



ROMÂNIA

Acord de Servicii de Asistență Tehnică Rambursabile privind
Consolidarea Sistemului Statistic Național (P167217)

REZULTAT Nr. 6d

**Raport privind serviciile de asistență tehnică oferite
Beneficiarului privind Metodologia pentru o evaluare continuă
și în timp real a calității datelor colectate din surse
administrative și geo-spațiale pentru perioadele intercenzitare**

Mai 2022



Clauză de limitare a responsabilității

Acest raport este un produs al Băncii Internaționale pentru Reconstrucție și Dezvoltare/Banca Mondială. Constatările, interpretările și concluziile exprimate în acest document nu reflectă în mod obligatoriu părerile directorilor executivi ai Băncii Mondiale sau ale guvernelor pe care aceștia le reprezintă. Banca Mondială nu garantează acuratețea datelor incluse în prezentul document.

Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului României.

Drepturile de autor

Materialul inclus în această publicație este protejat de legea drepturilor de autor. Copierea și/sau transmiterea unor secțiuni din această lucrare fără permisiune poate constitui o încălcare a legislației aplicabile.

Pentru a obține permisiunea de a fotocopiază sau retipări orice secțiune din prezenta lucrare, vă rugăm să trimiteți o solicitare conținând informațiile complete către: (i) Institutul Național de Statistică din România (Bd. Libertății nr. 16, Sector 5, București, România) sau (ii) Grupul Băncii Mondiale în România (str. Vasile Lascăr nr. 31, etaj 6, București, România).

Acest raport a fost livrat în luna mai 2022 în temeiul Acordului de Servicii de Asistență Tehnică Rambursabile privind Consolidarea Sistemului Statistic Național (P167217), semnat între Institutul Național de Statistică din România și Banca Internațională pentru Reconstrucție și Dezvoltare la data de 17 septembrie 2019. Acesta face parte din Rezultatul 6d din acordul menționat mai sus.

Mulțumiri

Acest raport a fost întocmit de Tiziana Tuoto și Carlo Vaccari sub coordonarea lui Michael Wild, Statistician senior, Banca Mondială, cu sprijinul expertului echipei locale, respectiv Antoniadă Ciprian Alexandru. Echipa dorește, de asemenea, să-și exprime recunoștința funcționarilor guvernamentali ai Institutului Național de Statistică (INS), doamnelor Silvia Pisică (Manager de proiect), Lavinia Bălțeanu (Responsabil proiect), Ruxandra Moldoveanu (Expert), Laura Ichim (Expert), Manuela Vlaicu (Expert), Mihaela Ram-Chand (Expert), Florica Frîncu (Expert) și echipei sale de specialiști pentru colaborarea constructivă.

Cuprins

Cuprins.....	4
Listă de figuri.....	5
Listă de acronime.....	5
Rezumat	6
Glosar de bază:.....	7
Context:.....	8
Un cadru model pentru evaluarea calității datelor cu etape și hiper-dimensiuni.....	8
Etape și hiper-dimensiuni ale evaluării calității.....	8
A. Analiza utilității surselor administrative și geo-spațiale în raport cu scopul propus	11
1. Identificarea utilizării preconizate a surselor administrative și geo-spațiale.....	11
2. Dimensiunile calității pentru sursă.....	11
3. Identificarea/proiectarea instrumentului adecvat pentru evaluarea utilității surselor	12
B. Menținerea metadatelor asociate cu datele administrative	14
1. Hiper-dimensiunea Metadata a surselor administrative și geo-spațiale.....	14
2. Un șablon pentru a evalua calitatea datelor	14
C. Identificarea și măsurarea diferitelor tipuri de erori.....	16
D. Recomandări.....	17
Lista de Anexe:.....	18
Anexa A. Lista de verificare a calității pentru etapa de introducere – Sursă.....	19
Anexa A1. Lista de verificare a calității pentru etapa de introducere – Sursa completată pentru Sursa „Agenție fiscală”	24
Anexa B. Lista de verificare a calității pentru etapa de introducere – Metadata	29
Anexa B1. Lista de verificare a calității pentru etapa de introducere – Metadata completată pentru Sursa „Agenție fiscală”.....	35
Anexa C. Șablon de informații despre metadata.....	41
Anexa D. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori.....	42
Anexa D1. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori completată în Sursa „Agenție fiscală”.	52

Listă de figuri

Figura 1 - Principalele etape ale ciclului de viață al datelor	9
Figura 2 - Dimensiunea calității etapei Sursă	12
Figura 3 - Dimensiunea calității în etapa Date.....	15

Listă de acronime

INS	Institutul Național de Statistică
ONS	Oficiu National de Statistică
RAS	Servicii de Asistență Tehnică Rambursabile
BM	Banca Mondială

Rezumat

Scopul acestui raport este de a oferi îndrumare și recomandări pentru o evaluare continuă și în timp real a calității datelor colectate din surse administrative și geo-spațiale în perioadele intercenzitare, cu obiectivul de a crea un cadru complet bazat pe registre și recensământ pentru calcularea numărului de locuitori. Documentul oferă atât măsuri calitative, cât și cantitative pentru evaluarea calității. Instrumentele de operare pentru evaluarea calității sunt reprezentate de trei liste de verificare și șablonurile corespunzătoare, furnizate în anexe. Documentul este conceput ca un ghid pentru utilizarea și compilarea listelor de verificare.

Acesta face parte din livrabilele din cadrul Acordului privind Serviciile de Asistență Tehnică Rambursabile (RAS) privind consolidarea Sistemului Statistic Național al României (proiectul nr. P167217). Proiectul este implementat de Institutul Național de Statistică cu sprijinul Băncii Mondiale.

Acest raport este împărțit în 4 (patru) secțiuni și include șapte anexe care fac parte integrantă din raport.

Secțiunea Context introduce conceptele de „Etapă” și „Hiper-dimensiune” utilizate în raport. Secțiunea A oferă o descriere a utilității datelor administrative și geo-spațiale în perioadele intercenzitare, concentrându-se pe dimensiunile calității surselor și pe instrumentele propuse pentru evaluarea calității.

Secțiunea B oferă o descriere a hiper-dimensiunii Metadate și a listei de verificare propuse pentru evaluarea calității metadatelor. Secțiunea C prezintă câteva recomandări pentru INS pentru a identifica și măsura diferitele tipuri de erori care pot apărea în gestionarea unui cadru bazat pe recensământ/registre.

Anexele completează raportul cu lista de verificare și șabloanele propuse INS pentru a evalua calitatea datelor și a metadatelor. De asemenea, au fost compilate împreună trei liste de verificare pe o sursă pilot de către grupul de lucru INS-BM, pentru a finaliza asistența tehnică practică și a verifica împreună corespondența instrumentelor cu nevoile INS.

Glosar de bază:

Cadru model: orice listă, material sau dispozitiv care delimitează, identifică și permite accesul la elementele populației țintă. Un exemplu specific este un registru statistic, precum și un registru administrativ.

- *Registrul populației:* un registru statistic și un cadru model al persoanelor de obicei rezidente (definite în orice mod) într-o țară. În plus, acesta oferă, adesea, unele caracteristici demografice ale persoanelor.

Unitate: cea mai mică entitate statistică la care se referă orice element de date. Unitățile se pot referi la persoane individuale, gospodării, clădiri sau locuințe.

- *Unitate administrativă:* unitățile pentru care sunt înregistrate date administrative. Acestea pot fi sau nu aceleași cu cele necesare pentru rezultatul statistic (care sunt denumite unități statistice). În literatura de specialitate (de exemplu, Zhang 2012), termenul „obiect” este folosit pentru a se referi la unitățile dintr-un set de date administrative. Termenul este folosit pentru a face distincția între unitățile din datele administrative și unitățile statistice după ce aceste date sunt transformate într-un anumit mod. Acest lucru este deosebit de relevant în cazurile în care unitatea (sau „obiectul”) din registrul administrativ diferă de unitatea statistică țintă. De exemplu, în cazul în care un registru fiscal, în care unitățile unei declarații fiscale anuale (adică, aceeași persoană poate face mai multe declarații într-unul sau mai mulți ani) sunt convertite în „persoane” individuale
- *Unitate compusă:* unități compuse din una sau mai multe unități individuale. Exemplu: gospodărie, care este compusă din persoane (în statisticile afacerilor, avem și întreprindere/companie compusă din una sau mai multe unități).
- *Relația dintre diferitele unități țintă*

Variabilă: caracteristici/atribute socio-demografice sau economice referitoare la o unitate administrativă sau statistică pentru care sunt necesare informații în scopul cadrului

- *Variabile de bază:* variabile care identifică și permit accesul la elementele populației țintă. Exemple: pentru actualizarea și întreținerea unui cadru bazat pe administrare/recensământ, variabilele principale sunt CNP, numele, prenumele, sexul, locul și data nașterii, informații de contact precum adresa de reședință (de preferință geo-referențiată), municipiul, adresa de e-mail, telefonul. În acest fel, puteți afla statisticile demografice de bază (alcătuirea populației în funcție de vârstă și sex), precum și informațiile de contact pentru anchetele prin sondaj. Este necesară o discuție specifică pentru variabila „statut de rezident” care specifică o caracteristică a apartenenței la populația țintă.
- *Variabile secundare:* orice alte variabile care nu contribuie la identificarea și accesul la elementele populației țintă. Exemple: pentru actualizarea și menținerea unui cadru bazat pe registru/recensământ, variabilele secundare sunt nivelul de studii, starea civilă, statutul de angajare etc. Aceste variabile pot fi obținute din surse administrative, dar nu sunt esențiale pentru identificarea și contactarea unităților țintă. Uneori, informațiile disponibile din sursele administrative nu corespund perfect definițiilor/cerințelor statistice, care trebuie derivate prin anchete prin sondaj.
- *Variabilă derivată:* o nouă variabilă formată prin utilizarea datelor din alte variabile. Variabila „statut de rezident” este adesea o variabilă derivată în cadrul statistic

Context:

Un cadru model pentru evaluarea calității datelor cu etape și hiper-dimensiuni

Calitatea unui cadru model produs folosind surse administrative este deosebit de dificil de evaluat și/sau măsurat din cauza complexității și multidimensionalității datelor utilizate. Unii factori care afectează calitatea nu sunt măsurabili cantitativ. Putem distinge între evaluarea calității, adică o evaluare calitativă, și măsurarea calității – adică atașarea unei valori cantitative acestei evaluări a calității. Acolo unde nu este posibil să se producă indicatori pentru măsurarea cantitativă sau acolo unde aceștia nu au fost încă elaborați, recomandăm o evaluare calitativă a impactului acestora asupra calității.

Un cadru care utilizează surse de date din recensământ și administrative se bazează parțial pe date care au fost produse în afara sistemului statistic, într-o organizație diferită asupra căreia INS nu deține de obicei controlul. Din acest motiv, impactul utilizării acestor *surse* exterioare trebuie luat în considerare cu atenție. Din acest motiv, putem aplica un cadru de calitate înainte de utilizarea statistică a datelor administrative, adică în etapa de introducere, pentru a determina dacă o sursă de date administrative poate fi utilizată în scopuri statistice și dacă da, în ce mod. Pentru a clarifica și sistematiza etapele (hiper-dimensiunile) unui cadru de evaluare a calității datelor secundare, putem consulta un standard internațional.

