



ROMÂNIA

Acord de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217)

REZULTAT Nr. 4.2b

Recomandări și îndrumări privind planul documentat destinat sistemului integrat pentru perioadele intercenzitare (lansare în producție)

Ianuarie 2023



Declinarea responsabilității

Acest raport este un produs al personalului Băncii Mondiale. 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 și nu își asumă responsabilitatea pentru orice erori, omisiuni sau inconsecvențe în ceea ce privește informațiile sau răspunderea cu privire la utilizarea sau neutilizarea informațiilor, metodelor, proceselor sau concluziilor prezentate. Granițele, culorile, denumirile și alte informații afișate pe orice hartă din acest document nu implică nicio apreciere din partea Băncii Mondiale cu privire la statutul juridic al oricărui teritoriu sau la aprobarea sau acceptarea acestor granițe. Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului Român.

Declarație privind drepturile de autor

Materialul din această publicație este protejat de drepturi de autor. Copierea și/sau transmiterea unor porțiuni din această lucrare fără permisiune poate constitui o încălcare a legilor aplicabile.

Pentru permisiunea de a fotocopia sau imprima orice parte din această lucrare, vă rugăm să trimiteți o solicitare cu informațiile complete către: (i) Institutul Național de Statistică din România (Bd. Libertății nr. 16, Sector 5, București, România) sau (ii) Grupul Băncii Mondiale România (str. Vasile Lascăr nr. 31, etaj 6, București, România).

Acest raport a fost livrat în luna ianuarie 2023 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 4.2 din acordul menționat mai sus.

Cuprins

Listă de acronime	4
Lista figurilor	5
Introducere	6
1. Rezultate așteptate în perioada intercenzitară.....	7
2. Descrierea datelor de intrare și a algoritmului registrului populației	9
Variabile și dezagregări pentru perioada intercenzitară.....	9
Surse și algoritm pentru actualizarea regulată a statisticilor populației în perioada intercenzitară.....	11
3. Implementarea fluxului de realizare a registrului intercenzitar al populației.....	15
Extragerea datelor	16
Verificarea calității datelor	17
Tratarea / Manipularea datelor.....	17
Stocarea datelor de bază.....	18
Aplicarea algoritmilor – proces de producție.....	19
Registrul intercenzitar (rezultat principal)	19
4. Instrumente pentru actualizarea datelor în registrul populației.....	21
Anexe.....	24
Anexa 1: Fluxul de lucru al registrului intercenzitar al populației	25
Anexa 2: Surse de date administrative, indicatori de calitate, variabile, cronologie	26
Anexa 3: Structură Registru intercenzitar	27

Listă de acronime

ANAF	Agenția Națională de Administrare Fiscală
ANPDCA	Autoritatea Națională pentru Drepturile Copilului și AdopțieAdopții
ANPD	Autoritatea Națională pentru Persoanele cu Dizabilități
ANPIS	Agenția Națională pentru Plăți și Inspecție Socială
ANOFM	Agentia Națională pentru Ocuparea forței de Munca
APIA	Agenția de Plăți și Intervenții în Agricultură
CAPI	Intervievare personală asistată de calculator
CAWI	Intervievare online asistată de calculator
EC/CE	Comisia Europeană
CNAS	Casa Națională de Asigurări de Sănătate
CNP	Cod numeric personal
CNPP	Casa Națională de Pensii Publice
DEPABD	Directia pentru Evidenta Persoanelor si Administrarea Bazelor de Date
DGP	Direcția Generală de Pașapoarte
DOM	Domiciliu
DTS	Direcții teritoriale de statistică
ID	Identificator
IGI	Inspectoratul General pentru Imigrări
INS	Institutul Național de Statistică
IT	Tehnologia informației
ITM	Inspectoratul Teritorial de Muncă
RAS	Servicii de asistență tehnică rambursabile
RESID	Reședință
RNEP	Registrul Național de Evidență a Persoanelor
RMU	Registrul Matricol Unic
PHC	Recensământul populației și locuințelor 2021
SIIR	Sistemul Informatic Integrat al Învățământului din România
SOL	Semne de viață en. “signs of life”
EU/UE	Uniunea Europeană
VPN	Rețea virtuală privată
WB	Banca Mondială

Lista figurilor

Figura 1 - Flux de lucru registru intercenizar	15
Figura 2 - Flux extragere și verificare calitate date	16
Figura 3 - Fluxuri tratare și stocare date	18
Figura 4 - Conexiuni citire/scriere baza de date	18
Figura 5 - Fluxuri aplicare algoritmi, producție rezultat primar și rezultate secundare.....	20

Introducere

Acest raport prezintă **recomandări și îndrumări privind planul documentat (producția de lansare) pentru sistemul integrat pentru perioade intercenzitare**. Acest livrabil face parte din Rezultatul 4.2 din Acordul de servicii de asistență tehnică rambursabile (RAS) privind *Consolidarea Sistemului Statistic Național din România (P167217)*. Proiectul este implementat de Institutul Național de Statistică cu sprijinul Băncii Mondiale.

Raportul prezintă metoda și procesul de creare a unui sistem, care să permită obținerea de informații actualizate privind mărimea populației și caracteristicile acesteia la momentul extragerii sau la un moment de referință specificat, asigurând astfel o listă structurată de unități, fiecare dintre acestea conținând un număr de atribute, un mecanism de actualizare regulată și o componentă de producere a rezultatelor statistice în perioadele intercenzitare.

Raportul este structurat în patru secțiuni și anexele care completează informațiile expuse în cadrul acestuia.

Prima secțiune detaliază rezultatele așteptate în perioada intercenzitară, în contextul trecerii la stabilirea populației statistice bazată pe registre. Sunt stabilite astfel unitățile, relațiile dintre unități, sfera de cuprindere a populației statistice, frecvența de actualizare (lunară, semestrială, anuală), mecanismul de actualizare și componenta de producere a rezultatelor statistice.

A doua secțiune prezintă datele de intrare, variabilele la nivel de persoană, gospodărie și familie, sursele de date și algoritmul pentru actualizarea registrului din sursele de date disponibile în perioada intercenzitară.

Realizarea/implementarea registrului populației este prezentată în secțiunea a treia. Aceasta include fluxul de lucru integrat, respectiv extragerea și verificarea calității datelor, tratarea/manipularea și stocarea datelor, aplicarea algoritmului în procesul de producție pentru obținerea rezultatului primar – registrul intercenzitar și a rezultatelor secundare, precum și cronologia/momentele actualizării.

Secțiunea a patra descrie instrumentele și soluțiile utilizate pentru actualizarea datelor, respectiv infrastructura IT și limbajul de programare aplicabile, precum și funcțiile principale ale acestora. Această secțiune este completată cu o bibliotecă de resurse de informare pentru actualii și viitorii specialiști ai INS care sunt implicați în realizarea și menținerea registrului.

Informații suplimentare cu privire la metodologiile și instrumentele dezvoltate pentru perioada intercenzitară se regăsesc în seria de rapoarte dedicate Rezultatului nr. 6 implementate în cadrul aceluiași acord servicii de asistență tehnică rambursabile.

1. Rezultate așteptate în perioada intercenzitară

Populația statistică în România are la bază măsurările exhaustive realizate decadal pe baza recensămintelor precum și determinări statistice privind populația rezidentă. Considerând Recensământul Populației și Locuințelor (RPL), runda 2021, ca fiind cel mai avansat din punct de vedere al metodelor de culegere și a calității datelor (utilizarea metodelor CAWI-CAPI) și având în vedere nevoia de date actualizate cu o frecvență ridicată (lunară, semestrială, anuală) se poate lua în considerare trecerea la populația statistică bazată pe registre. Astfel este avut în vedere dezvoltarea unui sistem care să permită obținerea de informații actualizate privind mărimea populației și caracteristicile acesteia la momentul extragerii sau la un moment de referință specificat.

Registrul populației statistice va permite obținerea unei liste structurată de unități, fiecare dintre acestea conținând un număr de atribute, un mecanism de actualizare regulată și o componentă de producere a rezultatelor statistice (tabele intercenzitare pentru difuzare, interogări de date / tabele pentru publicațiile INS, cereri de date pentru alte echipe INS, solicitări de date de la terți).

Unități

Registrul va ține evidența următoarelor tipuri de unități:

- Persoana (unitate elementară)
- Gospodărie (unitate compusă)
- Familie (unitate compusă)

Fiecare unitate are un identificator unic.

Relațiile dintre unități

O gospodărie cuprinde toate persoanele care au reședința obișnuită la aceeași adresă și care se gospodăresc în comun (contribuie împreună la aprovizionarea și la consumul de alimente și de alte bunuri esențiale traiului - și participă integral sau parțial la bugetul de venituri și cheltuieli). Persoanele care locuiesc și se gospodăresc singure sunt considerate gospodării de o singură persoană. La fel și persoanele care deși ocupă o locuință împreună cu alte persoane, declară că se gospodăresc singure¹.

O gospodărie poate cuprinde, astfel, una sau mai multe persoane. O persoană poate aparține, la un moment dat, de o singură gospodărie.

O familie (nucleu familial) cuprinde toate persoanele – membri ale aceleiași gospodării - între care există relații de tip soț-soție, partener-parteneră, părinte-copil¹.

O familie cuprinde, astfel, cel puțin două persoane. O persoană poate aparține, la un moment dat, de o singură familie.

O gospodărie poate cuprinde:

- una sau mai multe familii
- una sau mai multe persoane care nu fac parte dintr-o familie (între care nu există relații de tip soț - soție, partener – parteneră, părinte – copil)
- o combinație între cele de mai sus.

În relație cu familia deci, o gospodărie poate cuprinde niciuna, una sau mai multe familii.

Populația statistică

¹ Doar la momentul recensământului această delimitare se poate realiza fără echivoc

Sfera de cuprindere a registrului este determinată de sfera de cuprindere a cercetărilor pe care le alimentează, respectiv:

- Principalele caracteristici socio-demografice pentru:
 - o Populația rezidentă a României
 - o Populația după domiciliu a României
- Migrația internațională
- Baza de sondaj pentru anchetele în gospodării

În concluzie, registrul va ține evidența:

- persoanelor care au reședința obișnuită sau domiciliul în România
- gospodăriilor formate din persoane cu reședința obișnuită în România
- familiilor formate din persoane cu reședința obișnuită în România

Cum una dintre regulile utilizate în construirea registrelor este „unitățile ieșite din sfera de cuprindere nu se șterg ci se marchează ca inactive”, în orice moment registrul va cuprinde toate persoanele/ gospodăriile/ familiile care au avut vreodată reședința obișnuită sau domiciliul pe teritoriul României de la 1 Decembrie 2021 (momentul de referință al RPL) până la momentul respectiv – cu elemente de delimitare a unităților care aparțin fiecărei sfere de cuprindere la un moment dat (adică marcate/ variabile de tip flag care fac posibilă selectarea doar a unităților care aparțin sferei de cuprindere la un moment dat).

2. Descrierea datelor de intrare și a algoritmului registrului populației

Primul pas în construirea registrului este crearea unei liste, cât mai complet posibil, a unităților de interes, cu variabilele disponibile la un moment inițial.

Varianta inițială a registrului se creează din:

- a) înregistrările din Baza de date a RPL2021
- b) persoanele care apar în RNEP² și nu apar în RPL2021 - se adaugă la lista unităților „Persoana” - cu variabile disponibile și statut de rezidență la momentul recensământului „Nerezident”

La momentul creării variantei inițiale a registrului, fiecare atribut al unei unități are o singură valoare (corespunzător situației din 1 decembrie 2021). Dacă registrul va păstra și date istorice (dimensiune temporală) - orice atribut care suportă actualizării va avea valori diferite, valabile în perioade diferite³. Astfel, având la dispoziție datele RPL2021 și sursele de date administrative disponibile dezvoltării și actualizării registrului populației statistice se pot stabili variabilele și dezagregările, precum și pașii pentru actualizarea regulată a datelor în perioada intercenzitară.

Variabile și dezagregări pentru perioada intercenzitară

Variabile necesare în registru (în zona de producție) sunt:

Persoană:

- ID-ul persoanei
- ID-ul mamei
- ID-ul tatălui
- ID-ul soțului
- ID-ul gospodăriei
- ID-ul familiei
- Gradul de rudenie
- Sex
- Data nașterii
 - Zi
 - An
 - Lună
- Tara de naștere
- Tara de cetățenie
- Starea civilă legală
- Nivelul de instruire
- Statutul activității curente
- Statut profesional
- Activitatea economică
- Ocupație

² RNEP furnizat cuprinde persoanele născute înainte de 01.11.2021; persoanele născute în perioada 01.11.2021 – 30.11.2021 lipsesc

³ Register-based Statistics: statistical methods for administrative data, Anders Wallgren, Britt Wallgren - pg. 58-59 și 83-86.

