

QUALITÀ E STATISTICHE UFFICIALI: METODI, STRUMENTI E SOLUZIONI ALL'ISTAT





QUALITÀ E STATISTICHE UFFICIALI: METODI, STRUMENTI E SOLUZIONI ALL'ISTAT

A cura di Mauro Scanu e Giorgia Simeoni.

Revisione redazionale: Roberta Pazzini.

Attività editoriali: Claudio Bava e Alfredina Della Branca.

Grafica: Sofia Barletta.

ISBN 978-88-458-2196-7

© 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.
Premessa	7
Introduzione	9
1. Qualità nella statistica ufficiale: contesto interno, nazionale ed europeo	13
1.1 Il concetto di qualità nella statistica ufficiale e il <i>framework</i> di riferimento	13
1.2 La qualità delle statistiche ufficiali rispetto al contesto istituzionale e normativo	17
1.3 Il Sistema statistico europeo	19
1.4 Le <i>Peer review</i> nel Sistema statistico europeo	20
1.4.1. <i>Evoluzione delle Peer review</i>	22
1.4.2. <i>Impegno richiesto e attuazione delle azioni di miglioramento</i>	26
1.4.3. <i>Benefici della Peer review, esperienza dell'Istat e sfide future</i>	27
2. Strumenti per la qualità	29
2.1 Il nuovo sistema dei metadati Metastat	29
2.1.1 <i>Esigenza di un nuovo sistema dei metadati</i>	29
2.1.2 <i>Metastat: un sistema, tre moduli chiave</i>	31
2.1.3 <i>Nuovo sistema dei metadati e standard internazionali</i>	34
2.1.4 <i>Processo di validazione dei metadati</i>	36
2.1.5 <i>Sfide future per il sistema Metastat</i>	37
2.2 La standardizzazione di metodi e strumenti	37
2.2.1 <i>Metodi e software del processo statistico e governance sullo sviluppo di strumenti open source</i>	38
2.2.2 <i>Sviluppi futuri: il Catalogo dei metodi per la produzione statistica come strumento per la qualità</i>	39
2.3 La <i>Quality Report Card</i> dei dati amministrativi, il sistema di documentazione della qualità dei dati amministrativi	41
2.3.1 <i>Quality Report Card: un sistema flessibile in evoluzione</i>	44
2.3.2 <i>Qualità dell'input per i dati amministrativi</i>	48
2.4 Il Dossier Qualità	49
2.4.1 <i>Ambito e contenuti del Dossier Qualità</i>	49
3. Metodi per la valutazione della qualità	53
3.1 La valutazione dei processi tradizionali	54
3.1.1 <i>Valutazione della conformità tramite checklist</i>	54
3.1.2 <i>Valutazione della qualità tramite audit</i>	58



	Pag.
3.2 La qualità dei registri del Sistema integrato dei registri	62
3.2.1 QSIR, il framework per la qualità del SIR	64
3.2.2 Sviluppo delle schede e alcuni esempi	66
3.2.3 Riflessioni sulla implementazione e su possibili sviluppi	71
3.3 Lo studio di un <i>framework</i> per la valutazione della qualità delle <i>Trusted Smart Statistics</i>	72
3.3.1 Aspetti normativi, istituzionali e tecnologici	73
3.3.2 Framework di riferimento per la qualità delle TSS	74
3.3.3 Verso un framework di qualità delle TSS orientato al processo	76
3.3.4 Identificazione dei primi elementi della qualità dell'input: la classificazione delle fonti	77
3.3.5 Approccio empirico per la validazione del modello concettuale del framework	78
3.3.6 TSS e intelligenza artificiale	80
3.3.7 Sviluppi futuri	82
4. Azioni volte ad assicurare la qualità nel Sistema statistico italiano	83
4.1 Il Codice italiano per la qualità delle statistiche ufficiali	86
4.2 La Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali	89
4.3 Il manuale per il <i>quality reporting</i>	92
4.4 Sviluppi futuri	97
5. Valorizzazione delle esigenze degli <i>stakeholder</i> della statistica ufficiale	101
5.1 L'attenzione agli utenti dell'informazione statistica	101
5.1.1 Servizio di assistenza agli utenti tramite il Contact Centre	102
5.1.2 Metodi per rilevare le esigenze informative degli utenti	104
5.1.3 Rilevazioni sulla soddisfazione	105
5.2 La raccolta dei dati e la qualità	107
5.2.1 Approccio centrato sul rispondente: il paradigma per le indagini statistiche	109
5.2.2 Metodologie per l'ottimizzazione dei questionari	110
5.2.3 Interviste cognitive	111
5.2.4 Feedback degli intervistatori	112
5.2.5 Feedback dei rispondenti	114
5.2.6 Analisi delle richieste di assistenza ricevute dal Contact Centre	114
5.2.7 Analisi dei paradata	115
5.2.8 Controlli di qualità come supporto alla compilazione	116
5.2.9 Servizio di assistenza ai rispondenti tramite Contact Centre	117
5.2.10 Servizio di promemoria telefonici tramite Contact Centre	120
5.2.11 Misura e azioni per la riduzione dell'onere statistico sulle imprese	120
5.2.12 Relazioni con gli enti titolari dei dati amministrativi acquisiti dall'Istat	123
5.3 L'investimento nella cultura della qualità	124
5.3.1 Iniziative di formazione sulla qualità dedicate al personale Istat	125
5.3.2 Focus sulle attività formative verso il Sistan	125
5.3.3 Iniziative di formazione sulla qualità in ambito internazionale	127
Riferimenti bibliografici	129

ACRONIMI

AIS	Automatic Identification System
Ateco	Classificazione delle attività economiche
BDQF	Big Data Quality Framework
CAPI	Computer-Assisted Personal Interviewing
CATI	Computer-Assisted Telephone Interviewing
CAWI	Computer-Assisted Web Interviewing
CDQ	Circoli di qualità
Comstat	Comitato di indirizzo e coordinamento dell'informazione statistica
Cuis	Commissione nazionale degli utenti dell'informazione statistica
DQ	Dossier Qualità
EFTA	European Free Trade Association
EMOS	European Master in Official Statistics
ESS	European Statistical System
ESTP	European Statistical Training Programme
GPS	Global Positioning System
GSBPM	Generic Statistical Business Process Model
GSIM	Generic Statistical Information Model
INS	Istituti Nazionali di Statistica
ISO	International Standard Organization
ML	Machine Learning
MNO	Mobile Network Operator
MRT	Mancate risposte totali
ONA	Other National Authorities
PET	Privacy Enhancing Technologies
PSN	Programma statistico nazionale
Puc	Punto unico di contatto
QAF	Quality Assurance Framework
QF4SA	Quality Framework For Statistical Algorithms
QRCA	Quality Report Card dei dati amministrativi
RS	Registri statistici
SAQ	Self-Assessment Questionnaire
Sidi-SIQual	Sistema informativo di documentazione delle indagini – Sistema informativo sulla qualità dei processi statistici
Sigma	Sistema per la gestione dei microdati amministrativi e statistici
SIM	Sistema integrato di microdati
SIMS	Single Integrated Metadata Structure
SIR	Sistema integrato dei registri
TSE	Total Survey Error
TSS	Trusted Smart Statistics
XAI	eXplainable Artificial Intelligence

PREMESSA

L'attenzione dell'Istituto Nazionale di Statistica verso la qualità costituisce un suo tratto distintivo a partire dagli anni Novanta del secolo scorso, quando prese forma un approccio organico alla documentazione dei processi e dei prodotti statistici e al presidio metodologico. Già in quella fase iniziale, fu avviato un percorso strutturato volto a garantire la trasparenza dei processi produttivi, la tracciabilità delle scelte metodologiche e la valutazione sistematica dei dati e delle statistiche diffuse. In questo contesto nacquero sia il Sistema informativo di documentazione delle indagini (Sidi) sia, successivamente, il Sistema informativo sulla qualità dei processi statistici (SIQual), strumenti che hanno rappresentato un'evoluzione significativa nella capacità dell'Istat non solo di descrivere i processi, ma anche di controllare, misurare e migliorare gli standard di qualità. Parallelamente allo sviluppo di tali sistemi, si consolidò la partecipazione dell'Istat ai principali tavoli internazionali dedicati alla qualità in ambito statistico, promossi dalle Nazioni unite, dalla United Nations Economic Commission for Europe (UNECE) e da Eurostat. Tale partecipazione ha consentito all'Istat di contribuire in modo diretto alla definizione di standard globali, di recepire tempestivamente le innovazioni metodologiche emergenti e di collocarsi stabilmente tra gli attori di riferimento nella costruzione del quadro concettuale della qualità nella statistica ufficiale. Gli investimenti effettuati in quegli anni hanno posto le basi per un percorso di lungo periodo, nel quale l'Istat ha progressivamente acquisito una posizione riconosciuta a livello internazionale per la robustezza del proprio sistema di garanzia della qualità.

L'avvio del processo di modernizzazione nel 2016 ha segnato una nuova fase, caratterizzata da un profondo ripensamento delle architetture, dei processi e delle modalità di produzione delle statistiche ufficiali. La crescente integrazione di fonti amministrative e statistiche, l'utilizzo di metodi innovativi, la diffusione di sistemi produttivi centralizzati e l'evoluzione dei bisogni informativi del Paese hanno reso evidente l'esigenza di un rinnovamento del quadro metodologico e degli strumenti di gestione della qualità. Per rispondere a tali sfide, nel 2020 l'Istat ha avviato un processo di definizione di una nuova politica per la qualità, approvata nel 2021, concepita per assicurare la piena coerenza tra l'evoluzione del modello produttivo e i principi sanciti dal *Codice delle statistiche europee* (Eurostat e SSE 2017). La nuova politica per la qualità ha introdotto un insieme di obiettivi, azioni e strumenti, finalizzati a presidiare in modo sistemico la qualità dei processi e dei prodotti statistici in un contesto profondamente rinnovato. Essa ha rafforzato il ruolo del Comitato Qualità, istituito con l'obiettivo di garantire coordinamento, integrazione e coerenza tra le molteplici iniziative presenti in Istat, e ha formalizzato la figura del *quality manager*, con funzioni di raccordo e presidio tecnico. Grazie a tale assetto organizzativo, l'Istat ha potuto affrontare con maggiore efficacia anche il ciclo 2021-2023 di *peer review*, assicurando una gestione coordinata delle attività preparatorie e delle successive azioni di miglioramento. A distanza di alcuni anni dall'adozione della politica, si può constatare come la maggior parte degli obiettivi fissati sia stata sostanzialmente raggiunta, delineando un quadro evolutivo coerente e allineato alle migliori pratiche del Sistema statistico europeo. Accanto a tali aspetti, negli ultimi anni l'Istat ha investito con decisione nella standardizzazione dei metodi e delle procedure, riconoscendo in essa un presupposto fondamentale, sia per assicurare l'efficienza dei processi, sia per garantire livelli omogenei di qualità in un contesto caratterizzato da elevata complessità e crescente integrazione delle fonti.

Lo sviluppo di metodi e strumenti generalizzati, l'aggiornamento degli standard metodologici interni, la definizione di linee guida coerenti con le evoluzioni internazionali e la revisione dei sistemi informativi dedicati ai metadati – come Metastat, destinato a sostituire Sidi-SIQual – rappresentano i punti cardine di tale investimento.

Particolare rilievo assume la crescente attenzione alla governance delle soluzioni *open source* utilizzate nei processi produttivi, con specifico riferimento all'ecosistema del software *R*. L'Istat ha promosso la creazione di una rete di collaborazione interna finalizzata allo sviluppo, alla manutenzione e alla condivisione di pacchetti e soluzioni metodologiche, integrata da politiche volte a garantire la qualità del software, la tracciabilità delle componenti e l'adozione di standard comuni. Tale investimento risponde all'esigenza di assicurare piena trasparenza dei metodi, replicabilità delle elaborazioni e sostenibilità dei processi, aspetti centrali in un contesto in cui l'uso di strumenti *open source* assume un ruolo crescente nella produzione ufficiale.

In linea con una tradizione consolidata, un ruolo strategico è stato svolto dal rapporto con il mondo accademico. Le collaborazioni scientifiche, le attività congiunte di ricerca, la partecipazione ai programmi formativi europei e nazionali, e il coinvolgimento di docenti ed esperti nelle attività metodologiche hanno continuato a rappresentare un elemento essenziale per l'aggiornamento continuo delle competenze e per la diffusione di una cultura della qualità radicata e condivisa. Questo legame stabile con l'accademia ha favorito l'introduzione di metodologie innovative, l'approfondimento di temi emergenti e la costruzione di un linguaggio comune che rafforza ulteriormente la credibilità della statistica ufficiale.

L'evoluzione dei sistemi di valutazione e documentazione della qualità è andata di pari passo con l'evoluzione dei processi di produzione dell'Istat verso sistemi sempre più multifonte: un nuovo *framework* per il monitoraggio della qualità dei registri statistici del Sistema integrato dei registri dell'Istat è stato recentemente formulato ed è già stato implementato in alcuni dei principali registri, mentre l'attenzione sempre crescente verso l'integrazione nei processi statistici correnti di nuove fonti di tipo *big data* ha spinto a investire verso una riflessione su metodi e misure per monitorare e documentare la qualità delle *trusted smart statistics*.

Questo ebook è il risultato della collaborazione e del contributo trasversale delle diverse componenti dell'Istat (metodologia, raccolta dati, comunicazione e diffusione, affari internazionali, territorio e Sistan), a testimonianza del fatto che la qualità permea e sottende qualunque aspetto del ciclo di vita dell'informazione statistica ufficiale. Illustra gli sviluppi sui vari fronti in maniera sistematica, descrivendo gli avanzamenti e le innovazioni realizzate all'Istat nell'ambito degli strumenti, delle metodologie e delle procedure per la qualità. Delineato il quadro definitorio, istituzionale e normativo di riferimento, si approfondiscono le innovazioni introdotte nei sistemi di metadati, nella standardizzazione dei metodi, nei modelli di valutazione dei processi e nella gestione della qualità dei registri e delle nuove fonti di dati.

Viene evidenziato, inoltre, il ruolo delle azioni rivolte al Sistema statistico nazionale e alla rete degli *stakeholder*, incluse le azioni rivolte alla diffusione della cultura statistica, elementi essenziali per garantire che il principio dell'impegno continuo verso la qualità permei l'intero sistema statistico pubblico. Attraverso questo percorso, contribuisce alla diffusione di una conoscenza condivisa delle trasformazioni in atto, fornendo agli operatori del settore un quadro aggiornato e una base solida per proseguire nel percorso di miglioramento continuo che caratterizza la missione istituzionale dell'Istat.

Orietta Luzi

Direttrice della metodologia e il disegno dei processi statistici

INTRODUZIONE¹

L'investimento dell'Istat nella gestione della qualità nasce almeno trenta anni fa, e grazie a tale investimento oggi l'Istat può contare su solide basi. Solo per fare degli esempi, strumenti come il Sistema informativo sulla qualità dei processi statistici (SIQual) o le *Linee guida sulla qualità dei processi statistici* (Signore *et al.* 2012, Brancato 2016) rappresentano dei riferimenti a livello internazionale, per cui l'Istat è considerato un pioniere e un'eccellenza nell'ambito della qualità della statistica ufficiale. Inoltre, l'Istat ha sempre potuto contare su una struttura organizzativa di riferimento che coordinasse le attività sulla qualità e, già nel periodo 2010-2016, ha condotto un ciclo di procedure di valutazione sui propri processi statistici (*audit* e autovalutazione).

Ciononostante, nel 2020 è stato rinnovato l'impegno per la qualità e per la promozione di una nuova fase di adeguamento degli strumenti e dei metodi di gestione della qualità alle esigenze dell'Istat nella fase successiva alla modernizzazione. A partire dal 2016, infatti, l'Istat ha iniziato un percorso di modernizzazione che ha cambiato profondamente il paradigma della produzione statistica rendendo ora necessaria un'evoluzione degli strumenti e dei metodi di gestione della qualità nella stessa direzione.

Il percorso di rinnovamento ha seguito la direzione indicata dal principio 4 del *Codice delle statistiche europee* (Prospetto 1; Eurostat e SSE 2017) e il primo passo è stato di rafforzare la struttura organizzativa per il coordinamento delle molteplici iniziative sulla qualità presenti all'Istat (Indicatore 4.1: "[...] Esiste una struttura organizzativa [...] per assicurare la gestione della qualità").

Prospetto 1 - Principio 4 del *Codice delle statistiche europee*

Principio 4 - Impegno a favore della qualità

La qualità è un imperativo per le autorità statistiche, che individuano sistematicamente e regolarmente i punti di forza e di debolezza al fine di migliorare costantemente la qualità dei processi e dei prodotti.

Indicatori

- 4.1 La politica per la qualità è definita e resa pubblica. Esiste una struttura organizzativa e sono disponibili strumenti adeguati per assicurare la gestione della qualità.
- 4.2 Sono in atto procedure per pianificare, monitorare e migliorare la qualità dei processi statistici, compresa l'integrazione di dati provenienti da più fonti.
- 4.3 La qualità dei prodotti è regolarmente monitorata e valutata tenendo conto dei possibili compromessi; essa è oggetto di relazioni elaborate in base ai criteri di qualità delle statistiche europee.
- 4.4 È prevista una regolare e approfondita valutazione dei principali prodotti statistici con il ricorso, se del caso, anche a esperti esterni.

Fonte: Eurostat e Sistema statistico europeo

È stato così costituito il Comitato Qualità (di seguito, Comitato), coordinato dal Direttore della Direzione centrale per la metodologia e il disegno dei processi (DCME) dell'Istat e composto da rappresentanti delle diverse direzioni. Il Comitato è supportato da una segreteria scientifica e dal *quality manager*, figura introdotta nel 2020.

¹ L'Introduzione è stata redatta da Giorgia Simeoni.



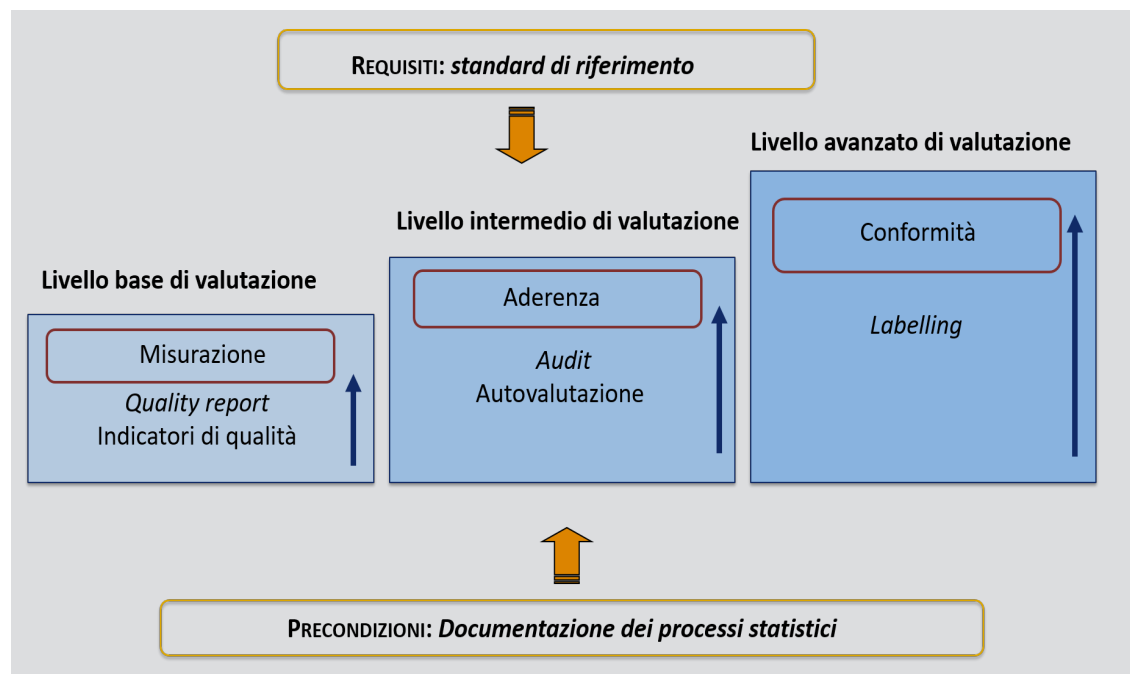
Il *quality manager*, esperto nella qualità della statistica ufficiale, è anche il responsabile dell'iniziativa "Metodi e strumenti per la qualità e la documentazione dei processi e dei prodotti statistici" nell'ambito della DCME, che costituisce il principale supporto operativo alle attività del Comitato. A quest'ultimo è attribuito un ruolo strategico di supervisione delle iniziative sulla qualità, con particolare riferimento a quelle connesse alla *peer review* sull'attuazione dei principi del Codice delle statistiche europee. La sua istituzione risale, infatti, al periodo di preparazione alla *peer review* 2021-2023; a seguito di tale processo, ne segue l'implementazione delle azioni di miglioramento individuate (per un approfondimento cfr. paragrafo 1.4). In linea con il principio 4 del Codice, il primo obiettivo è stato la definizione di una nuova strategia per la qualità della produzione statistica ufficiale (Indicatore 4.1: "la politica per la qualità è definita e resa pubblica [...]"). È stata così elaborata *La nuova politica per la qualità della produzione statistica* (Istat 2021), predisposta a cura del *quality manager* con il supporto del Comitato e approvata nel 2021 dal Comitato di Presidenza. *La nuova politica per la qualità della produzione statistica* è un documento strategico che propone delle azioni per il miglioramento della qualità per un arco temporale di cinque anni. Anche nella definizione della strategia per la qualità si è seguito il percorso tracciato dal principio 4 del Codice, approfondendo dapprima gli strumenti a supporto della qualità (Indicatore 4.1: "[...] sono disponibili strumenti adeguati per assicurare la gestione della qualità") e a seguire i metodi e le procedure per il monitoraggio e la valutazione della qualità dei processi e dei prodotti statistici (Indicatori 4.2, 4.3 e 4.4). Per ciascun ambito di competenza della qualità il documento riporta lo stato dell'arte, le azioni e le innovazioni da porre in essere.

A distanza di cinque anni dalla pubblicazione della nuova politica per la qualità, questo ebook presenta i risultati ottenuti con particolare riferimento all'ambito metodologico. Dopo il primo Capitolo introduttivo, che illustra i concetti e i principi fondamentali della qualità nella statistica ufficiale e il contesto nazionale ed europeo di riferimento, i Capitoli successivi riprendono la struttura del documento del 2021. In particolare, il Capitolo 2 approfondisce i risultati conseguiti nello sviluppo degli strumenti per la qualità, mentre il Capitolo 3, fulcro dell'analisi, descrive gli avanzamenti nelle procedure di valutazione dei processi statistici correnti.

L'approccio alla qualità dell'Istat (Figura 1; Istat 2022) è basato su metodi e strumenti di valutazione a diversi livelli, con l'obiettivo comune del miglioramento della qualità dei processi e dei prodotti statistici e fondato sulla documentazione dei processi statistici. Di conseguenza, gli strumenti a supporto della qualità consistono principalmente nei sistemi informativi per la documentazione della qualità, in primo luogo Metastat (cfr. paragrafo 2.1), il nuovo sistema di metadati in corso di sviluppo che sostituirà l'esistente e ormai datato Sidi-SIQual, e la *Quality Report Card* dei dati amministrativi (QRCA) (cfr. paragrafo 2.3; Di Bella 2021). I sistemi informativi per la documentazione della qualità permettono anche di analizzare la qualità dei processi multifonte che l'Istat ha sviluppato e continua a sviluppare a seguito della modernizzazione. Recentemente è stato anche realizzato il Dossier Qualità, strumento a supporto della dirigenza dell'Istat, che rappresenta un punto unico di accesso a tutta la documentazione metodologica e sulla qualità esistente e aggiornata per ogni indagine diretta (cfr. paragrafo 2.4).

Strumenti fondamentali per la valutazione della qualità sono, inoltre, gli standard di riferimento. L'Istat può contare su linee guida per la qualità dei processi statistici di tipo indagine (Signore *et al.* 2012) e di fonte amministrativa (Brancato 2016), ma la modernizzazione ha dato anche un forte impulso alla standardizzazione dei processi basati su metodologie e strumenti consolidati (cfr. paragrafo 2.2). Tali strumenti sono di supporto alle attività di gestione della qualità dei processi di produzione corrente, realizzate attraverso l'adozione di specifiche procedure e metodi di valutazione dei processi e dei prodotti statistici (Figura 1).

Figura 1 - L'approccio dell'Istat al *quality management*. Anno 2022

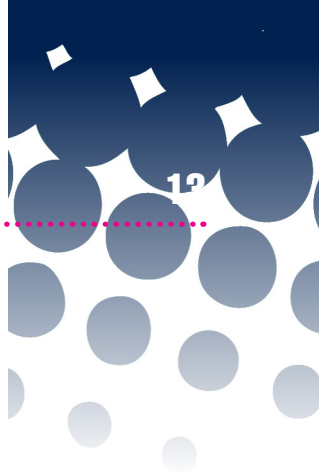


Fonte: Istat

I metodi e le procedure di valutazione della qualità sono differenziati in funzione delle diverse tipologie di processi di produzione: processi di tipo tradizionale (indagini dirette campionarie, censuarie o da fonte amministrativa), processi alla base del Sistema integrato dei registri e processi basati sulle nuove fonti di dati (Capitolo 3).

L'Istat è inoltre impegnato nell'estendere al Sistan l'impianto sviluppato per la garanzia della qualità; gli investimenti realizzati negli ultimi anni in questo ambito hanno portato allo sviluppo di una serie di prodotti, descritti nel paragrafo 1.2 e approfonditi nel Capitolo 4.

Come anche suggerito dai principi generali di gestione della qualità citati nel *Preambolo* del *Codice delle statistiche europee*, è fondamentale garantire alcune dimensioni della qualità, come la pertinenza e l'accessibilità, e altri requisiti come la riduzione del carico sui rispondenti, l'accessibilità ai dati diffusi e la qualità dei dati in input, esigenze che rendono necessario investire nella gestione dei rapporti con i diversi *stakeholder* (utenti interni ed esterni, rispondenti, *data provider* pubblici e privati). Il Capitolo 5 approfondisce i miglioramenti ottenuti nella gestione dei rapporti con i vari *stakeholder*, compresi quelli interni, che contribuiscono con la loro attività alla qualità dei processi e dei prodotti statistici degli uffici di statistica. A tale fine, un tassello fondamentale verso gli statistici ufficiali è rappresentato da una pervasiva e regolare attività di formazione sul tema della qualità.



1. QUALITÀ NELLA STATISTICA UFFICIALE: CONTESTO INTERNO, NAZIONALE ED EUROPEO¹

“La missione dell’Istituto Nazionale di Statistica è servire la collettività attraverso la produzione e la comunicazione di informazioni statistiche, analisi e previsioni di elevata qualità”².

In tale prospettiva, il concetto di qualità nell’ambito della statistica ufficiale è affrontato introducendo l’impianto definitorio e il quadro di riferimento per la garanzia della qualità sviluppati in ambito europeo e adottati dall’Istat.

Si illustra quindi brevemente il contesto nazionale (cfr. paragrafo 1.2) e quello europeo (cfr. paragrafo 1.3) in cui l’Istat opera nell’ambito della gestione della qualità. Il Capitolo si conclude con un approfondimento sulle *Peer review* condotte periodicamente a livello europeo, finalizzate a verificare il grado di attuazione dei principi fondamentali della qualità, come definiti nel *Codice delle statistiche europee* (di seguito, *Codice*) (Eurostat e SSE 2017).

1.1 Il concetto di qualità nella statistica ufficiale e il *framework* di riferimento

Il concetto di qualità è un concetto complesso, multidimensionale e con una componente soggettiva. Si pensi a un oggetto comune, come uno smartphone. La qualità di uno smartphone può essere valutata rispetto a diversi requisiti: la capacità di connettersi alla Rete per effettuare chiamate, inviare messaggi e consultare Internet, la memoria disponibile, la risoluzione delle fotografie, ma anche la facilità di uso, la compatibilità con modelli precedenti e con le diverse applicazioni disponibili, la facilità di acquisto, eccetera. Gli utenti possono dare peso diverso ai suddetti requisiti e quindi giudicare la qualità di uno stesso smartphone in modo differente. Inoltre, i livelli di qualità tendono a modificarsi nel tempo: un prodotto considerato di elevata qualità cinque anni fa può risultare obsoleto oggi. Non a caso, la definizione standard di qualità adottata dall’International Organization for Standardization (ISO) è piuttosto generica: infatti, secondo la ISO 9000:2015, la qualità è definita come “il grado in cui un insieme di caratteristiche intrinseche di un oggetto soddisfa i requisiti” (ISO 2015).

Il concetto di qualità applicato alla statistica ufficiale è egualmente complesso, multidimensionale, soggettivo e variabile nel tempo. Sicuramente un requisito fondamentale di una statistica ufficiale è che rappresenti correttamente la realtà, ovvero che sia accurata, ma anche accessibile, disponibile tempestivamente, presentata in maniera chiara e con tutte le informazioni necessarie a interpretarla correttamente, eccetera. I livelli di qualità possono variare nel tempo, ad esempio con la disponibilità di nuove tecnologie i tempi di produzione dei dati si sono notevolmente ridotti e i requisiti di tempestività tendono a essere più stringenti. Anche in questo caso utenti diversi possono avere diverse priorità, introducendo una componente soggettiva nella valutazione della qualità: per un decisore politico la tempestività può assumere un ruolo fondamentale a scapito di una minore precisione, mentre per un ricercatore universitario la disponibilità di dati più granulari e accurati può essere più rilevante.

¹ Il Capitolo 1 è stato redatto da: Giorgia Simeoni (paragrafo 1.1), Andrea Bruni (paragrafo 1.2), Elisabetta Parente e Marina Gandolfo (paragrafi 1.3 e 1.4).

² Cfr. Istat. <https://www.istat.it/istituto/>.

Non a caso, anche nell'ambito della statistica ufficiale, almeno a livello di Sistema statistico europeo (SSE) si adotta la suddetta definizione ISO 9000:2015 di qualità³.

Malgrado questi gradi di libertà, che comportano inevitabilmente complessità nella definizione di metodi e misure di valutazione della qualità della statistica ufficiale, l'insieme dei requisiti di qualità per la statistica ufficiale è oggi piuttosto consolidato, codificato e condiviso a livello internazionale.

Nel Prospetto 1.1 sono riportati, infatti, i criteri di qualità definiti nell'articolo 12 del regolamento (CE) n. 223/2009, successivamente modificato dal regolamento (UE) n. 759/2015 e più di recente dal regolamento (UE) n. 3018/2024, che ne ha aggiornato i contenuti per tenere conto dell'evoluzione della produzione statistica. Tali norme costituiscono la legge statistica europea (cfr. paragrafo 1.4 per ulteriori informazioni). Gli stessi criteri vengono proposti come principi da soddisfare nella gestione dei prodotti statistici nel *National Quality Assurance Framework* proposto da United Nations (2019).

Prospetto 1.1 - Criteri di qualità delle statistiche europee

Pertinenza: il grado in cui le statistiche rispondono alle esigenze attuali e potenziali degli utenti.

Accuratezza: la vicinanza tra le stime e i valori reali non noti.

Tempestività: il periodo che intercorre tra la disponibilità dei dati e l'evento o il fenomeno da essi descritto.

Puntualità: l'intervallo di tempo che intercorre tra la data di rilascio dei dati e la data obiettivo (data in cui avrebbero dovuto essere forniti).

Accessibilità e chiarezza: le condizioni e le modalità con cui gli utenti possono ottenere, utilizzare e interpretare i dati.

Comparabilità: la misurazione dell'impatto delle differenze tra i concetti statistici applicati, gli strumenti e le procedure di misurazione quando le statistiche si comparano per aree geografiche, ambiti settoriali o periodi di tempo.

Coerenza: la capacità dei dati di essere combinati attendibilmente secondo modalità diverse e per vari usi.

Fonte: Regolamento (CE) n. 223/2009

Come si vedrà, i criteri di qualità dell'output sono utili direttrici per la valutazione e la reportistica sulla qualità, ma per ottenere statistiche le cui caratteristiche intrinseche (come da definizione ISO) rispettino tali requisiti è necessario agire sui processi di produzione di tali output. Partendo dal presupposto che solo implementando processi statistici di qualità si possono produrre sistematicamente output di qualità, il concetto di qualità nella statistica ufficiale si estende dalla qualità del prodotto alla qualità del processo, evidenziando in particolare l'importanza di utilizzare solide metodologie e procedure appropriate, ma anche di perseguire l'efficienza dei processi rispetto ai costi.

All'Istat l'attenzione alla qualità del processo statistico ha portato con sé la tendenza alla standardizzazione per garantire efficienza e livelli generalizzati di qualità, ma anche l'approccio al miglioramento continuo, che si esplica:

- all'interno di ogni edizione di ogni processo statistico, con l'integrazione dei sistemi di monitoraggio e controllo della qualità nei processi produttivi;
- tra diverse edizioni di uno stesso processo, implementando misure e metodi per la valutazione della qualità per identificare punti di debolezza in cui innestare azioni di miglioramento per l'edizione successiva;
- tra diversi processi, estendendo metodologie, strumenti e tecnologie che si rivelano migliorative della qualità in maniera sistematica a tutti i processi simili.

3 Per il termine "quality" incluso nel Glossario di Eurostat, cfr. https://showvoc.op.europa.eu/#/datasets/ESTAT_ESS_QUALITY_GLOSSARY/data?resId=http:%2F%2Fdata.europa.eu%2Fvig%2Fglossary%2Fc_43a00b7a.

1. Qualità nella statistica ufficiale: contesto interno, nazionale ed europeo

Tutto questo diventa possibile anche grazie a una cultura diffusa della qualità nell'Istat, costruita attraverso un programma di formazione sulla qualità consolidato ormai da decenni e allo sviluppo di sistemi informativi che facilitano il monitoraggio, la documentazione e la valutazione della qualità e la condivisione di informazioni.

Il perseguimento dei principi di qualità di prodotto e di processo risulta comunque non ancora sufficiente per garantire la qualità della produzione statistica ufficiale ed è necessario estendere ulteriormente il concetto di qualità fino a considerare alcuni elementi a livello istituzionale. Banalmente, senza indipendenza professionale delle autorità statistiche non ci può essere nessuna garanzia di qualità dei risultati; se le autorità statistiche non hanno un chiaro mandato riconosciuto legalmente per acquisire i dati, non possono garantire la stabilità e l'affidabilità della produzione statistica. Tutti gli aspetti menzionati devono essere considerati nel quadro di riferimento per la garanzia della qualità (*Quality Assurance Framework* - QAF) di un istituto di statistica nazionale come l'Istat. Come membro del Sistema statistico europeo (SSE) (cfr. paragrafo 1.3), l'Istat ha contribuito a costruire e adottato il Quadro comune per la qualità del Sistema statistico europeo, che è fondato sul *Codice*, lo strumento di autoregolamentazione basato su sedici principi (cfr. Figura 1.1; Eurostat e SSE 2017; la Figura è stata ripresa da Cammarota 2024) organizzati in tre aree: contesto istituzionale (principi dall'1 al 6, incluso l'1bis), processi statistici (principi dal 7 al 10) e prodotti statistici (principi dall'11 al 15, che coincidono con i criteri di qualità definiti nella legge statistica europea).

Figura 1.1 - I sedici principi del *Codice delle statistiche europee*



Fonte: Eurostat e Sistema statistico europeo

Nel *Codice*, ogni principio è accompagnato da un insieme di indicatori, in tutto 84, di migliori pratiche e standard che costituiscono allo stesso tempo una guida all'attuazione e il riferimento per la valutazione dell'applicazione del *Codice*.

Il *Codice* è stato adottato nel 2005 con una raccomandazione della Commissione europea (COM/2005/0217 del 25 maggio 2005). Le due successive revisioni del *Codice*, nel 2011 e nel 2017, sono state approvate dal Comitato del SSE.



Particolarmente importante per la gestione della qualità è il principio 4 nell'area del contesto istituzionale, che riguarda l'impegno a favore della qualità. I suoi indicatori indicano una serie di azioni da implementare che possono essere utilizzate per definire una strategia verso il miglioramento costante della qualità. Esse sono state di ispirazione nella definizione della politica della qualità dell'Istat degli ultimi anni, come descritto nell'Introduzione (cfr. Prospetto 1).

Il *Codice* è poi accompagnato dal *Quality Assurance Framework of the European Statistical System* (ESS QAF), documento che suggerisce i metodi da implementare, sia al livello istituzionale, sia a livello di processo/prodotto statistico, per soddisfare i principi del *Codice* (ESS 2019). A titolo di esempio, nel Prospetto 1.2 sono riportati i metodi suggeriti per implementare l'indicatore 4.1 del *Codice*.

Prospetto 1.2 - Metodi dell'ESS QAF per l'Indicatore 4.1 del *Codice delle statistiche europee*

Principle 4: Commitment to Quality. Statistical authorities are committed to quality. They systematically and regularly identify strengths and weaknesses to continuously improve process and output quality.

Indicator 4.1: Quality policy is defined and made available to the public. An organisational structure and tools are in place to deal with quality management.

Institutional methods

1. **A quality policy/commitment statement.** A quality policy / commitment statement sets out principles, practices and commitments related to quality in statistics, consistent with the goals set out in the Mission and Vision statements of the European Statistical System. The policy/statement is publicly available.
2. **An organisational structure for managing quality.** There is a clear organisational structure for managing quality within the statistical authorities. Examples of such a structure are:
 - Quality Committee
 - Quality Manager
 - Centralised Quality Unit
 - Other structures (e.g. members of staff trained to act as quality coaches).
3. **Quality guidelines.** Guidelines are defined on how to implement elements of quality assurance related to the statistical production process, comprising:
 - A description of the different phases of the statistical production process and links to relevant reference documentation for each phase, following the Generic Statistical Business Process Model (GSBPM) or any other equivalent process representation.
 - A description of the methods to assure the quality of each phase of the statistical production process.
4. **Availability of quality guidelines.** Quality guidelines are used and publicly available at least in a summary version.
5. **An infrastructure for documentation.** An infrastructure and resources are in place in order to maintain updated documentation on quality.
6. **Quality culture.** A quality culture is spread in the organisation by means of regular training programmes supporting the implementation of the quality policy, training on the job, regular training courses, workshops and other initiatives.
7. **Risk management.** Risk management is implemented in the organisation and is applied at different levels, for example:
 - Strategic level, i.e. the risks of not following the strategic values and goals
 - European statistics Code of Practice level, i.e. the risks of not being compliant with the European statistics Code of Practice level as a whole or for specific principles
 - Process, output and project level.
8. **Risk and quality management.** Risk and quality management are closely coordinated.
9. **Availability of the European Statistics Code of Practice.** The European statistics Code of Practice is prominently displayed on websites.
10. **References to the European Statistics Code of Practice.** Where appropriate, the European statistics Code of Practice is referred to in press releases that relate to disseminated statistics.

Fonte: European Statistical System

Come già esposto, l'Istat ha contribuito attivamente alla definizione e alle successive revisioni del *Codice*, lo ha ratificato e periodicamente viene monitorato sul livello di aderenza ai principi del *Codice* attraverso il meccanismo delle *Peer review* (cfr. paragrafo 1.3).

Completano il Quadro comune per la qualità del Sistema statistico europeo alcuni principi

generali di gestione della qualità, quali “l’interazione costante con gli utenti, l’impegno a esercitare la leadership, il partenariato, la soddisfazione del personale, il miglioramento continuo, l’integrazione e l’armonizzazione” (Eurostat e SSE 2017). Anche di questi principi si è tenuto conto nella definizione della strategia per la qualità dell’Istat.

L’Istat promuove l’implementazione del *Codice* non solo al suo interno, ma anche nelle Altre autorità nazionali (*Other National Authorities* - ONA)⁴, gli enti italiani che, insieme all’Istat, producono statistiche europee. Infatti, non essendo in genere gli Istituti nazionali di statistica gli unici produttori di statistica ufficiale, è necessario che il QAF sia esteso al Sistema statistico nazionale e che il ruolo di guida e coordinamento dell’Istituto nel Sistema statistico nazionale (Sistan) sia riconosciuto in modo formale. La situazione italiana è da questo punto di vista complessa. Le ONA sono soggette al *Codice delle statistiche europee* come l’Istat, per gli altri membri del Sistan, è stato appositamente sviluppato il *Codice italiano per la qualità della statistica ufficiale* (Sistan e Istat 2022, *Codice italiano* nel seguito), che riflette il *Codice europeo*, adattandolo all’eterogenea realtà degli enti del Sistan.

L’Istat è poi membro di un sistema più ampio, il Sistema statistico europeo (cfr. paragrafo 1.3), che impone vincoli e richieste da soddisfare in termini di statistiche europee da produrre con specifici elevati livelli di qualità al nostro sistema di produzione statistica nazionale, ma funziona anche come leva e supporto e fornisce molte opportunità di crescita al nostro sistema statistico.

1.2 La qualità delle statistiche ufficiali rispetto al contesto istituzionale e normativo

La qualità delle statistiche ufficiali non può essere considerata un mero attributo tecnico o metodologico, ma rappresenta un valore pubblico essenziale, radicato in un solido quadro istituzionale e normativo. Un dato di qualità svolge una funzione strategica: rafforza la fiducia dei cittadini nelle istituzioni, favorisce processi decisionali fondati su evidenze, consente il monitoraggio e la valutazione delle politiche pubbliche e promuove la trasparenza e la partecipazione democratica. La produzione statistica ufficiale deve quindi avvenire in condizioni che assicurino indipendenza, affidabilità, pertinenza, coerenza e comparabilità dei dati, trasparenza e robustezza metodologica, nel rispetto di regole condivise e sotto la responsabilità di soggetti pubblici legittimati a produrli.

In Italia, il riferimento istituzionale della statistica ufficiale è costituito dal Sistan, istituito con il decreto legislativo n. 322 del 1989, che configura un insieme integrato e cooperativo di soggetti pubblici e privati deputati alla produzione, elaborazione, analisi e diffusione di informazioni statistiche di interesse pubblico. Ne fanno parte l’Istituto Nazionale di Statistica, gli uffici di statistica delle amministrazioni dello Stato e di altri enti pubblici, delle regioni e delle province autonome, delle camere di commercio, delle province, dei comuni, singoli o associati, e di altre istituzioni che svolgono funzioni rilevanti per la produzione statistica ufficiale.

Il Sistan è chiamato ad assicurare l’erogazione di dati statistici ufficiali tempestivi, affidabili e di elevata qualità, adeguati a rispondere ai bisogni informativi di cittadini, decisori pubblici, istituzioni e comunità scientifica. In tale contesto, la qualità costituisce un elemento strategico per il rafforzamento del Sistan: è fondamentale che le informazioni messe a disposizione della collettività siano contraddistinte dai più alti standard, in conformità con le direttive europee e internazionali.

⁴ Per l’elenco delle ONA italiane cfr. <https://ec.europa.eu/eurostat/web/european-statistical-system>.

All'interno di tale assetto, l'Istat esercita una funzione di indirizzo e coordinamento delle attività statistiche degli enti e degli uffici, attraverso azioni di assistenza tecnica e valutazione dell'adeguatezza dell'operato degli enti rispetto agli obiettivi del Programma statistico nazionale (PSN), lo strumento operativo per la pianificazione e l'organizzazione della produzione statistica ufficiale in Italia.

È compito dell'Istat, nel suo ruolo di guida del Sistan, definire e promuovere un quadro di azione condiviso per la qualità, in grado di garantire l'efficacia della produzione statistica ufficiale, servizio pubblico essenziale, e la sua coerenza con i bisogni conoscitivi del Paese.

Con il d.p.r. 166/2010, l'Istat ha assunto inoltre la responsabilità del coordinamento delle attività connesse allo sviluppo, alla produzione e alla diffusione delle statistiche europee, in quanto membro del Sistema statistico europeo e *contact point* della Commissione (Eurostat) per le questioni statistiche.

In questa prospettiva, l'Istat aderisce al quadro comune europeo per la garanzia della qualità, fondato sui principi e indicatori del *Codice delle statistiche europee*, richiamato anche nella legge statistica europea (cfr. paragrafo 1.1).

L'entrata in vigore del *Codice* ha rappresentato un passaggio decisivo per consolidare l'autorevolezza della statistica ufficiale prodotta dal Sistema statistico nazionale. Il valore riconosciuto al *Codice* come strumento di certificazione della qualità ha indotto l'Istat e il Comitato di indirizzo e coordinamento dell'informazione statistica (Comstat), l'organo di governo del Sistan, a promuoverne l'applicazione anche a livello nazionale, mediante un processo di adattamento condiviso e partecipativo che ha coinvolto vari attori del Sistema.

A partire dalla prima adozione del *Codice*, l'Istat ha investito con convinzione nella definizione di un linguaggio condiviso sulla qualità, ritenendo essenziale recepire, a livello nazionale, i contenuti e le modalità di attuazione dei dettami europei. Tuttavia, mentre il *Codice* si applica all'Istat e alle ONA, il contesto italiano richiede un'estensione più ampia, poiché la produzione di statistiche ufficiali coinvolge un insieme molto articolato di enti. È stato dunque necessario sviluppare strumenti nazionali paralleli a quelli europei, capaci di tradurre il quadro di riferimento del SSE in regole e pratiche adatte a tutto il Sistan, in modo da assicurare coerenza e uniformità a un sistema caratterizzato da una pluralità di attori.

Al termine di un processo consultivo e di sensibilizzazione che ha coinvolto i diversi soggetti del Sistan, l'impegno dell'Istat ha trovato compiuta espressione, nel 2010, con l'approvazione del *Codice italiano* per la qualità delle statistiche ufficiali, uno strumento di regolamentazione che riprende la struttura e la logica del *Codice europeo*, adattandola alla specificità organizzativa e operativa del Sistan. Il provvedimento è stato adottato con la direttiva Comstat n. 10 del 17 marzo 2010, pubblicata nella Gazzetta Ufficiale della Repubblica Italiana n. 240 del 13 ottobre 2010.

Il *Codice italiano* rappresenta una buona pratica riconosciuta anche a livello internazionale: l'Istat è stato il primo istituto nazionale di statistica in Europa ad adottare formalmente un codice di qualità ispirato all'esperienza comunitaria.

I due Codici, europeo e italiano, condividono la medesima finalità: rafforzare la fiducia nella statistica ufficiale attraverso la definizione di un quadro di principi e valori a garanzia della indipendenza, integrità e responsabilità dei produttori di statistica, nonché fornire uno strumento di autoregolamentazione che permetta il miglioramento continuo della qualità delle statistiche che si producono e si diffondono.

Negli anni successivi, il percorso per il miglioramento continuo della qualità si è arricchito di ulteriori strumenti di indirizzo e coordinamento.

Il Comstat nella seduta del 16 dicembre 2011 ha emanato le “Linee guida per il miglioramento della qualità della diffusione delle statistiche ufficiali da parte dei soggetti del Sistema statistico nazionale”, con l’obiettivo di fornire concrete indicazioni operative agli uffici di statistica, per facilitare una fase della diffusione dei dati statistici trasparente e coerente con i principi del *Codice italiano*.

Un’altra delibera chiave del Comstat è costituita dall’atto di indirizzo n. 3 (comunicato pubblicato sulla GU n.112 del 16 maggio 2018) che nel 2018 definisce i requisiti di qualità, le forme e le modalità di diffusione dei dati che gli enti e gli uffici del Sistan, in qualità di organi intermedi di rilevazione, raccolgono e trasmettono all’Istituto Nazionale di Statistica per l’esecuzione dei lavori compresi nel PSN di titolarità dell’Istat, prima della validazione da parte dell’Istituto stesso.

Nello stesso anno, l’Istat ha pubblicato anche le *Linee Guida per la Qualità delle Statistiche del Sistema Statistico Nazionale* (Istat 2018), che enuncia principi e buone pratiche da applicare nei processi statistici basati su indagini dirette o su dati di fonti amministrative. Tali linee guida vengono adottate, tra il 2018 e il 2021, come riferimento per la conduzione da parte dell’Istat di un programma di valutazione della qualità, attraverso procedure di *audit* e autovalutazione, condotto sui processi statistici alla base delle statistiche europee prodotte dalle ONA. Tali iniziative hanno contribuito a consolidare la cultura della qualità e a rafforzare la responsabilità diffusa nella produzione statistica.

Nel 2020-2021 l’Istat ha rilanciato il proprio impegno strategico attraverso l’approvazione della *Nuova politica per la qualità della produzione statistica* (Istat 2021), che valorizza il ruolo del coordinamento del Sistan e promuove l’allineamento tra il quadro europeo e quello italiano. La nuova politica ha introdotto un insieme coerente di strumenti e documenti operativi, specificamente calibrati sul contesto nazionale ma in piena sintonia con le evoluzioni normative e metodologiche del SSE. A livello europeo, infatti, il *Codice delle statistiche europee* (Eurostat e SSE 2017) è stato rivisto e rafforzato nel 2017 proprio nella componente del coordinamento del Sistema statistico nazionale, attraverso l’introduzione del principio 1bis “Coordinamento e Cooperazione”; inoltre, nel 2019 il Comitato del Sistema statistico europeo ha approvato la versione 2.0 dell’ESS QAF (ESS 2019), che include metodi a dettaglio di ogni principio e indicatore del *Codice*.

L’Italia ha risposto a questa evoluzione del quadro europeo attraverso un aggiornamento del proprio sistema di riferimento. Nel 2022 è stata approvata la seconda edizione del *Codice italiano* (Sistan e Istat 2022), direttiva Comstat n. 12/2022 pubblicata nella GU n. 23 del 29 gennaio 2022 e, nel 2023, è stata pubblicata la *Guida per l’implementazione del Codice italiano* (Istat 2023) che costituisce il corrispettivo nazionale dell’ESS QAF europeo. Quest’ultimo documento rappresenta un vero e proprio manuale operativo per gli enti del Sistan, offrendo strumenti, esempi e raccomandazioni utili a tradurre nella prassi quotidiana i principi del *Codice italiano* con l’obiettivo di favorire una qualità sistemica, continua e partecipata.

1.3 Il Sistema statistico europeo

La disponibilità di statistiche sulla situazione economica, sociale e ambientale dell’UE è un presupposto necessario per il processo di integrazione europea. Le statistiche europee costituiscono un contributo essenziale allo sviluppo della capacità informativa per tutti i decisori e i cittadini europei per valutare l’avanzamento delle iniziative politiche europee; esse rappresentano pertanto un bene pubblico e costituiscono la base per il buon funzionamento della democrazia.



Un sistema efficiente e fondato su un chiaro impegno sulla qualità per la produzione di statistiche europee richiede una base giuridica stabile e trasparente in grado di garantire l'indipendenza, l'integrità e la responsabilità delle autorità statistiche. Allo stesso tempo, il sistema deve essere sufficientemente flessibile da consentire di rispondere alle sfide attuali e future, ad esempio per quanto riguarda le esigenze politiche in rapida evoluzione, i vincoli finanziari, la riduzione dell'onere per i rispondenti e gli sviluppi nelle tecnologie dell'informazione.

In tale senso, il SSE è stato istituito dalla legge statistica europea (regolamento (CE) n. 223/2009 del Parlamento europeo e del Consiglio relativo alle statistiche europee e sue successive modifiche). Esso svolge un ruolo strategico nello sviluppo, nella produzione e nella diffusione delle statistiche europee, contribuendo anche al rafforzamento della struttura di governance. Il SSE è un partenariato tra l'autorità statistica comunitaria (Eurostat, Commissione europea), gli Istituti nazionali di statistica (INS) e le ONA che contribuiscono allo sviluppo, alla produzione e alla diffusione delle statistiche europee.

La legge statistica europea stabilisce gli attori, i principi e le regole di base per il funzionamento del SSE; risponde a domande quali "chi dovrebbe fare cosa", "chi decide e come". Tuttavia, non specifica quali statistiche debbano essere prodotte; ciò viene stabilito nella legislazione settoriale. In questa direzione, la legge statistica europea evidenzia il ruolo di coordinamento degli INS, un maggiore utilizzo dei dati amministrativi per la modernizzazione del processo statistico, l'attuazione del *Codice* e un maggiore impegno affinché si creino le condizioni necessarie per attuare efficacemente i suoi principi. La questione della qualità e della credibilità delle statistiche ufficiali è diventata il punto centrale del dibattito sulla governance statistica e un elemento decisivo anche per la governance economica. La legge stabilisce norme volte a garantire l'indipendenza, l'integrità e la responsabilità delle autorità statistiche e ne disciplina la cooperazione con Eurostat e all'interno del SSE per la produzione di statistiche europee di qualità.

Nel 2024 la legge è stata rivista. Le modifiche rafforzano l'accesso alle fonti amministrative e alle nuove fonti di dati, incluso l'accesso ai dati detenuti da privati, con l'obiettivo di ridurre l'onere per i rispondenti e per le autorità stesse, aprendo gli orizzonti anche verso statistiche sperimentali.

Le statistiche europee sono sempre più disciplinate attraverso l'adozione di regolamenti, o altri atti legislativi, che sanciscono l'obbligatorietà della fornitura dei dati e stabiliscono gli standard di contenuto e trasmissione, rispondendo ai principi statistici e seguendo criteri di qualità.

L'impegno per rafforzare la qualità e la credibilità delle statistiche europee è sempre più incisivo, e le azioni intraprese negli ultimi anni, con la revisione del *Codice* e la sua attuazione, confermano l'importanza del ruolo del *Codice* anche in futuro, a supporto del rafforzamento del sistema statistico nel suo complesso. L'appartenenza del SSE rappresenta un valore per i paesi e ha un impatto diretto sul nostro quadro di riferimento sulla qualità e assetto istituzionale.

1.4 Le *Peer review* nel Sistema statistico europeo

Le *Peer review* rappresentano uno strumento fondamentale all'interno della strategia del SSE per garantire e migliorare la qualità delle statistiche prodotte a livello sia europeo sia nazionale. Questo meccanismo si basa su un processo di valutazione indipendente e trasparente, volto a verificare come i paesi membri attuino i principi del *Codice* (Eurostat e SSE 2017). L'obiettivo è rafforzare l'integrità, l'indipendenza professionale e la responsabilità delle istituzioni coinvolte, assicurando così gli utenti sull'impegno del sistema e rafforzando la loro fiducia nelle statistiche europee.

Le *Peer review* vengono condotte periodicamente, circa ogni cinque anni, seguendo una metodologia e strumenti condivisi a livello di SSE. Si basano su un approccio sistematico che combina autovalutazioni, analisi documentale e visite in loco nei paesi membri e in Eurostat.

Il processo di *Peer review* si articola in diverse fasi: innanzitutto, Eurostat, gli INS e le ONA compilano un questionario di autovalutazione (*Self-Assessment Questionnaire*, SAQ) basato sui principi e indicatori del *Codice*. Successivamente, i *team* di *Peer review* analizzano le informazioni raccolte e la documentazione di supporto, e pianificano una visita nel paese membro. Durante questa visita, si approfondiscono le aree di interesse e si verifica sul campo l'effettiva attuazione dei principi del *Codice*.

Al termine di questa fase, i membri dei team di *Peer review* preparano un rapporto che riassume i risultati dell'analisi, mettendo in evidenza i punti di forza, le criticità e formulando raccomandazioni per migliorare l'attuazione del *Codice*. Il paese, in risposta alle raccomandazioni, definisce un piano di azioni di miglioramento da attuare con tempistiche precise, che vengono monitorate annualmente da Eurostat. Questo rapporto, insieme alle azioni di miglioramento, viene pubblicato sia sul sito dell'INS, sia su quello di Eurostat, garantendo trasparenza e assunzione di responsabilità (Figura 1.2)⁵.