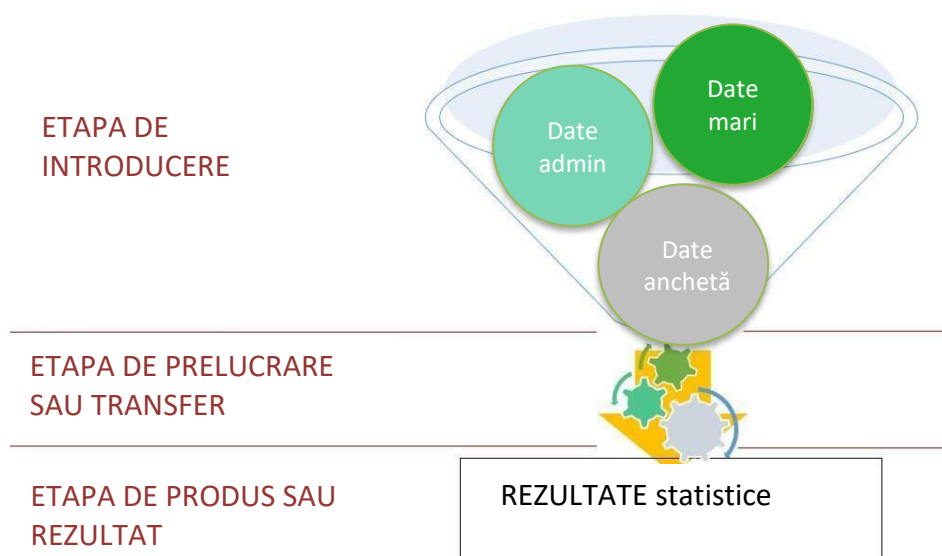
Cadrul de evaluare a calității are unele caracteristici dorite. În special, evaluarea este probabil să înceapă fără date disponibile. Ar trebui să existe un pas inițial în cadrul de evaluare care să ghideze achiziția (obținerea, preluarea) inițială de date și care să poată fi aplicat cu sau fără date disponibile. În plus, evaluarea ar trebui să ofere un instrument care ia în considerare mai multe utilizări potențiale statistice ale datelor administrative. În cele din urmă, evaluarea ar trebui să ofere puncte comune cu dimensiunile regăsite în cadrele recunoscute de evaluare a calității.

Etape și hiper-dimensiuni ale evaluării calității

Pentru a asigura o evaluare a calității completă și ușor de urmărit, putem lua în considerare etapele principiilor ale ciclului de viață al surselor administrative. Ele sunt aplicabile indiferent de utilizarea prevăzută. În timp ce proiectarea proceselor statistice nu este niciodată complet liniară, identificarea modului de a realiza evaluarea calității în acest fel, ar trebui, să ne permită să identificăm rapid considerentele cheie de calitate cele mai relevante pentru fiecare circumstanță. În plus, acest lucru ne permite să personalizăm fiecare etapă pentru a acoperi mai multe dimensiuni ale calității și indicatori de calitate asociați.

Principalele etape ale ciclului de viață al datelor secundare sunt prezentate în figura 1.

Figura 1 - Principalele etape ale ciclului de viață al datelor



Etapele sunt:

(a) **Etapa de introducere:** în această etapă recunoaștem hiper-dimensiunea *Sursă* și hiper-dimensiunea *Metadata* și trebuie să descriem atât evaluarea calității sursei, cât și pe cea bazată pe metadata pentru sursele administrative noi sau furnizate din nou care urmează să fie utilizate. Această Etapă nu necesită ca INS să se afle în posesia datelor reale, dar este crucială pentru etapele care urmează. În cazurile în care datele sunt deja disponibile, se recomandă finalizarea evaluării calității pentru *Sursă* și *Metadata*, deoarece contribuie la documentarea aspectelor de calitate care pot fi esențiale pentru luarea unei decizii finale cu privire la adecvarea pentru utilizare a sursei de date administrative, având în vedere utilizările propuse.

(b) **Etapa de prelucrare:** în această etapă, ne propunem să evaluăm calitatea datelor administrative brute furnizate de autoritățile administrative (furnizorii de date) și recunoaștem hiper-dimensiunea *Date* și hiper-dimensiunea *Prelucrare*. Acest lucru impune Oficiilor Naționale de Statistică (ONS) să valideze datele furnizate în raport cu lecțiile învățate din Etapa *Sursă*. Pe lângă validarea de bază, această Etapă include orice prelucrare necesară pentru a stabili calitatea datelor furnizate în raport cu ceea ce era de așteptat și comparații cu surse alternative. Hiper-dimensiunea *Prelucrare* are ca scop evaluarea calității mai multor procese desfășurate asupra surselor de date administrative, pentru a transforma datele pentru utilizare statistică și/sau pentru a îmbunătăți calitatea. Procesele identificate includ:

- (i) Înregistrarea corelării,
- (ii) Metodologia „Signs-of-life”,
- (iii) Rezolvarea/decizia cu privire la conflicte între surse și
- (iv) Editarea și imputarea.

Transformările la care sunt supuse datele administrative pentru a garanta utilizarea lor statistică nu fac obiectul prezentului raport. Unele dintre acestea vor fi discutate în Rezultatul 6b - *Metodologia privind metodele de eșantionare atunci când se trece la alt tip de cadru de eșantionare (de exemplu, de la eșantionul principal la cadrul bazat pe înregistrare) pentru anchete și statistici în perioadele*

intercenziare, în care vor fi evidențiate aspectele legate de calitate.

(c) **Etapă de rezultate:** evaluarea generală a calității rezultatelor statistice produse cu ajutorul datelor administrative. Acest lucru ar putea să nu fie atât de diferit din punct de vedere conceptual de evaluarea rezultatelor anchetelor tradiționale și, prin urmare, nu face obiectul prezentului raport.

A. Analiza utilității surselor administrative și geo-spațiale în raport cu scopul propus

1. Identificarea utilizării preconizate a surselor administrative și geo-spațiale

În perioadele intercenzitare (inter-recensământ), putem identifica trei scopuri principale:

- a) actualizarea cadrului de registru/recensământ în ceea ce privește acoperirea unităților țintă și informațiile de bază ale acestora (variabile de bază).
- b) actualizarea cadrului de registru/recensământ în raport cu celelalte informații utile disponibile în sursele pentru producerea statisticilor necesare (variabile secundare).
- c) deducerea calculelor aferente unităților țintă defalcate după unele variabile de bază fundamentale.

Institutul Național de Statistică se poate concentra asupra primului scop, deoarece acesta este cel mai important și sursele administrative joacă într-adevăr un rol crucial pentru acesta; pe de altă parte, vom oferi o perspectivă asupra celui de-al doilea scop, care este un produs secundar al celui dintâi, ținând cont de faptul că statisticile pentru variabilele secundare nu pot fi întotdeauna derivate prin cadrul bazat pe registru/recensământ, deși uneori au nevoie de sondaje prin eșantion dedicate (acesta este un caz obișnuit pentru statisticile privind ocuparea forței de muncă, de exemplu).

Al treilea scop, respectiv derivarea, contează pentru unele variabile de bază, este un produs secundar al celui dintâi, spre exemplu un registru /un cadru bazat pe recensământ, actualizat pentru a fi utilizat în scopuri diferite. Cu toate acestea, derivarea numerelor agregate poate fi considerată și o etapă preliminară în comparație cu primul scop și atins într-un stadiu incipient, al cadrului bazat pe registru/recensământ, în conformitate cu unele ipoteze de lucru simplificate.

2. Dimensiunile calității pentru sursă

Conform celor mai acceptate și utilizate standarde internaționale, dimensiunile calității pentru sursele de date pot fi rezumate în Tabelul 1. Este util să se investigheze dimensiunile calității din Tabelul 1 chiar înainte ca INS să aibă acces direct la sursele de date, într-o etapă de descoperire a etapei de achiziție a datelor.

Figura 2 - Dimensiunea calității etapei Sursă

ETAPA SURSĂ	DIMENSIUNEA CALITĂȚII	DEFINIȚIE
	Relevanță și Acuratețe	Gradul în care sursa de date administrative satisface nevoile recensământului. Acesta acoperă suprapunerea dintre populația țintă a recensământului, concepte și definiții (relevanță) și gradul în care datele descriu corect fenomenele pe care au fost concepute să le măsoare (acuratețe).
	Actualitate	Intervalul dintre sfârșitul perioadei de referință la care se referă informațiile și data la care informațiile devin disponibile pentru ONS.
	Coerență și Comparabilitate	Gradul în care sursa administrativă poate fi combinată cu succes cu alte surse utilizate în recensământ, inclusiv posibilitatea de corelare.
	Accesibilitate și Interpretabilitate	Ușurința cu care ONS pot obține datele administrative, acoperind impactul oricăror restricții, confidențialitatea și securitatea, acceptabilitatea publică a utilizării, ușurința transferului și primirii de date și disponibilitatea metadatelor.
	Mediul Instituțional	Factori organizaționali care afectează capacitatea furnizorului de date de a furniza date la calitatea așteptată. Acoperă puterea relației, experiența anterioară, existența unor acorduri formale, riscurile asociate cu statutul furnizorului și standardele de calitate ale furnizorului.

3. Identificarea/proiectarea instrumentului adecvat pentru evaluarea utilității surselor

Analiza utilității în raport cu scopul actualizării și menținerii cadrului bazat pe registru/recensământ poate fi evaluată chiar și fără sursa de date disponibilă, iar acesta este un mare avantaj pentru focalizare și priorități în ceea ce privește etapa de achiziție a datelor.

Un instrument util pentru evaluarea utilității unei surse de date este furnizat în Anexa A. „**Lista de verificare a calității pentru etapa de introducere – Sursa**”. Acest instrument este util și atunci când scopul utilizării este încă neclar la început și Oficiul Național de Statistică efectuează încă anchete/interogatorii printr-o etapă de achiziție a datelor. Acest instrument poate fi adaptat la cazul specific al INS.

În Anexa A1, oferim **Lista de verificare a calității pentru etapa de introducere – Sursa** completate pentru Sursa „Agenție fiscală” de către echipa BM și INS în timpul unei întâlniri virtuale dedicate.

Vă recomandăm să efectuați evaluarea surselor de date în trei pași ulteriori folosind:

- 1) Lista de verificare a sursei;
- 2) Lista de verificare a metadatelor; și, în final,
- 3) Lista de verificare a datelor (o evaluare mai aprofundată a calității).

În general, un rezultat pozitiv într-unul dintre cei trei pași ar indica faptul că trebuie efectuat următorul

pas. Listele de verificare a Sursei și a Metadatelor pot fi efectuate chiar și fără accesarea datelor efective. Listele de verificare sunt un instrument de lucru pentru transpunerea în cifre a conceptelor de calitate. În listele de verificare unele concepte pot fi omise (de exemplu, Coerența și Comparabilitatea nu sunt incluse în Lista de verificare a sursei), altele pot fi subliniate, iar lista de verificare poate fi ajustată în funcție de nevoi și situații specifice.

După cum era anticipat, fiecare pas acoperă mai multe dimensiuni ale calității și indicatori de calitate asociați. În listele de verificare, fiecărui indicator de calitate trebuie să i se atribuie un scor în raport cu elementul de calitate. O atenție deosebită trebuie acordată la alocarea scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare indicând un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate care este evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi folosită pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui să fie luați în considerare în evaluare.

După cum s-a menționat deja, lista de verificare oferă un instrument util chiar și atunci când INS explorează datele fără o utilizare specifică preconizată și face acest lucru pentru a descoperi utilizări potențiale, iar evaluarea servește în continuare verificării interpretabilității și a caracterului complet al informațiilor furnizate.

Este de remarcat faptul că tabelul rezumativ de la sfârșitul listei de verificare oferă o asistență prețioasă pentru a identifica punctele slabe și punctele forte ale surselor luate în considerare.

B. Menținerea metadatelor asociate cu datele administrative

1. Hiper-dimensiunea Metadata a surselor administrative și geo-spațiale

În etapa de Introducere, putem avea o evaluare a calității bazată pe metadata a surselor administrative noi sau refurnizate care urmează să fie utilizate, chiar și atunci când INS nu se află în posesia datelor efective. După cum era de anticipat, atunci când datele sunt deja disponibile, se recomandă finalizarea evaluării calității metadatelor, deoarece contribuie la aspectele legate de calitatea documentelor.

Dimensiunile calității care trebuie luate în considerare pentru hiper-dimensiunea Metadata sunt legate de nivelul de informații furnizate pentru a evalua interpretabilitatea, relevanța și coerența/acuratețea sursei.

Un instrument util pentru evaluarea calității hiper-dimensiunii Metadata este încă o listă de verificare. Anexa B furnizează o **Listă de verificare a calității pentru etapa de introducere – Metadata**, care poate fi personalizată pentru a ține cont de specificul INS.

În Anexa B1, furnizăm **Lista de verificare a calității pentru etapa de introducere – Metadata** completată pentru Sursa „Agenție fiscală” de către echipa BM și INS în timpul unei întâlniri virtuale dedicate.

2. Un șablon pentru a evalua calitatea datelor

Odată ce datele se află în posesia INS, în etapa de Prelucrare, un șablon ar putea ajuta la colectarea și organizarea evaluării calității (hiper-dimensiunea) datelor.

