vengono integrati nella piattaforma MaBI (Malattie Batteriche Invasive) i dati forniti da piattaforme regionali e viene attivato un processo di consolidamento del dato mediante controlli sulla qualità del dato con il supporto delle Regioni e della Pubblica amministrazione.

La qualità del lavoro in Italia (IAP-00009)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

La procedura allocativa ha previsto un CV massimo del 20 per cento per stime del 10 per cento a livello regionale. Nella Tavola 2.5 si riporta la distribuzione del campione di occupati per regione e il valore effettivo del CV.

Tavola 2.5 - Distribuzione del campione di occupati per regione e coefficiente di variazione (CV) effettivo. Anno 2015

REGIONI	N (Occupati)	n	Varianza	CV (p=0,1)
Piemonte	1.772.755	975	0,0001	0,0961
Valle d'Aosta/Vallée d'Aoste	55.102	240	0,0004	0,1932
Lombardia	4.236.657	2.250	0,0000	0,0632
Trentino-Alto Adige/Südtirol	474.441	365	0,0002	0,1570
Veneto	2.064.390	1.455	0,0001	0,0786
Friuli-Venezia Giulia	494.521	370	0,0002	0,1559
Liguria	599.110	355	0,0003	0,1592
Emilia-Romagna	1.911.189	1.460	0,0001	0,0785
Toscana	1.534.186	1.180	0,0001	0,0873
Umbria	348.695	285	0,0003	0,1776
Marche	625.045	470	0,0002	0,1383
Lazio	2.302.073	1.685	0,0001	0,0731
Abruzzo	475.927	335	0,0003	0,1638
Molise	100.644	250	0,0004	0,1895
Campania	1.559.142	925	0,0001	0,0986
Puglia	1.143.042	670	0,0001	0,1159
Basilicata	182.418	255	0,0004	0,1877
Calabria	522.584	335	0,0003	0,1639
Sicilia	1.321.288	785	0,0001	0,1070
Sardegna	548.095	355	0,0003	0,1592
Italia	22.271.302	15.000	0,0000	0,0245

Fonte: Istituto nazionale per le analisi delle politiche pubbliche (INAPP)

Oltre a distribuire il campione totale nelle regioni italiane, controllando la precisione delle stime prodotte su ciascuna regione, è stata predisposta un'allocazione target che prendesse in considerazione ulteriori caratteristiche degli individui da intervistare, in modo tale da tenere sotto controllo la precisione delle stime per sesso, classi di età e titolo di studio degli occupati. Tale allocazione target è stata tenuta in considerazione durante la fase di campo e ha garantito che per ciascuno strato (definito dalla concatenazione di regione, sesso, classe di età e titolo di studio delle unità campionate) fosse realizzato un numero minimo di interviste atto a garantire il livello di precisione delle stime fissato a priori.

Indagine sull'offerta pubblica e privata di servizi sociali (IAP-00025)

INAPP - Istituto nazionale per le analisi delle politiche pubbliche

Il campione teorico di istituzioni cui sottoporre il questionario di indagine ha subito, durante la fase di campo, delle cadute. Tali cadute sono avvenute sia in ragione della obsolescenza delle informazioni contenute nel *sampling frame* (che riportava informazioni relative al 2011), sia per rifiuto esplicito all'intervista. Inoltre, a esito della compilazione del questionario alcune interviste sono state scartate a causa dell'incompletezza delle informazioni rilasciate.

2.10 Attendibilità - Revisione dei dati

Codice	Nome del concetto	Definizione	Linee guida
10	Attendibilità - Revisione dei dati	L'attendibilità del dato è definita come la vicinanza tra il valore iniziale stimato di una quantità e il valore della stessa quantità stimato successivamente. La revisione è definita come la differenza tra una stima rilasciata precedentemente e una successiva.	Riportare l'eventuale politica di revisione applicata nel processo, cioè se vengono rilasciate stime provvisorie, revisionate e definitive. Se vengono diffusi solo risultati definitivi, si riporti che questo concetto non è applicabile. Nel caso ci siano diversi rilasci, riportare il calendario e i motivi di revisione.

2. Guida alla compilazione

Qualora l'ente disponga di una politica generale sulla revisione dei dati che copre tutti o gran parte dei processi statistici, essa deve essere riportata o, in alternativa, è possibile fornire un link al documento dove essa può essere consultata.

In ogni caso, le revisioni dei dati devono seguire procedure standard, consolidate e trasparenti. Gli eventuali diversi rilasci di una statistica devono essere annunciati preventivamente, indicando le cause di revisione. Ad esempio, potrebbe esserci un primo rilascio basato su un primo insieme di rispondenti per soddisfare vincoli di tempestività, seguito da un secondo rilascio revisionato che utilizza i dati di tutti i rispondenti e, infine, un ultimo rilascio definitivo che utilizza una fonte amministrativa disponibile solo successivamente.

Quando si implementa una politica di revisione dei rilasci, si presume generalmente che i rilasci successivi siano più accurati (cioè, vicini al valore vero) rispetto ai precedenti, perché basati su informazioni aggiuntive o controlli più approfonditi. Per questo motivo è fondamentale analizzare le revisioni, verificando in particolare se le stime precedenti presentano una significativa variabilità rispetto a quelle successive o tendono a essere sistematicamente inferiori o superiori, segnalando possibili problemi di distorsione. Se è stata effettuata un'analisi delle revisioni, ad esempio attraverso indicatori di qualità come Revisione media assoluta o Revisione media assoluta relativa⁴, occorre riportare i risultati.

Esempi

Sorveglianza delle malattie batteriche invasive (ISS-00028)

ISS - Istituto superiore di sanità

Per questo rapporto, l'analisi dei dati del 2022 è stata eseguita nel secondo semestre del 2023; si fa notare che i dati potrebbero subire delle variazioni e/o essere ulteriormente integrati in rapporti successivi. Nel presente rapporto sono stati aggiornati anche i dati degli anni precedenti, includendo le segnalazioni pervenute successivamente ai rapporti già pubblicati ed eventuali correzioni.

Rilevazione delle tipologie e caratteristiche dei clienti negli esercizi ricettivi (EMR-00028)

Regione Emilia-Romagna

È prevista una cadenza mensile per la verifica e il rilascio dei dati in forma provvisoria; tali attività si concludono solitamente entro 30 giorni dal termine del mese di riferimento. Si procede poi a una complessiva revisione annuale, svolta di solito nei mesi di febbraio e marzo dell'anno successivo a quello di rilevazione. I dati provvisori mensili vengono sistematicamente comunicati all'Istat e pubblicati sul portale di statistica della regione Emilia-Romagna; il dato annuale viene inoltrato all'Istat per una validazione finale: in attesa di quest'ultima verifica, il dato viene pubblicato sul portale regionale di statistica come "consolidato in attesa di validazione". Mensilmente si possono verificare casi di mancata risposta che, nel corso dell'anno, vengono solitamente recuperati a seguito di contatti con la struttura. Per alcune strutture non rispondenti è possibile comunque procedere alla stima di arrivi e di presenze basandosi sull'andamento del movimento che le stesse hanno comunicato nei periodi precedenti: il numero delle strutture stimate, solitamente decresce via via che ci si avvicina alla chiusura annuale della rilevazione e che i ripetuti contatti con i gestori degli esercizi ricettivi producono il proprio effetto.