Figura 1.2 - *Peer review* in sintesi



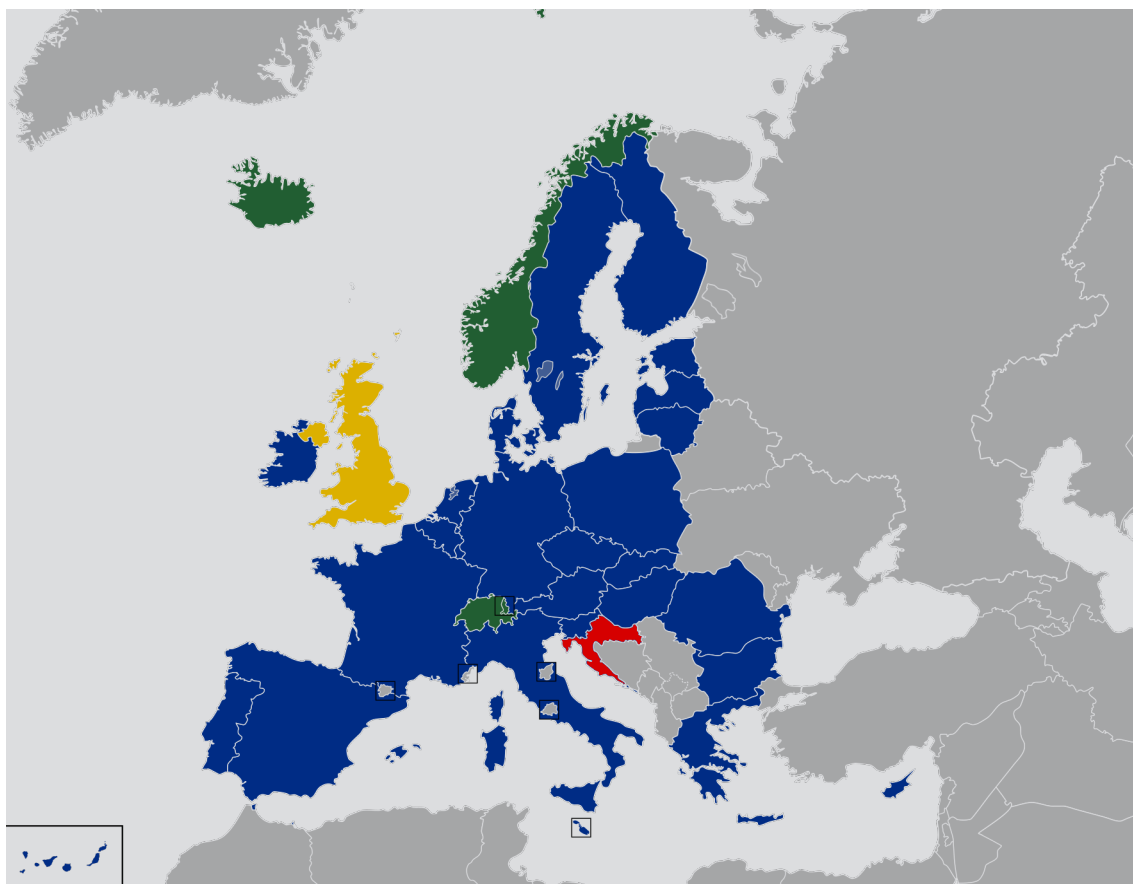
Fonte: Eurostat

5 cfr. Eurostat. <https://ec.europa.eu/eurostat/web/quality/peer-reviews>.

1.4.1 Evoluzione delle Peer review

Dall'adozione del *Codice* nel 2005, il SSE ha condotto tre cicli di *Peer review* nei paesi membri: il primo tra il 2006 e il 2008, il secondo tra il 2013 e il 2015, e il terzo tra il 2021 e il 2023. Il Regno Unito ha partecipato alle prime due *Peer review*, mentre la Croazia ha partecipato alle ultime due *Peer review*.

Figura 1.3 - Paesi sottoposti a *Peer review*. Anni 2006-2023 (a)



Fonte: Istat, Elaborazioni da Wikipedia, *European Health Insurance Card participating nations*

(a) In blu dal 2006 i paesi UE per tutto il periodo; in giallo il Regno Unito nella UE fino al 2015; in rosso la Croazia nella UE dal 2013; in verde i paesi EFTA.

Il primo ciclo (2006-2008) ha coinvolto i paesi dell'Unione europea (Figura 1.3, i paesi in blu e giallo), quelli aderenti all'Associazione europea di libero scambio (Aels, o in inglese EFTA – *European Free Trade Association*, ossia i paesi in verde) ed Eurostat. In questa fase iniziale, l'attenzione si è concentrata sul contesto istituzionale e sulle pratiche di diffusione delle statistiche (principi da 1 a 6 e 15 del *Codice*), limitandosi agli INS e al loro ruolo nel Sistema statistico nazionale. Dopo avere condotto due *Peer review* pilota all'interno del SSE, la prima *Peer review* ufficiale a regime è stata realizzata in Italia, presso l'Istat, nel mese di ottobre 2006. I *report* sui risultati, redatti alla fine della visita, hanno incluso anche le azioni di miglioramento riguardanti tutti i principi del Codice, la cui attuazione è stata monitorata annualmente da Eurostat (Eurostat 2006) (Prospetto 1.3)⁶.

⁶ Cfr. Eurostat. <https://ec.europa.eu/eurostat/web/quality/peer-reviews/first-round>.

1. Qualità nella statistica ufficiale: contesto interno, nazionale ed europeo

Prospetto 1.3 - Principali conclusioni del rapporto di *Peer review* per l'Italia del 2006

Primo ciclo di *Peer review* - report Italia: principali conclusioni

L'Italia ha un solido quadro giuridico che garantisce l'indipendenza professionale, il mandato per la raccolta dei dati e la protezione della riservatezza.

Controlli di qualità rigorosi, ma è ancora necessario implementare un approccio più orientato all'utente;

In generale, l'Istat dispone di un quadro legislativo adeguato in relazione all'imparzialità e all'obiettività. La principale sfida sembra essere lo sviluppo di una documentazione più sistematica e orientata all'utente, accessibile anche alla comunità internazionale (in inglese).

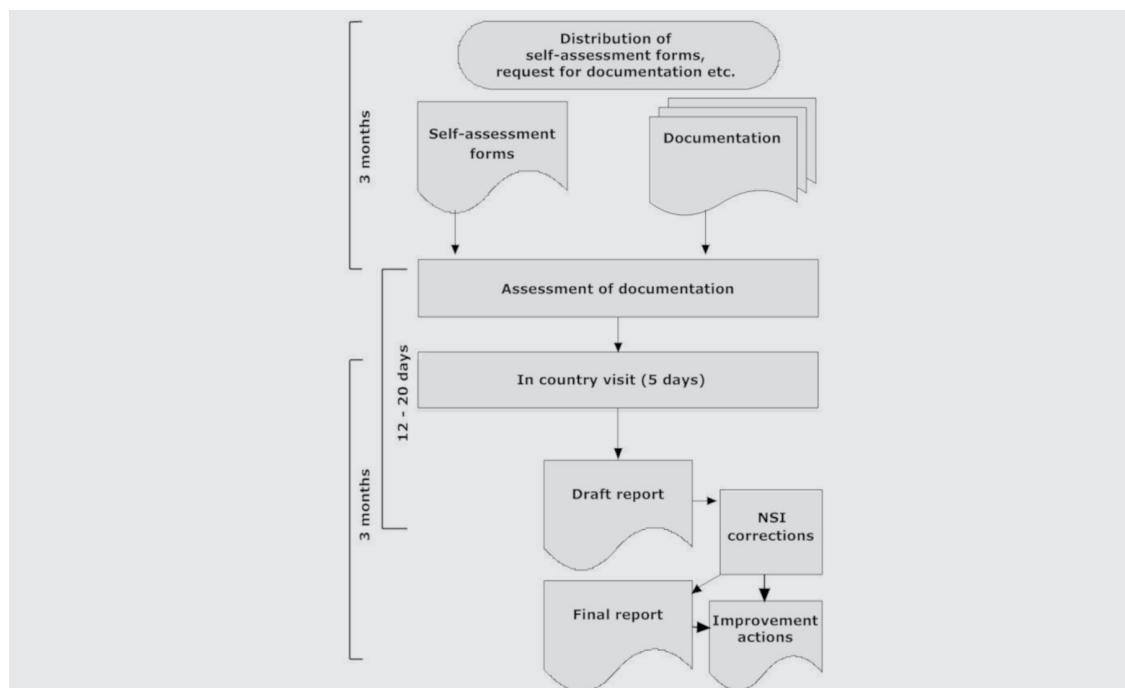
L'Istat ha un ampio programma di diffusione, ma sono necessari miglioramenti per soddisfare meglio le esigenze degli utenti riguardo a dati e metadati.

Fonte: Eurostat (traduzione degli autori)

Nel maggio 2006, Eurostat ha pubblicato il rapporto generale sui risultati della *Peer review* nei paesi. L'analisi ha evidenziato che il SSE presentava un elevato livello di conformità e solidità nelle aree relative all'indipendenza professionale, al mandato per la raccolta dei dati, alla riservatezza statistica e all'adozione di metodologie solide. Tuttavia, miglioramenti erano certamente possibili e necessari nell'area relativa al principio 4 "Impegno per la qualità".

Il secondo ciclo (2013-2015) ha avuto l'obiettivo di rafforzare i sistemi statistici a livello sia nazionale sia europeo, in modo da aumentare l'affidabilità e la credibilità delle statistiche europee e di mettere il SSE in condizione di affrontare le sfide future. L'adozione, nel maggio 2015, del regolamento che ha modificato per la prima volta il regolamento (CE) n. 223/2009, rafforzando in modo significativo l'indipendenza degli INS e il loro ruolo di coordinamento, ha messo in luce l'importanza di questo ciclo di *Peer review* quale strumento fondamentale per la modernizzazione del SSE (Prospetto 1.5). Anche questo ciclo ha coinvolto i paesi dell'UE (in questo frangente, inclusi anche della Croazia, in rosso nella Figura 1.3), quelli aderenti all'EFTA ed Eurostat, ma si è differenziato da quello precedente sotto molteplici aspetti.

Figura 1.4 - Fasi principali della *Peer review* 2013-2015 (approccio *audit*)



Fonte: European Statistical System

In primo luogo, è stato avviato un esame di tipo *audit* su tutti i 15 principi del *Codice*, il che significa che le risposte fornite ai questionari non sono state lasciate al semplice racconto o alla dichiarazione di buona volontà, ma sono state supportate da evidenze concrete e verificabili.

Questo approccio rigoroso ha l'obiettivo di garantire che le informazioni fornite siano basate su dati e documenti reali, aumentando così la trasparenza e l'affidabilità del processo di valutazione. In secondo luogo, sono state coinvolte, e sottoposte a valutazione, anche alcune ONA seppure non nella stessa misura degli INS (fino a tre ONA per paese) (Figura 1.4)⁷.

Prospetto 1.4 - Ulteriori differenze tra le *Peer review* del 2006 e del 2015

Ulteriori differenze del secondo ciclo di *Peer review* rispetto al precedente

- Gli INS hanno compilato altri due questionari, sul ruolo di coordinamento dei rispettivi Sistemi di statistica nazionali e sulla cooperazione e livello di integrazione del SSE.
- *Team* di *Peer review* composti da valutatori esterni agli INS del SSE.
- Sono state evidenziate le pratiche innovative nell'attuazione del *Codice*.
- Gli INS hanno fornito documentazione preliminare: descrizione del Sistema statistico nazionale, normativa nazionale sulla statistica, strategie, politiche e informazioni sulle ONA.

Fonte: Istat

Gli INS ed Eurostat, in risposta alle raccomandazioni dei revisori, hanno formulato complessivamente 929 azioni di miglioramento che sono state completate entro la fine del 2019.

Prospetto 1.5 - Principali conclusioni del rapporto di *Peer review* per l'Italia del 2015

Secondo ciclo di *Peer review* – report Italia: principali conclusioni

- L'Istat gode di un solido mandato legale che ne garantisce la professionalità e l'indipendenza;
- L'Istat ha costruito e mantenuto rapporti molto stretti con la comunità accademica;
- L'Istat ha realizzato un efficiente sistema di gestione della qualità, basato su una visione integrata della qualità, che va dall'applicazione delle definizioni a controlli di qualità dettagliati di procedure e prodotti;
- L'Istat si è impegnato nell'ampio utilizzo di dati amministrativi al fine di migliorare la propria efficienza e organizzare la propria produzione statistica.

Oltre agli evidenti punti di forza dimostrati dall'Istat, il *team* di *Peer review* ha individuato quattro tematiche in cui la conformità al *Codice delle statistiche europee* potrebbe essere migliorata: ridefinizione del Sistema statistico nazionale e rafforzamento del suo coordinamento, miglioramento dell'economicità e della politica delle risorse umane e proseguimento degli sforzi per soddisfare le esigenze degli utenti.

Fonte: Eurostat (traduzione degli autori)

I risultati hanno evidenziato diversi punti di forza. In particolare, sono emersi aspetti positivi riguardo la solida base giuridica che supporta l'indipendenza professionale, l'imparzialità e l'obiettività, ampiamente riconosciute e indiscusse, delle autorità statistiche. Tuttavia, riduzioni delle risorse umane e finanziarie sono emerse come aspetti di rischio per affrontare le sfide future (Eurostat 2015a) (Prospetto 1.6)⁸.

Il terzo ciclo (2021-2023) si riferisce alla più recente revisione del *Codice* che è stata realizzata nel 2017: rispecchia gli ultimi cambiamenti e innovazioni nella produzione di statistiche europee, quali le nuove fonti di dati, l'uso di nuove tecnologie e l'aggiornamento del quadro giuridico. In questa ultima versione, è stato inserito anche un nuovo principio (1bis) sul ruolo di coordinamento degli INS e la cooperazione inteso a rafforzare il contesto istituzionale, portando

⁷ Cfr. European Statistical System. <https://ec.europa.eu/eurostat/documents/64157/4372828/6-Guide-for-peer-reviewers.pdf>.

⁸ Cfr. Eurostat. <https://ec.europa.eu/eurostat/web/quality/peer-reviews/second-round>.

1. Qualità nella statistica ufficiale: contesto interno, nazionale ed europeo

i principi del *Codice* a 16. Questa versione rivista del *Codice* ha fatto da leva per l'avvio del terzo ciclo di *Peer review*, condotte negli stati membri dell'UE, paesi EFTA e presso Eurostat (con l'inserimento della Croazia nella UE dal 2013 e l'uscita del Regno Unito nel 2020; Figura 1.4). L'obiettivo è stato di migliorare la qualità e la fiducia nelle statistiche europee, valutando la conformità del SSE ai principi del *Codice* riveduto e formulando raccomandazioni orientate al futuro.

Tale ciclo ha visto, inoltre, un maggiore coinvolgimento delle ONA. La loro partecipazione è ritenuta molto importante: da un lato, perché favorisce una collaborazione più stretta all'interno del Sistema statistico nazionale, e dall'altro, perché aiuta le ONA a sensibilizzarsi sul loro ruolo nella produzione di statistiche europee di qualità, sottolineando l'importanza di garantire che le statistiche siano accurate e conformi ai principi del *Codice* (Prospetto 1.6).

La selezione delle ONA nei vari paesi si è basata su criteri definiti nella metodologia, che prevedeva la partecipazione da tre a sei ONA per paese, e stabiliva che la scelta finale delle ONA fosse responsabilità dell'INS, che poteva decidere il numero e i criteri di selezione.

Prospetto 1.6 - Principali differenze tra le *Peer review* del 2015 e del 2021

Principali differenze del terzo ciclo di *Peer review* rispetto al precedente

- Focus sui progressi rispetto al secondo ciclo.
- Il campo di analisi ha incluso gli INS e alcune ONA, ma con uno sguardo orientato al Sistema statistico nazionale.
- Il *team* composto da quattro esperti (interni ed esterni al SSE) uno di Eurostat con il ruolo di osservatore.
- Inserite domande aggiuntive per individuare pratiche innovative e orientate al futuro.
- Inserite domande sui punti di forza e di debolezza.
- Le tre aree del questionario (Istituzionale, Processi, Prodotti) hanno previsto domande sui progressi degli ultimi cinque anni e sui piani di azione attuali e futuri.
- Raccomandazioni orientate al futuro, suddivise in due categorie: quelle di importanza fondamentale per garantire la conformità al *Codice* (*compliance-relevant*) e quelle meno critiche, orientate al miglioramento (*improvement-related*).
- Richiesta dell'elenco e la descrizione di tutte le ONA.
- Sviluppata una campagna di comunicazione sulla *Peer review* a livello di SSE e declinata a livello di Sistema statistico nazionale.

Fonte: Istat

L'Istat ha considerato le *Peer review* un'importante opportunità per le ONA di valutare i propri processi produttivi, evidenziare punti di forza e criticità, e mostrare eventuali innovazioni adottate per rispondere meglio alle nuove esigenze di dati. Per questo motivo ha ritenuto fondamentale coinvolgere tutte le ONA nel processo, chiedendo loro di compilare il questionario di autovalutazione (SAQ *light*). Successivamente, l'Istat ha selezionato tra ministeri e altri enti quattro ONA a partecipare alla visita del *team* di *Peer review*.

Prospetto 1.7 - Principali conclusioni del rapporto di *Peer review* 2023 per l'Italia

Terzo ciclo di *Peer review* – report Italia: principali risultati

- L'Istat è un istituto statistico con una solida reputazione.
- È anche un ente moderno e innovativo: negli ultimi cinque anni ha investito significativamente nello sviluppo di sistemi e processi, favorendo la raccolta centralizzata dei dati e l'utilizzo di fonti amministrative.
- L'Istituto sta sperimentando l'uso di nuove fonti di dati.

Nel complesso, il Sistema statistico nazionale mostra un elevato livello di conformità al *Codice delle statistiche europee*. Tuttavia, il *team* di *Peer review* ha individuato 20 raccomandazioni: tre sono di rilevanza per la conformità, mentre le altre 17 riguardano il miglioramento complessivo nell'attuazione del *Codice*. Queste raccomandazioni sono suddivise in sei principali categorie:

- Rafforzare la governance, la pianificazione e il coordinamento
- Modernizzare il sistema statistico
- Ottimizzazione delle attività statistiche
- Migliorare il coinvolgimento degli utenti
- Ottimizzare l'uso delle risorse
- Migliorare la condivisione delle informazioni

Fonte: Eurostat (traduzione degli autori).



Per valutare efficacemente i sistemi statistici nazionali, è stata adottata una combinazione dei due approcci complementari: l'approccio di tipo *audit* e quello di revisione paritaria. Questa combinazione permette di sfruttare i punti di forza di entrambi i metodi, creando un processo di valutazione più completo e affidabile, integrando elementi di verifica rigorosa e di confronto collaborativo (Eurostat 2023a) (Prospetto 1.7)⁹.

A livello di SSE il terzo ciclo di *Peer review* ha dimostrato che, nel complesso, l'adesione al *Codice* è molto elevata. Sono state formulate 581 raccomandazioni (una media di 19 raccomandazioni per paese), di cui solo 68 rilevanti ai fini della conformità al *Codice* (*compliance-relevant*). Complessivamente sono state definite 1.724 azioni di miglioramento da parte degli INS. Sono stati conseguiti miglioramenti significativi rispetto al precedente ciclo di *Peer review*, soprattutto nei settori della governance, delle politiche per la qualità, della metodologia, del coordinamento del Sistema statistico nazionale e dell'accesso alle fonti di dati amministrative. Molte ONA sono diventate attori chiave nei Sistemi statistici nazionali. Tuttavia, saranno necessari ulteriori sforzi per attuare pienamente il ruolo di coordinamento degli INS e per garantire che le ONA aderiscano al *Codice* in modo sostenibile.

Sebbene la conformità al *Codice* sia stata generalmente alta, le *Peer review* hanno individuato aree di ulteriore miglioramento, in particolare in ambiti quali la revisione del quadro giuridico nazionale, la comunicazione e l'uso di nuove fonti di dati.

1.4.2 Impegno richiesto e attuazione delle azioni di miglioramento

Il Sistema statistico europeo ha l'obiettivo di garantire che le statistiche siano affidabili e di alta qualità in tutti i paesi membri. Tuttavia, questo processo comporta degli oneri sia per gli INS sia per le ONA.

Uno degli oneri principali riguarda la necessità di valutare attentamente e riflettere sulle pratiche statistiche attuali, in relazione ai principi stabiliti dal *Codice*. Tale autovalutazione richiede un significativo impegno nella raccolta, analisi e documentazione delle procedure esistenti, dei metodi di raccolta dati e delle misure adottate per garantire la qualità. Inoltre, la preparazione di risposte dettagliate ai questionari di autovalutazione e la presentazione della relativa documentazione comportano un uso intenso di risorse.

Le istituzioni coinvolte devono inoltre affrontare un aumento del carico di lavoro legato all'attuazione delle azioni di miglioramento identificate durante la *Peer review*. Ciò include lo sviluppo e l'attuazione di piani di azione. Questo rappresenta l'onere più rilevante, in quanto tali iniziative spesso richiedono investimenti in risorse umane, formazione del personale e modifiche organizzative, elementi difficili da gestire con budget e risorse limitate.

Va considerato anche l'impegno amministrativo continuo legato al monitoraggio e alla rendicontazione annuale dei progressi a Eurostat. Questo processo richiede personale dedicato e si estende per diversi anni, risultando particolarmente impegnativo, poiché prevede l'attuazione simultanea di più azioni.

Nonostante queste difficoltà, è importante sottolineare che il processo di *Peer review* offre anche importanti opportunità. Tra queste vi sono il rafforzamento delle capacità istituzionali, un aumento della credibilità e un migliore allineamento agli standard europei. L'impegno nel superare le criticità individuate conduce a sistemi statistici più robusti, dati più affidabili e una maggiore fiducia da parte degli utenti.

9 Cfr. Eurostat. <https://ec.europa.eu/eurostat/web/quality/peer-reviews/third-round>.

In sintesi, sebbene il processo di *Peer review* imponga agli INS e alle ONA un carico di lavoro significativo – dalla fase di valutazione iniziale fino all’implementazione e al monitoraggio delle azioni di miglioramento – rappresenta una sfida fondamentale per il raggiungimento dell’eccellenza nelle statistiche ufficiali.

Gli sforzi e le risorse investiti sono compensati dai benefici a lungo termine: dati di qualità superiore, maggiore resilienza istituzionale e fiducia accresciuta nelle statistiche nazionali ed europee.

1.4.3 Benefici della *Peer review*, esperienza dell’Istat e sfide future

Il terzo ciclo di *Peer review* 2021-2023 sull’attuazione del *Codice* si è inserito nelle attività degli INS in un momento in cui il contesto esterno stava attraversando rapidi cambiamenti, grazie all’emergere di un nuovo ecosistema di dati e ai processi di modernizzazione in corso. La crescente disponibilità di dati presentava ai produttori di statistiche ufficiali nuove sfide, in particolare per l’accesso e l’uso di fonti di dati innovative, l’accesso a dati privati e l’integrazione di informazioni provenienti da più fonti.

Le *Peer review* rappresentano per l’INS un’importante opportunità per mettere in luce i progressi fatti nel rispondere a queste sfide, a livello sia istituzionale sia di processi statistici e di prodotti. Si tratta di un momento strategico in cui si riconosce il ruolo fondamentale della statistica ufficiale nel migliorare le capacità statistiche di un paese, così come il ruolo dell’INS nella società. Questo esercizio permette di sistematizzare le informazioni, verificare i bisogni attuali e futuri di risorse, e rafforzare la cooperazione e il coordinamento con le ONA, oltre che con tutti gli *stakeholder* rilevanti del Sistema statistico nazionale. L’esperienza del terzo ciclo di *Peer review* in Italia ha fornito una visione dell’impatto di questo processo sulla qualità delle attività dell’Istat, dei progressi compiuti in questi ultimi anni e delle sfide future.

Le *Peer review* si sono rivelate un’esperienza formativa per l’Istat e le ONA. Innanzitutto, ha rafforzato il ruolo dell’Istat come punto di riferimento nel Sistan, soprattutto considerando i cambiamenti avvenuti rispetto al ciclo precedente. L’Istat ha dimostrato, inoltre, un grande impegno nell’innovazione e nell’approfondimento metodologico e statistico. Spesso, però, queste eccellenze non vengono comunicate sufficientemente all’esterno, se non in ambiti specializzati. L’aspetto positivo di questa esperienza è che, grazie alle *Peer review*, sono emerse tutte le potenzialità delle professionalità e delle procedure interne, valorizzando il lavoro svolto e favorendo la condivisione non solo a livello nazionale, ma anche internazionale ed europeo.

Lo stesso vale per le ONA, che spesso non sono pienamente consapevoli della qualità del proprio lavoro e delle buone pratiche adottate. Condividere queste esperienze può essere di grande supporto anche ad altri enti del Sistan.

Le *Peer review* hanno rivelato che i membri del SSE sono costantemente volti alle innovazioni per rispondere alle esigenze degli utenti e puntano a una maggiore efficienza, anche in termini di costi. Tuttavia, è emerso chiaramente che le risorse disponibili per affrontare nuovi compiti e sfruttare le opportunità tecnologiche sono ancora limitate. Le analisi delle *Peer review* sottolineano, infatti, l’urgenza di aumentare gli investimenti in progetti di innovazione e di cogliere appieno i vantaggi delle nuove tecnologie, con l’obiettivo di continuare a offrire dati e servizi statistici di elevata qualità.

Le *Peer review* hanno anche dimostrato che il SSE adotta un approccio lungimirante, volto a migliorare continuamente le proprie prestazioni. Gli effetti delle azioni di miglioramento si vedranno nei prossimi anni, confidando nel loro contributo a mantenere il SSE rilevante e competitivo nel nuovo ecosistema dei dati.



Tra le principali sfide emergenti, vi è la revisione del *Codice delle statistiche europee*. Il terzo ciclo di *Peer review* ha avuto anche l'obiettivo di rilevare elementi innovativi che possono essere fondamentali per una futura modifica del *Codice* stesso. Questo processo di aggiornamento è importante per garantire che il *Codice* rimanga al passo con le evoluzioni del sistema statistico e delle esigenze degli utenti.

Un aspetto cruciale del *Codice* è che dovrà anche riflettere le nuove esigenze di accesso ai dati, soprattutto quelli detenuti da soggetti privati e considerare la regolamentazione delle statistiche sperimentali. Il concetto di *data stewardship* assume, inoltre, crescente rilevanza, configurandosi come una gestione responsabile e strategica dei dati, necessaria per affrontare la complessità dell'attuale ecosistema informativo.

A seguito della revisione del regolamento (CE) 223/2009, infine, sarà fondamentale aggiornare le regole e le pratiche di gestione dei dati, rendendole più flessibili e adatte alle sfide attuali. In conclusione, il futuro richiede un *Codice* più innovativo, aperto e in grado di rispondere alle nuove dinamiche del mondo dei dati, garantendo sempre qualità, affidabilità e accessibilità.

2. STRUMENTI PER LA QUALITÀ¹

Al fine di perseguire la qualità, gli INS si sono dotati di strumenti che sono in grado di monitorare e descrivere gli input, i processi e i prodotti statistici che ne conseguono, facilitando così anche l'integrazione dei dati, la standardizzazione dei metodi e degli strumenti e la valutazione della qualità dei processi stessi. Negli ultimi anni l'Istat ha investito su diversi di questi strumenti, tra cui i più rilevanti sono i seguenti:

1. **Metastat.** È un nuovo sistema di metadati, sia referenziali sia strutturali, caratterizzato dall'attenzione agli aspetti semantici (utili a rendere i metadati chiari e non ambigui) e corredato da indicatori di qualità. È in fase avanzata di realizzazione, integrerà e sostituirà sistemi già esistenti;
2. **Standardizzazione dei metodi e degli strumenti.** L'Istat raccoglie e rende disponibili da tempo metodi e strumenti software sviluppati attraverso il software *R* e in generale con software *open source*; di recente, ha definito una chiara governance per la costruzione, l'adozione e il mantenimento di software *open source* per la produzione statistica ufficiale;
3. **Quality Report Card for Administrative Data (QRCA).** È un sistema di documentazione delle fonti amministrative usate per condurre processi di produzione statistica, che include indicatori di qualità.
4. **Dossier Qualità.** È uno strumento agile per facilitare l'accesso ai documenti sulla qualità delle indagini già disponibili presso l'Istat.

2.1 Il nuovo sistema dei metadati Metastat

2.1.1 Esigenza di un nuovo sistema dei metadati

Nel tempo, l'Istat ha affrontato il tema dei metadati, sia referenziali sia strutturali, dotandosi di diversi strumenti dedicati alla descrizione dei processi (metadati referenziali), alla documentazione e interpretazione dei dati prodotti (metadati strutturali) o agli aspetti semantici (tipici di un glossario). Alcuni di questi strumenti sono stati dei punti di riferimento anche in ambito internazionale. I motivi principali per la progettazione del nuovo sistema di metadati, denominato Metastat, sono di diverso tipo: da un lato, il sistema in uso dedicato alla descrizione dei processi, progettato più di 25 anni fa, è ormai obsoleto sia rispetto agli aspetti tecnico-informatici (con difficoltà di manutenzione o di modifica) sia rispetto all'evolversi dei modelli, degli standard internazionali e del processo di produzione statistica; dall'altro, si coglie l'opportunità per avviare una semplificazione delle infrastrutture a supporto della produzione statistica, ad esempio facendo sì che metadati referenziali, strutturali e glossario siano trattati in un'infrastruttura unica, senza ridondanze, utilizzando un insieme di termini comune.

¹ Il Capitolo 2 è stato redatto da: Mauro Scanu (paragrafo 2.1), Gabriele Ascari e Tiziana Pichiorri (paragrafi 2.2 e 2.2.2), Marcello D'Orazio (paragrafo 2.2.1), Grazia Di Bella (paragrafo 2.3), Francesca Rossetti e Giorgia Simeoni (paragrafo 2.4).

Il miglioramento continuo dei processi di produzione statistica, realizzato nel rispetto dei principi e degli indicatori del *Codice*, è stato uno degli elementi ispiratori del nuovo sistema. In particolare, l'attenzione è stata focalizzata sugli indicatori riportati nel Prospetto 2.1 (per informazioni generali su principi e indicatori del *Codice* cfr. paragrafo 1.1), tutti pertinenti ai metadati, che indicano quanto sia stretta l'interrelazione tra qualità e metadati.

Nel Prospetto 2.1 (Eurostat e SSE 2017) sono evidenziate alcune delle funzioni che Metastat deve svolgere. Più in dettaglio, il sistema deve consentire:

- il monitoraggio dei processi, documentando ogni processo nelle fonti e nei metodi e descrivendo gli output di ogni fase con gli opportuni indicatori di qualità a corredo;
- il confronto tra processi, sia per le singole fasi in cui è possibile scomporli sia rispetto agli output di ogni fase, anche al fine di modificare un processo nelle fonti, nei metodi o nei metadati, riutilizzando quanto considerato utile da altri processi;
- l'armonizzazione di concetti, definizioni e classificazioni, raccordandoli agli standard europei, in modo da favorire l'integrazione o il confronto tra dati;
- la comprensione dei dati da parte di un ampio pubblico; a tale fine si ritiene fondamentale corredare i metadati con definizioni chiare e non ambigue, in modo da consentire a tutti gli utenti, a partire da quelli interni, di interpretare i dati in modo preciso e univoco.

Prospetto 2.1 - Indicatori del *Codice delle statistiche europee* connessi ai metadati

Indicatore	Descrizione
Indicatore 4.3	La qualità dei prodotti è regolarmente monitorata e valutata tenendo conto dei possibili compromessi; essa è oggetto di relazioni elaborate in base ai criteri di qualità delle statistiche europee.
Indicatore 6.4	Le informazioni sulle fonti di dati nonché sui metodi e sulle procedure applicati sono disponibili a tutti.
Indicatore 7.2	Sono in atto procedure volte a garantire che l'autorità statistica nel suo insieme applichi in maniera coerente concetti, definizioni, classificazioni standard nonché altri tipi di standard.
Indicatore 7.4	Tra i sistemi nazionali di classificazione e i corrispondenti sistemi a livello europeo esiste un preciso raccordo.
Indicatore 8.3	I processi statistici sono regolarmente monitorati e, qualora necessario, riveduti.
Indicatore 8.4	I metadati relativi ai processi statistici sono gestiti durante l'intero processo statistico e opportunamente diffusi.
Indicatore 15.5	I metadati relativi ai prodotti sono gestiti e diffusi dall'autorità statistica conformemente agli standard europei.
Indicatore 15.7	Gli utilizzatori sono informati costantemente sulla qualità dei prodotti statistici in accordo con i criteri di qualità delle statistiche europee.

Fonte: Eurostat, Sistema statistico europeo

Da quanto già esposto, risulta evidente quanto sia importante l'aspetto semantico, ossia integrare il glossario opportunamente rafforzato sulla semantica ai sistemi di metadati referenziali e strutturali. Il sistema Metastat sarà quindi caratterizzato da due componenti fondamentali: l'anagrafica dei processi statistici e il vocabolario controllato. Ogni metadato statistico avrà un processo statistico che lo gestisce e sarà usato da più processi per descrivere i dati nei vari output, non solo finali ma anche intermedi (Figura 2.1, Brunini e Recchini 2024).

2. Strumenti per la qualità

Figura 2.1 - Esempio di interrelazione tra metadati strutturali, referenziali e i termini del glossario

Fonte: Istat

2.1.2 Metastat: un sistema, tre moduli chiave

I moduli chiave del sistema sono quindi tre, rappresentati rispettivamente dalla raccolta terminologica, l'impiego dei termini per descrivere i processi di produzione statistica (metadati referenziali) e i loro output (metadati strutturali). Le descrizioni dei tre moduli si basano sulle considerazioni tratte da: Brunini (2023a, 2023b), Brunini e D'Alessandro (2025), D'Alessandro *et al.* (2025) e Recchini (2023) per il modulo sulla terminologia; Budano e Simeoni (2024), e Casagrande *et al.* (2024) per il modulo dei metadati referenziali; Signore *et al.* (2015) e Casagrande e Olivieri (2024) per il modulo sui metadati strutturali.

Figura 2.2 - Esempio di due termini, costituiti da lemma e definizione, che costituiscono una polisemia (stesso lemma, diversa definizione)

METASTAT - Sistema Informativo per la Gestione dei Metadati				
Home	Consultazione	Creazione e Gestione	Supervisione	
Consultazione \ Metadati semantici				
Cerca: Lemma Vai Azioni Filtri Excel				
Lemma				
Lemma	Definizione		Acronimo	Processo Responsabile
① Disoccupati	Le persone non occupate tra i 15 e i 74 anni che: - hanno effettuato almeno un'azione attiva di ricerca di lavoro nelle quattro settimane che precedono la settimana a cui le informazioni sono riferite e sono disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive; oppure - inizieranno un lavoro entro tre mesi dalla settimana a cui le informazioni sono riferite e sarebbero disponibili a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive, qualora fosse possibile anticipare l'inizio del lavoro.			Rilevazione sulle forze di lavoro
① Disoccupati (censimento della popolazione e delle abitazioni 2011)	Le persone non occupate di 15 anni e più che: - hanno effettuato almeno un'azione attiva di ricerca di lavoro nelle quattro settimane (11 settembre-8 ottobre) che precedono la data del censimento (9 ottobre 2011) e sono disponibili a lavorare (o ad avviare una attività autonoma) entro le due settimane successive; oppure che - inizieranno un lavoro entro tre mesi dalla settimana a cui le informazioni sono riferite e sarebbero disponibili a lavorare (o ad avviare una attività autonoma) entro le due settimane successive, qualora fosse possibile anticipare l'inizio del lavoro.			Censimento generale della popolazione e delle abitazioni

Fonte: Istat



Raccolta terminologica. Il modulo sulla terminologia ha come obiettivo la gestione semantica dell'insieme di termini in uso nella statistica ufficiale italiana. I termini costituiscono l'unità minimale utile per la descrizione di processi e dati. La raccolta terminologica include in sé un vero e proprio vocabolario, nel senso che questo insieme di termini può comunque esistere a prescindere dall'esistenza di dati e processi. Al suo interno trovano spazio tutte le locuzioni tramite le quali si costruisce il linguaggio complessivo in uso nella statistica ufficiale. L'unità minimale utilizzata per la raccolta terminologica in Metastat è rappresentata dal termine, dotato di un lemma, e della corrispondente definizione (Figura 2.2, Brunini e Recchini 2024). Questa coppia di elementi è inscindibile e il modo in cui tali elementi si associano descrive la ricchezza e la varietà della raccolta terminologica in uso all'Istat. In particolare, al suo interno possono esistere polisemie, cioè termini diversi, caratterizzati dallo stesso lemma, ma da diversa definizione (per esempio, al primo gennaio 2025 al lemma "fatturato" sono assegnate tre definizioni diverse in tre regolamenti europei, che devono essere rispettati da tre diversi processi di produzione dell'Istat), o sinonimie, cioè termini sostanzialmente uguali nel significato (e quindi con identica definizione) anche se con lemmi diversi (per esempio, i lemmi "persona" e "individuo", entrambi ampiamente rappresentati nell'ambito delle statistiche ufficiali italiane con lo stesso significato). Il governo di questi aspetti semantici è di primaria importanza in un istituto di statistica per garantire la chiarezza delle informazioni a corredo dei dati e la loro univocità: una polisemia può riferirsi a statistiche basate su gruppi di unità statistiche sostanzialmente diverse e/o a variabili osservate su unità diverse, giustificando così una diversità nei numeri statistici a parità di tempo, nonostante l'uguaglianza nei lemmi.

Oltre alle già descritte relazioni di polisemia e sinonimia, i termini possono essere collegati tra loro in quanto termini sovraordinati (*broader term* in inglese, ad esempio le regioni includono le province) o viceversa subordinati (*narrower term* in inglese). Entrambe le relazioni sono fondamentali, ad esempio, nel contesto delle classificazioni (per maggiori approfondimenti cfr. Brunini 2023a e 2023b; Recchini 2023).

La raccolta terminologica rappresenta quindi una grandissima novità nell'ambito della statistica ufficiale in quanto diventa lo strumento di governo tramite il quale mettere in relazione termini usati da processi e temi statistici totalmente diversi.

Metadati referenziali. Il modulo sui metadati referenziali ha l'obiettivo di descrivere il processo produttivo statistico e le sue caratteristiche (contenuto informativo, input e output, fasi, sottoprocessi e metodi adottati), oltre alle informazioni sulla qualità, in termini di attività di prevenzione, monitoraggio e valutazione degli errori campionari e non campionari, di informazioni rilevanti sulle diverse dimensioni della qualità e di indicatori di qualità standard orientati al processo e al prodotto. Integrando i metadati referenziali con i necessari metadati strutturali, possono poi derivare tutte le reportistiche necessarie per accompagnare i dati, ad esempio i *quality report* per Eurostat e le note metodologiche che accompagnano i microdati messi a disposizione degli utenti.

Dall'inizio degli anni Duemila, l'Istat si è dotato di un sistema per i metadati referenziali e la documentazione della qualità ricco e articolato, Sidi-SIQual, con lo scopo di monitorare la qualità dei processi di produzione statistica e consentire la loro valutazione.

Sidi-SIQual (Brancato *et al.* 2004, Brancato *et al.* 2006) è infatti attualmente il sistema informativo ufficiale dell'Istat per la documentazione dei metadati referenziali e degli indicatori di qualità dei processi statistici. Operativo dal 2001, il sistema è orientato alla qualità e documenta gli aspetti, sia qualitativi sia quantitativi, dei processi statistici. È composto da due sottosistemi: Sidi, sistema di input, utilizzato internamente all'Istat per inserire le informazioni, e SIQual, sistema per la consultazione, disponibile in due versioni, una pubblica

sul sito Istat (<https://siqua.istat.it>) e una interna, in cui sono disponibili gli indicatori di qualità. Il sistema descrive obiettivi e contenuti dei processi statistici dell'Istat, e come viene svolto ogni processo, attraverso un modello organizzato in fasi e operazioni che può essere mappato sul *Generic Statistical Business Process Model* (Brancato e Simeoni 2012), sebbene risulti meno flessibile di quest'ultimo. Sono ovviamente documentate anche le azioni di controllo della qualità attuate nei processi, così come gli input e gli output dei processi, fornendo ad esempio collegamenti a comunicati stampa e documentazione rilevante (questionari, regolamenti). Nella versione interna sono disponibili indicatori di qualità relativi alle dimensioni della qualità come la tempestività, la comparabilità e l'accuratezza, tra cui la maggior parte degli indicatori di qualità standard del Sistema statistico europeo, in particolare quelli richiesti nei *quality report*. È possibile infatti attraverso Sidi produrre anche i *quality report* da trasmettere a Eurostat (Simeoni 2013) secondo lo standard *Statistical Data and Metadata eXchange* (SDMX). Tuttavia, Sidi-SIQual, oltre a essere tecnologicamente obsoleto, ha una struttura rigida delle fasi del processo statistico, con un ordine predefinito delle operazioni (ad esempio l'integrazione precedente al controllo e alla correzione) e non consente di descrivere diversi flussi paralleli nello stesso processo.

Oltre a sostituire Sidi-SIQual, Metastat è chiamato a svolgere il ruolo di "anagrafe dei processi statistici" condotti dall'Istat. Il sistema è adatto a rappresentare le indagini tradizionali, ma risulta meno efficace per i processi statistici innovativi, ad esempio i processi statistici multi-fonte o sperimentali. Tale problematica è affrontata dal nuovo sistema Metastat.

Il sistema mantiene quindi un elenco costantemente aggiornato dei processi statistici attivi, con alcune informazioni di base, e conserva le informazioni relative ai processi sospesi o cessati. Si osserva che l'Istat, a oggi, non dispone di un'anagrafe dei processi statistici, con conseguenti discrepanze tra i diversi sistemi. In generale, i lavori statistici di interesse pubblico sono inseriti nel Programma statistico nazionale (PSN) in fase di pianificazione, ma non è detto che poi vengano effettuati. Allo stesso tempo, non tutti i processi statistici svolti dall'Istat sono nel PSN. I processi statistici effettivamente svolti sono documentati in termini di metadati referenziali e qualità nel sistema di documentazione attualmente disponibile, Sidi-SIQual, ma solo al termine del ciclo produttivo. Pertanto, neanche Sidi-SIQual rappresenta un sistema di riferimento per conoscere i processi attivi a un dato istante di tempo poiché ha obiettivi diversi e può risultare parziale per motivi di sfasamento temporale o di ambito di interesse. Questa nuova funzione di Metastat, l'anagrafe dei processi statistici, è disponibile in diverse forme per gli altri sistemi e per il personale dell'Istat. L'inserimento di un processo in Metastat consente la generazione dei relativi metadati strutturali e referenziali all'interno del sistema, semplificando l'interazione tra utente e sistemi di supporto. L'anagrafe aggiornata correntemente prenderà avvio con il rilascio in produzione di Metastat. Le informazioni già disponibili in Sidi-SIQual sui processi statistici, sebbene incomplete ai fini di Metastat, vengono migrate, sia perché la sostituzione e la presa in carico delle funzionalità di Sidi-SIQual rappresentano uno degli obiettivi di Metastat, sia per ridurre l'onere di documentazione dei processi già presenti. La parte del sistema Metastat dedicata ai metadati referenziali è disegnata con una struttura modulare, in modo da implementare in tempi successivi le sue diverse componenti:

- modulo relativo alle informazioni generali e ai contenuti principali di un processo statistico;
- modulo relativo alla descrizione degli input (fonti) e degli output (prodotti) di ogni processo statistico;
- modulo relativo alle fasi e ai sottoprocessi di un processo statistico, anch'esso corredato delle indicazioni sugli input, gli output e il processo di elaborazione di ogni fase o sottoprocesso (inclusi gli indicatori di qualità), capace quindi di descrivere in dettaglio come è stato condotto un processo statistico.

Metadati strutturali. I metadati strutturali sono essenziali per comprendere il significato delle informazioni statistiche, soprattutto quando organizzate in forma strutturata, dalle semplici tabelle di dati fino agli ipercubi presenti nei *data warehouse*. Essi forniscono informazioni sulla disposizione gerarchica o sequenziale dei dati, inclusi il formato dei file, i tipi di dato e le relazioni tra le diverse componenti informative. Questo tipo di metadati è quindi fondamentale per comprendere l'organizzazione dei dati e per facilitarne il riutilizzo e l'integrazione.

In Metastat si è scelto di rappresentare le informazioni tramite metadati non solo per descrivere il contenuto delle tabelle, ma anche per facilitarne l'uso e il riuso a fini statistici. I dati devono quindi essere corredati da metadati che facilitino l'individuazione di una popolazione statistica, ossia un insieme di unità statistiche, su cui sono rilevate una o più variabili (chiamate così perché variano da unità a unità della popolazione). Il compito dello statistico è di sintetizzare le informazioni estremamente dettagliate delle variabili osservate su tutte le unità presenti nel *dataset*, riferite all'intera popolazione nel caso di una rilevazione esaustiva oppure a una sua porzione rappresentativa nel caso di un campione probabilistico, producendo un dato aggregato calcolato o stimato per l'intera popolazione. Per calcolare tale aggregato si fa riferimento a un metodo che deve rientrare tra quelli standard disponibili a livello di metadati referenziali. Le variabili possono essere di diverso tipo, a seconda del modo in cui si associano alle unità. Se l'associazione tra variabile e unità viene fatta in base a un numero finito e limitato di opzioni, la variabile è chiamata "categoriale", le opzioni associabili alle unità si chiamano "modalità" o "categorie", tra loro mutuamente esclusive (incompatibili) e contemporaneamente esaustive per ogni unità della popolazione (universali). L'insieme delle modalità così fatto prende il nome di classificazione e ogni unità può essere associata esclusivamente a una delle modalità in classificazione. Le restanti variabili si definiscono "descrivibili", ossia è possibile stabilire una regola per rappresentare il modo in cui la variabile si associa alle unità della popolazione. In questo gruppo rivestono particolare rilevanza le variabili "numeriche", in quanto è possibile applicarvi metodi particolarmente informativi (ad esempio, le medie). I concetti fin qui citati (unità, popolazione, variabile, classificazione, aggregato, metodo, eccetera) sono le componenti per la descrizione completa, sintetica ed esaustiva dei dati organizzati in forma tabellare.

I metadati strutturali, corredati da lemma e definizione, fanno parte della raccolta terminologica e sono associati ai processi, in quanto ne supportano la definizione o sono utilizzati per la descrizione dei dati prodotti. Il fatto che Metastat gestisca congiuntamente la raccolta terminologica e i metadati strutturali e referenziali rappresenta un elemento centrale del sistema, volto a chiarire il legame tra ambiti finora non connessi e a rendere disponibile all'utente, sia quello che li definisce sia quello che li utilizza, un'unica interfaccia semplificata per la gestione integrata di tali componenti.

2.1.3 Nuovo sistema dei metadati e standard internazionali

Uno dei compiti principali affidati al nuovo sistema dei metadati è il rispetto degli standard internazionali, strumento essenziale per garantire non solo l'allineamento e la coerenza, ma anche l'interoperabilità con i dati e i metadati di altre istituzioni nazionali, internazionali e sovranazionali. In principio, l'attenzione fu posta in particolare allo standard SDMX, requisito per la comunicazione dei dati e metadati verso Eurostat (Signore *et al.* 2015). Di seguito si citano i principali standard di riferimento per il sistema.

ISO 1087:2019 e ISO 25964:2013. Gli aspetti fondanti la semantica e la terminologia sono stati trattati in due documenti dell'*International Organization for Standardization* (ISO). Il primo, ISO 1087:2019, *Terminology work and terminology science* (ISO 2019), fornisce una descrizione sistematica dei concetti del lavoro terminologico, chiarendo l'uso dei termini in questo ambito. Brunini (2023a) definisce l'impianto concettuale per la terminologia in Metastat coerentemente a ISO 1087:2019, inserendo come ulteriore aspetto il fatto che un termine è specificato dalla coppia fornita dal lemma e dalla definizione. La concettualizzazione presente nella ISO 25964:2013, *Information and documentation - Thesauri and interoperability with other vocabularies* (ISO 2013), è stata invece utilizzata per documentare correttamente le connessioni semantiche. Si noti che le ISO non sono definite con specifico riferimento ai processi statistici e Metastat è il primo esempio di una loro esplicita applicazione nell'ambito dei sistemi di metadati statistici.

Generic Statistical Business Process Model (GSBPM). Questo standard descrive e definisce l'insieme di processi funzionali alla produzione di statistiche ufficiali (UNECE 2025). In quanto tale, il modello GSBPM fornisce il quadro di riferimento standard per dettagliare e descrivere i processi di produzione statistica. La sua definizione è posteriore allo sviluppo del Sistema Istat di documentazione delle indagini (Sidi). Il nuovo sistema dei metadati Metastat rappresenta l'opportunità per allineare i sistemi dell'Istat all'impianto concettuale del GSBPM, applicandolo ed estendendolo quando necessario, anche ai processi non tradizionali (cfr. paragrafo 3.2). Il modello dettaglia il processo in otto fasi ulteriormente distinte in sottoprocessi. I sottoprocessi GSBPM, per quanto elementi di dettaglio, possono a loro volta essere complessi e la loro descrizione non consente di rappresentare fedelmente un processo statistico già condotto, ossia i singoli passi di esecuzione che trasformano ogni input (di qualsiasi natura) in un output. A tale fine, Metastat ha l'obiettivo di applicare non solo lo standard GSBPM, ma anche, congiuntamente, lo standard *Generic Statistical Information Model* (GSIM) per l'impianto concettuale relativo al processo (UNECE 2024). Tale attività è stata favorita dalla partecipazione dell'Istat al Gruppo di lavoro UNECE relativo all'utilizzo congiunto dei due modelli (Rizzolo *et al.* 2021).

Generic Statistical Information Model (GSIM). Se il modello GSBPM è uno standard per i processi statistici, il GSIM (UNECE 2024) si pone come standard per la descrizione degli input e degli output dei passi di cui si compone ogni processo statistico. A tale scopo, questo modello è complementare al GSBPM. Da questo modello, Metastat ha rilevato l'impianto concettuale tradizionale statistico (ossia i concetti di unità, variabile, popolazione statistica, eccetera). Tramite questi concetti diventano espliciti i significati dei dati nelle tabelle (o strutture dati) che vengono prodotte durante un processo, arricchendo dove necessario l'impianto concettuale con ulteriori elementi utili alla descrizione dei dati, come ad esempio la formalizzazione di cosa si misura nelle tavole di macrodati statistici (Recchini 2024). Da questo punto di vista, il GSIM può essere considerato come il modello ispiratore dell'impianto concettuale di Metastat per i metadati strutturali. Metastat ha inoltre utilizzato alcune classi del gruppo *business* del GSIM (per esempio, quelli di *process step*, *process input*, *process output*, eccetera; la Figura 2.1 illustra un esempio di *unit type*, concetto ripreso da GSIM per rappresentare le unità statistiche dal punto di vista concettuale) per descrivere gli elementi di ogni passo dei sottoprocessi definiti nel GSBPM. Se il GSBPM dettaglia la sequenza di fasi e sottoprocessi di un generico processo statistico, il GSIM definisce in modo generalizzato gli elementi di un qualsiasi processo a un qualsiasi livello di dettaglio.

Si noti che il GSIM è stato preso in considerazione come impianto concettuale di riferimento standard, ma che nell'implementazione del sistema Metastat vi sono stati diversi adattamenti e sono stati formalizzati concetti e relazioni che nel GSIM non sono disponibili. Ad esempio, il concetto di misura in un *dimensional dataset* (ossia nella struttura dati di un ipercubo di macrodati statistici) è stato declinato in modo differente a seconda che l'obiettivo dell'aggregato sia un calcolo su una distribuzione statistica (relativa quindi solo a una popolazione statistica) oppure rappresenti un confronto tra aggregati (densità, rapporti, numeri indice, variazioni percentuali, eccetera).

Single Integrated Metadata Structure (SIMS). Eurostat nell'ambito del SSE ha definito SIMS, ossia un *template* unico contenente i concetti utili alla descrizione della qualità statistica e dei metadati dei processi statistici. Questa nuova struttura riunisce tentativi precedenti (quali i modelli *European Standard Quality Report Structure*, ESQRS, ed *Euro-SDMX Metadata Structure*, ESMS) con lo scopo di assicurare uniformità e coerenza delle informazioni sulla qualità dei processi statistici in tutto l'SSE. Metastat ha il compito di gestire ogni concetto di SIMS per ogni processo, al fine di consentirne la compilazione automatica (per ulteriori dettagli sul SIMS e per la definizione della versione italiana da implementare nel Sistan cfr. paragrafo 4.2).

Metastat si pone quindi come un sistema che utilizza, allineandole, parti di diversi modelli standard. Mentre i modelli scaturiti da gruppi di lavoro UNECE nell'ambito dell'*High-Level Group for the Modernization of Official Statistics* (HLG-MOS) sono stati già ampiamente integrati (Rizzolo *et al.* 2021) e SIMS è ampiamente usato dai paesi membri dell'Unione europea, la novità del sistema Metastat è di gestire tutta la terminologia con l'ausilio della semantica (e quindi agli standard ISO), evidenziando relazioni tra termini, come quelle di omonimia, sinonimia, polisemia, eccetera. Gli aspetti semantici rafforzano il sistema di metadati come strumento di integrazione e di confronto tra i metadati usati nei diversi processi di produzione, allargando il vocabolario da singoli insiemi di termini spesso conosciuti solo in ambiti settoriali.

2.1.4 Processo di validazione dei metadati

Metastat si pone come punto di snodo tra sistemi che, dalla fase di pianificazione a quella di diffusione dei dati, gestiscono anche metadati. A tale fine, è stata definita una procedura di governance che assicura la supervisione centralizzata di ogni operazione sui metadati da parte della struttura preposta al controllo e alla validazione dei contenuti di Metastat. Questa struttura garantisce la disponibilità di tutti i metadati prodotti per gestori e utenti interni ed esterni all'Istat. Ne garantisce, inoltre, il corretto aggiornamento, la standardizzazione, l'armonizzazione, la coerenza e l'integrazione.

È noto che un sistema di gestione dei metadati rappresenta un elemento chiave per la modernizzazione degli INS. Le Nazioni unite (United Nations 2025b) suggeriscono che tale sistema debba essere un'infrastruttura statistica comune indipendente da ciascun processo produttivo e capace di supportarli tutti.

L'idea è che i termini del glossario, con il loro specifico significato, sono usati dai processi per dare significato ai propri dati. Termini, processi e metadati strutturali sono quindi strettamente interconnessi in Metastat.

Per ogni metadato, i processi possono svolgere il ruolo di responsabili e utilizzatori. I primi generano nuovi metadati e ne diventano responsabili, i secondi utilizzano metadati già esistenti e non possono modificarli. Tutte le attività sui metadati sono eseguite dal responsabile

del processo e dal suo *team*: inseriscono i metadati referenziali, usano termini già presenti per dare significato ai dati (metadati strutturali) e, quando non sono disponibili, definiscono (divenendone responsabili) nuovi termini. Si occupano, inoltre, dei cambiamenti e della definizione degli stati durante tutto il ciclo di vita dei metadati (ad esempio, definendo una data dopo la quale il termine non dovrebbe più essere usato).

Il ruolo di responsabilità del processo sui metadati è l'elemento di collegamento tra l'anagrafica dei processi e i metadati di riferimento con i metadati strutturali, ossia tra lo standard GSBPM, utilizzato per modellare le fasi del processo, e lo standard GSIM utilizzato per modellare i metadati strutturali.

Metastat è strutturato per supportare l'interoperabilità semantica. Per questo motivo è dotato di una componente terminologica dove ogni termine ha un proprio ciclo di vita ed è collegato ai metadati strutturali e quindi a quello referenziale (tramite i processi statistici). Il processo di validazione quindi valida ogni termine sia nella sua componente semantica sia rispetto al ruolo statistico che il termine svolge nelle strutture dati generate dai processi statistici.

2.1.5 Sfide future per il sistema Metastat

Un sistema di gestione dei metadati evoluto come infrastruttura comune presenta molti vantaggi chiave. Ogni elemento dell'infrastruttura può supportare tutti i processi statistici e questi possono utilizzare le risorse per guadagnare efficienza e ridurre i costi. Un ulteriore vantaggio di un'infrastruttura statistica comune è di promuovere l'armonizzazione tra i processi statistici, grazie all'opportunità di utilizzare metodi e standard comuni.