Anexa C furnizează un șablon care este conceput pentru a surprinde aspectele cheie ale calității pentru un anumit set de date într-un mod organizat. Este o parte a evaluării mai ample a calității datelor. Acest șablon trebuie personalizat în funcție de nevoile specifice INS, deoarece este posibil să nu fie utilă fiecare casetă din acest șablon. În plus, nivelul de detaliu poate fi adaptat nevoilor proiectului sau anchetei specifice. Cu toate acestea, vă recomandăm să rețineți că informațiile pe care echipa INS le introduce în acest șablon vor fi foarte valoroase pentru orice alte persoane care utilizează setul de date în viitor. Din acest motiv, ca și în listele de verificare anterioare, vă recomandăm să folosiți un limbaj ușor de înțeles, să includeți toate detaliile și definițiile și să nu faceți prea multe presupuneri despre ceilalți cititori.

Șablonul conține și elemente pentru identificarea și măsurarea diferitelor tipuri de erori. Din nou, nivelul de detaliu la care erorile pot fi măsurate și raportate poate fi adaptat la situația specifică INS.

Dimensiunea calității de luat în considerare în etapa Date poate fi rezumată în figura 3.

Figura 3 - Dimensiunea calității în etapa Date

DIMENSIUNEA CALITĂȚII		DEFINIȚIE
ETAPA DATE	Validare și Armonizare	Fișierele de date furnizate ONS sunt într-un format lizibil. Alte măsuri de validare și armonizare a datelor sunt în vigoare la transferul datelor către ONS. Acest lucru se face pentru a confirma faptul că au fost furnizate variabilele/unitățile/perioada de referință/formatele așteptate și pentru a se asigura că prelucrarea datelor de către ONS este consecventă în cazurile de utilizare în cadrul recensământului.
	Acuratețe și Relevanță	Acuratețea, caracterul complet (pentru variabile și acoperirea populației) și coerența datelor furnizate corespund cerințelor cazului de utilizare specific al recensământului pentru care vor fi utilizate. Comparațiile cu surse alternative relevă măsurarea nivelurilor acceptabile sau erori reprezentative.
	Actualitate și Punctualitate	Actualitatea și punctualitatea datelor furnizate corespund cerințelor cazului de utilizare specific al recensământului pentru care vor fi utilizate.
	Posibilitatea de corelare	Sunt disponibile variabile adecvate pentru corelare (și anume, fie identificatori unici comuni, fie o combinație de variabile care permit identificarea) și acestea sunt de o calitate suficientă pentru a permite corelarea datelor.

O versiune mai scurtă a șablonului este „**Lista de verificare a calității pentru etapa de achiziție - Date**” furnizată în Anexa D. Aceasta colectează cele mai relevante aspecte ale șabloanelor anterioare, în special cele legate de principalele tipuri de erori care ar putea compromite analiza bazată pe datele administrative și geo-spațiale. În Anexa D1, vă oferim **Lista de verificare a calității pentru etapa de achiziție – Date**, completată pentru Sursa „Agenție fiscală” de către echipa BM și INS în timpul unei întâlniri virtuale dedicate.

C. Identificarea și măsurarea diferitelor tipuri de erori

Un cadru bazat pe recensământ/registre trebuie să îndeplinească anumite criterii de calitate în ceea ce privește actualizarea, acoperirea și acuratețea informațiilor conținute în acesta. Cadrul ideal ar trebui să îndeplinească următoarele cerințe:

- Să fie alcătuit numai din unitățile aparținând populației de interes la momentul anchetei.
- Să includă fiecare unitate a populației o singură dată.
- Să conțină date actualizate și corecte privind informațiile de identificare (nume și adresă) și orice informații descriptive de bază (variabile de bază) ale unităților.

Aceste caracteristici permit utilizarea cadrului dintr-o perspectivă strict teoretică a eșantionului, precum și pentru ținta redusă de calcul al unităților țintă prin unele defalcări specifice.

Posibilele situații de abatere de la situația ideală au drept cauză anumite „erori” pe care le-am putea clasifica drept:

- **subacoperire**, adică unele unități ale populației nu sunt incluse în cadru.
- **supraacoperire**, adică unele unități ale cadrului sunt inexistente și/sau nu aparțin populației țintă; un caz special de supraacoperire este *duplicarea* unor unități, dacă unele elemente ale populației sunt prezente de mai multe ori în cadru.
- **valori inexacte și lipsă** în variabilele de bază.
- **clustere de unități**, când unele elemente ale cadrului conțin clustere de elemente ale populației, de exemplu, avem un cadru de gospodării, în timp ce populația țintă este reprezentată de persoane.

Listele de verificare și șabloanele prezentate în secțiunile anterioare conțin indicatori relevanți pentru măsurarea erorilor menționate mai sus.

Este demn de remarcat faptul că, în ceea ce privește conceptele de subacoperire și supraacoperire, acestea sunt strict legate de locația unităților. De fapt, putem distinge cazurile care lipsesc complet din cadru (subacoperire pură), precum și cazurile localizate greșit, de exemplu, unei persoane îi este atribuită o adresă din Sofia în timp ce ea este rezidentă într-un alt oraș/sat. Aceste din urmă cazuri reprezintă subacoperirea locală compensată la nivel național prin supraacoperirea corespunzătoare în altă parte. Aceste subacoperiri și supraacoperiri locale trebuie luate în considerare (și măsurate) cu atenție atât în scopul principal de a stabili și menține un cadru complet, cât și în scopul redus de a stabili dimensiunea populației printr-o anumită defalcare. Prin urmare, deoarece erorile de acoperire trebuie reduse cât mai mult posibil, este oportun să se includă în evaluarea calității toate informațiile geospațiale care pot fi obținute de sursele secundare, pentru o evaluare clară a impactului acestora în ceea ce privește erorile de acoperire.

D. Recomandări

Mai multe recomandări pot fi considerate pentru punerea în aplicare pentru sursele de date administrative, respectiv:

- (a) Identificarea surselor administrative relevante și promițătoare pentru utilizare în actualizarea cadrului pe bază de recensământ/registru.
- (b) Stabilirea în mod clar a populației țintă, variabilelor și conceptelor necesare, împreună cu rezultatele anticipate pentru utilizarea unei surse administrative pe care să se bazeze evaluarea.
- (c) Înțelegerea restricțiilor și provocărilor legate de achiziționarea și integrarea surselor administrative, inclusiv acolo unde ar putea fi necesare modificări ale metodelor, proceselor și sistemelor de calcul ale INS.
- (d) Construirea și menținerea clarității și caracterului cuprinzător al metadatelor care captează toate informațiile relevante de calitate despre o sursă (acest lucru va oferi o resursă valoroasă pentru INS). Este importantă structurarea metadatelor folosind un format standard de metadata adecvat și convenit în prealabil.
- (e) Dezvoltarea unei bune înțelegeri a furnizorului de date, a contextului și a scopului colectării datelor, precum și a standardelor de calitate pe care le respectă.
- (f) Construirea de relații puternice cu furnizorul de date, pentru a asigura schimbul eficient de informații – construind o înțelegere comună a nevoilor celuilalt.
- (g) Punerea în aplicare de acorduri oficiale, care subliniază în mod clar cerințele, rolurile și responsabilitățile INS și ale furnizorilor de date.
- (h) Evaluarea cu atenție a valorii de achiziție și utilizare a unei surse administrative, în raport cu orice riscuri și costuri. Acest lucru poate fi legat de stabilitatea unei surse în timp și de riscul ca un furnizor de date să nu furnizeze datele la timp sau la calitatea așteptată.
- (i) Asigurarea că există o bază legală solidă pentru primirea și utilizarea unei surse administrative, cu garanții efective pentru a proteja confidențialitatea persoanelor vizate.
- (j) Asigurarea clarității și transparenței cu privire la utilizarea datelor administrative, prezentând dovezi că beneficiile depășesc orice preocupări legate de confidențialitate.
- (k) Acceptarea că obiectele, definițiile, conceptele și perioadele de timp de referință dintr-o sursă administrativă pot să nu se alinieze cu obiectivele statistice. Prin urmare, va fi necesar să se transforme datele și să se emită judecăți asupra nivelurilor de nealiniere acceptabile.
- (l) Evaluarea calității în mod continuu (folosind procesul și instrumentele prezentate) – răspunzând la orice modificări anticipate sau cunoscute ale unei surse. Documentarea și publicarea punctelor forte și punctelor slabe aferente surselor administrative, astfel încât utilizatorii de date să aibă încredere în date și să poată ține cont de orice limitări.
- (m) Pregătirea pentru faptul că va dura pentru a înțelege și a achiziționa sursele de date administrative, în special acolo unde este necesar un plan de lucru pentru a dezvolta registre.

Lista de Anexe:

Anexa A. Lista de verificare a calității pentru etapa de introducere – Sursă

Anexa A1. Lista de verificare a calității pentru etapa de introducere – Sursă completată pentru Sursa „Agenție fiscală”

Anexa B. Lista de verificare a calității pentru etapa de introducere – Metadate

Anexa B1. Lista de verificare a calității pentru etapa de introducere – Metadate completată pentru Sursa „Agenție fiscală”

Anexa C. Șablon de informații despre metadate

Anexa D. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori

Anexa D1. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori completată pentru Sursa „Agenție fiscală”

Anexa A. Lista de verificare a calității pentru etapa de introducere – Sursă

3.0	Lista de verificare a calității pentru etapa de descoperire - Sursă			
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, metadatele și datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv la unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribue un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date.			

3.1	Mediul instituțional			
		Descriere	Scor	Clasificare
3.1.1	Statusul furnizorului de date (1) Evaluati riscul general ca furnizorul de date să nu îndeplinească cerințele de calitate ale INS. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	Descrieți tipul de organizație (publică, privată, constituită sau nu, reputație etc.).		
3.1.2	Statusul furnizorului de date (2)	Descrieți factorii care ar putea afecta sustenabilitatea în timp (de exemplu,		

	Evaluati riscul ca ADS sa nu mai fie produsă sau pusă la dispoziție de către furnizorul de date. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	schimbarea priorităților, schimbarea legislației, instabilitatea afacerii etc.).		
3.1.3	Statusul collectorului de date Dacă furnizorul de date colectează datele de la alte organizații, evaluați riscurile suplimentare aduse de acești colectori de date atât cu privire la furnizarea, cât și la calitatea ADS (astfel cum este descris în 3.1.1 și 3.1.2). Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu, NA: nu este cazul (nu există colectori de date distincți)			

3.2	Interpretabilitate			
		Descriere	Scor	Clasificare
3.2.1	Documentație Există informații disponibile și accesibile pentru următoarele? 1. Aspectul înregistrării fișierului 2. Dicționar de date pentru fișiere (obiecte și variabile) 3. Perioada de referință 4. Perioada de timp asociată cu datele (punctul istoric de început și de final) 5. Proceduri de colectare a datelor 6. Proceduri de prelucrare a datelor (inclusiv captură, codificare, editare, imputare) 7. Proporția obiectelor și variabilelor lipsă 8. Rata de imputare 9. Modificări istorice ale procedurilor de colectare și prelucrare a datelor 10. Altele Oferiți un scor pentru fiecare element enumerat mai sus. Scor: 1: nu, 3: da, NS: nu știu	Descrieți la nivel global tipul de documentație și metadata disponibile (concept, colectare de date, prelucrare etc.).		
3.2.2	Feedback Are INS posibilitatea de a contacta furnizorul sursei de date pentru a pune întrebări sau pentru a obține			

	clarificări cu privire la informațiile furnizate atunci când se consideră necesar?			
	Scor: 1: nu, 2: poate, 3: da, NS: nu știu			

3.3	Accesibilitate			
		Descriere	Scor	Clasificare
3.3.1	Restricții Evaluați riscul potențial de a se confrunța cu restricții în accesul și utilizarea datelor de către INS. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	Enumerați legea, <i>norma</i> sau altă dispoziție legală sau de reglementare în temeiul căreia sursa de date este colectată și păstrată de colectorul de date și obținută de furnizorul de date. Indicați orice restricții sau limitări (pot fi precizate ca termeni și condiții) privind utilizarea datelor de către INS și diseminarea rezultatelor statistice rezultate.		
3.3.2	Transmiterea fișierelor Sunt acordurile de transmitere a datelor către INS acceptabile pentru INS atât în ceea ce privește securitatea transmisiei, cât și echipamentul necesar? Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Descrieți formatul prevăzut (de exemplu, fișier plat, bază de date relațională, formate SAS/Sybase) și procedura de transmitere a datelor prevăzută. Specificați dacă este necesar vreun software specific pentru a avea acces la date.		
3.3.3	Percepția publică (1) Înainte ca INS să utilizeze datele, va fi necesar să se efectueze o evaluare a impactului asupra confidențialității sau consultări publice? Scor: 1: da, 2: poate, 3: nu, NS: nu știu			
3.3.4	Percepția publică (2) Dacă este probabil să fie necesare consultări publice, cum evaluați riscul ca acest lucru să încetinească în mod semnificativ procesul de achiziție sau să îi			

	afecteze relevanța (de exemplu, dacă este limitat accesul la date)? Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu, NA: nu este cazul (nu sunt necesare consultări publice)			
3.3.5	Ușurința accesării Se așteaptă ca datele să fie ușor de citit odată ce sunt transmise către INS? Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Descrieți nivelul eforturilor de a citi datele odată transmise.		

