



ROMÂNIA

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

REZULTAT Nr. 13

**Raport privind serviciile de asistență tehnică oferite
Beneficiarului privind Recomandări și orientări pentru
planul documentat al sistemului integrat dedicat
implementării RPL 2021 (inclusiv gestionarea
utilizatorilor, criptarea ID-urilor, recuperare în urma
dezastrelor, menținerea disponibilității sistemului în caz de
avarie)**

Ianuarie 2023



*Proiect co-finanțat din Fondul Social European prin Programul Operațional
Capacitate Administrativă 2014-2020*

Declinarea responsabilității

Acest raport este un produs al personalului Băncii Mondiale. Constatările, interpretările și concluziile exprimate în acest document nu reflectă în mod obligatoriu punctele de vedere ale 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 României.

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 fie către: (i) Institutul Național de Statistică din România (B-dul Libertății nr. 16, Sector 5, București, România) sau (ii) Grupul Băncii Mondiale România (Strada Vasile Lascăr nr. 31, etaj 6, București, România).

Acest raport a fost livrat în luna ianuarie 2023 în cadrul Acordului de servicii de asistență tehnică rambursabile privind Consolidarea Sistemului Statistic Național (P167217), semnat la data de 17 septembrie 2019 între Institutul Național de Statistică din România și Banca Internațională pentru Reconstrucție și Dezvoltare. Acesta face parte din Rezultatul 13 din acordul menționat mai sus.

Cuprins

Tabel de figuri.....	4
Listă de tabele.....	4
Lista acronimelor.....	5
Introducere.....	6
1. Prezentare generală a sistemului integrat RPL	8
2. Fluxul de date în toate etapele RPL	10
2.1. Fluxul datelor colectate din surse administrative	10
2.2. Fluxul de date în cadrul CAWI.....	11
2.3. Fluxul de date în cadrul CAPI	12
3. Infrastructura RPL efectivă pentru colectarea datelor	13
3.1. Arhitectura generală.....	13
Sisteme de echilibrare a proceselor	14
Servere ale bazelor de date	14
Servere de aplicații	14
Server de monitorizare	15
3.2. Disponibilitate ridicată.....	15
3.3. Performanță	16
3.4. Backup	16
3.5. Securitate	16
3.6. Gestiunea jurnalelor.....	17
3.7. Recuperare în urma dezastrelor	17
4. Monitorizare și raportare pe durata colectării datelor RPL.....	20
4.1. Monitorizare și raportare zilnică.....	20
4.2. Monitorizarea și raportarea activității personalului de recenzare în vederea plății	22
5. Procesele de validare a datelor colectate	23
5.1. Prelucrarea datelor pe durata colectării datelor CAWI, CAPI.....	23
5.2. Prelucrarea datelor ulterior colectării datelor	24
6. Editarea și producția de date în scopul diseminării	25
6.1. Distribuția pe griduri a datelor RPL	25
Script R pentru transformare	26
Aplicația Shiny	28
6.2. Diseminarea (diseminare tabelată) a datelor RPL	30
Cadru general și logică	30
Cerințe tehnice și componente software utilizate	31
Conceptul soluției și fluxul de proces	31
Cerințe și condiții în materie de date	33
6.3. Cerințe tehnice pentru aplicații de diseminare a RPL.....	35
Cerințe preinstalate pe scurt	35
Alte resurse necesare	35
7. RPL2021 - concluzii privind implementarea efectivă	36
8. Anexe	38
Anexa 1 - Sistemul informatic de colectare a datelor RPL – Raport conform cu execuția ...	39
Anexa 2 – Modele de rapoarte de monitorizare zilnică (CAWI, CAPI).....	40
Anexa 3 - Modele de rapoarte de plăți aferente CAPI.....	41
Anexa 4 - Scripturi de validare și aprobare a colectării datelor RPL (CAWI, CAPI).....	42

Tabel de figuri

Figura 1 - Procesul sistemului integrat al RPL	8
Figura 2 - Diagrama de proces pentru RPL efectiv	10
Figura 3 - Flux de lucru de validare a Chestionarului RPL în CAWI realizată prin Survey Solutions ..	11
Figura 4 - Flux de lucru de validare a Chestionarului RPL în CAPI realizată prin Survey Solutions...	12
Figura 5 - Mediul de producție al RPL	13
Figura 6 - Mecanism de disponibilitate ridicată a sistemelor de echilibrare a proceselor RPL.....	15
Figura 7 - Mecanism de disponibilitate ridicată a bazei de date RPL	15
Figura 8 - Configurarea inițială a Survey Solutions pentru RPL.....	18
Figura 9 - Exemplu de raport zilnic de date CAWI colectate (pe județe).....	20
Figura 10 - Exemplu de raport zilnic de date CAWI colectate (pe UAT-uri)	20
Figura 11 - Exemplu de raport zilnic de date CAWI colectate (pe ARA).....	21
Figura 12 - Exemplu de raport zilnic cu tipuri de erori/mesaje înregistrate în timpul CAWI.....	21
Figura 13 - Exemplu de raport zilnic de date CAPI colectate (țară, județ, recenzor).....	21
Figura 14 - Exemplu de raport zilnic de date CAPI colectate (județ, recenzor).....	21
Figura 15 - Exemplu de raport de plată pentru activitatea recenzorului	22
Figura 16 - Exemplu de raport de plată pentru activitatea recenzorului, pe județe și UAT-uri.....	22
Figura 20 - Fluxuri automate pentru validare (faza CAWI) / aprobare (faza CAPI).....	23
Figura 21 - Aplicație pe bază de R Shiny pentru vizualizarea gridurilor de populație.....	28
Figura 22 - Hartă în format .pdf disponibilă pentru descărcare.....	29
Figura 23 - Fluxul procesului de diseminare	32
Figura 24 - Format de date diseminate salvat	33

Listă de tabele

Tabelul 1 - Resurse IT RPL alocate.....	14
Tabelul 2 - Resurse ale centrului secundar de date RPL	17
Tabelul 3 - Categoriile temei „recensământ” în Eurostat	26
Tabelul 4 - Categoriile suplimentare (8) ale temei „recensământ” impuse de INS	27