4 Per maggiori informazioni cfr. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Short-term_business_statistics_-_revisions

Agenzia delle dogane: Organizzazione, attività e statistica (MST-00007)

Agenzia delle dogane e dei monopoli

Alcuni dati del bollettino statistico, già pubblicati in precedenti edizioni della reportistica (ad esempio il *Libro blu*), potrebbero aver subito aggiornamenti a seguito di rettifiche e integrazioni. Tali aggiustamenti, infatti, in base alle disposizioni vigenti, possono essere apportati, per taluni dati, anche a distanza di tempo (ad esempio, per la rettifica dati di dichiarazioni doganali, aggiornamento a sistema delle schede di sequestro). Ovviamente, al decorrere del tempo si riduce la probabilità che vengano operate tali rettifiche. Conseguentemente, trascorso un certo intervallo temporale, i dati vengono consolidati e sono successivamente destinati a non mutare ulteriormente.

2.11 Tempestività e puntualità

Codice	Nome del concetto	Definizione	Linee guida
11	Tempestività	Periodo di tempo che intercorre tra la disponibilità dei dati e l'evento o il fenomeno descritto dai dati.	Indicare la durata temporale (in numero di giorni o settimane o mesi) intercorsa dal periodo di riferimento dei dati al momento in cui essi sono pubblicati.
	Puntualità	Intervallo di tempo tra l'effettiva consegna/rilascio dei dati e la data obiettivo prevista.	Indicare la durata temporale (in numero di giorni o settimane o mesi) intercorsa tra l'effettiva consegna/rilascio dei dati e la data obiettivo prevista.

La misura più comune di tempestività è il tempo di produzione, definito come l'intervallo di tempo che intercorre tra la fine del periodo (o punto nel tempo) di riferimento a cui si riferiscono i dati e il giorno della pubblicazione delle statistiche, qualsiasi sia il formato e il mezzo. L'unità di tempo appropriata (giorno, settimana, mese) dipende dal periodo di riferimento ed è a discrezione di chi compila il *report*. È opportuno spiegare le ragioni della durata del tempo di produzione e gli sforzi fatti per ridurlo, soprattutto quando esso potrebbe essere considerato dagli utenti molto lungo. Laddove uno standard di tempestività sia specificato in un regolamento, dovrebbe essere utilizzato come punto di riferimento e dovrebbe essere indicato il rapporto o la differenza tra il tempo di produzione effettivo e il tempo di produzione specificato nel regolamento.

In merito alla puntualità, bisogna distinguere se è previsto o meno un piano di rilascio dei dati. Nel caso in cui l'ente non abbia messo a disposizione degli utenti il piano (e qualora tale piano non fosse previsto da alcun regolamento), il *report* dovrà indicare che la puntualità non può essere misurata. Sarà necessario fornire, inoltre, le motivazioni per l'assenza del piano di rilascio e delineare le azioni previste per renderlo disponibile in futuro.

Nel caso, invece, esista un piano di rilascio dei dati, la puntualità è misurata dal tempo intercorso tra la data di consegna/rilascio effettiva dei dati e la data prevista dal piano. In caso di rilasci non puntuali, spiegare le ragioni e delineare le azioni atte a ridurre lo sfasamento temporale in futuro.

Si possono riportare in forma discorsiva i valori degli indicatori di tempestività e puntualità.

Per i processi che si ripetono annualmente o più frequentemente, i valori possono essere calcolati in media su un dato numero di cicli di processo o su un periodo di tempo, ad esempio tre anni. Il numero di cicli o il periodo di tempo considerato è a discrezione di chi scrive il *report*.

Alcune statistiche vengono rilasciate in due o più versioni, ad esempio versioni preliminari e finali, oppure versioni preliminari, revisionate e finali. In questo caso ogni versione

2. Guida alla compilazione

ha il proprio tempo di produzione. Questi rilasci dovrebbero essere distinti e la tempestività e la puntualità di ciascuno presentate separatamente.

Esempi

Cooperazione allo sviluppo - Impegni ed erogazioni per Paese destinatario (MAE-00023)

Ministero degli Affari esteri e della Cooperazione internazionale

L'intervallo di tempo tra la raccolta e diffusione dei dati definitivi è di 180 giorni.

Disoccupazione nei settori non agricoli (IPS-00006)

INPS - Istituto nazionale di previdenza sociale

Si distingue la tempestività della diffusione tra i dati provvisori e i dati definitivi. L'intervallo di tempo tra la raccolta e diffusione dei dati provvisori è di 60 giorni, dei dati definitivi 90 giorni.

2.12 Comparabilità nel tempo

Codice	Nome del concetto	Definizione	Linee guida
12	Comparabilità nel tempo	Grado di confrontabilità nel tempo dell'informazione statistica.	<p>Fornire informazioni sulle possibili limitazioni nell'uso dei dati per i confronti temporali.</p> <p>Distinguere tre situazioni:</p> <ol style="list-style-type: none">1. non ci sono stati cambiamenti;2. ci sono stati alcuni cambiamenti, ma non abbastanza da determinare un'interruzione nella serie;3. ci sono stati cambiamenti sufficienti da determinare un'interruzione nella serie.

La comparabilità nel tempo è un aspetto qualitativo cruciale per tutti i dati statistici pubblicati nelle diverse edizioni temporali del medesimo processo statistico. Per molti utenti, i cambiamenti nel tempo dei fenomeni economici o sociali sono gli aspetti più interessanti delle statistiche, e la comparabilità nel tempo è essenziale se si vuole che i dati riflettano i reali cambiamenti economici o sociali che si sono verificati. Indipendentemente dal fatto che le statistiche siano pubblicate direttamente sotto forma di serie storiche o che gli utenti debbano costruire autonomamente le proprie serie storiche partendo da dati di base, è necessario informare gli utenti sulle possibili limitazioni nell'uso dei dati per i confronti temporali. Ad esempio, un resoconto dettagliato delle differenze nelle fonti e nei metodi utilizzati nei diversi periodi di tempo è di grande beneficio se si vuole combinare dati provenienti da periodi di tempo diversi.

L'indicatore che è possibile documentare nel *report* è la "Lunghezza della serie storica comparabile", cioè il numero di periodi di riferimento in serie storica dall'ultima rottura della serie.

Dove ci sono state modifiche lievi, tali cioè da non provocare l'interruzione della serie, il *report* dovrebbe comunque registrare tali cambiamenti nei metadati che descrivono il processo. Laddove si siano verificati cambiamenti sufficienti a determinare un'interruzione nella serie, gli utenti devono essere debitamente informati dell'interruzione e ricevere le informazioni necessarie per gestirne le conseguenze.