3.4	Punctualitate			
		Descriere	Scor	Clasificare
3.4.1	Livrarea fișierului (1) Condițiile de livrare sunt acceptabile pentru INS? De exemplu, luați în considerare oportunitatea pentru disponibilitatea datelor ținând cont de cerințele INS (de exemplu, ciclul de producție). Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Documentați termenii potențiali de livrare a fișierelor (timpul și frecvența).		
3.4.2	Livrarea fișierului (2) Estimați riscul pentru INS ca sursa de date să nu fie livrată la timp. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, 0: nu știu			

3.5	Relevanță			
		Descriere	Scor	Clasificare
3.5.1	Confidențialitate Având în vedere politica de confidențialitate a INS, cum evaluați riscul ca aceasta să limiteze utilizarea preconizată a datelor? Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, 0: nu știu			

3.5.2	Obiecte Conține fișierul tipul de obiecte și seturi de obiecte necesare pentru a îndeplini cerințele INS pentru un anumit rezultat statistic? Scor: 1: nu, 2: parțial, 3: da, 0: nu știu	Descrieți la nivel global tipurile de obiecte și seturile de obiecte ale ADS.		
3.5.3	Variabile Conține fișierul variabilele necesare pentru a îndeplini cerințele INS pentru un anumit rezultat statistic? Scor: 1: nu, 2: parțial, 3: da, 0: nu știu	Descrieți variabilele relevante găsite în ADS		

3.6	Revizuirea sursei
3.6.1	Când informațiile de la furnizorul sursei de date sunt incomplete, colectorul sursei de date trebuie contactat dacă este diferit de furnizorul sursei de date. Este necesar să contactați colectorul sursei de date? Dacă da, adresați-vă colectorului de surse de date. Completați informațiile care lipsesc în lista de verificare de mai sus. Luați în considerare reflectarea costurilor suplimentare legate de această etapă suplimentară în rândul 2.3.1 din șablonul „Cerințele INS și utilizarea preconizată a datelor”.
3.6.2	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un scor anume și o anumită clasificare de mai sus. Privind tabelul rezumativ de mai jos, sunt cerințele importante îndeplinite (și anume, clasificările pe locuri superioare nu au scor Scăzut sau Nu știu)? Nu: puteți decide să opriți evaluarea și să raportați constatările; Da: continuați cu lista de verificare METADATA.

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzută						
Fără clasificare						
TOTAL						

Anexa A1. Lista de verificare a calității pentru etapa de introducere – Sursa completată pentru Sursa „Agenție fiscală”

3.0	Lista de verificare a calității pentru etapa de descoperire - Sursă
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, metadatele și datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv la unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribue un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date.

SURSĂ: Agenția Națională de Administrare Fiscală

Compileri: Moldoveanu Ruxandra și Antoniadă Ciprian Alexandru

Data: 04-02-2022

3.1	Mediul instituțional			
		Descriere	Scor	Clasificare
3.1.1	Statusul furnizorului de date (1) Evaluați riscul general ca furnizorul de date să nu îndeplinească cerințele de calitate ale INS. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	Descrieți tipul de organizație (publică, privată, constituită sau nu, reputație etc.).	3	
3.1.2	Statusul furnizorului de date (2)	Descrieți factorii care ar putea afecta sustenabilitatea		

	Evaluati riscul ca ADS sa nu mai fie produsă sau pusă la dispoziție de către furnizorul de date. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	în timp (de exemplu, schimbarea priorităților, schimbarea legislației, instabilitatea afacerii etc.).	3	
3.1.3	Statusul collectorului de date Dacă furnizorul de date colectează datele de la alte organizații, evaluați riscurile suplimentare aduse de acești colectori de date atât cu privire la furnizarea, cât și la calitatea ADS (astfel cum este descris în 3.1.1 și 3.1.2). Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu, NA: nu este cazul (nu există colectori de date distincți)		3	

3.2	Interpretabilitate			
		Descriere	Scor	Clasificare
3.2.1	Documentație Sunt disponibile și accesibile informații pentru următoarele? 1. Aspectul înregistrării fișierului 2. Dicționar de date pentru fișiere (obiecte și variabile) 3. Perioada de referință 4. Perioada de timp asociată cu datele (punctul istoric de început și de final) 5. Proceduri de colectare a datelor 6. Proceduri de prelucrare a datelor (inclusiv captură, codificare, editare, imputare) 7. Proportia obiectelor și variabilelor lipsă 8. Rata de imputare 9. Modificări istorice ale procedurilor de colectare și prelucrare a datelor 10. Altele Oferiți un scor pentru fiecare element enumerat mai sus. Scor: 1: nu, 3: da, NS: nu știu	Descrieți la nivel global tipul de documentație și metadate disponibile (concept, colectare de date, prelucrare etc.). * nu știm ce transformare efectuează la sfârșitul exercițiului financiar pentru DEPOZIT	3 3 3 3 3 2* 3** N.A. -3 3	
3.2.2	Feedback Are INS posibilitatea de a contacta furnizorul sursei de date pentru a pune întrebări sau pentru a obține		3	

	clarificări cu privire la informațiile furnizate atunci când se consideră necesar?			
	Scor: 1: nu, 2: poate, 3: da, NS: nu știu			

3.3	Accesibilitate			
		Descriere	Scor	Clasificare
3.3.1	Restricții Evaluați riscul potențial de a se confrunța cu restricții în accesul și utilizarea datelor de către INS. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu	Enumerați legea, <i>norma</i> sau altă dispoziție legală sau de reglementare în temeiul căreia sursa de date este colectată și păstrată de colectorul de date și obținută de furnizorul de date. Indicați orice restricții sau limitări (pot fi precizate ca termeni și condiții) privind utilizarea datelor de către INS și diseminarea rezultatelor statistice rezultate.	3* Confidențialitatea a fost soluționată prin criptare	
3.3.2	Transmiterea fișierelor Sunt acordurile de transmitere a datelor către INS acceptabile pentru INS atât în ceea ce privește securitatea transmisiei, cât și echipamentul necesar? Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Descrieți formatul prevăzut (de exemplu, fișier plat, bază de date relațională, formate SAS/Sybase) și procedura de transmitere a datelor prevăzută. Specificați dacă este necesar vreun software specific pentru a avea acces la date.	3	
3.3.3	Percepția publică (1) Înainte ca INS să utilizeze datele, va fi necesar să se efectueze o evaluare a impactului asupra confidențialității sau consultări publice? Scor: 1: da, 2: poate, 3: nu, NS: nu știu		3	
3.3.4	Percepția publică (2) Dacă este probabil să fie necesare consultări publice, cum evaluați riscul ca acest lucru să încetinească în mod semnificativ procesul de achiziție sau să îi		N.A.	

	afecteze relevanța (de exemplu, dacă este limitat accesul la date)? Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, NS: nu știu, NA: nu este cazul (nu sunt necesare consultări publice)			
3.3.5	Ușurința accesului Se așteaptă ca datele să fie ușor de citit odată ce sunt transmise către INS? Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Descrieți nivelul eforturilor de a citi datele odată transmise.	3	

3.4	Punctualitate			
		Descriere	Scor	Clasificare
3.4.1	Livrarea fișierului (1) Condițiile de livrare sunt acceptabile pentru INS? De exemplu, luați în considerare oportunitatea pentru disponibilitatea datelor ținând cont de cerințele INS (de exemplu, ciclul de producție). Scor: 1: nu, 2: parțial, 3: da, NS: nu știu	Documentați termenii potențiali de livrare a fișierelor (timpul și frecvența).	3	
3.4.2	Livrarea fișierului (2) Estimați riscul pentru INS ca sursa de date să nu fie livrată la timp. Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, 0: nu știu		3	

3.5	Relevanță			
		Descriere	Scor	Clasificare
3.5.1	Confidențialitate Având în vedere politica de confidențialitate a INS, cum evaluați riscul ca aceasta să limiteze utilizarea preconizată a datelor? Scor: 1: risc ridicat 2: risc mediu, 3: risc scăzut, 0: nu știu		3	

3.5.2	Obiecte Conține fișierul tipul de obiecte și seturi de obiecte necesare pentru a îndeplini cerințele INS pentru un anumit rezultat statistic? Scor: 1: nu, 2: parțial, 3: da, 0: nu știu	Descrieți la nivel global tipurile de obiecte și seturile de obiecte ale ADS.	2* * în funcție de rezultatul pe care ne concentrăm	
3.5.3	Variabile Conține fișierul variabilele necesare pentru a îndeplini cerințele INS pentru un anumit rezultat statistic? Scor: 1: nu, 2: parțial, 3: da, 0: nu știu	Descrieți variabilele relevante găsite în ADS	La fel ca mai sus	

3.6	Revizuirea sursei
3.6.1	Când informațiile de la furnizorul sursei de date sunt incomplete, colectorul sursei de date trebuie contactat dacă este diferit de furnizorul sursei de date. Este necesar să contactați colectorul sursei de date? Dacă da, adresați-vă colectorului de surse de date. Completați informațiile care lipsesc în lista de verificare de mai sus. Luați în considerare reflectarea costurilor suplimentare legate de această etapă suplimentară în rândul 2.3.1 din șablonul „Cerințele INS și utilizarea preconizată a datelor”.
3.6.2	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un scor anume și o anumită clasificare de mai sus. Privind tabelul rezumativ de mai jos, sunt cerințele importante îndeplinite (și anume, clasificările pe locuri superioare nu au scor Scăzut sau Nu știu)? Nu: puteți decide să opriți evaluarea și să raportați constatările; Da: Continuați cu lista de verificare METADATA.

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzută						
Fără clasificare	18	2				
TOTAL						

Anexa B. Lista de verificare a calității pentru etapa de introducere – Metadate

4.0	Lista de verificare a calității pentru etapa de descoperire - Metadate
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, metadatele și datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv într-unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribue un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date. Pentru listele de verificare pentru metadate și date, se utilizează o abordare orientată pe obiecte, deoarece permite aplicarea cadrului la orice tip de date. Componenta abordării orientate pe obiecte este reprezentată de seturile de obiecte, care includ toate obiectele de același tip (i), cărora le sunt asociate elemente (k) și (j) atribute sau variabile. De exemplu, un set de date care conține persoane fizice și întreprinderi poate fi definit ca incluzând două seturi de obiecte, primul set fiind compus din persoane, $Obiect_{(1)}$ și al doilea set de activități, $Obiect_{(2)}$. La elementul de la litera „k” al $Obiectului_{(1)}$, adică persoana fizică de la litera „k”, sunt asociate variabile (j).