Lista acronimelor

ARA	Asistent de autorecenzare
UAT	Unități administrativ-teritoriale
CAPI	Intervievare personală asistată de calculator
CAWI	Intervievare online asistată de calculator
CNP	Număr personal de identificare (Cod numeric personal)
CONRENA	Proiectul „Consolidarea Sistemului Statistic Național și modernizarea proceselor de producție statistică pentru efectuarea recensămintelor naționale”
SCR	Sistem de Coordonate de Referință
BD	Bază de date
DTS	Direcții Teritoriale de Statistică
EPSG	Grupul European de Prospekțiune a Zăcămintelor de Țiței
ETRS	Sistemul de Referință Terestru European
UE	Uniunea Europeană
GIS	Sistemul Informațional Geografic
GPS	Sistemul de Poziționare Globală
INS	Institutul Național de Statistică
INSPIRE	Infrastructură pentru Informații Spațiale în Comunitatea Europeană
IP	Protocol Internet
IT	Tehnologia informației
LAEA	Proiecție azimutală echivalentă Lambert
PIN	Număr personal de identificare (Cod numeric personal)
RPL2021	Recensământul Populației și al Locuințelor, etapa 2021
RAS	Servicii de Asistență Tehnică Rambursabile
SICCA	Soluție pentru Arhitectura și Calculul Costurilor asociate Produselor și Activităților INS
SIRUTA	Sistemul Informatic al Registrului Unităților Teritorial-Administrative
SDC	Controlul privind divulgarea informațiilor statistice
STS	Serviciul de Telecomunicații Speciale
VPN	Rețea virtuală privată
UI	Interfață cu utilizatorul
BM	Banca Mondială

Introducere

Scopul prezentului raport este de a prezenta *recomandările și orientărilor privind planul documentat pentru sistemul integrat dedicat implementării RPL 2021 (inclusiv gestionarea utilizatorilor, criptarea ID-urilor, recuperare în urma dezastrelor, menținerea disponibilității sistemului în caz de avarie)*. Aceste acțiuni fac parte din livrabilele din cadrul Acordului de Servicii de Asistență Tehnică Rambursabile (RAS) privind Consolidarea Sistemului Statistic Național (P167217). Proiectul este implementat de către Institutul Național de Statistică (INS), cu sprijinul Băncii Mondiale.

Prezentul Rezultat prezintă sprijinul acordat Institutului Național de Statistică (INS) pentru implementarea Recensământului Populației și Locuințelor, runda 2021, și documentarea sistemului de colectare și producție a datelor pentru această lucrare importantă de statistică socială.

România a implementat în premieră un Recensământ al Populației și Locuințelor (RPL) complet digitalizat (100% informatizat) (2022, 14 martie – 31 iulie 2022). Acesta a permis colectarea eficientă a datelor pentru 95,4% (18,15 milioane) din populația estimată a României cu ajutorul unui mediu online dedicat și a asigurat evitarea oricăror pierderi de date colectate cauzate de erori de citire manuală sau automată inerente recensământului tradițional cu creion și hârtie.

Cunoștințele și experiența dobândite de INS pe durata actualului RPL creează fundamentele implementării unor viitoare anchete în domeniul statisticii sociale și ale tranziției la statistica intercenzitară a populației și locuințelor în cadrul proceselor anuale pe bază de registre. Logistica aferentă colectării datelor în baza combinației optime identificate (CAWI-CAPI) și implementate pe durata RPL este reflectată în prezentul raport, cuprinzând șase secțiuni.

Prezentarea generală a fluxului sistemului integrat RPL, prima secțiune, include principalele faze din organizarea proceselor de colectare a datelor, utilizarea mediilor fizice și online/internet (mediile IT ale INS și STS, soluții și aplicații, proceduri), precum și principalele sarcini ale instituțiilor responsabile, care permit exploatarea surselor de date în scopurile recensământului.

A doua secțiune prezintă o sinteză a fluxului de date pentru organizarea colectării datelor în toate etapele RPL, inclusiv detalii privind procesele de obținere a datelor din surse administrative și colectarea datelor în baza unor secvențe logice din cadrul fazelor CAWI și CAPI.

Sistemul de colectare a datelor RPL, prezentat în a treia secțiune, expune arhitectura generală (componente hardware, sisteme de operare, aplicații software, baze de date și servicii electronice), structura disponibilității ridicate, conceptul axat pe performanță, soluția de back-up, precum și procesul de securitate și testare pentru sistemul de colectare a datelor. Configurația și instalarea detaliate ale întregului sistem sunt descrise în raportul de realizare propriu-zisă, conform cu execuția, din anexă (inclusiv codul pentru comenzi), dedicat specialiștilor IT și administratorilor de sistem ai INS.

Rapoartele privind monitorizarea colectării datelor RPL sunt prezentate în a patra secțiune, care oferă informații legate de proces și exemple de rapoarte de monitorizare a colectării datelor în cadrul fazelor CAWI și CAPI, cu valori generale și pe județe, precum și rapoartele referitoare la activitatea personalului de recenzare. Machetele rapoartelor de monitorizare și scripturile utilizate pentru a calcula volumul de chestionare colectate și sumele de plată sunt prevăzute în anexe și arhivate în format electronic.

Datele RPL colectate au fost prelucrate pe durata recenzării sau în urma recenzării, în scopuri de validare, rectificare a erorilor, deduplicare, imputare a datelor lipsă sau incomplete, precum și de calculare a indicatorilor statistici incluși în producția rezultatelor recensământului. Scripturile aplicate pentru validarea datelor colectate în cadrul fazei CAWI, precum și cele pentru validarea și aprobarea datelor colectate în cadrul fazei CAPI, ca elemente ale asistenței tehnice acordate, sunt prevăzute ca anexe la prezentul raport și arhivate în format electronic în vederea unor operațiuni ulterioare de instalare și rulare de către personalul INS. Cu toate acestea, scripturile sunt instalate în mediul IT al INS („*Serverul de prelucrare a datelor*”). Etapele de prelucrare sunt descrise în a cincea secțiune.