La quantità delle informazioni fornite può variare da molto completa a minimale. Il trattamento più completo consiste nel portare avanti entrambe le serie per un periodo di tempo e/o nel ricostruire le serie a ritroso, ovvero convertire le vecchie serie in quello che sarebbe stato con il nuovo approccio, duplicando la misurazione in un periodo di tempo utilizzando le definizioni/metodi originali e riveduti. Un'altra possibilità è quella di fornire agli utenti

fattori di aggiustamento in modo che possano procedere in autonomia nella lettura dei dati. Il trattamento meno oneroso consiste nel descrivere semplicemente i cambiamenti che si sono verificati e fornire solo valutazioni qualitative del loro probabile impatto sulle stime. Ovviamente, questo è il meno soddisfacente per l'utente.

Esempi

Sistema di sorveglianza delle nuove diagnosi di infezione da HIV (ISS-00056)

ISS - Istituto superiore di sanità

La serie storica delle nuove diagnosi HIV, riportata nei fascicoli del *Notiziario* dal 2017, risulta leggermente inferiore alla serie storica riportata nei volumi precedenti perché a partire dal 2017 si eliminano le doppie segnalazioni riscontrate in regioni diverse e relativamente alla serie storica disponibile per ciascuna regione, ossia per tutte le diagnosi segnalate al Centro operativo AIDS dall'anno di inizio della raccolta dati in ogni regione fino al 31 dicembre dell'anno interessato.

Relazione al Parlamento sul Patrocinio a spese dello Stato nel processo penale (ai sensi dell'art. 294 del DPR n.115/02) (MGG-00048)

Ministero della Giustizia

Si precisa infine che eventuali discordanze tra i dati della presente relazione e la precedente sono dovute ad alcune revisioni e correzioni cui vengono solitamente sottoposte le banche dati, oltre alla circostanza che alcuni uffici non riescono a rispondere in tempo utile per la stesura delle relazioni, ma solo in seguito. Pertanto, poiché le stime sono state effettuate solo per gli anni 2005-2020, i dati degli anni 1995-2004 non risultano pienamente confrontabili con tale periodo successivo.

Intervalli di quotazioni immobiliari in Euro/mq (FIN-00010)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

Le quotazioni immobiliari semestrali individuano, per ogni delimitata zona territoriale omogenea, ovvero una zona OMI (Osservatorio del Mercato Immobiliare) di ciascun comune, un intervallo minimo/massimo, per unità di superficie in euro al metro quadrato, dei valori di mercato e locazione, per tipologia immobiliare e stato di conservazione.

Le quotazioni OMI, disponibili in un semestre, sono relative ai comuni censiti negli archivi catastali. È quindi, possibile che l'elenco dei comuni presenti in banca dati differisca nei diversi semestri per effetto di variazioni circoscrizionali (in esito alle quali sono costituiti nuovi comuni e soppressi altri).

2.13 Accessibilità e chiarezza

Accessibilità e chiarezza, attributi fondamentali dell'output statistico, misurano la facilità con cui gli utenti possono “acquisire e comprendere” le informazioni statistiche prodotte. In considerazione del ruolo chiave svolto ai fini della concreta attuazione dei principi enunciati dal Codice europeo e dal Codice italiano delle statistiche ufficiali, la politica di diffusione della produzione statistica del Sistan deve seguire precise regole, rimuovendo ogni barriera all'utilizzo dei dati. Specificamente, nei citati codici, le nozioni di accessibilità e chiarezza sono combinate in un unico principio, secondo cui le statistiche europee devono essere presentate in una forma chiara e comprensibile, diffuse in modo adeguato e agevole, disponibili e accessibili su base imparziale, con metadati e linee guida di supporto.

2. Guida alla compilazione

In particolare:

- l'accessibilità descrive l'insieme delle condizioni e delle modalità con cui gli utenti possono ottenere i dati e i relativi metadati; è documentata descrivendo ciascun mezzo di diffusione e valutando la loro efficacia nel facilitare l'accesso ai dati;
- la chiarezza si riferisce alla misura in cui i dati sono accompagnati da metadati comprensibili, incluse informazioni riguardanti la qualità dei dati stessi, nonché alla disponibilità di assistenza supplementare fornita agli utenti per facilitare la comprensione dei dati; è documentata attraverso i metadati che accompagnano i dati, la pertinenza di tali metadati e la loro comprensibilità.

Pertanto, le statistiche ufficiali e i relativi metadati devono essere presentati e documentati in una forma che faciliti la corretta interpretazione, ovvero i servizi di diffusione devono utilizzare metodi, piattaforme e tecnologie dell'informazione e della comunicazione moderni, nonché gli standard in materia di *open data*. I formati di diffusione hanno infatti un impatto significativo sull'accessibilità e sulla chiarezza. In questo contesto, "formato" non si riferisce ai tipi di file elettronici (il documento *Word* o il foglio *Excel*, per esempio), bensì ai mezzi e meccanismi attraverso i quali i risultati statistici e i relativi metadati vengono diffusi agli utenti (ad esempio, tramite comunicato stampa, pubblicazione stampata, pubblicazione elettronica e banca dati online).

L'accessibilità e la chiarezza includono la capacità di soddisfare le esigenze delle varie classi di utenti, così come sono classificati nella sezione "Pertinenza".

Il riscontro degli utenti è essenziale per valutare l'accessibilità e la chiarezza. Le domande relative all'esperienza degli utenti riguardo alla facilità di accesso ai dati e alla loro interpretazione dovrebbero essere incluse nella progettazione delle indagini sulla soddisfazione degli utenti.

I risultati di tali indagini, insieme a qualsiasi altro riscontro ottenuto dagli utenti, devono essere documentati nel *report*. Ulteriori indicatori utili includono: il conteggio e le variazioni del numero di abbonati alle pubblicazioni (sia cartacee sia elettroniche), le vendite di pubblicazioni e il numero di consultazioni degli utenti sui siti web.

2.13.1 Metodi e formati di diffusione

Codice	Nome del concetto	Definizione	Linee guida
13.1	Calendario di diffusione	Il calendario delle date di pubblicazione delle statistiche e le modalità di diffusione.	Indicare se esiste un calendario di rilascio dei risultati del processo statistico e se questo calendario è accessibile al pubblico. Se possibile, fornire un link o un riferimento.
13.2	Comunicati stampa	Comunicati stampa periodici o ad hoc collegati ai dati.	Elencare i comunicati stampa regolari o ad hoc collegati ai dati nell'ultimo anno.
13.3	Pubblicazioni	Pubblicazioni periodiche o ad hoc in cui i dati sono resi disponibili al pubblico.	Elencare i titoli di eventuali pubblicazioni, incluso l'editore, l'anno e i collegamenti a documenti online (se disponibili). Fornire il numero di abbonamenti/acquisti per ciascuno dei principali <i>report</i> .
13.4	Portale online	Informazioni sulle banche dati online nelle quali è possibile accedere ai dati diffusi.	Fornire il nome di dominio, il link al database online (se presente) e il numero di accessi per periodo temporale significativo.
13.5	Altro	Riferimenti alle altre più importanti diffusioni di dati effettuate.	Descrivere eventuali altri importanti meccanismi di diffusione, ad esempio <i>working paper</i> .