4.1	Interpretabilitate (1), relevanță (2)			
		Descriere	Scor	Clasificare
4.1.1	Seturi și tipuri de obiecte (1) Sunt furnizate și suficiente informațiile pentru a descrie seturile de obiecte și tipurile de obiecte? 1: nu, 2: parțial, 3: da	Descrieți diferitele seturi și tipuri de obiecte.		
4.1.2	Seturi și tipuri de obiecte (2) Se potrivesc seturile și tipurile de obiecte cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se utiliza dacă scorul 4.1.1 =1)			
4.1.3	Obiecte relaționale (1) O relație între două tipuri de obiecte poate fi privită ca un tip special de tip de obiect numit obiect relațional. Cheile de identificare sunt un tip de obiecte relaționale care pot permite corelarea între obiecte. Sunt furnizate și suficiente informațiile pentru a descrie obiectele relaționale? 1: nu, 2: parțial, 3: da, NA: nu este cazul (nu este necesar pentru obiecte relaționale în utilizările potențiale de interes)	Descrieți obiectele relaționale, dacă există.		
4.1.4	Obiecte relaționale (2) Se potrivesc obiectele relaționale cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se utiliza dacă scorul 4.1.3 =1), NA: nu este cazul (nu este nevoie de obiecte relaționale)			
4.1.5	Acoperirea setului de obiecte (1) Sunt furnizate și suficiente informațiile pentru a descrie acoperirea setului de obiecte? 1: nu, 2: parțial, 3: da	Descrieți acoperirea setului de obiecte.		
4.1.6	Acoperirea setului de obiecte (2) Se potrivește acoperirea setului de obiecte (de exemplu, unități ale populației și unități statistice,			

	evenimente) cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se folosi dacă scorul 4.1.5=1)			
4.1.7	Set de obiecte - dimensiune de timp - perioadă de referință (1) Sunt furnizate și suficiente informațiile pentru a determina perioada de referință? 1: nu, 2: parțial, 3: da	Descrieți perioada de referință.		
4.1.8	Set de obiecte - dimensiune de timp - perioadă de referință (2) Având în vedere posibilele utilizări, este perioada de referință adecvată? 1: nu, 2: parțial, 3: da, NS: nu știu (a se folosi dacă scorul 4.1.7=1)			
4.1.9	Set de obiecte - dimensiune de timp¹ – modificări istorice (1) Sunt furnizate și suficiente informațiile pentru a determina schimbări în timp care afectează definirea obiectelor? 1: nu, 2: parțial, 3: da, NA: nu este cazul (de exemplu, ADS moment unic în timp)	Descrieți schimbările istorice legate de definirea seturilor de obiecte.		
4.1.10	Set de obiecte - dimensiune de timp – modificări istorice (2) Dacă apar modificări în timp, cum apreciați limitările privind utilizările potențiale? 1: ridicate, 2: medii, 3: scăzute, NS: nu știu (a se folosi dacă scorul 4.1.9 = 1), NA: nu este cazul (de exemplu, ADS moment unic în timp)			
4.1.11	Variabile (1) Sunt furnizate și suficiente informațiile pentru a descrie variabilele de interes? 1: nu, 2: parțial, 3: da	Descrieți variabilele (de interes).		
4.1.12	Variabile (2) Cât de apropiate sunt variabilele de cele necesare pentru utilizările potențiale de interes? NS: nu știu (descrierea lipsește sau este insuficientă) 1: Nu sunt la fel și conversia este imposibilă			

¹ **Dimensiunea sensibilitate** este utilizată în principal pentru a determina efectul modificărilor dependente de timp în compoziția populației asupra calității datelor (Daas et al., 2008b).

	2: Nu sunt la fel, dar conversia este posibilă 3: Identice			
4.1.13	Variabile - dimensiunea timpului - modificări istorice (1) Sunt furnizate și suficiente informațiile pentru a determina schimbările în timp care afectează definirea variabilelor? 1: nu, 2: parțial, 3: da, NA: nu se aplică (de exemplu, moment unic în variabila de timp)	Describeți modificările istorice legate de definirea variabilelor.		
4.1.14	Variabile - dimensiunea timpului - modificări istorice (2) Dacă apar modificări ale variabilei de-a lungul timpului, cum apreciați limitările privind utilizările potențiale? 1: ridicate, 2: medii, 3: scăzute, NS: nu știu (a se folosi dacă scorul 4.1.13 = 1), NA: nu se aplică (de exemplu, ADS moment unic în timp)			

4.2	Coerență, acuratețe			
		Descriere	Scor	Clasificare
4.2.1	Variabile - combinație unică Există o combinație de variabile prezentă care poate fi utilizată pentru a identifica în mod unic obiectele din setul de obiecte? 1: nu, 2: parțial, 3: da			
4.2.2	Variabile - proceduri de colectare Pe baza informațiilor pe care le aveți despre procedurile de colectare (cu referire la 3.1.2, elementul 5), există limitări importante care ar afecta calitatea datelor? 1: da, 2: poate, 3: nu, NS: nu știu (a se folosi dacă punctajul 3.1.2 elementul 5 = 1)	Describeți metodele de colectare utilizate. (Includeți frecvența și momentul colectării). Documentați dacă instrumentul de colectare (chestionar sau formular) și modul de colectare au fost testate și dacă răspunsul împuternicitului este acceptat. Luați în considerare dacă există motive juridice sau financiare pentru care este probabil ca unele domenii de interes să fie foarte de încredere.		

4.2.3	Variabile - captarea și codificarea datelor Există procese de asigurare a calității la captarea datelor? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți modul în care datele sunt captate și codificate, precum și modul în care este asigurată calitatea în timpul acelor etape.		
4.2.4	Variabile - editare Sunt valorile care au eșuat la editare ușor de identificat în fișier? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți metodele utilizate pentru verificările de editare (editări de consecvență, detectarea valorilor aberante etc.).		
4.2.5	Variabile - imputare Sunt valorile modificate ușor de identificat în fișier? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți metodele de imputare.		
4.2.6	Variabile - tratament – modificări în timp Dacă apar schimbări importante de-a lungul timpului pentru diferitele procese de prelucrare descrise (colectarea datelor, codificare, editare și imputare), cum apreciați limitările privind utilizările potențiale de interes? 1: ridicate, 2: medii, 3: scăzute, NS: nu știu, NA: nu se aplică (de exemplu, fără modificări în timp)			

4.3	Rezumatul evaluării calității metadatelor
4.3.1	Când informațiile de la furnizorul sursei de date sunt incomplete, colectorul sursei de date trebuie contactat dacă este diferit de furnizorul sursei de date. Este necesar să contactați colectorul sursei de date? Dacă da, adresați-vă colectorului de surse de date. Completați informațiile care lipsesc în lista de verificare de mai sus. Luați în considerare reflectarea costurilor suplimentare legate de această etapă suplimentară în rândul 2.3.1 din șablonul „Cerințele INS și utilizarea preconizată a datelor.

4.3.2	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un scor anume și o anumită clasificare de mai sus. Privind tabelul rezumativ de mai jos, sunt cerințele importante îndeplinite (și anume, clasificările pe locuri superioare nu au scor Scăzut sau Nu știu)? Nu: puteți decide să opriți evaluarea și să raportați constatările; Da: continuați cu lista de verificare METADATE.
-------	---

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzute						
Fără clasificare						
TOTAL						

Anexa B1. Lista de verificare a calității pentru etapa de introducere – Metadate completată pentru Sursa „Agenție fiscală”

4.0	Lista de verificare a calității pentru etapa de descoperire - Metadate
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, metadatele și datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv într-unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribuie un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date. Pentru listele de verificare pentru metadate și date, se utilizează o abordare orientată pe obiecte, deoarece permite aplicarea cadrului la orice tip de date. Componenta abordării orientate pe obiecte este reprezentată de seturile de obiecte, care includ toate obiectele de același tip (i), cărora le sunt asociate elemente (k) și (j) atribute sau variabile. De exemplu, un set de date care conține persoane fizice și întreprinderi poate fi definit ca incluzând două seturi de obiecte, primul set fiind compus din persoane, Obiect₍₁₎ și al doilea set de activități, Obiect₍₂₎. La elementul de la litera „k” al Obiectului₍₁₎, adică persoana fizică de la litera „k”, sunt asociate variabile (j). Data: 18.02.2022 Compilatori: Moldoveanu Ruxandra și Antoniadă Ciprian Alexandru Sursa: Agenția fiscală

4.1	Interpretabilitate (1), relevanță (2)			
4.1.1	Seturi și tipuri de obiecte (1) Sunt furnizate și suficiente informațiile pentru a descrie seturile de obiecte și tipurile de obiecte? 1: nu, 2: parțial, 3: da	Descrieți diferitele seturi și tipuri de obiecte.	3	
4.1.2	Seturi și tipuri de obiecte (2) Se potrivesc seturile și tipurile de obiecte cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se utiliza dacă scorul 4.1.1 =1)	Utilizare prevăzută: actualizarea cadrului în perioada intercenzitară Deja utilizate: lunar după statistici salariale/statistici de afaceri (pentru verificare)	3	
4.1.3	Obiecte relaționale (1) O relație între două tipuri de obiecte poate fi privită ca un tip special de tip de obiect numit obiect relațional. Cheile de identificare sunt un tip de obiecte relaționale care pot permite legătura între obiecte. Sunt furnizate și suficiente informațiile pentru a descrie obiectele relaționale? 1: nu, 2: parțial, 3: da, NA: nu este cazul (nu este necesar pentru obiecte relaționale în utilizările potențiale de interes)	Descrieți obiectele relaționale, dacă există. Relația: Angajator - salariat Relația: persoană – gospodărie, DAR NU DIN ACESTE DATE VERIFICĂRI SUPLIMENTARE PENTRU ALCĂTUIREA GOSPODĂRIILOR	3	
4.1.4	Obiecte relaționale (2) Se potrivesc obiectele relaționale cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se utiliza dacă scorul 4.1.3 =1), NA: nu este cazul (nu este nevoie de obiecte relaționale)	Principala utilizare a acestei surse va fi actualizarea numărului de rezidenți din România. Pentru aceasta, nu avem nevoie de obiecte relaționale. Sursa nu oferă informații despre obiectele relaționale pentru utilizarea suplimentară ca șablon de cadru. SUNT NECESARE CONSIDERAȚII SUPLIMENTARE		
4.1.5	Acoperirea setului de obiecte (1) Sunt furnizate și suficiente informațiile pentru a descrie acoperirea setului de obiecte? 1: nu, 2: parțial, 3: da	Descrieți acoperirea setului de obiecte.	3	

4.1.6	Acoperirea setului de obiecte (2) Se potrivește acoperirea setului de obiecte (de exemplu, unități ale populației și unități statistice, evenimente) cu utilizările potențiale de interes? 1: nu, 2: parțial, 3: da, NS: nu știu (a se folosi dacă scorul 4.1.5=1)		3	
4.1.7	Set de obiecte - dimensiune de timp - perioadă de referință (1) Sunt furnizate și suficiente informațiile pentru a determina perioada de referință? 1: nu, 2: parțial, 3: da	Descrieți perioada de referință.	3	
4.1.8	Set de obiecte - dimensiune de timp - perioadă de referință (2) Având în vedere posibilele utilizări, este perioada de referință adecvată? 1: nu, 2: parțial, 3: da, NS: nu știu (a se folosi dacă scorul 4.1.7=1)	Date lunare = venituri din salariu Date trimestriale = venituri din salariu Date anuale = alte venituri, dar nu din salariu	3 3 2	
4.1.9	Set de obiecte - dimensiune de timp¹ – modificări istorice (1) Sunt furnizate și suficiente informațiile pentru a determina schimbări în timp care afectează definirea obiectelor? 1: nu, 2: parțial, 3: da, NA: nu este cazul (de exemplu, ADS moment unic în timp)	Descrieți schimbările istorice legate de definiția seturilor de obiecte. Trebuie să facă distincția între modificările în structurile de date și modificările privind obiectele	3	
4.1.10	Set de obiecte - dimensiune de timp – modificări istorice (2) Dacă apar modificări în timp, cum apreciați limitările privind utilizările potențiale? 1: ridicate 2: medii, 3: scăzute, NS: nu știu (a se folosi dacă scorul 4.1.9 = 1), NA: nu este cazul (de exemplu, ADS moment unic în timp)		3	
4.1.11	Variabile (1) Sunt furnizate și suficiente informațiile pentru a descrie variabilele de interes? 1: nu, 2: parțial, 3: da	Descrieți variabilele (de interes).	3	
4.1.12	Variabile (2) Cât de apropiate sunt variabilele de cele necesare pentru utilizările potențiale de interes? NS: nu știu (descrierea lipsește sau este insuficientă) 1: Nu sunt la fel și conversia este imposibilă și	Scop: determinarea statutului de rezident Scop: determinarea venitului net /vârstei Scop: determinarea genului	1 2 3	

	sunt necesare analize elaborate 2: Nu sunt la fel, dar conversia este posibilă 3: Identice			
4.1.13	Variabile - dimensiunea timpului - modificări istorice (1) Sunt informațiile furnizate și suficiente pentru a determina schimbările în timp care afectează definirea variabilelor? 1: nu, 2: parțial, 3: da, NA: nu se aplică (de exemplu, punct unic în variabila de timp)	Descrieți modificările istorice legate de definirea variabilelor.	3	
4.1.14	Variabile - dimensiunea timpului - modificări istorice (2) Dacă apar modificări ale variabilei de-a lungul timpului, cum apreciați limitările privind utilizările potențiale? 1: ridicate 2: medii, 3: scăzute, NS: nu știu (a se folosi dacă scorul 4.1.13 = 1), NA: nu se aplică (de exemplu, ADS punct unic în timp)		3	