INS diseminează rezultatele RPL și a considerat util pentru viitorii utilizatori ai datelor RPL să automatizeze accesul prin intermediul unei aplicații care este prezentată în a șasea secțiune. Diseminarea datelor RPL cuprinde trei soluții specifice: hipercuburi de date, care sunt publicate în mod obligatoriu pe portalul Eurostat; distribuția pe griduri a datelor RPL, ce permite diseminarea unui număr de 21 de indicatori (13+8) pe griduri (INSPIRE) de 1 km² pentru întreg teritoriul României; și diseminarea tabelată a rezultatelor RPL. În timp ce echipa INS gestionează hipercuburile de date pregătind mediul IT și instalând aplicațiile necesare furnizate de Eurostat cu scopul de a prelua date din antrepozitul de date RPL organizat în mediul IT al INS, această secțiune prezintă distribuția pe griduri și tabelele dinamice dezvoltate în cadrul proiectului. Structura, funcțiile, intrările, ieșirile și fluxurile de date descriu aplicațiile dezvoltate. Aplicațiile sunt instalate în mediul IT al INS („*Serverul de prelucrare a datelor*”) pentru a fi ulterior rulate de către personalul INS.

A șaptea secțiune a raportului prezintă o serie de concluzii pe marginea RPL, etapa 2021. RPL a reprezentat un eveniment statistic deosebit pentru România și primul recensământ informatizat, pregătit și implementat în cadrul unui efort comun al INS și al altor autorități publice pe parcursul ultimilor cinci ani. La rândul său, acesta a necesitat asistență tehnică substanțială în cadrul acordului RAS curent, livrată sub forma unui număr de cincisprezece (15) rezultate, dintre care zece (10) au reprezentat asistență directă dedicată RPL iar cinci (5) au asigurat suport încrucișat pentru alte componente ale proiectului CONRENA (RGA, perioada intercenzitară și SICCA).

Instrumentele de suport pentru producția de date și pentru sustenabilitatea proceselor și a aplicațiilor implementate spre a fi utilizate de către INS sunt prezentate în a opta secțiune a raportului. Informații suplimentare sunt disponibile în alte rapoarte livrate în baza RAS, fiind făcută referire la acestea în secțiunile prezentului raport.

1. Prezentare generală a sistemului integrat RPL

RPL etapa 2021 a reprezentat cea mai vastă operațiune statistică desfășurată de către INS în ultima decadă cu scopul de a colecta, prelucra și disemina date privind structura populației rezidente și a locuințelor din România. Scopul RPL a fost recenzarea integrală a populației, a unităților locative, a unităților locative și a gospodăriilor care locuia/existau la data de 1 decembrie 2021 în România.

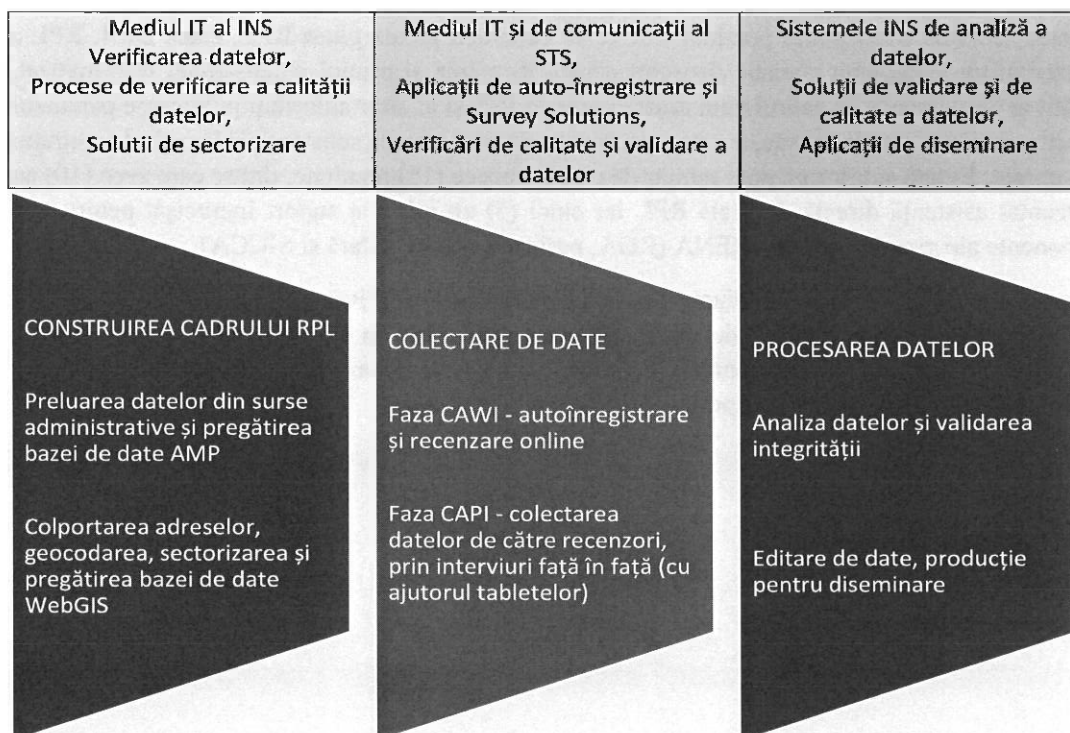
RPL 2021 a necesitat și a aplicat instrumente optimizate, respectiv pentru:

- întocmirea și elaborarea cadrului de desfășurare a recensământului
- colectarea electronică integrală a datelor în două moduri diferite, CAPI și CAWI, inclusiv geo-referențierea integrală a tuturor datelor colectate și un management permanent al calității aplicat în fiecare etapă a procesului de colectare a datelor
- prelucrarea datelor colectate și producția unor statistici în vederea diseminării

În concordanță cu scopul asumat al RPL, INS și alte entități-omolog responsabile au implementat proceduri și instrucțiuni administrative, infrastructură și mijloace de comunicare fizice și informatice, precum și instrumente de control pentru pregătirea și implementarea colectării datelor.

Sistemul integrat proiectat și implementat (a se vedea Figura 1) a îndeplinit așteptările procesului de colectare a datelor și a asigurat o rată foarte ridicată a datelor colectate de la 18,15 milioane de persoane (95,4% din populația domiciliată țintă a României), oferind o listă consecventă de variabile și rezultate calitative obținute în timp util.