Per ciascun meccanismo di diffusione, descrivere:

- le politiche dei prezzi e il loro probabile effetto sull'accesso degli utenti;
- come è stata valutata l'accessibilità (ad esempio mediante consultazione degli utenti) e riassumere i risultati di tale valutazione;
- eventuali modifiche che verranno apportate per migliorare l'accesso nel prossimo futuro;
- come è stata valutata la chiarezza e i risultati della valutazione;
- le modifiche che verranno apportate per migliorare la chiarezza.

Nel campo 13.4:

- descrivere il processo di registrazione per l'accesso al database online (se presente) e commentare la sua facilità di uso;
- fornire e analizzare i valori dell'indicatore “Numero di consultazioni di tabelle di dati all'interno di un dominio statistico” per mese/trimestre/anno e riassumere le conclusioni.

Nel campo 13.5:

- riassumere le limitazioni all'accesso che sono conseguenza delle norme sulla riservatezza e come queste sono state spiegate agli utenti, con un rimando alla sezione “Politica della riservatezza”;
- riassumere l'accessibilità e la chiarezza dei dati associati ai vari formati di diffusione, includendo i risultati delle consultazioni degli utenti e gli effetti delle politiche di prezzo e delle disposizioni sulla riservatezza;
- fornire informazioni sul supporto agli utenti per l'accesso e l'interpretazione dei dati;
- descrivere i dati eventualmente forniti all'Istat, ad altri enti Sistan o ad altri soggetti, che non sono già descritti nella sezione “Mandato istituzionale”;
- descrivere la fornitura interna di dati ad altri uffici all'interno dell'ente di appartenenza che in ultima analisi portano alla diffusione di tali dati.

Esempi

Analisi statistiche dei dati derivanti dalle dichiarazioni dei redditi delle persone fisiche (FIN-00004)

Ministero dell'Economia e delle Finanze - Dipartimento delle Finanze

13.1

Calendario di diffusione su:

<https://www.finanze.gov.it/it/statistiche-fiscali/calendario-delle-pubblicazioni/>

Il calendario di diffusione è reperibile nella sezione “Dati e statistiche”; il file annovera le date nell'anno in corso e il periodo di riferimento dei dati. Per conferire maggiore visibilità, nella *homepage* del Dipartimento del Ministero dell'Economia e delle Finanze è in evidenza l'ultima statistica in uscita prevista da calendario.

13.2

Comunicato stampa: analisi sintetica e analisi estesa delle statistiche sulle dichiarazioni fiscali-Irpef. Nell'analisi sintetica sono incluse anche le dichiarazioni Iva.

13.3

Pubblicazione: Gli immobili in Italia (ricchezza, reddito e fiscalità immobiliare) su: https://www1.finanze.gov.it/finanze/immobili/public/contenuti/immobili_2023.pdf.

13.4

Catalogo *open data* Dichiarazioni:

https://www1.finanze.gov.it/finanze/analisi_stat/public/index.php?opendata=yes.

2. Guida alla compilazione

13.5

Il Dipartimento delle Finanze diffonde e promuove i dati statistici sulle dichiarazioni fiscali in formato aperto, in accordo con il codice dell'Amministrazione digitale e le linee guida indicate dall'Agenzia per l'Italia digitale. Le distribuzioni statistiche sono esportabili in PDF, Excel, CSV e nel formato a 5 stelle RDF, contenente metadati che contestualizzano i dati (*linked open data*).

Al fine di agevolare la consultazione e il riuso, è stata predisposta una raccolta dei dati dell'ultima annualità disponibile, organizzata tramite filtri di ricerca dinamici, il catalogo degli *open data* delle dichiarazioni.

Il Dipartimento rende inoltre disponibile un file contenente i principali dati fiscali di tutti i comuni. L'esportazione dei dati comunali è possibile nei formati CSV e RDF e sono presenti i valori *missing* al fine di evitare rischi di identificabilità dei soggetti, nel rispetto della vigente normativa per la tutela della *privacy*. Per una corretta interpretazione delle informazioni, nelle pagine dedicate alle dichiarazioni fiscali sono disponibili le analisi dei dati e le novità normative sui principali valori contenuti nelle statistiche. I dati sono aperti con licenza di uso *Creative Commons 3.0* e diffusi mediante l'uso delle attuali tecnologie digitali e dei più diffusi standard del web: le pagine web sono accessibili e validate secondo le indicazioni del *World Wide Web Consortium* (W3C). *Working paper* su:

<https://www.finanze.gov.it/it/il-dipartimento/collana-di-lavori-e-di-ricerca/working-papers>.

Prestazioni pensionistiche e beneficiari del sistema pensionistico italiano (IPS-00081)

INPS - Istituto nazionale di previdenza sociale

13.1

Il calendario è disponibile al pubblico nella sezione Osservatori statistici e altre statistiche. La pagina web dedicata contiene le date delle pubblicazioni per l'intero anno in corso, il periodo di riferimento dei dati e i link diretti ai singoli osservatori, così da garantire modalità di accesso trasparenti e più agevoli nel reperire le informazioni. L'integrazione dei dati di questo osservatorio, denominato "Casellario centrale dei pensionati" con quelli degli altri osservatori statistici consente di elaborare un quadro complessivo della dinamica e delle caratteristiche e di pensioni e pensionati italiani, che confluiscono in portali e pubblicazioni.

<https://www.inps.it/it/dati-e-bilanci/osservatori-statistici-e-altre-statistiche/calendario-osservatori-statistici.html>.

13.2

Link ai comunicati stampa: <https://www.inps.it/it/it/inps-comunica/ufficio-stampa/comunicati-stampa.html>.

13.3

Pubblicazione: "Rapporto annuale"

<https://www.inps.it/it/dati-e-bilanci/rapporti-annuali/xxiii-rapporto-annuale.html>.

Statistica breve, corredata da una nota metodologica

servizi2.inps.it/servizi/osservatoristatistici/api/getAllegato/?idAllegato=1007.

13.4

Banche dati: dati disponibili in formato tabellare e grafico, esportabili in formato CSV su: <https://servizi2.inps.it/servizi/osservatoristatistici/4>.

Open data su <https://www.inps.it/it/dati-e-bilanci/open-data/scarica-gli-open-data.html>.

13.5

Analisi e approfondimenti tematici, *report* e studi su mercato del lavoro, pensioni, prestazioni a sostegno del reddito, mortalità: <https://www.inps.it/it/dati-e-bilanci/osservatori-statistici-e-altre-statistiche/analisi-e-approfondimenti-tematici.html>; *working*

paper, studi e analisi per approfondire particolari aspetti sociali e previdenziali del nostro Paese: <https://www.inps.it/it/it/dati-e-bilanci/attivit--di-ricerca/pubblicazioni/working-papers.html>.

È data la possibilità agli utenti di esprimere un parere sul portale INPS: “Il tuo parere è importante. Tutti i suggerimenti, i *feedback* e le opinioni degli utenti che navigano il portale sono fondamentali per migliorarlo e soddisfare le vostre esigenze”.