Fino alla prima metà del 2026, il sistema rilasciato viene progressivamente popolato con i principali termini e con i relativi metadati strutturali e referenziali. Questi primi elementi costituiscono la base per un processo di standardizzazione progressivo rispetto a quanto finora definito e utilizzato autonomamente dai processi statistici. Il sistema è inoltre corredato da indicatori di qualità relativi non solo al processo e all'output statistico, ma anche ai metadati stessi, che consentono di valutare il grado di allineamento agli standard e di individuare i processi e i metadati su cui intervenire per la standardizzazione.

2.2 La standardizzazione di metodi e strumenti

I principi della qualità definiti nel *Codice delle statistiche europee* (Eurostat e SSE 2017) pongono una particolare enfasi sugli aspetti riguardanti i metodi statistici applicati nei processi di produzione: dalla definizione di solide metodologie alla standardizzazione dei metodi rispetto ai processi di produzione. Nell'ambito dell'assetto organizzativo dell'Istat, definito a seguito della modernizzazione avviata nel 2016 e basato su un sistema di servizi erogati da strutture trasversali alle unità di produzione, tale attività è svolta dalla Direzione centrale per la metodologia e il disegno dei processi statistici (DCME). Tale Direzione è fornitrice ai settori di produzione di servizi che coprono diversi aspetti del trattamento e dell'elaborazione dei dati: tra questi, l'implementazione di disegni di campionamento, la correzione dei dati raccolti, l'integrazione dei *dataset* disponibili, l'applicazione di tecniche di tutela della riservatezza.

In un quadro così variegato, sia per il numero di servizi forniti sia per la diversità delle esigenze dei settori di produzione, emerge in modo naturale l'esigenza di documentazione delle metodologie e degli strumenti statistici impiegati. È disponibile sul sito web dell'Istat



il *repository* Metodi e software del processo statistico² dedicato ai metodi e agli strumenti consolidati usati nei processi statistici e che è possibile riutilizzare per la produzione di statistiche ufficiali, anche al di fuori dell'Istat. Rispetto agli strumenti informatici sviluppati internamente per la produzione statistica, di recente è stata anche definita una governance relativa agli strumenti costruiti con software *open source*, non solo per consentire il loro riutilizzo generalizzato, ma anche per assicurarne la manutenzione e certificarne la qualità dei risultati.

Negli ultimi anni, infine, è stato ideato il Catalogo dei metodi per la produzione statistica, uno strumento attualmente in corso di progettazione, il cui obiettivo è raccogliere informazioni non solo sui *tool* statistici impiegati all'Istat nei diversi processi di produzione, ma anche su come questi vengono usati descrivendo i dettagli necessari affinché il personale, inclusi i non esperti, possa orientarsi nella scelta e nell'applicazione di tecniche statistiche spesso complesse, favorendone al tempo stesso l'apprendimento. Il Catalogo, sebbene nelle fasi iniziali di sviluppo e non ancora disponibile, presenta già delle significative potenzialità per scopi di documentazione, formazione e diffusione di buone pratiche, contribuendo così al miglioramento della qualità dei processi dell'Istat.

2.2.1 Metodi e software del processo statistico e governance sullo sviluppo di strumenti *open source*

I software statistici rivestono un ruolo cruciale nella produzione delle statistiche ufficiali. Nel corso degli ultimi anni l'Istat ha affiancato il software proprietario *Statistical Analysis System* (SAS) con soluzioni *open source* come *R* e *Python*. Questa evoluzione si è resa necessaria per permettere il riuso degli strumenti in ambito Sistan, per aumentare la flessibilità, promuovere l'innovazione e ridurre i costi. La scelta risponde anche alla necessità di ottemperare a diverse raccomandazioni a livello nazionale, come la direttiva ministeriale emanata dal Ministro per l'innovazione e le tecnologie il 19 dicembre 2003 o le raccomandazioni dell'AgID del 9 maggio 2019 (GURI, Serie generale, n.119 del 23 maggio 2019), ed europeo, in particolare con il documento *Open Source Software Strategy*, rilasciato nel 2014 per il periodo 2014-2017 e, successivamente aggiornato nell'ottobre 2020, per il periodo 2020-2023 (European Commission 2020).

L'investimento su software statistici *open source* è stato particolarmente fruttuoso perché, grazie alla flessibilità e modularità degli strumenti adottati, è stato possibile sviluppare numerosi *tool*, che implementano metodologie all'avanguardia per risolvere problematiche frequenti nel processo di produzione delle statistiche. I *tool* sono stati diffusi attraverso il già citato *repository* Metodi e software del processo statistico, ospitato sul sito istituzionale dell'Istat.

L'adozione di strumenti *open source*, anche in altri istituti nazionali di statistica europei, ha ampliato significativamente l'offerta di strumenti per le elaborazioni che caratterizzano il processo di produzione delle statistiche ufficiali. In una fase di crescente disponibilità di software statistici *open source* e di sperimentazione di tecniche di *machine learning*, appare importante definire una governance per lo sviluppo e l'uso di tali strumenti, al fine di sfruttarne appieno i vantaggi e mitigarne gli svantaggi. Ciò consentirebbe inoltre di aggiornare il *repository* istituzionale, anche attraverso la sostituzione di strumenti destinati a diventare obsoleti dal punto di vista sia metodologico sia tecnologico. La governance (D'Orazio *et al.* 2025) offre una guida per l'implementazione di metodi statistici nei software *open source*, con l'obiettivo di migliorare l'efficienza e l'efficacia dei processi di produzione statistica.

2 Cfr. Istat. <https://www.istat.it/classificazioni-e-strumenti/metodi-e-software-del-processo-statistico/>.

Si forniscono alcune raccomandazioni per la ricerca di strumenti esistenti, la valutazione della fattibilità dello sviluppo ex novo, la definizione dei requisiti, lo sviluppo e la modifica dei *tool*, nonché la fase di test. In particolare, viene descritta una procedura per la conduzione di test a difficoltà crescente e per la valutazione dell'idoneità del *tool* all'uso in produzione. La documentazione dei test e delle nuove funzionalità è cruciale per garantire la qualità e l'affidabilità del software. Il documento si conclude con indicazioni sul rilascio dei *tool* sviluppati o modificati nel *repository* istituzionale e sulle relative procedure di manutenzione per garantirne l'aggiornamento e la continuità operativa.

2.2.2 Sviluppi futuri: il Catalogo dei metodi per la produzione statistica come strumento per la qualità

Il Catalogo dei metodi per la produzione statistica si prefigge l'obiettivo di fornire la descrizione e la documentazione teorica di dettaglio delle fasi metodologiche, che corrispondono alle fasi di un processo produttivo con spiccato contenuto metodologico e per le quali la DCME ha tra i suoi compiti istituzionali quello di fornire supporto ai servizi tecnici per la realizzazione. Nel disegno del Catalogo, ciascuna fase viene articolata in "passi", concetto equivalente a *process step* secondo il modello GSIM (UNECE 2024). I passi sono il livello di dettaglio più fine del Catalogo, descrivono in maniera precisa la sequenza di operazioni da svolgere per l'esecuzione di una fase metodologica. Ogni passo viene risolto attraverso l'impiego di uno o più metodi statistici (o *process method*, secondo la terminologia GSIM), i quali a loro volta possono essere implementati attraverso l'utilizzo di oggetti software, costituiti da: i) software generalizzati specifici per l'applicazione di una determinata tipologia di metodi; ii) specifiche funzioni o pacchetti di linguaggi di programmazione; iii) procedure informatiche disegnate ad hoc non ancora sufficientemente generalizzate.

Ciascuna procedura, definita come l'implementazione di uno specifico metodo, con un particolare oggetto software, per la soluzione di un determinato passo, è corredata da una sezione di esempi particolarmente ricca e da un elenco di controlli suggeriti per verificarne il corretto e il regolare svolgimento. Tale Catalogo può quindi rendere trasparenti e facilmente replicabili i processi di produzione dell'Istat.

Le fasi metodologiche del Catalogo sono:

- campionamento delle indagini sociali;
- campionamento delle indagini economiche;
- controllo e correzione;
- trattamento delle mancate risposte totali (MRT) mediante imputazione;
- trattamento delle MRT mediante riponderazione;
- stima per piccole aree;
- tutela della riservatezza (microdati);
- tutela della riservatezza (tabelle);
- integrazione.

Il Catalogo può implementare azioni di controllo centralizzato e di miglioramento della qualità. Ad esempio, l'organizzazione dei passi secondo le usuali fasi di un processo di produzione può servire come riferimento per meglio comprendere l'organizzazione interna delle attività e, al contempo, fornire un elemento di confronto con altre organizzazioni internazionali. I software descritti in *Awesome list of statistical software* (ten Bosch 2022)³ sono raggruppati

³ Cfr. <https://github.com/SNStatComp/awesome-official-statistics-software>.

per fasi e sottoprocessi GSBPM: disporre di una mappatura tra le attività metodologiche così come organizzate all'Istat e il GSBPM può in questo caso rappresentare un punto di partenza per capire quali metodi adottati al di fuori dell'Istituto possono essere acquisiti nelle pratiche interne.

Proseguendo nel percorso delle informazioni previste per il Catalogo, la presenza degli stessi metodi e relative descrizioni permetterà di identificare le procedure comuni a più fasi metodologiche. Ciò può consentire di diffondere soluzioni e *best practice* per problemi che possono essere comuni a più fasi. D'altra parte, la possibile presenza di passi comuni a più fasi può incoraggiare la standardizzazione delle relative operazioni applicate, qualora non sia già stata avviata, con un miglioramento dell'efficienza e della trasparenza dei processi stessi. In questa sede è inoltre opportuno segnalare che indicatori di qualità in particolari indagini o fasi metodologiche possono essere documentati nel Catalogo, mostrando i software per calcolarli, i relativi controlli e i passi del processo in cui è opportuno che vengano applicati. Da questo punto di vista il Catalogo può affiancare i sistemi a supporto della qualità già esistenti all'Istat, integrandone le funzionalità da una prospettiva metodologica e di processo.

Per alcuni metodi o passi, infine, sono riportate le azioni di controllo finalizzate alla verifica della corretta applicazione del metodo e della coerenza dei risultati con quanto atteso. Si tratta in genere di regole di controllo specifiche per i metodi utilizzati, per le indagini a cui sono applicate e per i parametri adottati. Considerando tale specificità, queste azioni rappresentano comunque una buona pratica, che permette di monitorare la qualità di un processo a un livello di dettaglio che altrimenti rischierebbe di essere ignorato; il Catalogo può quindi fornire un'idea di quanto siano diffuse tali azioni di controllo e dove sia prioritario introdurne altre.

In sintesi, pur non essendo pensato in modo specifico come strumento per la qualità, il Catalogo può rappresentare un utile strumento di supporto, promuovendo la standardizzazione delle metodologie adottate, il monitoraggio della qualità e la diffusione di buone pratiche nel quadro delle attività metodologiche dell'Istat.

Tra le molteplici possibilità di utilizzo del Catalogo, di primaria importanza è la funzione di strumento di documentazione. Al suo interno l'utente può infatti trovare descrizioni complete delle metodologie consolidate per risolvere le principali fasi caratterizzate da uno spiccato contenuto metodologico nella conduzione dei processi produttivi. Le caratteristiche del Catalogo appena descritte rendono tale strumento utile ai seguenti obiettivi:

- confronto con le attuali pratiche di indagine. I responsabili dei processi statistici possono avere la possibilità di documentarsi e confrontare le pratiche correnti seguite nell'ambito dei processi di propria competenza con le indicazioni contenute nel Catalogo, verificare se la sequenza dei passi seguita è assimilabile a quella ritenuta ottimale, l'esistenza di metodi statistici eventualmente più avanzati o di procedure informatiche e software specifici.
- Formazione per i metodologi nelle attività di supporto nell'ambito di una fase di cui non si erano occupati in precedenza. Lo *staff* nei settori di produzione, con esigenze di approfondimento o applicazione dei metodi in questione, può trovare tutta la documentazione e le procedure da seguire.
- Documentazione software. È possibile eseguire delle interrogazioni per capire quali software sono in via di dismissione e di quali invece si sta incrementando l'utilizzo.
- *Repository* di metodi e strumenti. Nella prospettiva di un ampliamento della documentazione a corredo attualmente disponibile, il Catalogo può costituire un prezioso *repository* integrato di metodi statistici, strumenti informatici e descrizioni di fasi metodologiche a beneficio di tutto il personale dell'Istat.

Il programma di sviluppo del Catalogo prevede diverse fasi. La prima fase riguarda il cosiddetto livello 1, nel quale le procedure metodologiche sono descritte da una prospettiva

2. Strumenti per la qualità

ad alto livello, senza scendere nei dettagli degli specifici processi. Ciò avviene con lo sviluppo dei passi successivi attraverso l'implementazione dei livelli 2 e 3. Nel livello 2 ciascun referente di processo può confrontare le pratiche in atto in ciascun processo produttivo di propria competenza con le *best practice*, indicando per ogni replicazione dell'indagine quali macropassi, passi e metodi siano stati eseguiti e quali i software utilizzati. Nel livello 3 sono indicati anche i singoli valori dei parametri utilizzati in ciascuna procedura, contribuendo alla replicabilità dei singoli processi. In sintesi, le espansioni del Catalogo possono avvenire sia in una direzione "orizzontale", in particolare ampliando il numero delle fasi e dei metodi descritti, sia in senso "verticale", scendendo nelle specificità dei processi condotti all'Istat.

2.3 La *Quality Report Card* dei dati amministrativi, il sistema di documentazione della qualità dei dati amministrativi

La gestione sistematica della documentazione della qualità dei dati amministrativi acquisiti dall'Istat a scopi statistici è affidata alla *Quality Report Card* dei dati amministrativi (QRCA), un portale *statistics producer-oriented* disponibile sulla Intranet dell'Istat (Di Bella 2021).

Il primo prototipo della QRCA è stato pubblicato nel 2019 e, nel tempo, si è arricchito di nuove informazioni e indicatori definiti in base alle esigenze degli utenti. Dalle statistiche di accesso al sito risulta che nel 2025 gli utenti medi mensili diversi sono stati 111 con un numero medio di accessi pari a 283 (medie calcolate sui mesi del 2025, a esclusione di agosto). Si accede con le credenziali di Istituto e si seleziona, attraverso una ricerca per parola chiave, l'archivio amministrativo di interesse tra i circa 350 gestiti dall'Istat. Per ciascun archivio si possono visualizzare le caratteristiche delle forniture acquisite accedendo a 22 *report* che mostrano complessivamente 61 misure declinate per opportune variabili di classificazione e arricchite dai vari attributi disponibili⁴.

La formulazione degli indicatori presenti nella QRCA è frutto della condivisione di buone pratiche a livello internazionale e si ispira alle linee guida contenute nei principali manuali di reportistica (Eurostat 2021). L'idea originaria del sistema nasce nell'ambito del progetto internazionale *BLUE-Enterprise and Trade Statistics* (BLUE-ETS) (Daas *et al.* 2011) e si sviluppa successivamente all'Istat (Cerroni *et al.* 2014, Di Bella 2022) tenendo conto delle specificità organizzative dell'Istituto. La centralizzazione in un'unica direzione delle operazioni di raccolta dei dati sia da indagine sia da fonti amministrative – elemento fondamentale nel processo di modernizzazione – rende necessario adottare una prospettiva della qualità focalizzata sugli input dei processi. Ciò implica non solo un controllo dei processi di raccolta, ma anche sulla qualità dei dati prodotti, i quali come input vengono impiegati direttamente o in forma integrata nei processi di produzione statistica (cfr. paragrafo 5.2).

Le dimensioni della qualità che permettono l'organizzazione degli indicatori e delle misure utilizzate nella QRCA sono specifiche dei dati di input, in particolare quelli di origine amministrativa, e si differenziano necessariamente dalle dimensioni riferite alla qualità dell'output statistico. Il *framework* teorico della qualità adottato dall'Istat riflette tale distinzione (Prospetto 2.2).

⁴ Per "archivio amministrativo" si intende la struttura di un *dataset* di dati amministrativi definiti come sottoinsieme di una fonte amministrativa, in termini di unità/eventi e di variabili. Con "fornitura di dati amministrativi" ci si riferisce a uno o più *dataset* amministrativi ricevuti dall'Istat da parte del titolare dei dati amministrativi in seguito a una specifica richiesta, quindi, con la caratterizzazione dei riferimenti temporali dei dati e dell'eventuale versione (dati provvisori/definitivi). Per ulteriori dettagli sulla formalizzazione delle informazioni cfr. Di Bella 2021.

Prospetto 2.2 - *Framework* della qualità dei dati amministrativi adottato all'Istat

IPERDIMENSIONE	DIMENSIONE
Fonte Informazioni necessarie a gestire il processo di acquisizione dei dati con lo scopo di valutare e migliorare la qualità dei dati acquisiti	Informazioni di base Rilevanza e usi statistici della fonte amministrativa Riservatezza e protezione dei dati Accordi con l'ente titolare, relazioni e <i>feedback</i>
METADATI Informazioni per la valutazione della qualità a livello concettuale e di processo	Chiarezza e interpretabilità Comparabilità tra concetti amministrativi e concetti statistici Stabilità temporale dei concetti amministrativi Descrizione del processo di acquisizione e trattamento da parte del titolare
DATI Valutazione della qualità dei dati acquisiti	Stato delle forniture Aspetti temporali Controlli tecnici Integrazione Accuratezza e coerenza interna Completezza in termini di unità (copertura) e di variabili (valori mancanti)

Fonte: Istat

Nell'iperdimensione della Fonte sono presentate le informazioni attinenti all'origine dei dati e le caratteristiche principali come, ad esempio, la presenza di dati personali, la serie storica disponibile, il tipo di pretrattamento operato centralmente sui dati. L'iperdimensione dei Metadati mostra per ciascun archivio amministrativo i contenuti informativi in termini di unità, di variabili e di classificazioni delle variabili categoriali.

Nell'iperdimensione dei Dati si mostrano: le caratteristiche delle forniture di dati (denominazione, periodicità, data di acquisizione prevista) e il monitoraggio dello stato di acquisizione; gli Aspetti temporali quali la tempestività e la puntualità e, come descritto più avanti, i nuovi indicatori di dinamica delle popolazioni; gli indicatori quantitativi in termini di controlli tecnici (ad esempio il numero di valori *missing*, la presentazione delle distribuzioni in serie storica); gli indicatori di integrazione dei dati.

Le misure della qualità adottate nell'ambito della QRCA riflettono l'organizzazione della struttura centralizzata dedicata alla raccolta dati articolata nelle seguenti fasi operative:

- analisi dei fabbisogni informativi dei processi di produzione statistica che utilizzano dati amministrativi;
- richiesta dei dati agli enti che ne sono titolari e loro acquisizione attraverso il Sistema di acquisizione di archivi amministrativi e dati di indagine (Arcam) (Drovandi *et al.* 2017);
- analisi e caricamento dei microdati tramite il Sistema integrato dei microdati (SIM) (Runci *et al.* 2016);
- pseudonimizzazione primaria delle unità statistiche a cura del SIM;
- messa a disposizione dei microdati di input agli utenti interni per la produzione delle statistiche tramite il Sistema di gestione dei microdati amministrativi e statistici (Sigma) (Frattarola e Spirito 2025).

Particolare rilevanza riveste il processo di pseudonimizzazione. Esso prevede una prima fase di riconoscimento e codifica delle unità statistiche presenti nei dati amministrativi (individui e unità economiche) attraverso procedure di *record linkage*, seguite da operazioni di gestione dei dati personali mediante tecniche di ricodifica cifrata. Tali operazioni, recentemente finalizzate dal nuovo sistema Sigma, permettono di effettuare l'integrazione dei microdati per i processi di tipo multifonte e, nello stesso tempo, garantire la protezione dei dati personali secondo la normativa vigente.

2. Strumenti per la qualità

La QRCA fornisce la documentazione delle attività effettuate in termini di indicatori di qualità del processo e del prodotto (per ulteriori informazioni sulle attività di raccolta dei dati cfr. paragrafo 5.2). Nel Prospetto 2.3 è riportato l'elenco dettagliato delle misure presenti attualmente nella QRCA, organizzate nel *framework* della qualità.

Prospetto 2.3 - Elenco delle misure presenti attualmente nella *Quality Report Card* dei dati amministrativi

INDICATORE	Informazione/Misura
Report FONTE - Informazioni di base	
Identificazione	Denominazione archivio
Identificazione	Denominazione dell'ente titolare della fonte amministrativa
Riservatezza e protezione dei dati	Presenza di dati personali (regolamento (UE) 2016/679) (Sì/No)
Riservatezza e protezione dei dati	Presenza di dati rientranti in particolari categorie (ex sensibili) (Sì/No)
Riservatezza e protezione dei dati	Presenza di dati relativi a condanne penali e reati (ex giudiziari) (Sì/No)
Lunghezza della serie dei dati	Serie storica dei dati disponibili (anni disponibili)
Tipologia di trattamento	Tipologia di trattamento a cui è sottoposto l'Archivio
Lunghezza della serie dei dati	Serie storica disponibile nel SIM (anni)
Tipologia di trattamento	Codifica degli indirizzi nel Registro statistico di base dei luoghi (Sì/No)
Lunghezza della serie dei dati	Anni di codifica nel Registro statistico di base dei luoghi (RSBL)
Gestione dell'acquisizione	Referente Istat per l'acquisizione
Report FONTE - Rilevanza e usi statistici	
Rilevanza	Numero dei lavori Istat del PSN che utilizzano i dati dell'archivio per anno di aggiornamento del PSN (serie storica dal 2017)
Rilevanza	Elenco e codice dei lavori Istat del PSN che utilizzano i dati dell'archivio per anno (serie storica dal 2017)
Rilevanza	Numero di normative UE il cui adempimento dipende dall'uso dell'archivio a fini statistici per lavoro PSN per anno (serie storica dal 2017)
Rilevanza	Elenco della normativa UE il cui adempimento dipende dall'uso dell'archivio a fini statistici per lavoro PSN per anno (serie storica dal 2017)
Report METADATI - Chiarezza e interpretabilità dei dati	
Descrizione dei dati (unità)	Popolazioni e tipi di unità presenti nell'archivio
Descrizione dei dati (variabili)	Elenco delle variabili dell'archivio fornite dall'ente e loro caratteristiche
Descrizione dei dati (variabili)	Elenco strutturato delle viste di SIM per i rilasci interni e delle variabili in esse contenute
Descrizione dei dati (variabili)	Elenco delle tabelle <i>Oracle</i> che compongono l'archivio e delle variabili in esse contenute
Descrizione dei dati (variabili)	Grafico delle strutture dei dati (tabelle e chiavi <i>Oracle</i>)
Descrizione dei dati (variabili)	Classificazioni delle variabili categoriche presenti nell'archivio
Descrizione dei dati	Altra documentazione disponibile
Report DATI - Stato delle forniture	
Descrizione dei <i>dataset</i>	Nome della fornitura
Descrizione dei <i>dataset</i>	Periodicità della fornitura
Descrizione dei <i>dataset</i>	Numero di ordine del <i>dataset</i> nel caso di forniture multiple nell'anno
Descrizione dei <i>dataset</i>	Riferimento puntuale dei dati
Descrizione dei <i>dataset</i>	Anno di riferimento
Descrizione dei <i>dataset</i>	Data minima per la ricezione della fornitura
Descrizione dei <i>dataset</i>	Data massima per la ricezione della fornitura
Descrizione dei <i>dataset</i>	Modalità di acquisizione
Descrizione dei <i>dataset</i>	Trattamento previsto
Descrizione dei <i>dataset</i>	Stato del monitoraggio
Descrizione dei <i>dataset</i>	Data dello stato del monitoraggio
Descrizione dei <i>dataset</i>	Note
Report DATI - Aspetti temporali	
Puntualità	Confronto tra la data di acquisizione della fornitura e l'intervallo temporale concordato per l'invio dei dati
Tempestività dell'Ente	Intervallo temporale tra la data di acquisizione dei dati e l'ultimo evento registrato nei dati della fornitura
Tempestività complessiva	Intervallo temporale tra la data di disponibilità dei dati pretrattati centralmente e l'ultimo evento registrato nei dati della fornitura
Dinamica delle popolazioni di individui	Tassi del numero di individui nuovi, persistenti e uscenti rispetto all'anno precedente (serie storica dei dati annuali)

Fonte: Istat

Prospetto 2.3 segue - Elenco delle misure presenti attualmente nella *Quality Report Card* dei dati amministrativi

INDICATORE	Informazione/Misura
Report DATI – Controlli tecnici	
Conformità	Numero di <i>record</i> per ciascun file che compone la fornitura, valori assoluti in serie storica
Conformità	Percentuale dei campi valorizzati sul totale dei <i>record</i> valorizzati in serie storica per ogni variabile fornita
Conformità	Frequenze percentuali delle modalità in serie storica per ogni variabile categorica
Conformità	Modalità con decodifiche mancanti e corrispondente numero di <i>record</i> in serie storica per ogni variabile categorica
Report DATI – Integrazione	
Linkabilità delle unità	Numero ed elenco delle variabili disponibili e utilizzate come chiavi di <i>linkage</i> per l'identificazione e la pseudonimizzazione delle unità
Linkabilità delle unità	Percentuale di valori mancanti sul totale dei record per ogni variabile di <i>linkage</i> disponibile nella fornitura
Accuratezza delle unità/monitoraggio del processo di integrazione	Serie storica della distribuzione percentuale di <i>record</i> per tipo di abbinamento in SIM (ultimi due anni)
Accuratezza delle unità/monitoraggio del processo di integrazione	Grafico in serie storica della percentuale di record con individui che non si sono abbinati in SIM al momento del <i>record linkage</i> e a cui è stato associato un nuovo codice individuo
Accuratezza delle unità/monitoraggio del processo di integrazione	Grafico in serie storica della percentuale di record con individui che si sono abbinati in SIM al momento del <i>record linkage</i>
Accuratezza delle unità/monitoraggio del processo di integrazione	Grafico in serie storica della percentuale di record con individui abbinati in SIM al momento del <i>record linkage</i> e persistenti nell'archivio
Accuratezza delle unità/monitoraggio del processo di integrazione	Grafico in serie storica della percentuale di record con individui abbinati in SIM al momento del <i>record</i> e non persistenti nell'archivio (nuovi abbinati)
Accuratezza delle unità/Copertura	Numero di individui comuni a due fonti amministrative

Fonte: Istat

2.3.1 *Quality Report Card*: un sistema flessibile in evoluzione

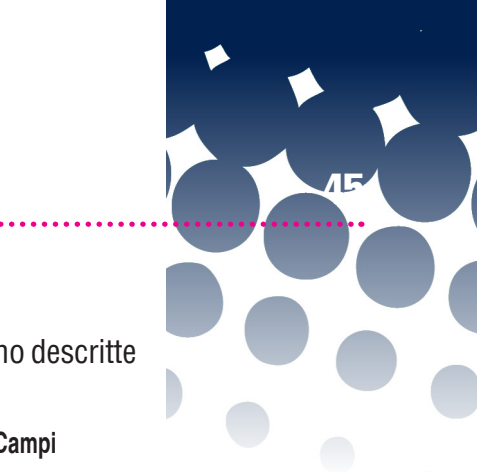
La strategia di produzione della documentazione che permette il rispetto dei vincoli di tempestività necessaria e di efficienza è il riuso dei metadati dei sistemi di gestione dei dati amministrativi ovvero il sistema di acquisizione Arcam, il sistema di trattamento dei dati amministrativi SIM, il sistema di gestione dei microdati Sigma e il database del PSN per l'associazione tra i processi statistici e i dati amministrativi.

In breve, la QRCA legge i metadati di processo ed elabora alcuni macrodati per il calcolo degli indicatori. Un database di passaggio alimenta i *report* prodotti con la piattaforma *Visual analytics* di MicroStrategy che vengono poi pubblicati nel portale della QRCA supportato da un'applicazione Java. Questo meccanismo consente di accedere a informazioni sempre aggiornate in tempo reale (con un ritardo massimo di 2-3 ore)⁵. Appena una fornitura viene caricata, si rendono disponibili nei *report* della QRCA gli indicatori di qualità.

Un esempio di *report*, inserito nell'iperdimensione Dati e nella dimensione Controlli tecnici, è riportato nella Figura 2.3. Esso mostra i risultati del controllo preliminare effettuato sul *dataset* dei Lavoratori parasubordinati dell'Inail (Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro) esposto in serie storica, sia in forma tabellare sia di grafico. In particolare si tratta del numero di campi valorizzati per ciascuna variabile della fornitura dei dati che forniscono un primo indicatore di qualità dei dati acquisiti. Tale indicatore viene calcolato, attraverso procedure parametrizzate, a monte della fase di caricamento dei dati nel database *Oracle*.

Dalla versione prototipale del 2019 che comprendeva 39 misure, la QRCA ha compiuto notevoli passi in avanti. La versione 2.0, rilasciata nel 2021, ha introdotto 12 nuove misure volte ad agevolare l'adeguamento alle disposizioni del Regolamento generale sulla protezione

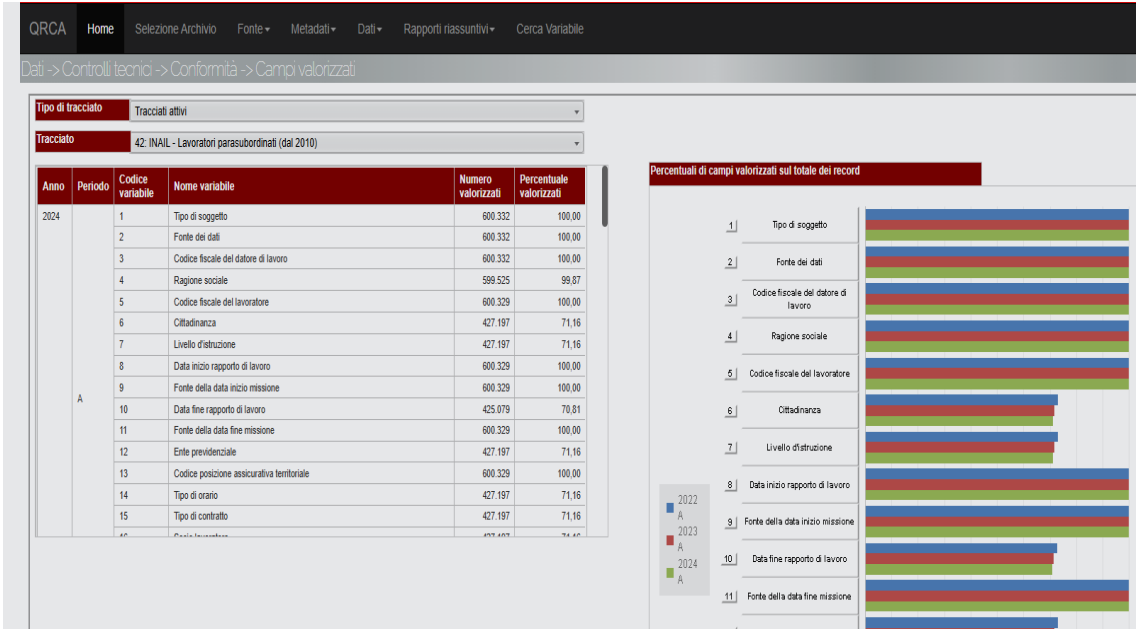
⁵ Il database della QRCA comprende infatti delle viste sui database dei sistemi che gestiscono i dati amministrativi, permettendo un aggiornamento automatico in tempo reale. In alcuni casi, per rendere più efficiente il sistema, sono state create delle tabelle che vengono aggiornate 5 volte al giorno.



2. Strumenti per la qualità

dei dati (regolamento (UE) 2016/679). Le misure pubblicate fino a tale versione sono descritte in Di Bella (2021).

Figura 2.3 - Portale QRCA - Archivio dei Lavoratori parasubordinati: *Report* “Dati->Controlli tecnici->Campi valorizzati”. Anni 2022-2024



Fonte: Istat

Con la versione 2.1, rilasciata nel 2023, è stato reso disponibile un nuovo strumento di ricerca. Oltre alla consueta selezione dell'archivio amministrativo di interesse, su tutti i *dataset* acquisiti è stata implementata la ricerca di variabili amministrative attraverso parole chiave; come risultato della ricerca accanto alle denominazioni delle variabili trovate, vengono fornite informazioni sul *dataset* di appartenenza. In tale modo, è più incisiva la ricerca di possibili fonti di interesse nella funzione di Analisi dell'usabilità dei dati amministrativi. Con la nuova versione, inoltre, nel caso in cui le informazioni siano dettagliate a livello di singolo file acquisito, è stato aggiunto un filtro che permette l'accesso ai metadati anche delle versioni precedenti, qualora le forniture siano mutate nel tempo per esigenze statistiche o per variazione delle informazioni disponibili. Su richiesta degli utenti, sono state aggiunte ulteriori misure di dettaglio, mentre diverse migliorie tecniche hanno ottimizzato l'*editing* di tabelle e grafici in relazione all'estensione delle serie storiche dei dati amministrativi disponibili (maggiore di dieci anni per alcuni indicatori e archivi).

Successivamente alla pubblicazione della versione 2.1, è stato possibile inserire nella QRCA – all'iperdimensione Dati, dimensione Aspetti temporali – il nuovo *report* “Dinamica delle popolazioni di individui”. Tale *report* descrive le variazioni nel tempo delle popolazioni di individui, ovvero le unità statistiche di base enucleate in seguito al trattamento di identificazione (integrazione logica) e pseudonimizzazione primaria, presenti nei *dataset* amministrativi.

La QRCA riporta per ogni *dataset* amministrativo acquisito dall'Istat le seguenti misure.

- Numerosità totale di individui distinti:
$$N(t) = \text{Numero di individui distinti presenti nel dataset dell'anno } t.$$
- Numerosità dell'insieme degli individui persistenti tra l'anno di riferimento e l'anno precedente:
$$P(t) = \text{Numero di individui distinti presenti nel dataset dell'anno } t \text{ e dell'anno } t-1.$$



- Numerosità dell'insieme dei nuovi individui che nell'anno entrano nel *dataset*, cioè che non risultano presenti nell'anno precedente:

$$E(t) = \text{Numero di individui distinti presenti nel dataset dell'anno } t \text{ e non presenti nell'anno } t-1.$$
- Numerosità del sottoinsieme degli individui definiti “in uscita”, ovvero presenti nell'anno di riferimento ma non nell'anno successivo:

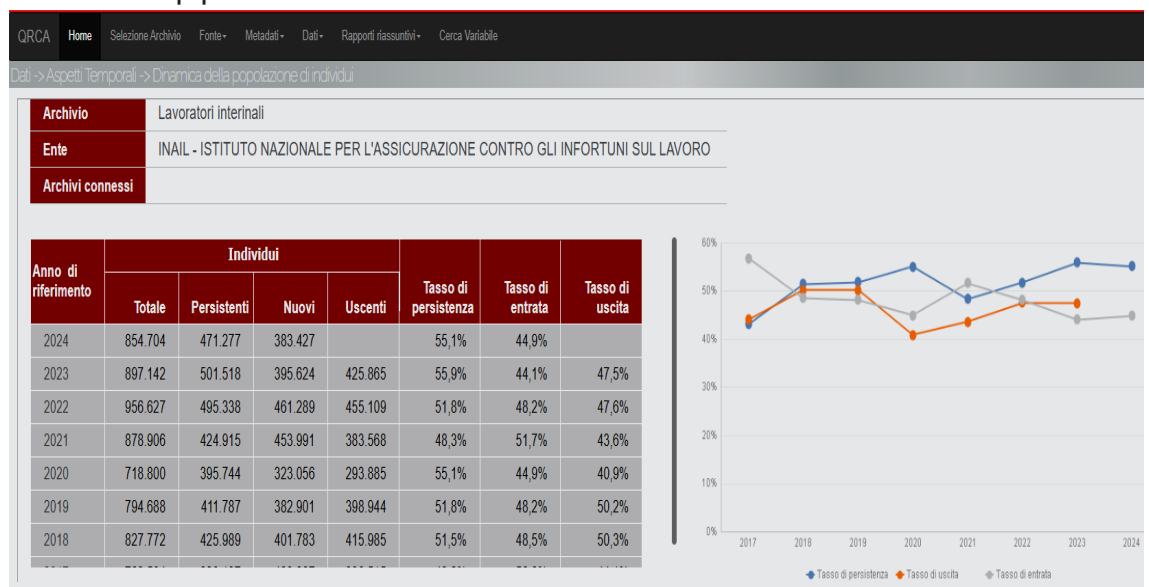
$$U(t) = \text{Numero di individui distinti presenti nel dataset dell'anno } t \text{ e non presenti nell'anno } t+1.$$

Si riportano, inoltre, il tasso di persistenza, il tasso di entrata e il tasso di uscita:

- Tasso di persistenza = $P(t)/N(t) * 100$;
- Tasso di entrata = $E(t)/N(t) * 100$;
- Tasso di uscita = $U(t)/N(t) * 100$.

Si fornisce un esempio del *report* per l'Archivio dei lavoratori interinali dell'Inail che l'Istat acquisisce annualmente dal 2009 (Figura 2.4).

Figura 2.4 - Portale QRCA - Archivio dei lavoratori interinali dell'Inail: Report “Dati->Aspetti temporali->Dinamica delle popolazioni di individui”. Anni 2017-2024



Fonte: Istat

Recentemente, grazie alle informazioni presenti in Sigma, sono state inserite alcune nuove misure. La prima riguarda la classificazione delle variabili adottata per la protezione dei dati personali, ovvero la classificazione in “variabili identificative” e “variabili tematiche” con classificazioni più specifiche che ne caratterizzano la natura. Nella sezione dei Metadati della QRCA vengono pubblicate le varie tipologie di dati presenti nella fornitura oggetto del *report*. La seconda riguarda le popolazioni amministrative le cui unità sono state codificate e pseudonimizzate: se prima le unità erano distinte solo secondo le tre tipologie di “Individui”, “Unità economiche” e “Luoghi”, ora è possibile riportare anche la descrizione della popolazione amministrativa ricavata dai metadati presenti nei *dataset* e gestita dal sistema Sigma. La Figura 2.5 mostra il nuovo *report* della QRCA per l'Archivio delle dichiarazioni contributive delle imprese per i lavoratori dipendenti (Uniemens) in cui sono presenti due popolazioni, i Lavoratori dipendenti e le Aziende, evidenziando la caratteristica di dati di tipo LEED (*Linked Employer-Employee Data*).

2. Strumenti per la qualità

Figura 2.5 - Portale QRCA - Archivio UNIEMENS – Parte contributiva dell’INPS: *Report* “Metadati->Chiarezza-> Interpretabilità -> Popolazioni amministrative”



QRCA Home Selezione Archivio Fonte Metadati Dati Rapporti riassuntivi Cerca Variabile

Metadati -> Chiarezza/Interpretabilità -> Popolazioni amministrative

Si riportano le **Popolazioni amministrative** presenti nell'archivio come unità statistiche di base nelle tipologie:

- Individui
- Unità economiche
- Individui e unità economiche
- Luoghi

La descrizione della popolazione è presente quando le unità sono state codificate (pseudonimizzate); per gli Individui e le Unità economiche attraverso il **Codice SIM**, per i Luoghi attraverso il **Codice Unico Indirizzo – CUI**. Quando una stessa popolazione comprende sia Individui (persone fisiche) che Unità economiche (persone giuridiche) la tipologia è mista e il trattamento in SIM assegna ai primi il Codice Individuo e ai secondi il Codice Unità economica. Le informazioni sono aggiornate all'ultima fornitura trattata.

Archivio	UNIEMENS - Parte contributiva
Ente	INPS - ISTITUTO NAZIONALE DELLA PREVIDENZA SOCIALE
Archivi connessi	

Popolazioni amministrative codificate (pseudonimizzate) in unità statistiche di base

Descrizione	Tipologia
Lavoratori dipendenti	Individui
Aziende con lavoratori dipendenti	Unità economiche

Fonte: Istat

L'ultimo *report* pubblicato nella QRCA comprende alcuni indicatori che mostrano le potenzialità di uso dei dati amministrativi per le statistiche multifonte. Per ciascun archivio che contiene come unità statistiche di base gli individui, si riporta la numerosità delle unità comuni a due archivi amministrativi, ovvero alla numerosità dell'intersezione tra l'archivio selezionato nella QRCA e gli altri archivi di confronto considerando le rispettive forniture dello stesso anno.

Siano:

- $N(A_t)$ il numero di individui distinti contenuti nella fornitura dell'anno t dell'archivio A selezionato e oggetto di analisi;
- $N_t(B_{it})$ il numero di individui distinti contenuti nella fornitura dell'anno t dell'archivio di confronto B_{it} per $(i = 1, \dots, nt)$ dove nt è il numero di archivi che contengono individui nell'anno t , a esclusione dell'archivio in esame;
- $I(A_t, B_{it})$ la numerosità dell'intersezione ovvero degli individui appartenenti sia all'insieme A_t sia all'insieme B_{it} . Si tratta degli individui su cui è possibile integrare le informazioni.

Per analizzare l'intersezione si mostrano nella QRCA due indicatori calcolati rapportando la numerosità dell'intersezione al totale di ciascun archivio:

- Indicatore 1, percentuale delle unità dell'archivio selezionato contenute nell'archivio di confronto per l'anno t

$$(A_t, B_{it}) = I(A_t, B_{it}) / N(A_t) * 100 \quad \text{per ogni } i, t,$$

- Indicatore 2, percentuale delle unità dell'archivio di confronto contenute nell'archivio selezionato

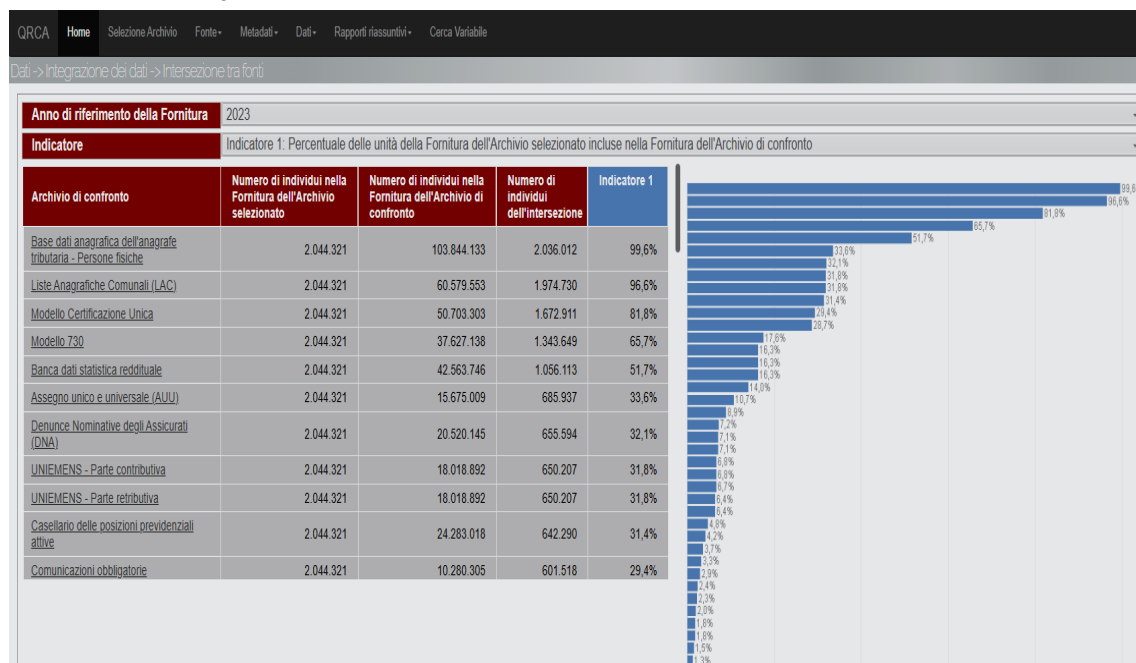
$$(A_t, B_{it}) = I(A_t, B_{it}) / N(B_{it}) * 100 \quad \text{per ogni } i, t.$$

Le numerosità sono calcolate come numero di individui distinti.

Gli indicatori possono fornire una valutazione preliminare sulle possibili integrazioni delle variabili tra due fonti amministrative e, quindi, sulla sottocopertura o sovracopertura della popolazione target di interesse. Di seguito un esempio del nuovo *report* che mostra quanti individui della fornitura 2023 dell'archivio “Anagrafe degli studenti universitari - Iscritti e iscrizioni” sono contenuti anche negli altri archivi acquisiti a fini statistici a parità

di anno di riferimento dei dati (Figura 2.6). Il segnale di presenza dello studente può dare informazioni sulle attività svolte dal collettivo di individui in esame.

Figura 2.6 - Portale QRCA - Archivio dell'Anagrafe nazionale degli studenti universitari - Iscritti e iscrizioni: Report-> Dati->Integrazione dei dati-> Intersezione tra fonti. Anno 2023



Fonte: Istat

2.3.2 Qualità dell'input per i dati amministrativi

Poiché la qualità dell'output statistico dipende necessariamente dalla qualità dell'input utilizzato, per le statistiche basate sull'uso delle fonti amministrative è possibile tracciarne le interrelazioni. Diversi elementi dell'input influenzano direttamente aspetti fondamentali dell'accuratezza dell'output, come la presenza di valori mancanti nei *dataset* amministrativi, gli errori di misurazione e la copertura delle unità target. Inoltre, la tempestività dei *dataset* amministrativi acquisiti (calcolata come la distanza dalla data di acquisizione e l'ultimo evento registrato nei dati) e la puntualità dei titolari dei dati (rispetto delle scadenze di consegna dei dati concordate) incidono sulle medesime dimensioni dell'output statistico, *mutatis mutandis*.

L'accuratezza dei dati amministrativi utilizzati come input dei processi statistici è ulteriormente condizionata dalle operazioni di integrazione logica operate centralmente presso l'Istat. Queste ultime implicano il riconoscimento delle unità statistiche attraverso procedure di *record linkage* e la conseguente pseudonimizzazione dei dati. In presenza di errori nelle chiavi di *linkage* utilizzate, ovvero dei dati identificativi delle unità, tali operazioni possono generare possibili errori di mancati o falsi abbinamenti.

Il percorso di valutazione del processo dall'input all'output rende necessario operare un monitoraggio continuo della qualità dell'input considerando che i dati amministrativi sono prodotti al di fuori dell'Istat per finalità prevalentemente non statistiche e possono subire modifiche normative o gestionali. A tale scopo la QRCA, dove opportuno, pubblica nei *report* le misure in serie storica, intercettando possibili cambiamenti nei dati che possono derivare da cadute della qualità.

2. Strumenti per la qualità

Oltre alla funzione valutativa dell'input operata dalla QRCA, la misurazione della qualità dei dati amministrativi svolge un ruolo strategico per la prima analisi di usabilità di una fonte, laddove si prevede di inserire innovazioni nel processo statistico (ad esempio, sostituzione o integrazione dei dati amministrativi rispetto ai dati di indagine).

In generale, la QRCA si configura come un presidio irrinunciabile della qualità dei dati amministrativi, orientato non solo ai produttori di statistiche ma un utile supporto per il personale coinvolto nella gestione di acquisizione e trattamento, nonché per le attività dell'Istituto di rendicontazione sulla qualità e per i decisori istituzionali.

2.4 Il Dossier Qualità

I sistemi informativi sulla qualità presentati nei paragrafi precedenti non esauriscono la documentazione metodologica e sulla qualità che viene correntemente prodotta per i processi statistici dell'Istat. Di seguito alcuni esempi: i comunicati stampa pubblicati sul sito istituzionale sono in genere accompagnati da note metodologiche e glossari; una specifica sezione del sito contiene le schede standard di qualità per la maggioranza dei processi correnti; per tutte le statistiche sotto regolamento viene prodotto e trasmesso a Eurostat un *quality report* che viene spesso pubblicato a corredo dei dati diffusi da Eurostat stesso. Tali documenti hanno obiettivi, utenti e mezzi di diffusione diversi, ma possono anche contribuire a costruire un quadro completo e aggiornato della qualità del processo a cui si riferiscono. Si è di conseguenza avvertita l'esigenza di costruire un punto unico di accesso a tale documentazione. Il Dossier Qualità (DQ) è stato progettato e implementato nel corso del 2025 allo scopo di facilitare la consultazione della documentazione legata ai processi statistici già disponibile in Istituto. Si presenta come una mappa ragionata che raccoglie tutti i metadati relativi ai processi, descrivendone contenuti, contesto, qualità e struttura.

Il Dossier nasce quindi con l'obiettivo di tracciare in modo sistematico le azioni che l'Istat svolge nell'ambito della produzione statistica, offrendo agli utenti uno strumento utile per valutare la qualità delle informazioni prodotte.

In questa fase iniziale, il DQ è destinato esclusivamente agli utenti interni e consultabile tramite la Intranet dell'Istat. In prospettiva futura, si sta valutando l'opportunità e la fattibilità di renderlo disponibile anche sul sito istituzionale. Oltre a essere un supporto operativo, il Dossier rappresenta un importante strumento di trasparenza e di rendicontazione: consente di comprendere le metodologie adottate e i criteri utilizzati per garantire affidabilità e precisione dei dati, rafforzando così la fiducia nelle statistiche dell'Istat.

2.4.1 Ambito e contenuti del Dossier Qualità

Per quanto riguarda l'ambito di applicazione del DQ, si è scelto di procedere per gradi, disegnando in questa prima implementazione il Dossier con riferimento ai processi di indagine statistica, per rimandare a futuri sviluppi la definizione di strutture analoghe in relazione ad altre tipologie di processi quali i registri statistici e i processi di elaborazione dati, ad esempio quelli della contabilità nazionale.

Focalizzata l'attenzione sulle indagini statistiche, è stato necessario individuare l'elenco dei processi per cui produrre il Dossier.

Per la prima implementazione si è scelto di fare riferimento a tutti i processi continuativi



e ai processi periodici o occasionali per cui la fase di raccolta dati si è svolta nel triennio 2023-2025. L'elenco dei processi viene comunque integrato nel tempo. Il Dossier prevede una sezione in cui sono resi disponibili i documenti descrittivi del processo che fanno riferimento all'ultima edizione di indagine svolta e una sezione legata alle differenti edizioni di indagine così che, per i processi continuativi e/o periodici, permette di accedere ai documenti specifici delle singole edizioni con riferimento, ad esempio, al monitoraggio della raccolta dati.

Il DQ viene aggiornato periodicamente per includere la nuova documentazione disponibile offrendo una visione sempre attuale e completa delle attività.

Il Dossier, disponibile in un'apposita sezione della Intranet istituzionale, presenta innanzitutto per ogni indagine una serie di informazioni anagrafiche: Codice PSN, Denominazione processo, Struttura organizzativa competente e base normativa (Figura 2.7).

Figura 2.7 - Schermata del Dossier Qualità

Dossier Qualità													
Il dossier													
Codice PSN	Denominazione Processo	Direzione - Servizio competente	BASE NORMATIVA	Pagina rispondenti	SIQual	Scheda standard di qualità	Ultimo Quality Report pubblicato da Eurostat	Ultimo Quality Report validato da Eurostat - Non pubblicato	Ultimo Report - Nota metodologica e glossario	SIGMA	Ultimo Report di valutazione dell'audit	Check list	Edizioni del processo
IST-00697	AGRITURISMO - Agriturismo	DCAT-ATC	PSN										
IST-00907	AMBURB - Rilevazione Dati ambientali nelle città	DCAT-ATA	PSN										
IST-00175	ANNUALE LATTE - Indagine mensile e annuale su latte e sui prodotti lattiero caseari	DCAT-ATC	Regolamento UE										

Fonte: Istat

- A seguire, per ogni dossier è possibile accedere ai collegamenti alle seguenti informazioni:
- la pagina Informazioni sulla rilevazione che viene predisposta per ogni indagine per fornire informazioni ai rispondenti chiamati a partecipare all'indagine;
 - la sezione di presentazione del processo del sistema SIQual nella sua versione interna all'Istituto, da cui è possibile consultare anche gli indicatori di qualità disponibili;
 - la specifica Scheda standard di qualità, che rappresenta una versione in italiano, standardizzata, dei *quality report* trasmessi a Eurostat;
 - il più recente *quality report* trasmesso a Eurostat. Nel caso esso sia stato pubblicato da Eurostat, si trova il collegamento direttamente alla pagina del sito Eurostat, altrimenti viene caricato nel DQ il *quality report* una volta validato da Eurostat;
 - la più recente nota metodologica e glossario diffusa sul sito.

Sono inoltre caricati nel DQ alcuni documenti prodotti a seguito delle procedure di valutazione della qualità delle indagini (cfr. paragrafo 3.1) non già disponibili altrove:

- la *checklist* per la valutazione della conformità alle pratiche consolidate in Istituto compilata dall'indagine e l'esito ottenuto;

2. Strumenti per la qualità

- il *report* di valutazione dell'*audit*, in caso il processo sia stato sottoposto ad *audit* nell'ultimo ciclo di *audit* avviato nel 2024;
- il dominio specifico di integrazione (DSI) utilizzato dall'indagine e la data del primo utilizzo dei dati. I DSI sono ambienti controllati il cui scopo principale è consentire l'elaborazione statistica dei microdati amministrativi e statistici in modo sicuro, in linea con le indicazioni del Garante per la protezione dei dati personali. In ogni DSI sono presenti solo i dati necessari per conseguire una specifica finalità statistica, per il tempo strettamente necessario, e a cui hanno accesso esclusivamente le persone opportunamente autorizzate.

All'interno della sezione "edizioni del processo", sono disponibili le informazioni relative all'analisi del rischio di trattamento, alla valutazione di impatto sulla protezione dei dati e al *report* di monitoraggio.

L'analisi del rischio di trattamento (ADR) è un obbligo generalizzato per tutti i trattamenti di dati personali introdotto dal regolamento (UE) 2016/679 ed è uno degli strumenti attraverso il quale il titolare dimostra e documenta le scelte operate in ordine al trattamento, nonché l'adozione di idonee misure organizzative e tecniche di mitigazione del rischio, in ossequio al principio di *accountability*. Il Dossier riporta la documentazione relativa alla ADR che risulta accessibile ai direttori designati al trattamento dei dati e alle strutture di competenza.

La valutazione di impatto sulla protezione dei dati (VIP) (art. 35 del regolamento (UE) 2016/679 - GDPR), ha lo scopo di verificare che il trattamento dei dati personali rispetti i principi del regolamento. La documentazione relativa è riportata nel Dossier dell'indagine e accessibile alle strutture competenti.

Merita un approfondimento il *report* di monitoraggio, unico documento tra quelli del DQ di recente realizzazione. Esso ha l'obiettivo di sintetizzare le informazioni che consentono di valutare l'andamento del processo di raccolta dati e la qualità dei dati forniti. Per il 2024, per ciascuna indagine è stato caricato un *report* finale riassuntivo dell'anno predisposto sulla base delle caratteristiche dell'indagine. Per migliorare la comparabilità, a partire dal 2025, è stato predisposto un *report* standard con periodicità di calcolo e pubblicazione mensile: le tavole vengono calcolate con riferimento alle attività di raccolta dati svolte nell'intero mese. Poiché le indagini svolte dall'Istat possono essere difformi tra loro per molti aspetti (organizzazione, unità di rilevazione, periodicità, tecnica di indagine, eccetera), è prevista una personalizzazione delle tabelle da produrre in base alle caratteristiche specifiche dell'indagine. In generale, è presente una tabella complessiva dell'indagine per valutarne l'andamento nel mese in cui viene prodotto il *report*, e, ove possibile, le tabelle specifiche per tecnica di indagine, per *wave* e per tipologia di rispondenti. Le tabelle raccolgono informazioni sulle distribuzioni delle unità rispondenti per area geografica, per tipologia di esito, per sanzionabilità, classe di addetti, attività economica secondo la classificazione Ateco, ma anche sul numero di interviste complete per settimana di rilevazione. Nel *report* sono presenti, inoltre, tavole volte a monitorare l'efficacia dell'attività delle reti di rilevazione sia CATI (*Computer-Assisted Telephone Interviewing*), sia CAPI (*Computer-Assisted Personal Interviewing*), ove previste nel disegno di indagine.