Figura 1 - Procesul sistemului integrat al RPL



Procesul de colectare a datelor a fost organizat în format fizic și în medii online/pe internet, găzduite de INS și STS dar și pentru a permite folosirea surselor de date administrative în scopurile

recensământului. Sursele de date administrative obținute de la instituții publice au fost organizate în cadrul mediului INS. Colectarea datelor de la populația rezidentă a fost organizată în mediu online/internet în fazele CAWI și CAPI, prin intermediul auto-înregistrării și a aplicațiilor Survey Solutions. Datele recensământului au fost găzduite și securizate cu ajutorul sistemelor de colectare a datelor ale STS, în timp ce producția statistică a fost și este asigurată, până la publicarea rezultatelor, prin intermediul unor sisteme de analiză organizate în mediul INS.

Fazele axate pe colectarea datelor au fost și sunt asigurate prin controlul calității datelor, procese fizice și logice în materie de siguranță și securitate descrise în rapoartele mai multor rezultate livrate în cadrul aceluiași acord RAS (a se vedea Rezultatul 4.1b, Rezultatul 7a, Rezultatul 10a, Rezultatul 10b, Rezultatul 10c).

Desfășurarea RPL efectiv a avut la bază configurația definită în Rezultatul 4.1c: *Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detalii privind modul în care va fi realizată implementarea infrastructurii IT pentru RPL 2021)*.

Rolurile instituționale și principalele responsabilități privind configurarea sistemului integrat RPL au fost distribuite după cum urmează:

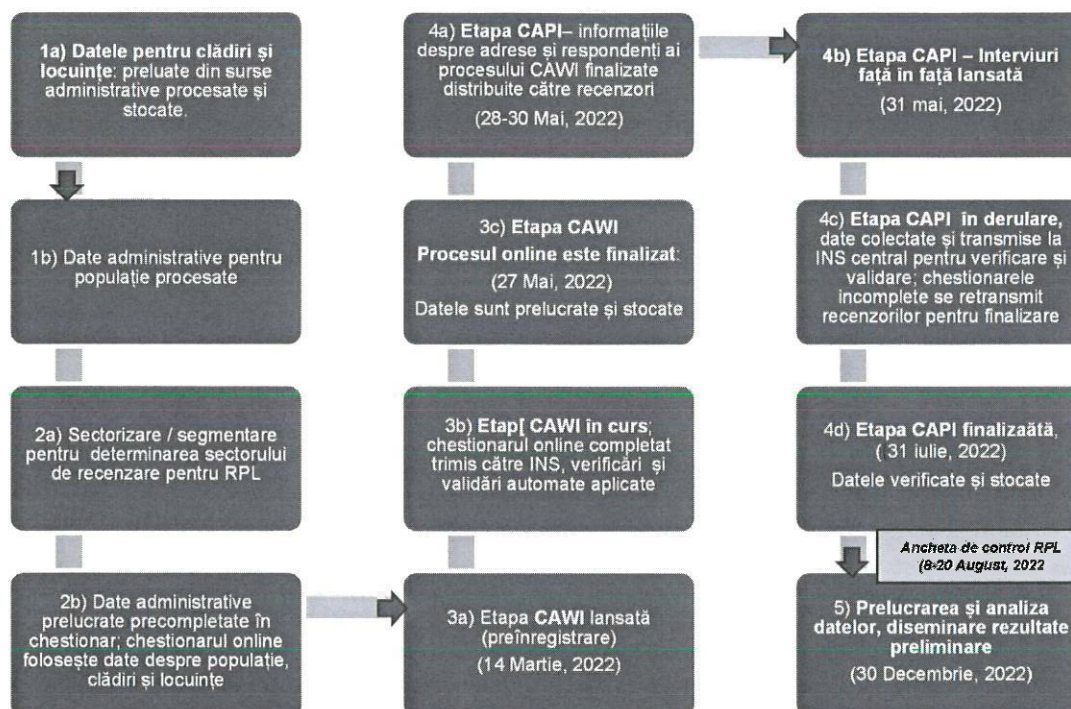
- INS
 - Asigură resurse pentru găzduirea, contractarea, instalarea, configurarea și recepția sistemului IT integrat al INS
 - Asigură resurse pentru exploatarea sistemului IT integrat anterior, pe parcursul și ulterior RPL2021
 - Furnizează conectivitate la rețea prin canal de fibră optică
 - Prevede criteriile de recepție și administrează testele de recepție
 - Oferă suport în procesul de testare a performanțelor
- STS
 - Asigură găzduire și furnizează mașini virtuale pentru aplicații software, servere ale bazelor de date și de echilibrare a proceselor pentru activități de colectare a datelor
 - Furnizează conectivitate la rețea prin canal de fibră optică și VPN între STS și INS
 - Configurează accesul la Internet al sistemului de echilibrare a proceselor din aplicație
 - Configurează FQDN (*nume de domeniu complet calificat*) pe domeniul insse.ro
 - Asigură interconectarea cu baza de date a populației în vederea verificării Codurilor Numerice Personale (CNP)
- BM
 - Operează ca integrator general de sisteme, oferind recomandări și orientări
 - Administrează procesul de instalare, configurare și testare a aplicațiilor
 - Instalarea, Configurarea și Optimizarea serverului Bazei de date PostgreSQL
 - Instalarea și configurarea serverelor aplicației Survey Solutions, sistemului de echilibrare a proceselor din aplicație, componentei software de monitorizare, setului de instrumente pentru teste și a portalului de auto-înregistrare
 - Desfășoară procesul de testare a performanțelor, analizează rezultatele și oferă recomandări
 - Oferă documentații conform cu execuția pentru instalarea aplicațiilor

Astfel, echipele INS, STS și BM s-au implicat în permanență, prin cooperare, în desfășurarea fazelor efective ale recensământului și monitorizare, respectiv în facilitarea și asigurarea funcționării integrale a sistemului integrat RPL, 24 de ore din 24, 7 zile din 7, din 14 martie până în 31 iulie 2022.