- Locul de munca - localizare - fix sau nu
- Locul de munca - localizare – tara
- Locul de munca - localizare – localitate
- A imigrat in cursul anului de referința
- Țara de origine (doar pentru imigranți)
- Data imigrării
- A emigrat in cursul anului de referința
- Țara de destinație (doar pentru emigranți)
- Data emigrării
- A avut vreodată reședința în străinătate
- Data stabilirii ultimei reședințe în România
 - An
 - Lună
- Adresa de reședința completa
 - ID-ul adresei
 - Strada
 - Număr
 - Bloc
 - Scara
 -
 - Etaj
 - Apartament
 - Municipiu/ oras/ comuna
- Adresa de domiciliu
 - Țara
 - Municipiu/ oras/ comuna
- Persoana cu domiciliu in Romania in perioada de referința
- Persoana cu reședința in Romania in perioada de referința
- Activ (personale care nu sunt decedate sau emigrate - active; restul - pasive)
- Cel mai recent eveniment pe baza căruia s-a actualizat DOM si RESID
- Si-a schimbat reședința in anul curent
- Statutul persoanei in gospodărie
- Statutul persoanei in familie

Gospodărie:

- ID-ul gospodăriei
- Aranjamentele de locuit (in ce fel de spațiu locuiește)
- Tipul gospodăriei
- Titlul sub care gospodăria ocupa locuința
- Mărimea gospodăriei
- Activ (gospodarii formate din persoane rezident - activ; foste gospodarii dizolvate ca urmare a decesului, emigrării, divorțului, mutării - pasiv)

Familii:

- ID-ul familiei
- ID-ul gospodăriei
- Mărimea familiei
- Tipul familiei

- Activ (familii formate din persoane rezidente - activ; foste familii dizolvate ca urmare a decesului, emigrării, divorțului, mutării - pasiv)

Posibilitatea de actualizare a variabilelor din registru depinde în mare măsură de existența și calitatea datelor de intrare. Primul pas însă este crearea procedurilor pentru obținerea listei inițiale de unități (persoane) și de actualizare a acesteia. În cadrul acestei componente, acest aspect a fost tratat în principal, dezvoltări ulterioare fiind posibile, pe același flux de lucru, fără achiziții suplimentare.

Surse și algoritm pentru actualizarea regulată a statisticilor populației în perioada intercenzitară

Actualizarea registrului din sursele de date disponibile – unitățile registrului sunt în continuă schimbare – în orice moment au loc evenimente care modifică lista unităților prin intrări (nașteri și imigrări) sau ieșiri (decese și emigrări), caracteristicile acestora (adresă de reședință sau domiciliu, stare civilă, statutul activității curente, nivel de educație etc.) sau relațiile dintre unități (crearea de noi gospodării/ familii sau dizolvarea lor).

În practică, datele privind modificările devin disponibile lunar, semestrial sau anual. Lista surselor de date disponibile, modul (principal) în care sunt utilizate pentru actualizarea registrului și data la care sunt disponibile la INS este prezentată mai jos:

- a) Surse utilizate în primul rând pentru actualizarea listei unități - cu evenimentele demografice
 - Adăugare nașcuți⁴ (INS) lunar sau semestrial :
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Marcare decedați⁵ (INS) – lunar sau semestrial :
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Marcare/ Adăugare⁶ imigranți (DEPABD) – semestrial :
 - Evenimentele⁷ raportate pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele raportate pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1

⁴ Nașcuți vii - în România sau în străinătate - a căror mama sau tată are domiciliul sau reședința în România

⁵ Decedați - în România sau în străinătate - care au domiciliul sau reședința în România

⁶ Unii dintre ei pot exista deja pentru că existau în RNEP la momentul creării registrului; aceștia nu mai trebuie adăugați ci doar marcați ca rezidenți

⁷ Evenimentele raportate sunt:

- Cetățeni români care și-au stabilit/restabilit domiciliul/reședința obișnuită în România în cursul perioadei de referință
- Persoane care au dobândit cetățenia română în cursul anului și care și-au stabilit reședința obișnuită în România
- Persoane care au redobândit cetățenia română în cursul anului și care și-au stabilit reședința obișnuită în România

- Marcare emigranți⁸ (DGP) - semestrial :
 - Evenimentele raportate pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele raportate pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Adăugare imigranți⁹ (IGI) si marcare emigranti anual:
 - Evenimentele raportate pentru perioada 1 ianuarie – 31 decembrie an N – disponibil pe 15 martie an N+1
- b) Surse utilizate în primul rând pentru actualizare atribute demografice de bază:
- Stabiliri de domiciliu¹⁰ (DEPABD) – semestrial
 - Evenimentele raportate pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele raportate pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Anunțări de reședință¹¹ (DEPABD) – semestrial
 - Evenimentele raportate pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele raportate pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Căsătorii¹² (INS) – semestrial
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1
 - Divorțuri¹³ (INS) – semestrial
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 ianuarie – 30 iunie - an N – disponibil pe 15 septembrie an N
 - Evenimentele înregistrate în buletinele statistice întocmite pentru perioada 1 iulie – 31 decembrie - an N – disponibil pe 15 martie an N+1

⁸ Evenimentele raportate sunt:

- Cetățenii români care și-au stabilit domiciliul în străinătate în cursul perioadei de referință

⁹ Evenimentele raportate sunt:

- Cetățenii Uniunii Europene, SEE/Confederației Elvețiene, cetățenii non-UE/SSE și cei fără cetățenie care au obținut în cursul anului de referință un act care atestă dreptul de ședere în România
- Cetățeni străini non-UE/SSE sau fără cetățenie care au solicitat azil sau sunt refugiați pe teritoriul României și au o durată efectivă de ședere de cel puțin 12 luni
- Cetățenii ai Uniunii Europene, SEE/Confederației Elvețiene, cetățenii non-UE/SSE și cei fără cetățenie care au fost excluși din baza de date a IGI în cursul anului de referință

¹⁰ Persoane care și-au schimbat domiciliu într-o altă localitate administrativă. Nu sunt raportate schimbările de domiciliu în interiorul aceluiași oraș, sector, comună. Pe viitor vor trebui obținute informații și despre persoanele care și-au schimbat domiciliul în cadrul aceleiași localități, inclusiv de la o adresă la alta.

¹¹ Persoane care și-au stabilit reședința într-o altă localitate administrativă decât cea de domiciliu, care la data de 1 ianuarie sau 1 iulie aveau înscrisă în actul de identitate și în fișele de evidență a populației mențiunea de stabilire a reședinței. Nu sunt raportate stabilirile de reședință la o adresă în interiorul aceluiași oraș, sector, comună (ca cel de domiciliu). Pe viitor vor trebui obținute informații și despre persoanele care și-au stabilit reședința în cadrul aceleiași localități, inclusiv de la o adresă la alta.

¹² Căsătorii – încheiate în țară sau în străinătate – între persoane dintre care cel puțin unul are domiciliul sau reședința în România. Nu se raportează căsătoriile din străinătate în cazul în care niciunul dintre soți nu are cetățenie română

¹³ Divorțurile cu hotărâri definitive pronunțate de instanțele judecătorești, prin procedură notarială, la Notarii Publici – din România

- c) Surse utilizate pentru aplicare algoritm pentru determinarea „signs of life” (SOL) – variabila RESID - și pentru actualizarea altor atribute:
- ANAF-D112¹⁴ - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +31
 - ANAF-Depozit – anual:
 - Raportările pentru anul N – disponibile la 30 octombrie anul N+1
 - CNPP lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +45
 - ITM lunar:
 - Contracte de muncă (inclusiv date istorice – adică situații a căror valabilitate a încetat; în plus variabila „spor” este multivaloare) înregistrate până la luna curentă (L) – disponibile la L+27
 - SIIR¹⁵ – bianual:
 - Evenimente (asocieri formațiuni de studiu) și status de la începutul anului școlar N – N+1 (pana la momentul extragerii)– disponibile pe 30 ianuarie an N+1
 - Evenimente (asocieri formațiuni de studiu) și status de la începutul pana la sfârșitul anului școlar N – N+1 – disponibile pe 15 noiembrie an N+1
 - RMU¹⁶ bianual:
 - Evenimente (operații) și școlaritate la începutul anului universitar N – N+1 – disponibile pe 28 februarie an N+1
 - Evenimente (operații) și școlaritate la sfârșitul anului universitar N – N+1 – disponibile pe 30 octombrie an N+1
 - ANPIS - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +30
 - APIA anual:
 - Raportările pentru anul N – disponibile la 30 octombrie anul N+1
 - ANDPDCA¹⁷ - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +30
 - ANOFM¹⁷ - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +30
 - CNAS¹⁷ - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +31
 - ANPD¹⁷ - lunar:
 - Raportările pentru luna 1 – luna curenta (L) – disponibile la L +30
- ci) Îmbunătățirea acoperirii registrului - aplicarea metodei „signs of life” (SOL) prin aplicare algoritmul descris în documentul **ALGORITHM-POPRESZ_17_Jan_2022_lb En.doc**, disponibil la INS http://xx.x.xxx.xx:8081/intercensus/intercensus-documentation/-/blob/main/ALGORITHM-POPRESZ_17_Jan_2022_lb%En.doc
- cii) Selectare populație rezidentă la 1 ianuarie an N cu atribute relevante¹⁸ pentru output-uri:

¹⁴ Se raportează date la nivelul asocierii contribuabil – persoană asigurată – tip venit pentru care se datorează CASS și detaliat

¹⁵ Primit pentru 3 perioade – pentru evaluare

¹⁶ Primit pentru o perioadă – pentru evaluare

¹⁷ Pana în prezent nu s-au primit date

¹⁸ Atât unitățile active cât și atributele unităților selectate se referă la 30 iunie respectiv 31 decembrie

- Rezultate cu grad redus de detaliere:
 - 31 august an N pentru 30 iunie an N
 - 28 februarie an N+1 pentru 31 decembrie an N
 - Rezultate detaliate
 - 31 decembrie an N+1 pentru 31 decembrie an N
- f) Selectare populație cu domiciliu la 1 ianuarie an N cu attribute relevante¹⁹ pentru rezultate
- g) Selectare gospodarii și persoane rezidente – stoc curent la data extragerii²⁰ – cu attribute relevante la data extragerii

Sursele de date menționate pot fi utilizate integral și permanent sau doar în parte la realizarea și actualizarea registrului populației, în funcție de calitatea informației furnizată de acestea și în baza procesului stabilit intern de INS. Totodată, pentru verificarea calității surselor de date se recomandă aplicarea metodelor descrise în rapoartele elaborate pentru producția statistică din perioadele intercenzitare.

- *Rezultat 6b: Raport privind serviciile de asistență tehnică oferite Beneficiarului privind Metodologia referitoare la metodele de eșantionare la trecerea la un alt tip de cadru de eșantion (de exemplu, de la eșantion de bază (eșantion master) la cadrul bazat pe registre) pentru anchete și statistici în perioadele intercenzitare*
- *Rezultat 6c: Raport privind serviciile de asistență tehnică oferite Beneficiarului privind Metodologia de evaluare a efectelor modificării metodei de colectare a datelor, în contextul trecerii la metode multimodale (CAPI/CAWI), pentru sondaje și statistici în perioadele intercenzitare*
- *Rezultatul 6d: Raport privind serviciile de asistență tehnică oferite Beneficiarului pentru Metodologia pentru o evaluare continuă și în timp real a calității datelor colectate din surse administrative și geo-spațiale pentru perioadele intercenzitare.*

¹⁹ Atât unitățile active cât și attributele unităților selectate se referă la 31 decembrie

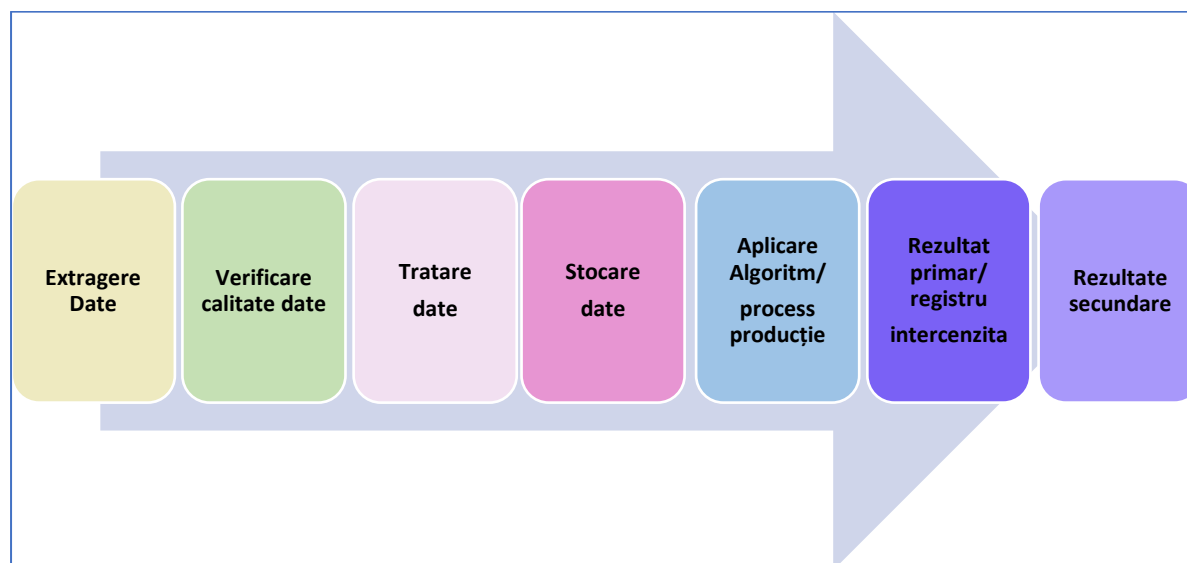
²⁰ Atât unitățile active cât și attributele unităților selectate se referă la 31 decembrie - actualizat cu cele mai recente date disponibile până la momentul extragerii (1 octombrie an N). Populația (persoane/ gospodării) astfel rezultată va fi diferită de cea cu moment de referință 1 ianuarie an N+1 (cea care ar fi de fapt necesară pentru extragerea eșantionului pentru anul N+1) deoarece unele evenimente nu s-au produs încă sau nu au fost înregistrate încă în sistem. „Stocul curent” este astfel doar cea mai bună aproximare posibilă a populației la 1 ianuarie an N+1, pe baza datelor disponibile, la momentul extragerii - adică cu 3 luni înainte de momentul de referință.