4.2	Coerenta, Acuratețe			
		Descriere	Scor	Clasificare
4.2.1	Variabile - combinație unică Există o combinație de variabile prezentă care poate fi utilizată pentru a identifica în mod unic obiectele din setul de obiecte? 1: nu, 2: parțial, 3: da		3 există CNP pentru persoane 1 pentru HH	
4.2.2	Variabile - proceduri de colectare Pe baza informațiilor pe care le aveți despre procedurile de colectare (cu referire la 3.1.2, elementul 5), există limitări importante care ar afecta calitatea datelor? 1: da, 2: poate, 3: nu, NS: nu știu (a se folosi dacă scorul 3.1.2 elementul 5 = 1)	Descrieți metodele de colectare utilizate. (Includeți frecvența și momentul colectării). Documentați dacă vehiculul de colectare (chestionar sau formular) și modul de colectare au fost testate și dacă răspunsul imputernicitului este acceptat. Luați în considerare dacă există motive juridice sau financiare pentru care este probabil ca unele domenii de interes să fie foarte de încredere.	3 Instrumentul de colectare este un formular, care include verificări	

4.2.3	Variabile - captarea și codificarea datelor Există procese de asigurare a calității la captarea datelor? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți modul în care datele sunt captate și codificate, precum și modul în care este asigurată calitatea în timpul acelor etape.	3	
4.2.4	Variabile - editare Sunt valorile care au eșuat la editare ușor de identificat în fișier? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți metodele utilizate pentru verificările de editare (editări de consecvență, detectarea valorilor aberante etc.).	1	
4.2.5	Variabile - imputare Sunt valorile modificate ușor de identificat în fișier? 1: nu, 2: parțial, 3: da, NS: nu știu (de utilizat dacă 3.1.2 elementul 6 = 1)	Descrieți metodele de imputare.	1 Verificați în continuare cu colegii dacă aceste informații sunt furnizate de Agenția fiscală	
4.2.6	Variabile - tratament – modificări în timp Dacă apar schimbări importante de-a lungul timpului pentru diferitele procese de prelucrare descrise (colectarea datelor, codificare, editare și imputare), cum apreciați limitările privind utilizările potențiale de interes? 1: ridicate, 2: medii, 3: scăzute, NS: nu știu, NA: nu se aplică (de exemplu, fără modificări în timp)		NS	

4.3	Rezumatul evaluării calității metadatelor			
4.3.1	Când informațiile de la furnizorul sursei de date sunt incomplete, colectorul sursei de date trebuie contactat dacă este diferit de furnizorul sursei de date. Este necesar să contactați colectorul sursei de date? Dacă da, adresați-vă colectorului de surse de date. Completați informațiile care lipsesc în lista de verificare de mai sus. Luați în considerare reflectarea costurilor suplimentare legate de această etapă suplimentară în rândul 2.3.1 din șablonul „Cerințele INS și utilizarea preconizată a datelor.			

4.3.2	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un scor anume și o anumită clasificare de mai sus. Privind tabelul rezumativ de mai jos, sunt cerințele importante îndeplinite (și anume, clasificările pe locuri superioare nu au scorul Scăzut sau Nu știu)? Nu: puteți decide să opriți evaluarea și să raportați constatările; Da: continuați cu lista de verificare DATE.
-------	---

¹ **Dimensiunea sensibilitate** este utilizată în principal pentru a determina efectul modificărilor dependente de timp în compoziția populației asupra calității datelor (Daas et al., 2008b).

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzută						
Fără clasificare						
TOTAL						

Anexa C. Șablon de informații despre metadate

Anexa C este furnizată în format Excel „6d app C admin-data-quality-metadata-info-template.xls”, împreună cu raportul. Un extras din fișa cu informații și instrucțiuni de utilizare a șabloanelor este prezentat mai jos.

Instrucțiuni pentru acest șablon de informații despre metadate

Acest șablon este conceput pentru a surprinde aspectele cheie ale calității pentru un anumit set de date într-un mod organizat. Utilizați-l ca parte a unei evaluări mai ample a calității datelor, astfel cum este descris în **Ghidul de raportare privind calitatea datelor administrative** (a se vedea PDF în „Fișiere disponibile”).

Nu trebuie să completați fiecare casetă din acest șablon pentru ca acesta să fie util. Adaptați nivelul de detaliere la nevoile proiectului sau anchetei dumneavoastră. Trebuie să rețineți că informațiile pe care le introduceți în acest șablon vor fi foarte valoroase pentru orice alte persoane care vor folosi setul de date în viitor. Încercați să utilizați un limbaj ușor de înțeles, includeți toate detaliile și definițiile și nu faceți prea multe presupuneri despre cititorii dumneavoastră.

Cele mai importante elemente pe care ar trebui să le completați mai întâi sunt:

Informații generale: elementele 1.1–1.6, inclusiv agenția sursă, scopul colectării, rezumatul variabilelor și intervalul de timp al datelor.

Populație: populația țintă, populația administrativă și unitățile de raportare. Este posibil să nu se poată răspunde la elementele legate de acoperire printr-o evaluare rapidă, dar notați tot ce cunoașteți.

Variabile: o scurtă descriere a variabilelor cheie. Pe măsură ce lucrul progresează, înregistrați conceptele țintă pentru variabilele investigate pe măsură ce devin cunoscute.

Colectare: informațiile despre moment/întârziere și metoda de colectare sunt importante și ar trebui să fie ușor de găsit și înregistrat.

Rețineți: este posibil să găsiți o mare parte din aceste informații pentru seturile de date deja utilizate la Statistics NZ în Colectica. Toate elementele vă vor ajuta să obțineți o înțelegere solidă a calității unui set de date și a problemelor care ar putea apărea în urma utilizării acestuia în alt scop. De exemplu, înțelegerea scopului inițial al colectării de date vă poate ghida către variabilele care ar putea fi de o calitate mai bună decât altele și către acoperirea probabilă a datelor.

Înregistrați orice informații utile pentru alte întrebări, dar ignorați casetele nerelevante din șablon. Dacă descoperiți informații relevante mai târziu în evaluare, adăugați-le - în mod ideal, șablonul cu informații despre metadate pentru un anumit set de date ar trebui îmbunătățit și extins pe măsură ce diferiți oameni din Statistics NZ află mai multe despre el. Ar trebui să existe un singur șablon de informații despre metadate pentru toți cei care utilizează un anumit set de date.

Publicat de Statistics New Zealand

xx martie 2016

www.stats.govt.nz

Anexa D. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori

5.0	Lista de verificare a calității pentru etapa de achiziție - Date
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, metadatele și datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv într-unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribue un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date. Pentru listele de verificare pentru Metadate și Date, se utilizează o abordare orientată pe obiecte, deoarece permite aplicarea cadrului la orice tip de date. Componenta abordării orientate pe obiecte este reprezentată de seturile de obiecte, care includ toate obiectele de același tip (i), cărora le sunt asociate elemente (k) și (j) atribuite sau variabile. De exemplu, un set de date care conține persoane fizice și întreprinderi poate fi definit ca incluzând două seturi de obiecte, primul set fiind compus din persoane, Obiect₍₁₎ și al doilea set de activități, Obiect₍₂₎. La elementul de la litera „k” al Obiectului₍₁₎, adică persoana fizică de la litera „k”, sunt asociate variabile (j).

5.1	Accesibilitate			
		Descriere	Scor	Clasificare
5.1.1	Lizibilitate Pot fi accesate toate datele din ADS? 1: nu, 2: parțial, 3: da, 0: nu știu	Raportați problemele întâlnite.		

5.2	Interpretabilitate			
		Descriere	Scor	Clasificare
5.2.1	Conformitatea metadatelor Analiza date versus metadata relevă anomalii care pun datele sub semnul întrebării și ar putea limita utilizarea datelor sau ar putea necesita solicitarea de clarificări de la furnizorul de date? 1: da, 2: parțial, 3: nu, NS: nu știu (metadata sau date absente sau insuficiente)	Descrieți anomaliile când datele nu sunt conforme cu definiția metadatelor. De exemplu, formatul datelor este diferit de cel așteptat sau datele conțin valori în afara intervalului de valori așteptat.		

5.3	Acuratețe, relevanță			
		Descriere	Scor	Clasificare
5.3.1	Seturi de obiecte - supraacoperire (1) Duplicatelor generează supraacoperire. Identificarea duplicatelor este un proces simplu dacă în fișier sunt prezenți identificatori unici la nivel de element pentru un anumit tip de obiect. Dacă nu este cazul, pot fi identificate un număr de variabile care să servească drept identificatori unici pentru elemente. Atunci când fișierul conține multe înregistrări asociate cu un anumit element (k), identificarea duplicatelor folosind această metodă poate necesita mai multe analize, deoarece înregistrările care par identice, atunci când se utilizează un subset de variabile,	Descrieți metoda folosită pentru a identifica duplicatelor, indicați procentul de duplicate și dacă acestea au fost eliminate din ADS.		

	ar putea fi asociate pe drept cu același element cu litera „k” pentru un anumit tip de obiect și să nu reprezinte duplicate. Calculați procentul de elemente duplicate pentru fiecare tip de obiect (și anume, în cadrul fiecărui set de obiecte). Procentul de elemente duplicate după tipul de obiect poate fi calculat ca: $\frac{\text{Nr. de elemente duplicate în ADS}}{\text{Nr. de elemente în ADS}} \times 100\%$ Odată ce sunt identificate duplicatele, acestea ar trebui eliminate din ADS. Fișierul mai curat trebuie utilizat în etapele ulterioare ale procesului de evaluare a calității. Duplicatele pot fi ușor identificate și îndepărtate? 1: nu, 2: parțial, 3: da, NS: nu știu (lipsa variabilelor pentru a identifica elementele în mod unic)			
5.3.2	Seturi de obiecte - supraacoperire (2) Dacă este disponibilă o sursă de date de referință (RDS), aceasta poate fi utilizată pentru a evalua supraacoperirea, adică elementele care sunt în afara domeniului de aplicare. Rețineți că elementele duplicate care nu au putut fi identificate și eliminate din ADS în 5.3.1 vor contribui, de asemenea, la supraacoperire. Identificarea elementelor ADS în afara domeniului de aplicare poate avea loc la nivel de element (nivel micro) prin corelare cu RDS sau la nivel agregat (nivel macro) pentru spațiul de date specific (de exemplu, subpopulații) de interes. De exemplu, puteți calcula procentul de supraacoperire în funcție de tipul de obiect: $\frac{\text{Nr. de elemente ADS care NU sunt în RDS}}{\text{Nr. de elemente din RDS}} \times 100\%$ Cum apreciați impactul erorii de supraacoperire, având în vedere utilizările potențiale de interes? 1: ridicat 2: mediu, 3: scăzut, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate pentru a calcula supraacoperirea.		

5.3.3	Seturi de obiecte - subacoperire Dacă este disponibilă o sursă de date de referință (RDS), aceasta poate fi utilizată pentru a evalua subacoperirea. Indicatorul trebuie calculat după tipul de obiect (adică, în cadrul fiecărui set de obiecte). De exemplu, $\frac{\text{Nr. de elemente ADS care nu intră în RDS}}{\text{Nr. de elemente din RDS}} \times 100\%$ Validarea elementelor poate avea loc la element (nivel micro) prin corelare cu RDS sau la nivel agregat (nivel macro) pentru spațiul de date specific (de exemplu, subpopulații) de interes. Cum apreciați impactul erorii de supraacoperire, având în vedere potențialele utilizări de interes? 1: ridicat 2: mediu, 3: scăzut, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate pentru a calcula subacoperirea		
5.3.4	Seturi de obiecte - selectivitate (favorizare) Dacă este disponibilă o RDS, indicatorii de selectivitate pot fi utilizați pentru a măsura gradul în care elementele din domeniul de aplicare incluse în ADS diferă de elementele din domeniul de aplicare care lipsesc din ADS (favorizare). Identificați în cadrul RDS elementele care sunt comune cu ADS (grupul A) și cele prezente numai în RDS (grupul B). Pentru grupul A și grupul B, comparați statisticile rezumate pentru atributele tipului de obiect (variabile) din spațiul de date de interes. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes (pot fi utilizate histograme, diagrame cu bare). Având în vedere rezultatele obținute cu privire la selectivitate, cum apreciați limitările datelor, date fiind utilizările potențiale? 1: ridicate 2: medii, 3: scăzute, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate.		