2. Fluxul de date în toate etapele RPL

Fluxul de proces și cel de date au fost implementate și au funcționat conform planificării din raportul Rezultatului 4.1c: *Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detalii privind modul în care va fi realizată implementarea infrastructurii IT pentru RPL 2021)*, a se vedea Figura 2 de mai jos.

Figura 2 - Diagrama de proces pentru RPL efectiv



Procesul este explicat suplimentar în această secțiune și completat cu cadrul de asigurare a calității aplicat datelor colectate, și cu instrumentele statistice de promovare a protecției și securității datelor (a se vedea informații detaliate în raportul Rezultatului 7a: *Metodologie de evaluare și promovare continuă și în timp real a calității și acoperirii datelor colectate RPL2021 și a protecției/securității datelor*; și în cel al Rezultatului 10c: *Stabilirea unor instrumente statistice provizorii pentru recensăminte folosind metode multimodale de promovare a protecției și securității datelor*).

Procesul de colectare a datelor a fost coordonat și implementat de către INS și autoritățile centrale și cele locale. Procesul s-a axat pe folosirea datelor disponibile pentru populație și clădiri, obținute din procesul propriu-zis de colectare a datelor (CAWI, CAPI), însă și pe maximizarea informațiilor oferite de date obținute din surse administrative.

2.1. Fluxul datelor colectate din surse administrative

Acest demers a reprezentat o colectare preliminară a datelor despre populație, clădiri și unități locative. Activitatea s-a realizat de către echipa INS în perioada de pregătire a recensământului. Caracteristicile pentru majoritatea clădirilor și unităților locative, inclusiv coordonatele GIS și adresele clădirilor, au fost colectate, prelucrate și stocate într-o bază de date inițială (a se vedea detalii despre pregătirea bazei de date WebGIS în raportul Rezultatului 7a). Clădirile și unitățile

locative, în special adresele și datele geografice, au fost necesare în sine și utilizate drept referințe inițiale în recensarea populației. Procesul de sectorizare pentru crearea sectoarelor (zone de recensare) pentru populația/zona de recensat s-a realizat și verificat de către personalul INS și DTS și a urmărit să faciliteze activitățile de eșantionare în cadrul recensământului și pe cele logistice.

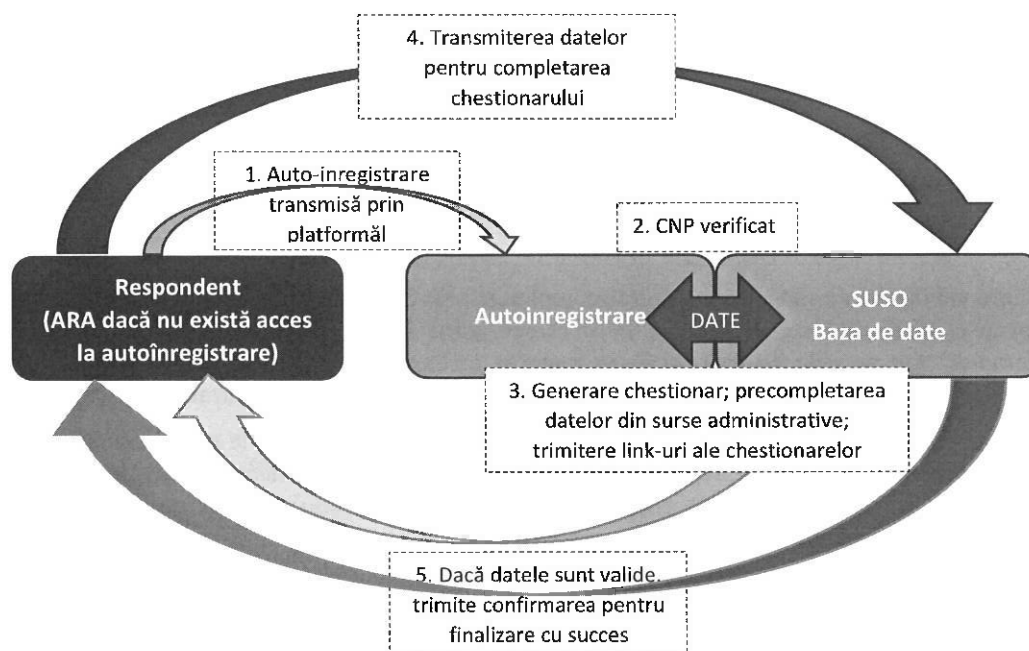
Datele din surse administrative privind populația au fost prelucrate pentru procesul de preîncărcare a chestionarelor personale în sistemul Survey Solutions de colectare a datelor, precum și pentru validarea datelor colectate în faza de prelucrare post-recensare. O parte din datele cu caracteristicile clădirilor și unităților locative au fost colectate de la un membru al gospodăriei. CNP-ul persoanei a fost transformat într-un Identificator Statistic Unic printr-o funcție de criptare HASH ireversibilă.

2.2. Fluxul de date în cadrul CAWI

Preînregistrare - respondenții au accesat aplicația de preînregistrare găzduită pe internet, iar capul gospodăriei a completat CNP-ul și adresa gospodăriei/gospodăriilor, adresa de e-mail și numărul de telefon, numărul membrilor gospodăriei și, după caz, CNP-ul și adresa de e-mail ale acestora. De asemenea, capul gospodăriei are posibilitatea de a declara adresa unor unități locative secundare. Un e-mail de confirmare a preînregistrării, trimis de domeniul insse.ro, conținând linkul către chestionarul/formularul individual și instrucțiuni de completare a chestionarului, a fost primit de către respondent. CNP-ul a fost criptat după preînregistrare, iar Identificatorul Statistic Unic a fost alocat fiecărui respondent. Capii de familie au primit în e-mail și linkuri către chestionarul/formularul pentru unitățile locative secundare.

Autorecensare – respondenții au accesat linkul chestionarului primit prin e-mail, au completat formularul online și au transmis informațiile către serverul principal de aplicații. Transmiterea și validarea datelor în fluxul de date CAWI sunt prezentate în figura de mai jos (sursă: raportul Rezultatului 7a).