3. Implementarea fluxului de realizare a registrului intercenzitar al populației

Registrul populației statistice are la bază fluxul de lucru prezentat în Figura 1 și detaliat pe faze în această secțiune (vezi fluxul integrat în **Anexa 1**). Acesta include mai multe faze, replicabile pentru fiecare sursă de date:

- **Extragerea datelor** - intrare date / input - Conexiune API / integrarea de fișiere pe baza de scheme și designuri tabelare clare (i.e. cerințe, format, stocare),
- **Validarea datelor** (Quality Check, QC) - procesul include verificarea calității datelor, revizuirea variabilelor de QC, actualizarea scripturilor, adăugarea de variabile de QC noi, dacă este cazul.
- **Manipularea datelor** - fiecare sursă de date presupune crearea de scripturi individuale de curățare și manipulare a datelor bazate pe obiectivele/nevoile fiecărei surse de date (i.e. obiectivul, configurare date)
- **Stocarea datelor (de bază)** - Datele de bază sunt stocate într-un format potrivit (e.g. PostgreSQL) pentru fi făcute disponibile pentru analiză și exporturi.
- **Aplicarea algoritmilor/proces de producție** - algoritmi principali (scripturi) sunt aplicați datelor extrase și curățate, se crează / actualizează registrul intercenzitar.
- **Generarea / actualizarea registrului intercenzitar (rezultat principal)** - în urma aplicării algoritmilor se stochează ultima versiune a registrului intercenzitar și se documentează modificările. Rezultatele obținute în procesul de producție determină/compun registrul intercenzitar al populației statistice ale cărui date se vor putea disemina ca **rezultate secundare**.

Figura 1 - Flux de lucru registru intercenzitar



Realizarea efectivă a registrului are la bază scripturi în R pentru fiecare etapă precizată disponibilă pe serverul 'Procesare Date' ca la <http://xx.x.xxx.xx:8081/intercensus> :

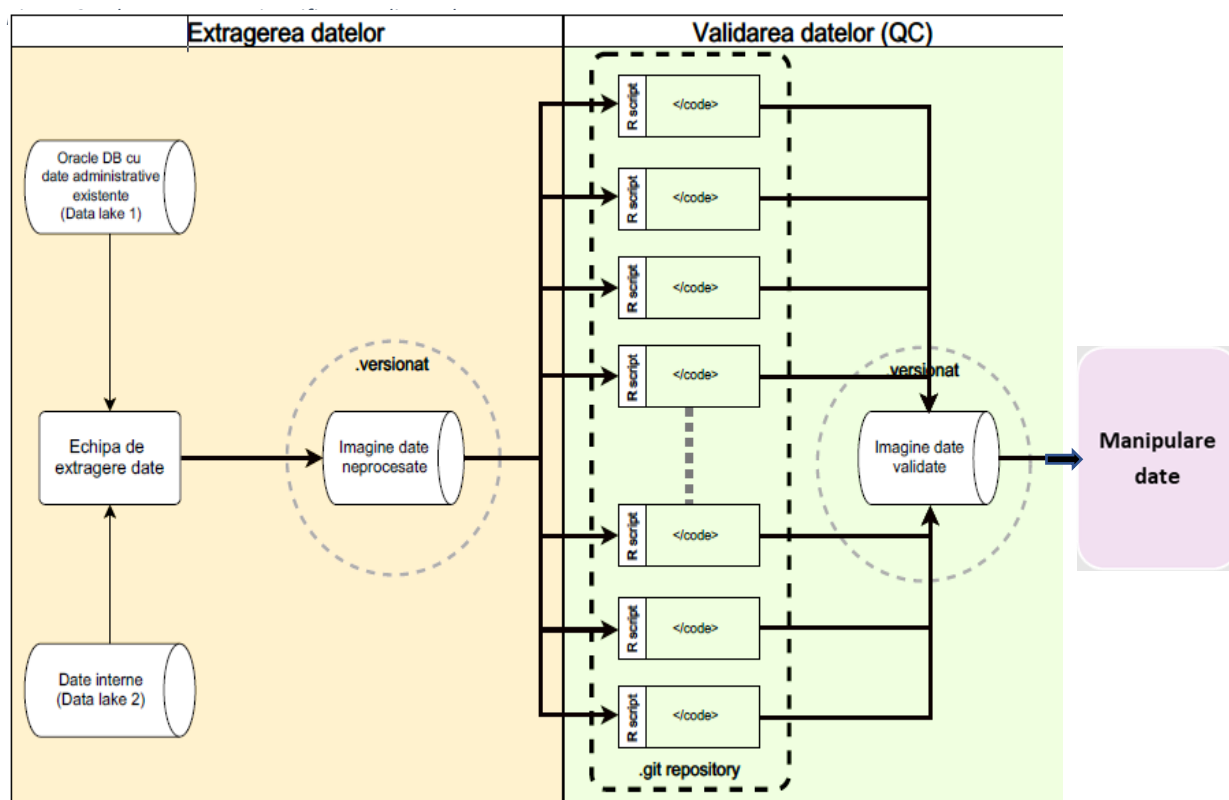
- | | |
|----------------------|-------------------------|
| - 01_Data extraction | - 05_Production_process |
| - 02_Quality_check | - 06_Primary_output |
| - 03_Data_wrangling | - 07_Secondary_output |
| - 04_Data_warehouse | |

Extragerea datelor

Extragerea datelor se realizează de către echipa specializată a INS cuprinzând două etape: prima, care are loc în afara sistemului de actualizare regulată a informației implementată de către specialiștii din Direcția generală IT; a doua, în care are loc preluarea datelor de interes în sistemului de actualizare regulată a informației, reprezentând în fapt începutul procesului subiect al acestui raport și care este descrisă în continuare.

Echipa INS realizează extragerea datelor din surse externe sau interne, realizează verificări minimale referitor la consistența datelor și verifică ca acestea să respecte schemele principale și design-urile tabelare prestabilite (vezi Figura 2). La acest nivel se pot realiza modificări în ceea ce privește schimbarea sursei sau a modului de extragere a datelor. Fiecare sursă de date este stocată separat și este necesară versionarea fiecărei iterații de extragere. Se va păstra un istoric (fișier log) al tuturor încărcărilor, prelucrărilor și validărilor. Versionarea se poate realiza fie prin utilizarea unui sistem dedicat acestui lucru, respectiv un log al modificărilor în structura codului - acestea apar în Gitlab, precum și în datele propriuzise, fie prin simpla datare a fișierelor sursă. Este important de notat faptul că datorită cantității mari de informații extrase, versionarea prin datarea extrasului complet poate deveni problematică în raport cu spațiul de stocare disponibil. Din acest motiv se recomandă utilizarea unui sistem care să poată extrage numai diferențele față de ultima iterație. În pregătirea și implementarea acestei faze, specialiștii INS stabilesc variabilele pe care urmează să le actualizeze în registru coroborat cu cronologia actualizării datelor de intrare pe surse de date și momentul actualizării (data calendaristică), un inventar al surselor secundare și indicatorii de calitate din surse administrative utilizate (vezi **Anexa 2 – Surse de date administrative, indicatori de calitate, variabile, cronologie**).

Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibil pe serverul ‘Procesare Date’ cale http://xx.x.xx.xx:xxx/intercensus/01_Data_extraction



Verificarea calității datelor

După extragerea datelor, acestea trebuie verificate în raport cu obiectivul urmărit și din punct de vedere al corectitudinii acestora în raport cu metadatele care descriu sursa de date (vezi Figura 2). Acest tip de verificare trebuie realizat de o persoană / echipă specializată care să cunoască foarte bine specificitățile sursei de date. În prezent, această verificare nu se poate realiza la nivelul procesului de extragere a datelor, deoarece acesta din urmă este implementat la nivelul departamentului IT, în acest caz nefiind necesară cunoașterea în detaliu a specificității datelor. Este recomandat ca fiecare sursă de date / instituție să fie atribuită unei persoane sau echipe care să cunoască detaliile și specificitățile acestora. De asemenea, trebuie asumată poziția de "lead" / responsabil pentru fiecare sursă de date. Persoana responsabilă alături de echipă, atunci când este cazul, va trebui să realizeze scripturile necesare verificării datelor și să asigure mentenanța acestora. Acest lucru include:

- crearea scripturilor necesare pentru citirea/scrierea datelor (read-in / write-out)
- realizarea de rapoarte de calitate a datelor
- crearea / gestionarea indicatorilor de calitate și extinderea listei de indicatori atunci când este cazul
- realizarea modificărilor necesare pentru asigurarea calității datelor
- documentarea proceselor și a calendarului/cronologiei (time-line) modificărilor.

Se recomandă ca toate scripturile să fie versionate utilizând Git (vezi în **Secțiunea 4** a raportului instrumentele pentru actualizarea datelor) care oferă o serie de avantaje, precum reproductibilitate, trasabilitatea modificărilor, cronologie sau posibilitatea întoarcerii la versiuni precedente.

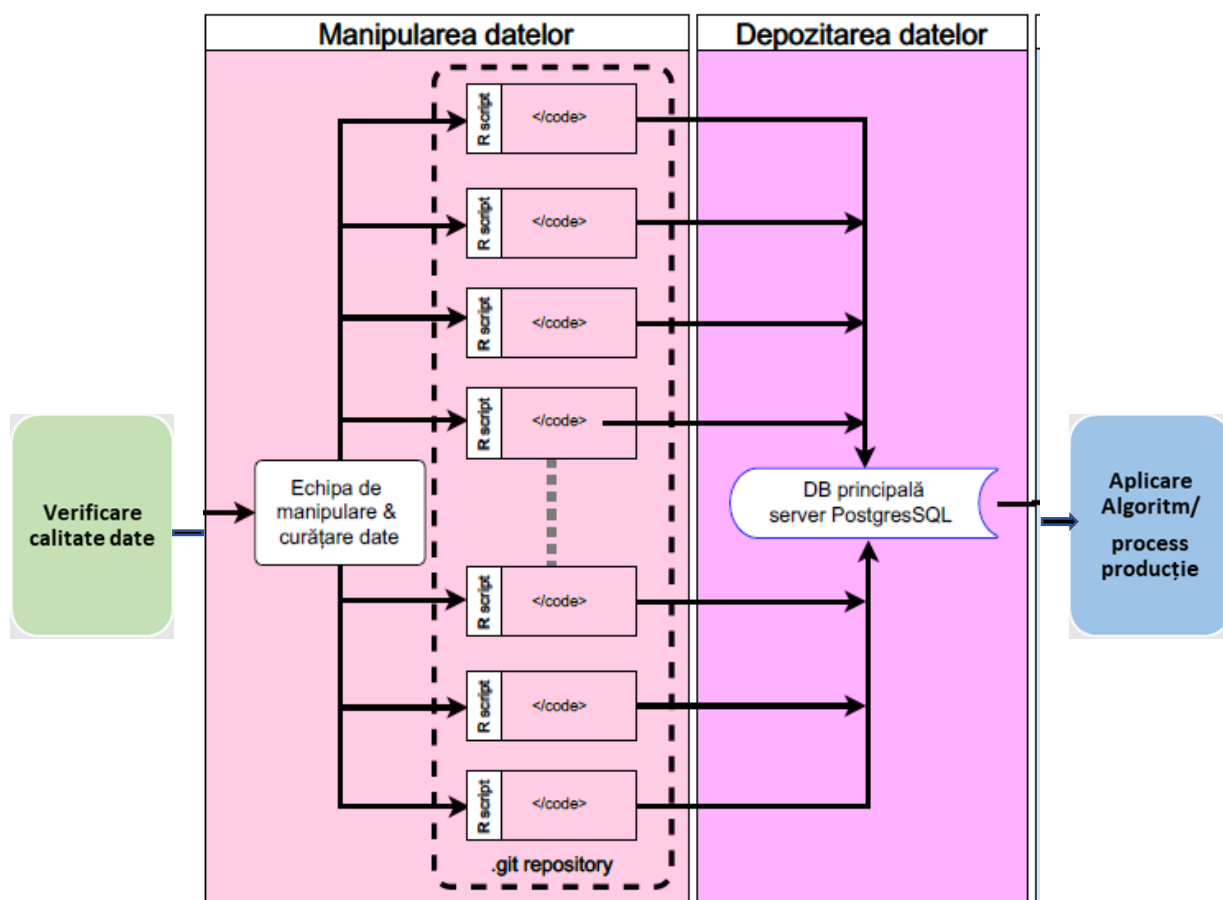
Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibile pe serverul 'Procesare Date' cale http://xx.x.xx.xx:xxx//intercensus/02_Quality_check

Tratarea / Manipularea datelor

Ulterior verificării datelor, acestea trebuie reconfigurate / transformate astfel încât să poată deservi obiectivul final și rezultatele așteptate (vezi Figura 3). Acest proces poate include: reconfigurarea coloanelor și rândurilor, standardizarea valorilor în funcție de nevoi, transformarea variabilelor, etc. Și în acest caz este necesar ca persoana / echipa să cunoască atât sursa de date (pentru a putea asigura o bună interpretare a acestora) cât și obiectivul final al procesului. Persoanele responsabile vor crea/ întreține seturi de scripturi necesare transformării / manipulării datelor și vor asigura mentenanța acestora, fie prin optimizări ocazionale fie prin adăugarea de transformări sau utilități noi. Este important ca aceste scripturi să fie versionate (de exemplu utilizând git) și documentate pentru ca transferul în echipă și între echipe să fie facil. De asemenea, este important ca datele rezultate să fie și acestea versionate utilizând o opțiune potrivită pentru dimensiunea datelor rezultate (i.e. versionare cu dată pentru fișiere până în 1GB, versionare folosind Git LFS pentru fișiere cu dimensiune mai mare de 1GB).

Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibile pe serverul 'Procesare Date' cale http://xx.x.xx.xx:xxx//intercensus/03_Data_wrangling

Figura 3 – Fluxuri tratare și stocare date



Stocarea datelor de bază

Datele rezultate în urma tratamentelor și manipulării datelor, seturile de date rezultate se stochează într-un format unic (final) alături de metadatele necesare, fiind astfel pregătite pentru interogări în procesele ulterioare (vezi Figura 3). Datele pot fi stocate ca fișiere individuale dateate/versionate folosind Git și Git LFS, însă se recomandă stocarea acestora într-o bază de date PostgreSQL (deja instalat pe serverul de analiză). Unul dintre avantajele acestei metode este creșterea vitezei de citire/scriere în cazul seturilor mari de date. De asemenea, crearea unei baze de date facilitează crearea unui punct unic de acces pentru seturile de date necesare pentru analiză, registrul intercenzitar și exporturi. Din punct de vedere al metodei de lucru, prin intermediul pachetelor adiționale dedicate conexiunilor SQL , e.g. DBI, RPostgreSQL, se pot realiza foarte facil conexiuni de citire scriere către o bază de date (exemplu in Figura 4).

Figura 4 - Conexiuni citire/scriere baza de date

```

library(DBI)

con <- dbConnect(RPostgres::Postgres(),
  host = "localhost",
  port = "5555",
  dbname = "some_db",
  user = "user",
  password = "password")

```

În cazul în care se optează pentru stocarea datelor în fișiere separate datate, se recomandă utilizarea formatului *.parquet*. Acest format permite utilizarea mai eficientă a spațiului și o mai mare eficiență în execuția codului în cazul fișierelor de mari dimensiuni. Formatul *.parquet* stochează datele în format columnar, se poate utiliza în orice sistem de operare, poate stoca metadate și poate partiționa fișierele. În plus, aplicarea query-uri asupra datelor în acest format nu sunt evaluate până la comanda explicită a acestui proces. În R, lucrul cu fișiere *.parquet* este facilitat de pachetele {arrow} și {dplyr}. Pentru informații suplimentare se poate consulta manualul {arrow}: <https://arrow.apache.org/docs/r/index.html>.

Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibile pe serverul 'Procesare Date' la http://xx.x.xx.xx:xxx/intercensus/04_Data_warehouse

Aplicarea algoritmilor – proces de producție

În cazul fiecărei surse și set de date se aplică o serie de algoritmi principali (scripturi) în limbajul de programare R disponibile pe serverul 'Procesare Date' la http://xx.x.xx.xx:xxx/intercensus/05_Production_process. Algoritmii se aplică asupra datelor extrase și curățate, iar apoi se actualizează registrul intercenzitar și se stochează ultima versiune a bazei de date (vezi Figura 5). Structura registrului intercenzitar (schemele) bazei de date este prezentată în **Anexa 3 - Structură Registrul Intercenzitar** la prezentul raport sau la <http://xx.x.xx.xx:xxx/intercensus/intercensus-documentations/#intercensus-registry-schemas>, 'Server procesare date' al INS. Aplicarea algoritmilor se face conform necesităților INS și a standardelor prestabilite.

Este recomandat ca scripturile să implementeze algoritmii pe baza a 4 pași principali:

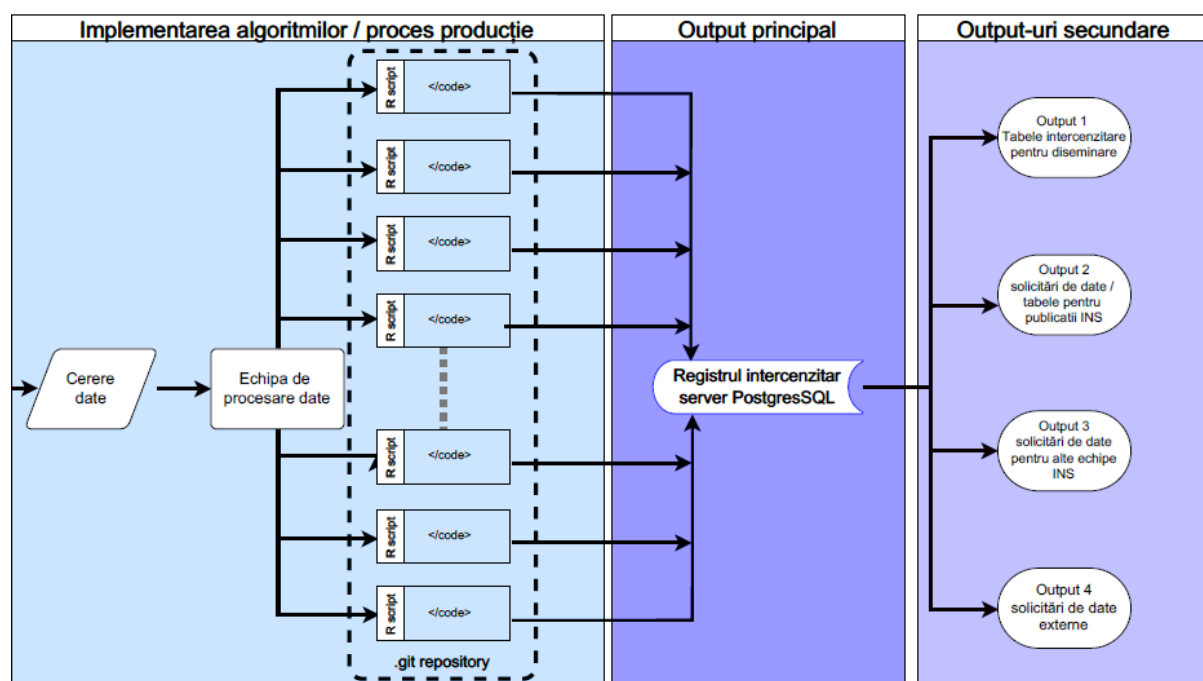
- solicitare date din DB principală;
- aplicarea algoritmului;
- testarea output-urilor;
- stocarea output-urilor în baza de date intercenzitară.

Registrul intercenzitar (rezultat principal)

Registrul intercenzitar este rezultatul principal al procesului descris mai sus (vezi Figura 5). În acesta se stochează toate datele și metadatele necesare. Este recomandată stabilirea unei persoane sau a unei echipe în funcția de custode principal al registrului intercenzitar. Custodele trebuie să stăpânească principalele scripturi în urma cărora rezultă generarea și actualizarea registrului. De asemenea persoanele care se ocupă cu mentenanța registrului trebuie să stăpânească limbajul de programare R, tipologia mediilor de stocare utilizate, limbajul SQL și să poată realiza acțiunile administrative necesare asupra bazei de date. Pentru stocarea datelor din registrul intercenzitar se recomandă utilizarea unei baze de date PostgreSQL, așa cum a fost deja menționat, acest lucru poate fi foarte ușor integrat în scripturi și procese prin intermediul limbajului R și a pachetelor suplimentare. Registrul intercenzitar trebuie versionat pentru a se stoca un număr prestabilit de versiuni și trebuie create backup-uri off-site pentru asigurarea redundanței bazei de date în cazul unor accidente sau incidente care pot afecta infrastructura.

Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibile pe serverul 'Procesare Date' la http://xx.x.xx.xx:xxx/intercensus/06_Primary_output

Figura 5 - Fluxuri aplicare algoritm, producție rezultat primar și rezultate secundare



Rezultate secundare

Rezultatele obținute în procesul de producție determină/compun registrul intercenzitar al populației statistice ale carui date se vor putea disemina ca rezultate secundare. Datele de bază vor fi făcute disponibile pentru diseminare, interogări de date, tabele pentru publicațiile INS, cereri de date pentru alte echipe INS, solicitări de date de la terți (vezi Figura 5).

Pentru realizarea acestor procese se vor stabili persoane / echipe responsabile pentru acest lucru, care vor centraliza cererile și le vor redirecționa către persoanele potrivite spre soluționare. Se recomandă ca în cazul deschiderii accesului spre terți prin intermediul solicitărilor de date, acestea să se realizeze standardizat utilizând formulare în format electronic pentru a limita variabilitatea cererilor.

Intregul proces se realizează prin aplicarea de scripturi în limbajul de programare R disponibile pe serverul 'Procesare Date' la http://1xx.x.xx.xx:xxx//intercensus/07_Secondary_output

4. Instrumente pentru actualizarea datelor în registrul populației

Actualizarea datelor se realizează respectând metodologia INS, a fluxului de lucru și principiilor prezentate mai sus.

Pentru o abordare coerentă și pentru a asigura consistența datelor se recomandă:

- stabilirea de persoane / echipe responsabile pentru implementare / actualizare;
- documentarea proceselor, funcțiilor și a codului utilizat pentru a facilita transferul ușor între membrii echipelor;
- respectarea pașilor de actualizare a datelor și a variabile de actualizat, i.e. a se vedea *Anexa 2 - Surse de date administrative, indicatori de calitate, variabile, cronologie*;
- respectarea fluxurilor de proces pentru fiecare sursă de date / instituție și documentarea acestora, incluzând cronologia modificărilor și versionarea utilizând instrumentele potrivite, e.g. git.
- utilizarea consistentă a aceluiași limbaj de programare - R - și documentarea codului și a funcțiilor suplimentare.

Pentru implementarea proceselor aferente perioadei intercenzitare se va utiliza infrastructura existentă în cadrul INS. Astfel, serverul de analiză / procesare date este principalul instrument de lucru. . Serverul este instalat în incinta INS și poate fi accesat numai din interiorul instituției. Adresa IP de acces din rețeaua INS este xx.x.xxx.xx.

Pentru buna funcționare a întregului flux sunt utilizate următoarele instrumente, instalate în prealabil pe serverul de analiză:

1. RStudio Server - mediul principal de lucru
2. Gitlab (alături de Git LFS)- sistem de versionare
3. PostgreSQL - bază de date

Pentru instalarea sistemului de operare și a softwarerului necesar se poate consulta raportul Rezultatului nr. 4.2a. Serverul RStudio este disponibil la adresa <http://xx.x.xx.xx:xxx/>. Fiecărui membru al echipei i-a fost creat un user și i-au fost transmise credențialele de acces de către echipa IT a INS. RStudio Server reprezintă principalul mediu de lucru pentru perioada intercenzitară, utilizatorii având acces la toate resursele disponibile la nivelul sistemului.

Serverul Gitlab a fost instalat pentru a facilita lucrul în echipă și a reduce conflictele în timpul implementării. Serverul Gitlab se accesează la adresa: <http://xx.x.xx.xx:xxx/>. Aici toți membrii echipelor intercenzitare își vor crea useri și se va realiza o ierahie de grupuri și subgrupuri care să asigura buna colaborare în implementarea fluxurilor descrise mai sus. Repository-ul principal al serverului de Gitlab se regăsește pe serverul de analiză în directorul **/arhiva1/gitlabrep**. Se recomandă ca utilizatorii să nu își creeze directoarele de lucru în directorul /home aferent ci în directorul **/arhiva2/intercensus/nume_utilizator**. Acest aspect este recomandat, deoarece lucrul în perioada intercenzitară include seturi de date foarte mari, iar directoarele /home se regăsesc pe partiția principală a serverului, care are o dimensiune mai redusă. Ulterior creării directoarelor de lucru, utilizatorii vor clona repository-urile de pe serverul Gitlab.