Rilevazione delle compravendite immobiliari (FIN-00009)

Agenzia delle Entrate - Ministero dell'Economia e delle Finanze

13.1

Calendario delle pubblicazioni dell’Osservatorio del mercato immobiliare (OMI) su:

https://www.agenziaentrate.gov.it/portale/documents/d/guest/calendario_pubblicazioniomi_2025-1;

https://www.agenziaentrate.gov.it/portale/documents/20143/262451/Calendario_pubblicazioniOMI_2024.pdf.

Il calendario è menzionato nella pagina principale dell’area tematica dell’OMI. Il calendario presenta a tutti gli utenti informazioni molto dettagliate: alle date di pubblicazione si aggiunge una descrizione sintetica dei prodotti, con il riferimento temporale e spaziale dei dati.

13.2

Elenco dei più recenti comunicati stampa:

<https://www.agenziaentrate.gov.it/portale/web/guest/agenzia/agenzia-comunica/comunicati-stampa>;

Archivio dei comunicati stampa:

<https://www.agenziaentrate.gov.it/portale/web/guest/archivio/agenzia-comunica/archivio-comunicati-stampa>.

13.3

Nel calendario delle pubblicazioni sono riportate in dettaglio, per ogni mese, le date delle presentazioni dei nuovi prodotti, delle conferenze stampa e delle pubblicazioni sul sito web dell’Agenzia.

- Statistiche trimestrali sull’andamento del mercato immobiliare.
- Rapporti mutui ipotecari con allegato statistico.
- Rapporti immobiliari residenziali con allegato statistico, con il contributo dell’ABI (Associazione bancaria italiani).
- Rapporti immobiliari non residenziali, con il contributo di Assilea (Associazione nazionale delle società di *leasing*).
- Statistiche regionali.
- Statistiche catastali.

13.4

Le banche dati disponibili sono: “Quotazioni immobiliari”, portale con ricerca testuale o ricerca su mappa tramite il servizio di navigazione territoriale Geopoi, attraverso il quale è anche disponibile il *download* gratuito dei perimetri vettoriali delle zone OMI; “Valori agricoli medi” (PDF adatto per la consultazione); “Volumi di compravendita” (CSV adatto per l’elaborazione successiva); “Stock catastale” (sono disponibili gratuitamente, a seguito di specifica e motivata richiesta con finalità di studio e ricerca); “Locazioni immobiliari” (file CSV).

13.5

È disponibile il servizio “Forniture dati OMI”, che consente il *download* gratuito delle quotazioni immobiliari, dei perimetri delle zone OMI e dei volumi di compravendita (dettaglio co-

2. Guida alla compilazione

munale) dell'OMI attraverso l'accesso ai servizi telematici disponibili nell'area riservata, aperta a tutti, dell'Agenzia. I dati non disponibili nelle forniture dati OMI possono essere richiesti da:

- i Consulenti Tecnici di Ufficio (CTU), se richieste secondo le modalità indicate nella sezione "Accesso ai dati per i Consulenti tecnici di ufficio";
- gli studenti universitari e i dottorandi, se richieste per finalità di studio e ricerca, secondo le modalità indicate nella sezione "Accesso ai dati per gli studenti universitari";
- i Comuni, utilizzando il servizio "Estrazione dati OMI" disponibile nell'ambito di quelli forniti attraverso il portale per i comuni.

Per chi naviga con *smartphone* o *tablet* è disponibile l'applicazione *OMI Mobile*.

Per maggiori informazioni si consiglia di consultare la *Guida alla consultazione delle quotazioni OMI*.

2.13.2 Documentazione metodologica

Codice	Nome del concetto	Definizione	Linea guida
13.6	Documentazione metodologica	Documenti che descrivono i metodi utilizzati per la produzione dell'output statistico.	Elencare gli eventuali documenti metodologici o disponibili, relativi al processo statistico di interesse.

Fornire, in questa voce del *quality report*, l'elenco dei documenti e dei file relativi alle metodologie applicate. Il contenuto dettagliato della documentazione va invece descritto nella sezione "Gestione della qualità". I documenti metodologici cui fare riferimento possono essere:

- note metodologiche generali sul processo;
- documenti di approfondimento relativi a una fase specifica, ad esempio la metodologia adottata per il campionamento, per la stima, per l'integrazione, per la codifica, eccetera;
- documenti relativi alla metodologia per uno specifico output prodotto, ad esempio un indicatore particolarmente complesso.

Si possono riportare anche link a lavori presentati a convegni o pubblicati su riviste scientifiche.

Esempi

Registro nazionale degli eventi coronarici e cerebrovascolari maggiori (ISS-00039)

ISS - Istituto superiore di sanità

Le procedure e le metodologie del sistema di registrazione e validazione sono descritte nel protocollo dello studio: <https://www.cuore.iss.it/indicatori/pdf/01-8.pdf>.

PLUS: Partecipation Labour Unemployment Survey - Indagine campionaria nazionale sulle caratteristiche e le aspettative degli individui sul lavoro (IAP-00004)

INAPP - Istituto nazionale per l'analisi delle politiche pubbliche

Materiali relativi alle singole annualità - nota metodologica:

<https://www.inapp.gov.it/wp-content/uploads/Non-organizzati/Nota-metodologica-Indagine-Plus-2022.pdf>;

<https://www.inapp.gov.it/wp-content/uploads/Non-organizzati/Nota-metodologica-Plus-2021.pdf>;

<https://www.inapp.gov.it/wp-content/uploads/2023/01/Plus-20014-Nota-Metodologica.pdf>;
<https://www.inapp.gov.it/wp-content/uploads/2023/01/Plus-2011-Nota-metodologica.pdf>;
<https://www.inapp.gov.it/wp-content/uploads/2023/01/Plus-2010-Nota-metodologica.pdf>.

Statistica annuale della produzione e del consumo di energia elettrica in Italia Terna - Rete elettrica nazionale S.p.A.

Link alla documentazione sulla qualità e sulla metodologia:

https://download.terna.it/terna/Nota_metodologica_8da42531c0ff392.pdf.

2.14 Costi e onere per i rispondenti

Codice	Nome del concetto	Definizione	Linee guida
14	Costi e oneri sui rispondenti	Costo associato alla raccolta e alla produzione di un prodotto statistico e onere per i rispondenti.	<p>a) Costo: fornire i costi operativi annuali del processo, con ripartizione per principale componente di costo. Descrivere i recenti e correnti sforzi effettuati per migliorare l'efficienza e commentare la misura in cui tramite l'uso della ICT (<i>Information and Communication Technology</i>) vengono automatizzate operazioni di routine, come codifica, validazione e imputazione.</p> <p>b) Onere: fornire una stima dell'onere per il rispondente imposto dalla raccolta dati diretta. Non si applica all'acquisizione di dati amministrativi o microdati di altra indagine, né all'uso di dati aggregati. Descrivere i mezzi adottati per ridurre al minimo il disturbo statistico.</p>

La misurazione e il monitoraggio dei costi e degli oneri sono necessari per analizzare e migliorare il processo statistico. Sebbene non siano aspetti correlati direttamente alla qualità dell'output, ne rappresentano un vincolo fondamentale. Pertanto, la loro valutazione è necessaria per prendere decisioni ponderate, considerando le risorse a disposizione e gli obiettivi prefissati.