Si può notare che la consultazione del DQ permette di seguire l'intero ciclo di vita delle indagini, a partire dal monitoraggio del lavoro sul campo delle reti di rilevazione fino alla produzione di *report* specifici sulla qualità dei risultati finali.

3. METODI PER LA VALUTAZIONE DELLA QUALITÀ¹

Come in tutti i sistemi di gestione della qualità, l'attività di valutazione ha un ruolo centrale per stimolare il miglioramento dei processi e dei prodotti. Una delle principali innovazioni introdotte dalla politica della qualità dell'Istat del 2021 è stata l'adozione di un sistema di valutazione differenziato in base alla tipologia e alla maturità dei processi.

Per i processi di tipo tradizionale (indagini ed elaborazioni), per cui l'Istat ha già una documentazione standard, degli indicatori di qualità e dei *quality report*, è stata definita una nuova procedura di valutazione: tale procedura integra l'applicazione di una *checklist*, che verifica la conformità di tutti i processi alle pratiche e metodologie consolidate presso l'Istat e assegna una *label* di qualità ai processi conformi, a una procedura di *audit* statistico interno, limitato a un numero ristretto di processi (cfr. paragrafo 3.1). Entrambe le metodologie di valutazione sono state definite in modo da tenere conto dell'organizzazione dell'Istat dopo la modernizzazione, in cui alcune fasi del processo statistico possono essere assegnate a diverse strutture, sebbene il coordinamento resti al responsabile tematico.

Un secondo tipo di processi statistici sono quelli alla base dei registri del Sistema integrato dei registri (SIR), che rappresentano attualmente una base informativa fondamentale per la produzione statistica corrente. Per produrre e aggiornare i registri del SIR, dati amministrativi e risultati di processi statistici vengono integrati in complessi processi multifonte. Gli indicatori di qualità standard definiti e correntemente calcolati per i processi tradizionali non riescono a cogliere le peculiarità di questi processi ed è per questo che è stato definito un apposito *framework* di qualità (cfr. paragrafo 3.2).

Negli ultimi anni l'Istat, infine, ha investito molto nella ricerca e nello sviluppo di processi basati sulle nuove fonti, per la produzione delle cosiddette *Trusted Smart Statistics* (TSS). Questo tipo di statistiche, basate su nuove fonti, non possono essere considerate statistiche ufficiali, ma sono definite statistiche sperimentali². Spesso, infatti, per le TSS non si è in grado di garantire il rispetto di tutti i requisiti di qualità richiesti dagli standard dell'Istat, in quanto molti metodi e misure per la valutazione della qualità tradizionali non sono applicabili a questo tipo di processi, oltre al fatto che entrano in gioco fattori specifici, che possono influire sulla qualità. Per questi motivi, l'Istat sta procedendo in primo luogo alla definizione di un *framework* di qualità di riferimento per questi processi, cercando di generalizzare i risultati delle esperienze internazionali e di quelle interne (cfr. paragrafo 3.3).

¹ Il Capitolo 3 è stato redatto da: Francesca Budano (paragrafo 3.1.1), Maria Francesca Loporcaro (paragrafo 3.1.2), Fabiana Rocci e Giorgia Simeoni (paragrafo 3.2), Giorgia Simeoni (paragrafi 3, 3.3, 3.3.1 e 3.3.7), Francesca Inglese (paragrafo 3.3.2, 3.3.3 e 3.3.4), Giorgia Simeoni e Francesca Inglese (paragrafo 3.3.5), Luca Mancini (paragrafo 3.3.6).

² Si fa presente la natura sperimentale delle statistiche citate, disponibili sulla pagina web dell'Istat dedicata a queste tipologie di dati: <https://www.istat.it/comunicati-e-analisi/statistiche-sperimentali/>.

3.1 La valutazione dei processi tradizionali

3.1.1 Valutazione della conformità tramite checklist

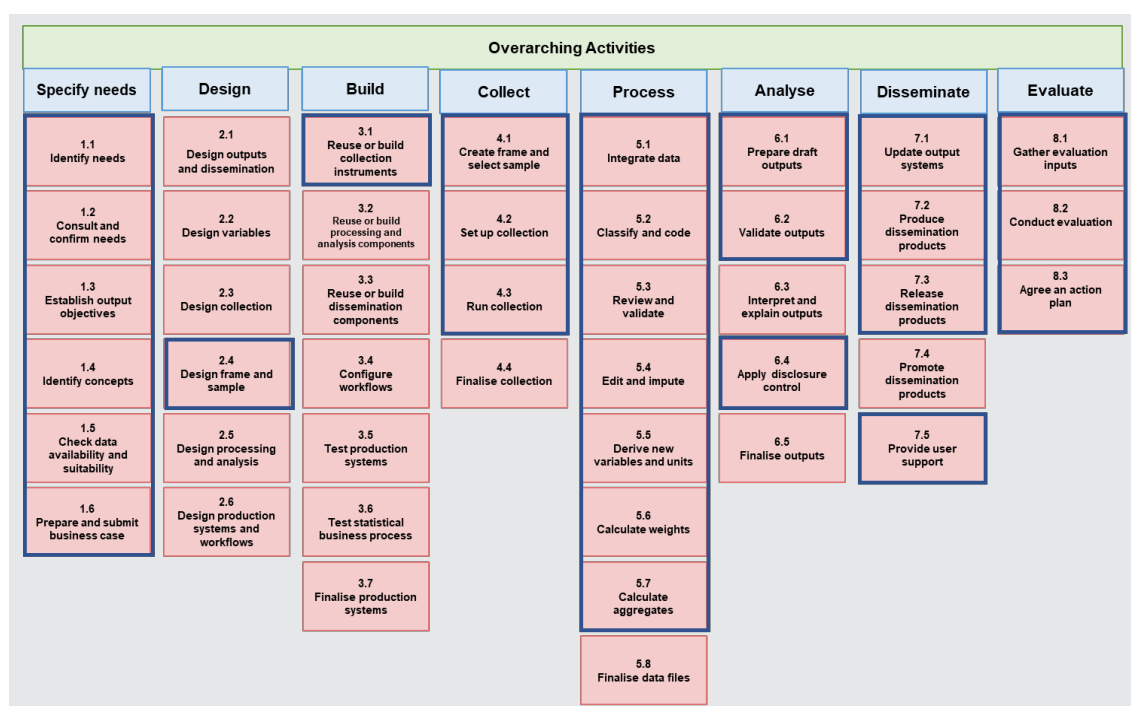
Considerando i processi tradizionali correnti (rilevazioni dirette, processi di fonte amministrativa o elaborazioni), è stato introdotto un approccio di *labelling* interno volto a favorire un ulteriore miglioramento della qualità. Tale soluzione ha consentito di effettuare una valutazione su larga scala, adottando però uno strumento snello: la *checklist*.

La *checklist* è stata applicata ai processi statistici di tipo tradizionale come le indagini dirette, ma anche alle elaborazioni (ad esempio quelle relative alle statistiche di contabilità nazionale) con l'obiettivo di verificarne la conformità alle metodologie e alle procedure consolidate presso l'Istat (considerate come standard), con particolare riferimento alle operazioni che possono influenzare la qualità degli output statistici, quali l'accuratezza delle stime, la coerenza con le altre fonti, la comparabilità nel tempo e l'affidabilità.

La *checklist* è stata progettata e realizzata da un sottogruppo del Comitato Qualità con il supporto di esperti esterni al comitato, considerando le principali fasi e sottoprocessi del processo statistico, in conformità al modello standard GSBPM (UNECE 2025). Nella Figura 3.1 sono rappresentate in blu le fasi e i sottoprocessi interessati dalla *checklist*. Per ogni fase e sottoprocessi sono stati considerati cinque aspetti fondamentali:

1. le metodologie e gli strumenti;
2. le azioni di controllo della qualità;
3. gli indicatori di qualità;
4. la documentazione;
5. l'autovalutazione.

Figura 3.1 - Rappresentazione delle fasi e dei sottoprocessi secondo il modello GSBPM (v5.2)



Fonte: United Nations Economic Commission for Europe - UNECE

3. Metodi per la valutazione della qualità

Sulla base delle informazioni raccolte mediante la *checklist* dai responsabili coinvolti, è stata attribuita una *label* di qualità ai processi conformi agli standard definiti; negli altri casi, sono state individuate azioni di miglioramento per le pratiche non ancora adottate.

In particolare, la *checklist* si articola in una parte relativa all'anagrafica del processo, che include il nominativo del responsabile, la struttura di appartenenza e ulteriori referenti, e in undici sezioni relative ad alcune fasi del processo statistico, con risposte prevalentemente chiuse ("Sì/No", ossia presenza/assenza dell'*item* descritto).

È inoltre prevista una sezione finale di valutazione dedicata a raccogliere il *feedback* dei responsabili sulla compilazione, nonché eventuali suggerimenti per migliorarla.

Poiché non tutti i processi considerati (rilevazioni dirette, esaustive o campionarie, di fonte amministrativa ed elaborazioni) seguono gli stessi sottoprocessi (ad esempio le elaborazioni non prevedono campionamento e non tutti i processi integrano più fonti) è stato fondamentale nella costruzione della *checklist* prevedere l'inserimento di domande filtro all'inizio di specifiche sezioni. Ciò ha consentito di identificare i sottoprocessi rilevanti per ciascun tipo di processo (Prospetto 3.1).

Prospetto 3.1 - Sezioni della *checklist* e domande filtro

Sezione del questionario	Domanda filtro
B - Lista di riferimento e disegno di campionamento	Il processo statistico è una rilevazione
D - Integrazione	Il processo statistico prevede l'integrazione di due o più <i>dataset</i>
E - Codifica e classificazione	Il processo statistico prevede una rilevante procedura di codifica di quesiti aperti
F - Identificazione e trattamento degli errori non campionari	Il processo statistico prevede una procedura di identificazione e trattamento degli errori non campionari
G - Derivazione unità	Il processo statistico prevede la derivazione di unità
H - Derivazione variabili	Il processo statistico prevede la derivazione di variabili
I - Calcolo dei pesi e degli aggregati	Il processo statistico prevede una procedura di calcolo e aggiustamento dei pesi
L - Validazione dei risultati	Vengono effettuate procedure di validazione dei risultati

Fonte: Istat

Vengono di seguito descritte sinteticamente le undici sezioni della *checklist*.

Sezione A - *Esigenze informative degli utenti/definizione dei concetti*

Questa sezione viene compilata da tutti i processi statistici e ha l'obiettivo di valutare la pertinenza dei risultati prodotti da ogni processo: a tale scopo si esamina se sono state raccolte e soddisfatte le esigenze informative degli utilizzatori e viene approfondito l'uso di concetti e classificazioni standard, che garantisce pertinenza e comparabilità dei risultati.

Sezione B - *Lista di riferimento e Disegno di campionamento*

Questa sezione è stata progettata per la compilazione del questionario da parte dei processi statistici di tipo "rilevazione" e approfondisce gli aspetti relativi alla lista di riferimento e al disegno di campionamento. In particolare, viene richiesta l'effettuazione di azioni di controllo preventive e di monitoraggio sulla lista e sui criteri alla base della scelta del disegno di indagine, in particolare se è stato adottato un disegno di campionamento non probabilistico.

Sezione C - *Acquisizione dei dati*

Questa sezione approfondisce la fase di acquisizione dei dati, cruciale per la qualità dei risultati, poiché è in questa fase che si originano molti errori non campionari con un impatto significativo sull'accuratezza. È necessario eseguire approfondimenti specifici in base alla tipologia di raccolta dei dati: raccolta diretta dei dati, acquisizione di archivi amministrativi e acquisizione di altre fonti. Le principali tecniche assistite da computer sono CATI

(*Computer-Assisted Telephone Interviewing*), CAWI (*Computer-Assisted Web Interviewing*), e CAPI (*Computer-Assisted Personal Interviewing*). Nel caso della raccolta diretta dei dati si rileva l'uso di tecniche assistite da computer (interviste telefoniche con tecnica CATI; via web con tecnica CAWI o CAPI, eccetera) per favorire la partecipazione. Viene inoltre richiesta l'applicazione di concetti e classificazioni standard nel questionario, nonché l'adozione di azioni di prevenzione e monitoraggio degli errori, il calcolo di indicatori di qualità per il monitoraggio e la valutazione delle misure adottate per contenere il carico statistico.

Per l'acquisizione di archivi amministrativi viene richiesto se l'acquisizione sia avvenuta tramite la struttura incaricata della raccolta dei dati oppure in maniera autonoma. Si verifica, inoltre, l'esistenza di accordi formali, il rispetto dei requisiti di sicurezza nella trasmissione dei dati e il soddisfacimento delle esigenze informative.

Per l'acquisizione di altre fonti, infine, si approfondiscono l'esistenza di accordi e la garanzia della sicurezza dei dati.

Sezione D - *Integrazione*

Sezione E - *Codifica e classificazione*

Sezione F - *Identificazione e trattamento degli errori non campionari*

Sezioni G e H - *Derivazione unità e variabili*

Sezione I - *Calcolo dei pesi e degli aggregati*

Sezione L - *Validazione dei risultati*

Sezione M - *Tutela della riservatezza*

Le sezioni dalla D alla M approfondiscono gli aspetti relativi al trattamento dei dati. Viene sottolineata l'importanza di utilizzare metodologie consolidate per garantire la qualità dei dati e indicatori di qualità per monitorare e migliorare i processi, l'importanza della prevenzione e della correzione degli errori, l'attività di validazione dei risultati e la protezione dei dati sensibili con procedure specifiche per garantire la riservatezza.

Sezione N - *Diffusione*

Questa sezione esamina vari aspetti relativi alla diffusione dei risultati, concentrandosi sugli elementi che influenzano maggiormente la qualità. È essenziale avere una chiara programmazione dei rilasci dei prodotti di diffusione per trasparenza verso gli utenti e, di conseguenza, un aspetto importante da monitorare è la puntualità (definita come il tempo tra la data di rilascio effettiva e quella programmata).

Sezione O - *Documentazione e valutazione*

Questa sezione approfondisce la disponibilità di documentazione per le varie fasi del processo e se il processo stesso sia stato oggetto di valutazione in precedenza.

Domande di autovalutazione

In ogni sezione sono state inserite domande di autovalutazione, con l'obiettivo di confrontare i risultati emersi dalla *checklist* con la percezione del responsabile del processo e di identificare eventuali criticità non emerse dalla struttura schematica della *checklist*.

Compilazione della *checklist*

Prima di richiedere ai referenti dei processi di compilare la *checklist*, una parte delle domande è stata precompilata sulla base delle informazioni disponibili nel sistema Sidi-SIQual, lasciando la possibilità di poterle confermare o modificare se necessario. Nella compilazione

della *checklist* il principale attore coinvolto è stato il responsabile tematico che ha una conoscenza approfondita del processo.

Laddove opportuno, sono stati coinvolti nella compilazione di specifiche sezioni o quesiti della *checklist* altre due tipologie di referenti:

- il referente della raccolta dati per la sezione relativa all'acquisizione dei dati;
- i referenti di altre direzioni trasversali, in particolare i referenti delle metodologie, laddove coinvolti in specifici sottoprocessi di competenza (controllo e correzione, campionamento, eccetera).

Analisi dei risultati e monitoraggio delle azioni di miglioramento

Dopo avere raccolto le informazioni attraverso la *checklist*, si è proceduto con l'analisi dei dati per ciascun processo e la conseguente attribuzione di un esito. Questa analisi è stata effettuata da un gruppo di esperti del Comitato Qualità, sulla base di una griglia di valutazione precedentemente concordata, per identificare eventuali criticità e proporre l'implementazione di eventuali azioni di miglioramento. È stato importante definire esiti armonizzati al fine di garantire coerenza e obiettività nella valutazione.

La fase di analisi delle *checklist* è stata molto complessa e non sempre ha consentito di raggiungere subito l'attribuzione di un esito; talvolta sono state necessarie richieste di chiarimento ai referenti dei processi. Si è tenuto traccia di tali scambi al fine di migliorare la *checklist* per le prossime edizioni.

Sebbene tutti i processi statistici dell'Istat che producono statistiche ufficiali rispettino stringenti requisiti di qualità, ai processi che sono risultati pienamente conformi agli standard consolidati per tutte le fasi è stata attribuita la *label* di "Statistica di qualità". L'esito della *checklist* è stato comunicato via email ai responsabili dei processi e i processi sono stati inclusi in un elenco ufficiale diffuso sulla intranet dell'Istat.

Per i processi risultati non completamente conformi, ad esempio a causa del mancato calcolo degli indicatori di qualità per qualche fase del processo o del mancato aggiornamento della documentazione, sono state proposte opportune azioni di miglioramento che una volta implementate, hanno consentito ai processi di ottenere la *label* di "Statistica di qualità" e di essere inseriti nell'elenco dei processi conformi. Tale meccanismo è stato infatti disegnato come stimolo al miglioramento continuo della qualità, obiettivo principale della valutazione.

Risultati dell'applicazione della *checklist*

La *checklist* è stata finalizzata e sviluppata con il software di Gestione indagini online (GINO), utilizzato in Istat per la conduzione di indagini CAWI, con il supporto della struttura apposita. Dopo essere stata precompilata con informazioni disponibili nel sistema di documentazione delle indagini Sidi-SIQual, la fase di compilazione si è svolta a cavallo tra il 2022 e il 2023.

Dall'analisi dei risultati è emerso che, su 210 processi sottoposti a *checklist*, inizialmente circa il 38% dei processi sono risultati conformi, mentre al termine dei due anni di tempo in cui si potevano implementare le azioni di miglioramento assegnate, la percentuale di conformi è salita all'86%. Per il restante 14% le azioni sono ancora in corso di implementazione o non è stato possibile implementarle ad esempio per carenza di risorse. Il monitoraggio costante dell'implementazione delle azioni di miglioramento ha consentito di aggiornare progressivamente la percentuale dei processi conformi.

Le informazioni raccolte attraverso la *checklist* oltre a permettere la valutazione di conformità dei processi, hanno fornito una preziosa panoramica su come vengono condotti i processi statistici, permettendo approfondimenti anche sulle attività trasversali e focaliz-

zando l'attenzione su specifiche fasi del processo. Si è potuto verificare ad esempio che circa il 36 per cento dei processi considerati come tradizionali non si basano solo su indagini dirette, ma prevedono l'integrazione di due o più fonti: in taluni casi si tratta di indagini dirette che utilizzano fonti aggiuntive per il trattamento dei dati mancanti (ad esempio, ciò accade per la Rilevazione sulle forze di lavoro che è una rilevazione campionaria diretta che utilizza fonti supplementari), in altri casi si tratta di processi che integrano i dati di indagine con quelli di fonti amministrative per recuperare da esse alcune variabili (come per l'Indagine sul Trasporto di merci su strada), o ancora processi basati su fonti amministrative che effettuano integrazioni di più archivi (come per la Rilevazione degli incidenti stradali con lesione a persone).

3.1.2 Valutazione della qualità tramite audit

La norma ISO 19011 (ISO 2018b) definisce l'*audit* in un contesto generale, non statistico, come un "processo sistematico, indipendente e documentato per ottenere evidenze oggettive e valutarle con obiettività al fine di determinare in quale misura i criteri dell'*audit* siano soddisfatti". Questo sistema di valutazione può essere applicato a un'organizzazione nel suo complesso oppure ai singoli processi statistici (Prospetto 3.2).

Nel primo caso l'obiettivo è quello di valutare l'assetto istituzionale e strutturale in cui i processi statistici sono inseriti. Si focalizza su aspetti quali:

- la strategia e la cultura della qualità;
- la gestione delle risorse (umane, tecniche, finanziarie);
- l'organizzazione interna (divisione dei compiti, comunicazione);
- la conformità a principi come l'indipendenza professionale, la trasparenza e la riservatezza.

Viene inoltre analizzato come questi aspetti influenzino la qualità e la sostenibilità del sistema statistico. Nel caso degli INS, il meccanismo di *peer review* (cfr. paragrafo 1.4) rappresenta un esempio di valutazione della qualità a livello di organizzazione.

Nel secondo caso l'obiettivo è quello di valutare l'efficienza, la coerenza e la conformità dei processi rispetto agli standard metodologici e alle buone pratiche. Gli aspetti su cui si focalizza sono:

- il flusso operativo della produzione statistica (ad esempio raccolta, codifica, validazione);
- l'applicazione coerente delle metodologie;
- l'identificazione di colli di bottiglia o ridondanze.

Prospetto 3.2 - *Audit* di processo e *audit* dell'organizzazione: schema riassuntivo delle caratteristiche

Aspetto	<i>Audit</i> di processo	<i>Audit</i> dell'organizzazione
Focus	Fasi e metodi della produzione dati	Struttura, governance e risorse
Scopo	Miglioramento dei processi operativi	Valutazione della capacità istituzionale
Approccio	Più tecnico e metodologico	Più strategico e organizzativo
Output tipico	Raccomandazioni per ottimizzare flussi	Raccomandazioni su gestione e struttura

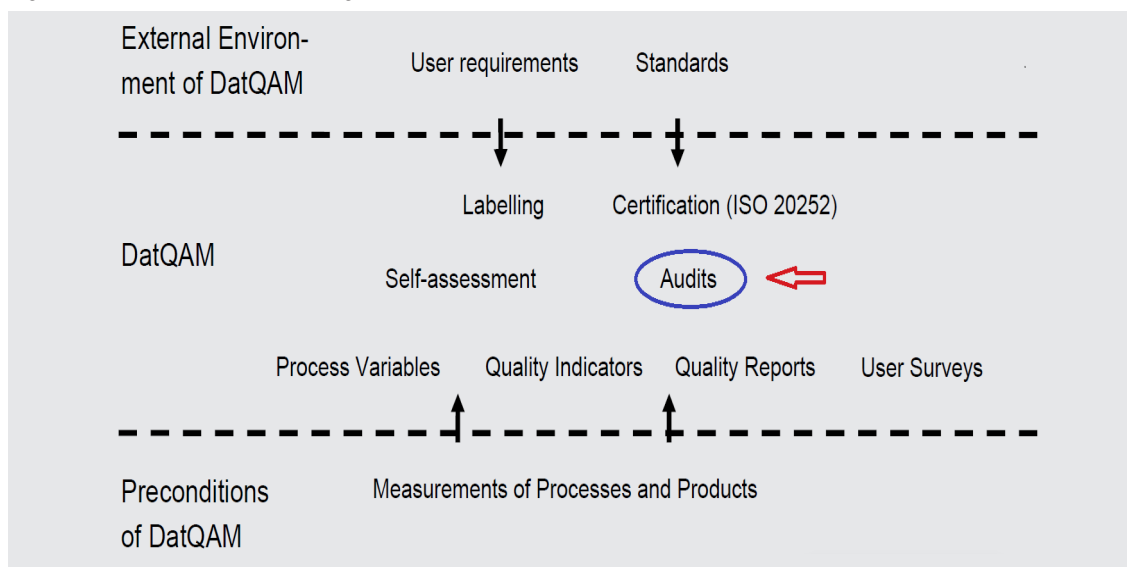
Fonte: Istat

Nell'ambito dei sistemi di valutazione presentati nel *Handbook on data quality assessment methods and tools* (Ehling e Körner 2007) l'*audit* è classificato come metodo di secondo livello (Figura 3.2) perché:

3. Metodi per la valutazione della qualità

- non valuta direttamente la qualità dei dati, ma dei processi. Esamina se i metodi, le procedure e i controlli messi in atto per produrre dati statistici rispettino standard, linee guida e normative. Questo approccio è indiretto in quanto presume che buoni processi producano buoni dati;
- ha una natura sistemica e qualitativa; non si basa su indicatori quantitativi, ma su valutazioni qualitative del sistema produttivo. Di conseguenza, non misura direttamente i singoli aspetti della qualità del dato, ma il contesto e le condizioni in cui i dati vengono generati;
- supporta, ma non sostituisce, le misure dirette. Viene considerato uno strumento complementare ai metodi di primo livello, come gli indicatori di qualità o i *quality report*. Questi ultimi forniscono evidenze dirette sulla qualità dei dati statistici, mentre l'*audit* fornisce una valutazione del livello di affidabilità dell'intero processo;
- è utile soprattutto per la valutazione complessiva e il miglioramento continuo.

Figura 3.2 - Mappa dei metodi e degli strumenti di valutazione della qualità



Fonte: Ehling e Körner (2007)

Si distinguono inoltre due principali forme di *audit*:

1. *audit interni*, condotti dall'organizzazione stessa per valutare la qualità dei propri processi e prodotti statistici;
2. *audit esterni*, eseguiti da esperti esterni o da altre istituzioni statistiche per fornire una valutazione indipendente della qualità.

La nuova procedura di *audit* in Istat

Nel 2023 l'Istat ha definito una nuova procedura di *audit* interno, da applicare ai processi di tipo tradizionale, che mira a verificare l'aderenza ai principi delle *Linee guida della Qualità* (Brancato 2016 e Signore *et al.* 2012), e a promuovere il miglioramento continuo nei processi produttivi. L'*audit* ha lo scopo di migliorare la qualità dei processi statistici, identificare punti di forza e criticità, favorire la diffusione delle buone pratiche e stimolare azioni correttive di miglioramento.

La nuova procedura vuole anche rappresentare uno strumento di verifica dell'integrazione tra le strutture tematiche e trasversali dell'Istituto coinvolte nella produzione statistica.

Gli attori coinvolti nella procedura sono:

- il responsabile tematico, che guida il processo statistico e ne è responsabile;
- i responsabili trasversali, che curano fasi specifiche del processo come, ad esempio, la raccolta dati o la fase di controllo e correzione;
- gli auditori, esperti interni con competenze in ciascuna delle macroaree in cui si articola il processo: esperto tematico, esperto di raccolta dati, esperto metodologo, esperto informatico, esperto di diffusione;
- un servizio di coordinamento che organizza e supporta l'attuazione della procedura.

Le fasi della procedura di *audit* consistono in una programmazione triennale che prevede l'*audit* di tre processi di produzione all'anno. La procedura si articola in tre macrofasi.

1. Selezione dei processi e formazione dei team di auditori

In questa fase, i direttori responsabili della produzione identificano una rosa di processi candidati per l'*audit* e, insieme ai direttori responsabili delle direzioni trasversali individuano esperti che possano rivestire il ruolo di auditori nelle varie aree in cui si articola il processo. Il Comitato Qualità, che sovrintende e coordina le attività relative alla qualità, seleziona i processi da sottoporre ad *audit* e forma i *team* di auditori per ciascun processo selezionato. Viene, quindi, organizzata una giornata di formazione per tutti gli attori coinvolti in cui si illustrano la procedura e gli strumenti a disposizione, e si definiscono i tempi e le modalità di realizzazione.

La selezione dei processi è basata sui seguenti criteri:

- rilevanza dei processi, ad esempio processi statistici legati all'attuazione della normativa dell'Unione europea e nazionale;
- presenza di criticità emerse dall'applicazione di altri sistemi di valutazione, ad esempio dalla *checklist* (cfr. paragrafo 3.1.1);
- processi multifonte, per esempio che utilizzano dati da indagine e da fonte amministrativa;
- presenza di criticità legate all'utilizzazione di più servizi trasversali che devono coordinarsi tra loro.

Sono esclusi dalla procedura di selezione processi la cui ultima edizione risale ai cinque anni precedenti l'inizio della procedura, quelli cessati e quelli pluriennali in corso di svolgimento nell'anno dell'*audit*. La procedura di selezione si conclude con l'individuazione di tre processi per ogni anno.

2. Realizzazione

In questa fase gli auditori studiano la documentazione di riferimento per la conduzione dell'*audit* (manuale di descrizione della procedura, linee guida e standard istituzionali) e il materiale disponibile sul processo (documentazione generale, indicatori di qualità, note metodologiche, prodotti di diffusione) e preparano l'intervista in base a una griglia di domande che approfondisce tutte le fasi del processo.

In generale, nel caso in cui sia stata precedentemente somministrata una *checklist* di valutazione, quest'ultima fornirà la base per l'intervista di approfondimento dell'*audit*. La fase di studio preliminare si conclude con la pianificazione e la realizzazione dell'intervista ai responsabili di processo.

Si passa quindi alla redazione del *report* di valutazione: la stesura della prima parte, che include la sintesi dell'intervista con punti di forza e le aree di criticità individuate, è a cura degli auditori; la seconda parte, che individua le azioni di miglioramento, è a cura dei responsabili di processo.

3. Metodi per la valutazione della qualità

3. Approvazione e monitoraggio

I direttori responsabili del processo prendono visione del *report* e approvano le azioni di miglioramento in accordo con i direttori delle direzioni dei servizi trasversali nel caso in cui le loro direzioni fossero coinvolte nelle azioni di miglioramento. Le azioni di miglioramento vengono inserite nella pianificazione operativa.

Successivamente alla programmazione delle azioni di miglioramento la struttura di coordinamento ne monitora semestralmente l'attuazione.

A conclusione della procedura, i *report* di valutazione sono presentati al Comitato Qualità che viene periodicamente informato sull'andamento del monitoraggio delle azioni di miglioramento.

Gli strumenti utilizzati come riferimenti sono:

- *Linee guida per la Qualità* (Brancato 2016 e Signore *et al.* 2012) che definiscono principi e metodi per garantire la qualità lungo le fasi del processo statistico e costituiscono un riferimento per individuare gli eventuali punti critici del processo produttivo esaminato. Si può inoltre considerare l'aderenza a standard internazionali e interni esistenti/de facto;
- *Checklist* che descrive le fasi del processo; strumento strutturato per valutare l'adozione di metodologie e pratiche consolidate;
- *Griglia strutturata di domande* per la conduzione dell'intervista, che investiga e approfondisce le fasi del processo;
- *Report di valutazione*, che sintetizza gli esiti dell'*audit* e propone azioni di miglioramento (Prospetto 3.3).

Prospetto 3.3 - Struttura del *report* di valutazione

Sintesi Breve riassunto che indichi i principali punti di forza e gli aspetti che necessitano di azioni di miglioramento.			
Risultati La relazione deve illustrare i principali punti di forza e gli aspetti che necessitano azioni di miglioramento facendo riferimento ai vari aspetti esaminati:			
A. Esigenze informative degli utenti/definizione dei concetti B. Lista di riferimento e disegno di campionamento C. Acquisizione dei dati D. Integrazione E. Codifica e classificazione F. Identificazione e trattamento degli errori non campionari G. Derivazione unità H. Derivazione variabili I. Calcolo dei pesi e degli aggregati L. Validazione dei risultati M. Tutela della riservatezza N. Diffusione O. Documentazione e valutazione			
Azioni di miglioramento Indicare quali aspetti del processo richiedono un intervento, la problematica individuata e l'eventuale azione di miglioramento proposta, indicare anche la necessità di supporto esterno alla direzione/dipartimento.			
Sezione (team di audit)	Problematica individuata (team di audit)	Azione di miglioramento proposta e strutture coinvolte (responsabili processo)	Tempistica (direzione tematica del processo sentiti eventualmente i direttori trasversali)
...	...	...	...
...	...	...	...

Fonte: Istat

Per concludere, si riportano nel Prospetto 3.4 alcuni suggerimenti per le attività che il team di auditori è chiamato a svolgere nelle diverse fasi della procedura. La sequenza parte da un'accurata preparazione dell'intervista e si conclude con la definizione e il monitoraggio delle azioni correttive.



Prospetto 3.4 - Suggerimenti per il *team* di auditori

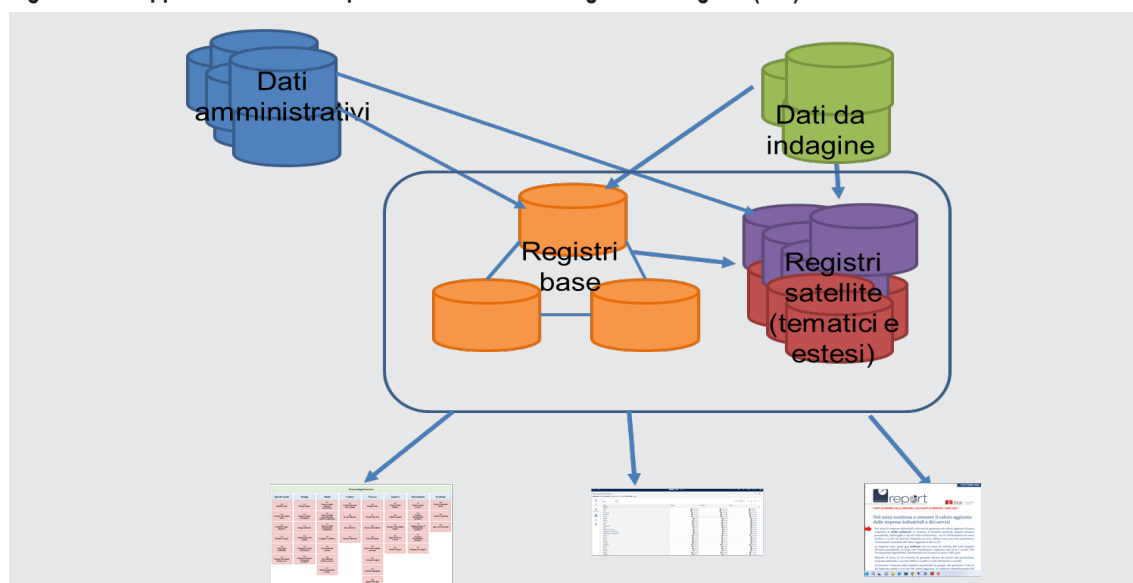
Fase dell'attività	Azioni da compiere	Output
Preparazione dell'intervista	Riunioni preparatorie per definire la traccia delle domande da strutturare secondo le fasi del processo da esaminare Invio delle domande ai responsabili del processo	Elenco strutturato di domande per l'intervista
Intervista	Svolgere l'intervista in una o due mezze giornate; registrare lo svolgimento dell'intervista	Registrazione e trascrizione
Redazione del <i>report</i>	Redigere la prima parte del <i>report</i> con la sintesi dell'intervista e l'individuazione di punti di forza e aree di miglioramento entro 15 giorni dallo svolgimento dell'intervista; decidere gli aspetti da evidenziare in accordo con i responsabili del processo	
Azioni di miglioramento	Individuare azioni concrete e con una tempistica ben definita	Diagramma di Gantt per l'implementazione delle azioni e il loro monitoraggio

Fonte: Istat

Particolare rilievo assume la fase preparatoria, nella quale la definizione condivisa delle domande e il loro invio anticipato ai responsabili del processo favoriscono una maggiore efficacia dell'intervista e consentono agli interlocutori di predisporre la documentazione necessaria. L'intervista rappresenta il momento centrale della raccolta delle evidenze e la registrazione del suo svolgimento garantisce la tracciabilità delle informazioni raccolte e facilita la successiva attività di analisi

3.2 La qualità dei registri del Sistema integrato dei registri

La sempre crescente disponibilità di dati amministrativi e da altre fonti non progettate a fini statistici ha aperto nuove frontiere per la statistica ufficiale, sia a supporto della produzione delle statistiche già esistenti, sia per rispondere a nuove esigenze informative su fenomeni emergenti. Questo ha comportato la necessità di riorganizzare i processi correnti, sempre più di tipo multifonte, gestendo l'integrazione di dati di diversa natura, che possono richiedere specifiche analisi, elaborazioni e trasformazioni per potere essere utilizzati a fini statistici secondo i requisiti di qualità alla base della produzione statistica ufficiale.

Figura 3.3 - Rappresentazione semplificata del Sistema integrato dei registri (SIR)

Fonte: Istat

3. Metodi per la valutazione della qualità

In questa ottica, il Sistema integrato dei registri (SIR) è stato realizzato per mettere a sistema dati derivanti da una pluralità di fonti, in particolare modo di tipo amministrativo, eventualmente combinati con dati provenienti da indagini dirette (Figura 3.3).

L'obiettivo è organizzare in modo strutturato e coerente le informazioni relative a specifiche popolazioni statistiche e ai fenomeni a esse associati, così da disporre di un sistema informativo statistico rilevante sia di per sé sia quale supporto a diversi processi statistici. Le basi metodologiche e operative per la costruzione del SIR sono state delineate con riferimento ad approcci condivisi a livello internazionale tra gli Istituti nazionali di statistica e le organizzazioni internazionali, quali Eurostat e UNECE (UNECE 2007; Wallgren e Wallgren 2014).

Il SIR consiste in registri statistici di diverse tipologie (base e satellite, questi ultimi distinti a loro volta in estesi e tematici) coerenti tra loro, per le cui definizioni si rinvia ad Ascari *et al.* 2023, che riprende e adatta al contesto dell'Istat l'impianto definitorio disponibile in letteratura (Wallgren e Wallgren 2014).

Si richiama di seguito la definizione più generale di registro statistico: “raccolta sistematica, strutturata, regolarmente aggiornata e autorizzata di dati e metadati (proprietà) a livello di unità o evento (oggetto) per una specifica popolazione effettuata esclusivamente per finalità di statistica ufficiale. Il registro statistico è realizzato da un ente facente parte del Sistan. Gli oggetti del registro sono determinati da definizioni e classificazioni che derivano da criteri statistici, connessi alle esigenze della statistica ufficiale, e sono identificati in maniera inequivocabile da un codice univoco. I registri statistici sono rappresentati da unità statistiche caratterizzate da variabili statistiche. Il registro statistico può essere creato e aggiornato a partire da una o più fonti anche combinate tra loro: registri amministrativi, registri statistici e indagini statistiche. Un registro statistico può essere produttore di output statistici in modo diretto oppure costituire la base di riferimento a supporto di altri processi che producono informazioni statistiche” (Ascari *et al.* 2023).

Rispetto alle diverse tipologie, in estrema sintesi, i registri base hanno l'obiettivo di definire le popolazioni di riferimento (individui, imprese e luoghi); i registri estesi arricchiscono i registri base di ulteriori contenuti, cioè variabili; i registri tematici, oltre ad avere l'obiettivo di aggiungere contenuti, hanno il compito di costruire unità derivate rispetto a quelle dei registri base, necessarie alla formalizzazione di alcuni contenuti specifici per la loro struttura multidimensionale; un esempio tipico è quello della posizione lavorativa a cui fa riferimento il Registro tematico del lavoro.

La realizzazione del SIR ha gettato le basi per un nuovo paradigma nella produzione statistica dell'Istat: ai processi di indagine diretta si affiancano processi statistici basati sull'integrazione di fonti amministrative e di indagine per la realizzazione di registri statistici.

Già nei primi anni di avvio del SIR, è stato evidente come tali processi presentassero caratteristiche peculiari e specifiche problematiche. Ad esempio, le variabili che compongono uno stesso registro possono essere originate dall'elaborazione di fonti diverse attraverso processi tra loro disgiunti e diversi. Questa evidenza ha fatto emergere la complessità di monitorare il processo di produzione di un registro statistico in modo univoco e di conseguenza la necessità di disegnare uno strumento di monitoraggio e documentazione della qualità adeguatamente flessibile per gestire tale complessità.

In questa ottica, l'Istat ha definito un *framework* innovativo specifico per la documentazione e il monitoraggio della qualità del SIR, che affianca ed estende i sistemi di monitoraggio e documentazione standard dei processi statistici adottati in Istituto. Si sottolinea che, rispetto alle diverse fasi del processo statistico, la creazione del registro statistico si

pone a valle della raccolta centralizzata dei dati amministrativi nonché della loro pseudonimizzazione, effettuata per garantire il rispetto della normativa sulla *privacy*, e a valle della validazione dei risultati (in genere microdati) delle indagini che vengono utilizzate come fonti. Di conseguenza, anche il *framework* di qualità si innesta nello stesso momento del processo, facendo riferimento alla QRCA (cfr. paragrafo 2.3) e a Sidi-SIQual, in futuro Metastat (cfr. paragrafo 2.1), per il corredo dei metadati ed eventuali misurazioni di qualità delle fonti di input.

L'obiettivo primario della definizione del *framework* era disporre di un sistema di indicatori utili al monitoraggio e alla valutazione della qualità dei processi di produzione e dei prodotti dei registri del SIR, anche in un'ottica di alimentazione del nuovo sistema di metadati in corso di sviluppo Metastat. Tale sistema di indicatori è stato poi corredato da un *set* di metadati che li rende interpretabili oltre che garantire la tracciabilità del processo stesso.

A seguire si illustrano le caratteristiche principali del *framework*, i criteri su cui si basa e gli strumenti che mette a disposizione per la mappatura dei processi e la misurazione della loro qualità. Infine, si condividono alcune riflessioni che accompagnano il lavoro di implementazione operativa e il rilascio sistematico della documentazione e del calcolo degli indicatori.

3.2.1 QSIR, il framework per la qualità del SIR

Il *framework* per la qualità del SIR (QSIR) si presenta come una proposta organica, strutturata e standardizzata per la documentazione dei registri del SIR nell'ottica del monitoraggio e della valutazione della qualità. Il suo disegno teorico completo è stato rilasciato a seguito di analisi approfondite dei processi alla base della produzione dei registri, nonché di applicazioni sperimentali effettuate al fine di testare ed eventualmente aggiungere ulteriori innovazioni e miglioramenti rispetto al disegno iniziale (Ascari *et al.* 2023). Il QSIR è infatti il risultato del lavoro di diversi gruppi interni all'Istat, che hanno messo a sistema le competenze trasversali necessarie per definire, testare, migliorare e implementare un quadro di riferimento complesso e articolato, capace di rispondere sia alle esigenze di documentazione e valutazione ex post della qualità sia a quelle dei responsabili dei registri, fornendo uno strumento operativo di supporto al monitoraggio corrente dei processi.

Nella definizione del QSIR si è fatto riferimento ai modelli standard sviluppati in ambito UNECE, cioè il GSBPM e il GSIM (cfr. paragrafo 2.1.3).

Il *framework* include:

- un *set* di metadati generali con informazioni anagrafiche, sugli obiettivi e sulle fonti del singolo registro statistico;
- un modello per la documentazione della qualità delle diverse fasi del processo del singolo registro, che include gli elementi di metadattazione e specifici indicatori di qualità per ogni fase;
- alcune proposte di metodi per la valutazione della qualità dell'output prodotto dal registro, applicabili in casi specifici;
- alcune proposte di indicatori per valutare la coerenza tra i diversi registri del SIR, considerati come possibili strumenti di validazione degli output.

Il nucleo centrale più consolidato è certamente quello relativo al modello di metadattazione e agli indicatori di qualità per le diverse fasi del processo. Il *framework* è stato in effetti costruito a partire dalla mappatura dei processi, finalizzata sia alla documentazione sia all'individuazione degli indicatori di qualità più appropriati per ogni singola fase del processo.

3. Metodi per la valutazione della qualità

Sono stati identificati i sottoprocessi GSBPM di maggior rilievo nei processi di produzione e di aggiornamento periodico di un registro statistico. Per ogni sottoprocesso, sono stati poi identificati i relativi indicatori di qualità, i metadati per documentarlo e per calcolare e interpretare correttamente gli indicatori stessi.

Per consentire l'implementazione degli indicatori il più possibile automatizzata, sono stati, infine, definiti chiaramente gli input e gli output necessari e le formule di calcolo.

Nel Prospetto 3.5 sono elencati i sottoprocessi che più si adattano e rappresentano i flussi dei vari registri e i corrispettivi del GSBPM.

Prospetto 3.5 - Sottoprocessi rilevanti per i registri del SIR e mappatura con GSBPM

SOTTOPROCESSO	CORRISPONDENZA CON GSBPM
Controllo della disponibilità delle fonti	<i>Check availability</i>
Acquisizione dei dati	<i>Run Collection</i>
Primi controlli sul <i>dataset</i>	<i>Review and validate</i>
Integrazione	<i>Data integration</i>
Ricodifica	<i>Classify and code</i>
Controllo e correzione	<i>Edit and impute</i>
Derivazione nuove unità e variabili	<i>Derive new variables and units</i>
Calcolo degli aggregati	<i>Calculate weights; calculate aggregates</i>
Validazione degli output	<i>Validate outputs</i>

Fonte: Istat

È importante sottolineare che in alcuni casi, tra le descrizioni dei sottoprocessi nel modello GSBPM e quelle relative a un registro statistico esiste un'elevata corrispondenza, mentre in altri casi sono state estese e riadattate, per meglio rappresentare alcune attività e controlli rilevanti in particolare nell'elaborazione dei dati amministrativi.

Come già sottolineato, obiettivo del QSIR è supportare il monitoraggio del processo di alimentazione e di aggiornamento corrente del registro e non il disegno iniziale del processo. È stato tuttavia mantenuto il sottoprocesso di controllo della disponibilità delle fonti perché può succedere che una fonte arrivi in ritardo rispetto a quanto pianificato o che non sia più disponibile, ed è importante essere in grado di valutare i possibili effetti sul registro di tali eventi. Deve essere, inoltre, continua l'attività di *scouting* per identificare nuove fonti che possono migliorare la qualità delle statistiche prodotte (ad esempio in termini di copertura della popolazione obiettivo del registro o dei fenomeni in esso disponibili).

Per quanto riguarda invece la fase di acquisizione dei dati, si fa riferimento ai sistemi già preposti, QRCA e Sidi-SIQual (cfr. paragrafo 2.3), per la valutazione della qualità dell'input per sé (cioè rispetto al suo obiettivo originario amministrativo o statistico), e si propongono ulteriori misure di qualità orientate a valutare l'adeguatezza dell'input a rappresentare correttamente i fenomeni statistici obiettivo del registro (Rocci *et al.* 2022).

Lo studio dei dati amministrativi in relazione al loro potenziale utilizzo statistico, prima della fase di integrazione, rappresenta di norma una parte rilevante delle analisi svolte nella progettazione del processo stesso. Il sottoprocesso relativo ai primi controlli sul *dataset* rappresenta, da una parte, una validazione dei dati di input e, dall'altra, una prima preparazione dei dati per il successivo trattamento; si tratta di un passaggio fondamentale per tutti i registri analizzati, sebbene non sia facilmente mappabile nel GSBPM.

Il modello, infatti, tende a non esplicitare questo passaggio del processo come sottoprocesso autonomo, ma include alcuni controlli in diversi sottoprocessi, ad esempio la deduplicazione nella fase di integrazione e il controllo dei valori mancanti nel sottoprocesso di controllo e correzione.

Seguono i sottoprocessi tradizionali della fase di trattamento dati, tra cui assume particolare rilevanza quello di integrazione, ma anche gli altri sottoprocessi possono assumere un peso considerevole e un'accezione diversa rispetto a quanto rappresentano nei processi di indagine: si pensi alla rilevanza che può assumere la ricodifica dei dati per armonizzare le classificazioni adottate nelle diverse fonti amministrative, oppure alla derivazione di nuove unità o nuove variabili, come avviene spesso nei registri satellite.

I sottoprocessi di calcolo degli aggregati e validazione degli output rappresentano, infine, l'estensione del *framework* per includere anche proposte iniziali di metodi e misure per valutare la qualità delle stime prodotte da registro e indicatori di coerenza tra registri. Si può dire che al momento rappresentino dei *placeholder*, con indicatori applicabili solo in scenari specifici, in vista di futuri sviluppi teorici.

È importante evidenziare che, nonostante i sottoprocessi siano presentati come sequenziali, la loro implementazione segue spesso un diverso ordine così come avviene per lo standard GSBPM. Può infatti accadere che alcuni sottoprocessi vengano svolti prima o contestualmente ad altri che sono rappresentati successivamente, che alcuni sottoprocessi non si applichino o che debbano essere ripetuti più volte. Pertanto, quando si descrive un processo statistico secondo i sottoprocessi del QSIR è necessario anche definire e descrivere in modo dettagliato il *workflow* che lega i vari sottoprocessi: quali sono quelli applicati, la sequenza tra essi e se essa è condizionata a qualche regola specifica.

3.2.2 Sviluppo delle schede e alcuni esempi

Il GSBPM fornisce per ogni sottoprocesso la descrizione delle azioni che ci si aspetta possano essere fatte. È importante sottolineare che nell'ambito del QSIR tali descrizioni sono state contestualizzate con descrizioni quanto più possibile adattate ai processi dei registri.

Per permettere la documentazione del processo in modo standardizzato è stato disegnato un *template* strutturato, che individua gli elementi di metadatazione utili a rappresentare le caratteristiche di ogni sottoprocesso con i relativi input e output (Prospetto 3.6). Il *template* prende spunto da un lavoro che integra i due modelli standard GSBPM e GSIM (Rizzolo *et al.* 2021).

Il *template* generico è stato poi declinato per ognuno dei sottoprocessi GSBPM ritenuti rilevanti per i registri creando un *set* di schede di metadatazione di riferimento per l'applicazione a ogni registro. Per ogni sottoprocesso, nella voce Qualità e *performance* sono definiti gli indicatori di qualità (corredati dalle formule di calcolo non riportate per brevità) utili al monitoraggio e alla valutazione del sottoprocesso stesso. I risultati raccolti rappresentano essi stessi parte della documentazione del processo, ma possono anche fornire indicazioni per il miglioramento dei metodi statistici applicati e quindi della qualità sia del processo sia del prodotto.

Il modello permette di organizzare e sistematizzare gli appropriati metadati strutturali e referenziali e gli indicatori di qualità utili per il monitoraggio e la valutazione di ogni passo del processo.

Il *template*, le schede di metadatazione e gli indicatori di qualità sono il risultato di un accurato processo di verifica sperimentale sulle diverse tipologie di registro statistico del SIR.

3. Metodi per la valutazione della qualità

67

Prospetto 3.6 - Struttura del *template* per la documentazione del sottoprocesso

Macroelemento	Nome oggetto (corrispondente GSIM in parentesi)	Definizione ed esempio
Input	Input principale (<i>Core input</i>)	Oggetti forniti in input al sottoprocesso per essere da esso modificati. Esempio: nel sottoprocesso di controllo e correzione, il <i>dataset</i> (con la sua struttura) che è sottoposto a controllo e correzione
	Parametri (<i>Parameter</i>)	Oggetti forniti in input al sottoprocesso per configurare il sottoprocesso stesso. Esempio: i parametri di un modello di stima
	Input ausiliario (<i>Process support input</i>)	Oggetti in input al sottoprocesso, necessari ma non soggetti a modifica da parte di quest'ultimo. Esempio: la classificazione standard usata in un sottoprocesso di codifica
Sottoprocesso	Obiettivo (<i>Business function</i>)	Obiettivo del sottoprocesso o attività. Esempio: "Ottenere un <i>dataset</i> completo e coerente" o più di dettaglio "Correggere gli errori sistematici"
	Fase GSBPM (<i>Business process GSIM</i>)	Concetto GSI che si può mappare con la fase GSBPM e che per definizione è composto da una serie di sottoprocessi o attività (<i>process step</i>). Esempio: "Trattamento" (<i>Process</i>)
	Sottoprocesso GSBPM o attività (<i>Process step</i>)	Sottoprocesso (GSBPM) o attività più elementare che realizza un obiettivo specifico. Può a sua volta essere formato da altri passi a un maggior livello di dettaglio. Esempio: "Controllo e correzione" o più di dettaglio "Individuazione degli outlier"
	Metodologia (<i>Process method</i>)	Descrizione della metodologia usata. Esempio: nel controllo e correzione, applicazione del metodo del donatore
	Regola (<i>Rule</i>)	Regola algebrica o matematica che si può applicare nel metodo o per gestire il passaggio da un sottoprocesso al successivo. Esempio: se $x_1 < V_1 < x_2$ then $V_1 = \text{errore}$
	Software utilizzato	Applicativo utilizzato per svolgere l'attività o linguaggio in cui è stata sviluppata una procedura ad hoc. Esempio: pacchetto statistico specifico oppure procedura (in SAS, R, ...) ad hoc
Output	Output principale (<i>Core output</i>)	Oggetti creati o modificati dal sottoprocesso. Esempio: Il <i>dataset</i> imputato o una nuova variabile creata
	Qualità e performance (<i>Process metric</i>)	Sottoprodotto del sottoprocesso, utile a documentare, misurare e valutare l'esecuzione del sottoprocesso stesso. Esempio: Tasso di imputazione
	Log della procedura (<i>Process execution log</i>)	Rapporto di esecuzione del sottoprocesso, essenzialmente legato all'esecuzione pratica (<i>log</i> del programma, tempo di esecuzione della procedura). I dati registrati nel <i>log</i> della procedura possono essere utilizzati per la costruzione di una process metric. Esempio: <i>log</i> dell'esecuzione di Selemix

Fonte: Istat

Di seguito sono riportate delle schede di esempio, selezionate tra quelle più rappresentative delle peculiarità dei processi alla base dei registri del SIR.

Sottoprocesso: primi controlli sul *dataset* (Prospetto 3.7)

La fase iniziale di analisi dei dati amministrativi che alimentano un registro statistico è di fondamentale importanza, non solo in fase di disegno e prima realizzazione del registro stesso, ma anche nella produzione delle sue successive edizioni: infatti, oltre alla necessità di una efficace comunicazione tra gli enti su eventuali cambiamenti e/o ritardi dei dati, la validazione iniziale dei dati riveste un ruolo cruciale preliminare alle elaborazioni al fine di accertare che i prerequisiti di completezza e copertura attesi siano soddisfatti. Nei complessi processi di produzione dei registri, gli stessi controlli possono essere rimandati o ripetuti sui *dataset* frutto di integrazione di due o più fonti.

Descrizione

In questo sottoprocesso vengono effettuati i primi controlli e valutazioni su un *dataset*, che può essere proveniente dalle singole fonti oppure essere un *dataset* integrato. Viene effettuata la deduplicazione e vengono identificati ed eliminati i *record* non utilizzabili in



quanto completamente mancanti delle informazioni. Vengono verificati i valori mancanti delle variabili. Se è disponibile una fonte ausiliaria appropriata (ad esempio nel caso dei registri estesi la lista delle unità della popolazione è disponibile dal registro base), può essere stimata la copertura del *dataset* in esame, sia sulla base di confronti aggregati sia su abbinamenti a livello micro. Alla stima generale della copertura si affianca la valutazione della copertura rispetto a sottopopolazioni con caratteristiche particolari, ad esempio: nel caso della popolazione delle imprese italiane, presenza o assenza di imprese di grandi dimensioni; nel caso della popolazione dei percorsi di formazione, presenza o assenza di università con corsi di laurea unici o rari, eccetera.

Gli indicatori calcolati in questa fase sulle singole fonti, in particolare quelli relativi alla deduplicazione, possono essere un utile *feedback* per la fase di identificazione delle unità effettuata a monte dell'acquisizione dei dati per la costruzione del registro per l'attribuzione del codice pseudonimo, visto che tali errori possono essere propri della fonte o dovuti al processo di pseudonimizzazione.