Întreg procesul de conectare, clonare și lucru cu repositoryurile git a fost descris în detaliu în cadrul sesiunii de instruire. Acesta este disponibilă aici: https://worldbankgroup.webex.com/webappng/sites/worldbankgroup/meeting/postinfo/0f3e850fe6f74bf581c468b298aaca71_I_248559732106732301?tab=recordings.

De asemenea, pe serverul Gitlab se regăsește un repository care include un tutorial de bază care poate fi dezvoltat ulterior: <http://xx.x.xx.xx:xxx//aghita/gitlab-rstudio-tutorial>.

Pentru detalii suplimentare, descrierea comenzilor git utilizate prin intermediul consolei se pot consulta:

- <https://git-scm.com/docs/git>
- <https://training.github.com/downloads/github-git-cheat-sheet/>

Serverul PostgreSQL este instalat de asemenea pe serverul de analiză. Pentru conectarea prin intermediul scripturilor acesta poate fi accesat la adresa IP locală a serverului utilizând pachetele DBI sau RPostgreSQL. Pentru realizarea acțiunilor administrative serverul poate fi accesat de asemenea prin intermediul unor software-uri suplimentare, e.g. pgAdmin.

La dispoziția specialiștilor INS este disponibilă o bibliotecă cu înregistrările întâlnirilor de lucru privind procesul de stabilire a fluxurilor de lucru din registrul intercenizar:

- Noiembrie 8, 2022
<https://worldbankgroup.webex.com/recordingservice/sites/worldbankgroup/recording/a5ef9fe04193103bbbfe00505681c6e0/playback>
- Noiembrie 15, 2022
<https://worldbankgroup.webex.com/webappng/sites/worldbankgroup/recording/c9e690644713103b9bff0050568195cf/playback>
- Noiembrie 22, 2022
<https://worldbankgroup.webex.com/recordingservice/sites/worldbankgroup/recording/c9e690644713103b9bff0050568195cf/playback>
- Noiembrie 29, 2022
<https://worldbankgroup.webex.com/webappng/sites/worldbankgroup/recording/204facf51eb103b9b5f005056816ca1/playback>
- Decembrie 13, 2022
<https://worldbankgroup.webex.com/webappng/sites/worldbankgroup/recording/c24b1ea15d14103bbb900505681f6c7/playback>
- Ianuarie 10, 2023 (GitLab training)
<https://worldbankgroup.webex.com/recordingservice/sites/worldbankgroup/recording/7fa65d267315103b991f00505681a659/playback>
- Ianuarie 17, 2022
<https://worldbankgroup.webex.com/recordingservice/sites/worldbankgroup/recording/a6fd6c57787d103b9fee005056818d2f/playback>

Această bibliotecă este completată cu înregistrările întâlnirilor de lucru pentru dezvoltarea scripturilor pentru fazele fluxului de dezvoltare a registrului, detaliate în Secțiunea 3 a raportului, disponibile pe „Serverul de procesare Date” <http://xx.x.xx.xx:xxx//interensus/interensus-documentation#training-materials>.

Totodată, se recomandă parcurgerea informațiilor privind metodologii și instrumente dedicate producției statisticilor în perioadele intercenizare disponibile în rapoartele Rezultatului 6:

- Rezultatul 6d: *Raport privind serviciile de asistență tehnică oferite Beneficiarului pentru Metodologia pentru o evaluare continuă și în timp real a calității datelor colectate din surse administrative și geo-spațiale pentru perioadele intercenizare.*
- Rezultat 6b: *Raport privind serviciile de asistență tehnică oferite Beneficiarului privind Metodologia referitoare la metodele de eșantionare la trecerea la un alt tip de cadru*

de eşantion (de exemplu, de la eşantion de bază (eşantion master) la cadrul bazat pe registre) pentru anchete şi statistici în perioadele intercenzitare

- *Rezultat 6c: Raport privind serviciile de asistenţă tehnică oferite Beneficiarului privind Metodologia de evaluare a efectelor modificării metodei de colectare a datelor, în contextul trecerii la metode multimodale (CAPI/CAWI), pentru sondaje şi statistici în perioadele intercenzitare*
- *Rezultat 6a: Raport privind serviciile de asistenţă tehnică oferite Beneficiarului pentru recomandări şi bune practici pentru implementarea unui sistem documentat de modernizare a metodelor de colectare a datelor pentru anchetele statistice efectuate în perioadele de colectare intercenzitare (CAWI/CAPI) pentru anchetele prin sondaj la scară largă*

Anexe

Anexa 1: Flux de lucru al registrului intercenzitar al populației

Anexa 2: Surse de date administrative, indicatori de calitate, variabile, cronologie

Anexa 3: Structură registru intercenzitar

Anexa 1: Fluxul de lucru al registrului intercenzitar al populației

Anexa este arhivată și livrată în format electronic la prezentul raport



Output 4.2b - Annex
1 - Intercensus_registri

Anexa 2: Surse de date administrative, indicatori de calitate, variabile, cronologie

Anexa este arhivată și livrată în format electronic la prezentul raport și conține următoarele fișiere:



01_Inventar_surse_se
cundare.xlsx



02_Extras indicatori
de calitate.xlsx



03_Variabile de
actualizat in registru.x



04_Cronologie.xlsx

Anexa 3: Structură Registru intercenzitar

Anexa este arhivată și livrată în format electronic la prezentul raport.

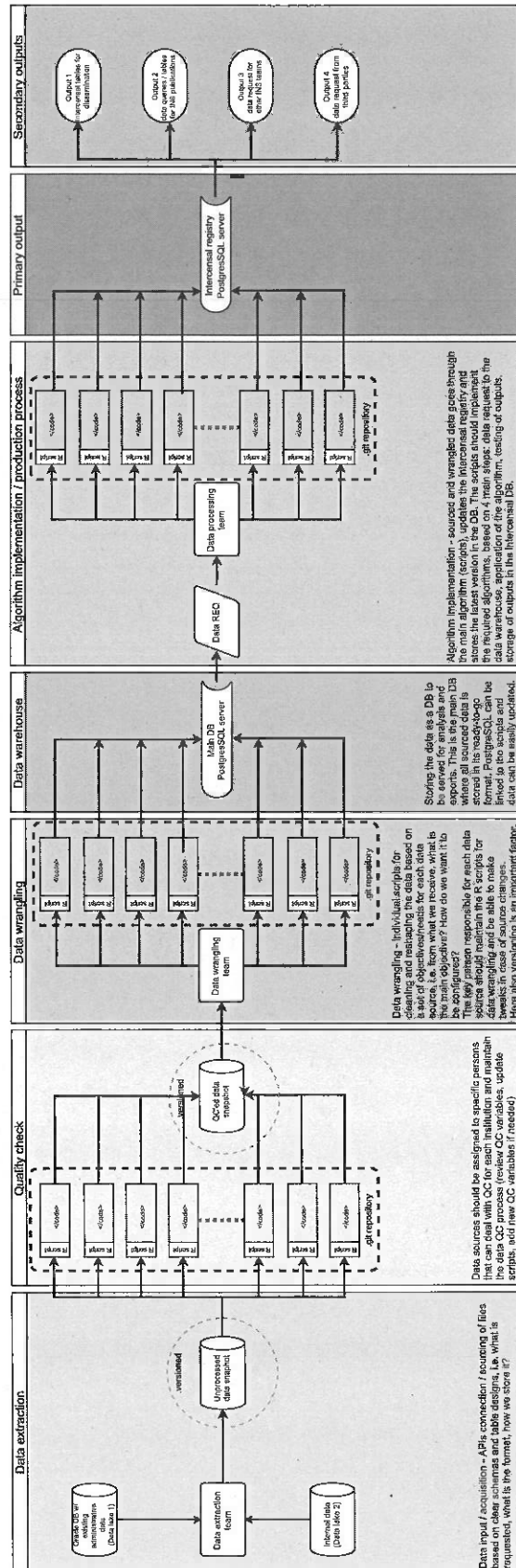


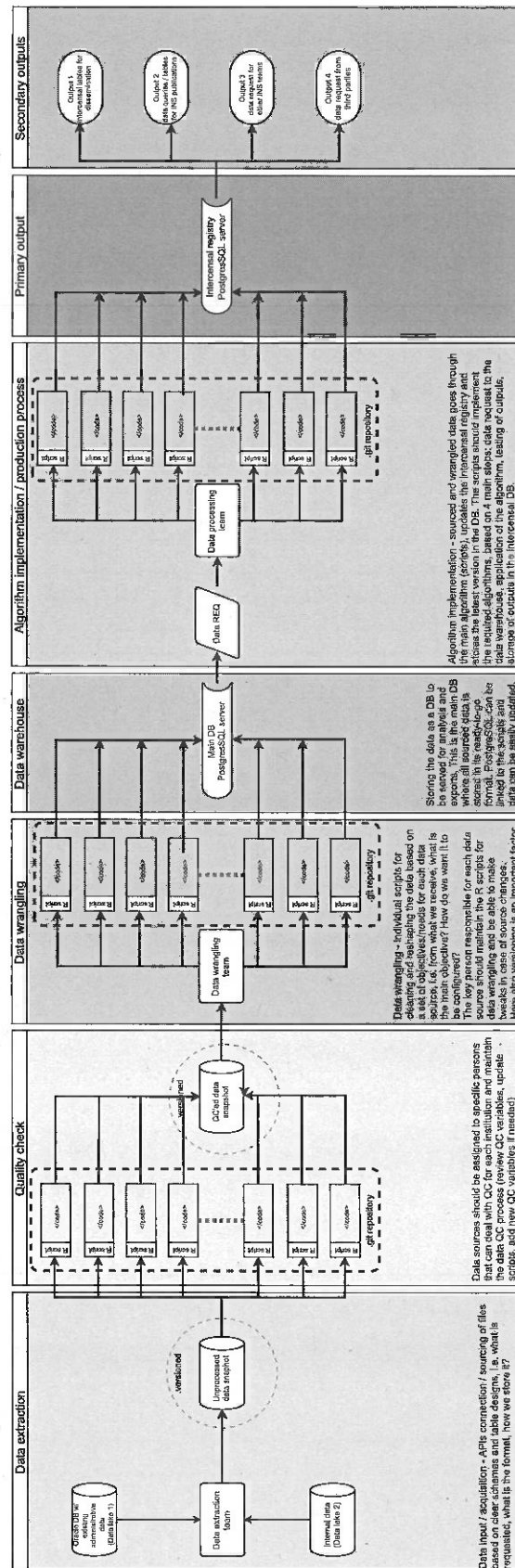
Output 4.2b - Anexa
3 - Structura Registru



Competența face diferența!

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





Inventar surse

Denumire sursa	In algoritm	Date disponibile in mod curent	Comentarii	In afara de datele primite cu regularitate, pentru RPL s-au primit, o singura data, date in baza OUG19/2020
INS-Nascuti vii	Da	Da	Datele nu au adresa exacta	
INS-Decese	Da	Da	Datele nu au adresa exacta	
INS-Casatorii	Da	Da	Datele nu au adresa exacta	
INS-Divorturi	Da	Da	Datele nu au adresa exacta	
DPABD-Stabiliri de domiciliu	Da	Da	Datele nu au CNP si adresa adresa exacta deci nu pot fi folosite momentan	
DPABD-Anunțări de reședință	Da	Da	Datele nu au CNP si adresa adresa exacta deci nu pot fi folosite momentan	
DPABD-RNEP	Da	Nu	Nu exista conventie semnata; datele s-au primit doar in 2014 si mai nou pentru RPL 2021	
DPABD- migratie externa	Da	Da	Datele nu au CNP si adresa adresa exacta deci nu pot fi folosite momentan	
DGP-migratie externa	Da	Da	Datele nu au CNP si adresa adresa exacta deci nu pot fi folosite momentan	
IGI-migratie externa	Da	Da	Datele nu au CNP si adresa adresa exacta deci nu pot fi folosite momentan	
ANAF-D112	Da	Da		
ANAF-Depozit	Da	Da		
CNPP	Da	Da		
SIIR	Da	Da	Date primite pentru testare	
RMU	Da	Da	Date primite pentru testare	
ANPIS	Da	Da		
IM	Nu	Da		

APIA	Da	Da	Datele pe CD la directia de IT	
ANDPCCA	Da	Nu		
ANOFM	Da	Nu		
CNAS	Da	Nu	Nu exista conventie semnata	
ANPD	Da	Nu		

Date pe care se poate lucra. Restul, fie nu sunt disponibile, fie le lipsesc elemente esentiale fara de care nu se pot face legaturi la nivel individual.