5.3.5	Element – non-răspuns Elementul non-răspuns apare atunci când lipsesc toate datele pentru variabilele de interes pentru un anumit element. Zerourile valide nu sunt considerate lipsă. Calculați procentul de elemente cu toate datele lipsă pentru toate variabilele de interes. Acest indicator poate fi calculat separat pentru principalele variabile de interes. $\frac{\text{Nr. de elemente ADS fără date pentru toate variabilele de interes}}{\text{Nr. de elemente ADS}} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați impactul elementului non-răspuns asupra utilizărilor potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere datele ADS disponibile)			
5.3.6	Set de obiecte - non-răspuns - selectivitate Prin definiție, elementele non-răspuns ADS au valori lipsă pentru variabilele principale ale studiului, dar pot conține informații pentru alte variabile care sunt legate de variabilele principale ale studiului. Dacă este cazul, pot fi utilizați indicatori de selectivitate pentru a măsura gradul în care elementele cu răspuns ADS diferă de elementele non-răspuns ADS. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes. Comparați statisticile rezumative pentru atributele tipului de obiect disponibile în spațiul de date de interes pentru elementele cu răspuns și cele non-răspuns (pot fi utilizate histogramele sau diagramele cu bare). Având în vedere rezultatele obținute, cum apreciați limitările datelor privind potențialele utilizări? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.3.7	Variabile - non-răspuns Variabila non-răspuns apare atunci când valorile pentru variabile specifice de interes lipsesc pentru unele			

	elemente. Nu numărați zerourile valide ca lipsă. Când se efectuează această analiză, poate fi luat în considerare gradul în care valorile lipsesc pentru o anumită variabilă.. De exemplu, se poate considera că o variabilă are valori lipsă numai atunci când datele lipsesc pentru un anumit procent de elemente (adică, gradul de elemente lipsă). Calculați procentul de elemente cu date lipsă pentru anumite variabile incluse în spațiul de date de interes. Acestea ar trebui să fie calculate după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori lipsă pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.3.8	Obiecte - variabile non-răspuns - selectivitate Evaluati dacă obiectele cu variabila non-răspuns au caracteristici similare cu obiectele cu date care nu lipsesc. Acestea ar trebui să fie calculate după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). Comparați statisticile rezumative pentru atributele obiectelor disponibile în spațiul de date de interes pentru obiectele cu variabila non-răspuns și obiectele fără variabilă non-răspuns (pot fi utilizate histogramele sau diagramele cu bare. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes. Se pot folosi histograme sau diagrame cu bare. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			

5.4 Coerență				
		Descriere	Scor	Clasificare
5.4.1	Variabile - valori inconsecvente Calculați procentul de elemente care au încălcat regulile de editare pentru intervalul de valori acceptabile pentru o anumită variabilă de interes. Acestea pot fi calculate și după tipul de obiect (adică, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu o valoare validă pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.4.2	Set de variabile - coerență Calculați procentul de elemente care au încălcat regulile de editare pentru coerența dintre variabile. Includeți numai variabilele care fac parte din spațiul de date de interes. Acestea pot fi calculate și după tipul de obiect (adică, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori incoerente}}{\text{Nr. de obiecte ADS}} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.4.3	Variabilă – valori aberante Pentru variabilele de interes care nu au intervale de valori cunoscute, examinați distribuția valorilor pentru a identifica valori aberante. Luați în considerare utilizarea diferitelor tehnici de detectare a valorii exterioare, cum ar fi metoda quartilelor. Valorile aberante ar trebui semnalate. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte).			

	$\frac{\text{Nr. de elemente ADS cu valori aberante pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.4.4	Variabile - ajustări de prelucrare a datelor de către furnizorul sursei de date Calculați procentul elementelor cu valori ajustate (editate) de furnizorul de date pentru principalele variabile de interes. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori editate pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere informațiile disponibile despre variabilele editate)			
5.4.5	Variabile - imputare de către furnizorul sursei de date Calculați procentul elementelor cu valori imputate de furnizorul de date pentru principalele variabile de interes. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori imputabile pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere informațiile disponibile despre variabilele imputate)			
5.4.6	Variabile – rotunjire Rotunjirea poate afecta distribuția valorii.			

	Există vreo dovadă că a avut loc o rotunjire pentru principalele variabile de interes? Acest lucru poate fi detectat prin producerea de statistici rezumative și diagrame sau histograme. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, nu se poate determina în funcție de informațiile disponibile)			
--	---	--	--	--

5.5	Corelare			
		Descriere	Scor	Clasificare
5.5.1	Obiecte relaționale – calitatea corelării (1) În multe cazuri, corelarea cu alte surse de date (ODS), cum ar fi cadrele populației, este o activitate planificată pentru a exploata pe deplin ADS. Pentru aceste cazuri, trebuie acordată o atenție specială calității variabilelor utilizate pentru a efectua corelarea. Rețineți că pot apărea și alte erori în timpul procesului de corelare, totuși aceste tipuri de erori nu fac obiectul acestei liste de verificare. 1. Calculați procentul de elemente: a) Elemente corelate atât în ADS, cât și în ODS $\frac{\text{Nr. de elemente ADS corelate în ODS}}{\text{Nr. de elemente în ADS}} \times 100\%$ $\frac{\text{Nr. de elemente ADS corelate în ODS}}{\text{Nr. de elemente în ODS}} \times 100\%$ b) Procentul de elemente corelate fără ambiguitate (corelare puternică) c) Procentul de elemente corelate cu o corelare slabă (cerințe de corelare pe care le-am redus pentru a corela mai multe elemente) Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, nu se poate determina în funcție de informațiile disponibile)			

5.5.2	Obiecte relaționale – calitatea corelării (2) Calculați procentul de elemente necorelate: d) Reziduu ADS: elemente ADS necorelate $\frac{\text{Nr. de elemente ADS necorelate}}{\text{Nr. de elemente ADS}} \times 100\%$ e) Reziduu ODS: elemente ODS necorelate $\frac{\text{Nr. de elemente ODS necorelate}}{\text{Nr. de elemente ODS}} \times 100\%$ Reziduurile ODS de mari dimensiuni sunt un indiciu al limitărilor privind integrarea datelor. Reziduurile ADS și ODS pot fi un indiciu al erorii de acoperire fie în ADS, fie în ODS. Comparațiile cu RDS, dacă sunt disponibile, pot ajuta la rezolvarea acestui aspect. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind potențialele utilizări de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu			
-------	--	--	--	--

5.6	Rezumatul evaluării calității datelor
5.6.1	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un anumit scor și o anumită clasificare de mai sus.

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzută						
Fără clasificare						
TOTAL						

5.7	Evaluarea finală și decizia de achiziție și utilizare a ADS O decizie finală cu privire la achiziția și utilizarea datelor ar trebui să țină cont de evaluările calității făcute în cele patru liste de verificare (cerința INS, sursă, metadate și date).
5.7.1	Explicați pe ce bază se ia decizia de a achiziționa sau de a nu achiziționa datele și cum vor fi utilizate datele.

Anexa D1. Lista de verificare a calității pentru achiziția datelor, cu identificarea și măsurarea diferitelor tipuri de erori completată în Sursa „Agenție fiscală”.

5.0	Lista de verificare a calității pentru etapa de achiziție - Date
	Vă recomandăm să efectuați evaluarea în trei pași ulteriori folosind: 1) Lista de verificare Sursă, 2) Lista de verificare Metadate și, în final, 3) Lista de verificare Date (o evaluare mai aprofundată a calității). Sursa, Metadatele și Datele pot fi denumite și hiper-dimensiuni. În general, un rezultat pozitiv într-unul dintre cei trei pași ar indica că trebuie continuat cu pasul următor. Listele de verificare Sursă și Metadate constituie etapa de descoperire. Acestea pot fi efectuate înainte de achiziționarea datelor. În cazurile în care datele sunt deja disponibile, se recomandă completarea listelor de verificare Sursă și Metadate, deoarece contribuie la documentarea aspectelor de calitate care pot fi cruciale pentru luarea unei decizii finale cu privire la adecvarea utilizării Sursei de date administrative (ADS), având în vedere utilizarea preconizată de către INS. Fiecare pas acoperă o serie de dimensiuni ale calității și indicatori de calitate asociați. Fiecărui indicator de calitate trebuie să i se atribue un scor pentru a evalua adecvarea ADS pentru utilizare în ceea ce privește elementul de calitate. O atenție specială ar trebui acordată alocării scorului, deoarece scala este adaptată fiecărei întrebări, cu un scor mare care indică un rezultat pozitiv. INS poate, de asemenea, să prioritizeze importanța fiecărui indicator de calitate a fi evaluat, atribuindu-i o clasificare aleasă dintre înaltă, medie și scăzută. Clasificarea poate fi utilizată pentru a identifica un set minim de indicatori de calitate (cei mai importanți) care ar trebui luați în considerare în evaluare. Coloana „Descriere” este folosită pentru a adăuga informații relevante și, de asemenea, pentru a oferi indicații când informațiile nu sunt disponibile. Rețineți că, atunci când INS explorează datele fără o utilizare specifică în minte și face acest lucru pentru a descoperi mai degrabă utilizări potențiale, evaluarea servește în continuare pentru a evalua interpretabilitatea și caracterul complet al informațiilor furnizate. Aceste informații vor fi utile dacă în viitor apar utilizări potențiale pentru același fișier de date. Pentru listele de verificare pentru Metadate și Date, se utilizează o abordare orientată pe obiecte, deoarece permite aplicarea cadrului la orice tip de date. Componenta abordării orientate pe obiecte este reprezentată de seturile de obiecte, care includ toate obiectele de același tip (i), cărora le sunt asociate elemente (k) și (j) atribute sau variabile. De exemplu, un set de date care conține persoane fizice și întreprinderi poate fi definit ca incluzând două seturi de obiecte, primul set fiind compus din persoane, Obiect₍₁₎ și al doilea set de activități, Obiect₍₂₎. La elementul de la litera „k” al Obiectului₍₁₎, adică persoana fizică de la litera „k”, sunt asociate variabile (j). Compileri: Ruxandra și Tiziana Data: 04.03.2022 Date: Agenția fiscală, ultima prevedere anuală (din 2021)

5.1	Accesibilitate			
		Descriere	Scor	Clasificare
5.1.1	Lizibilitate Pot fi accesate toate datele din ADS? 1: nu, 2: parțial, 3: da, 0: nu știu	Raportați problemele întâlnite.	3	

5.2	Interpretabilitate			
		Descriere	Scor	Clasificare
5.2.1	Conformitatea metadatelor Analiza date versus metadata relevă anomalii care pun datele sub semnul întrebării și ar putea limita utilizarea datelor sau ar putea necesita solicitarea de clarificări de la furnizorul de date? 1: da, 2: parțial, 3: nu, NS: nu știu (metadata sau date absente sau insuficiente)	Descrieți anomaliile când datele nu sunt conforme cu definiția metadatelor. De exemplu, formatul datelor este diferit de cel așteptat sau datele conțin valori în afara intervalului de valori așteptat.	3	