Prospetto 3.7 - Scheda di metadatazione e indicatori di qualità per il sottoprocesso "Primi controlli sul *dataset*"

Macroelemento	Nome oggetto	Elemento o valore
Input	Input principale	<i>Dataset</i> da revisionare e/o valutare (fonte o integrato), (unità e variabili)
	Parametri	Chiavi di <i>linkage</i> (pseudonimo o codice identificativo)
	Input ausiliario	<i>Dataset</i> di riferimento (<i>benchmark</i> : potrebbe essere un registro base per un registro esteso, il registro di riferimento se si sta valutando la copertura di una specifica variabile, un'indagine di controllo per un registro base)
		Definizione popolazione <i>dataset</i> da valutare
Sottoprocesso	Obiettivo	Deduplicare <i>dataset</i> , identificare <i>record</i> inutilizzabili verificare dati mancanti e per quali variabili, verificare copertura e rappresentatività del <i>dataset</i> rispetto a <i>benchmark</i> .
	Fase GSBPM	Trattamento (5. <i>Process in GSBPM</i>) o 8. <i>Evaluate</i>
	Sottoprocesso GSBPM/attività	5.3 <i>Review and Validate</i>
	Metodologia	Identificazione duplicazioni, identificazione <i>record</i> inutilizzabili, Stima della copertura attraverso confronti aggregati, stima della copertura attraverso abbinamento dei microdati, regole per la costruzione di tassi di valori mancanti per variabile.
	Regola	Regola di identificazione delle deduplicazioni Regola di identificazione dei valori mancanti Modello di integrazione, Relazione 1-1, n-1, n-n
	Software utilizzato	Procedure ad hoc (ad esempio SAS, R), Pacchetti generalizzati come <i>Fastlink</i> , <i>StatMatch</i>
Output	Output principale	<i>Dataset1</i> al netto delle duplicazioni e dei <i>record</i> inutilizzabili
	Qualità e <i>performance</i>	<u>Indicatori relativi alla deduplicazione</u> 3.1. Percentuale di <i>record</i> duplicati 3.2. Percentuali di <i>record</i> duplicati sulla chiave di aggancio 3.3. Percentuale di <i>record</i> duplicati su un sottoinsieme di variabili rilevanti 3.4. Discordanza dell'informazione presente tra i <i>record</i> doppi <u>Indicatori relativi ai valori mancanti</u> 3.5. Tasso di <i>missing</i> per le principali variabili 3.6. Percentuali di <i>record</i> eliminati perché non utilizzabili (eventualmente per motivo) <u>Indicatori relativi a copertura e rappresentatività</u> 3.7. Stima della copertura del <i>dataset</i> rispetto al <i>dataset</i> di riferimento 3.8. Confronto tra statistiche (medie, totali, eccetera) e distribuzioni di variabili tra <i>dataset</i> da valutare e relativa sottopopolazione nel <i>dataset</i> di riferimento 3.9. Confronto tra statistiche (medie, totali, eccetera) e distribuzioni di variabili tra <i>dataset</i> da valutare e la popolazione totale del <i>dataset</i> di riferimento 3.10. Sottocopertura fonti rispetto al <i>dataset</i> di riferimento
	Log della procedura	Tempi di elaborazione

Fonte: Istat

3. Metodi per la valutazione della qualità

Sottoprocesso: integrazione (Prospetto 3.8)

La complessità dei processi di produzione dei registri nasce dalla loro natura multifonte, pertanto il sottoprocesso di integrazione dei dati è generalmente centrale nella costituzione del registro. Da notare che tra gli indicatori di qualità proposti per questo sottoprocesso ne sono stati definiti anche due sulla copertura della singola fonte amministrativa di input rispetto alla popolazione target. Questi indicatori sono stati introdotti con l'obiettivo di fornire un *feedback* al sistema QRCA (cfr. paragrafo 2.3) che valuta la qualità delle fonti amministrative, con l'obiettivo di innescare un circolo virtuoso di miglioramento continuo della qualità.

Descrizione

In questo sottoprocesso avviene l'integrazione tra le fonti che alimentano il registro con l'obiettivo di costruire un *dataset* più completo possibile rispetto alle unità e/o alle variabili obiettivo del registro stesso, quindi anche con l'obiettivo di aumentare la copertura del registro integrando fonti diverse.

Prospetto 3.8 - Scheda di metadatazione e indicatori di qualità per il sottoprocesso "Integrazione"

Macroelemento	Nome oggetto	Elemento o valore
Input	Input principale	<i>Dataset1, Dataset2...</i> (unità e variabili)
	Parametri	Soglie, chiavi di <i>linkage</i> , variabili di <i>blocking</i>
	Input ausiliario	Ulteriori variabili identificative oltre alle chiavi o di controllo per l'abbinamento. Definizioni delle popolazioni di riferimento, delle variabili, informazioni sul processo di costruzione delle variabili e loro riferimenti temporali
Sottoprocesso	Obiettivo	Integrazione orizzontale, verticale o sia orizzontale che verticale
	Fase GSBPM	Trattamento (5. <i>Process</i> in GSBPM)
	Sottoprocesso GSBPM/attività	Integrazione (5.1. <i>Integrate</i> in GSBPM)
	Metodologia	<i>Record linkage</i> deterministico, deterministico gerarchico, probabilistico, <i>privacy preserving and predictive linkages (classification or regression techniques)</i> , <i>statistical matching</i> Procedure di <i>appending</i> <i>Data pooling</i> Strategia di integrazione sulla base di un ordine di priorità fra le fonti
	Regola	Modello di integrazione Regole di scelta gerarchica tra fonti Regole di trasformazione delle unità e delle variabili Abbinamento attraverso codici pseudonimi
	Software utilizzato	Procedure ad hoc (per esempio in SAS, R, ...) Pacchetti generalizzati come <i>Fastlink</i> , <i>StatMatch</i>
Output	Output principale	<i>Dataset</i> integrato, <i>Dataset</i> dei <i>record</i> non abbinati;
	Qualità e <i>performance</i>	<u>Indicatori relativi alle <i>performance</i> dell'integrazione</u> 4.1. Valori mancanti o errati nella variabile di <i>linkage</i> 4.2. Tasso di link (o <i>match rate</i>) 4.3. Tasso di falsi link 4.4. Tasso di falsi non-link <u>Indicatori relativi alle unità</u> 4.5. Percentuale di unità derivanti dai differenti <i>dataset</i> di input sul totale delle unità 4.6 Indicatore di sottocopertura del <i>dataset</i> amministrativo 4.7 Indicatore di sovracopertura del <i>dataset</i> amministrativo <u>Indicatori relativi variabili</u> 4.8 Percentuale di variabili derivanti dai differenti <i>dataset</i> di input sul totale delle variabili nel <i>dataset</i> integrato 4.9 Distanze tra distribuzioni sulle variabili nel <i>dataset</i> integrato e nei <i>dataset</i> di input 4.10 Numero variabili "costruite" a valle del processo di integrazione 4.11 Discordanza per i <i>record</i> che si agganciano tra l'informazione presente nelle diverse fonti
	Log della procedura	Tempi di integrazione

Fonte: Istat

L'integrazione può avvenire in diversi momenti del processo di produzione (subito dopo la raccolta dati oppure a valle di altri sottoprocessi di trattamento delle singole fonti); può coinvolgere tutte le fonti contemporaneamente o sottoinsiemi di fonti in sequenza e quindi può essere ripetuta più volte all'interno di un processo produttivo statistico. Possibili obiettivi dell'integrazione sono:

- migliorare la copertura delle unità della popolazione obiettivo del registro (integrazione orizzontale a livello micro);
- aumentare la disponibilità di informazioni sulle variabili obiettivo del registro (integrazione verticale a livello micro);
- integrare i dati in modo sia orizzontale sia verticale a livello micro;
- derivare nuove unità o nuove variabili statistiche non presenti come tali nelle singole fonti amministrative ma ricavabili da esse mediante funzioni che coinvolgono dati di fonti diverse.

L'integrazione può riguardare anche diversi registri del SIR, ad esempio con scopi di validazione, di analisi statistica, eccetera.

Sottoprocesso: validazione degli output e verifica della coerenza del SIR (Prospetto 3.9)

Questo sottoprocesso rappresenta il momento in cui i contenuti di un registro sono messi a confronto con quelli degli altri registri del Sistema. Il *set* di indicatori di coerenza attualmente inserito nel *framework* è applicabile in maniera sufficientemente generalizzata ai vari registri, sebbene le analisi svolte abbiano evidenziato quanto lo studio della coerenza dipenda dagli specifici contenuti tematici, dalle definizioni adottate, dalle fonti e dal processo di costruzione seguito, rafforzando la necessità e evidenziando l'importanza di disporre di una documentazione approfondita di ogni registro per essere in grado di valutare la coerenza del sistema. Il *set* di indicatori per questo sottoprocesso rappresenta una base per i futuri sviluppi metodologici in questa area.

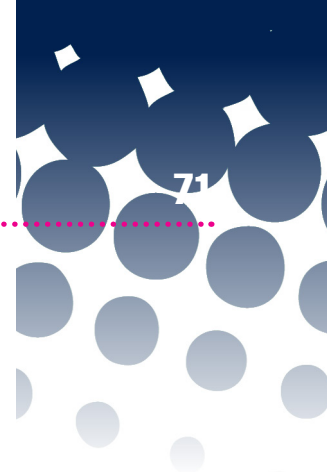
Descrizione

In questo sottoprocesso si verifica la coerenza dei contenuti informativi di un registro (variabili disponibili) con quelli degli altri registri del SIR a esso collegati, in relazione alle popolazioni o sottopopolazioni di riferimento.

In particolare, si verifica:

- i. la coerenza di una stessa variabile presente in più registri del SIR; in questa fase intermedia di realizzazione del SIR, infatti, può ancora avvenire che la stessa variabile sia presente in più registri, condizione che andrà scomparendo man mano che il SIR consolida le sue componenti;
- ii. la coerenza di variabili diverse legate (pseudo)funzionalmente e presenti in diversi registri di SIR.

Assume importanza fondamentale comprendere se, a fronte di uguali concetti, i processi di costruzione o i riferimenti temporali adottati in registri diversi portino a incoerenze tra registri. Si deve valutare, inoltre, se le incoerenze riscontrate possono dipendere dalla fase di identificazione delle unità a monte della costruzione di ciascun registro.



3. Metodi per la valutazione della qualità

Prospetto 3.9 - Scheda di metadattazione e indicatori di qualità per il sottoprocesso “Validazione dei dati”

Macroelemento	Nome oggetto	Elemento o valore
Input	Input principale	<i>Dataset</i> che integra il registro di interesse con uno o più altri registri
	Parametri	(pseudo)funzioni che legano variabili di diversi registri
	Input ausiliario	Definizioni delle popolazioni di riferimento, delle variabili, informazioni sul processo di costruzione delle variabili e loro riferimenti temporali
Sottoprocesso	Obiettivo	Valutare la coerenza del SIR
	Fase GSBPM	Analisi (6. <i>Analyse</i> in GSBPM)
	Sottoprocesso GSBPM/attività	Validazione degli output (6.2 <i>Validate output</i> in GSBPM)
	Metodologia	Verifica della corrispondenza delle definizioni delle popolazioni e delle variabili, Verifica della coerenza dei riferimenti temporali delle variabili in diversi registri, Confronto sui diversi processi di costruzione delle stesse variabili in diversi registri Calcolo e analisi di indicatori di coerenza a livello micro e macro
	Regola	Formule degli indicatori
	Software utilizzato	Procedure ad hoc (per esempio in SAS, R, ...)
Output	Output principale	<i>Dataset</i> validato Prospetti comparativi dei metadati rilevanti: per ogni variabile in analisi, vengono specificate le eventuali differenze nelle definizioni, nel processo di costruzione e nelle fonti per i registri che si comparano
	Qualità e <i>performance</i>	9.1. Coerenza in termini di conteggio dei valori diversi 9.2. Coerenza in termini di distanza relativa tra i valori 9.3. Coerenza in termini di conteggio dei valori diversi oltre un prefissato <i>range</i> 9.4. Coerenza in termini di valori assoluti delle distanze 9.5. Coerenza a livello di aggregati 9.6. Coerenza in termini di unità: copertura di un registro rispetto all'altro 9.7. Incompatibilità tra registri 9.8. Impatto sulle stime delle correzioni effettuate in un registro per renderlo coerente ad altri
	Log della procedura	Tempi di calcolo degli indicatori

Fonte: Istat

3.2.3 Riflessioni sulla implementazione e su possibili sviluppi

Il QSIR nasce come sistema di indicatori di qualità per il monitoraggio e la valutazione dei prodotti e dei processi di produzione dei registri del SIR, anche in un'ottica di alimentazione del nuovo sistema di documentazione dei metadati in via di realizzazione.

Il modello e le schede proposte sono stati analizzati anche da un punto di vista tecnologico-informatico, per porre le basi dell'architettura per l'implementazione, e anche questa attività fornisce *feedback* per migliorare la proposta di *framework*. Per quanto riguarda l'architettura, è stata proposta una soluzione ibrida in cui gli indicatori di qualità per il monitoraggio vengono integrati nei sistemi di gestione dei registri, mentre i metadati e gli indicatori di qualità per la documentazione e la valutazione a posteriori vengono raccolti nel nuovo sistema trasversale di metadati dell'Istat (cfr. paragrafo 2.1).

Le attività di studio condotte nel percorso di definizione del *framework* sono molteplici e hanno contribuito allo sviluppo ed evoluzione del modello fino alla sua forma attuale. I prodotti rilasciati hanno diversi livelli di maturità e hanno permesso in alcuni casi di procedere all'implementazione del calcolo degli indicatori proposti nei processi di produzione di alcuni registri (Filippini *et al.* 2024). In altri casi, ad esempio le prime proposte metodologiche per la valutazione della qualità dell'output, sono applicabili solo sotto specifiche condizioni, ma il loro inserimento nell'impianto complessivo rende possibile estendere facilmente il *framework* integrando eventuali ulteriori sviluppi metodologici (Di Zio *et al.* 2023). In particolare, per quanto riguarda la valutazione della qualità dell'output, è stata prioritariamente analizzata la definizione dei prodotti statistici (registro o stime prodotte a partire dal registro) e la relativa letteratura metodologica; sulla base di tale analisi, sono

state formulate alcune proposte di misure di qualità, sottoposte a specifiche sperimentazioni. Per la coerenza, sono state studiate le relazioni tra registri in termini di chiavi di *linkage* e proposti possibili indicatori di qualità.

Tali proposte sono state integrate nel *framework* teorico complessivo; tuttavia, non si prevede che gli indicatori di coerenza e le misure di qualità dell'output proposti siano implementati come indicatori standard in Metastat, sia perché tali indicatori sono strettamente dipendenti dal tematismo cui ogni registro risponde, sia perché richiedono l'integrazione di diversi registri o di registri e altre fonti, oppure l'applicazione di modelli statistici complessi. Tali indicatori e misure risultano pertanto più adatti a essere applicati in una fase di valutazione *ex post* che nella costruzione corrente del registro. Potranno comunque essere conservati nel sistema di metadati.

I diversi registri su cui viene applicato il *framework* QSIR si trovano attualmente in fasi di sviluppo diverse e, anche se l'architettura del SIR è in larga misura armonizzata, vi sono alcune differenze nella disponibilità di documentazione che hanno spinto a seguire diversi approcci nell'implementazione del QSIR. In particolare, se è già disponibile una dettagliata descrizione del processo, si focalizza l'attenzione sullo sviluppo degli indicatori di qualità e si procede parallelamente alla mappatura della documentazione disponibile alla forma standard delle schede di metadattazione; nel caso in cui non sia disponibile tale documentazione, si investe prima nella produzione delle schede e nel disegno del *workflow*, e solo successivamente si approfondiscono gli indicatori di qualità. In questo contesto è importante il coordinamento delle implementazioni effettuato tra gli esperti che seguono i vari processi (Casagrande *et al.* 2024).

Per ridurre l'onere della sua applicazione sulla produzione corrente, un altro obiettivo perseguito durante l'implementazione del QSIR è di automatizzare il più possibile il processo di calcolo degli indicatori di qualità. Al momento, il calcolo degli indicatori è in fase di automazione, il prossimo passo potrebbe essere la definizione di un sistema di *quality gate* (ABS 2016), fissando soglie per i valori degli indicatori che generano avvisi solo in caso di valori fuori controllo, per facilitare ulteriormente il compito di monitoraggio del personale addetto alla gestione dei registri.

3.3 Lo studio di un *framework* per la valutazione della qualità delle *Trusted Smart Statistics*

Le *Trusted Smart Statistics* (TSS) sono per definizione prodotti statistici generati con l'ausilio di sistemi intelligenti - architetture tecnologiche avanzate che integrano componenti più o meno automatizzati - che devono essere sviluppati secondo principi e metodi della statistica ufficiale, e quindi verificabili e trasparenti.

Il termine "trust" indica la garanzia per l'utente finale che tali prodotti presentino la stessa integrità metodologica delle statistiche tradizionali, disponendo inoltre degli strumenti necessari per verificarne la qualità.

Derivate in gran parte da *big data* (BD), queste statistiche arricchiscono l'offerta informativa tradizionale offrendo risposte rapide e granulari sui fenomeni sociali, ambientali ed economici. Oltre a potenziare la statistica ufficiale, le TSS hanno il vantaggio di avvicinare cittadini e imprese al valore dei dati che producono, promuovendone un coinvolgimento attivo e consapevole, e riducendone il carico statistico. Si favorisce così un rapporto collaborativo tra istituzioni statistiche, cittadini e imprese. La possibilità di riutilizzare fonti digitali prodotte da cittadini e imprese costituisce un vantaggio sia per questi ultimi, in termini di riduzione del carico statistico, sia per l'Istat, in termini di riduzione dei costi di rilevazione.

La produzione di TSS comporta una trasformazione sostanziale del modello tradizionale di produzione statistica. In un contesto in cui le fonti e i processi di elaborazione possono provenire da soggetti esterni – come piattaforme digitali o aziende private – diventa necessario aggiornare gli strumenti metodologici, tecnici, legali e organizzativi, per garantire comunque il rispetto degli standard di qualità del SSE. L'Istat ha pertanto avviato un progetto, internamente alla direzione metodologica, per la definizione di un *framework* di riferimento metodologico rispetto al quale valutare e documentare la qualità delle TSS al fine di una loro integrazione nella produzione statistica ufficiale. Il percorso per la definizione del *framework* non si è ancora concluso, ma sono ormai chiari la struttura e gli elementi che dovranno essere sviluppati, così come le principali criticità che sarà necessario affrontare. Il presente capitolo illustra il lavoro fin qui svolto e i risultati conseguiti in questo ambito.

3.3.1 Aspetti normativi, istituzionali e tecnologici

Le TSS introducono un approccio completamente innovativo per la produzione statistica ufficiale. Le differenze riguardano la natura delle fonti, le tecniche di raccolta ed elaborazione dei dati, le strutture organizzative e gli attori coinvolti. Il *framework* attualmente in fase di definizione è prevalentemente orientato agli aspetti metodologici; tuttavia, è fondamentale tenere in considerazione anche alcuni requisiti di natura normativa, istituzionale e tecnica che ne costituiscono il contesto di riferimento.

La mancanza di una definizione giuridica delle TSS (e dei *big data* stessi) riflette la difficoltà di delinearne con precisione i contenuti, sia dal punto di vista tecnico sia normativo. Si tratta di un ambito in forte evoluzione, soggetto a continui adattamenti dettati dal rapido cambiamento tecnologico. In questo scenario, la recente modifica della legge statistica europea (cfr. paragrafo 1.3 per la revisione del regolamento (CE) n. 223/2009) rappresenta un'opportunità concreta per favorire l'integrazione di queste nuove fonti nei processi ufficiali, tracciando la strada verso una statistica pubblica più moderna, agile e connessa alla realtà digitale. Una delle modifiche più significative al Regolamento statistico europeo riguarda infatti l'introduzione dell'obbligo per i detentori di dati privati (*private data holder*) di rendere accessibili i propri dati per la produzione di statistiche ufficiali, nel rispetto delle normative sulla *privacy* e sulla sicurezza. Sebbene questo obbligo non sia risolutivo delle criticità legate alla qualità e alla disponibilità dei dati, rappresenta un mandato legale fondamentale che rafforza la posizione delle autorità statistiche pubbliche nel dialogo con i soggetti privati. Il mandato legale rappresenta solo un tassello, pur indispensabile, per lo sviluppo delle TSS soprattutto con l'obiettivo di produrre statistiche ufficiali e non soltanto statistiche sperimentali. Il contesto istituzionale entro cui si inserisce questa evoluzione è disciplinato dai principi del *Codice delle statistiche europee*, ad esempio il *Principio 2* sul mandato per la rilevazione e l'accesso ai dati o il *Principio 5* sulla riservatezza statistica e sulla protezione dei dati.

Un altro aspetto fondamentale è rappresentato dalle relazioni con i detentori dei dati. È necessario creare solidi partenariati con questi nuovi attori, ormai con un ruolo fondamentale nella statistica ufficiale, non solo affinché sia consentito l'accesso ai dati e il loro utilizzo per la statistica pubblica, ma anche perché, dovendo spesso delegare a questi soggetti una parte del trattamento dei dati, è necessario instaurare un rapporto di estrema collaborazione e trasparenza.

Trasparenza è una parola chiave. Già con l'uso dei dati amministrativi a fini statistici, gli INS hanno perso il controllo diretto sulla generazione del dato ed è stato necessario

introdurre una serie di controlli sui dati di input. Nel caso delle TSS spesso anche una pre-elaborazione dei dati viene effettuata fuori dal controllo dell'INS ed è quindi necessario definire nuove modalità per il monitoraggio della qualità di tale parte del processo. Questo può essere fatto innanzitutto chiedendo ai fornitori dei dati massima trasparenza rispetto alle elaborazioni effettuate e richiedendo di fornire anche metadati e misure di qualità di accompagnamento. Un approccio ottimale è che sia proprio l'Istituto a fornire le indicazioni metodologiche su come procedere nella fase elaborativa (Eurostat 2023b).

Altro aspetto fondamentale per instaurare e mantenere la fiducia sia con i fornitori dei dati sia, più in generale, con gli utenti, i cittadini e le imprese cui tali dati si riferiscono, è la garanzia della riservatezza e la protezione dei dati. I dati di alcune fonti TSS possono essere considerati molto sensibili, ed è a volte necessario costruire sistemi che garantiscano un maggiore livello di protezione alle informazioni. Questo è uno dei motivi per cui, in alcuni casi, una parte delle elaborazioni deve essere effettuata dal fornitore dei dati, in quanto non è possibile acquisire i dati elementari per ragioni di sicurezza e di riservatezza. Da questo punto di vista, a livello europeo e internazionale si sta investendo molto nello sviluppo di *Privacy enhancing technologies* (PETs)³.

Il *Principio 3* del *Codice delle statistiche europee* (Eurostat e SSE 2017) stabilisce che le autorità statistiche devono disporre di risorse sufficienti, sia in termini quantitativi sia qualitativi, per soddisfare le esigenze statistiche. Tali risorse includono quelle finanziarie, tecnologiche e umane. Tralasciando gli aspetti finanziari, che comunque possono assumere rilievo quando viene richiesto al fornitore dei dati viene richiesto di svolgere elaborazioni onerose in termini di tempo e risorse tecniche, la gestione delle TSS può rendere necessario l'acquisizione di nuove competenze, non ancora presenti in Istituto, e l'adeguamento delle infrastrutture tecnologiche per l'acquisizione e il trattamento di grandi moli di dati generati quasi in tempo reale.

Tali fattori devono essere tenuti in considerazione anche nella stesura degli accordi che si stipulano con i *private data holder*, dove saranno stabiliti non solo i contenuti dei dati forniti, ma anche le caratteristiche della fornitura, la frequenza, la modalità e i metadati di accompagnamento. In un recente documento prodotto da United Nations (2025a) si ipotizza addirittura che il fornitore dei dati possa produrre un *quality report* in collaborazione con l'autorità statistica.

3.3.2 Framework di riferimento per la qualità delle TSS

La qualità è un concetto intrinsecamente multidimensionale e difficilmente misurabile in modo univoco. A livello europeo, il riferimento consolidato è il *Quality Assurance Framework of the European Statistical System* (ESS QAF), fondato sul *Codice delle statistiche europee* (Eurostat e SSE 2017), che identifica le dimensioni fondamentali per valutare la qualità dell'output statistico: pertinenza; accuratezza e affidabilità; tempestività e puntualità; comparabilità e coerenza; accessibilità e chiarezza.

Tali dimensioni, riferendosi appunto all'output, mantengono validità indipendentemente dalla fonte di dati. Tipicamente le nuove fonti portano dei miglioramenti nella rilevanza e nella tempestività delle statistiche prodotte, ma possono comportare dei rischi in termini di affidabilità. Per altre dimensioni tradizionali della qualità, come la comparabilità e la coerenza, diventano rilevanti altri aspetti: ovvero la stabilità nel tempo e la coerenza del risultato

3 Si veda ad esempio il progetto Europeo JOCONDE, attualmente in corso <https://cros.ec.europa.eu/joconde>.

statistico ottenuto da nuove fonti con il risultato statistico prodotto con fonti tradizionali. L'utilizzo di nuove fonti di dati influisce in particolare sull'accuratezza, che dipende dagli errori presenti nei dati di input o che possono essere introdotti nel processo di produzione delle statistiche. La valutazione dell'accuratezza delle statistiche prodotte diventa più complessa rispetto alle fonti tradizionali (ad esempio dalle indagini campionarie) con le nuove fonti di dati come i BD o con le TSS. Ad esempio, gli indicatori di qualità e *performance* richiesti da Eurostat nei *quality report* risultano spesso non applicabili in contesti non convenzionali, in quanto pensati prevalentemente per le statistiche derivate da indagini tradizionali.

Per affrontare queste sfide, il modello del *Total Survey Error* (TSE) (Groves *et al.* 2009), può rappresentare un punto di partenza, in quanto offre una visione integrata delle diverse fonti di errori che possono compromettere l'accuratezza e l'efficienza dei risultati statistici, introducendo distorsione ed elevata variabilità nelle stime prodotte. Il TSE mappa e definisce le fonti di errore lungo tutto il ciclo di vita del dato e favorisce l'individuazione di strategie di prevenzione e di correzione degli errori.

Il modello, originariamente sviluppato per le indagini statistiche, e successivamente esteso alle statistiche di fonte amministrativa (Zhang 2012) si adatta in realtà con difficoltà all'analisi dei BD, a causa delle differenze sostanziali nella struttura, nell'accesso e nei processi di gestione dei dati. Il modello TSE, infatti, presuppone una classica matrice di dati unità-variabili, una condizione non sempre riscontrabile nei BD, soprattutto quando questi presentano formati non strutturati.

Per superare tali limiti, è stata proposta una rivisitazione del *framework* dell'errore totale specificamente pensato per i BD (Amaya *et al.* 2020). Tale approccio mira a valutare la qualità complessiva del prodotto finale piuttosto che del singolo dato e consente di identificare, classificare e analizzare le fonti di errore lungo l'intero ciclo di vita del dato, dalla raccolta all'analisi, tenendo conto sia della distorsione sia della varianza. Nel contesto dei BD, il processo di produzione si articola in due fasi principali: l'identificazione delle fonti, che può essere considerata analoga alla definizione del disegno campionario nelle indagini statistiche; l'estrazione, trasformazione e caricamento dei dati, finalizzata alla loro preparazione per l'analisi. Ciascuna di queste fasi comporta potenziali fonti di errore, che possono emergere durante il reperimento, la trasformazione o l'interpretazione dei dati. Ogni passaggio, dalla generazione iniziale dei dati fino all'analisi finale, può introdurre errori le cui cause variano in funzione del tipo di dato e del contesto applicativo.

Sebbene il modello del TSE, opportunamente riformulato per l'uso di BD, rappresenti un riferimento utile, esso non è sufficiente, da solo, per definire un *framework* di qualità adeguato alle TSS. In particolare, tende a risultare eccessivamente generico e a semplificare in modo significativo la complessità delle fasi di trattamento dei dati.

Per questo motivo, nella definizione del *framework* per le TSS, è fondamentale fare riferimento a esperienze consolidate in ambiti affini, che offrono approcci metodologici e strumenti già validati a livello internazionale. In particolare, si è dato rilievo ai *framework* sviluppati per l'utilizzo dei BD a fini statistici, che rappresentano una base concettuale solida e adattabile al contesto delle TSS. Tra questi, assumono particolare importanza il *Big Data Quality Framework* (BDQF) (UNECE 2014) e le *Quality Guidelines for the Acquisition and Usage of Big Data* (Eurostat 2020).

Il BDQF propone una visione della qualità articolata sulle tre fasi del processo statistico, già utilizzate per strutturare i *framework* di qualità per le statistiche basate su dati amministrativi: *Input* (origine e caratteristiche dei dati); *Throughput* (trasformazione e

trattamento dei dati); *Output* (prodotto statistico finale). Il BDQF, inoltre, si integra perfettamente con il GSBPM (UNECE 2025) e organizza le dimensioni della qualità in tre iperdimensioni: Fonti, Metadati, Dati, anch'esse già presenti nei modelli per dati amministrativi (cfr. paragrafo 2.3).

Le *Quality Guidelines for the Acquisition and Usage of Big Data* rappresentano uno dei contributi più rilevanti a livello internazionale sul tema della qualità nell'uso dei BD per scopi statistici. A differenza di approcci generalisti, queste linee guida non propongono un modello unico valido per tutte le tipologie di BD, ma offrono indicazioni specifiche per ciascuna fonte (ad esempio, dati testuali, da sensori, immagini), strutturate secondo un modello comune basato sulle fasi del processo produttivo. Tra le innovazioni più significative introdotte dalle *Guidelines*, vi è l'evoluzione del modello *Input-Throughput-Output*, in cui la fase di *throughput* viene suddivisa in due sottofasi distinte:

- la fase di *Lower throughput*, che comprende il trattamento iniziale dei dati grezzi e le operazioni preliminari necessarie alla loro successiva elaborazione statistica. Quando i dati sono forniti da terze parti, queste attività sono spesso svolte direttamente dal *data provider*,
- la fase di *Upper throughput*, che include l'applicazione di metodi statistici tradizionali (ad esempio validazioni, calcolo di aggregati) e, ove opportuno, l'integrazione con fonti statistiche ufficiali.

Questa articolazione consente di mappare in modo più dettagliato le attività e le potenziali fonti di errore, permettendo di adattare i criteri di valutazione della qualità alle diverse caratteristiche delle fonti utilizzate. Si tratta, dunque, di un approccio flessibile e strutturato, coerente con la complessità e la varietà dei dati utilizzati nelle TSS.

3.3.3 Verso un framework di qualità delle TSS orientato al processo

La definizione del *framework* di qualità per le TSS deve basarsi su una prospettiva che riprende ed estende il *Total Survey Error* e deve adottare un approccio orientato al processo statistico, secondo l'articolazione *input-lower and upper throughput-output quality* proposto in Eurostat (2020). Tale impostazione rappresenta non solo un fondamento metodologico solido, ma anche un valido punto di partenza per la costruzione di un sistema di monitoraggio strutturato, supportato da indicatori di qualità coerenti con le peculiarità delle fonti e dei processi coinvolti.

In questo contesto, l'approccio risulta particolarmente rilevante poiché le *Quality Guidelines* forniscono indicazioni specifiche per ciascuna fonte (ad esempio, dati testuali, sensori, immagini). Uno degli elementi distintivi delle TSS è, infatti, l'eterogeneità delle fonti di dati, che si differenziano profondamente per struttura, origine, finalità, livello di controllo e qualità intrinseca. Questa eterogeneità comporta una serie di implicazioni dirette sulla progettazione di un *framework* di qualità adeguato, che deve essere flessibile, adattabile e in grado di tenere conto delle caratteristiche peculiari di ciascuna fonte.

Le differenze strutturali tra i dati – che possono essere strutturati, come quelli provenienti da sensori, semistrutturati, come nel caso dei dati generati dai *social media*, o non strutturati, come testi, immagini e video – implicano l'utilizzo di strumenti di trattamento specifici, criteri di qualità distinti, e la necessità di definire indicatori e metadati coerenti con la natura dei dati stessi. Questo si riflette direttamente sulla costruzione del *framework*, che dovrebbe prevedere moduli differenziati per ciascuna tipologia di fonte,

3. Metodi per la valutazione della qualità

evitando approcci generalisti o uniformi, sebbene evidenziando le caratteristiche e le problematiche comuni, laddove presenti. La varietà delle fonti in termini di provenienza – ad esempio dati da sensore, dati generati dagli utenti tramite *social media* e *app*, dati da dispositivi mobili, da piattaforme di commercio elettronico – ha forti implicazioni sulla qualità, in particolare perché implica la presenza di attori molto diversi tra loro, quali istituzioni, aziende private e cittadini. Tale pluralità determina livelli eterogenei di accessibilità, trasparenza, governance e controllo sul dato, elementi che devono necessariamente essere riflessi nel *framework* di qualità.

Anche i processi di acquisizione e trasformazione dei dati variano sensibilmente in funzione della tipologia di fonte. I dati possono infatti essere ricevuti in forma grezza, pretrattati dal fornitore, o essere accessibili in tempo reale o su base periodica. Di conseguenza, è fondamentale prevedere un monitoraggio continuo dei processi di ingestione e trasformazione, supportato da un uso esteso dei metadati per documentare ogni passaggio e garantirne la tracciabilità.

Alcune fonti di BD, come quelle provenienti da sensori o *social media*, sono caratterizzate da una forte dinamicità, risultando poco stabili nel tempo. È dunque necessario che il *framework* contempli una dimensione temporale della qualità, includendo elementi quali la frequenza di aggiornamento, la tempestività e la coerenza longitudinale, adattandoli alle specifiche caratteristiche delle fonti utilizzate.

L'uso combinato di fonti eterogenee comporta, inoltre, rischi significativi legati a incongruenze tra i *dataset*, duplicazioni, o perdite di coerenza semantica. Questi rischi impongono la necessità di includere nel *framework* criteri specifici per la valutazione della qualità nei processi di integrazione, come quelli legati al *linkage* e all'armonizzazione dei dati, oltre a criteri relativi alla disponibilità e all'aggiornamento di opportuni metadati di accompagnamento, che rendano possibile valutare la coerenza dei dati ricevuti con i concetti statistici.

In sintesi, se da un lato l'eterogeneità delle fonti offre un potenziale informativo senza precedenti, dall'altro impone la progettazione di un *framework* di qualità che sia modulare e flessibile.

3.3.4 Identificazione dei primi elementi della qualità dell'input: la classificazione delle fonti

In un contesto dove i processi possono divergere enormemente tra le varie fonti di dati, così come i potenziali errori, è importante fare riferimento prima a una classificazione delle fonti e dei dati, e quindi di processi simili, e poi collegare le tipologie di errore a ciascuna classe omogenea così individuata.

Una classificazione di tipo gerarchico in cui i dati sono raggruppati sia per tipo di fonte sia per struttura dei dati, può rappresentare un punto di partenza per affrontare le dimensioni della qualità.

Una prima e utile classificazione delle fonti può basarsi sulla distinzione tra: sorgenti web, sensori e dati provenienti da transazioni elettroniche (ad esempio, bonifici bancari, addebiti diretti, carte di credito e di debito, servizi di pagamento online). Di seguito si riportano dei prospetti relativi alle prime due categorie di fonti con una descrizione sintetica delle informazioni di interesse che è possibile derivarne (Prospetti 3.10 e 3.11).

Prospetto 3.10 - Sorgenti web per tipologia di fonte e informazioni di interesse (esempi)

Fonte	Informazioni di interesse
Portali web (esempio: siti per attività di <i>e-commerce</i> ; siti per la pubblicazione di annunci di lavoro; siti di vendita online, cataloghi di prodotti online; portali sul turismo e strutture ricettive; portali degli aeroporti)	Caratteristiche imprese, altre informazioni
	Mercato del lavoro lato domanda
	Acquisizione di informazioni sui prezzi di beni e servizi
	Statistiche sul turismo (posti letto, ristorazione, servizi di trasporto, prezzi)
Siti aziendali	Acquisizione di informazioni sulle tratte aeroportuali
	Caratteristiche imprese, attività di impresa alternative alla classificazione delle attività economiche Ateco, presenza su <i>social media</i>
Servizi web cartografici (<i>Web Map Services - WMS</i>)	Acquisizione di informazioni sulla rete stradale per la costruzione di indicatori di incidentalità con nuovi denominatori - per arco ed estesa stradale e punti di traffico
<i>Social media/network</i> come <i>Facebook, Instagram, LinkedIn, YouTube, X, Flickr, Xing</i> , altri	Acquisizione di informazioni sui <i>point of interest</i>
	Indici di <i>sentiment</i> su varie tematiche

Fonte: Istat

Prospetto 3.11 - Dati da sensore per tipologia di fonte e informazioni di interesse (alcuni esempi)

Tipologie di sensori	Informazioni di interesse
Satelliti per il telerilevamento (<i>satellite imagery</i>)	Copertura dei suoli e riforestazione Aree urbane verdi
Reti di telecomunicazione degli operatori di telefonia mobile (<i>Mobile Network Operator - MNO signaling data</i>)	Informazioni su popolazione, turismo, trasporti
<i>Sistema globale</i> per il tracciamento delle navi (<i>Automatic Identification System - AIS data</i>) <i>Global positioning system (GPS) technology</i>	Trasporto marittimo
Sensori meteorologici e per il monitoraggio dell'inquinamento	Dati metereologici
<i>Smart meter</i> per il monitoraggio dei consumi energetici	Monitoraggio/rilevazione sui consumi energetici Individuazione delle abitazioni a seconda che siano occupate o non utilizzate sempre o solo in alcuni periodi dell'anno

Fonte: Istat

3.3.5 Approccio empirico per la validazione del modello concettuale del framework

Il modello generale ipotizzato per strutturare le fasi del processo e le dimensioni della qualità è stato testato empiricamente attraverso esperienze sviluppate all'interno dell'Istat (cfr. paragrafo 3.3). L'obiettivo era verificare in che misura le specificità dei singoli processi e prodotti risultassero coerenti e integrabili in una configurazione concettuale preliminare del *framework*. A tale fine è stata condotta una ricognizione basata sulla compilazione di schede informative standardizzate, curate dai referenti dei diversi processi sperimentali. Tali schede erano strutturate secondo il modello proposto in Eurostat (2020), che distingue le fasi di *input*, *lower throughput*, *upper throughput* e *output*.

Ogni scheda era articolata in tre sezioni principali:

- la sezione di *input*, dedicata alla descrizione della fonte e dei dati grezzi;
- la sezione di *lower throughput*, finalizzata a documentare i processi utilizzati per derivare gli output intermedi e i relativi aspetti di qualità, compresi metodi e algoritmi applicati;
- la sezione di *upper throughput*, relativa alla produzione degli output statistici finali e ai criteri di qualità adottati (ad esempio validazione, accuratezza, coerenza).

Attraverso queste schede sono state analizzate numerose tipologie di TSS, tra cui: dati provenienti da *social network* (ad esempio *X*, precedentemente denominato *Twitter*), dati web (come gli annunci di lavoro online), telerilevamento, *shape* vettoriali in formato *Geographic Information System (GIS)*, dati *Automatic Identification System (AIS)*, dati meteo-climatici, dati da operatori di rete mobile e transazioni elettroniche.

3. Metodi per la valutazione della qualità

L'analisi delle schede ha prodotto risultati rilevanti, confermando diverse ipotesi iniziali. La classificazione delle fonti si è rivelata fondamentale per una più efficace comprensione delle loro caratteristiche distintive; è emersa l'utilità di considerare nella classificazione, non solo la natura della fonte, ma anche due aspetti fondamentali: la tipologia del dato (strutturato, semistrutturato, non strutturato) e la modalità di acquisizione (interna, tramite terze parti, o attraverso portali/API). Nel complesso, l'analisi ha confermato la solidità dell'impostazione metodologica adottata, evidenziando al contempo la necessità di un *framework* di qualità strutturato e modulare, capace di rappresentare in modo sistematico l'eterogeneità delle fonti e i processi di trasformazione necessari per rendere i dati utilizzabili nell'ambito della produzione statistica ufficiale.

Il Prospetto 3.12 riporta le principali tipologie di errore emerse dall'analisi delle schede per tipo di fonte.

Prospetto 3.12 - Fonti analizzate e principali tipologie di errore

Fonti analizzate	Tipologie di errori nei processi
1. Sorgenti web	
- X (già Twitter)	Sottocopertura Campione non rappresentativo della popolazione Mancanza di rilevanza statistica, errori di <i>text cleaning</i>
- Annunci di lavoro online	Sottocopertura e sovracopertura Campione non rappresentativo Duplicati, errori di classificazione, errori di <i>text cleaning</i>
2. Sensori	
- Telerilevamento	Valori mancanti, errori di misurazione
- Dati meteorologici e idrologici	Errori nei formati dei dati, errori di misurazione, valori mancanti
- AIS	Valori mancanti e <i>outlier</i>
- <i>shape vettoriali in formato GIS</i> (<i>Geographic Information System</i>)	Errori di misurazione
- MNO	Sottocopertura e sovracopertura Valori mancanti, errori di specificazione (mancanza di corrispondenza precisa con le definizioni statistiche)
3. Transazioni elettroniche	Sottocopertura Campione non rappresentativo della popolazione

Fonte: Istat

Per le fonti riportate nel Prospetto emergono diverse tipologie di errore che incidono significativamente sulla qualità del dato statistico prodotto. Si rilevano frequentemente problemi di rappresentatività della fonte rispetto alla popolazione obiettivo, dovuti a problemi di copertura, selezione o autoselezione tipici di molte fonti di BD. In effetti, una peculiarità riscontrata praticamente in tutte le casistiche analizzate è che l'oggetto a cui si riferisce il dato grezzo non corrisponde all'unità statistica di interesse. Ciò comporta che una parte del processo di trattamento del dato è dedicato a trasformare questi oggetti per derivare le unità di interesse, spesso con approssimazioni che comportano, appunto, problemi di rappresentatività. Si evidenziano inoltre criticità metodologiche legate specificatamente ai processi di trattamento e trasformazione del dato, in particolare all'uso di modelli di *machine learning* o di altri algoritmi complessi. Tali metodi, se non adeguatamente validati, possono introdurre errori sistematici o amplificare distorsioni preesistenti, compromettendo l'accuratezza e l'affidabilità dei risultati statistici finali (cfr. paragrafo 3.3.7).

Pertanto, è fondamentale che il *framework* di qualità consideri non solo le caratteristiche intrinseche della fonte, ma anche le potenziali implicazioni metodologiche derivanti dagli strumenti/metodi analitici utilizzati nel processo produttivo. Questo conduce alla riflessione sulla qualità del *throughput*. Il *lower throughput* risulta spesso la parte più delicata, sia perché è in genere la più innovativa, per cui non esistono ancora metodologie consolidate e

misure standard per la valutazione della qualità, sia perché viene comunemente implementata dai *data provider*, senza la possibilità di essere monitorata direttamente dall'autorità statistica. Pertanto, è necessario nel *framework* di qualità non solo definire nuovi metodi e misure di controllo della qualità, ma tenere in considerazione anche gli impianti organizzativi e le possibilità di dialogo e di interazione con i fornitori dei dati.

D'altro canto, l'*upper throughput* rappresenta l'opportunità per l'autorità statistica di migliorare la qualità del dato pre-elaborato ricevuto, spesso combinandolo con altre fonti tradizionali. Il *framework* di qualità dovrà essere dunque orientato a fornire raccomandazioni su come procedere, oltre a proporre misure per verificare di non introdurre ulteriori errori. La grande sfida sarà, infine, di proporre misure per la valutazione della qualità di un output derivante da processi complessi come quelli delle TSS. L'orientamento sarà quello di sviluppare misure di validazione rispetto ad altre fonti disponibili. Si valuterà anche un possibile cambio di prospettiva, in cui la tradizionale indagine statistica viene utilizzata per valutare la qualità dell'output della TSS, invece di essere la principale fonte della statistica ufficiale.

3.3.6 TSS e intelligenza artificiale

L'elaborazione di nuove tipologie di dati spesso rende necessario l'impiego dell'intelligenza artificiale (IA), aspetto di cui si deve tenere in conto in un *framework* di qualità per le TSS.

L'IA di ultima generazione si basa su modelli di *machine learning* (ML) che permettono di analizzare grandi moli di dati per individuare regolarità e strutture non immediatamente osservabili. In questo contesto, le reti neurali rappresentano una delle principali tecniche utilizzate: si tratta di modelli computazionali ispirati ad architetture biologiche semplificate, capaci di apprendere dai dati attraverso un processo iterativo di ottimizzazione dei parametri e di adattamento delle connessioni interne del modello. Rispetto ai metodi statistici tradizionali, la capacità predittiva che contraddistingue gli algoritmi di ML è dunque ottenuta attraverso architetture di calcolo complesse che rendono complicato se non impossibile ricostruire il processo che ha generato l'output. I risultati così ottenuti, a differenza di quelli prodotti dai modelli statistici tradizionali, sono generalmente sprovvisti di una misura di incertezza che consenta di valutarne l'errore di stima o di previsione.

Tutto ciò costituisce un ostacolo alla loro adozione nella statistica ufficiale. Per questo motivo, il *Quality Framework for Statistical Algorithms* (QF4SA) (Yung *et al.* 2022) raccomanda di puntare su un'IA in grado di rendere espliciti non solo i risultati delle stime o delle previsioni, ma anche le modalità e, ove possibile, le ragioni che le hanno generate (UNECE 2021).

La comprensione da parte di utente umano di un meccanismo di previsione, sia esso un modello parametrico o un algoritmo di autoapprendimento, è cruciale quando i risultati prodotti devono supportare decisioni delicate o scelte strategiche. Una previsione molto accurata senza margini di incertezza potrebbe risultare tale per le ragioni sbagliate con conseguenze potenzialmente nefaste se estesa a soggetti o a contesti diversi da quelli originari su cui l'algoritmo è stato addestrato. I rischi associati ad algoritmi poco espressivi o troppo ermetici hanno recentemente orientato parte della ricerca verso la costruzione di un'intelligenza artificiale più comprensibile (*explainable AI* o XAI) (Vilone e Longo 2020, Das e Rad 2020). Metodi di ML per l'imputazione di esiti potenziali che riconciliano predizione e inferenza causale (Daud e Dabhashy 2023), algoritmi che addestrano modelli lineari ad approssimare localmente stime ottenute con metodi non lineari più complessi (Ribeiro *et al.* 2016), tecniche di randomizzazione che perturbano individualmente le *feature* utilizzate in input dall'algoritmo per verificarne la rilevanza

sulla predizione finale attraverso statistiche test (Burns *et al.* 2020) sono solo alcuni degli approcci proposti dalla letteratura XAI.

La trasparenza è dunque un requisito basilare in quanto funzionale alle altre dimensioni della qualità di processo. Oltre alla robustezza e all'affidabilità delle stime prodotte essa è strumentale per la riproducibilità dei risultati che contraddistingue i modelli usati tradizionalmente nella statistica ufficiale. L'opacità degli algoritmi di ML, unitamente alle limitazioni di accesso al codice e ai dati di *training*, appaiono al momento incompatibili con la replicabilità degli output intermedi e/o finali generati dai nuovi modelli di AI. Allo stesso modo, l'interpretabilità dell'output risente negativamente del fatto che le nuove architetture basate sul ML privilegiano l'accuratezza predittiva alla spiegazione del meccanismo inferenziale e causale, legittimando il ricorso a interpretazioni ad hoc o post-hoc non sorrette da basi teoriche e scientifiche solide e condivise.

Nonostante le criticità evidenziate, lo sviluppo di sistemi di IA efficaci e interpretabili è possibile e può contribuire all'avanzamento della conoscenza scientifica. Esistono in letteratura casi studio di successo che incoraggiano a ricercare ambiti specifici di applicazione delle nuove tecnologie (Keane *et al.* 2021). Previsioni favorevoli (o sfavorevoli) giudicate ambigue o *borderline* dall'analista umano potrebbero rivelarsi di fatto corrette a seguito di un esame più approfondito e, di conseguenza, contribuire ad aggiornare il *gold standard* codificato in certo processo produttivo o protocollo terapeutico.

Nel contesto dell'applicazione di algoritmi di ML, l'approccio TSE può assumere un ruolo di rilevanza strategica; la sua integrazione nel processo di sviluppo dei modelli non è da considerarsi accessoria, ma costituisce un passaggio fondamentale per garantire la validità dei modelli predittivi, l'integrità metodologica, la robustezza dell'intero sistema analitico. I modelli di ML sono particolarmente sensibili alla validità delle assunzioni e alla qualità dei dati utilizzati nelle fasi di *training*. La presenza di errori di rappresentatività dovuti alla scelta inappropriata del modello e a problemi nei dati (copertura, dati mancanti, selezione non casuale o campioni sbilanciati) e di misurazione dovuti a classificazioni errate, variabili mal definite, a errori nelle *feature*, può compromettere significativamente le prestazioni predittive, introducendo distorsione e riducendo la capacità di generalizzazione del modello. In questo scenario, l'approccio TSE si configura come una struttura concettuale chiara e sistematica per identificare, quantificare e gestire le principali fonti di errore, già nelle fasi preliminari alla modellizzazione e può contribuire a rendere il processo di modellizzazione più trasparente, tracciabile e replicabile, facilitando la documentazione delle scelte metodologiche e aumentando la fiducia nei risultati prodotti dai modelli. In definitiva, può rappresentare una condizione abilitante per un impiego consapevole, rigoroso e metodologicamente solido del ML, non solo in alcuni processi delle indagini tradizionali, ma anche in un contesto informativo sempre più eterogeneo e complesso come quello dei BD.

Oltre alla pubblicazione periodica di statistiche sperimentali come gli indicatori del *social mood* sull'andamento dell'economia⁴ (Istat 2024) e quelli sull'incidentalità (Broccoli e Bruzzone 2021, Istat 2025b)⁵, l'Istat, sotto l'impulso delle collaborazioni internazionali per la modernizzazione della statistica ufficiale, ha recentemente sperimentato tecniche di ML con reti neurali per l'imputazione di dati mancanti sui propri registri statistici (De Fausti *et al.* 2022).

4 Si sottolinea la natura sperimentale delle statistiche citate, disponibili sulla pagina web dell'Istat dedicata a queste tipologie di dati: <https://www.istat.it/statistica-sperimentale/social-mood-on-economy-index/>.

5 Si sottolinea la natura sperimentale delle statistiche citate, disponibili sulla pagina web dell'Istat dedicata a queste tipologie di dati: <https://www.istat.it/statistica-sperimentale/utilizzo-di-open-street-map-per-il-calcolo-di-indicatori-per-lincidentalita-stradale-sulla-rete-viaria-italiana-anno-2023/>.

Stabilire un collegamento tra la cultura statistica ipotetico-deduttiva tradizionale, basata su modelli semplificati dei fenomeni oggetto di indagine, e quella algoritmico-induttiva che analizza i dati disponibili alla ricerca di associazioni e relazioni latenti, appare una condizione necessaria per delimitare il perimetro di utilizzo del *machine learning* nella statistica ufficiale e per costruire un sistema di monitoraggio della qualità. Tale sistema consentirebbe di valutare la possibilità di promuovere la statistica sperimentale e l'architettura sottostante al rango di statistica ufficiale, sulla base delle esperienze maturate in seno all'Istat.

3.3.7 Sviluppi futuri

Il lavoro fin qui svolto ha permesso di approfondire la conoscenza delle TSS, di evidenziare l'eterogeneità che le contraddistingue ma anche di identificare i tratti comuni, e soprattutto i maggiori rischi per la qualità degli output prodotti.

È ormai chiara anche la struttura che dovrà assumere il *framework* per la valutazione della qualità di questa tipologia di processi e i principali aspetti, dimensioni ed errori che devono essere tenuti in considerazione. Il prossimo passo sarà quello di definire metodi e misure che consentano il monitoraggio e la valutazione della qualità delle TSS.

Visto che, come noto, la documentazione del processo è un prerequisito per la valutazione, e che, anche in questo caso, si intende seguire un approccio empirico, si intende procedere documentando in maniera dettagliata delle TSS "pilota". Si cercherà di utilizzare i modelli standard internazionali per la descrizione dei processi statistici, come GSBPM, verificandone anche l'applicabilità o la necessità di estensione. Insieme con la metadattazione del processo si proporranno metriche e indicatori di qualità per il monitoraggio e la valutazione della qualità nelle diverse fasi del processo. L'obiettivo sarà poi generalizzare quanto sviluppato per produrre delle linee guida metodologiche che forniranno raccomandazioni su come valutare la qualità di una TSS e così decidere se rispetta i requisiti di qualità necessari per essere considerata statistica ufficiale. Tale *framework* risulterà quindi particolarmente utile per le attuali statistiche sperimentali⁶ al fine di fare diventare statistiche ufficiali quelle basate su TSS.

⁶ Si sottolinea la natura sperimentale delle statistiche citate, disponibili sulla pagina web dell'Istat dedicata a queste tipologie di dati: <https://www.istat.it/comunicati-e-analisi/statistiche-sperimentali/>.

4. AZIONI VOLTE AD ASSICURARE LA QUALITÀ NEL SISTEMA STATISTICO ITALIANO¹

Assicurare l'elevata qualità delle statistiche ufficiali rappresenta una condizione imprescindibile per consolidare la fiducia dei cittadini e delle istituzioni nei confronti della statistica pubblica, per salvaguardarne la credibilità e per accrescerne la reputazione.

Accanto al necessario inquadramento normativo e regolatorio, la promozione della qualità nella statistica ufficiale richiede un impegno operativo concreto, che si realizza attraverso l'adozione di strumenti tecnici, standard metodologici e linee guida che consentano agli enti del Sistan di rendere operativi i principi della qualità, adattandoli al proprio contesto organizzativo, alle risorse disponibili e alla natura delle attività svolte. La qualità deve infatti essere gestita in modo sistematico e integrato, non come adempimento formale o risposta a obblighi esterni, ma come componente strutturale di ogni fase del processo statistico, dalla progettazione alla raccolta, dall'elaborazione alla diffusione.

Affinché ciò sia possibile, i sistemi di statistica ufficiale devono fondarsi su modelli di governance chiari, condivisi e orientati al principio di responsabilità, capaci di assicurare la coerenza operativa tra i diversi attori istituzionali, la trasparenza delle scelte metodologiche e la piena tracciabilità dei processi produttivi e dei relativi output. In questa direzione si orienta il Sistan, che adotta in modo sistematico il quadro concettuale elaborato a livello europeo: un riferimento strutturato che fornisce principi guida, definizioni operative, raccomandazioni metodologiche e strumenti di monitoraggio, promuovendo al contempo un linguaggio comune e criteri condivisi per la misurazione e la comunicazione della qualità statistica. In ambito italiano vi è una diffusa consapevolezza, condivisa dagli organi di indirizzo e di controllo del Sistan, che l'adesione costante ai principi contenuti nel *Codice delle statistiche europee* (di seguito, *Codice*) (Eurostat e SSE 2017) rappresenti una leva fondamentale per il miglioramento continuo della qualità statistica. In tale senso, l'allineamento degli standard e delle pratiche degli enti del Sistan a quelli definiti a livello europeo rappresenta un fattore cruciale per rafforzare il funzionamento complessivo del Sistema. Questo parallelismo consente, da un lato, di promuovere l'indipendenza istituzionale e funzionale degli enti produttori di statistica, tutelandoli da interferenze esterne e assicurando l'imparzialità dell'informazione statistica; dall'altro, permette di affermare l'integrità, la trasparenza e la responsabilità come tratti distintivi dell'azione statistica pubblica, valorizzando il ruolo della funzione statistica anche all'interno delle amministrazioni.

È tuttavia necessario considerare che il *Codice*, per sua natura e per le finalità per cui è stato concepito, si applica formalmente ai soli organismi che concorrono alla produzione di statistiche europee. Esso costituisce il riferimento deontologico e metodologico per Eurostat, per gli Istituti nazionali di statistica degli Stati membri e per le ONA designate alla produzione di statistiche europee. Ne consegue che la sfera di applicazione effettiva del *Codice* è circoscritta a un sottoinsieme degli attori statistici italiani, lasciando al di fuori gran parte degli enti che compongono il Sistan, i quali operano prevalentemente su scala nazionale, con funzioni che, pur essenziali, non rientrano nel perimetro delle statistiche europee. In altri termini, se da un lato il *Codice* costituisce un riferimento imprescindibile per la qualità delle statistiche ufficiali, dall'altro lato la sua efficacia formale non si estende automaticamente all'intera rete degli enti Sistan, che pure svolgono un ruolo strategico nella produzione e diffusione dell'informazione statistica nel Paese.

¹ Il Capitolo 4 è stato redatto da: Andrea Bruni (paragrafi 4, 4.1, 4.2 e 4.4) e Paola Giordano (paragrafo 4.3).

A titolo esemplificativo, si rinvia alla Tavola 4.1, che offre una rappresentazione aggiornata della composizione del Sistan, evidenziando come soltanto una frazione assai ridotta degli enti del sistema – corrispondente alle tredici ONA italiane riconosciute a livello europeo – ricada nell’ambito di applicazione diretta del *Codice*.