Figura 3 - Flux de lucru de validare a Chestionarului RPL în CAWI realizată prin Survey Solutions

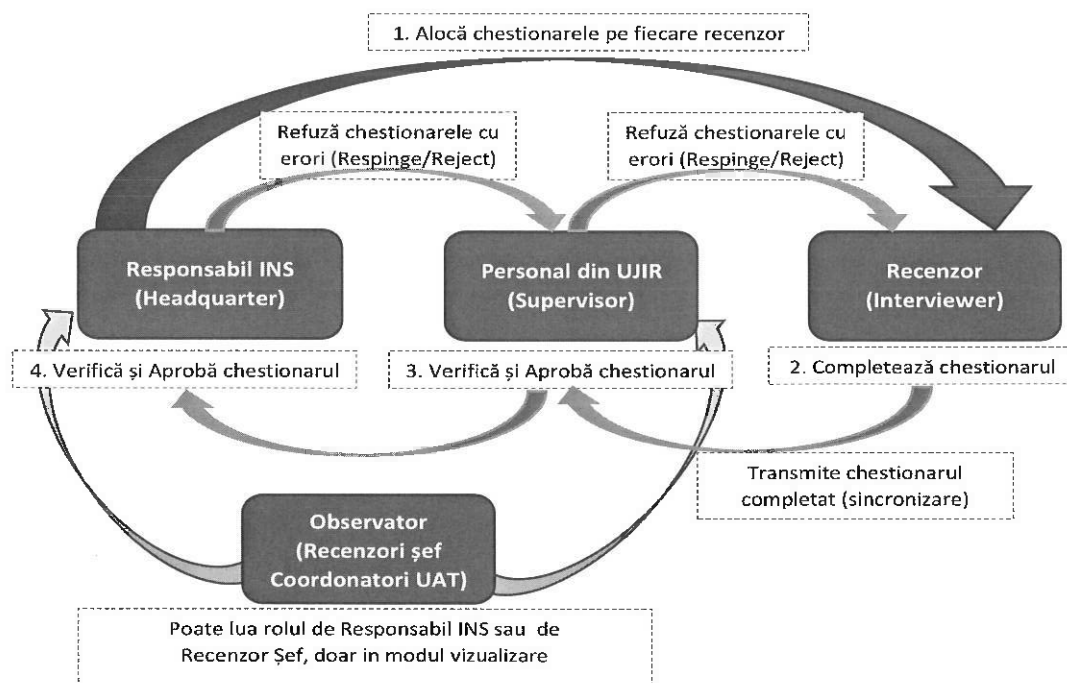


Pe serverul principal s-a executat o procedură automată de validare, prin verificarea acceptabilității informațiilor furnizate. Dacă, potrivit cerințelor, faza de autorecenzare a procesului a fost încheiată/finalizată, chestionarul online a fost trimis la INS pentru stocarea finală a datelor. În caz contrar, în faza CAPI, recenzorul a avut sarcina de a contacta respondenții în faza ulterioară CAPI de colectare față în față a datelor. Pentru chestionare validate, respondenții au putut primi o confirmare online a reușitei procesului de autorecenzare accesând o platformă web dedicată.

2.3. Fluxul de date în cadrul CAPI

INS a efectuat interviuri CAPI față în față cu persoane care nu s-au autorecenzat sau au furnizat informații incomplete cu ocazia autorecenzării, apelând la recenzori care au înregistrat răspunsurile pe dispozitivele lor portabile. Transmiterea și validarea datelor, în fluxul de date CAPI, sunt prezentate în figura de mai jos (sursă: raportul Rezultatului 7a).

Figura 4 – Flux de lucru de validare a Chestionarului RPL în CAPI realizată prin Survey Solutions



Pe baza validării automate, chestionarele completate prin CAPI au fost aprobate la Sediul central și transmise bazei de date centrale a INS. Supervizorii (de la DTS) au verificat chestionarele incomplete sau pe cele cu erori și le-au realocat recenzorilor pentru a colecta datele lipsă sau incorecte. După validare, chestionarele au fost transmise bazei de date centrale a INS.

Din nou, INS a verificat gradul de acceptabilitate al acestor informații, la sediul central. Dacă ele au fost acceptabile, faza de recenzare a procesului a fost finalizată/încheiată. În caz contrar, supervizorii recenzorilor au fost contactați pentru a opera modificări, iar modificările au fost realizate de către recenzori la solicitarea supervizorilor.

Informațiile, odată considerate acceptabile, au fost prelucrate și stocate.

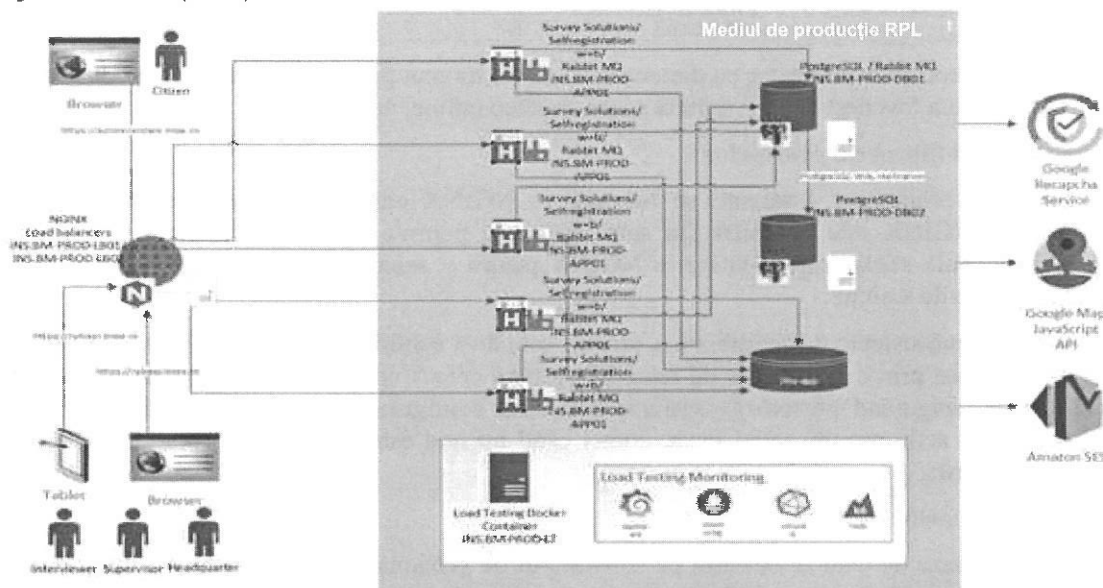
3. Infrastructura RPL efectivă pentru colectarea datelor

Rapoartele privind dezvoltarea și implementarea RPL au furnizat informații detaliate despre infrastructura (hardware, aplicații și comunicații) pusă în aplicare pentru procesul de colectare a datelor (a se vedea rapoartele de la Rezultatele 4.1b, 4.1c, 4.2a, 7a, 10a, 10b, 10c sub același RAS¹). În continuare sunt prezentați pașii și conținutul infrastructurii RPL efectiv ca raport de execuție.