Nume utilizat	ANAF - D112	ANAF - Depozit	ANAF - Depozit	CNPP	ANPIS	ITM	SIIR	RMUR
Schema	SRS_ANAFD112	SRS_ANAFDEPO	SRS_ANAFDEPO	SRS_CNPP	SRS_ANPIS	SRS_ITM	SRS_SIIR	SRS_MECS
Tabel	CONTRIB_ASIG_D112	PFDW_D205_aa	PFDW_DECO_aaaa	CNPP_INS_Ilaa	ANPIS_aaaall	ITM_Ilaa	SIIR_2020_2021	RMUR_2021_FAC
Statistici								
Numar inregistrari								
Numar inregistrari perioada curenta/Numar inregistrari perioada precedenta								
Numar CNP-uri unice								
Numar CNP-uri unice perioada curenta/Numar CNP-uri unice perioada precedenta								
5.2.1 Corespondenta cu structura asteptata cf. metadatelor								
Numar (%) campuri in fisier care nu ar trebui sa existe cf. structurii asteptate								
Lista campuri in fisier care nu ar trebui sa existe cf. structurii asteptate								
Numar (%) campuri in fisier care nu exista desi ar trebuie cf. structurii asteptate								
Lista campuri in fisier care nu exista desi ar trebuie cf. structurii asteptate								
5.3.1 Supra-acoperire - duplicate								
Numar inregistrari duble - adica toate campurile identice								
Numar aparitii ale combinatiilor (adica cate combinatii au o singura aparitie, cate combinatii au doua aparitii etc); in principiu toate combinatiile ar trebui sa apara o singura data (daca am identificat corect campurile si deci "semnificatia" unei inregistrari)								
	CNP_CRIPT, TIP_ASIGURAT, SECTIUNE_DECLARAT IE, CONTRIBUABIL_ID, LUNA_RAPORTARE, AN_RAPORTARE	CNP_CRIPT, MARCAJ_SPECIFIC IC, CATEGORIE_VE NIT, AN_FISCAL	AN_FISCAL, CIF_PLATITOR, CATEGORIE_VENIT, BAZA_CALCUL_IM POZIT, IMPOZIT, CNP_CRIPT	CNP_CRIPT, CFP, DOSAR	CNP_BENEFICIAR R_CRIPT, PRESTATIE_ID, DOSAR_ID, AN_LUNA_DREPT	CNP_CRIPT, TIP_SPOR, SALARIU_DE_BAZA, DATA_SFARSIT_DETASARE_CONTRACT, DATA_INCEPUT_DETASARE_CONTRACT, COD_COR_CONTRACT, STARE_CONTRACT, TIP_NORMA, DURATA_CONTRACTULUI, TIPUL_CONTRACTULUI, DATA_CONTRACT, NUMAR_CONTRACT, CUI_ANGAJATOR	"an_studiu_final", "situatie_scolara_final_an", "operatie_final_an", "an_studiu_inceput_an", "situatie_scolara_inceput_an", "operatie_inceput_an", "program_studiu", "an_calendaristic", "strain_codpp", "strain_codci", "strain_codssn", SHA1	

5.3.5 Non-raspuns la nivel de unitate

Numarul (%) de inregistrari fara date pentru niciuna dintre variabilele esentiale*

CNP_CRIPT, TIP_ASIGURAT, SECTIUNE_DECLARAT_IC, IE, CONTRIBUABIL_ID, LUNA_RAPORTARE, AN_RAPORTARE	CNP_CRIPT, MARCAI_SPECIFIC, AN_FISCAL, CIE_PLATTOR, CATEGORIE_VENIT, CATEGORIE_VE_NIT, AN_FISCAL, BAZA_CALCUL_IMPOZIT, CNP_CRIPT, LUNA_RAPORTARE	CNP_CRIPT, TCFP, CFP, DOSAR, MODPL	CNP_BENEFICIA, R_CRIPT, PRESTATIE_ID, DOSAR_ID, AN_LUNA_DREP	CNP_CRIPT, TIP_SPOR, SALARIU_DE_BAZA, DATA_SFARSIT_DETASARE_CONTRACT, DATA_INCEPUT_DETASARE_CONTRACT, COD_COR_CONTRACT, STARE_CONTRACT, TIP_NORMA, DURATA_CONTRACTULUI, TIPUL_CONTRACTULUI, DATA_CONTRACT, NUMAR_CONTRACT, CUI_ANGAJATOR, CETATENIE_SALARIAT, CONTRACT_RADIAT	ISU, CETAT, TARA, "an_studiu_final", "situatie_scolara", "final_an", "operatie_final_an", "an_studiu_inceput_an", "situatie_scolara_inceput_an", "operatie_inceput_an", "program_studiu", "an_calendaristic", "strain_codpp", "stain_cocdi", "strain_codssn", SHA1, cetatenie,
--	--	------------------------------------	--	---	--

5.3.7 Non-raspuns la nivel de variabila Numarul (%) de inregistrari fara date pentru variabila esentiala*

CNP_CRIPT, TIP_ASIGURAT, SECTIUNE_DECLARAT_IC, IE, CONTRIBUABIL_ID, LUNA_RAPORTARE, AN_RAPORTARE	CNP_CRIPT, MARCAI_SPECIFIC, AN_FISCAL, CIE_PLATTOR, CATEGORIE_VENIT, CATEGORIE_VE_NIT, AN_FISCAL, BAZA_CALCUL_IMPOZIT, CNP_CRIPT, LUNA_RAPORTARE	CNP_CRIPT, TCFP, CFP, DOSAR, MODPL	CNP_BENEFICIA, R_CRIPT, PRESTATIE_ID, DOSAR_ID, AN_LUNA_DREP	CNP_CRIPT, TIP_SPOR, SALARIU_DE_BAZA, DATA_SFARSIT_DETASARE_CONTRACT, DATA_INCEPUT_DETASARE_CONTRACT, COD_COR_CONTRACT, STARE_CONTRACT, TIP_NORMA, DURATA_CONTRACTULUI, TIPUL_CONTRACTULUI, DATA_CONTRACT, NUMAR_CONTRACT, CUI_ANGAJATOR, CETATENIE_SALARIAT, CONTRACT_RADIAT	ISU, CETAT, TARA, "an_studiu_final", "situatie_scolara", "final_an", "operatie_final_an", "an_studiu_inceput_an", "situatie_scolara_inceput_an", "operatie_inceput_an", "program_studiu", "an_calendaristic", "strain_codpp", "stain_cocdi", "strain_codssn", SHA1, cetatenie,
--	--	------------------------------------	--	---	--

5.4.1 Variabile cu valori inconsistente

Numar (%) inregistrari care au si valori in afara domeniului asteptat
(vezi descriere si nomenclatoare)*

CNP_CRIPT, TIP_ASIGURAT, SECTIUNE_DECLARAT IE, CONTRIBUABIL_ID, LUNA_RAPORTARE, AN_RAPORTARE	CNP_CRIPT, MARCAI_SPECIF IC, CATEGORIE_VE NIT, AN_FISCAL, LUNA	AN_FISCAL, CIF_PLATITOR, CATEGORIE_VENIT, BAZA_CALCUL_IM POZIT, IMPOZIT, CNP_CRIPT, LUNA	CNP_CRIPT, TCFP, CFP, DOSAR, MODPL	CNP_BENEFICIA R_CRIPT, PRESTATIE_ID, DOSAR_ID, AN_LUNA_DREP T	CNP_CRIPT, TIP_SPOR, SALARIU_DE_BAZA, DATA_SFARSIT_DETASARE_ CONTRACT, DATA_INCEPUT_DETASARE_ CONTRACT, COD_COR_CONTRACT, STARE_CONTRACT, TIP_NORMA, DURATA_CONTRACTULUI, TIPUL_CONTRACTULUI, DATA_CONTRACT, NUMAR_CONTRACT, CUI_ANGAJATOR, CETATENIE_SALARIAT, CONTRACT_RADIAT	ISU, CETAT, TARA, ", "situatie_scolara _final_an", "operatie_final_a n", "an_studiu_ince put_an", "situatie_scolara _inceput_an", "operatie_incepu t_an", "program_studiu ", "an_calendaristic ", "strain_codpp", "stain_codci", "strain_codssn", SHA1, cetatenie, tara_nastere
--	---	---	--	--	--	---

Lista inregistrari care au si valori in afara domeniului asteptat (vezi
descriere si nomenclatoare)*

CNP_CRIPT, TIP_ASIGURAT, SECTIUNE_DECLARAT IE, CONTRIBUABIL_ID, LUNA_RAPORTARE, AN_RAPORTARE	CNP_CRIPT, MARCAI_SPECIF IC, CATEGORIE_VE NIT, AN_FISCAL, LUNA	AN_FISCAL, CIF_PLATITOR, CATEGORIE_VENIT, BAZA_CALCUL_IM POZIT, IMPOZIT, CNP_CRIPT, LUNA	CNP_CRIPT, TCFP, CFP, DOSAR, MODPL	CNP_BENEFICIA R_CRIPT, PRESTATIE_ID, DOSAR_ID, AN_LUNA_DREP T	CNP_CRIPT, TIP_SPOR, SALARIU_DE_BAZA, DATA_SFARSIT_DETASARE_ CONTRACT, DATA_INCEPUT_DETASARE_ CONTRACT, COD_COR_CONTRACT, STARE_CONTRACT, TIP_NORMA, DURATA_CONTRACTULUI, TIPUL_CONTRACTULUI, DATA_CONTRACT, NUMAR_CONTRACT, CUI_ANGAJATOR, CETATENIE_SALARIAT, CONTRACT_RADIAT	ISU, CETAT, TARA, ", "situatie_scolara _final_an", "operatie_final_a n", "an_studiu_ince put_an", "situatie_scolara _inceput_an", "operatie_incepu t_an", "program_studiu ", "an_calendaristic ", "strain_codpp", "stain_codci", "strain_codssn", SHA1, cetatenie, tara_nastere
--	---	---	--	--	--	---

* variabilele esențiale sunt cele precizate în coloana ofertei fiecărei surse

Actualizarea datelor din registru

Variabila	Fenomen/ Baza de date	Camp din baza de date	Observatii
Sex	RNEP	SEX	
	Nascuti-vii	SEX	
	Imigranti	SEX	
Data nasterii	RNEP	DATA_NASTERE	
	Nascuti-vii	DATAN	
	Imigranti	DATAN	
Tara de naștere	RNEP	LOC_NASTERE	
	Nascuti-vii	TARAN	
	Imigranti	TARAN	
Tara de cetățenie	RNEP	CET	
	RPL	CET	
	Nascuti-vii	CET	
	Imigranti	CET	
	Casatorii	CET_M si CET_F	
	Divort	CET_M si CET_F	
Starea civila legală	Nascuti-vii	se trece NECASATORIT cod 1	
	Imigranti	STCIV	
	Casatorii	se trece CASATORIT cod 2 pt cele 2 persoane	
	Divort	se trece DIVORTAT cod 3 pt cele 2 persoane	

Deces	se trece VADUV cod 4 pentru sotul/sotia supravietuitoare (in caz	Nu exista CNP sotul/sotia supravietuitoare, doar data nasterii. (de verificat cu familia din baza RPL)
-------	---	---

Nivelul de instruire

RPL	SCA	
RNEP	NIVEL_EDUCATIE	
Nascuti-vii	se trece FARA EDUCATIE.	
Imigranti	~^d^o	
Casatorii	SCOALA_M si SCOALA_F	
Divort	SCOALA_M si SCOALA_F	
Anuntari de resedinta	-	
Schimbari de domiciliu	SCOALA	
SIIR	NIVEL_INVATAMANT	nivelul de instruire va fi cel anterior NIVEL_INVATAMANT. (ex. Elev in gimnaziu, nivel de instruire este primar)

RMU	IDCicludestudiu	nivelul de instruire va fi cel anterior IDCicludestudiu. (ex. Persoana la ciclu de licenta, nivel de instruire este liceal, Persoana la master, nivel este - licenta). Exceptii pot exista in cazul persoanelor care urmeaza nivel de studiu similar cu cel absolvit (a doua facultate, inca un program de master...)
-----	-----------------	---

Statutul activității curente

(nomenclator CAS pt EUROSTAT)

RPL	lucrat
Nascuti-vii	cod 2
Imigranti	-
Casatorii	ACTIV_M si ACTIV_F
Divort	ACTIV_M si ACTIV_F
Anuntari de resedinta	-
Schimbari de domiciliu	ACTIV

Statut profesional	RPL	STAP
--------------------	-----	------

D112

Activitatea economică

RPL

ACT

Nascuti-vii

cod 12

(nomenclator IND pt EUROSTAT)

ITM

COD_CAEN

Ocupație

RPL

OCUP

RNEP

OCUPATIE

Nascuti-vii

-

Imigranti

OCUP

Casatorii

-

Divort

-

Anuntari de resedinta

-

Schimbari de domiciliu

-

Data stabilirii ultimei reședințe în RPL

AA_URR, LL_URR

România

RNEP

DATA_STABILIRE_RESEDINTA, DATA_EXPIRARE_RESEDINTA

Nascuti-vii

data nasterii - DATAN

Imigranti

DATE

Casatorii

-

Divort

-

Anuntari de resedinta

AN, LUNA

Schimbari de domiciliu

DATAM

Adresa de reședință completă	RNEP	CLA_RESEDINTA	campuri suplimentare: CS_RESEDINTA, ATS_RESEDINTA, DS_RESEDINTA, NI1_RESEDINTA, LI1_RESEDINTA, NI2_RESEDINTA, LI2_RESEDINTA, B_RESEDINTA, S_RESEDINTA, E_RESEDINTA, A_RESEDINTA
	Nascuti-vii	JUDD_MR, LOCAD_MR. Daca mama nu are reședință în România, se ia reședință tatălui	NU exista adresa completă
	Imigranți	JUD, SIRUTA	NU exista adresa completă
	Anunțări de reședință	JUDDS, LOCADS	NU exista adresa completă
	Schimbări de domiciliu	JUDV, LOCAY	NU exista adresa completă
Adresa de domiciliu	RNEP	CL_DOMICILIU	campuri suplimentare: CS_DOMICILIU, ATS_DOMICILIU, DS_DOMICILIU, NI1_DOMICILIU, LI1_DOMICILIU, NI2_DOMICILIU, LI2_DOMICILIU, B_DOMICILIU, S_DOMICILIU, E_DOMICILIU, A_DOMICILIU
Gospodării	Nascuti-vii	JUDD_M, LOCAD_M. Daca mama nu are domiciliul în România, se ia domiciliul tatălui	NU exista adresa completă
ID-ul gospodăriei constituire gospodărie prin:	RPL		
	Nastere	CNP_CRIPT	se adaugă născutul în ID_gospodărie
	Imigranți	-	
	Casătorii	CNP_M, CNP_F	ID_gospodărie nou