Tavola 4.1 - Uffici di statistica del Sistan al 31 dicembre 2024 (valori assoluti e percentuali)

TIPOLOGIA DI ENTE	Numero	%
Ministeri e Presidenza del Consiglio dei ministri	17	0,5
Prefetture - Uffici territoriali del governo (UTG)	99	3,0
Enti e amministrazioni pubbliche centrali	23	0,7
Regioni e province autonome	21	0,6
Province	64	1,9
Città metropolitane	12	0,4
Comuni capoluogo o con almeno 30 mila abitanti	249	7,5
Comuni non capoluogo con meno di 30 mila abitanti	2.691	81,4
Camere di commercio	64	1,9
Altre amministrazioni locali	54	1,6
Altri soggetti (soggetti privati)	11	0,3
Totale	3.305	100,0

Fonte: Rilevazione sugli elementi identificativi, risorse e attività degli uffici di statistica del Sistan (Eup)

Tale configurazione istituzionale ha reso evidente, già da tempo, l’esigenza di sviluppare strumenti e modelli nazionali specifici per la promozione e la gestione della qualità, capaci di garantire l’adozione di principi e pratiche comuni a tutto il Sistan, indipendentemente dalla partecipazione alla produzione di statistiche europee. È dunque emersa con forza la necessità di dotarsi, in ambito nazionale, di un quadro di riferimento coerente con il *Codice*, ma adattato alle specificità organizzative, giuridiche e operative del contesto italiano, così da favorire il rafforzamento della qualità in modo omogeneo e inclusivo, coinvolgendo l’intera rete del Sistan.

Nonostante la distinzione formale tra gli ambiti europeo e nazionale, sin dalle prime fasi del dibattito sulla qualità, è stata condivisa l’opportunità – e anzi la necessità – di estendere i principi e gli strumenti europei all’intero universo degli enti Sistan, superando ogni distinzione fondata sulla destinazione finale dei dati prodotti. Questa scelta strategica è basata sull’assunto che la qualità non sia un requisito da applicare selettivamente, ma rappresenti un valore universale che deve caratterizzare ogni statistica ufficiale, a prescindere dal fatto che sia utilizzata a livello sovranazionale o meno.

A fondamento di tale approccio vi è il riconoscimento del *Codice* come strumento fondamentale di governo della funzione statistica pubblica, capace di orientare le politiche di produzione verso standard elevati di affidabilità, indipendenza e trasparenza. Estendere l’ispirazione e l’applicazione del *Codice* a tutto il Sistan ha significato, pertanto, fare della qualità un principio ordinatore dell’intero sistema, rafforzando l’integrazione tra le diverse componenti istituzionali, e favorendo una visione condivisa e responsabile del ruolo della statistica nella società contemporanea.

Già nel novembre 2007, il Comstat riconosceva la rilevanza strategica di un allineamento strutturale tra il Sistan e il quadro europeo di riferimento in materia di qualità. Con lungimiranza, veniva dunque avviato un processo di recepimento e adattamento dei principi già consolidati a livello europeo, al fine di rafforzare la coerenza interna del sistema italiano e di promuovere una visione condivisa della qualità statistica anche al di fuori del perimetro delle statistiche europee. In quest’ottica, si è progressivamente delineato un impianto

metodologico nazionale, parallelo e coerente con l'architettura europea, volto a rendere cogenti e applicabili su scala nazionale i criteri e le raccomandazioni del *Codice*.

Tale orientamento ha risposto all'esigenza di estendere l'ambito di influenza del modello europeo a tutta la produzione statistica nazionale, non solo a quella realizzata dalle ONA. L'obiettivo era costruire un riferimento formalizzato, oggettivo e condiviso, capace di orientare l'azione dei numerosi enti del Sistan, fornendo una base comune su cui fondare le politiche e le pratiche per la qualità. Il processo di recepimento e adattamento ha richiesto un importante sforzo di armonizzazione, finalizzato a garantire che gli standard qualitativi internazionali potessero essere declinati in modo efficace all'interno dell'eterogeneo e articolato sistema statistico nazionale.

L'adozione di una versione italiana del *Codice* per l'intero Sistan ha comportato fin da subito implicazioni rilevanti sul piano culturale e organizzativo: richiede, infatti, non soltanto un impegno formale, ma una riflessione consapevole e critica sui punti di forza e di debolezza dei processi statistici, stimolando gli operatori della statistica ufficiale a riconsiderare modalità operative consolidate, introducendo logiche valutative e approcci basati sulle evidenze quantitative. Attraverso attività strutturate di valutazione e monitoraggio della qualità, è possibile restituire una rappresentazione dinamica e documentata delle capacità organizzative, delle soluzioni metodologiche adottate e dei risultati conseguiti. Ciò consente non solo di individuare tempestivamente le criticità e di intervenire in modo mirato, ma anche di valorizzare le buone pratiche e favorire l'apprendimento diffuso all'interno del sistema. In questo modo si innesca un circolo virtuoso di crescita, basato su un principio di miglioramento continuo, che rafforza nel tempo la qualità complessiva della produzione statistica e la sua coerenza con gli standard europei e internazionali.

L'attuazione dei principi di qualità, dunque, non deve essere interpretata come un mero esercizio di conformità o adempimento formale, ma come un'opportunità concreta per promuovere la cultura della qualità nella statistica ufficiale, facendo leva su meccanismi già sperimentati a livello internazionale. Tale trasparenza concorre a consolidare la visibilità del sistema e la credibilità della statistica ufficiale. A complemento di questo impianto concettuale e metodologico, l'Istat ha messo a disposizione degli enti del Sistan una serie di strumenti operativi mirati, pensati per facilitare l'attuazione concreta dei principi e per rendere effettiva la gestione della qualità nei contesti operativi più diversi. Tra questi, un ruolo di primo piano è svolto dalle *Linee guida sulla qualità delle statistiche del Sistema statistico nazionale* (Istat 2018). Si tratta di un documento di carattere applicativo, indirizzato ai responsabili dei processi statistici degli enti, che fornisce indicazioni pratiche per la gestione della qualità lungo tutte le fasi del ciclo di vita della produzione statistica, indipendentemente dalla natura delle fonti utilizzate, siano esse indagini dirette o archivi amministrativi.

Le *Linee guida* analizzano in maniera sistematica ogni fase del processo produttivo, dalla progettazione alla diffusione, fornendo un quadro ragionato delle principali fonti di errore non campionario e delle strategie per prevenirle, controllarle o correggerle. Oltre alla componente descrittiva e orientativa, il documento include anche un corredo di indicatori e strumenti per il monitoraggio in tempo reale delle attività, nonché per la valutazione ex post dei risultati ottenuti. Il contributo di queste *Linee guida* risiede nella capacità di tradurre i principi teorici (le raccomandazioni nazionali e internazionali) in pratiche operative accessibili e adattabili anche agli enti con minore dotazione tecnico-organizzativa. Il documento favorisce in tale modo un'applicazione flessibile ma rigorosa dei criteri di qualità, sostenendo la diffusione di competenze e consapevolezza all'interno del Sistan.

Le sezioni del manuale sono sviluppate in conformità con il modello GSBPM (UNECE 2025, cfr. paragrafo 2.1.3), per descrivere le fasi del processo produttivo statistico. Tale mappatura viene corredata da suggerimenti operativi e riferimenti bibliografici utili per l'attuazione pratica.

Per garantire che tali indicazioni restino aggiornate e aderenti all'evoluzione normativa, metodologica e tecnologica, il Comitato Qualità dell'Istat ha incluso tra le priorità del prossimo triennio una revisione sostanziale delle *Linee guida*. L'aggiornamento sarà orientato non solo al recepimento dei più recenti sviluppi internazionali in materia di qualità statistica, ma anche al miglioramento della fruibilità del documento, attraverso l'inserimento di esempi pratici, casi di uso, esperienze maturate sul campo, e indicazioni applicative per l'impiego delle nuove tecnologie nella gestione della qualità. Questa revisione consentirà di rafforzare ulteriormente il ruolo delle *Linee guida* come strumento centrale del sistema di qualità nazionale, mantenendone la coerenza con il contesto europeo e accrescendone l'utilità per gli enti del Sistan.

4.1 Il *Codice italiano* per la qualità delle statistiche ufficiali

Il *Codice italiano* per la qualità delle statistiche ufficiali (di seguito: *Codice italiano*) (Istat 2021) rappresenta il quadro di riferimento essenziale per l'attuazione dei principi e degli standard relativi alla progettazione, alla produzione e alla diffusione delle statistiche ufficiali in Italia. Si configura come uno strumento strategico volto al consolidamento della governance del Sistan, alla valorizzazione e alla standardizzazione dell'attività statistica, nonché alla promozione di una cultura condivisa del dato e del suo riutilizzo informato.

Nel dettaglio, il *Codice italiano* persegue una pluralità di obiettivi interconnessi:

- accrescere la fiducia nell'indipendenza, nell'integrità e nella responsabilità dei produttori di statistica ufficiale, contribuendo a rafforzarne la credibilità e a migliorare la qualità dei prodotti diffusi;
- promuovere l'adozione dei principi, delle pratiche e dei metodi più avanzati da parte di tutti i produttori di statistica ufficiale, al fine di accrescere la qualità dei processi e dei dati prodotti;
- fornire un contributo significativo per migliorare il funzionamento complessivo del Sistan, con particolare riguardo al consolidamento dell'indipendenza scientifica e operativa degli enti e degli uffici che lo compongono.

Il *Codice italiano* svolge, inoltre, una funzione di supporto alla visibilità e al riconoscimento istituzionale degli uffici di statistica, facilitando il loro ruolo strategico all'interno degli enti di appartenenza. In quanto documento ufficiale e condiviso, il *Codice italiano* può essere presentato ai vertici amministrativi e agli altri uffici coinvolti nella produzione di dati come strumento autorevole per promuovere pratiche di miglioramento continuo della qualità e rafforzare la funzione statistica nel contesto organizzativo interno.

L'adozione del *Codice italiano* implica un percorso fondato sull'autovalutazione degli assetti istituzionali, dei processi e dei prodotti statistici, finalizzato alla individuazione di azioni concrete per il miglioramento continuo della qualità. Questo approccio deve essere concepito non come una fotografia statica, bensì come una rappresentazione dinamica e progressiva della maturità organizzativa e della capacità del Sistema di adattarsi e di evolvere per affrontare nuove sfide informative.

La struttura concettuale del *Codice*, di tipo gerarchico, si articola in tre livelli principali: macroaree (ambito istituzionale, processi produttivi, risultati statistici), principi e indicatori. Il *Codice italiano* ha progressivamente adottato lo stesso impianto teorico.

La prima versione del *Codice italiano*, risalente al 2010, insieme alle macroaree recepi-va inizialmente solo il secondo livello, quello dei principi, ritenuto sufficientemente generale da poter essere applicato trasversalmente, anche in presenza di una marcata eterogeneità degli attori del Sistan, le cui strutture e funzioni sono spesso molto diverse da quelle degli Ins. Per ogni principio, il *Codice italiano* fornisce una definizione chiara e un campo di applicazione ben delineato, facendo riferimento, laddove possibile, alle disposizioni della normativa statistica ufficiale oppure, in assenza di queste, a criteri consolidati della buona pratica. Alcuni contenuti sono stati, fin dalla prima edizione, estesi o adattati per rispondere a esigenze emergenti, sia a livello nazionale sia europeo.

Il terzo livello concettuale, quello degli indicatori, non fu inizialmente formalizzato nel *Codice italiano*. Si preferì, invece, attribuire ai documenti attuativi, come i questionari di autovalutazione o le guide operative, il compito di tradurre in operatività i principi, per loro natura generici e di ampio respiro. Tale scelta rifletteva l'intenzione di garantire una maggiore flessibilità e aderenza al contesto nazionale, attraverso la calibrazione diretta sul campo.

La seconda edizione del *Codice italiano* ha segnato un punto di svolta significativo, con la ricezione delle più recenti evoluzioni sia sul piano europeo sia a livello nazionale. Tra le principali innovazioni vi è l'inserimento del principio 1.bis "Coordinamento e cooperazione", che promuove la collaborazione a fini statistici sia all'interno degli enti sia tra enti diversi e con l'Istat. Questo principio promuove una duplice funzione: da un lato, il supporto interno all'amministrazione di appartenenza attraverso l'offerta di informazioni statistiche utili, ad esempio, per la programmazione o la valutazione della *performance*; dall'altro, la partecipazione attiva a reti di collaborazione e coordinamento con gli altri attori del Sistan, incluso l'Istat.

Un'ulteriore novità riguarda l'introduzione, nel testo del *Codice italiano*, del riferimento all'utilizzo statistico di fonti innovative, come i dati detenuti da soggetti privati, riconoscendo il crescente ruolo che queste fonti possono assumere nella produzione statistica ufficiale. In seguito a un confronto con il Garante per la protezione dei dati personali, il testo è stato, inoltre, aggiornato per includere disposizioni coerenti con il quadro normativo nazionale ed europeo in materia di riservatezza statistica e tutela dei dati personali, rafforzando le tutele a garanzia del corretto utilizzo delle informazioni raccolte a fini statistici.

La principale innovazione della seconda edizione riguarda la definizione del secondo livello concettuale del modello, attraverso l'introduzione dei criteri guida, che ha permesso di completare l'allineamento strutturale con il *Codice*. Questa evoluzione ha consentito di rafforzare la misurazione oggettiva del livello di aderenza e conformità ai principi: i criteri rappresentano una riformulazione degli *indicator* europei, selezionati, semplificati e adattati al contesto nazionale. L'obiettivo è stato di assicurare una maggiore coerenza rispetto alla varietà organizzativa e alla diversa complessità produttiva che caratterizza gli enti del Sistan, in termini di dimensioni, organizzazioni e dotazioni. I criteri permettono di definire parametri standard per la valutazione comparabile delle evidenze raccolte, offrendo un quadro interpretativo in cui fare emergere tanto aspetti trasversali quanto elementi specifici. Essi si concentrano, in particolare, su:

- il contesto politico-istituzionale e il quadro normativo di riferimento;
- il coordinamento tra i diversi attori della statistica ufficiale;
- le procedure operative e i meccanismi di controllo interno;
- le risorse disponibili e la qualificazione del personale;
- l'impegno formale per il miglioramento della qualità;
- l'accesso alle informazioni e la capacità di soddisfare le esigenze informative degli utenti;

- il dialogo con i rispondenti e la protezione dei dati personali;
- l'organizzazione dei processi e la qualità dei prodotti.

La struttura attuale del *Codice italiano* si articola secondo un impianto fondato su tre macroaree (Contesto istituzionale, Processi statistici, Prodotti statistici), ciascuna composta da un insieme di principi, a loro volta declinati in un numero variabile di criteri guida. Questa articolazione è riportata sinteticamente nella Tavola 4.2, che offre un quadro numerico della struttura con cettuale del *Codice italiano* nella sua seconda edizione.

Tavola 4.2 - I numeri del *Codice italiano*

AREA	Principi	Criteri
Contesto istituzionale	1 Indipendenza professionale	5
	1bis Coordinamento e cooperazione	2
	2 Mandato per la rilevazione di dati e l'accesso ai dati	5
	3 Adeguatezza delle risorse	3
	4 Impegno a favore della qualità	5
	5 Riservatezza statistica e protezione dei dati	6
Processi statistici	6 Imparzialità e obiettività	4
	7 Solida metodologia	3
	8 Procedure statistiche appropriate	3
	9 Onere non eccessivo sui rispondenti	5
	10 Efficienza rispetto ai costi	4
Prodotti statistici	11 Pertinenza	2
	12 Accuratezza e attendibilità	3
	13 Tempestività e puntualità	3
	14 Coerenza e confrontabilità e integrazione	4
	15 Accessibilità e chiarezza	4

Fonte: Direttiva Comstat n. 12, pubblicata nella GURI, Serie Generale, n. 23 del 29 gennaio 2022

L'introduzione dei criteri ha favorito l'adozione di un approccio di autoregolazione, volto strategicamente a stimolare il dialogo e il confronto tra i produttori di statistica ufficiale, incoraggiando la condivisione di buone pratiche e il trasferimento di competenze. In questa prospettiva, il *Codice italiano* si configura come uno strumento abilitante per una produzione statistica coerente, coordinata ed efficace su scala sia locale sia nazionale. Come sottolineato nel 2021 dal Comitato di garanzia per l'informazione statistica (Cogis), nella sua espressione di parere favorevole alla nuova edizione del *Codice italiano*, i criteri permettono di "realizzare una programmazione locale delle statistiche di interesse, basata su criteri validi anche a livello internazionale, la cui qualità possa essere verificata in maniera trasparente secondo i principi della statistica ufficiale".

Un elemento di ulteriore interesse, infine, è rappresentato dal confronto tra gli *indicator* previsti dal *Codice* e i criteri elaborati nel contesto del *Codice italiano*, come riportato nella Tavola 4.3.

Tavola 4.3 - Confronto tra *indicator* europei e criteri nazionali

MACROAREA	Principi	N.ro <i>Indicator</i>	Criteri	Differenza %
Contesto istituzionale	1 - 6	37	30	-18,9
Processi statistici	7 - 10	24	15	-37,5
Prodotti statistici	11 - 15	23	16	-30,4
Totale	1 - 15	84	61	-27,4

Fonte: Istat

Dalla comparazione emerge una riduzione significativa nel processo di adattamento a livello nazionale. Tale riduzione, pari complessivamente al 27,4 per cento, non riflette un indebolimento del quadro normativo nazionale, bensì una selezione mirata e ponderata delle disposizioni europee maggiormente rilevanti e applicabili nel contesto italiano.

A differenza degli indicatori europei, concepiti per autorità statistiche nazionali generalmente dotate di una missione statistica esplicita, i criteri italiani devono risultare funzionali per la molteplicità dei soggetti del Sistan, caratterizzati da livelli di maturità statistica molto eterogenei e nei quali spesso la funzione statistica ha un ruolo sussidiario rispetto ad altre competenze istituzionali. Di conseguenza, i criteri italiani sono stati riformulati in modo da preservare la coerenza con i principi del *Codice*, ma anche per essere applicabili in modo efficace all'interno delle realtà operative dei singoli enti del Sistan, tenendo conto della diversità in termini di dimensioni, struttura organizzativa, disponibilità di risorse e vocazione statistica.

Si osserva, in particolare, che la riduzione più contenuta si riscontra nella macroarea del Contesto istituzionale, che da sola raccoglie circa la metà del totale dei criteri previsti dal *Codice italiano*. Tale evidenza riflette la centralità strategica di questa dimensione nell'ambito della governance del Sistan, che la rende particolarmente rilevante e meritevole di un'attenzione specifica. Non a caso, durante la fase di consultazione pubblica che ha preceduto la definizione del *Codice italiano*, la maggior parte delle osservazioni e dei contributi ricevuti dagli enti coinvolti ha riguardato proprio i principi relativi a questa macroarea. Tale attenzione riflette una chiara consapevolezza del ruolo strategico giocato da elementi quali l'indipendenza professionale, la cooperazione tra enti, la protezione dei dati, l'allocazione delle risorse e l'impegno formale verso la qualità, nel sostenere la credibilità e l'autorevolezza della statistica pubblica in Italia.

4.2 La Guida per l'implementazione del *Codice italiano* per la qualità delle statistiche ufficiali

Il Quadro comune per la qualità del Sistema statistico europeo (*Common Quality Framework*) si fonda su una struttura integrata a tre livelli, concepita per garantire coerenza, trasparenza e affidabilità nella produzione delle statistiche europee. Non si esaurisce nell'adozione del *Codice*, che con i suoi 16 principi e corrispondenti 84 indicatori fornisce i primi due livelli della gerarchia (cfr. paragrafo 1.1).

Il terzo livello, invece, non è contenuto all'interno del *Codice*, ma è sviluppato attraverso un documento separato: il *Quality Assurance Framework of the European Statistical System* - ESS QAF (ESS 2019). Esso ha la funzione di tradurre i principi e gli indicatori in 439 metodi applicativi e buone pratiche, offrendo così una guida concreta per l'implementazione del quadro europeo per la qualità da parte delle autorità statistiche nazionali (Figura 4.1).

L'intenzione di sviluppare anche in Italia questo terzo livello operativo, e quindi completare la convergenza all'impianto europeo, è stata rafforzata dai risultati di un test pilota realizzato con il coinvolgimento diretto delle sedi territoriali dell'Istat.

L'iniziativa ha interessato un campione di enti locali distribuiti su tutto il territorio nazionale e di diversa scala istituzionale, con l'obiettivo di verificare l'appropriatezza, comprensibilità e applicabilità dei criteri contenuti nella seconda edizione del *Codice italiano*. I risultati del test hanno fornito indicazioni incoraggianti, mettendo in luce una buona capacità di comprensione del contenuto concettuale dei criteri da parte degli enti partecipanti.

Figura 4.1 - La struttura del Quadro comune per la qualità del Sistema statistico europeo



Fonte: Istat

Tuttavia, è emersa in maniera trasversale, in particolare presso i comuni, una certa difficoltà nel tradurre i criteri in azioni operative concrete, soprattutto in assenza di esempi pratici o strumenti specifici di supporto. Questo riscontro ha evidenziato la necessità, largamente condivisa dagli interlocutori istituzionali, di rafforzare il supporto applicativo al processo di implementazione del *Codice italiano*.

Nella fase di approvazione della seconda edizione del *Codice italiano*, tutti gli organismi istituzionali coinvolti nel processo Comstat, Cogis e il Dipartimento della funzione pubblica hanno sottolineato l'opportunità di affiancare al *Codice italiano* due documenti interpretativi di supporto: una guida per l'implementazione del *Codice italiano*, finalizzata a facilitare la concreta attuazione dei principi e dei criteri; un manuale per la standardizzazione del *quality reporting*, concepito per uniformare le modalità con cui gli enti del Sistan rendicontano la qualità delle statistiche prodotte (cfr. paragrafo 4.3).

In risposta a tali esigenze, l'Istat ha predisposto e pubblicato la *Guida per l'implementazione del Codice italiano per la qualità delle statistiche ufficiali* (Istat 2023), un documento ad alto contenuto operativo, che costituisce il complemento applicativo del *Codice italiano*. La *Guida* si configura come strumento operativo di supporto per gli enti del Sistan, offrendo indicazioni concrete, esempi applicativi, buone pratiche già collaudate e strumenti di lavoro, al fine di agevolare la traduzione operativa dei contenuti del *Codice italiano*. Per ciascuno dei 61 criteri guida previsti dal *Codice italiano*, la *Guida* associa uno o più metodi applicativi, definiti sulla base dell'esperienza maturata a livello nazionale e internazionale. Complessivamente, il documento raccoglie 301 metodi, articolati in modo da garantire flessibilità nell'adattamento ai contesti organizzativi eterogenei che caratterizzano il Sistan.

4. Azioni volte ad assicurare la qualità nel Sistema statistico italiano

Ciascun metodo rappresenta una modalità concreta con cui il criterio può essere realizzato, declinandolo in azioni osservabili, descrivibili e, se possibile, documentabili.

Il Prospetto 4.1 illustra, a titolo esemplificativo, la struttura gerarchica del *framework* italiano: in corrispondenza del principio 2 del *Codice italiano* (Mandato per la rilevazione dei dati e l'accesso ai dati), il criterio 2.1 viene declinato in quattro metodi operativi, che ne specificano le modalità di attuazione.

Prospetto 4.1 - Esempio di struttura del *framework* italiano per la qualità

PRINCIPIO	Criterio	Indicatori
2 Mandato per la rilevazione di dati e l'accesso ai dati	2.1 Il mandato per la rilevazione di dati provenienti da più fonti e per l'accesso a tali informazioni ai fini dello sviluppo, della produzione e della diffusione delle statistiche ufficiali è sancito dalla legge	2.1.1 La normativa statistica nazionale stabilisce il mandato dell'ufficio di statistica di raccogliere e accedere ai dati provenienti da più fonti, compresi quelli di fonte amministrativa, per finalità statistiche
		2.1.2 La normativa relativa al mandato di raccolta e accesso ai dati è disponibile nella pagina web dell'ufficio di statistica.
		2.1.3 L'US rende pubblici i documenti che specificano i requisiti metodologici, tecnologici, amministrativi che giustificano la necessità di raccolta e accesso ai dati.
		2.1.4 L'ufficio di statistica esplora regolarmente nuove fonti di dati, comprese quelle di proprietà privata, e ne valuta la possibilità di accesso e di utilizzo per la produzione di statistiche ufficiali.

Fonte: Istat (2023)

La *Guida* rappresenta il corrispettivo nazionale del QAF europeo. Si tratta, in sostanza, di un repertorio strutturato di raccomandazioni e suggerimenti operativi, che l'ente del Sistan può adattare in funzione delle proprie caratteristiche organizzative, del livello di maturità statistica raggiunto e del ruolo svolto all'interno del sistema.

Oltre a svolgere una funzione di supporto all'applicazione pratica dei principi e dei criteri, la *Guida* si configura anche come uno strumento utile per effettuare un'autovalutazione del grado di adesione dell'ente al *Codice italiano*. In tale senso, essa può essere utilizzata sia per orientare interventi correttivi, sia per evidenziare punti di forza e buone pratiche replicabili in altri contesti.

Va tuttavia sottolineato che i metodi descritti nella *Guida* non sono da intendersi in alcun modo vincolanti o prescrittivi: non si tratta di un insieme di obblighi normativi, ma di un ventaglio di opzioni, tra le quali gli enti possono selezionare quelle più coerenti con le proprie capacità e priorità strategiche. Alcuni metodi, infatti, risultano maggiormente pertinenti per enti dotati di strutture statistiche consolidate, con un margine significativo di autonomia progettuale nella produzione statistica (attività autodiretta); altri, invece, possono apparire meno rilevanti o applicabili per enti che si occupano unicamente di attività statistiche di tipo esecutivo o routinario. Proprio per questo motivo, il documento è stato concepito per accompagnare gli enti lungo un percorso graduale e realistico, nel quale i criteri del *Codice italiano* rappresentano traguardi da raggiungere progressivamente, senza imporre standard irrealistici o penalizzare le differenze strutturali esistenti all'interno del Sistan. In questa prospettiva, anche i metodi più ambiziosi o avanzati – inizialmente percepiti come difficilmente realizzabili – possono fungere da *benchmark* utili per definire le traiettorie di crescita, innovazione e rafforzamento della funzione statistica all'interno degli enti, contribuendo a una visione condivisa della qualità.

Una particolare attenzione, nell'elaborazione della *Guida*, è stata riservata alla dimensione territoriale della statistica ufficiale. Gli enti locali rappresentano infatti snodi fondamentali per l'arricchimento dell'offerta informativa statistica, in grado di cogliere con maggiore dettaglio e tempestività i bisogni conoscitivi espressi a livello subnazionale, tanto dai decisori pubblici quanto dalla cittadinanza attiva.



Sul piano metodologico, la redazione della *Guida* si è ispirata ai principali riferimenti europei ed extraeuropei in materia di *quality framework*, assicurando una piena coerenza con i modelli più accreditati a livello globale. In particolare, sono stati presi in considerazione, oltre all'ESS QAF, il *Data Quality Assessment Framework* (IMF 2012), lo *United Nations Quality Assurance Frameworks Manual For Official Statistics* (United Nations 2019) e il *Code of Practice for Statistics* UK (UK Statistics Authority and Office for Statistics Regulation 2018).

Il processo di elaborazione è stato caratterizzato da una marcata impronta partecipativa. All'interno dell'Istat, la redazione della *Guida* ha visto il coinvolgimento coordinato di più strutture e competenze, in un approccio interdipartimentale fondato sulla condivisione delle conoscenze e delle esperienze. Parallelamente, anche gli enti del Sistan sono stati attivamente coinvolti, attraverso consultazioni mirate e momenti di confronto che hanno permesso di raccogliere contributi, osservazioni e proposte di integrazione. Questo lavoro collegiale ha consentito di calibrare i contenuti della *Guida* sulle esigenze reali degli utilizzatori, rafforzandone la rilevanza e la praticabilità.

Al termine di questo percorso articolato, la *Guida* è stata formalmente approvata dal Comstat nel maggio 2023. Il documento è oggi disponibile sul sito istituzionale dell'Istat, quale risorsa accessibile al servizio di tutto il Sistan.

4.3 Il manuale per il *quality reporting*

Nell'ambito del Sistema statistico europeo, il *quality report* è il documento, separato dal prodotto statistico, che raccoglie i metadati referenziali e rappresenta un elemento fondamentale per garantire la trasparenza e la comparabilità dei dati statistici, nel rispetto dei principi di qualità.

Nel tempo, l'armonizzazione della struttura e dei contenuti della documentazione relativa ai metadati referenziali e alle informazioni sulla qualità dell'output ha portato alla costruzione di un *set* di concetti: a livello europeo, lo standard di riferimento attuale per la compilazione e lo scambio di metadati e informazioni sulla qualità è rappresentato dal *Single Integrated Metadata Structure* (SIMS) v2.0 (Eurostat 2015b), ulteriormente dettagliato nello *European Statistical System handbook for quality and metadata reports* (Eurostat 2021).

L'attuale versione 2.0 di SIMS include 19 concetti e 80 sottoconcetti, organizzati gerarchicamente come esposto nel Prospetto 4.2. Il percorso di adeguamento del quadro di riferimento italiano ha necessitato di un ultimo passo: iniziare dagli standard europei in materia di *quality reporting* per definire un *template* italiano per la reportistica sulla qualità, con la redazione di un manuale che ne definisse le linee guida per la compilazione.

Tale manuale rappresenta il documento operativo di riferimento, in ambito Sistan, per la standardizzazione della raccolta del *set* essenziale di metadati sulla qualità dei processi e dei prodotti statistici, per la corretta interpretazione dei dati diffusi. In particolare, gli obiettivi 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 i prodotti;
- garantire una diffusione chiara e standardizzata dei dati, consentendone un corretto utilizzo e interpretazione.

4. Azioni volte ad assicurare la qualità nel Sistema statistico italiano

Prospetto 4.2 - La struttura del SIMS

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

Fonte: Eurostat (2015)

Ultimata la fase redazionale, il manuale è attualmente in corso di pubblicazione. Una volta pubblicato, assicurando la coerenza dei *quality report*, consentirà il raggiungimento della piena conformità con il *framework* europeo, migliorando la qualità e la fruibilità dell'informazione statistica prodotta nell'ambito del Sistan.

Circa il processo di definizione del documento, considerata l'eterogeneità delle realtà statistiche nel Sistan, si è ritenuto opportuno non limitarsi ad adottare gli standard europei, bensì procedere a un esame capillare delle pratiche relative alla diffusione dei metadati presso i vari enti e arrivare a un sottoinsieme di metadati più utile e adatto per gli specifici scopi nazionali.

L'attività si è svolta secondo le seguenti fasi:

- 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 partire da settembre 2023, l'analisi si è focalizzata sui metadati presenti nel Programma statistico nazionale (PSN). Ci si è concentrati sulle statistiche non Istat di maggiore rilevanza, conducendo un'analisi dettagliata su 379 lavori.

Nella pratica, sono stati esaminati i metadati dei lavori pubblicati sui siti ufficiali dagli enti del Sistan. L'obiettivo era raccogliere tutti i metadati disponibili e, al contempo, identificare quali fossero i metadati ritenuti più rilevanti dagli uffici di statistica e quindi necessari per la pubblicazione. Al termine di questa attività, si è scelto di concentrarsi su un insieme ridotto di 37 concetti sui 99 complessivamente previsti dal SIMS v2.0, con l'obiettivo di rendere più efficace e sostenibile il processo di redazione del *quality report* italiano.

Tale selezione risponde a 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 ricorrenza di tali concetti nei metadati già in uso dai soggetti del Sistan, valorizzando elementi informativi già consolidati nella prassi nazionale.

Il Prospetto 4.3 presenta il *template* italiano e la corrispondente 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.

Una volta definito il *template*, si è proceduto a utilizzare le informazioni relative ai lavori del PSN già raccolte dall'Istat. Per ogni lavoro contenuto nel PSN, l'ente titolare deve compilare una apposita scheda, contenente le principali informazioni. Tale scheda svolge il ruolo di *repository* di metadati, dove i responsabili degli uffici di statistica degli enti Sistan forniscono metadati fondamentali per la definizione del PSN. Si è proceduto quindi a mappare la scheda PSN con i 99 *item* di SIMS e, di conseguenza, con i 37 concetti del *quality report* italiano.

4. Azioni volte ad assicurare la qualità nel Sistema statistico italiano

Prospetto 4.3 - Template italiano e mappatura con i concetti del SIMS

Versione inglese		Versione italiana		SIMS
1 st level	2 nd 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		Email	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

Come illustrato nel Prospetto 4.4, risulta che 25 (il 68 per cento) dei 37 concetti che compongono il *quality report* italiano sono già presenti nella scheda PSN. La presenza di una tale quota di concetti comuni tra PSN e SIMS, già standardizzati, ha portato all'esigenza di sfruttare il patrimonio informativo del PSNPlus² in due modi: 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*.

² PSNPlus è una piattaforma dedicata al PSN. La piattaforma consente l'accesso ai sistemi informativi di supporto alla raccolta, gestione e consultazione di dati della programmazione statistica nazionale puntando a integrare le informazioni a livello di ente titolare dei lavori statistici e di ufficio di statistica.

Prospetto 4.4 - Campi del *quality report* italiano presenti nella scheda PSN

Primo livello	Secondo livello	Presente nella scheda PSN
Contatti	Ente	Sì
	Unità organizzativa	Sì
	Responsabile	Sì
	Funzione del responsabile	Sì
	Email	Sì
	Telefono	Sì
Ultimo aggiornamento		
Descrizione dei dati	Caratteristiche dei dati	Sì
	Classificazioni	Sì
	Variabili statistiche	Sì
	Unità statistica	Sì
	Popolazione statistica	Sì
	Territorio di riferimento	Sì
	Copertura temporale	
	Unità di misura	
Descrizione del processo statistico	Periodo temporale di riferimento	
	Fonte dei dati	Sì
	Periodicità della raccolta dei dati	Sì
	Periodicità della diffusione	Sì
	Raccolta dei dati	Sì
Mandato istituzionale	Trattamento dei dati	Sì
		Sì
		Sì
		Sì
		Sì
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	Sì
	Pubblicazioni	Sì
	Portale online	Sì
	Altro	Sì
Documentazione metodologica		
Costi e onere sui rispondenti		
Costi		
Commenti		

Fonte: Istat

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

Dopo l'individuazione del sottoinsieme di contenuti utile ai fini Sistan, in analogia con l'*European Statistical System handbook for quality and metadata reports*, sono state predisposte le linee guida operative per la compilazione del *quality report* secondo il formato standardizzato.

4.4 Sviluppi futuri

Con il completamento del processo di allineamento dell'impianto nazionale al *framework* europeo, è attualmente allo studio lo sviluppo di procedure operative per il monitoraggio della qualità nel Sistan. Si tratta di un passaggio essenziale per consolidare un linguaggio comune sulla qualità, promuovere la consapevolezza del valore della trasparenza e della replicabilità dei processi statistici e costruire un sistema coerente di garanzie metodologiche, etiche e organizzative.

È opportuno recuperare, con i necessari adattamenti, l'esperienza di monitoraggio condotta a valle della prima edizione del *Codice italiano*. Tali attività contribuirono sensibilmente alla diffusione della cultura della qualità all'interno degli uffici di statistica, creando un ambiente favorevole alla cooperazione istituzionale e al miglioramento continuo. Pur disponendo unicamente degli enunciati dei principi, fu allora possibile avviare un processo strutturato di autovalutazione, mediante la somministrazione di un questionario di autovalutazione e di una traccia di intervista semistrutturata per le visite di *Peer review*, finalizzate alla valutazione dell'effettivo grado di recepimento del *Codice italiano* nelle attività degli uffici.

A distanza di anni e in un contesto istituzionale profondamente evoluto, sono stati messi a punto e formalizzati gli strumenti illustrati nei precedenti paragrafi, più completi e integrati, che consentono di attuare una procedura sistematica e rigorosa di valutazione della qualità. Il futuro prossimo richiede la piena messa in opera degli strumenti predisposti dall'Istat, in un'ottica integrata e strategica, e in maniera incessante e incisiva.

Tra i principi guida del piano programmatico, riveste particolare valenza quello della gradualità, da intendersi come costruzione progressiva e sostenibile di un percorso capace di generare impatti misurabili nel breve, medio e lungo periodo. In questa logica, le azioni iniziali dovranno essere concrete e pragmatiche, privilegiando interventi semplici ma ad alta efficacia, in grado di generare risultati immediati e tangibili, creando le condizioni favorevoli per l'attivazione di dinamiche virtuose.

Nel breve termine, rivestono un ruolo centrale le iniziative di comunicazione e sensibilizzazione, quali conferenze e seminari a livello nazionale e internazionale. Tali eventi mirano a favorire una piena comprensione degli strumenti metodologici e operativi recentemente elaborati. È fondamentale che tali strumenti non solo siano conosciuti, ma anche compresi in profondità e utilizzati con consapevolezza da parte degli operatori, nell'ambito delle attività quotidiane degli uffici di statistica.

Una delle prime azioni da promuovere in questa direzione riguarda l'avvio dell'utilizzo del *template* unico per il *quality reporting*, per la documentazione standardizzata dei processi e dei prodotti statistici.

L'impiego di questo strumento rafforza la capacità degli enti di descrivere le operazioni statistiche in modo trasparente, organico e confrontabile. Si tratta di un'attività chiaramente definita nei suoi contorni operativi, priva di ambiguità interpretative, che proprio per questo può costituire un primo passo efficace verso l'integrazione sistematica delle pratiche orientate alla qualità. Nel caso dei lavori previsti nel PSN, l'utilizzo del *template* consente di armonizzare e valorizzare informazioni già fornite dagli enti nella compilazione delle schede PSN. In tale modo, si promuove una maggiore coerenza tra gli strumenti di programmazione e quelli di documentazione, facendo leva su un formato collaudato e pienamente conforme allo standard europeo SIMS. In prospettiva, la pubblicazione dei *quality report* come *set* di metadati associati ai dati pubblicati sul portale *Open Data Hub* della statistica

pubblica³ costituirebbe un'ulteriore evoluzione, rafforzando la trasparenza, la visibilità e il riconoscimento dell'impegno degli enti.

Un altro principio guida deve essere la valorizzazione sistematica delle informazioni già disponibili presso l'Istat, il cui impiego dovrebbe rappresentare la prassi nella definizione delle relazioni con le diverse tipologie di enti del Sistan. Salvo casi eccezionali e adeguatamente motivati, è essenziale evitare di imporre ulteriori oneri agli uffici di statistica, nel rispetto del principio di proporzionalità e sostenibilità operativa. In questa logica, i risultati della Rilevazione sugli elementi identificativi, risorse e attività degli uffici di statistica del Sistan (Eup) offrono un patrimonio conoscitivo prezioso, che merita di essere pienamente analizzato e utilizzato.

I dati di tale rilevazione, infatti, consentono di sviluppare un'analisi articolata delle attività svolte dagli uffici di statistica, delle modalità di impiego e diffusione delle informazioni prodotte, e soprattutto del grado di conoscenza e applicazione delle competenze statistiche all'interno degli enti. Tali evidenze costituiscono un solido punto di riferimento per orientare con maggiore efficacia le iniziative di supporto, formazione e sviluppo, adattandole ai bisogni concreti e specifici degli enti.

Queste attività di breve periodo rappresentano la base preparatoria, propedeutica all'avvio di una successiva valutazione strutturata, attualmente in fase di progettazione, finalizzata a verificare l'effettivo grado di allineamento delle pratiche statistiche adottate dagli enti rispetto al quadro concettuale delineato.

In una prospettiva di medio-lungo periodo - secondo quanto auspicato dal Comitato Qualità dell'Istat, che indica il triennio 2026-2028 per la definizione e l'avvio della fase attuativa - sarà necessario avviare un sistema strutturato di monitoraggio dei livelli di adeguatezza delle attività statistiche degli enti e degli uffici del Sistan, assegnando massima priorità agli obiettivi del PSN. L'obiettivo è identificare punti di forza, rilevare pratiche virtuose replicabili e far emergere criticità e spazi di miglioramento. A tale fine, occorre definire un disegno di osservazione sistematica del rispetto, della attuazione e della verifica dello stato di avanzamento nell'applicazione dei principi, dei criteri e dei metodi previsti dal *framework* nazionale per la qualità.

La filosofia sottesa questo approccio è quella del miglioramento continuo: la fase di valutazione dello stato dell'arte deve essere seguita da un'azione concreta, orientata al consolidamento dei risultati positivi e all'innovazione dei processi, mediante interventi mirati alla risoluzione delle criticità emerse. In analogia con le esperienze delle *Peer review* europee (cfr. paragrafo 1.4), il piano prevede che al termine di ogni ciclo completo siano formulate raccomandazioni, individuate priorità e delineate linee di intervento, funzionali a un progressivo innalzamento degli standard qualitativi. La reiterazione regolare di queste fasi consentirà di monitorare l'effettiva implementazione delle azioni di miglioramento e di valutarne l'impatto. In questo processo, i sistemi di misurazione rivestono un'importanza strategica, poiché forniscono evidenze concrete su cui basare decisioni informate. Consentono di individuare in modo oggettivo le aree critiche, aiutano a rilevare le deviazioni dagli standard e misurano l'efficacia delle azioni intraprese.

Il monitoraggio assume anche una funzione di indirizzo strategico, contribuendo a individuare strumenti e indicazioni operative da sottoporre a consultazione pubblica e da sistematizzare in linee guida, a cura dell'Istat o del Comstat, con l'obiettivo di assicurarne la diffusione e l'adozione a livello di sistema.

3 Cfr. <https://sistanhub.istat.it/databrowser/>.

4. Azioni volte ad assicurare la qualità nel Sistema statistico italiano

In ogni caso, è importante ribadire con chiarezza che il miglioramento della qualità della statistica ufficiale non può essere inteso come un adempimento formale, né come un obiettivo da raggiungere una tantum. Si tratta, piuttosto, di un impegno strutturale e permanente, che deve tradursi in azioni coerenti, risorse dedicate e visione strategica. La disponibilità di strumenti metodologici avanzati, per quanto necessaria, non è di per sé sufficiente: è solo attraverso un investimento continuo, sistematico e incisivo che è possibile promuovere un effettivo consolidamento della qualità nei processi e nei prodotti statistici.

L'adozione del nuovo impianto valutativo assume una valenza strategica di particolare rilievo. Da un lato, esso rappresenta uno strumento trasparente di rendicontazione nei confronti degli utenti finali e dei portatori di interesse, rafforzando la legittimazione e la credibilità dell'attività statistica pubblica. Dall'altro lato, esso costituisce una preziosa opportunità per gli enti coinvolti, che possono avvalersene per analizzare in profondità le proprie pratiche, individuare le aree di forza e di debolezza, valorizzare le esperienze virtuose già maturate e orientare con maggiore efficacia i futuri interventi di innovazione e miglioramento.

In sintesi, il piano di azione per il rafforzamento della qualità nel Sistan può essere riassunto nel Prospetto 4.5.

Prospetto 4.5 - Qualità Sistan: obiettivi e azioni dal breve al lungo periodo

Periodo	Azioni principali
Breve	Comunicazione e sensibilizzazione sugli strumenti per la qualità. Valorizzazione dei dati esistenti in Istat. Integrazione tra programmazione (schede PSN) e documentazione (<i>template di quality reporting</i>) Potenziamento della formazione per neoassunti in Istat. Progettazione del sistema strutturato di valutazione della qualità.
Medio	Attivazione dei percorsi di formazione e qualificazione professionale Avvio dell'uso del <i>template</i> unico di <i>quality reporting</i> . Aggiornamento linee guida Sistan del 2018 Finalizzazione del sistema strutturato di valutazione della qualità
Lungo	Implementazione del sistema di valutazione della qualità - Avvio cicli valutativi sul rispetto del <i>framework</i> - Formulazione di raccomandazioni - Intervento a valle dei cicli Pubblicazione dei <i>quality report</i> Linee guida su criticità emerse a cura di Istat/Comstat

Fonte: Istat

In conclusione, il rafforzamento del sistema di garanzie della qualità passa attraverso un disegno integrato di strumenti, metodologie e iniziative formative, ispirato ai valori dell'indipendenza, della responsabilità e della cooperazione interistituzionale. Solo attraverso un impegno congiunto e continuativo sarà possibile consolidare nel tempo un Sistan coeso, competente e capace di rispondere con efficacia alle nuove sfide conoscitive poste dalla società e dal contesto internazionale.

5. VALORIZZAZIONE DELLE ESIGENZE DEGLI *STAKEHOLDER* DELLA STATISTICA UFFICIALE¹

La statistica ufficiale nasce con l'obiettivo di produrre informazioni affidabili e di qualità al servizio della collettività. Per raggiungere tale obiettivo è fondamentale riconoscere e valorizzare le esigenze degli *stakeholder* – cittadini, ricercatori, istituzioni, imprese, media – che utilizzano i dati statistici. La capacità di ascoltare, interpretare e integrare i loro bisogni rappresenta il presupposto per garantire che i prodotti e i servizi offerti dall'Istat risultino pertinenti, utili e di reale impatto. A tale fine, il *Codice delle statistiche europee* (Eurostat e SSE 2017) annovera tra i suoi principi la pertinenza, intesa come capacità delle statistiche di rispondere alle esigenze attuali e potenziali degli utenti, e l'accessibilità, relativa alla facilità con cui gli utenti raggiungono l'informazione statistica di interesse. In questa prospettiva, l'Istat ha avviato diverse attività che riguardano gli utenti finali dei prodotti delle statistiche ufficiali (cfr. paragrafo 5.1). Pur essendo le statistiche ufficiali concepite per rispondere alle esigenze informative degli utenti, un INS è chiamato a valorizzare anche gli ulteriori portatori di interesse coinvolti. Questo Capitolo estende la propria analisi anche ai fornitori di dati, ovvero cittadini, famiglie, imprese, enti e amministrazioni pubbliche rispondenti alle indagini e, non ultimi, gli enti titolari dei dati amministrativi acquisiti dall'Istat a fini statistici. La loro collaborazione è fondamentale per la qualità dei dati di input dei processi di produzione delle statistiche. Nell'ambito della raccolta dei dati, sono diversi i principi di qualità coinvolti: dalla tempestività all'accuratezza, fino alla riduzione dell'onere sui rispondenti (principio 9 del *Codice delle statistiche europee*), solo per citarne alcuni (cfr. paragrafo 5.2). Il processo produttivo è inoltre condotto dal personale dell'Istat, ulteriore interlocutore nella produzione delle statistiche ufficiali. Questi ultimi devono essere informati, coinvolti e sensibilizzati continuamente sulla qualità delle statistiche, anche attraverso una sistematica attività di formazione, affinché la qualità stessa diventi parte integrante del lavoro quotidiano (cfr. paragrafo 5.3). Tra gli elementi da evidenziare, si noti come l'obiettivo di standardizzare metodi e strumenti abbia portato alla definizione di un unico punto di contatto, il *Contact Centre*, sia con gli utenti finali delle statistiche ufficiali (cfr. paragrafo 5.1.1), sia con i rispondenti (cfr. paragrafo 5.2), con notevoli benefici tanto nell'acquisizione quanto nella diffusione dei dati (si pensi, ad esempio, alla possibilità di informare i rispondenti alle indagini rispetto alle statistiche prodotte grazie alle loro risposte e a ulteriori informazioni correlate).

5.1 L'attenzione agli utenti dell'informazione statistica

L'Istat pone da sempre grande attenzione alle esigenze degli utenti dell'informazione statistica, consapevole del ruolo che la statistica ufficiale svolge a supporto della collettività. I dati statistici rappresentano infatti strumenti indispensabili per comprendere la realtà, orientare le politiche pubbliche, sostenere le decisioni individuali o collettive e, in questo modo, promuovere una cittadinanza informata e consapevole.

¹ Il Capitolo 5 è stato redatto da: Roberta Roncati (paragrafi 5.1, 5.1.1 e 5.1.3), M. Francesca Loporcaro (paragrafo 5.1.2), Grazia Di Bella (paragrafi 5.2 e 5.2.12), Gabriella Fazzi (paragrafo 5.2.1), Barbara M. R. Lorè (paragrafi 5.2.2, 5.2.3, 5.2.6 e 5.2.8), Sabrina Barcherini (paragrafi 5.2.4, 5.2.5 e 5.2.7), Paola Bosso (paragrafo 5.2.9), Silvana Curatolo (paragrafo 5.2.10), Giampaola Bellini (paragrafo 5.2.11), Gabriele Ascari (paragrafi 5.3, 5.3.1, 5.3.3), Andrea Bruni (paragrafo 5.3.2).

Attraverso le attività di ascolto, le consultazioni e il monitoraggio delle richieste l'Istat raccoglie e analizza costantemente i bisogni informativi espressi dagli utenti. Questa attenzione ha portato l'Istituto ad ampliare e progressivamente diversificare i canali di accesso ai dati e alle informazioni, con l'obiettivo di garantire una fruizione più semplice, tempestiva e personalizzata. Attraverso portali tematici, *open data*, visualizzazioni interattive, pubblicazioni digitali e strumenti di analisi l'Istat risponde a una domanda conoscitiva sempre più articolata e differenziata.

Ricercatori, istituzioni, imprese, media, studenti e cittadini trovano nell'offerta dell'Istat un patrimonio informativo ampio e aggiornato, reso disponibile con linguaggi e formati pensati per soddisfare le esigenze di pubblici diversi.

In particolare, una quota significativa degli utenti è rappresentata da ricercatori e accademici, che impiegano grandi volumi di dati a elevato dettaglio territoriale per analisi scientifiche e approfondimenti tematici; questi ricorrono prevalentemente al *data warehouse* dell'Istat, con migliaia di tavole statistiche, navigabili per tema e territorio, e di file di microdati.

A questi si aggiungono le imprese e i professionisti che necessitano di informazioni quantitative per valutare mercati, orientare investimenti o predisporre piani di sviluppo; a tale fine, le risorse più utilizzate sono il sito istituzionale e le banche dati tematiche, in particolare quelle censuarie.

Media e comunicatori utilizzano i dati per produrre informazione tempestiva a beneficio dell'opinione pubblica; fanno uso dei comunicati stampa, di tutta la produzione editoriale dell'Istat e dei sistemi informativi (*Noi Italia*, *Bes*, *Rapporto SDGs*, *Rapporto annuale*) che offrono un quadro sintetico dei principali indicatori economici, sociali e ambientali.

Le pubbliche amministrazioni utilizzano i dati per programmare, monitorare e valutare le politiche pubbliche, in particolare a livello locale, privilegiando dati in formato aperto e leggibile automaticamente.

Cittadini, studenti e associazioni accedono, infine, ai dati dell'Istat per conoscere fenomeni economici, sociali e ambientali che li riguardano direttamente o di cui si occupano; per questa tipologia di utenti il sito istituzionale e gli strumenti di visualizzazione – mappe interattive, infografiche e *dashboard* per rappresentazioni intuitive – sono i canali di accesso più utilizzati.

Ciò premesso, i paragrafi seguenti approfondiscono i principali strumenti e processi attraverso cui l'Istat promuove una relazione efficace con i propri utenti. Il primo focus riguarda l'evoluzione del *Contact Centre*, il servizio multicanale che consente di fornire assistenza qualificata a diversi tipi di utenza. Viene poi affrontato il tema dell'individuazione delle esigenze informative, inquadrato nel contesto del *Codice delle statistiche europee* e analizzato attraverso gli strumenti di consultazione attivati a livello nazionale. Si esamina, infine, il ruolo delle rilevazioni sulla soddisfazione dell'utenza, che rappresentano un elemento chiave per orientare il miglioramento continuo della qualità dei prodotti e dei servizi statistici.

5.1.1 Servizio di assistenza agli utenti tramite il *Contact Centre*

Semplificazione e trasparenza delle comunicazioni con gli utenti sono alla base della creazione del *Contact Centre*, il servizio a cui cittadini, imprese, pubbliche amministrazioni e qualunque altro soggetto può rivolgersi per chiedere supporto nella ricerca di dati, anche europei e storici, possibilità di acquisto di prodotti editoriali, rilascio di microdati, informazioni su eventi e concorsi, segnalazione di errori e malfunzionamenti, presentazione di reclami.

Il primo *Contact Centre* dell'Istat, aperto nel maggio del 2008, rappresentava un progetto innovativo per l'epoca, che anticipò persino l'analogo *Contact Service* di Eurostat, e si articolava in una piattaforma web attraverso la quale gli utenti potevano richiedere informazioni e dati statistici, nonché ricevere assistenza e consulenza personalizzate. Ogni richiesta veniva registrata con un numero progressivo per anno e affidata a un operatore specializzato. Un canale dedicato era riservato all'assistenza nella ricerca di statistiche europee e al supporto nelle metodologie e classificazioni comunitarie; un altro era riservato ai giornalisti, alle cui richieste rispondeva l'ufficio stampa, fornendo dati e organizzando interviste anche attraverso i canali tradizionali, quali email e telefono.

In aggiunta al *Contact Centre*, l'Istat metteva a disposizione due servizi raggiungibili esclusivamente via email: uno sportello per i cittadini, attraverso cui richiedere informazioni su servizi e rilevazioni o presentare reclami e segnalazioni; un servizio dedicato alla vendita di prodotti editoriali e alla sottoscrizione di abbonamenti.

La molteplicità dei canali di contatto con l'Istat – email, telefono, web – non ha mai costituito una criticità: i cittadini sono abituati a comunicare con la pubblica amministrazione attraverso strumenti diversi, così come avviene con i soggetti privati, dal telefono ai *social*, dalle piattaforme web alle email e ai *chatbot*. Tuttavia, un insieme di canali non omogenei nonché presidiati da operatori diversi ha comportato una gestione delle richieste e una modalità di comunicazione con gli utenti a tratti complessa e spesso ridondante.

Si è optato quindi per un approccio multicanale e coordinato, volto a semplificare la relazione con i destinatari e a ottimizzare le risorse interne. Da questa impostazione è nato il nuovo *Contact Centre*, concepito per offrire un servizio uniforme e integrato a tutte le categorie di utenti (ricercatori, docenti, imprese, pubbliche amministrazioni, giornalisti e cittadini) che si rivolgono all'Istat per esigenze di comunicazione e diffusione statistica (cfr. paragrafo 5.2 per le attività di supporto ai rispondenti delle indagini statistiche).

Il nuovo *Contact Centre* dell'Istat è stato realizzato nel 2022 con la messa in esercizio dei servizi di assistenza nella ricerca dati, fornitura di microdati, elaborazioni personalizzate, Eurostat *user support*, servizio di ricerche storiche e bibliografiche, acquisto di prodotti editoriali, sportello per i media e sportello per i cittadini. I servizi sono forniti attraverso una piattaforma accessibile da qualsiasi dispositivo e conforme alla normativa vigente in materia di *privacy*, accessibilità e autenticazione (Figura 5.1).

Figura 5.1 - Home page del *Contact Centre* dell'Istat



Fonte: Istat

Durante lo sviluppo del progetto sono stati definiti alcuni requisiti fondamentali per la nuova piattaforma: la riconoscibilità dell'Istat attraverso una linea grafica coerente; il rispetto della normativa sul trattamento dei dati personali e la garanzia dell'accessibilità digitale per utenti e operatori; la possibilità di dialogo del *back-end* con i sistemi Istat di produzione, ragioneria e amministrazione, in raccordo con fatturazione elettronica, ciclo attivo e Pago PA; l'organizzazione della *knowledge base* al servizio degli utenti; il monitoraggio delle attività e dei servizi erogati dalla piattaforma.

Ciascuno dei servizi offerti agli utenti presenta un percorso differenziato e una dettagliata sezione descrittiva e informativa. L'assistenza è prevista sempre in lingua italiana e inglese. A ogni richiesta viene assegnato un numero progressivo e un operatore dedicato, che segue l'utente fino alla completa risoluzione.

5.1.2 Metodi per rilevare le esigenze informative degli utenti

Il principio 11 del *Codice delle statistiche europee* riguarda la pertinenza e sottolinea l'importanza per i produttori di statistiche ufficiali di soddisfare le esigenze informative attuali e potenziali degli utenti. Questo principio stabilisce che le statistiche devono essere sviluppate, prodotte e diffuse in modo da fornire risposte adeguate alle priorità informative degli utenti, siano essi istituzioni pubbliche, imprese, ricercatori o cittadini. Il *Codice* prevede inoltre meccanismi di consultazione sistematica con gli utenti, indispensabili per identificarne le esigenze, valutarne la soddisfazione e individuare eventuali lacune nei dati disponibili, al fine di assicurare la rilevanza e l'utilità delle statistiche europee. La pertinenza costituisce infatti una dimensione fondamentale della qualità delle statistiche ufficiali, in quanto esprime la capacità dei dati prodotti di rispondere in modo adeguato ai bisogni informativi degli utenti. Una statistica è pertinente quando fornisce informazioni utili, tempestive e coerenti con le esigenze espresse da decisori pubblici, imprese, ricercatori e cittadini, consentendo loro di comprendere i fenomeni di interesse e di prendere decisioni consapevoli. La valutazione della pertinenza richiede un dialogo continuo con gli utenti e un'attenta analisi delle loro priorità, così da orientare i processi di produzione statistica verso la rilevazione di aspetti realmente significativi per la società e l'economia.