3.1. Arhitectura generală

Pe baza observațiilor obținute din pilot și a estimărilor volumului de muncă în timpul testării performanței (a se vedea *Anexa 1: Planificarea, execuția și rezultatele testului RPL*, Rezultatul 4.2a) pentru a se asigura că sunt îndeplinite capacitatea planificată a procesului de colectare a datelor și timpul de răspuns convenit, și coroborat cu testarea de securitate pentru componentele expuse la Internet ale sistemului (subiectul activităților experților STS acoperit pe protocolul STS – INS și rapoartele de securitate existente), sistemul actual de colectare a datelor PHC a fost construit pe o arhitectură cu trei niveluri, constând într-o pereche de 2 echilibratori de încărcare care deservesc cererile web, 5 noduri de server de aplicații și o pereche de 2 noduri de bază de date. Valorile semnificative sunt capturate de la toate nodurile și trimise la un nod de monitorizare.

Figura 5 – Mediul de producție al RPL



¹ Rezultatul nr. 4.1b: *Recomandări privind elaborarea documentației tehnice cu scopul de a organiza achiziția unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocare de date, protecție și securitate a datelor pentru rularea activităților dezvoltate în baza RAS (GIS, RPL2021, RGA2020, perioade intercențitare, SICCA)*
Rezultatul 4.1c: *Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detalii privind modul în care va fi realizată implementarea infrastructurii IT pentru RPL 2021);*
Rezultatul nr. 4.2a: *Recomandări și orientări oferite Beneficiarului pentru rularea sistemului IT integrat cu componentele sale, hardware (inclusiv tablete) și software (licențe), inclusiv stocare de date, protecție și securitate a datelor pentru rularea activităților dezvoltate în baza RAS (GIS, RPL2021, RGA2020, perioade intercențitare, SICCA);*
Rezultatul 7a: *Raport privind serviciile de asistență oferite Beneficiarului privind Metodologia de evaluare și promovare continuă și în timp real a calității și acoperirii datelor colectate în cadrul RPL2021, precum și a protecției/securității datelor;*
Rezultatul 7c: *Raport privind serviciile de asistență oferite Beneficiarului privind Recomandările și bunele practici de implementare a sistemului de management al datelor pentru datele geospațiale RPL2021 și sectorizarea efectivă;*
Rezultatul 10a: *Raport privind Recomandările oferite Beneficiarului privind modul în care se desfășoară procesul de pilotare a RPL2021;*
Rezultatul 10b: *Raport privind serviciile de asistență oferite Beneficiarului privind Asistența tehnică și recomandările de bune practici pentru SDC, confidențialitatea datelor, modalități de a securiza microdate și date agregate; și*
Rezultatul 10c: *Raport privind serviciile de asistență oferite Beneficiarului privind Setul de instrumente statistice provizorii pentru recensământ folosind metode multimodale de promovare a protecției și securității datelor; și raportul propriu-zis*

Nodurile sunt dispozitive virtuale găzduite la sediul STS. Următoarele resurse au fost dedicate fiecărui nod:

Tabelul 1 – Resurse IT RPL alocate

Denumire	Spațiu necesar	SO invitat	Memorie	Procesoare	Adresă IP
INS.BM-PROD-APP01	1000 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP02	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP03	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP04	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-APP05	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-PROD-DB01	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-PROD-DB02	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-PROD-LB01	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.BM-PROD-LB02	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.VM-PROD-SI	98.09 GB	Ubuntu Linux 20.04 (64-bit)	48 GB	16	

Un spațiu de stocare suplimentar cu dimensiunea de 2T montat prin NFS pe primul nod de server de baze de date a fost destinat a fi utilizat pentru backup online, după caz.

Sisteme de echilibrare a proceselor

Sistemele de echilibrare a sarcinii se bazează pe NGINX ediția comunitară versiunea 1.18. Software-ul NGINX este compilat din sursă pentru a permite personalizări viitoare. Au fost adăugate module sticky suplimentare la NGINX pentru a asigura persistența sesiunii folosind module cookie de sesiune.

Rolul celor două sisteme de echilibrare a sarcinii este de a echilibra cererile web provenite de pe Internet (dirijate prin echipamente de rețea STS) către cele 5 noduri de aplicație într-o manieră consecventă, asigurând persistența sesiunii. Ele sunt configurate într-o configurație de înaltă disponibilitate activ-pasivă, astfel încât, atunci când un nod este nefuncțional, celălalt să poată prelua solicitările.

Servere ale bazelor de date

Serverele de baze de date se bazează pe software-ul de gestionare a bazelor de date PostgreSQL, ediție comunitară 12. PostgreSQL 12 este instalat din depozitul specializat dedicat versiunii actuale a sistemului de operare.

Nodurile bazei de date servesc ca stocare a informațiilor pentru păstrarea sarcinilor, interviurilor și a informațiilor de gestionare a sistemului și a utilizatorilor. Nodurile bazei de date utilizează replicarea activ-pasivă, astfel încât un nod (MASTER) să fie în modul citire-scriere, în timp ce al doilea nod (BACKUP) să fie în modul „doar citire”. Este implementată opțiunea de reluare automată astfel încât, în cazul în care MASTER este nefuncțional, nodul BACKUP să fie promovat la rang de master. De asemenea, baza de date a fost configurată pentru backup online, ceea ce face posibilă recuperarea punctului în timp în caz de inconsecvență a informațiilor.