Divort

se dizolva gospodaria. Personele formează gospodarii individuale, sau în fct de adresa lor cu alte rude

Deces

CNP_M, CNP_F
CNP_CRIPT

Persoane este exclusa din gospodarie

Tipul gospodariei (nomenclator HST - EUROSTAT)
Marimea gospodariei

Familie

RPL

RNEP

constituire familie prin:

Nascuti-vii

Imigranti

Casatorii

Dizolvare familie prin:

Divort

Deces

Anuntari de resedinta

Schimbari de domiciliu

CNP_CRIPT, CNP_MAMA_CRIPT, CNP_TATA_CRIPT

CNP
Mai este nevoie de CNP_MAMA si CNP_TATA

-

CNP_M, CNP_F

CNP_M, CNP_F

CNP_CRIPT

-

-

se dizolva familia

se dizolva familia

Tipul familiei

Marimea familiei

TIMELINE/Cronologie

Termene transmiŃtere date de intrare

Denumire sursa	Anul crt.											
	Ian	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
INS-Masculini vii	15 (Nov - anul ant.)	15 (Dec - anul ant.)	15 (Ian)	15 (Feb)	15 (Mar)	15 (Apr)	15 (May)	15 (Jun)	15 (Jul)	15 (Aug)	15 (Sept)	15 (Oct)
INS-Decese	15 (Nov - anul ant.)	15 (Dec - anul ant.)	15 (Ian)	15 (Feb)	15 (Mar)	15 (Apr)	15 (May)	15 (Jun)	15 (Jul)	15 (Aug)	15 (Sept)	15 (Oct)
INS-Casatori	15 (Nov - anul ant.)	15 (Dec - anul ant.)	15 (Ian)	15 (Feb)	15 (Mar)	15 (Apr)	15 (May)	15 (Jun)	15 (Jul)	15 (Aug)	15 (Sept)	15 (Oct)
INS-Divortii	15 (Nov - anul ant.)	15 (Dec - anul ant.)	15 (Ian)	15 (Feb)	15 (Mar)	15 (Apr)	15 (May)	15 (Jun)	15 (Jul)	15 (Aug)	15 (Sept)	15 (Oct)
DPABD-Stabiliti de domiciliu			15 (semestrul 2 an ant.)						15 (semestrul 1 an crt.)			
DPABD-AnunŃari de reŃedinŃă			15 (semestrul 2 an ant.)						15 (semestrul 1 an crt.)			
DPABD- RNEP												
DPABD- migraŃie externa		15 (1 Ianuarie)							15 (semestrul 1 an crt.)			
DGP-migraŃie externa			15 (semestrul 2 an ant.)						15 (semestrul 1 an crt.)			
IGI-migraŃie externa			15 (an ant.)									
ANAF-D112	31 (Dec - anul ant.)	31 (Ian)	31 (Feb)	31 (Mar)	31 (Apr)	31 (May)	31 (Jun)	31 (Jul)	31 (Aug)	31 (Sept)	31 (Oct)	31 (Nov)
ANAF-Depozit										30 (an-1)		
CNPP	45 (Dec)	45 (Ian)	45 (Feb)	45 (Mar)	45 (Apr)	45 (May)	45 (Jun)	45 (Jul)	45 (Aug)	45 (Sept)	45 (Oct)	45 (Nov)
SIIR	30 (an scolar crt.)									15 (an scolar)		
RMU		28 (an universitar crt.)								30 (an universitar)		
ANPIS	30 (Dec)	30 (Ian)	30 (Feb)	30 (Mar)	30 (Apr)	30 (May)	30 (Jun)	30 (Jul)	30 (Aug)	30 (Sept)	30 (Oct)	30 (Nov)
IMI	27 (Dec)	27 (Ian)	27 (Feb)	27 (Mar)	27 (Apr)	27 (May)	27 (Jun)	27 (Jul)	27 (Aug)	27 (Sept)	27 (Oct)	27 (Nov)
APIA										30 (an-1)		
ANDPDA	30 (Dec)	30 (Ian)	30 (Feb)	30 (Mar)	30 (Apr)	30 (May)	30 (Jun)	30 (Jul)	30 (Aug)	30 (Sept)	30 (Oct)	30 (Nov)
ANOFM	30 (Dec)	30 (Ian)	30 (Feb)	30 (Mar)	30 (Apr)	30 (May)	30 (Jun)	30 (Jul)	30 (Aug)	30 (Sept)	30 (Oct)	30 (Nov)
CNAS												
ANPD	30 (Dec)	30 (Ian)	30 (Feb)	30 (Mar)	30 (Apr)	30 (May)	30 (Jun)	30 (Jul)	30 (Aug)	30 (Sept)	30 (Oct)	30 (Nov)

Termene output

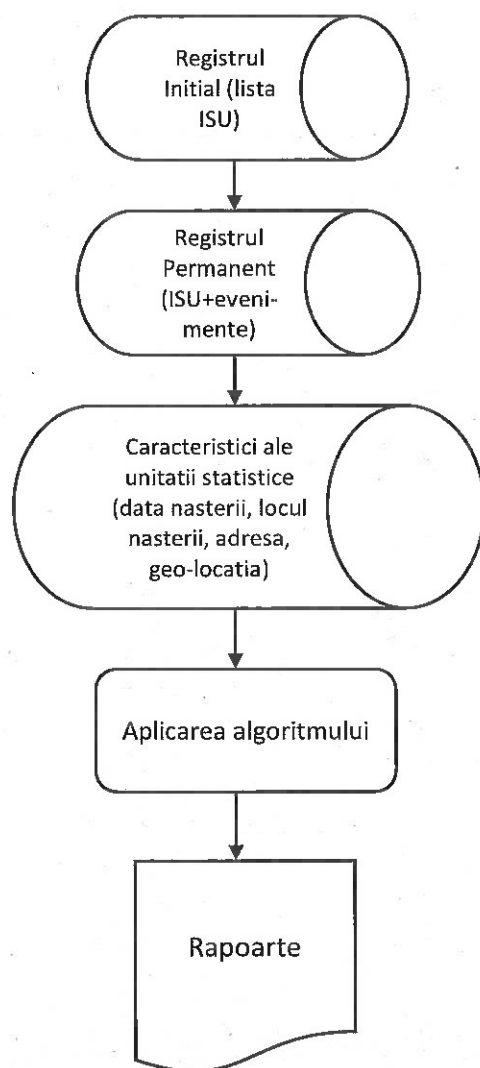
	An curent											
	Ian	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Pop rezidentia la 1 Ian - an crt.*						30 (provizorii)						
MigraŃie externa in perioada 1 Ianuarie - 31 decembrie - an ant.*						30 (provizorii)						
Populatia cu domiciliu in RO la 1 Ianuarie - an crt.												
Populatia cu domiciliu in RO la 1 Iulie - an crt.												
Esantion - stoc curent la data extragerii									30	15 (provizorii)		

* Termenele sunt cele prevazute de legislatia actuala . In Proiectul de Regulament (Anexa 3) este specificat

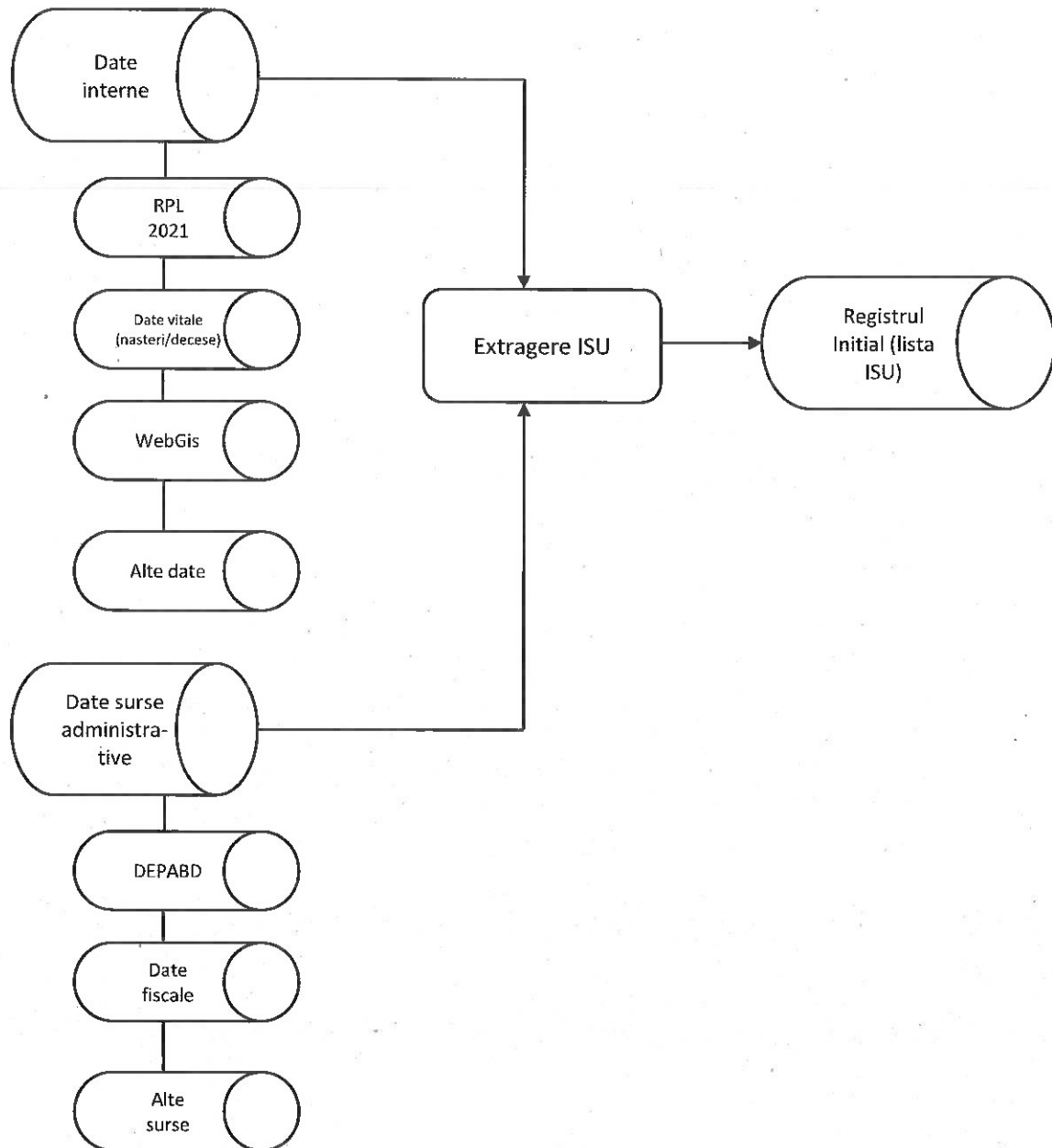
1. For the 'population' domain, tabular data and metadata shall be provided to the Commission (Eurostat) monthly and annually.
3. For the 'migration domain' tabular data and metadata shall be provided to the Commission (Eurostat) monthly and annually.