In questo contesto, l'Istat sviluppa una pluralità di attività rivolte agli utenti. Gli utilizzatori dei dati vengono consultati sia per conoscerne le esigenze informative attuali e potenziali, sia per valutarne la soddisfazione rispetto ai prodotti e ai servizi di diffusione.

In fase di avvio di ogni processo statistico, l'ascolto sistematico dei bisogni informativi consente di raccogliere, analizzare e definire le esigenze degli utenti, orientando la progettazione dei processi statistici verso la realizzazione di statistiche pertinenti e utili. Questo coinvolgimento iniziale è fondamentale per garantire che le statistiche rispondano a domande concrete e supportino le decisioni di istituzioni, imprese, ricercatori e cittadini. Alla conclusione del processo statistico, la valutazione della soddisfazione degli utenti permette di verificare l'effettiva adeguatezza delle statistiche prodotte rispetto alle aspettative e ai bisogni espressi, fornendo un riscontro prezioso per migliorare la qualità futura e mantenere elevata la pertinenza delle informazioni diffuse.

Il legame tra il programma di lavoro per la produzione delle statistiche e le esigenze degli utenti è attualmente demandato alla procedura di predisposizione del PSN tramite il coinvolgimento dei Circoli di Qualità (CDQ), gruppi di lavoro istituiti per ambiti tematici e composti da esperti Istat (tematici e della raccolta dati) e da rappresentanti degli uffici di statistica del Sistan, che rivestono il duplice ruolo di produttori e di utilizzatori dei dati.

I CDQ sono incaricati di riunirsi per analizzare la domanda e l'offerta di informazione statistica per il PSN; rappresentano uno strumento fondamentale nell'ambito del PSN, in quanto costituiscono momenti di confronto strutturato tra produttori e utilizzatori di statistiche ufficiali. Riunendosi periodicamente, i Circoli analizzano in modo approfondito la domanda e l'offerta di informazione statistica, individuando esigenze emergenti, criticità nei dati disponibili e opportunità di miglioramento. Questo processo consente di allineare i contenuti del PSN alle reali priorità informative degli utenti, favorendo lo sviluppo di statistiche più pertinenti e coerenti con l'evoluzione dei fenomeni sociali ed economici. I Circoli di Qualità rappresentano un canale importante per la rilevazione delle esigenze informative, poiché offrono un contesto di confronto diretto tra produttori di statistiche e un'ampia gamma di utenti qualificati. Tuttavia, la partecipazione a questi Circoli è generalmente limitata a soggetti istituzionali o a categorie di utenti organizzati, e non sempre riesce a intercettare in modo capillare le esigenze informative di tutti i potenziali utilizzatori delle statistiche, quali i singoli cittadini o le piccole imprese. Sebbene i Circoli di qualità siano quindi uno strumento efficace per approfondire le esigenze di utenti strategici e per migliorare l'allineamento del PSN alle priorità emerse, è tuttavia necessario integrarli con altri strumenti di consultazione più ampi, come indagini sulla soddisfazione, questionari online o *forum* pubblici, per garantire una rilevazione completa e rappresentativa dei bisogni informativi dell'intera collettività.

Nel 2011, è stata istituita la Commissione nazionale degli utenti dell'informazione statistica (Cuis) per contribuire al miglioramento della produzione statistica ufficiale. Nell'ambito della Cuis, attiva fino al 2018, si sono svolte diverse attività per la rilevazione delle esigenze informative; la Cuis è stata anche consultata in occasione della ristrutturazione di indagini e i suoi membri hanno partecipato agli incontri dei circoli di qualità, per la predisposizione del PSN. La costituzione di uno *user council* è, inoltre, prevista tra le azioni di miglioramento della *Peer review* 2023 (cfr. paragrafo 1.3).

Un'analisi, seppure indiretta, delle esigenze informative degli utilizzatori dell'informazione statistica e dell'uso che fanno di determinati prodotti e servizi offerti dall'Istat è possibile attraverso i dati sugli accessi ai siti web generati dalle piattaforme di *web analytics*, frutto di combinazioni eterogenee di metriche (dati quantitativi) e dimensioni (attributi qualitativi dei dati).

5.1.3 Rilevazioni sulla soddisfazione

Nel quadro descritto di consolidata attenzione agli utenti, considerati non più meri destinatari dell'azione amministrativa ma interlocutori attivi nel miglioramento della qualità dei prodotti e dei servizi offerti, si inseriscono le rilevazioni sulla soddisfazione per i prodotti e i servizi offerti.

Le rilevazioni di *customer satisfaction*, oltre a essere uno strumento raccomandato nel *Quality Assurance Framework of the European Statistical System* (ESS QAF), sono indicate dalla normativa vigente quali strumenti strategici per orientare i percorsi di miglioramento della Pubblica amministrazione. In particolare, vanno ricordati la direttiva del Ministro della funzione pubblica del 24 marzo 2004 (*Rilevazione della qualità percepita dai cittadini*) e il decreto legislativo 7 marzo 2005, n. 82 (*Codice dell'amministrazione digitale*), aggiornato dalla legge 21 aprile 2023, n. 41, sottolinea la necessità di ascoltare e comprendere le esigenze espresse dal cittadino, prestare attenzione costante al suo giudizio e sviluppare la capacità di dialogo e di relazione tra chi eroga il servizio e chi lo riceve. L'articolo 7 del *Codice dell'amministrazione digitale* dispone inoltre che le pubbliche amministrazioni consentano

agli utenti di esprimere la propria soddisfazione sulla qualità del servizio reso – anche in termini di fruibilità, accessibilità e tempestività – e che i dati risultanti siano pubblicati sui propri siti istituzionali.

Le prime indagini per conoscere la soddisfazione degli utenti rispetto ai prodotti e ai servizi erogati sono state condotte dall'Istat a partire dagli anni Novanta del secolo scorso attraverso questionari cartacei distribuiti nei punti di contatto con il pubblico (biblioteca, centri di informazione statistica). Nel 2004 è stata effettuata per la prima volta una rilevazione online. Dal 2013 le indagini sulla soddisfazione degli utenti per i prodotti e i servizi erogati tramite il sito web istituzionale sono svolte con regolarità, attraverso un questionario online sottoposto ai visitatori del sito. Nel 2021 la rilevazione è stata profondamente rinnovata con l'obiettivo, da un lato, di focalizzare l'attenzione sulla soddisfazione per i dati e i metadati consultati sul sito istituzionale, dall'altro lato, di delineare ipotesi di gruppi di utenza omogenei secondo la tipologia di prodotto utilizzato, degli interessi, delle preferenze, delle abitudini di uso e delle competenze statistiche e informatiche.

Nell'ultima edizione della rilevazione, ai visitatori del sito web dell'Istat è stato sottoposto un questionario composto da 14 domande al fine di conoscere la motivazione della navigazione, i prodotti utilizzati e il livello di soddisfazione per ciascuno di essi, nonché alcune caratteristiche degli utenti, quali la competenza statistica e la frequenza di utilizzo dei dati.

In particolare, è stato chiesto ai rispondenti di indicare il motivo per il quale hanno visitato il sito www.istat.it (studio, ricerca, lavoro, curiosità) e l'argomento prevalente consultato (popolazione e famiglie, salute e sanità, imprese eccetera). È stato, inoltre, chiesto di specificare i prodotti consultati e il livello di soddisfazione rispetto alla facilità di reperimento, l'adeguatezza dei contenuti (pertinenza), la tempestività degli aggiornamenti, la chiarezza della presentazione e, solo per i prodotti contenenti dati, l'accuratezza e l'affidabilità e la possibilità di confronto con altri dati.

Per valutare la soddisfazione è stata scelta una scala usata in genere per valutare atteggiamenti, opinioni o percezioni degli individui (scala di Likert) a parziale autonomia semantica, basata su quattro categorie di affermazioni che esprimono un giudizio da positivo a negativo: " pienamente", "abbastanza", "poco" e "per niente soddisfatto".

Una domanda ha riguardato il grado di fiducia nelle statistiche prodotte dall'Istat e, anche in questo caso, è stata offerta una scala basata su quattro affermazioni che vanno da "piena fiducia" a "nessuna fiducia".

È stato chiesto a ciascun rispondente di indicare quattro caratteristiche utili a profilarlo:

- il tipo di prodotto consultato (*report* e analisi, banche dati, file di microdati);
- la frequenza di consultazione del sito istituzionale (due o tre volte l'anno, mensile o trimestrale, settimanale o quotidiana);
- il livello di alfabetizzazione statistica (basso, intermedio, elevato);
- la competenza informatica (bassa, media, buona).

Altri due quesiti si sono concentrati sui formati dei file (xls, csv, sql, json, eccetera) e i software (SAS, R, Python, eccetera) utilizzati per la lettura e l'elaborazione dei dati. Nell'ottica di individuare le caratteristiche e, soprattutto, le esigenze degli utenti dell'informazione statistica, gli ultimi due quesiti hanno permesso di conoscere l'età e il settore di attività.

L'insieme di queste informazioni – la fiducia nella statistica, la frequenza di uso, la competenza statistica e informatica, le tipologie di file e software utilizzati, il settore di attività – è estremamente utile per delineare gruppi di utenza omogenei rispetto ai prodotti e ai servizi che l'Istat realizza e rende disponibili.

Una sezione a risposta aperta ha infine consentito agli utenti di lasciare commenti, suggerimenti o segnalazioni di qualsiasi tipo. Ogni anno questi testi vengono classificati e analizzati per raccogliere indicazioni precise sulle esigenze degli utilizzatori dell'informazione statistica e sui possibili miglioramenti dei prodotti disponibili sul sito istituzionale.

Dai risultati dell'ultima rilevazione emerge un livello di soddisfazione complessivamente buono. Gli utenti apprezzano soprattutto i grafici e le infografiche, le collane editoriali e il glossario statistico, mentre le statistiche sperimentali risultano meno apprezzate. Gli stessi utenti riconoscono che i dati dell'Istat sono altamente accurati, affidabili e chiari nei contenuti; lamentano invece qualche limite nella facilità di reperimento e nella scarsa confrontabilità tra fonti diverse.

Da maggio 2024 anche sul *Contact Centre* è presente una breve indagine per registrare il livello di soddisfazione riguardo al servizio offerto. Attraverso tale questionario, inviato via email agli utenti automaticamente all'interno del messaggio di chiusura della richiesta, si desidera misurare il grado di usabilità della piattaforma, la chiarezza e completezza delle informazioni fornite dagli operatori, la pertinenza delle informazioni ricevute, la congruità dei tempi di risposta.

Anche da questa rilevazione emerge che gli utenti sono complessivamente soddisfatti dei servizi ricevuti tramite il *Contact Centre*, con particolare riguardo alla chiarezza nelle risposte fornite dagli operatori e ai tempi di attesa contenuti.

5.2 La raccolta dei dati e la qualità

Nel 2016 l'Istat ha avviato un processo di modernizzazione che, tra le diverse iniziative intraprese, ha comportato l'istituzione di una Direzione centrale per la raccolta dati (DCRD) a supporto dei processi statistici dell'Istituto, accentrando le attività relative sia alle indagini statistiche sia all'acquisizione sistematica di dati da fonti amministrative.

La centralizzazione di tali funzioni ha comportato un rilevante sforzo di coordinamento con le diverse componenti dell'Istituto, inclusi i settori tematici di produzione, la comunicazione istituzionale, nonché gli ambiti giuridici ed economici. Tuttavia, questa riorganizzazione ha rappresentato un'opportunità strategica per ottimizzare i processi di raccolta, migliorandone l'efficienza e la qualità. L'analisi che segue si colloca nel più ampio filone metodologico dedicato alla gestione della qualità dell'input statistico e intende sistematizzare le principali dimensioni operative affrontate nel presente contributo.

Negli ultimi anni, gli sforzi si sono concentrati in particolare sulla standardizzazione e sull'automazione dei processi, con l'obiettivo di ridurre il rischio di errore e di razionalizzare tempi e risorse. In tale ottica, sono stati implementati sistemi di controllo e monitoraggio automatico, concepiti per attivare meccanismi di *feedback* in grado di rilevare tempestivamente eventuali anomalie o disfunzioni, consentendo interventi correttivi puntuali.

Un ulteriore ambito di intervento ha riguardato la tracciabilità dei processi, finalizzata a garantire la documentazione sistematica di tutte le fasi di raccolta dei dati. Questo aspetto si rivela particolarmente rilevante per i settori tematici, che utilizzano i dati raccolti come input per la produzione finale delle statistiche e, soprattutto, nel contesto della produzione multifonte, in cui l'integrazione e la coerenza delle informazioni rappresentano elementi chiave per la valutazione della qualità dell'output statistico. La tracciabilità costituisce, inoltre, un elemento essenziale per la corretta gestione del trattamento dei dati personali e per supportare le attività di *audit* e di verifica *ex post*.

A sostegno delle attività di raccolta dati così centralizzate sono stati sviluppati e consolidati nel tempo diversi strumenti informatici (IT *tools*), oggi pienamente operativi, che si configurano come leve fondamentali per il miglioramento continuo della qualità dei processi e dei dati. Nello schema riportato nel Prospetto 5.1 vengono presentati i principali attualmente in uso.

La *Quality Report Card for Administrative Data* (QRCA) è un vero e proprio sistema di monitoraggio e documentazione sistematica della qualità dei dati amministrativi utilizzati come input dei processi (cfr. paragrafo 2.3).

Sigma è il nuovo Sistema di gestione dei microdati amministrativi e statistici: permette il trattamento statistico dei dati di input dei processi nel rispetto della protezione dei dati personali, in particolare modo per i processi di tipo multifonte che richiedono l'integrazione di microdati. Lo svolgimento del trattamento avviene all'interno dei domini specifici di integrazione (DSI), ovvero un'area in cui sono presenti soltanto i dati necessari a conseguire una specifica finalità statistica, per il tempo strettamente necessario, e a cui hanno accesso esclusivamente le persone autorizzate a svolgerlo (Frattarola e Spirito 2025).

Il Sistema integrato dei microdati SIM (Runci *et al.* 2016) svolge le funzioni di caricamento e pseudonimizzazione dei microdati amministrativi e verrà inglobato in Sigma a completa garanzia della protezione dei dati personali secondo la normativa vigente.

Gli altri strumenti riportati nel Prospetto 5.1 sono funzionali allo svolgimento delle attività proprie della direzione e sono richiamati nel presente paragrafo.

In termini di qualità, la struttura centralizzata dell'Istat dedicata alla raccolta dei dati rappresenta un presidio a garanzia della qualità dei dati di input dei processi. Tale qualità condiziona quella dell'output, con impatti sulle dimensioni dell'accuratezza, della tempestività e della puntualità. Gli aspetti della qualità dell'input dei dati grezzi di indagine in termini di accuratezza si sostanziano nella capacità di ottimizzare il rapporto con i rispondenti e, in particolare, gli esiti della rilevazione sia per la riduzione delle mancate risposte totali e parziali, sia per la qualità stessa delle risposte fornite. La tempestività e la puntualità afferenti alle fasi di raccolta dei dati sono continuamente ottimizzate al fine di ridurre il peso nel computo finale degli indicatori della qualità dell'output.

Prospetto 5.1 - Principali strumenti informatici adottati in Istat per la raccolta dati

IT Tool	Scopo
QRCA - <i>Quality Report Card for Administrative Data</i>	Portale web per la documentazione della qualità dei dati amministrativi acquisiti dall'Istat per scopi statistici, disponibile sulla intranet dell'istituto
Sigma - Sistema di gestione dei microdati amministrativi e statistici	Fornisce un ambiente per il trattamento dei dati di input dei processi, in particolare modo i processi multifonte, in modo da ottemperare alle norme vigenti in tema di protezione dei dati personali
SIM - Sistema integrato dei microdati	Effettua i caricamenti e l'integrazione logica dei dati amministrativi (in fase di evoluzione rispetto al nuovo sistema Sigma)
Arcam - Sistema di acquisizione di archivi amministrativi e dati di indagine	Portale web per l'acquisizione dei dati amministrativi composto da un'interfaccia esterna per gli enti e una con gli amministratori, il database dei metadati di gestione, il <i>repository</i> e il sistema di rilascio per le forniture che non contengono dati personali
Statistica & Imprese - Portale Imprese	Sistema centralizzato che offre servizi a supporto dell'impresa nello svolgimento degli adempimenti statistici a partire dall'accesso alla compilazione di tutti i questionari in cui ciascuna l'impresa è coinvolta
GINO - Gestione indagini online	Piattaforma utilizzata per la raccolta dati di indagini statistiche in modalità CAWI (<i>Computer-Assisted Web Interviewing</i>) permettendo agli utenti di compilare e trasmettere i questionari online, in modo sicuro ed efficiente.
SGI - Sistema di gestione dell'indagine	Sistema informatico di supporto alla conduzione della rilevazione
Puc - Punto unico di contatto <i>Salesforce</i>	Piattaforma multicanale per la gestione delle richieste di assistenza

Fonte: Istat

Per quanto riguarda i dati amministrativi, che per definizione sono prodotti all'esterno dell'Istat, si applica il concetto di "usabilità statistica" dei dati che coniuga gli indicatori di accuratezza in relazione alle finalità statistiche e non amministrative, gli indicatori di tempestività e puntualità rispetto all'ente titolare dei dati e aggiunge altre dimensioni specifiche della qualità (cfr. paragrafo 2.3).

Gli *stakeholder* cui il settore della raccolta dati dedica particolare attenzione sono i rispondenti delle indagini statistiche e gli enti che forniscono dati amministrativi a fini statistici. Le strategie adottate nei loro confronti sono illustrate nei paragrafi seguenti.

5.2.1 Approccio centrato sul rispondente: il paradigma per le indagini statistiche

I rispondenti alle indagini sono la fonte primaria per raccogliere informazioni su temi e aspetti per i quali le fonti amministrative sono assenti, parziali o non utilizzabili a scopo statistico. Sia nelle indagini su individui e famiglie sia in quelle rivolte a imprese o istituzioni, è sempre un individuo, o più individui, a compilare i questionari.

Se i rispondenti si rifiutassero di rispondere, per mancanza di fiducia o di motivazione, o se rispondessero in maniera poco accurata, i dati raccolti potrebbero essere non rappresentativi della variabile di interesse sulla popolazione di riferimento a causa di cadute eccessive e non omogenee nel campione o dati incompleti o inaffidabili. Analogamente, anche i rispondenti motivati potrebbero fornire risposte non fedeli qualora il compito richiesto risultasse eccessivo in termini di tempo o di risorse cognitive da impiegare.

A partire da queste considerazioni e in linea con la letteratura consolidata nel campo del *service design*, da alcuni anni l'Istat ha adottato un approccio che pone il rispondente al centro del processo di progettazione delle indagini. Le linee guida per la progettazione di indagini *respondent centred*, definite da Wilson e Dickinson (2022), individuano i processi da attivare fin dalle fasi iniziali, riconoscendo le esigenze non solo degli *stakeholder* tradizionali (Eurostat, normativa europea e nazionale, esigenze istituzionali) ma anche dei rispondenti.

La relazione deve essere curata in tutte le fasi del percorso che ogni rispondente affronta con l'Istat: dalla ricezione della lettera informativa, che lo avvisa di essere incluso in un campione, alla ricerca di maggiori informazioni attraverso il sito Istat o il *Contact Centre*, fino alla fase cruciale di *login* e compilazione del questionario web o di risposta all'intervistatore.

Come vedremo nei prossimi paragrafi, mettere il rispondente al centro non deve essere solo una dichiarazione di intenti, ma deve tradursi in azioni concrete. A titolo di esempio, sono state semplificate e standardizzate le lettere di sollecito per la partecipazione alle indagini, inviate alle unità non rispondenti del campione nel corso della fase di raccolta dati.

Il processo di creazione della pagina web del *Contact Centre* ha preso avvio dalla ricognizione e analisi delle esigenze che profili diversi di utenti potrebbero manifestare, creando un insieme di potenziali *personas*, ossia utenti tipo, secondo un approccio centrale nella *user research* (Cooper 1999) e raccomandato dalle linee guida di Eurostat per la progettazione di strumenti di comunicazione (Karlberg e Gauckler 2020). Il nuovo *Contact Centre* o Punto unico di contatto (Puc) dell'Istat, infatti, integra in un unico spazio le richieste di tutti gli utenti che necessitano di entrare in contatto con l'Istat, siano essi fruitori dei dati (cfr. paragrafo 5.2.1.8) oppure potenziali rispondenti a un'indagine.

Particolare attenzione viene posta, inoltre, alla fase di progettazione e ottimizzazione dei questionari di indagine. L'intervista o la compilazione di un questionario è infatti un "evento cognitivo" (Gobo 2002). Il modello cognitivo del processo di risposta inizia con la comprensione della domanda, il recupero delle informazioni rilevanti, la formulazione di un

giudizio, l'adattamento del giudizio alla struttura prevista e, infine, alla formulazione della risposta finale (Tourangeau e Rasinski 1988). È costante l'attenzione a semplificare le fasi del modello per l'intervistato.

In generale, l'obiettivo è ridurre il *burden*, ovvero il disturbo statistico del rispondente: come descritto nei paragrafi successivi, per le indagini su individui e famiglie il raggiungimento di questo obiettivo è ricercato attraverso il ricorso a interviste cognitive, ascolto dei *feedback* di intervistatori e intervistati, analisi dei paradata² e delle richieste di supporto. Per le imprese questo obiettivo si declina anche attraverso la messa a disposizione di un sistema unico e semplificato di accesso, il portale *Statistica & Imprese* (anche denominato Portale Imprese), avente l'obiettivo di agevolare lo svolgimento dei vari adempimenti statistici a partire dall'accesso fino alla compilazione dei questionari delle indagini in cui le imprese sono coinvolte. Il sistema centralizzato permette di controllare, attraverso la cosiddetta "analisi del pervenuto", il numero di indagini in cui le imprese sono coinvolte. L'analisi dei paradata e di informazioni appositamente raccolte a chiusura del questionario consente, infine, di calcolare il carico statistico sulle unità rispondenti.

Ciascuna fase del percorso andrebbe sistematicamente testata con tecniche qualitative o quantitative su un campione di utenti rappresentativi, operazione che non sempre è praticabile per vincoli di tempo e di risorse. L'obiettivo è estendere progressivamente il ricorso a test utente (cognitivi, di usabilità e altri) a un numero crescente di fasi e di processi statistici, al fine di rilevare le difficoltà e le esigenze dei rispondenti e migliorare la loro esperienza di partecipazione.

5.2.2 Metodologie per l'ottimizzazione dei questionari

Il disegno del questionario rappresenta una fase cruciale del processo di raccolta dei dati statistici, nella quale si concentrano molte delle potenziali fonti di errore di misura. In particolare, emergono la difficoltà di comprensione delle domande, la complessità nella navigazione del questionario, l'inadeguatezza delle categorie di risposta e l'eccessivo carico cognitivo imposto ai rispondenti. Tali criticità possono compromettere la qualità dei dati raccolti, generando risposte distorte, abbandoni della compilazione e una riduzione della partecipazione.

Come affermato da Sudman e Bradburn (1982), nemmeno l'esperienza consente di redigere un questionario perfetto: la qualità della progettazione deve essere garantita da un processo strutturato e iterativo. Pertanto, l'adozione di un approccio *respondent-centred* si configura come una buona pratica metodologica.

Il concetto di usabilità, mutuato dal campo del *design* e dell'ingegneria dell'interazione uomo-macchina, si è rivelato centrale anche nella progettazione dei questionari elettronici. L'ISO 9241-11:2018 definisce l'usabilità come "il grado in cui un prodotto può essere usato da utenti specifici per raggiungere obiettivi specifici con efficacia, efficienza e soddisfazione in uno specifico contesto d'uso" (ISO 2018a). Nell'ambito delle indagini statistiche, ciò implica la costruzione di strumenti che supportino il rispondente nel compito di fornire informazioni corrette, riducendo gli sforzi richiesti e minimizzando gli errori. Un questionario ben progettato dovrebbe operare come un assistente virtuale, in grado di guidare il rispondente, svolgere operazioni per suo conto, segnalare le azioni già effettuate e fornire istruzioni chiare sui passaggi successivi (Nielsen 1993, United Nations 2022).

² Nel processo di rilevazione dei dati con questionari elettronici, per paradata si intendono le informazioni di *log* memorizzate dal software relative all'operatività dell'utente. Ad esempio, numero di accessi, durata delle sessioni, punti di interruzione, percorso di compilazione.

Numerosi studi hanno dimostrato che l'attenzione al rispondente nel processo di progettazione — attraverso tecniche come le interviste cognitive, l'analisi dei paradata, i *de-briefing* degli intervistatori, il *feedback* post-compilazione e il monitoraggio delle richieste di assistenza — consente di identificare e correggere tempestivamente i punti critici del questionario, migliorandone la fruibilità e, di conseguenza, la qualità dei dati raccolti (Barcherini *et al.* 2022, Converse e Presser 1986). Tali evidenze sono particolarmente rilevanti in un contesto di crescente digitalizzazione delle rilevazioni, in cui la modalità di autocompilazione assistita da computer CAWI (*Computer-Assisted Web Interviewing*) è sempre più utilizzata. In questo scenario, i rispondenti si trovano spesso da soli di fronte a un compito che richiede abilità linguistiche, competenze digitali e capacità di *problem solving* non sempre disponibili in maniera omogenea nella popolazione.

Fattori come il livello di istruzione, la familiarità con gli strumenti digitali e il capitale sociale influenzano significativamente la capacità di completare con successo la compilazione di un questionario elettronico. Le analisi condotte dall'Istat mostrano che, le persone con bassa scolarizzazione o prive di una rete di supporto come gli anziani soli, incontrano maggiori difficoltà nella compilazione ma, con strumenti progettati ad hoc e se adeguatamente supportati, sono in grado di portare a termine il compito richiesto (Barcherini *et al.* 2024a). La progettazione del disegno del questionario rappresenta, quindi, un nodo metodologico e organizzativo centrale nel processo di rilevazione statistica, con impatti diretti sulla qualità del dato e sull'efficienza della raccolta dati. Nel 2016, in occasione della costituzione della direzione centralizzata per la raccolta dati, l'Istat ha introdotto una struttura organizzativa che distingue chiaramente i ruoli tra esperti tematici, incaricati della definizione dei contenuti informativi, ed esperti della raccolta dati, responsabili dell'ottimizzazione del disegno del questionario. Per garantire uniformità e tracciabilità nell'intero processo, è stata sviluppata una procedura articolata in fasi, che definisce *step* operativi, input, output, attori coinvolti e rispettive competenze. Tale modello consente una pianificazione più efficace, facilita l'integrazione di competenze multidisciplinari e assicura la documentazione sistematica delle scelte progettuali.

Nei seguenti paragrafi verranno illustrate le strategie messe in atto dall'Istat per ottimizzare la raccolta dei dati da indagine, con un focus particolare sulla scelta delle tecniche di rilevazione e sul disegno dei questionari. L'adozione di un approccio sistematico e multidisciplinare, fondato sull'ascolto del rispondente, si è rivelato fondamentale per garantire dati di alta qualità, riducendo al contempo il *burden* e promuovendo la partecipazione.

Porre il rispondente al centro della progettazione del questionario non è solo una scelta metodologica, ma una condizione essenziale per migliorare la qualità statistica dell'informazione raccolta. Le esperienze dell'Istat mostrano come un approccio sistematico basato su test, analisi e strumenti tecnologici permetta di migliorare l'usabilità dei questionari e la qualità dei dati, minimizzando l'onere per il rispondente. Per perseguire questi obiettivi l'Istat impiega un insieme articolato di tecniche qualitative e quantitative.

5.2.3 Interviste cognitive

Le interviste cognitive rappresentano uno strumento chiave, in quanto consentono di migliorare la qualità del dato attraverso una più profonda comprensione dei meccanismi mentali che guidano le risposte degli intervistati.

Le interviste cognitive trovano le loro radici teoriche nel paradigma dello *Human information processing* (Neisser 1967) e sono state inizialmente sviluppate per migliorare l'attendibilità delle testimonianze in ambito giudiziario (Fisher e Geiselman 1992).



Attualmente sono ampiamente utilizzate nelle scienze sociali e nella statistica ufficiale per testare i questionari. Si tratta di un'intervista in profondità, di tipo qualitativo, che permette di analizzare come i rispondenti interpretano i quesiti, quali strategie cognitive utilizzano per richiamare alla memoria le informazioni richieste e come selezionano le risposte. L'obiettivo è identificare eventuali difficoltà di comprensione, ambiguità o formulazioni fuorvianti e correggerle prima della messa in campo della rilevazione.

Nel contesto dell'Istat, le interviste cognitive sono parte integrante del processo di progettazione e ottimizzazione dei questionari, quando le risorse e i tempi lo consentono. Possono essere realizzate sia in modalità tradizionale (*face to face*), sia in forma remota tramite videointervista, o anche attraverso tecniche autosomministrate come il *web probing*. Le tecniche principalmente utilizzate sono il *think-aloud* (i rispondenti verbalizzano spontaneamente i loro pensieri mentre rispondono ai quesiti) e il *verbal probing* (l'intervistatore pone domande specifiche per esplorare i processi cognitivi del rispondente). Questi approcci, spesso utilizzati in combinazione, offrono una panoramica dettagliata sulle difficoltà insite nella formulazione dei quesiti e sulle possibili fonti di errore.

Un esempio innovativo di utilizzo di questo metodo è rappresentato dall'Indagine Discriminazioni in ambito lavorativo e condizioni di vita delle persone LGBT+ (in unione civile o già unite) - Anno 2020-2021, realizzata nel 2020 grazie a un accordo di collaborazione tra Istat e Ufficio nazionale antidiscriminazioni razziali (UNAR). Data la sensibilità del tema, il questionario è stato progettato in modalità CAWI e le interviste cognitive sono state condotte online, anche in virtù delle restrizioni dovute alla pandemia da Covid-19. Questa modalità ha facilitato il reclutamento di soggetti appartenenti alla popolazione target, ha garantito maggiore *privacy* agli intervistati e ha permesso di individuare quesiti problematici in termini di linguaggio, memoria e struttura. Per ogni domanda selezionata, è stato costruito un percorso di intervista con tecniche e *probe* dedicati, garantendo risultati qualitativi accurati anche in remoto (United Nations 2022; Barcherini *et al.* 2023).

Un altro caso rilevante riguarda l'Indagine multiscopo sugli aspetti della vita quotidiana, dove è stato sperimentato un approccio misto, combinando test quantitativi con approfondimenti cognitivi qualitativi autosomministrati via CAWI. Questo metodo innovativo ha permesso di testare efficacemente e su un ampio campione³ quesiti problematici come quelli sulle preoccupazioni per l'ambiente e quelli sulle abitudini di lettura, rivelando criticità di comprensione, incoerenze nelle risposte e difficoltà legate alla formulazione delle istruzioni (Bologna *et al.* 2020).

Queste esperienze dimostrano che le interviste cognitive, integrate in una logica *respondent-centred*, consentono non solo di migliorare la qualità dei dati raccolti, ma anche di rafforzare la fiducia dei cittadini nella statistica ufficiale.

5.2.4 Feedback degli intervistatori

Tra i metodi per valutare un questionario la consultazione sistematica della rete di rilevazione riveste un ruolo cruciale in quanto l'intervistatore è un depositario privilegiato di informazioni sulle criticità delle indagini, che vanno da aspetti più generici relativi all'organizzazione a quelli più particolari come, ad esempio, la comprensibilità delle domande (Balì *et al.* 2019).

³ Tale approccio ha consentito di intervistare un campione di dimensioni non consuete per questo metodo (920 individui), rendendo così possibile sia lo studio di fenomeni di risposta rari sia l'integrazione di analisi quantitative con approfondimenti qualitativi.

Raccogliere e valorizzare tali segnalazioni rappresenta un passaggio essenziale in un'ottica di miglioramento continuo della qualità e uno tra gli esempi più significativi è rappresentato dal nuovo Censimento permanente della Popolazione e delle abitazioni, che ha preso avvio nel 2018. A conclusione della prima e dell'ultima edizione del primo ciclo censuario (2018 e 2021), è stata effettuata una consultazione che ha coinvolto la rete di rilevazione comunale. Dalla prima consultazione, sono emerse alcune problematiche connesse alla gestione degli strumenti informatici, alla chiarezza del questionario elettronico e alla tutela della *privacy* (Bali e Cangialosi 2024). Tali *feedback* hanno portato all'implementazione di tre interventi volti a migliorare l'usabilità del questionario *computer-assisted* del Censimento (Barcherini *et al.* 2024a):

- i. un *restyling* grafico, finalizzato a ottimizzare la leggibilità a schermo mediante l'eliminazione di elementi visivi distraenti, come l'uso eccessivo del colore rosso, e l'evidenziazione delle parole chiave per facilitare la focalizzazione dell'attenzione;
- ii. una semplificazione della navigazione, ottenuta attraverso la riorganizzazione del menu e l'introduzione di etichette per i pulsanti più intuitive, con l'obiettivo di ridurre il carico cognitivo dell'utente;
- iii. un supporto guidato alla compilazione, che ha razionalizzato l'accesso ai questionari individuali e reso più lineare la procedura di invio finale.

L'impatto delle azioni di miglioramento apportate al questionario è stato valutato nel 2021, in occasione della seconda consultazione della rete di rilevazione comunale. I risultati hanno evidenziato, in maniera pressoché unanime, un giudizio positivo sull'usabilità e la funzionalità del questionario. Più del 90 per cento degli operatori ha dichiarato l'assenza di difficoltà in ogni componente valutata, a conferma dell'efficacia delle migliorie implementate. È emersa inoltre l'importanza dell'assistenza tecnica ricevuta, valutata positivamente nella maggior parte dei casi. Tale risultato è stato ulteriormente corroborato dalla drastica riduzione del numero di segnalazioni ricevute dal *Contact Centre*, con solo lo 0,5 per cento dei ticket riferiti a problematiche di compilazione.

Un altro esempio di applicazione è rappresentato dalla consultazione effettuata a novembre 2020 e rivolta alla rete di rilevazione CAPI della Rilevazione sulle forze di lavoro e dell'Indagine sulle spese delle famiglie. Gli intervistatori sono stati invitati a compilare un questionario di *debriefing* per raccogliere le loro opinioni sulle interviste svolte dall'inizio del periodo di emergenza da Covid-19. Durante questo periodo l'Istat ha dovuto ridefinire le modalità di conduzione delle interviste, integrando la tecnica faccia a faccia con quella telefonica. La consultazione ha avuto lo scopo di rilevare l'impatto e gli effetti di questa nuova modalità, sia sulla relazione degli intervistatori con le famiglie sia sullo svolgimento del loro lavoro in generale. Ciò al fine di evidenziare possibili aree di miglioramento ed eventuali soluzioni.

Le analisi sui risultati raccolti hanno dimostrato che, in generale, la situazione emergenziale vissuta nel 2020 ha reso più difficile il compito dell'intervistatore a causa della maggiore diffidenza da parte delle famiglie, sia per la mancanza di un contatto diretto e visivo, sia per una carenza nella comunicazione ufficiale (informativa da parte dell'Istat, lettera da parte del comune, tesserino di riconoscimento). I risultati confermano quindi molto chiaramente l'importanza strategica della lettera informativa per ottenere la collaborazione delle famiglie e motivarle alla partecipazione.

Nell'intervista condotta al telefono vengono meno le strategie della comunicazione visiva e gestuale che, oltre a permettere di stabilire un rapporto empatico con il rispondente, facilitano il superamento delle barriere linguistiche nel caso di famiglie straniere.

La maggiore difficoltà nell'instaurare un clima di fiducia tra l'intervistatore e l'intervistato, rende inoltre più faticoso somministrare i quesiti ritenuti intrusivi e di natura sensibile come quelli sul reddito e sulle spese sanitarie.

Tuttavia, a fronte di un peggioramento nell'organizzazione del lavoro rispetto al periodo precedente, causato da un generale aumento della complessità dell'organizzazione del lavoro e del carico di lavoro, alcuni intervistatori dichiarano un miglioramento grazie a una riduzione dei tempi di spostamento e al fatto che l'intervista telefonica risulta meno intrusiva rispetto a una visita a domicilio.

5.2.5 *Feedback dei rispondenti*

Un ulteriore strumento che consente di monitorare e analizzare il carico cognitivo e l'esperienza del rispondente, fornendo dati essenziali per ottimizzare l'usabilità del questionario senza incrementare il *burden* complessivo associato alla compilazione è rappresentato dal *feedback* dei rispondenti. Questo può essere analizzato mediante l'inserimento di alcune domande di valutazione standardizzate poste a fine questionario che permettono di rilevare eventuali difficoltà incontrate nella compilazione e raccogliere le preferenze degli utenti rispetto alle modalità di rilevazione adottate.

Nel caso del Censimento permanente della Popolazione e delle abitazioni, è stato fondamentale il contributo della sezione finale del questionario rivolta direttamente ai rispondenti. Le domande standardizzate, relative alle modalità di compilazione, alle difficoltà incontrate e all'eventuale richiesta di assistenza, hanno permesso un'analisi mirata dell'esperienza utente (Barcherini *et al.* 2024a). Il confronto tra i dati del 2019 e del 2021 ha evidenziato un miglioramento in tutti gli indicatori: è aumentata sia la quota di questionari compilati autonomamente dai componenti della famiglia sia la percentuale di famiglie che non ha avuto bisogno di alcun tipo di supporto. Tra gli anziani, notoriamente considerati il gruppo più vulnerabile, si è osservato un incremento nell'autonomia di compilazione e nell'utilizzo della modalità CAWI, segnale del successo delle modifiche introdotte.

L'inclusione di una domanda aggiuntiva nell'edizione del 2021, volta a rilevare i motivi della richiesta di aiuto, ha permesso di identificare che circa un quarto dei rispondenti, soprattutto anziani, ha sperimentato difficoltà riconducibili all'usabilità, come problemi di navigazione o invio. Ciò conferma la necessità di un'attenzione costante al disegno del questionario, con particolare riguardo ai sottogruppi vulnerabili. In sintesi, il *feedback* sistematico dei rispondenti, integrato nelle fasi finali del questionario, si è rivelato uno strumento cruciale per monitorare criticità, valutare l'impatto degli interventi correttivi e orientare lo sviluppo di soluzioni progettuali centrate sull'utente.

5.2.6 *Analisi delle richieste di assistenza ricevute dal Contact Centre*

Una fonte preziosa per comprendere le difficoltà incontrate dai rispondenti nella compilazione dei questionari è rappresentata dalle richieste di assistenza ricevute dal *Contact Centre*. Queste segnalazioni costituiscono una miniera di informazioni utili per individuare criticità nel disegno dei questionari. In fase di revisione degli strumenti di rilevazione, i ticket generati in fase di richiesta vengono selezionati in base alla categoria in cui sono stati classificati (ad esempio, difficoltà di accesso, comprensione delle istruzioni, interpretazione delle domande) e analizzati.

La classificazione delle richieste di assistenza permette di concentrare l'attenzione esclusivamente sui ticket relativi al questionario, facilitando così l'individuazione rapida e mirata dei punti che hanno generato incertezze o errori da parte degli utenti. Un esempio significativo è rappresentato dal processo di revisione del questionario del Censimento permanente delle imprese 2022. Su circa 46 mila segnalazioni ricevute durante l'edizione precedente, circa il 5 per cento riguardava problemi direttamente connessi al questionario. È stato possibile isolare queste segnalazioni e ricavarne indicazioni preziose per migliorare la formulazione delle domande, disambiguare alcuni concetti e perfezionare le istruzioni per la compilazione.

L'impiego sistematico dell'analisi delle richieste di assistenza rappresenta dunque uno strumento efficace per rilevare le esigenze dei rispondenti, migliorare l'esperienza di compilazione, ottimizzare il processo di raccolta dati e, al contempo, la qualità dell'informazione statistica. Con l'attivazione del Puc (cfr. paragrafo 5.2.9), la classificazione delle richieste di assistenza è stata ulteriormente perfezionata, rendendo queste informazioni ancora più facilmente accessibili e integrabili nei processi di progettazione e revisione dei questionari.

5.2.7 Analisi dei paradata

L'analisi dei paradata derivanti dal processo di compilazione del questionario CAWI rappresenta un'altra fonte di informazione preziosa per valutare la fruibilità del questionario, individuare punti critici e stimare la propensione alla risposta.

A titolo esemplificativo, si riportano i risultati dell'analisi effettuata per la seconda edizione dell'Indagine su comportamenti, atteggiamenti e progetti futuri di ragazzi e ragazze tra 11 e 19 anni finalizzata a valutare l'efficacia del disegno del questionario e identificare margini di miglioramento nella qualità del dato (Barcherini *et al.* 2024c). Le informazioni sul numero di accessi al questionario, sul dispositivo utilizzato, sul punto di interruzione nella compilazione e sulle risposte parziali, hanno consentito di tracciare *pattern* comportamentali specifici dei giovani rispondenti. Dall'analisi è emerso che la quasi totalità degli utenti che accedevano al questionario (90 per cento) lo completavano, e nella maggior parte dei casi ciò avveniva in un'unica sessione. Tuttavia, le interruzioni nella compilazione si concentravano prevalentemente nelle prime sezioni, suggerendo che la rinuncia fosse più riconducibile a un deficit di *engagement* iniziale che a effettive difficoltà cognitive o alla lunghezza del questionario. Inoltre, il tasso di accessi multipli, significativamente più alto tra stranieri e giovanissimi, ha evidenziato possibili ostacoli legati all'usabilità e alla comprensione, suggerendo l'esigenza di interventi mirati su tali target.

Particolarmente rilevante è stata l'analisi delle mancate risposte parziali. Nello specifico, percentuali di mancata risposta basse nelle sezioni facoltative del questionario, hanno confermato la buona usabilità dello strumento. Al tempo stesso, scostamenti significativi in presenza di domande sensibili o strutture complesse (come le domande a matrice), specie tra gli stranieri, si sono rivelati input preziosi per miglioramenti futuri.

La sezione conclusiva del questionario, finalizzata a individuare eventuali difficoltà nella compilazione e a raccogliere suggerimenti e spunti per gli sviluppi futuri, ha evidenziato un elevato livello di apprezzamento da parte dei rispondenti, con oltre il 64 per cento che ha dichiarato di non volere apportare modifiche al questionario. Tuttavia, sono stati segnalati alcuni ambiti di miglioramento, tra cui la lunghezza dello strumento e gli aspetti relativi alla fruibilità e alla gradevolezza (*layout*).

In sintesi, l'integrazione dei paradata all'interno del processo valutativo ha dimostrato

un forte potenziale diagnostico, fornendo evidenze oggettive e granulari sui comportamenti di compilazione, che si rivelano essenziali per il perfezionamento continuo degli strumenti di rilevazione rivolti a popolazioni difficili da raggiungere come i giovani.

5.2.8 Controlli di qualità come supporto alla compilazione

Nel contesto delle indagini statistiche che utilizzano questionari *computer-assisted* (CAWI, CATI, CAPI), i controlli di coerenza e di completezza sono da sempre utilizzati come strumenti per garantire l'accuratezza e l'affidabilità dei dati raccolti. Gli errori vengono identificati nel momento stesso in cui vengono commessi e i rispondenti sono responsabili della loro correzione, diventando così artefici della qualità del dato. Questo può essere oneroso e fastidioso, soprattutto nel caso dell'autocompilazione, quando il contributo dell'intervistatore non è disponibile.

L'evoluzione delle pratiche di progettazione dei questionari avvenuta in Istat nel corso dell'ultimo decennio ha progressivamente orientato l'utilizzo dei *quality check* non solo come strumenti finalizzati alla correzione degli errori, ma anche – e soprattutto – come metodo di prevenzione degli stessi, in un'ottica di supporto alla compilazione.

Il cambiamento di paradigma si fonda sull'idea che la qualità dei dati migliori se il rispondente viene messo nelle condizioni di svolgere il compito con maggiore consapevolezza e facilità. Aiutare i rispondenti a fornire dati corretti al primo inserimento evita che questi vengano bloccati dai controlli di qualità, che vengano loro richieste azioni aggiuntive e che, di conseguenza, vengano gravati da un ulteriore *burden*. Come conseguenza, si riduce il rischio di abbandono definitivo del questionario (che è sempre presente quando la compilazione viene interrotta dalla richiesta di correggere un errore) e migliora l'efficienza complessiva del processo di rilevazione.

Ciò è particolarmente rilevante nei questionari ad alta complessità informativa, come quelli che prevedono l'inserimento ripetuto degli stessi dati quantitativi, sia in forma aggregata sia ripartiti secondo varie classificazioni, con la richiesta al rispondente di mantenere la coerenza nel corso di tutto il questionario. In questi casi, il rischio di commettere errori aumenta notevolmente e la richiesta di sanarli può risultare onerosa e frustrante. Al contrario, fornire indicatori in tempo reale – come somme parziali, totali calcolati automaticamente, richiami visivi dei dati inseriti precedentemente – permette al rispondente di monitorare la correttezza delle risposte e ridurre il rischio di errore già in fase di inserimento dei dati.

Un esempio significativo è rappresentato dal questionario del Censimento generale dell'agricoltura 2021, nel quale i rispondenti sono chiamati a dichiarare l'estensione delle superfici aziendali secondo tre criteri distinti: la tassonomia di utilizzo del suolo, articolata in oltre 90 categorie, il titolo di possesso e il metodo di irrigazione. In tale contesto, l'introduzione di funzionalità di calcolo automatico dei totali, la predisposizione di schermate riepilogative dei dati forniti e l'inserimento di domande di riconciliazione agevolano la compilazione, riducendo il numero di errori, limitando il *burden* e migliorando la qualità dei dati rilevati.

Un ulteriore esempio proviene dalla Rilevazione statistica sulla ricerca e sviluppo nelle imprese, nella quale è stato introdotto un sistema di validazione dei dati economici che richiede al rispondente di confermare i totali calcolati automaticamente al termine della compilazione della prima tabella, prima di procedere con le successive (Figura 5.2). I totali, presentati in euro anche quando l'inserimento richiesto è in migliaia di euro, hanno facilitato il controllo visivo da parte del compilatore, contribuendo a ridurre gli errori legati all'errato utilizzo dell'unità di misura, che rappresentava una delle principali criticità delle edizioni precedenti (Barcherini *et al.* 2024b).

L'introduzione di questi accorgimenti nel 2023 ha comportato una riduzione degli errori dal 13,5 per cento nell'edizione del 2022 al 9,2 nell'annualità successiva, fino a raggiungere il 3,1 per cento dopo tre anni di rilevazione.

Figura 5.2 - Esempio di quesito di validazione

B1.1 - Nel 2021 qual è stata la spesa in migliaia di Euro sostenuta dall'impresa per attività di R&S *INTRA-MUROS* per le seguenti voci economiche?

Indicare zero per le spese non sostenute.
Indicare la spesa arrotondata alla cifra intera (esempio: per 135.543 euro inserire 136; per 135.473 euro inserire 135)

	SPESA 2021 in migliaia di Euro
VOCE ECONOMICA	
Spese per personale interno impegnato in R&S	
1. Ricercatori	50
2. Tecnici	75
3. Altro personale	0
Spese per il personale esterno	
4. Consulenti	0
10. Software	0
11. Diritti di brevetto industriale e diritti di sfruttamento di opere dell'ingegno	0
TOTALE	125

Il totale delle spese sostenute dall'impresa nel 2021 per le attività di R&S *INTRA-MUROS* risulta pari a **125.000** Euro. Lo conferma?

☐ Confermo

☐ Non confermo

Fonte: Rilevazione statistica sulla ricerca e sviluppo nelle imprese. Anno 2023

Queste esperienze confermano che i *quality check*, se progettati in coerenza sia con le esigenze informative dell'indagine sia con quelle dei rispondenti, possono costituire uno strumento efficace di miglioramento della qualità statistica, oltre che di contenimento del *burden*. Per assolvere a tale funzione, è necessario che siano calibrati con attenzione rispetto alla complessità del questionario, al grado di familiarità del target con il contenuto informativo e alla modalità di rilevazione impiegata. In tale modo, i controlli non si configurano più come meri vincoli di compilazione, ma come dispositivi attivi di facilitazione del processo di risposta, in grado di valorizzare il ruolo del rispondente come coproduttore dell'informazione statistica.

5.2.9 Servizio di assistenza ai rispondenti tramite *Contact Centre*

L'assistenza alle unità statistiche coinvolte nelle rilevazioni dell'Istat è essenziale per garantire la qualità e l'efficacia della raccolta dati. Questo ruolo risulta particolarmente rilevante nelle rilevazioni basate su tecniche autocompilate CAWI, in cui l'assenza dell'intervistatore rende indispensabile un supporto qualificato ai rispondenti. In senso più generale, il servizio di assistenza favorisce la partecipazione alla raccolta dati, offrendo rassicurazioni sull'ufficiatà dell'indagine e sulla trasparenza dei diversi attori coinvolti nelle reti di rilevazione.

Il servizio è centralizzato e in Istat si colloca all'interno della Direzione che si occupa della raccolta dei dati. Supporta in maniera trasversale un consistente numero di rilevazioni dell'Istat, circa 80 tra rilevazioni campionarie e censimenti. La strategia è di armonizzare e standardizzare le procedure, generalizzandole il più possibile in riferimento alle diverse tipologie di unità di rilevazione coinvolte, pur mantenendone le specificità. I soggetti a cui è rivolto il servizio sono: individui, famiglie, imprese, istituzioni, aziende agricole, comuni. Il servizio è multicanale e multilingue.

Nel 2024, l'Istat ha avviato una profonda ristrutturazione del servizio, passando da un modello completamente esternalizzato - affidato a operatori economici specializzati in *contact centre inbound* - a un Puc basato su una piattaforma multicanale interna e personalizzata sulle specifiche esigenze dell'Istituto (la piattaforma utilizzata è *Salesforce*). Il passaggio da una soluzione gestita totalmente in *outsourcing* a un sistema interno personalizzato, consente all'Istat di adattare dinamicamente il servizio alle evoluzioni dei processi di raccolta dati e di valorizzare pienamente il patrimonio informativo e organizzativo accumulato nel tempo.

Il nuovo modello di assistenza è stato progettato con un forte orientamento al rispondente, ponendo al centro le sue esigenze di chiarezza, accessibilità e autonomia. Il Puc mette a disposizione dei rispondenti una pluralità di canali di accesso integrati - portale web, numero unico, caselle email dedicate, Pec (posta elettronica certificata), *chatbot* - per soddisfare target caratterizzati da diversi livelli di digitalizzazione ed esigenze.

Dal punto di vista della multicanalità, le principali innovazioni riguardano l'introduzione del canale web e il numero unico di pubblica utilità 1510.

Il portale web - accessibile tramite Spid, Cie o accreditamento con email - consente ai rispondenti di aprire autonomamente richieste di assistenza (ticket) e monitorare in tempo reale lo stato di avanzamento delle lavorazioni. Tutte le richieste riferite a una stessa unità di rilevazione sono aggregate in un'area personale unica, dove è possibile consultare sia le segnalazioni attive che lo storico. Questo approccio migliora l'efficienza operativa e garantisce una comunicazione tracciabile e trasparente tra rispondente e Istituto.

L'introduzione del numero unico 1510, in sostituzione di molteplici numeri verdi dedicati, ha inoltre contribuito alla standardizzazione delle comunicazioni con i rispondenti e ha rafforzato la riconoscibilità dell'Istituto, offrendo un numero di contatto unico e facilmente identificabile.

Per i rispondenti, queste innovazioni si traducono in un vantaggio in termini di accesso semplificato all'assistenza, maggiore autonomia nella gestione delle richieste e maggiore trasparenza nelle interazioni con l'Istituto.

Dal punto di vista organizzativo, il Puc adotta un modello orientato all'efficienza nella risoluzione delle richieste e contestualmente all'ottimizzazione delle risorse. La lavorazione delle richieste si articola su tre livelli di specializzazione progressiva: operatore AI (livello 1), operatore esterno (livello 2) ed esperto Istat (livello 3). L'instradamento delle richieste attraverso i tre livelli successivi dipende dalla complessità e dalla specificità dei contenuti delle stesse.

Al Livello 1, l'operatore AI, basato su tecnologie di intelligenza artificiale, gestisce in autonomia le richieste altamente standardizzate, utilizzando contenuti strutturati all'interno della *knowledge base*, come FAQ (*Frequently Asked Questions*), manuali, guide operative e altra documentazione relativa alle indagini. L'introduzione dell'operatore AI, di recente attivazione, ha l'obiettivo di intercettare e risolvere in autonomia una quota significativa di richieste ricorrenti, generando nel tempo un effetto di *case deflection*, ovvero una riduzione progressiva del carico sui livelli successivi.

Il livello 2 prevede il coinvolgimento di operatori di *contact centre* esterni, selezionati tramite procedure pubbliche, che intervengono nei casi in cui l'operatore AI non sia in grado di fornire risposte esaustive. Questi operatori sono abilitati all'uso di specifiche funzionalità dei sistemi gestionali dell'Istat, ad esempio il portale *Statistica & Imprese* e il Sistema di gestione dell'indagine (SGI), e operano principalmente tramite il supporto delle FAQ. La loro attività è sostenuta da un programma di formazione continua, che assicura un aggiornamento costante sulle indagini in corso e sulle procedure operative, contribuendo al mantenimento di elevati standard qualitativi.

Il livello 3 è costituito da esperti dell'Istat, referenti per la raccolta dati e la produzione statistica, ai quali vengono indirizzate solo le richieste più complesse, provenienti dal livello 2, che richiedono conoscenze specialistiche.

Per i processi operativi che sottendono la gestione delle richieste di assistenza, molti sforzi sono stati condotti in termini di automatizzazione dei flussi di lavorazione. Le segnalazioni vengono preventivamente classificate all'ingresso nel sistema, grazie a strumenti come l'IVR (*Interactive Voice Response*), e instradate agli operatori con le competenze più adeguate. Anche la priorità di lavorazione delle richieste viene definita con regole automatiche, considerando criteri come la data di scadenza dell'indagine o l'eventuale sanzionabilità dell'unità di rilevazione.

Grazie all'impiego di soluzioni di intelligenza artificiale e alle regole automatiche di gestione e instradamento delle richieste, il servizio centralizzato di assistenza, nella sua attuale configurazione, è in grado di garantire risposte tempestive e pertinenti, migliorando significativamente la qualità del servizio offerto ai rispondenti.

Si sottolinea che l'elemento fondamentale per garantire un servizio di assistenza di qualità ai rispondenti è la disponibilità di una *knowledge base* completa, standardizzata e costantemente aggiornata, a supporto degli operatori. Il Puc supporta la raccolta dati di numerose rilevazioni, caratterizzate da contenuti eterogenei e da sistemi di acquisizione con funzionalità diverse. In questo contesto è essenziale che la produzione della documentazione di supporto, in particolare delle FAQ, sia esaustiva e in grado di recepire con tempestività i cambiamenti delle fasi di raccolta dati, che spesso non sono prevedibili in anticipo.

Numerosi sforzi sono stati condotti per superare le criticità della passata gestione delle FAQ che, a causa di un'eccessiva frammentazione, produceva disomogeneità dei contenuti, presenza di ridondanze e ritardi degli aggiornamenti. Con il nuovo Puc, tali criticità sono state superate grazie all'internalizzazione della piattaforma che consente una gestione diretta e tempestiva degli aggiornamenti attraverso un coordinamento centrale tra i vari referenti dei processi statistici.