5.3	Acuratețe, relevanță			
		Descriere	Scor	Clasificare
5.3.1	Seturi de obiecte - supraacoperire (1) Duplicatale generează supraacoperire. Identificarea duplicatelor este un proces simplu dacă în fișier sunt prezenți identificatori unici la nivel de element pentru un anumit tip de obiect. Dacă nu este cazul, pot fi identificate un număr de variabile care să servească drept identificatori unici pentru elemente. Atunci când fișierul conține multe înregistrări asociate cu un anumit element (k), identificarea duplicatelor folosind această metodă poate necesita mai multe analize, deoarece înregistrările care par identice, atunci când se utilizează un subset de variabile, ar putea fi asociate pe drept cu același element cu litera „k” pentru un anumit tip de obiect și să nu reprezinte duplicate. Calculați procentul de elemente duplicate pentru	Descrieți metoda folosită pentru a identifica duplicatale, indicați procentul de duplicate și dacă acestea au fost eliminate din ADS.	3	

	fiecare tip de obiect (și anume, în cadrul fiecărui set de obiecte). Procentul de elemente duplicate după tipul de obiect poate fi calculat ca: $\frac{\text{Nr. de elemente duplicate în ADS}}{\text{Nr. de elemente în ADS}} \times 100\%$ Odată ce sunt identificate duplicatele, acestea ar trebui eliminate din ADS. Fișierul mai curat trebuie utilizat în etapele ulterioare ale procesului de evaluare a calității. Duplicatele pot fi ușor identificate și îndepărtate? 1: nu, 2: parțial, 3: da, NS: nu știu (lipsa variabilelor pentru a identifica elementele în mod unic)			
5.3.2	Seturi de obiecte – supraacoperire (2) Dacă este disponibilă o sursă de date de referință (RDS), aceasta poate fi utilizată pentru a evalua supraacoperirea, adică elementele care sunt în afara domeniului de aplicare. Rețineți că elementele duplicate care nu au putut fi identificate și eliminate din ADS în 5.3.1 vor contribui, de asemenea, la supraacoperire. Identificarea elementelor ADS în afara domeniului de aplicare poate avea loc la nivel de element (nivel micro) prin corelare cu RDS sau la nivel agregat (nivel macro) pentru spațiul de date specific (de exemplu, subpopulații) de interes. De exemplu, puteți calcula procentul de supraacoperire în funcție de tipul de obiect: $\frac{\text{Nr. de elemente ADS care NU sunt în RDS}}{\text{Nr. de elemente din RDS}} \times 100\%$ Cum apreciați impactul erorii de supraacoperire, având în vedere utilizările potențiale de interes? 1: ridicat 2: mediu, 3: scăzut, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate pentru a calcula supraacoperirea. Scop: numărarea persoanelor rezidente RDS: listă pop rezidenți din Afaceri interne	3	

5.3.3	Seturi de obiecte - subacoperire Dacă este disponibilă o sursă de date de referință (RDS), aceasta poate fi utilizată pentru a evalua subacoperirea. Indicatorul trebuie calculat după tipul de obiect (adică, în cadrul fiecărui set de obiecte). De exemplu, $\frac{\text{Nr. de elemente ADS care nu intră în ADS}}{\text{Nr. de elemente din RDS}} \times 100\%$ Validarea elementelor poate avea loc la element (nivel micro) prin corelare cu RDS sau la nivel agregat (nivel macro) pentru spațiul de date specific (de exemplu, subpopulații) de interes. Cum apreciați impactul erorii de subacoperire, având în vedere potențialele utilizări de interes? 1: ridicat 2: mediu, 3: scăzut, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate pentru a calcula subacoperirea RDS nu este ușor disponibilă, putem folosi un împrumut al acestora, obținând un procent brut de subacoperire	1-2	
5.3.4	Seturi de obiecte - selectivitate (favorizare) Dacă este disponibilă o RDS, indicatorii de selectivitate pot fi utilizați pentru a măsura gradul în care elementele din domeniul de aplicare incluse în ADS diferă de elementele din domeniul de aplicare care lipsesc din ADS (favorizare). Identificați în cadrul RDS elementele care sunt comune cu ADS (grupul A) și cele prezente numai în RDS (grupul B). Pentru grupul A și grupul B, comparați statisticile rezumate pentru atributele tipului de obiect (variabile) din spațiul de date de interes. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes (pot fi utilizate histograme, diagrame cu bare). Având în vedere rezultatele obținute cu privire la selectivitate, cum apreciați limitările datelor, date fiind utilizările potențiale? 1: ridicate 2: medii, 3: scăzute, NS: nu știu (de exemplu, nu este disponibilă o RDS de încredere)	Includeți rezultatele și o descriere a metodelor utilizate. Sunt necesare mai multe investigații pentru a evalua riscul de favorizare	NS	

5.3.5	Element – non-răspuns Elementul non-răspuns apare atunci când lipsesc toate datele pentru variabilele de interes pentru un anumit element. Zerourile valide nu sunt considerate lipsă. Calculați procentul de elemente cu toate datele lipsă pentru toate variabilele de interes. Acest indicator poate fi calculat separat pentru principalele variabile de interes. $\frac{\text{Nr. de elemente ADS fără date pentru toate variabilele de interes}}{\text{Nr. de elemente ADS}} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați impactul elementului non-răspuns asupra utilizărilor potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere datele ADS disponibile)		2	
5.3.6	Set de obiecte - non-răspuns - selectivitate Prin definiție, elementele non-răspuns ADS au valori lipsă pentru variabilele principale ale studiului, dar pot conține informații pentru alte variabile care sunt legate de variabilele principale ale studiului. Dacă este cazul, pot fi utilizați indicatori de selectivitate pentru a măsura gradul în care elementele cu răspuns ADS diferă de elementele non-răspuns ADS. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes. Comparați statisticile rezumative pentru atributele tipului de obiect disponibile în spațiul de date de interes pentru elementele cu răspuns și cele non-răspuns (pot fi utilizate histogramele sau diagramele cu bare). Având în vedere rezultatele obținute, cum apreciați limitările datelor privind potențialele utilizări? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)	Selectivitatea trebuie luată în considerare cu mai multă atenție	NS	
5.3.7	Variabile - non-răspuns Variabila non-răspuns apare atunci când valorile pentru variabile specifice de interes lipsesc pentru unele elemente. Nu numărați zerourile valide ca lipsă. Când se efectuează această analiză, poate fi luat în considerare gradul în care valorile lipsesc pentru o		2-3	

	anumită variabilă. De exemplu, se poate considera că o variabilă are valori lipsă numai atunci când datele lipsesc pentru un anumit procent de elemente (adică, gradul de lipsă). Calculați procentul de elemente cu date lipsă pentru anumite variabile incluse în spațiul de date de interes. Acestea ar trebui să fie calculate după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori lipsă pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.3.8	Obiecte - variabile non-răspuns - selectivitate Evalueați dacă obiectele cu variabila non-răspuns au caracteristici similare cu obiectele cu date care nu lipsesc. Acestea ar trebui să fie calculate după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). Comparați statisticile rezumative pentru atributele obiectelor disponibile în spațiul de date de interes pentru obiectele cu variabila non-răspuns și obiectele fără variabilă non-răspuns. În mod ideal, variabilele utilizate sunt în relație cu rezultatul sau variabilele de studiu de interes. Se pot folosi histogramme sau diagrame cu bare. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)	La fel ca în cazul precedent		

5.4 Coerență				
		Descriere	Scor	Clasificare
5.4.1	Variabile - valori inconsecvente Calculați procentul de elemente care au încălcat regulile de editare pentru intervalul de valori acceptabile pentru o anumită variabilă de interes. Acestea pot fi calculate și după tipul de obiect (adică, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu o valoare validă pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)		1	
5.4.2	Set de variabile - coerență Calculați procentul de elemente care au încălcat regulile de editare pentru coerența dintre variabile. Includeți numai variabilele care fac parte din spațiul de date de interes. Acestea pot fi calculate și după tipul de obiect (adică, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori incoerente}}{\text{Nr. de obiecte ADS}} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)		NS	
5.4.3	Variabilă – valori aberante Pentru variabilele de interes care nu au intervale de valori cunoscute, examinați distribuția valorilor pentru a identifica valori aberante. Luați în considerare utilizarea diferitelor tehnici de detectare a valorii exterioare, cum ar fi metoda quartilelor. Valorile aberante ar trebui semnalate. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori aberante pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați		3	

	limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere variabilele disponibile)			
5.4.4	Variabile - ajustări de prelucrare a datelor de către furnizorul sursei de date Calculați procentul elementelor cu valori ajustate (editate) de furnizorul de date pentru principalele variabile de interes. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori editate pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere informațiile disponibile despre variabilele editate)		NS	
5.4.5	Variabile - imputare de către furnizorul sursei de date Calculați procentul elementelor cu valori imputate de furnizorul de date pentru principalele variabile de interes. Acestea pot fi calculate și după tipul de obiect (și anume, în cadrul fiecărui set de obiecte). $\frac{\text{Nr. de elemente ADS cu valori imputabile pentru variabila } i}{\text{Nr. de elemente ADS în domeniul de aplicare pentru variabila } i} \times 100\%$ Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, imposibil de evaluat având în vedere informațiile disponibile despre variabilele imputate)		NS	

5.4.6	Variabile – rotunjire Rotunjirea poate afecta distribuția valorii. Există vreo dovadă că a avut loc o rotunjire pentru principalele variabile de interes? Acest lucru poate fi detectat prin producerea de statistici rezumative și diagrame sau histograme. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, nu se poate determina în funcție de informațiile disponibile)		3	
-------	---	--	---	--

5.5	Corelare			
		Descriere	Scor	Clasificare
5.5.1	Obiecte relaționale – calitatea corelării (1) În multe cazuri, corelarea cu alte surse de date (ODS), cum ar fi cadrele populației, este o activitate planificată pentru a exploata pe deplin ADS. Pentru aceste cazuri, trebuie acordată o atenție specială calității variabilelor utilizate pentru a efectua corelarea. Rețineți că pot apărea și alte erori în timpul procesului de corelare, totuși aceste tipuri de erori nu fac obiectul acestei liste de verificare. 1. Calculați procentul de elemente: a) Elemente corelate atât în ADS, cât și în ODS $\frac{\text{Nr. de elemente ADS corelate în ODS}}{\text{Nr. de elemente în ADS}} \times 100\%$ $\frac{\text{Nr. de elemente ADS corelate în ODS}}{\text{Nr. de elemente în ODS}} \times 100\%$ b) Procentul de elemente corelate fără ambiguitate (corelare puternică) c) Procentul de elemente corelate cu o corelare slabă (cerințe de corelare pe care le-am redus pentru a corela mai multe elemente) Având în vedere rezultatele obținute, cum apreciați limitările datelor privind utilizările potențiale de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu (de exemplu, nu se poate determina în funcție de informațiile disponibile)			

5.5.2	Obiecte relaționale – calitatea corelării (2) Calculați procentul de elemente necorelate: d) Reziduu ADS: elemente ADS necorelate $\frac{\text{Nr. de elemente ADS necorelate}}{\text{Nr. de elemente ADS}} \times 100\%$ e) Reziduu ODS: elemente ODS necorelate $\frac{\text{Nr. de elemente ODS necorelate}}{\text{Nr. de elemente ODS}} \times 100\%$ Reziduurile ODS de mari dimensiuni sunt un indiciu al limitărilor privind integrarea datelor. Reziduurile ADS și ODS pot fi un indiciu al erorii de acoperire fie în ADS, fie în ODS. Comparațiile cu RDS, dacă sunt disponibile, pot ajuta la rezolvarea acestui aspect. Având în vedere rezultatele obținute, cum apreciați limitările datelor privind potențialele utilizări de interes? 1: impact ridicat 2: impact mediu, 3: impact limitat, NS: nu știu			
-------	--	--	--	--

5.6	Rezumatul evaluării calității datelor
5.6.1	Completați tabelul rezumativ de mai jos cu numărul de indicatori de calitate care au obținut un anumit scor și o anumită clasificare mai sus.

Rezumat – Numărul de indicatori de calitate date fiind scorul și clasificarea lor.

Clasificare	Scor					TOTAL
	3	2	1	NS	NA	
Înaltă						
Medie						
Scăzută						
Fără clasificare						
TOTAL						

5.7	Evaluarea finală și decizia de achiziție și utilizare a ADS O decizie finală cu privire la achiziția și utilizarea datelor ar trebui să țină cont de evaluările calității făcute în cele patru liste de verificare (cerința INS, sursă, metadata și date).
5.7.1	Explicați pe ce bază se ia decizia de a achiziționa sau de a nu achiziționa datele și cum vor fi utilizate datele.



Competența face diferența!

Proiect selectat în cadrul Programului Operațional Capacitate Administrativă, cofinanțat de Uniunea Europeană din Fondul Social European