I costi sono spesso difficili da calcolare: un approccio semplice consiste nel misurare i principali costi "diretti" (i costi che possono essere attribuiti direttamente al processo statistico, tipicamente i giorni lavorativi dedicati).

La scelta appropriata dell'approccio di misurazione dipende dal sistema di contabilità dei costi nell'ente/istituzione. Se vi fossero adeguate ed esaustive informazioni, sarebbe preferibile un approccio "completo", comprensivo dei costi "indiretti", cioè i costi condivisi con altri processi statistici (ad esempio i costi di un registro statistico) e i costi generali (contratti di locazione e utenze dell'ufficio, eccetera).

Nel *PsnPlus*, nella definizione della stima delle spese previste dai lavori PSN, si applica il concetto di "costo aggiuntivo", che l'ente titolare sostiene per realizzare il lavoro e che è conseguente all'inserimento del lavoro nel PSN. Sono, cioè, da escludersi dal computo i costi delle attività gestionali e amministrative strettamente connesse alle attività istituzionali, che il soggetto titolare sosterrebbe comunque, anche se collegate alla produzione di informazione statistica.

Si può considerare allo scopo il seguente prospetto, mutuato dalla Scheda Costi del *PsnPlus*.

2. Guida alla compilazione

Prospetto 2.1 - Documentazione dei costi

COSTI ANNUALI	
(1)	Giornate persona impegnate dall'ufficio di statistica dell'ente
(2)	Giornate persona impegnate da altro Ufficio dell'ente incaricato, diverso dall'Ufficio di statistica
(3)	Giornate persona impegnate in altri Uffici/Strutture dell'ente coinvolte nel progetto o di Uffici di altro ente
(4)	Totale giornate/persona (1 + 2 + 3)
(5)	Costo medio giornata/persona
(6)	Costo del Personale (4 * 5)
(7)	Costi o quote di costi per l'acquisto di beni e servizi correnti utilizzati per la realizzazione del lavoro PSN
(8)	Quote di costi pluriennali imputati al progetto
(9)	Costo complessivo del progetto (6 + 7 + 8)

Fonte: *PsnPlus*

Nella voce (7) devono, ad esempio, considerarsi:

- spese per la raccolta, l'elaborazione e la diffusione di dati statistici;
- acquisto di cancelleria e materiali di consumo;
- consulenze, locazioni e noleggi;
- spese di spedizione, trasporto e facchinaggio.

Nella voce (8) ricadono i costi o le quote dei costi pluriennali attinenti al lavoro statistico che, sostenute in un esercizio finanziario, si estendono per più esercizi, come ad esempio:

- quote di ammortamento dei beni materiali e immateriali;
- manutenzioni straordinarie;
- campagne promozionali;
- acquisti di software;
- ricerche e studi.

L'onere statistico grava sui rispondenti (soggetti pubblici o privati, individuali o collettivi) solo nell'indagine diretta, ovvero un processo statistico che prevede la raccolta dei dati direttamente dalle unità rispondenti; non è presente, infatti, nell'acquisizione dei dati da fonte di tipo amministrativo o da nuove fonti di dati (*big data*) o da microdati provenienti da altre indagini, o da rielaborazione e/o aggregazione di dati già disponibili.

La compilazione di un questionario impone infatti un carico di risposta sulle unità di rilevazione, che deve essere misurato e ridotto al minimo allo scopo di limitare le mancate risposte (totali o parziali). Nel rispetto dei principi fondamentali della produzione di statistica ufficiale, tale *burden* deve essere proporzionato alle esigenze degli utenti e non deve essere eccessivo per i rispondenti. L'istituzione ha pertanto la responsabilità di verificare la gravosità dell'onere arrecato ai rispondenti e stabilire un programma per la sua riduzione nel tempo, in particolare evitando la "sovraposizione", sia temporale sia tematica, con altre indagini dirette. Quest'ultima indicazione deve essere inquadrata nella più ampia possibilità di collaborazione tra le istituzioni per la progettazione di indagini e/o la condivisione di dati (microdati e dati aggregati) nell'ambito di specifici protocolli di intesa, così da evitare il duplicarsi di processi statistici da cui scaturiscono i medesimi risultati, ovvero per realizzare soluzioni sinergiche capaci di innalzare l'efficienza complessiva del sistema.

È pertanto fondamentale assicurare ai rispondenti che sono stati adottati tutti gli accorgimenti necessari per minimizzare tale carico, per esempio prevedendo l'inserimento di quesiti precompilati nei questionari, oppure esplicitando che l'insieme e il dettaglio dei dati raccolti siano stati limitati a quanto necessario per soddisfare gli obiettivi dichiarati dell'indagine, o le ragioni per cui non è stato possibile sostituire la raccolta dei dati con altre fonti

amministrative o altre fonti non dirette. Anche il supporto di una adeguata rete di rilevatori sul campo o in *back office*, valido contributo per favorire la partecipazione all'indagine e contenere le mancate risposte, rappresenta uno strumento utile al fine di alleggerire il carico statistico. Il peso sui rispondenti può essere stimato mediante il tempo impiegato (in numero totale di ore) per la compilazione del questionario. Esso può essere derivato da apposita domanda sul questionario in cui si chiede il tempo impiegato per la compilazione, può essere osservato dall'intervistatore nel caso di un'intervista "faccia a faccia", può essere derivato in indagini web, oppure semplicemente stimato attraverso un esercizio o simulazione interna. Questa misurazione deve però tener conto anche del grado di difficoltà nel reperimento delle informazioni richieste da parte del rispondente, legato all'utilizzo di una terminologia non corrente per il rispondente o alla necessità di consultazione di altri documenti o archivi amministrativi.

Potrebbe essere utile definire e misurare anche l'onere percepito dai rispondenti. Ciò può includere la chiarezza delle domande, la facilità di navigazione nel questionario e la semplicità con cui gli intervistati possono accedere alle informazioni richieste per rispondere alle domande. La misurazione dell'onere percepito può essere effettuata utilizzando un breve set di domande alla fine del questionario o, successivamente, tramite un contatto di *follow-up* con un sottocampione della popolazione di indagine.

Esempi

Rilevazione delle tipologie e caratteristiche dei clienti negli esercizi ricettivi (EMR-00028)
Regione Emilia-Romagna

La Regione Emilia-Romagna comunica nella pagina web dedicata alle rilevazioni statistiche sul turismo (<https://statistica.regione.emilia-romagna.it/metadati/rilevazioni/turismo>) che, per agevolare e rendere più tempestiva la raccolta dati sul movimento turistico e ridurre il disturbo per i rispondenti, mette a disposizione delle strutture ricettive *Ross1000*. L'applicazione web di acquisizione dei dati offre anche strumenti utili all'analisi della gestione dell'impresa (quali statistiche di sintesi, *benchmarking*, studi di settore) e permette di disporre di dati più funzionali alle nuove politiche di programmazione e promozione turistica.