I criteri alla base della gestione centrale delle FAQ sono:

- l'identificazione di referenti specifici, responsabili del coordinamento e del monitoraggio del ciclo di vita delle FAQ (creazione, aggiornamento, archiviazione) relative a una singola indagine o a un gruppo di indagini (ad esempio, congiunturali);
- la predisposizione di tracciati standardizzati, volti a uniformare la struttura e la categorizzazione dei contenuti, indipendentemente dall'indagine di riferimento.

I referenti individuati sono incaricati della gestione delle FAQ afferenti alle rilevazioni di propria competenza, fungendo da raccordo tra i contenuti tematici, prodotti dagli esperti di produzione, e quelli non tematici, a cura degli esperti di raccolta dati. Il processo di caricamento delle FAQ nel sistema avviene in autonomia, accedendo direttamente alla piattaforma.

L'assegnazione di centri di responsabilità dedicati e l'adozione di tracciati strutturati assicurano che le informazioni fornite ai rispondenti siano coerenti, tempestivamente ag-

giornate e prive di ridondanze informative. La standardizzazione, inoltre, facilita il lavoro dei referenti, promuovendo l'uso di linee guida condivise e strumenti uniformi.

5.2.10 Servizio di promemoria telefonici tramite Contact Centre

Un ulteriore elemento su cui il settore della raccolta dei dati dell'Istat ha puntato al fine di migliorare la comunicazione con i rispondenti è la standardizzazione dei promemoria, ovvero la comunicazione rivolta alle unità di indagine, finalizzata a richiamare l'attenzione sulla partecipazione all'indagine, a fornire indicazioni utili riguardo alle modalità di raccolta dei dati, ai tempi entro cui rispondere e agli eventuali obblighi normativi o procedurali connessi alla rilevazione.

La scelta del canale più adeguato all'invio dei promemoria si basa su criteri qualitativi che tengono conto sia delle caratteristiche dell'unità di rilevazione sia dell'obiettivo comunicativo. Per unità istituzionali o imprese si utilizza principalmente la Pec, per famiglie e individui si predilige la posta tradizionale. Poiché i promemoria sono solleciti informali, possono essere utilizzati a integrazione anche canali più snelli e immediati, come l'email o telefonate di cortesia (campagne *outbound*), a seconda della disponibilità di indirizzi elettronici e numeri di telefono aggiornati.

Le campagne *outbound* vengono effettuate dal numero unico 1510, che consente di rendere riconoscibile l'Istat e favorisce una migliore collaborazione con i rispondenti. Oltre a essere impiegate per i promemoria standard, tali campagne sono utilizzate anche per attività di supporto alla raccolta dati per indagini che prevedono la collaborazione di organi intermedi, nonché per le fasi preliminari di avvio di alcuni censimenti. Per questionari brevi e indagini specifiche la chiamata in *outbound* è inoltre offerta come supporto diretto alla compilazione.

Tale approccio differenziato consente di adattare il livello di comunicazione alle esigenze dell'unità di rilevazione e dell'indagine.

Per ciascuna tipologia è stata sviluppata una scheda contatto standardizzata, contenente uno specifico *script* per ciascuna determinata attività e tipologia di unità di rilevazione. La scheda contatto definisce lo schema di dialogo che l'operatore segue durante il contatto con il rispondente e presenta le istruzioni per la gestione della chiamata. Le schede contatto sono uno strumento chiave per assicurare che ogni comunicazione sia chiara, precisa e professionale, migliorando così la qualità complessiva dell'interazione con il rispondente.

Un elemento centrale nella distribuzione dei contatti tra gli operatori è la definizione delle priorità, stabilite dai referenti dell'Istat in base alle caratteristiche specifiche dell'indagine e dell'unità coinvolta. Le unità da contattare sono, inoltre, associate agli operatori in funzione delle loro competenze.

I sistemi prevedono anche un'integrazione tra i servizi *inbound* e *outbound*, un aspetto fondamentale per migliorare la qualità percepita dal rispondente. Qualora durante il contatto il rispondente richieda esplicitamente chiarimenti o assistenza, la richiesta viene inoltrata al servizio *Inbound* senza che sia necessario per il rispondente contattare autonomamente un altro servizio, garantendo così un'esperienza più fluida, immediata e soddisfacente.

5.2.11 Misura e azioni per la riduzione dell'onere statistico sulle imprese

Nonostante il costante incremento nel ricorso alle fonti amministrative, le imprese restano ancora coinvolte in numerose rilevazioni, con un grado via via maggiore al crescere delle

dimensioni, sia in termini di numero di addetti sia di fatturato. Tali rilevazioni, infatti, ricorrono alla strategia campionaria nel caso delle piccole e medie imprese, mentre per le grandi il coinvolgimento è generalmente esaustivo, in particolare per quelle con almeno 250 addetti.

La misura dell'onere statistico sulle imprese (D'Urzo *et al.* 2019) e le iniziative per la sua riduzione sono state affrontate a diverso livello in Istat nel corso del tempo. Un ulteriore impulso al complesso delle azioni messe in campo è rappresentato dal processo di centralizzazione della raccolta dati in Istat (cfr. paragrafo 5.2), che ha rinnovato l'attenzione a questa problematica che rimane una componente essenziale per la riduzione del tasso di risposta e, indirettamente, per il miglioramento della qualità dei dati prodotti.

Le componenti del carico statistico sono molteplici e dipendono dalla tecnica di indagine, dalla semplicità nell'accesso e nella navigazione dei sistemi di acquisizione e dei questionari, dai contenuti del questionario stesso (lunghezza, complessità delle informazioni richieste, distanza tra informazione richiesta e quella disponibile nei database dell'impresa, periodicità della rilevazione), nonché dal grado di coinvolgimento nelle rilevazioni previste dal Sistema statistico nazionale.

Nell'ultimo decennio un passaggio chiave di semplificazione dell'accesso ai questionari dell'Istat è stato rappresentato dalla creazione del portale per le imprese *Statistica & Imprese* (Fazio e Mottura 2015), quale punto unico di accesso per tutte le rilevazioni economiche dell'Istat alle quali l'unità di rilevazione è chiamata a partecipare. Una sola chiave di accesso consente pertanto all'incaricato all'interno dell'impresa, il responsabile portale, di verificare tutti gli adempimenti statistici di sua competenza e le relative scadenze, lo stato dell'adempimento, nonché di entrare nei relativi questionari. Tale utente diventa quindi l'interlocutore privilegiato della comunicazione con l'Istat, che è in grado di contattarlo anche nelle successive occasioni di coinvolgimento in nuove rilevazioni.

È importante ricordare che il portale si alimenta con le informazioni anagrafiche del Registro statistico delle imprese attive (Consalvi *et al.* 2025) corrente (solitamente aggiornato al tempo $t-2$) che, a sua volta, è aggiornato costantemente dal portale in relazione alle variazioni delle variabili anagrafiche segnalate dagli utenti. Tale sinergia è possibile grazie alle funzionalità implementate che permettono il costante aggiornamento – previa validazione interna – delle variabili anagrafiche, eliminando tutte le ridondanze già esistenti relative a richieste di aggiornamento reiterate in più rilevazioni, spesso contemporaneamente e a carico delle medesime imprese. Tale attività centralizzata consente, a sua volta, di migliorare la qualità delle liste di indagine successivamente estratte, riducendo l'inclusione di imprese coinvolte in procedure concorsuali irreversibili o già cancellate dal registro delle camere di commercio. Per tale motivo, il portale comporta una riduzione dell'onere di risposta di grande rilevanza.

Recentemente l'Istat è stata impegnata nella messa in esercizio di ulteriori semplificazioni all'accesso ai questionari delle imprese con il rilascio di una nuova versione del portale. In attesa dell'introduzione dello Spid per l'accesso come impresa o utente fisico, sono state migliorate la trasparenza e la *performance* della procedura di *login* (con standardizzazione controllata del nome utente), nonché facilitate le procedure di recupero/modifica della *password*. La possibilità di personalizzare la *password* da parte dell'utente ne consente, infatti, una più facile conservazione e memorizzazione. L'attenzione verso il rispondente è stata alta anche nel processo di semplificazione e razionalizzazione della rinnovata visualizzazione dell'elenco delle rilevazioni in cui l'impresa è stata ed è coinvolta, garantendo maggiore visibilità a quelle in corso, invece che a quelle già concluse, ed evidenziandone il relativo stato di adempimento. L'analisi delle richieste di assistenza effettuate nel tempo ha

infatti evidenziato che la stragrande maggioranza dei contatti al numero gratuito dell'Istat, da parte di imprese coinvolte in rilevazioni, siano da attribuire alle problematiche connesse all'accesso al portale (smarrimento credenziali, richiesta nuova *password*, eccetera) e alle rilevazioni (individuazione degli adempimenti statistici in termini di rilevazioni in corso e scadenze connesse). Si stima pertanto che i miglioramenti adottati ridurranno drasticamente il ricorso al servizio di assistenza dell'Istat, liberando risorse che potranno essere impiegate in attività di assistenza a più alta competenza.

L'attività sulla misura dell'onere statistico procede su aspetti e con approcci diversi. Attualmente l'attività è focalizzata sulla sistematizzazione e sulla accessibilità delle misure, in parte basate sui dati derivati da quesiti ad hoc inseriti nel questionario di indagine, a cui l'utente è invitato ma non obbligato a rispondere, in parte derivate da informazioni presenti sul portale e sui sistemi di acquisizione (esiti di indagine e paradati). Le misure calcolate di volta in volta alimentano un sistema di indicatori che afferiscono ai due temi base dell'onere statistico: quello del carico statistico generato dalla singola indagine e quello generato dal sistema statistico, nel suo complesso, sulla singola impresa.

Premesso che le rilevazioni sulle imprese sono quasi tutte CAWI, il *set* di indicatori sul carico statistico per rilevazione è stato definito in maniera standardizzata, in riferimento ad alcune caratteristiche di indagine: lunghezza del questionario, complessità di navigazione nel *form* online, complessità delle informazioni richieste, periodicità della rilevazione. Gli indicatori vengono popolati grazie ai quesiti posti in una sezione specifica alla fine del questionario.

Le informazioni richieste riguardano il numero di persone coinvolte nel processo di risposta, il tempo impiegato per reperire le informazioni richieste, il tipo di difficoltà incontrata nella compilazione del questionario, con personalizzazione dei quesiti nel caso di indagini strutturali e congiunturali, distinguendo tra queste ultime le mensili dalle trimestrali. Vengono, inoltre, utilizzati alcuni paradati, ossia dati memorizzati dal sistema di acquisizione che si riferiscono all'operatività dell'utente sul questionario e ai salvataggi effettuati, dall'inizio della compilazione fino all'invio finale. In tale attività, la convergenza dei sistemi di acquisizione verso due sole piattaforme ha contribuito a standardizzare il processo di conservazione dei microdati in un unico *repository*. I dati relativi al salvataggio del questionario, dall'inizio alla fine della compilazione, possono essere utilizzati per il calcolo del tempo di compilazione, seppure inteso in senso ampio e inclusivo dei tempi di non operatività dell'utente.

La collocazione ottimale per l'accessibilità degli indicatori per indagine è rappresentata dal Sistema di documentazione della qualità Sidi-SIQual (in futuro Metastat, cfr. paragrafo 2.1), che li conserva per singola edizione. In questo modo gli indicatori standardizzati prodotti, potranno essere consultati da utenti interni ed esterni, e potranno essere utilizzati per confronti tra rilevazioni e nel tempo.

L'analisi degli indicatori, oltre a consentire il confronto tra rilevazioni e a evidenziare le aree di maggiore criticità, può mettere in luce nel tempo gli effetti degli eventuali miglioramenti apportati al processo (quali il cambio di tecnica, la riduzione dei quesiti e l'ottimizzazione del questionario) sulla singola indagine.

L'indicatore sul carico statistico per impresa è invece calcolato facendo riferimento ai dati presenti nel sistema del pervenuto associato al citato Portale Imprese. Il sistema del pervenuto, infatti, costituisce un *repository* delle informazioni sulla presenza delle imprese nelle varie rilevazioni e sulla loro partecipazione alle stesse, intesa come consegna del questionario compilato, informazioni che consentono di calcolare l'onere statistico potenziale ed effettivo. L'indicatore può essere declinato secondo le variabili di classificazione disponibili, quali la dimensione fisica o economica (per classe di addetti o di fatturato), l'attività

economica prevalente svolta e altre caratteristiche di interesse, così da rendere possibili i confronti tra imprese con valori diversi della variabile di classificazione considerata. L'andamento nel tempo consente di verificare se il sistema statistico è stato in grado di ridurre il coinvolgimento delle imprese nelle rilevazioni in virtù delle strategie messe in atto. Una di queste è rappresentata dal coordinamento negativo tra campioni, che è sempre più adottato per le rilevazioni che selezionano la lista delle unità da includere su una base campionaria.

Considerata la rilevanza del ruolo della raccolta dati nel mantenere alta l'attenzione verso l'utente, è fondamentale dare impulso e consolidare le attività intraprese nell'ottica della loro messa a regime. Una corretta valorizzazione delle informazioni raccolte sarà sicuramente utile alla prosecuzione delle attività e alla individuazione delle ulteriori azioni da intraprendere, anche nel senso della riduzione dell'onere statistico.

5.2.12 Relazioni con gli enti titolari dei dati amministrativi acquisiti dall'Istat

L'Istat adotta pienamente il principio 9 del *Codice delle statistiche europee* (Eurostat e SSE 2017), che richiede la riduzione dell'onere sui rispondenti, facendo ampio uso dei dati amministrativi, quando possibile, e promuovendo l'integrazione dei dati al fine di massimizzarne il potere informativo. Tuttavia, se da un lato si garantisce il rispetto di tale principio, l'accesso ai dati amministrativi, sebbene in presenza di adeguate basi normative, è condizionato alla collaborazione con gli enti titolari di tali dati, richiedendo, quindi, una particolare cura nelle relazioni.

L'Istat acquisisce annualmente 490 forniture di dati amministrativi provenienti da 204 archivi, inviati da 60 enti fornitori (dati 2026). Questi dati sono utilizzati circa nel 50 per cento dei lavori dell'Istat inseriti nel Programma statistico nazionale (dati triennio 2020-2022, fonte PSN), rappresentando quindi una componente significativa nell'input dei processi di produzione delle statistiche ufficiali dell'Istituto. Di conseguenza, si rende essenziale un'attenzione particolare alle relazioni con gli enti fornitori al fine di favorire la loro partecipazione attiva alla produzione della statistica ufficiale e di promuovere un atteggiamento collaborativo.

Il coordinamento e la centralizzazione dell'acquisizione dei dati amministrativi costituiscono il principale risultato raggiunto per garantire il rispetto e la tutela dei fornitori. Nel corso degli anni, gli strumenti informatici adottati hanno consentito agli enti di inviare i dati nelle modalità più idonee, sempre nel rispetto della normativa sulla protezione dei dati personali. Dal 2016, il portale Arcam (Drovandi *et al.* 2017) rappresenta il punto di riferimento per i fornitori, che possono effettuare periodicamente l'*upload* delle forniture attese. La pianificazione anticipata delle attività di acquisizione permette di comunicare tempestivamente agli enti le esigenze informative dell'Istituto e di concordare gli aspetti tecnici, quali: l'elenco delle forniture necessarie nell'anno, il contenuto delle forniture in termini di unità, eventi e variabili, nonché le tempistiche di consegna.

Il principio di minimizzazione, nel rispetto dell'articolo 5 del regolamento (UE) 2016/679, guida costantemente la programmazione delle acquisizioni, ottimizzando il numero di forniture e variabili in relazione ai fabbisogni informativi effettivi dell'Istituto. Tale approccio mira a ridurre il carico di lavoro sugli enti, facilitando il loro adempimento. Il servizio incaricato dell'acquisizione mantiene, inoltre, un contatto continuo con i referenti degli enti, offrendo supporto e chiarimenti in ogni fase del processo. Quando possibile, si stipulano accordi, convenzioni e gruppi di lavoro che prevedono scambi di dati e collaborazioni, rafforzando l'impegno degli enti coinvolti.

Il calcolo sistematico dell'Indicatore di rilevanza, per ogni archivio acquisito a scopo statistico, permette di monitorare l'uso dei dati amministrativi nei processi. In particolare, i metodi di misura pubblicati sulla QRCA (cfr. paragrafo 2.3; Di Bella 2021) si declinano in due ambiti:

- rilevanza di un archivio amministrativo per la produzione delle statistiche ufficiali: numero ed elenco dei lavori Istat del Programma statistico nazionale che dichiarano di utilizzare l'archivio per anno (in serie storica);
- rilevanza di un archivio amministrativo per gli adempimenti della normativa europea: numero ed elenco delle normative il cui adempimento dipende dall'uso dell'archivio per anno (in serie storica).

Queste misure forniscono importanti informazioni sul grado di dipendenza dai dati di ciascun archivio e sottolineano la necessità di curare con particolare attenzione le relazioni con gli enti titolari degli archivi per cui l'indicatore assume valori elevati, al fine di garantire la continuità di acquisizione dei dati e la qualità attesa.

Per monitorare il carico di lavoro sugli enti fornitori, vengono calcolati i seguenti indicatori: numero di archivi e di forniture richiesti a ciascun ente per anno di programmazione delle acquisizioni. Nella Figura 5.3 si mostra il *report* della QRCA in cui sono presenti, a titolo esemplificativo, gli indicatori dell'Agenzia delle entrate, per l'Anno 2024 da cui emerge che l'ente fornisce annualmente all'Istat 29 archivi (ovvero predefinite strutture di dati) per un totale di 89 forniture, che comprendono gli invii anche trimestrali o semestrali dei dati. Nella parte destra della Figura è inoltre riportato un esempio dello stato del monitoraggio dell'acquisizione e del trattamento delle varie forniture nel corso dell'anno (per maggiori informazioni cfr. paragrafo 2.3).

Figura 5.3 - Esempi di indicatori del carico statistico sugli enti fornitori di dati amministrativi: numero di archivi e forniture da acquisire e monitoraggio delle forniture per stato di acquisizione per l'Agenzia delle entrate. Anno 2024

Enti fornitori		Stato Monitoraggio	Forniture	Percentuale forniture
Enti fornitori	1	In fase di richiesta all'ente	1	1,12
Archivi	29	Richiesta all'ente, in attesa di acquisizione	13	14,61
Forniture	89	Acquisita, da trattare in SIM	23	25,84
		Acquisita, trattata in SIM con prima assegnazione del Codice SIM Unità giuridica e con assegnazione codice SIM Individuo, in attesa di assegnazione del Codice SIM Unità economica	6	6,74
		Acquisita, trattata in SIM con prima assegnazione del codice SIM Unità giuridica, in attesa di assegnazione dei codici SIM	4	4,49
		Disponibile	36	40,45
		Disponibile con codici SIM	6	6,74
		Total	89	100,00

Fonte: Istat

5.3 L'investimento nella cultura della qualità

In rispondenza ai principi fondamentali del *quality management*, quali il miglioramento continuo e la diffusione della consapevolezza sulla qualità, una parte delle attività correnti legate alla qualità è dedicata alla formazione, sia interna all'Istituto sia rivolta all'esterno, in particolare al Sistema statistico nazionale e al Sistema statistico europeo.

In questo paragrafo si presentano le principali iniziative messe in atto nell'ambito della formazione. Verranno illustrate le attività correnti chiarendone gli obiettivi e i destinatari, in quanto possibile riferimento nella pianificazione delle attività future.

5.3.1 Iniziative di formazione sulla qualità dedicate al personale Istat

L'importanza del coinvolgimento del personale nel processo di miglioramento della qualità della produzione statistica è uno dei pilastri della politica della qualità dell'Istat, oltre che dei fondamenti della gestione della qualità. I corsi di formazione dedicati al personale dell'Istat sui temi della qualità sono molteplici e svolti con cadenza regolare, a partire dal corso "Introduzione alla qualità nella statistica ufficiale", che illustra i fondamenti della qualità ed è rivolto a tutto il personale dell'Istituto, statistico e non.

Tale corso fornisce i concetti fondamentali della qualità, inquadrandoli nel Quadro comune della qualità del SSE. Può essere considerato propedeutico ad attività formative più avanzate e specialistiche, pur non essendo requisito necessario per l'accesso alle altre opportunità formative offerte. Il corso introduttivo sulla qualità è generalmente offerto con cadenza annuale. Alcuni corsi sono invece organizzati in occasione di specifiche esigenze formative o per determinati gruppi di destinatari. Ad esempio, nel recente passato sono state realizzate alcune edizioni del corso "Uso integrato di dati statistici e amministrativi per gli indicatori territoriali: il percorso della qualità", rivolto al personale degli uffici territoriali dell'Istat. Con minore regolarità sono stati inoltre organizzati corsi avanzati sulla qualità e sul *quality reporting*, destinati al personale coinvolto in tali attività e caratterizzati da un livello di approfondimento orientato a un pubblico con competenze statistiche.

5.3.2 Focus sulle attività formative verso il Sistan

L'Istat è chiamato a progettare verso il Sistan un piano di intervento ampio e articolato, complesso dal punto di vista tecnico-organizzativo e impegnativo sul piano della mobilitazione delle risorse umane. Si tratta, tuttavia, di un investimento strategico per rilanciare il Sistan e valorizzare appieno il contributo di ciascun soggetto appartenente alla rete: una statistica ufficiale di qualità non può prescindere dalla presenza di professionalità solide, dotate di competenze tecniche costantemente aggiornate. In un ecosistema statistico in rapida trasformazione, caratterizzato da un'evoluzione continua delle tecnologie, delle fonti e delle aspettative degli utenti, diventa imprescindibile disporre di operatori opportunamente formati sul piano tecnico-metodologico, capaci di applicare con rigore i principi e gli standard, ma anche di leggere in modo critico e consapevole il contesto informativo, individuare tempestivamente i fabbisogni emergenti, selezionare le soluzioni più appropriate e interagire con i diversi *stakeholder* con trasparenza, competenza e senso di responsabilità.

Diventa quindi essenziale promuovere cicli regolari e strutturati di formazione, progettati secondo un approccio modulare e progressivo, fondato sull'analisi sistematica dei fabbisogni formativi, e coerenti con le priorità individuate dal *Codice italiano per la qualità delle statistiche ufficiali*. Particolare attenzione dovrà essere riservata a temi strategici quali le metodologie statistiche, la gestione e classificazione dei dati, l'ordinamento della statistica ufficiale, la protezione dei dati personali, la riservatezza statistica, nonché gli aspetti connessi alla governance della qualità e all'integrazione tra le fonti.

L'obiettivo è costruire una base solida di competenze condivise, capaci di generare impatti concreti sul piano operativo e di consolidare la qualità dell'intero sistema.

Investire in modo strutturato nella formazione rappresenta un progetto ambizioso, che richiede un impegno continuativo, ma che appare indispensabile per guidare il Sistan attraverso le trasformazioni in atto. Questo approccio deve necessariamente essere accompagnato da percorsi formativi idonei e calibrati, in grado di integrare proposte di natura generalista, utili a costruire un linguaggio comune e a diffondere una visione sistemica della statistica ufficiale, con moduli specialistici orientati al rafforzamento di competenze tecniche avanzate. L'elemento qualificante di questa offerta formativa dovrà essere la sua capacità di adattarsi in modo dinamico e puntuale ai fabbisogni concreti. Tali finalità si inseriscono pienamente nel quadro delle raccomandazioni, e conseguenti azioni di miglioramento, emerse a seguito dell'ultimo *round* di *Peer review* (cfr. paragrafo 1.4), che sollecitano l'Istat a consolidare il proprio ruolo di coordinamento nazionale, anche mediante l'organizzazione di sessioni di *training* dedicate all'implementazione degli standard e degli strumenti per la qualità sviluppati per il Sistan.

In tale contesto, l'Istat è stato esplicitamente invitato a intensificare il proprio impegno strategico nel promuovere standard metodologici condivisi, rafforzare i servizi di supporto tecnico agli enti, migliorare la gestione integrata dei metadati e definire un calendario di rilascio condiviso delle statistiche ufficiali. L'obiettivo prioritario è garantire coerenza, tempestività e qualità nella produzione statistica, in risposta alle crescenti esigenze conoscitive della collettività e alle istanze di trasparenza e *accountability* delle istituzioni pubbliche.

Il conseguimento di questi obiettivi richiede appunto la predisposizione di un programma di formazione dedicato, in cui l'Istat assuma pienamente il proprio ruolo di guida e sostegno. Allo stesso tempo, è fondamentale che gli altri enti del Sistan partecipino con spirito collaborativo, nel rispetto delle proprie attribuzioni e missioni istituzionali. Il successo dell'iniziativa dipenderà sia dalla capacità dell'Istat di garantire un'offerta formativa solida e qualificata, sia dalla disponibilità degli enti e degli uffici di statistica a contribuire attivamente, condividendo esperienze e investendo nel potenziamento delle proprie strutture statistiche.

Risulta tuttavia irrealistico ipotizzare un coinvolgimento generalizzato dell'intera platea di oltre tremila soggetti appartenenti al Sistan. È pertanto necessario, nella progettazione *in fieri*, circoscrivere l'ambito dei destinatari del programma di formazione e di rafforzamento della qualità. Una proposta efficace, almeno in una fase iniziale di avvio e consolidamento, dovrebbe rivolgersi ai soli soggetti produttori di statistica ufficiale, in particolare ai titolari di lavori PSN e ai soggetti inclusi nell'elenco delle ONA, in coerenza con le indicazioni della *Peer review* riguardante l'individuazione del *new core* del Sistan, vale a dire il sottoinsieme dei principali contributori alla produzione e diffusione di statistiche ufficiali. L'obiettivo è consolidare un nucleo del Sistan caratterizzato da autorevolezza, trasparenza e coerenza con gli standard europei e internazionali.

Nell'ottica di rafforzare il ruolo dell'Istat quale ente di coordinamento del Sistan, appare strategico investire anche sulla formazione interna all'Istat, con particolare riferimento alla conoscenza del contesto istituzionale e organizzativo in cui l'Istituto opera. In linea con le raccomandazioni emerse nella più recente *Peer review*, andrebbe potenziata l'offerta formativa rivolta al personale neoassunto, attraverso moduli specifici dedicati al funzionamento del Sistan, al PSN e alla governance complessiva della statistica ufficiale in Italia. Offrire una visione sistemica e integrata del Sistan, fin dai primi anni di carriera, significa

valorizzare le professionalità interne, rafforzare la cultura del coordinamento e creare una consapevolezza diffusa del ruolo che ciascuna componente dell'Istituto svolge all'interno di una rete più ampia e articolata.

5.3.3 Iniziative di formazione sulla qualità in ambito internazionale

Nel contesto della qualità, l'Istat ha assunto un ruolo di *leadership* riconosciuto a livello europeo, al punto che i due corsi sulla qualità di livello avanzato all'interno dell'*European Statistical Training Programme* (ESTP) sono da diversi anni tenuti da docenti dell'Istat: *Quality Framework, Process and Product Quality Measurement e Advanced Course on Quality Reporting*. Il primo corso ha un orientamento fortemente statistico e, oltre a presentare il contesto del Quadro comune europeo, illustra in dettaglio le tecniche di misurazione e correzione dei principali errori non campionari, come l'errore di mancata risposta e di misurazione. Il corso sulla reportistica di qualità, invece, si concentra sugli standard europei per la documentazione dei metadati, presentando in dettaglio i concetti previsti nel *template* SIMS (cfr. paragrafo 4.3). Entrambi i corsi menzionati prevedono lavori di gruppo ed esercitazioni pratiche come momento centrale di apprendimento e di coinvolgimento dei partecipanti.

Legati all'ambito delle attività del SSE sono i corsi per l'*European Master in Official Statistics* (EMOS), un curriculum universitario rivolto alla formazione di statistici interessati a lavorare nell'ambito della statistica ufficiale. All'interno di questo programma viene svolto annualmente, presso Sapienza Università di Roma, il corso *Data Quality and Other Topics of Official Statistics*, previsto all'interno del programma di laurea magistrale *Statistical Methods and Applications* della facoltà di Ingegneria dell'informazione, informatica e statistica.

Il contributo dell'Istat allo sviluppo delle competenze in materia di qualità non si esaurisce nell'ambito europeo. Tra le attività internazionali dell'Istat rientrano anche iniziative di cooperazione internazionale, quali missioni di assistenza tecnica, *study visit* e corsi di formazione organizzati ad hoc per il rafforzamento dei sistemi statistici dei paesi beneficiari, spesso nell'ambito di più ampi progetti di cooperazione tra l'Italia e tali paesi.

RIFERIMENTI BIBLIOGRAFICI

- Amaya, A., P. P. Biemer, and D. Kinyon. 2020. "Total Error in a Big Data World: Adapting the TSE Framework to Big Data". *Journal of Survey Statistics and Methodology*, Volume 8, N. 1: 89-119. <https://academic.oup.com/jssam/issue/8/1>.
- Ascari, G., G. Brancato, A. Bruni, S. Daddi, G. Di Bella, S. Giavante, F. Rocci, R. Sanzo, A. M. Sgamba, G. Simeoni, e S. Spirito. 2023. "Il monitoraggio e la valutazione della qualità del Sistema Integrato dei Registri". *Istat Working Papers*, N. 7/2023. <https://www.istat.it/produzione-editoriale/il-monitoraggio-e-la-valutazione-della-qualita-del-sistema-integrato-dei-registri/>.
- Australian Bureau of Statistics - ABS. 2016. *ABS Data Quality Manual. Part B – Quality Gates in the Statistical Process*. Canberra, Australia: ABS. https://unstats.un.org/unsd/methodology/dataquality/references/Australia_Part_B_Quality_Gates_in_the_statistical_process2.pdf.
- Bali, N., e D. Cangialosi. 2024. *L'approccio circolare alla formazione delle reti di rilevazione del primo ciclo del Censimento permanente della popolazione e delle abitazioni 2018-2021*. Letture Statistiche - Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/lapproccio-circolare-alla-formazione-delle-reti-di-rilevazione-del-primo-ciclo-del-censimento-permanente-della-popolazione-e-delle-abitazioni/>.
- Bali, N., S. Barcherini, G. Fazzi, D. Grassi, B. Lorè, e S. Macchia. 2019. "Metodologia integrata per l'ottimizzazione del disegno dei questionari di indagine". *Istat Working Papers*, N. 16/2019. <https://www.istat.it/produzione-editoriale/metodologia-integrata-per-lottimizzazione-del-disegno-dei-questionari-di-indagine/>.
- Barcherini, S., K. Bontempi, M. Bussola, B. Lorè, and S. Rosati. 2024a. "Evaluating computer-assisted questionnaire usability: the case of permanent census of the population and housing". *Rivista Italiana di Economia Demografia e Statistica*, Volume LXXVIII: 125-136. <https://doi.org/10.71014/sieds.v78i1.227>.
- Barcherini, S., B. Lorè, V. Mastrostefano, and S. Rosati. 2024b. "Differential benefits of questionnaire redesign: implications for data quality and statistical burden across different respondent profiles". *Rivista Italiana di Economia Demografia e Statistica*, Volume LXXVIII: 267-278. <https://doi.org/10.71014/sieds.v78i3.343>.
- Barcherini, S., S. Liani, S. Pecora and S. Rosati. 2024c. "Response behavior patterns among youth". *Rivista Italiana di Economia Demografia e Statistica*, Volume LXXVIII: 171-182. <https://doi.org/10.71014/sieds.v78i3.356>.
- Barcherini, S., K. Bontempi, S. Liani, B. Lorè, and S. Rosati. 2023. "Online cognitive interviewing, a contribution to data quality". *Rivista Italiana di Economia Demografia e Statistica*, Volume LXXVI: 173-182. <https://www.rieds-journal.org/rieds/article/view/113>.
- Barcherini, S., K. Bontempi, G. Fazzi, S. Liani, B. Lorè, S. Pietropaoli, and S. Rosati. 2022. *The respondent as the focus of the questionnaire design*. Paper presented at *UNECE Expert Meeting on Statistical Data Collection "Towards a New Normal?"*. Roma, 26-28 October 2022. <https://unece.org/statistics/events/DC2022>.
- Brancato, G. (a cura di). 2016. *Linee guida per la qualità dei processi statistici che utilizzano dati amministrativi. Versione 1.1*. Roma, Italia: Istat. <https://www.istat.it/classificazioni-e-strumenti/strumenti-per-la-qualita/linee-guida/>.
- Brancato, G., and G. Simeoni. 2012. *Istat Statistical Process Modelling and the Generic Statistical Business Process Model: a Comparison*. Paper presented at *European Conference on Quality in Official Statistics (Q2012)*. Athens, 29 May - 1 June 2012.

- Brancato, G., C. Pellegrini, G. Simeoni, and M. Signore. 2004. "Standardising, Evaluating and Documenting quality: the implementation of Istat information system for survey documentation – SIDI". *Proceedings of the European Conference on Quality and Methodology in Official Statistics* (Q2004). Mainz, 24-26 May 2004.
- Brancato, G., R. Carbinì, C. Pellegrini, M. Signore, and G. Simeoni. 2006. *Assessing Quality Through the Collection and Analysis of Standard Quality Indicators: the ISTAT Experience*. Paper presentato alla *European Conference on Quality in Survey Statistics* (Q2006). Cardiff, 24-26 April 2006. <https://webarchive.nationalarchives.gov.uk/ukgwa/20120715031240/http://www.ons.gov.uk/ons/media-centre/events/past-events/q2006---european-conference-on-quality-in-survey-statistics-24-26-april-2006/index.html>.
- Bologna, E., B. Loré, S. Macchia, e S. Orsini. 2020. "Un nuovo approccio per i test cognitivi: l'esperienza sull'indagine "Aspetti della vita quotidiana". *Istat Working Papers*, N. 2/2020. <https://www.istat.it/produzione-editoriale/un-nuovo-approccio-per-i-test-cognitivi-l'esperienza-sull'indagine-aspetti-della-vita-quotidiana/>.
- Broccoli, M., e S. Bruzzone. 2021. *Utilizzo di Open Street Map per il calcolo di indicatori per l'incidentalità stradale*. Paper presentato alla LVII Riunione Scientifica della SIEDS. Palermo, 27 Maggio 2021.
- Brunini, C. 2023a. "Il ruolo dei termini nel Catalogo Nazionale della Semantica dei Dati". In Bellitti, G., e M. Fedeli. *Trasformazione digitale della Pubblica amministrazione. Metodi per l'interoperabilità per lo sviluppo di e-service*: 36-37. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/trasformazione-digitale-della-pubblica-amministrazione-metodi-per-l'interoperabilita-per-lo-sviluppo-di-e-service/>.
- Brunini, C. 2023b. "Il termine come unità minima di una raccolta metodologica". In Bellitti, G., e M. Fedeli. *Trasformazione digitale della Pubblica amministrazione. Metodi per l'interoperabilità per lo sviluppo di e-service*: 37-38. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/trasformazione-digitale-della-pubblica-amministrazione-metodi-per-l'interoperabilita-per-lo-sviluppo-di-e-service/>.
- Brunini, C., and G. D'Alessandro. 2025. *The terminology management in official statistics*. Paper presented at the conference *Terminology & Ontology: Theories and applications* (TOTh 2025). Chambéry, 4-5 giugno 2025.
- Brunini, C., and E. Recchini. 2024. *The role of the term in defining centralised standard metadata*. Presentation at the conference *Smart Metadata for Official Statistics* (COSMOS2024), Paris, 11-12 April 2024. http://cosmos-conference.org/2024/slides/32_Brunini.pdf.
- Budano, F., and G. Simeoni. 2024. *Integrating international standards in the design of reference metadata component for the new Istat system METAstat*. Paper presented at the *European Conference on Quality in Official Statistics*. Estoril, 5-7 June 2024.
- Burns, C., J. Thomason, and W. Tansey. 2020. "Interpreting Black Box Models via Hypothesis Testing". In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference (FODS '20)*: 47-57. <https://doi.org/10.1145/3412815.3416889>.
- Cammarota, M. (a cura di). 2024. *Sistema integrato di diffusione delle statistiche dell'Istat: percorso e stato attuale*. *Lecture Statistiche - Metodi*. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/sistema-integrato-di-diffusione-delle-statistiche-dell'istat-percorso-e-stato-attuale/>.
- Casagrande, C., and V. Olivieri. 2024. *The formalization of cross concepts in statistical registers about units and variables: methods and practical applications in Istat*. Presentation at the conference *Smart Metadata for Official Statistics* (COSMOS2024). Paris, 11-12 April 2024. <http://cosmos-conference.org/2024/program.html>.
- Casagrande, C. S. Giavante, F. Rocci, and G. Simeoni. 2024. *Implementing the quality framework for the Istat Integrated System of Statistical Registers: challenges and solutions*. Paper presented at the *European Conference on Quality in Official Statistics*. Estoril, 5-7 June 2024.
- Cerroni, F., G. Di Bella, and L. Galiè. 2014. "Evaluating administrative data quality as input of the statistical production process". *Rivista di statistica ufficiale*, N. 1-2/2014: 117-146. <https://www.istat.it/produzione-editoriale/rivista-di-statistica-ufficiale-1-22014/>.

- Converse, J. M., and S. Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, California, U.S.: Sage Publications.
- Cooper, A. 1999. *The inmates are running the asylum*. Carmel, Indiana, U.S.: Sams Publishing.
- D'Alessandro, G., C. Brunini, A. Di Benedetto, and R.M. Aracri. 2025. *Terminology and Conceptualization*. Paper presented at the conference *Terminology & Ontology: Theories and applications* (TOTh 2025). Chambéry, 4-5 June 2025.
- D'Orazio, M., M. Di Zio, O. Luzi, and M. Scanu. 2025. "The journey towards a governance of open source statistical tools at Istat". *Statistical Journal of the IAOS*, Volume 41, N. 4: 1144-1153. <https://doi.org/10.1177/18747655251390274>.
- D'Urzo, A., A. Golino, S. Macchia, A. Nurra, C. Pascucci, M. Rinaldi, F. Rocci, C. Schiattone, and G. Simeoni. 2019. "A systematic and standardised burden measurement system for surveys on businesses". *Rivista di statistica ufficiale*, N. 1/2019. https://www.istat.it/it/files/2021/03/RSU_1-2019_Article_1.pdf.
- Daas, P.J.H., S.J.L. Ossen, M. Tennekes, L.-C. Zhang, C. Hendriks, K. Foldal Haugen, F. Cerroni, G. Di Bella, T. Laitila, A. Wallgren, and B. Wallgren. 2011. *Reports on methods preferred for the quality indicators of administrative data sources*. BLUE-ETS Workpackage 4, Deliverable 4.2.
- Das, A., and P. Rad. 2020. "Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey". *ArXiv*, Volume 2006, N.11371. <https://doi.org/10.48550/arXiv.2006.11371>.
- Daud, A., and D. Dubhashi. 2023. "Statistical Modeling: The Three Cultures". *Harvard Data Science Review*, N. 4.1, Winter 2023. <https://doi.org/10.1162/99608f92.89f6fe66>.
- De Fausti, F., M. Di Zio, R. Filippini, S. Toti, and D. Zardetto. 2022. "Multilayer perceptron models for the estimation of the attained level of education in the Italian Permanent Census". *Statistical Journal of the IAOS*, Volume 38, N. 2: 637-646. <https://doi.org/10.3233/SJI-210877>.
- Di Bella, G. 2022. *Systematic reporting on the quality of administrative data as input of the statistical production process: the Italian experience*. Paper presented at the *European Conference on Quality in Official Statistics* (Q2022). Vilnius, 8-10 June 2022. <https://q2022.stat.gov.it/scientific-information/papers-presentations/session-24>.
- Di Bella, G. (a cura di). 2021. *Il sistema di documentazione dei dati amministrativi in Istat*. Letture statistiche – Metodi. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/il-sistema-di-documentazione-dei-dati-amministrativi-in-istat/>.
- Di Zio, M., S. Falorsi, F. Rocci, and G. Simeoni. 2023. "Process and output quality evaluation measures for Istat Integrated System of Statistical Registers". Paper presented at the *European Establishment Statistics Workshop - EESW23*, Lisbon, 19-22 September 2023. <https://drive.google.com/file/d/1IjHPMI4NF-oggIEqIPvne3B0RwSh1Cj/view>.
- Drovandi, G., P. Giacomini, M. Giacommo, and E. Sibilio. 2017. "ARCA: reengineering of Admin Data acquisition". Paper presented at the conference *NTTS New techniques and technologies for statistics*. Bruxelles, 13-17 March 2017.
- Ehling, M., and T. Körner (eds.). 2007. *Handbook on Data Quality Assessment Methods and Tools*. Wiesbaden, Germany: Eurostat. <https://www.paris21.org/sites/default/files/dataqualityassessment-eurostat.pdf>.
- European Commission. 2020. *Communication to the Commission. Open Source Software Strategy 2020–2023. Think Open*. Bruxelles, Belgium: European Commission. https://commission.europa.eu/about/departments-and-executive-agencies/digital-services/open-source-software-strategy_en.
- European Statistical System - ESS. 2019. *Quality Assurance Framework of the European Statistical System*. Version 2.0. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V2.0-final.pdf>.
- Eurostat. 2023a. *Peer review report on compliance with the European Statistics Code of Practice and further improvement and development of the National Statistical System. Italy*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/documents/64157/13566711/italy-peer-review-report.pdf>.

- Eurostat. 2023b. *Reusing Mobile Network Operator data for Official Statistics: the case for a common methodological framework for the European Statistical System. 2023 edition*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-statistical-reports/w/ks-ft-23-001>.
- Eurostat. 2021. *European Statistical System handbook for quality and metadata reports. 2021 re-edition*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-gq-21-021>.
- Eurostat. 2020. *Revised Version of the Quality Guidelines for the Acquisition and Usage of BD. ESSnet Big Data II, Deliverable K3*. <https://cros.ec.europa.eu/ESSnetBDII-WPK>.
- Eurostat. 2015a. *Peer review report on compliance with the Code of Practice and the coordination role of the National Statistical Institute. Italy*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/documents/64157/4372828/2015-IT-report.pdf>.
- Eurostat. 2015b. *Single Integrated Metadata Structure v 2.0 (SIMS V2.0) and its underlying reporting structures. The ESS quality and reference metadata reporting standards ESMS 2.0 and ESQRS 2.0*. Luxembourg: Publications Office of the European Union. <https://ec.europa.eu/eurostat/web/metadata/reference-metadata-reporting-standards>.
- Eurostat. 2006. *Peer review of the Italian Statistical Office (ISTAT) on the implementation of the European Statistics Code of Practice*. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/eurostat/documents/64157/21811392/PEER_REVIEW_IT_2006-EN.PDF.pdf.
- Eurostat e Sistema statistico europeo - SSE. 2017. *Codice delle statistiche europee. Per le autorità statistiche nazionali ed Eurostat (autorità statistica dell'UE)*. Lussemburgo: Ufficio delle pubblicazioni dell'Unione europea. <https://ec.europa.eu/eurostat/documents/4031688/9394142/KS-02-18-142-IT-N.pdf>.
- Fazio, N., and G. Mottura. 2015. *The business portal: an integrated system to collect data and communicate with our statistical partners*. Paper presented at the *UNECE Workshop on statistical data collection: Riding the wave of data deluge*. Washington D.C., 29 April - 1 May 2015. <https://unece.org/statistics/events/DC2015>.
- Filippini, R., S. Giavante and G. Rocchetti. 2024. "The framework for quality assessment in the Istat Integrated System of Registers: An application to the estimation of the Attained Level of Education". Paper presented at the *European Conference on Quality in Official Statistics (Q2024)*, Estoril, 5-7 June 2024.
- Fisher, R.P., and R.E. Geiselman. 1992. *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL, U.S.: Charles C. Thomas.
- Frattarola, D., and S. Spirito. 2025. "SIGMA, a New Approach to Statistical Data Processing". In di Bella, E., V. Gioia, C. Lagazio, and S. Zaccarin (eds.). *Statistics for Innovation IV. SIS 2025, Short Papers, Contributed Session III*. Italian Statistical Society Series on Advances in Statistics. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-96033-8_86.
- Gobo, G. 2002. *Le risposte e il loro contesto. Processi cognitivi e comunicativi nelle interviste standardizzate*. Milano, Italia: Franco Angeli.
- Groves, R., F. Fowler, M. Couper, J. Lepkowski, E. Singer, and R. Tourangeau. 2009. *Survey Methodology. Second edition*. New York, NY, U.S.: Wiley.
- International Monetary Fund - IMF. 2012. *Data Quality Assessment Framework (DQAF) – Generic Framework*. Washington, DC, U.S.: IMF, Statistics Department. https://dsbb.imf.org/content/pdfs/dqrs_Genframework.pdf.
- International Organization for Standardization - ISO. 2019. *ISO 1087:2019 Terminology work and terminology science - Vocabulary*. Geneva, Switzerland: ISO.
- International Organization for Standardization - ISO. 2018a. *ISO 9241-11:2018. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts*. Geneva, Switzerland: ISO.
- International Organization for Standardization - ISO. 2018b. *ISO 19011:2018. Guidelines for auditing management systems*. Geneva, Switzerland: ISO.

- International Organization for Standardization - ISO. 2015. ISO 9000:2015, *Quality Management Systems*. Geneva, Switzerland: ISO.
- International Organization for Standardization - ISO. 2013. ISO 25964-2:2013, *Information and documentation - Thesauri and interoperability with other vocabularies - Part 2: Interoperability with other vocabularies*. Geneva, Switzerland: ISO.
- Istituto Nazionale di Statistica - Istat. 2025a. *Struttura e dimensione delle imprese secondo la nuova definizione - Registro Asia 2023*. Nota metodologica allegata alle Tavole di dati Registro statistico delle imprese attive - Anno 2023. Roma, Italia: Istat. <https://www.istat.it/tavole-di-dati/registro-statistico-delle-imprese-attive-anno-2023/>.
- Istituto Nazionale di Statistica - Istat. 2025b. *Utilizzo di Open Street Map per il calcolo di indicatori per l'incidentalità stradale sulla rete viaria italiana. Anno 2023*. Area dedicata del sito web sulla statistica sperimentale. Roma, Italia: Istat. <https://www.istat.it/statistica-sperimentale/utilizzo-di-open-street-map-per-il-calcolo-di-indicatori-per-lincidentalita-stradale-sulla-rete-viaria-italiana-anno-2023/>.
- Istituto Nazionale di Statistica - Istat. 2024. *Social Mood on Economy Index. Una misura del sentiment italiano sull'economia basata sui dati di X (ex Twitter)*. Area dedicata del sito web sulla statistica sperimentale. Roma, Italia: Istat. <https://www.istat.it/statistica-sperimentale/social-mood-on-economy-index/>.
- 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. 2022. *Istat Quality Management System*. Roma, Italia: Istat. <https://www.istat.it/wp-content/uploads/2011/11/Quality-management-system.pdf>.
- Istituto Nazionale di Statistica - Istat. 2021. *La nuova politica per la qualità 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 prodotte dal Sistema statistico nazionale*. Roma, Italia: Istat. <https://www.istat.it/classificazioni-e-strumenti/strumenti-per-la-qualita/strumenti-per-la-qualita-nel-sistan/>.
- Karlberg, M., and B. Gauckler. 2020. *Digital communication, User analytics and Innovative products (DIGICOM)*. Project-End Report. Luxembourg: European Commission. <https://circabc.europa.eu/d/d/versionStore/version2Store/5624ea0e-ad7d-4793-a0c0-8ff0bd722ea8/DIGICOM-FinalReport-VIG-2020-20-2B.pdf>.
- Keane, P. A., G. Zhang, D.J. Fu., B. Liefers, L. Faes, S. Ginton, S. Wagner, R. Struyven, N. Pontikos, and K. Balaskas. 2021. "Clinically relevant deep learning for detection and quantification of geographic atrophy from optical coherence tomography: a model development and external validation study". *The Lancet, Digital Health*, Volume 3, N. 10: e665-e675. [https://doi.org/10.1016/S2589-7500\(21\)00134-5](https://doi.org/10.1016/S2589-7500(21)00134-5).
- Neisser, U. 1967. *Cognitive Psychology*. New York, NY, U.S.: Appleton-Century-Crofts.
- Nielsen, J. 1993. *Usability Engineering*. Cambridge, MA, U.S.: Academic Press.
- Recchini, E. 2024. *Standardised statistical metadata and semantics in the international community*. Paper presented at the conference *Terminology & Ontology: Theories and applications (TOTH 2025)*. Chambéry, 4-5 June 2025.
- Recchini, E. 2023. "Relazioni semantiche tra termini". In Bellitti, G. e M. Fedeli (a cura di). *Trasformazione digitale della pubblica amministrazione - metodi per l'interoperabilità per lo sviluppo di e-service*: 38-42. Roma, Italia: Istat. <https://www.istat.it/produzione-editoriale/trasformazione-digitale-della-pubblica-amministrazione-metodi-per-linteroperabilita-per-lo-sviluppo-di-e-service/>.
- Ribeiro, M.T., S. Singh, and C. Guestrin. 2016. "Why Should I Trust You? Explaining the Predictions of Any Classifier". *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 1135-1144. <https://doi.org/10.1145/2939672.2939778>.

- Rizzolo, F., M. Signore, J. Lynch, A. Tyler, T. Caparelli, F. Kalonji, A. Hathoot, E. Kaukonen, M. Saloila, C. Ábry, E.H. Ender, Z. Vereczkei, C. Brunini, E. Recchini, M. Scanu, G. Simeoni, L. Tosco, C. Vaccari, J. Munoz, I. Ramos, A. Dlugosz, Y. Choi, C. Karling, P. Wahlgren, J. Greenfield, E. Greising, and I. Choi. 2021. *Information flow within GSBPM using GSIM*. Note on UNECE StatsWiki, prepared by the "Linking GSBPM and GSIM" Task Team within the HLG-MOS Supporting Standards Group. Geneva, Switzerland: UNECE. <https://statswiki.unece.org/plugins/servlet/mobile?contentId=330370507#content/view/330370507>.
- Rocci, F., R. Varriale, and O. Luzi. 2022. "Total Process Error: An Approach for Assessing and Monitoring the Quality of Multisource Processes". *Journal of Official Statistics*, Vol. 38, No. 2: 533 - 556. <https://doi.org/10.2478/jos-2022-0025>.
- Runci, M.C., G. Di Bella, e L. Galiè. 2016. "Il sistema di integrazione dei dati amministrativi in Istat". *Istat Working Papers*, N. 18/2016. <https://www.istat.it/produzione-editoriale/il-sistema-di-integrazione-dei-dati-amministrativi-in-istat/>.
- Signore, M., M. Scanu, and G. Brancato. 2015. "Statistical Metadata: a Unified Approach to Management and Dissemination". *Journal of Official Statistics*, Volume 31, N. 2: 325-47. <https://doi.org/10.1515/jos-2015-0020>.
- Signore, M., G. Brancato, R. Carbinì, M. D'Orazio, e G. Simeoni (a cura di). 2012. *Linee guida per la qualità dei processi statistici. Versione 1.1*. Roma, Italia: Istat <https://www.istat.it/classificazioni-e-strumenti/strumenti-per-la-qualita/linee-guida/>.
- Simeoni, G. 2013. *Implementing ESS standards for reference metadata and quality reporting at Istat*. Paper presented at *UNECE Work Session on Statistical Metadata*, Geneva, 6-8 May 2013. <https://unece.org/statistics/events/work-session-statistical-metadata>.
- Sistema statistico nazionale - Sistan, e Istituto nazionale di statistica - Istat. 2022. *Codice italiano per la qualità delle statistiche ufficiali*. Direttiva n. 12/Comstat. *Gazzetta Ufficiale Serie Generale* N. 23 del 29 gennaio 2022.
- Sudman, S., and N.M. Bradburn. 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA, U.S.: Jossey-Bass.
- ten Bosch, O. 2022. *The awesome list of official statistics software & FOSS best practices*. Presentation in the *European Task Force on Trusted Smart Statistics* (TF-TSS). <https://doi.org/10.5281/zenodo.7665189>.
- Tourangeau, R., and K.A. Rasinski. 1988. "Cognitive process underlying Context effect in attitude measurement". *Psychological Bulletin*, Volume 103, N. 3: 299-314. <https://www.scirp.org/reference/referencespapers?referenceid=1135105>.
- UK Statistics Authority, and Office for Statistics Regulation. 2018. *Code of Practice for Statistics. Ensuring official statistics serve the public. Edition 2.0*. London, UK: Office for Statistics Regulation. <https://code.statisticsauthority.gov.uk/wp-content/uploads/2018/02/Code-of-Practice-for-Statistics.pdf>.
- United Nations. 2025a. *Module for Quality Assurance when using Administrative and Other Data Sources to produce Official Statistics*. New York, NY, U.S.: UN. <https://unstats.un.org/unsd/methodology/dataquality/aos/>.
- United Nations. 2025b. *The Handbook on Management and Organization of National Statistical Systems: 4th Edition of the Handbook of Statistical Organizations*. New York, NY, U.S.: UN. <https://unstats.un.org/capacity-development/handbook/index.cshmtl>.
- United Nations. 2022. *Guidelines on the use of electronic data collection technologies in population and housing censuses*. New York, NY, U.S.: UN. <https://unstats.un.org/unsd/demographic/standmeth/handbooks/guideline-edct-census-v1.pdf>.
- United Nations. 2019. *United Nations National Quality Assurance Frameworks. Manual for Official Statistics. Including recommendations, the framework and implementation guidance*. Studies in Methods, Series M, N. 100. New York, NY, U.S.: UN. <https://unstats.un.org/unsd/methodology/dataquality/un-nqaf-manual/>.

- United Nations Economic Commission for Europe - UNECE. 2025. *Generic Statistical Business Process Model (GSBPM). (Version 5.2, May 2025)*. Geneva, Switzerland: UNECE. <https://unece.org/statistics/documents/2025/06/working-documents/generic-statistical-business-process-model-gsbpm>.
- United Nations Economic Commission for Europe - UNECE. 2024. *Generic statistical information model (GSIM): communication paper for a general statistical audience. Version 2.0*. Geneva, Switzerland: UNECE. <https://unece.org/statistics/modernstats/gsim>.
- United Nations Economic Commission for Europe - UNECE. 2021. *Machine Learning for Official Statistics*. Geneva, Switzerland: UNECE. <https://statswiki.unece.org/spaces/ML/overview>.
- United Nations Economic Commission for Europe - UNECE. 2014. *A Suggested Framework for the Quality of Big Data*. Geneva, Switzerland: UNECE. <https://unece.org/statistics/documents/2025/02/quality-framework-big-data-2014>.
- United Nations Economic Commission for Europe - UNECE. 2009. *Common Metadata Framework Part A: Statistical Metadata in a Corporate Context*. Geneva, Switzerland: UNECE. <https://unece.org/info/publications/pub/21877>.
- United Nations Economic Commission for Europe - UNECE. 2007. *Register-based Statistics in the Nordic Countries: Review of Best Practices with Focus on Population and Social Statistics*. New York, NY, U.S., and Geneva, Switzerland: United Nations - UN. <https://digitallibrary.un.org/record/609979?v=pdf>.
- United Nations Statistical Commission - UNSC, and United Nations Economic Commission for Europe - UNECE. 2000. *Terminology on Statistical Metadata*. Geneva, Switzerland: UNECE. <https://unece.org/DAM/stats/publications/53metadaterminology.pdf>.
- Vilone, G., and L. Longo. 2020. "Explainable Artificial Intelligence: A Systematic Review". *ArXiv*, Volume 2006, N.00093. <https://arxiv.org/abs/2006.00093>.
- Wallgren, A., and B. Wallgren. 2014. *Register-based Statistics - Statistical Methods for Administrative Data. Second edition*. New York, N.Y., U.S.: Wiley.
- Wilson, L., and E. Dickinson. 2022. *Respondent Centred Surveys. Stop, Listen and then Design*. London, UK: Sage.
- Yung, W., S.-M. Tam, B. Buelen, H. Chipman, F. Dumpert, G. Ascari, F. Rocci, J. Burger, and I. Choi. 2022. "A quality framework for statistical algorithms". *Statistical Journal of the IAOS*, Volume 38: 291-308. <https://doi.org/10.3233/SJI-210875>.
- Zhang, L.-C. 2012. "Topics of statistical theory for register-based statistics and data integration". *Statistica Neerlandica*, Volume 66, N. 1: 41-63. <https://doi.org/10.1111/j.1467-9574.2011.00508.x>.