Structura (schemele) Registrului Intercenzitar

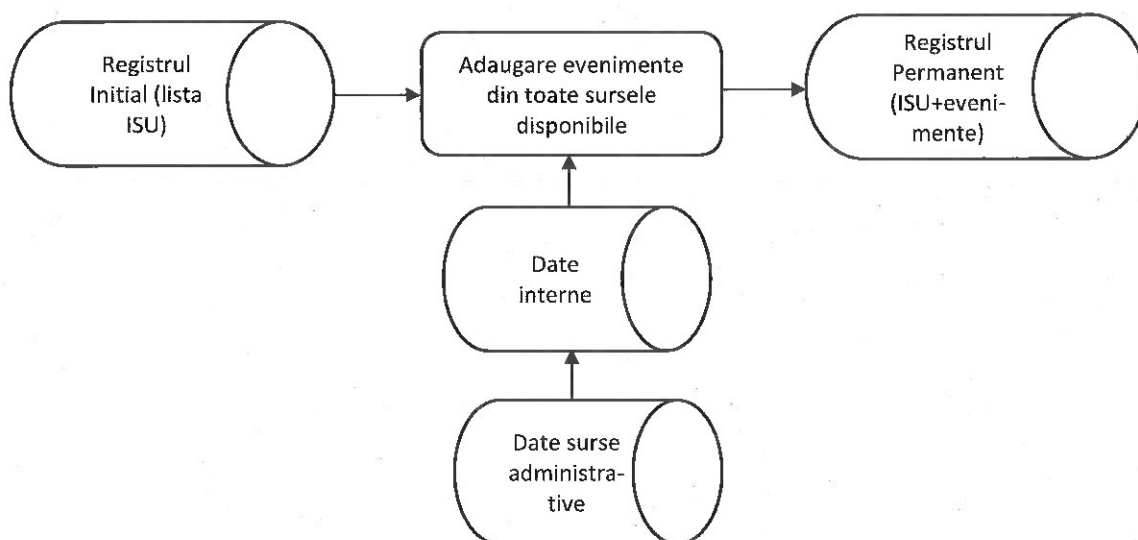
1. Fazele constructiei registrului



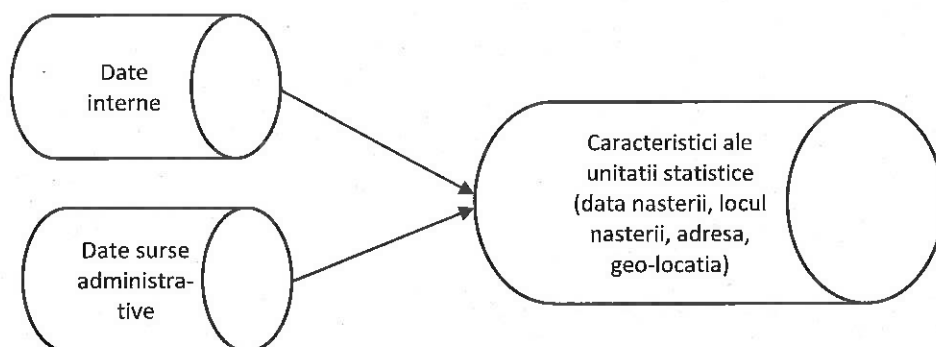
2. Construcția Registrului Initial (extragere date + verificare calitate date + tratare date + stocare date)



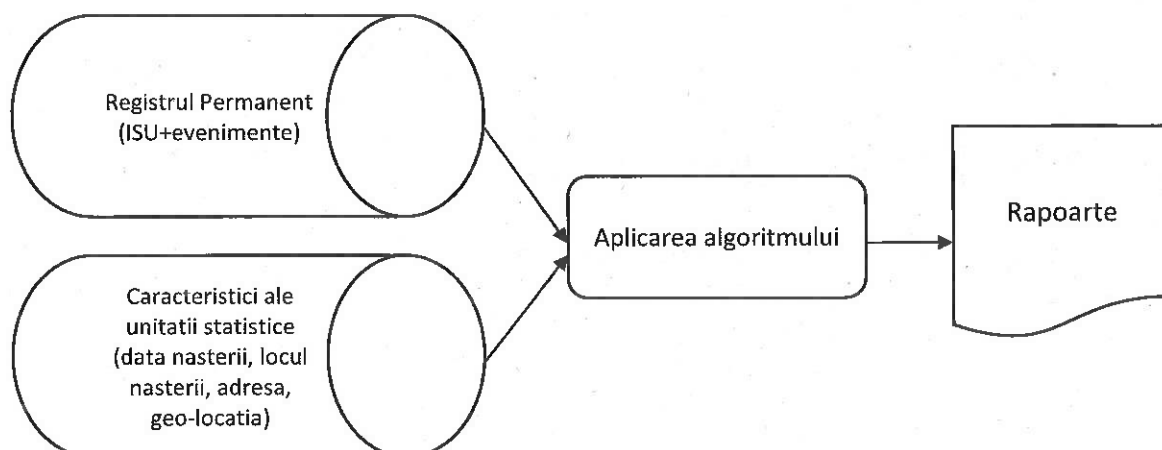
3. Registrul Permanent - actualizare (extragere date + verificare calitate date + tratare date + stocare date)



4. Colectarea caracteristicilor unitatii statistice (proces productie)

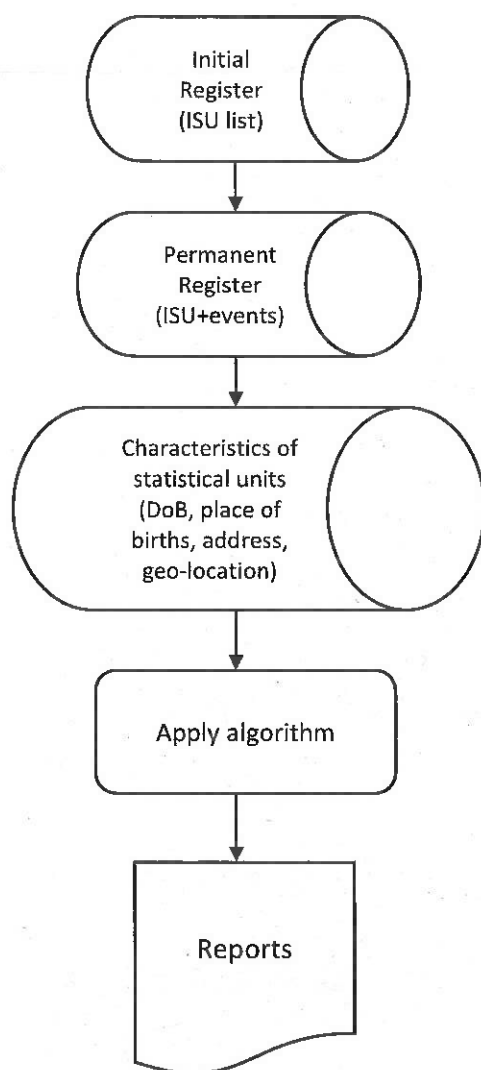


5. Rapoarte/diseminare (proces productie + rezultat primar + rezultate secundare)

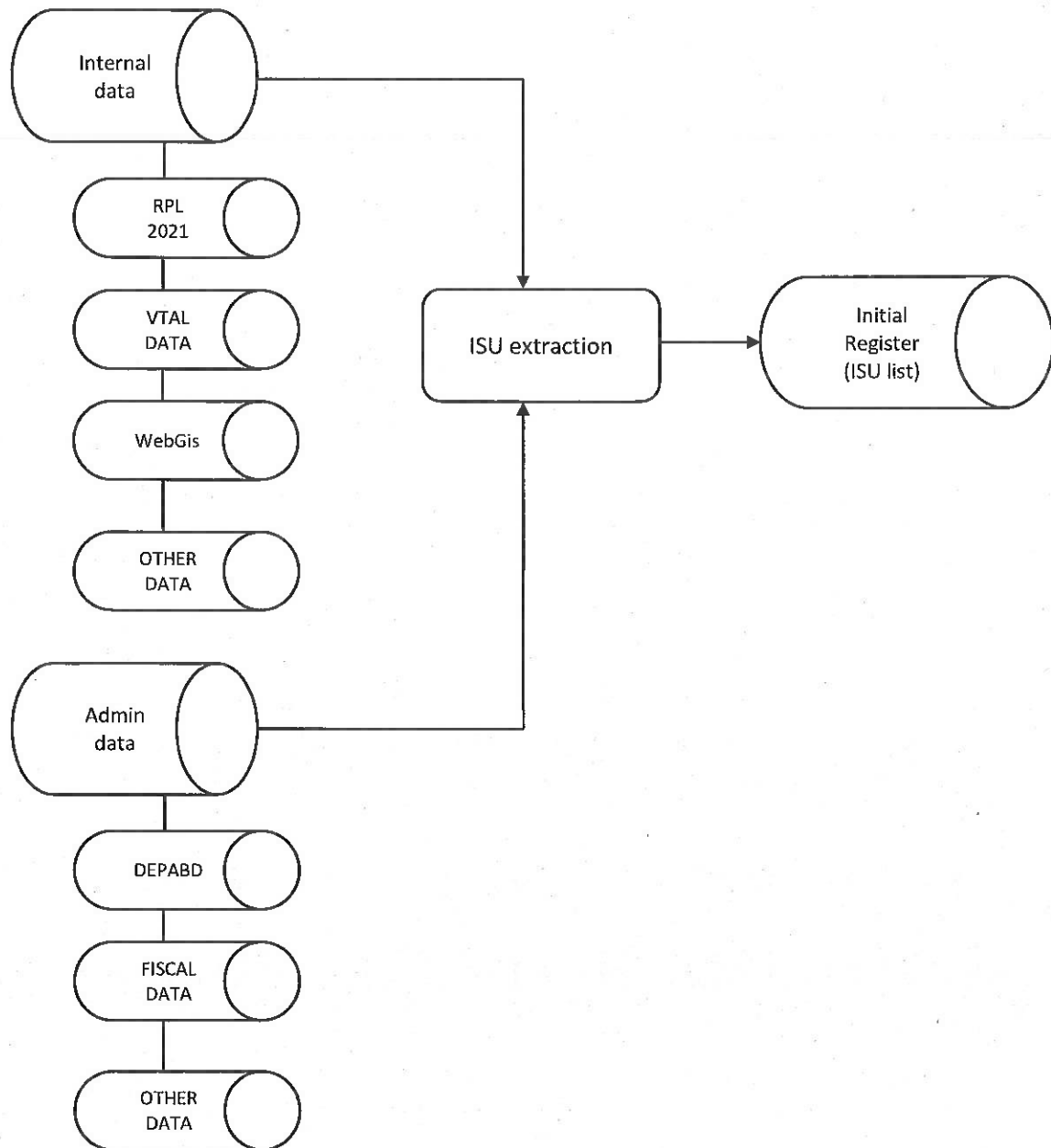


Intercensus Registry Structure (Schemas)

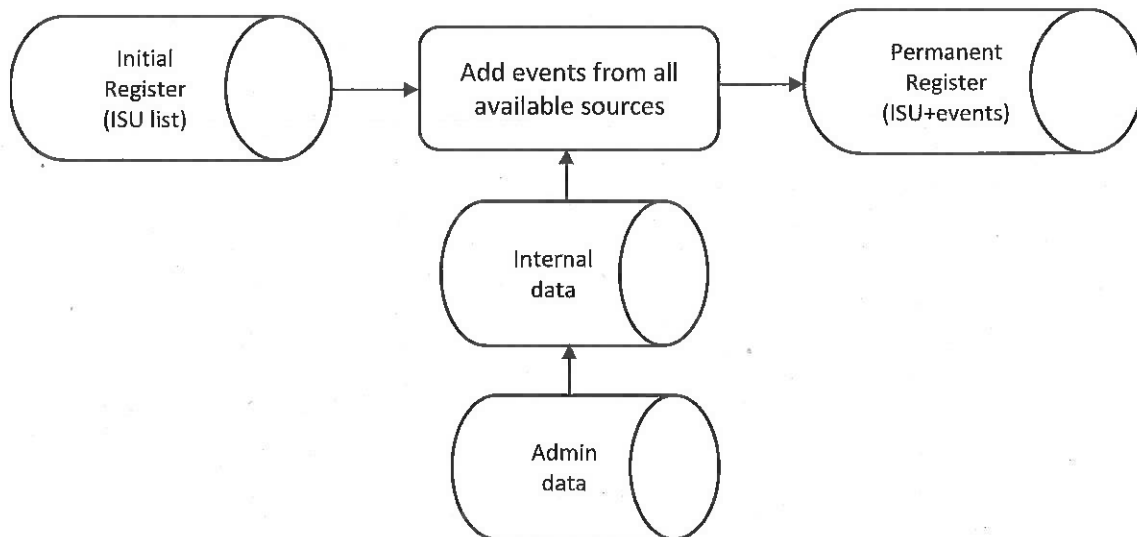
1. Registry's Construction Structure and Phases



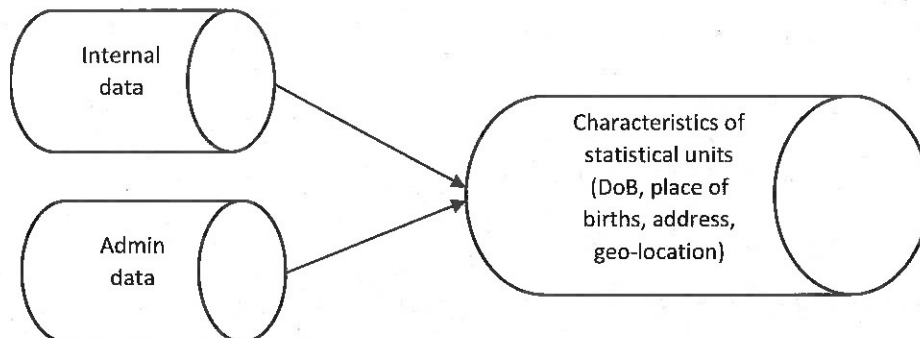
2. Initial Registry Construction (data extraction + quality check + data wrangling + data warehouse)



3. Permanent Registry Updates (data extraction + quality check + data wrangling + data warehouse)



4. Characteristics of Statistical Units Collection (production process)



5. Reports/dissemination phase (production process + primary outputs + secondary outputs)