2.15 Commenti

Codice	Nome del concetto	Definizione	Linea guida
15	Commenti	Testo descrittivo supplementare che accompagna dati o metadati.	Fornire informazioni pertinenti al processo statistico ma non adatte a uno qualsiasi degli altri concetti.

Questo ultimo campo del *quality report* può essere utilizzato per fare riferimento agli allegati che sono riferiti al processo.

APPENDICE 1 - ESEMPIO DI *QUALITY REPORT* PRECOMPILOT¹

Il seguente *quality report* costituisce il documento precompilato, grazie ai metadati della scheda PSN, da inviare all'ente Ismea relativamente al lavoro PSN “Panel delle aziende agricole” (ISM-00015).

Si noti che nella sezione Contatti i dettagli personali sono stati soppressi per preservare la *privacy*.

Prospetto A.1 - Precompilazione di un *quality report*

QUALITY REPORT		
1. Contatti	1.1 Ente	Ismea - Istituto di servizi per il mercato agricolo alimentare.
	1.2 Unità organizzativa	Direzione Filiera e Analisi dei Mercati - Unità operativa Analisi di Settore e Filiera Agroalimentari.
	1.3 Responsabile	xxxx
	1.4 Funzione del responsabile	Dirigente della Direzione Filiera e Analisi dei Mercati.
	1.5 Email	xxxx@ismeal.it
	1.6 Telefono	xxxx
2. Ultimo aggiornamento		(da inserire)
3. Descrizione dei dati	3.1 Caratteristiche dei dati	Raccogliere informazioni di tipo qualitativo sulle tendenze e sulle attese a breve termine della produzione, degli acquisti dei mezzi correnti di produzione e dei costi correnti totali, nonché sull'andamento del mercato dei prodotti agricoli.
	3.2 Classificazioni	Sono utilizzate classificazioni ufficiali: Territorio (Codici dei comuni, delle province e delle regioni), classificazione delle attività economiche - Ateco. Altre classificazioni: OTE (Orientamento tecnico-economico).
	3.3 Variabili statistiche	Valutazione sugli aspetti generali del settore di appartenenza; andamento del mercato, commercializzazione, vendite, prezzi di vendita, conferimenti, costi, scorte, problematiche di filiera; andamento del fatturato e degli investimenti; accesso al credito; previsioni di produzione, costi e prezzi di vendita.
	3.4 Unità statistica	Impresa agricola.
	3.5 Popolazione statistica	Universo delle imprese agricole nazionali.
	3.6 Territorio di riferimento	Ripartizioni territoriali, nazione.
	3.7 Copertura temporale	(da inserire)
	3.8 Unità di misura	(da inserire)
	3.9 Periodo temporale di riferimento	2024
4. Descrizione del processo statistico	4.1 Fonte dei dati	È prevista l'acquisizione di dati mediate raccolta diretta, campionaria, di tipo non probabilistico. Numerosità dell'universo di riferimento: 500.000. Numerosità campione teorico: 900. Lista utilizzata da fonte statistica: Istat - Registro Asia agricoltura.
	4.2 Periodicità della raccolta dei dati	Si rilevano trimestralmente informazioni qualitative per il settore agricolo.
	4.3 Periodicità della diffusione	Trimestrale.
	4.4 Raccolta dei dati	Soggetto che raccoglie le informazioni: ditta esterna. La raccolta delle informazioni avviene presso le imprese agricole. Modalità di raccolta delle informazioni: Cati, Cawi. È prevista l'effettuazione di un'indagine pilota.
	4.5 Trattamento dei dati	(da inserire)

¹ L'Appendice 1 è stata redatta da Loredana Amilcare e Paola Giordano.

Prospetto A.1 segue - Precompilazione di un *quality report*

QUALITY REPORT	
5. Mandato istituzionale	Il lavoro ha origine da Atto programmatico del titolare/Atto amministrativo/Circolare: Ipotesi di rinnovo accordo di programma Masaf-Ismea Agriquote. Non è imposto l'obbligo di risposta.
6. Politica della riservatezza	Dati di contatto del Responsabile protezione dati: xxx@ismeia.it. Non sono rilasciati dati elementari.
7. Gestione della qualità	(da inserire)
8. Pertinenza	Circolo di qualità cui afferisce il lavoro: Agricoltura, foreste e pesca
9. Accuratezza	Azioni per ridurre gli errori: - controlli sulle liste di unità di rilevazione; - prevenzione e monitoraggio degli errori dovute ai rilevatori (corsi di formazione, attivazione di <i>help desk</i> , eccetera); - iniziative per favorire la partecipazione dei rispondenti alla raccolta (lettere informative, garanzie di tutela della riservatezza, istruzioni chiare, solleciti); - monitoraggio della raccolta dati attraverso indicatori di qualità.
10. Attendibilità - Revisione dei dati	Non sono diffusi dati provvisori.
11. Tempestività e puntualità	Intervallo di tempo tra il periodo di raccolta dei dati e la prima diffusione dei risultati: 45 giorni (dati definitivi).
12. Comparabilità nel tempo	(da inserire)
13. Accessibilità e chiarezza	13.1 Calendario di diffusione (da inserire)
	13.2 Comunicati stampa (da inserire)
	13.3 Pubblicazioni (da inserire)
	13.4 Portale on line (da inserire)
	13.5 Altro (da inserire)
	13.6 Documentazione metodologica (da inserire)
14. Costi e onere sui rispondenti	Giornate persona impegnate dall'Ufficio di statistica: 2 Costo medio giornata persona: 220,0 Costo del personale: 440,0 Costo complessivo del progetto: 440,0 Il lavoro non richiede finanziamenti aggiuntivi rispetto a quanto previsto nel Bilancio dell'Ente.
15. Commenti	(da inserire)

APPENDICE 2 - GENERAZIONE SDMX DEL *QUALITY REPORT*²

Al fine di agevolare il lavoro degli utenti finali nella creazione degli artefatti SDMX, è stato sviluppato un file in formato Microsoft *Excel* che può essere utilizzato come base per strutturare e organizzare i dati in modo conforme. Questo *template* può essere generato all'interno del *framework* SDMX Istat StatKit, tramite lo strumento *ReferenceMetadataTool*, attraverso un'apposita funzione.

Una volta completata la creazione del *template*, si otterrà un foglio di lavoro strutturato secondo lo schema standard, con la definizione degli elementi in due lingue, inglese e italiano, per garantire una maggiore interoperabilità e accessibilità delle informazioni. Nel Prospetto A.2 è riportato un esempio di questa struttura, che include anche una gerarchia rappresentata a puro scopo esemplificativo.

Prospetto A.2 - *Template Excel* per SDMX

Type	ID	Code	Format	Concept Name (en)	Concept Name (it)
T	1	CONTACT		Contact	Contatti
E	1.1	CONTACT_ORG	String	Contact organisation	Ente
E	1.2	CONTACT_ORG_UNIT	String	Contact organisation unit	Unità organizzativa
E	2	META_LAST_UP	String	Metadata last update	Ultimo aggiornamento
T	3	STAT_PRES		Statistical presentation	Descrizione dei dati
E	3.1	DATA_DESC	String	Data description	Principali caratteristiche
E	3.2	CLASS_SYS	String	Classification system	Classificazioni

Le righe del *template* contrassegnate con *Type* = T rappresentano titoli che, nella fase finale di creazione del *report*, risulteranno non editabili. Al contrario, le righe con *Type* = E consentiranno l'inserimento delle occorrenze dei metadati descritti durante la generazione del *report*. È importante sottolineare che i titoli non sono obbligatori nella costruzione delle gerarchie e, pertanto, è possibile inserire righe con *Type* = E senza che siano precedute da un titolo, come illustrato nell'esempio alla riga con ID = 2.

Partendo da un *template Excel* creato, si procede alla generazione sequenziale degli artefatti SDMX, che costituiscono l'infrastruttura univoca per la creazione e l'esposizione di tutti i *report* di metadati necessari.

I primi due artefatti da creare sono la *MetadataStructure* e il *ConceptScheme*, che hanno l'obiettivo di replicare la struttura definita sul *template*.

Allo scopo, nel modulo *ReferenceMetadataTool* si richiama la funzione *Metadata Structure Definition* (MSD), che prevede il seguente flusso di attività:

1. importare il file *xlsx* relativo al *template*;
2. inserire l'ID della *MetadataStructure* (ad esempio, MS_SISTAN);
3. inserire il nome della *MetadataStructure* in inglese (per esempio, Sistan *metadata structure*) e in italiano (per esempio, Struttura dei metadati Sistan);
4. inserire l'ID del *ConceptScheme* (ad esempio, CS_SISTAN);
5. inserire il nome del *ConceptScheme* in inglese (per esempio, Sistan *metadata concept scheme*) e *ConceptScheme* in italiano (per esempio, Schema dei concetti metadati Sistan);

² L'Appendice 2 è stata redatta da Simone Coccia.

6. creare il file XML contenente *MetadataStructure* e *ConceptScheme*, per il quale dovrà essere impostato un nome (ad esempio, QR_SISTAN).

La MSD è quindi l'oggetto SDMX, cioè un file in formato XML, che replica il *template Excel*.

Importata la MSD nell'ambiente di produzione, diviene quindi possibile, nel modulo *MetaDataManager*, creare l'artefatto che funge da punto di accesso ai metadati referenziali, ovvero il *metadataflow*, che si basa sulla MSD appena creata. Seguendo la gerarchia imposta dallo standard, a valle del *metadataflow* si crea il *metadataset*, che rappresenta l'artefatto finale che contiene il *quality report*. Si può scegliere di definire un unico o diversi *metadataset* in base alle proprie esigenze organizzative.

Per la creazione di un *quality report* si hanno a disposizione due opzioni. Si può operare manualmente tramite l'interfaccia grafica offerta dal *MetaDataManager* oppure optare per un “caricamento rapido” tramite il *template Excel*. È quest’ultima l’ipotesi da preferirsi per la produzione dei *quality report* in ambito Sistan.

È una procedura che non comporta aggravio per l’ente Sistan, che si limiterà a gestire il tutto tramite il file *Excel*, validando o, eventualmente, modificando, i campi precompilati e inserendo le informazioni mancanti.

Centralmente, in Istat, si avrà cura di generare, dal file *Excel*, un formato *Json* del *metadataset*, che viene importato nell’ambiente di produzione e poi esposto via API (*Application Programming Interface*), per renderlo visibile all'esterno.

RIFERIMENTI BIBLIOGRAFICI

Cammarota, M. (a cura di). 2023. *Sistema integrato di diffusione delle statistiche dell'Istat: percorso e stato attuale*. Letture Statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/sistema-integrato-di-diffusione-delle-statistiche-dellistat-percorso-e-stato-attuale/>.

Comitato di indirizzo e coordinamento dell'informazione statistica - Comstat. 2018. *Requisiti per la diffusione da parte degli uffici di statistica degli Enti appartenenti al Sistema statistico nazionale dei dati di titolarità Istat, rilevati e inviati all'Istituto nazionale di statistica per le attività di cui all'art. art. 6, c.1, lett. a), b) e c) del d.lgs. n. 322/1989, non ancora validati dall'Istituto*. Atto di indirizzo n. 3 del 27 marzo 2018. https://www.sistan.it/fileadmin/Repository/Home/NORME_E PROCEDURE/ORGANIZZAZIONE_E_FUNZIONAMENTO/UFFICI_DI_STATISTICA/Comstat Atto%20di%20indirizzo%203_2018.pdf.

Comitato di indirizzo e coordinamento dell'informazione statistica - Comstat. 2011. *Linee guida per il miglioramento della qualità della diffusione delle statistiche ufficiali da parte dei soggetti del Sistema statistico nazionale*. Seduta Comstat del 16 dicembre 2011. https://www.sistan.it/fileadmin/Repository/Home/QUALITA_E_SVILUPPO/CODICE/MONITORAGGIO/Linee_guida.pdf.

European statistical system - ESS. 2020. *Quality Assurance Framework of the European Statistical System*. Version 2.0. Luxembourg, Luxembourg: Publications office of the European union. <https://data.europa.eu/doi/10.2785/847733>.

Eurostat. 2021. *European Statistical System Handbook for quality and metadata reporting. 2021 re-edition*. Luxembourg, Luxembourg: Publications office of the European union. <https://doi.org/10.2785/616374>.

Eurostat. 2015. *Single Integrated Metadata Structure V 2.0 (SIMS V2.0)*. European Statistical System Committee, Twenty-seventh Meeting, Luxembourg, 19 November. <https://ec.europa.eu/eurostat/documents/64157/4373903/SIMS-2-0-Revised-standards-November-2015-ESSC-final.pdf/47c0b80d-0e19-4777-8f9e-28f89f82ce18>.

Eurostat. 2003. *Item 4.2: Methodological documents - Definition of Quality in Statistics*. Working Group "Assessment of quality in statistics", Sixth Meeting, Luxembourg, 2-3 October 2003.

Eurostat, and European statistical system - ESS. 2018. *European Statistics Code of Practice for the national statistical authorities and Eurostat (EU statistical authority)*. Adopted by the European Statistical System Committee, 16th November 2017. Luxembourg, Luxembourg: Publications Office of the European Union. <https://doi.org/10.2785/798269>.

Istituto Nazionale di Statistica - Istat. 2023. *Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali*. Letture statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/guida-per-limplementazione-del-codice-italiano-per-la-qualita-delle-statistiche-ufficiali/>.

Istituto Nazionale di Statistica - Istat. 2021. *La nuova politica per la qualità della produzione statistica*. Roma, Italia: Istat. <https://www.istat.it/wp-content/uploads/2024/01/politica-qualita-2021-v2.pdf>.

Istituto Nazionale di Statistica - Istat. 2018. *Linee guida per la qualità delle statistiche del Sistema statistico nazionale*. Ver. 1.0. Marzo 2018. Roma, Italia: Istat. <https://www.istat.it/wp-content/uploads/2023/12/Linee-Guida-2.5-agosto-2018.pdf>.