Servere de aplicații

Serverele de aplicații constau în instanțe ale Survey Solution versiunea 21.1.7.30747, care este cea mai recentă versiune stabilă la momentul instalării. Fiecare nod comunică cu sistemele de

echilibrare a sarcinii pe portul 9700 prin http. O componentă dedicată numită ExportService este activată numai pe primul nod. Această componentă efectuează exportul de date și comunică cu toate celelalte noduri de pe portul 5001. Serviciul poate fi apelat doar pe partea dinspre server și nu are legătură la Internet. Datele din Survey Solutions pot fi exportate fie printr-un format de arhive specializate, folosind interfața web, fie prin conectarea la o altă instanță Survey Solutions prin API.

Server de monitorizare

Serverul de monitorizare se bazează pe un set docker de containere, cu Prometheus drept colector central de monitorizare, folosind și Grafana pentru vizualizarea valorilor și InfluxDB pentru înregistrarea valorilor. Sunt adăugate funcționalități suplimentare printr-o implementare k6 personalizată care poate fi utilizată pentru testarea performanței.

3.2. Disponibilitate ridicată

S-a luat în considerare înalta disponibilitate pentru a evita punctul unic de defecțiune în proiectarea sistemului RPL. De asemenea, a fost luat în considerare obiectivul punctului de recuperare de o oră (timp maxim în care anchetatorii pot rămâne deconectați de la sistem). Disponibilitatea ridicată este implementată în fiecare dintre cele 3 niveluri, după cum urmează:

- Pe sistemele de echilibrare a sarcinii, există o configurație activ-pasivă care afișează un al doilea nod în cazul în care serviciul NGINX al nodului curent este întrerupt.
- Pe serverele de aplicații, deoarece cererile sunt direcționate de la sistemele de echilibrare a sarcinii, o verificare a stării le poate indica sistemelor de echilibrare a sarcinii dacă nodul curent al serverului de aplicații este oprit, iar cererea poate fi redirecționată către un alt nod.
- Pe serverele de baze de date, există un nod MASTER activ care deservește toate cererile. Datele sunt replicate pe un al doilea nod (BACKUP) și există reluare automată. Mecanismul de restabilire este manual pentru un control mai bun al sursei defecțiunii.

În cadrul sistemelor de echilibrare a sarcinii, mecanismul de înaltă disponibilitate a fost implementat folosind pachetul software Keepalive. Keepalive a fost configurat să afișeze un IP virtual pe nodul MASTER atunci când serviciul NGINX de pe acest nod este funcțional, precum și să oprească IP-ul virtual pe MASTER și să îl afișeze pe nodul BACKUP în cazul în care NGINX are o defecțiune pe nodul MASTER.

Figura 6 – Mecanism de disponibilitate ridicată a sistemelor de echilibrare a proceselor RPL

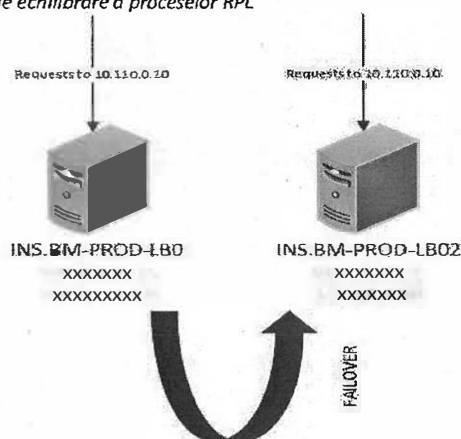
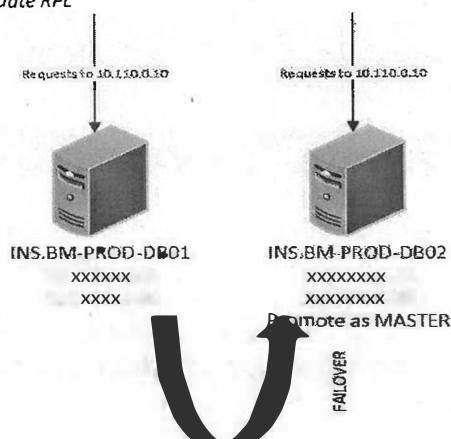


Figura 7 – Mecanism de disponibilitate ridicată a Bazei de date RPL



Un mecanism similar de înaltă disponibilitate a fost implementat pentru bazele de date. În cazul nodurilor cu baze de date, este necesară o acțiune suplimentară pentru a promova nodul BACKUP la rangul de MASTER pentru ca serverul să poată scrie și cererile către motorul SQL.

Pentru a verifica dacă primul nod acționează ca master, astfel încât să fie informat despre schimbările de stare, a fost implementat un mecanism personalizat. Pe nodul principal, un script de verificare actualizează informațiile dintr-un tabel utilizând data și ora curente la fiecare 5 secunde. Dacă informațiile din tabel (care sunt replicate în al doilea nod) diferă de data și ora curentă cu mai mult de 100 de secunde, înseamnă că fie serverul PostgreSQL este nefuncțional, fie serverul este în modul „doar citire”; în acest caz, poate fi deci declanșată reluarea automată.

În cazul serverelor de baze de date, mecanismul de restabilire este manual și este descris în capitolul referitor la operațiuni. Din cauza faptului că pot fi multiple cauze ale defecțiunilor, de la defecțiuni fizice la probleme de configurare software, se recomandă ca restabilirea să fie efectuată manual pentru a ajunge din nou la starea inițială.

Serverele de baze de date au fost configurate pentru replicare. Mecanismul de replicare ales a fost replicarea logică (a se vedea <https://www.postgresql.org/docs/12/different-replication-solutions.html>). Acest tip de replicare are avantajul de a fi un mecanism integrat, oferind un nivel ridicat de consistență și având o amprentă minimă asupra operațiunilor serverului principal.

3.3. Performanță

Sistemul a fost dimensionat pentru a permite unui număr de 25.000 de utilizatori/recenzori să lucreze concomitent. Deoarece configurația recomandată pentru un singur nod s-a dovedit că suportă cel puțin 6.000 de recenzori, atunci 5 noduri de aplicație ar trebui să fie mai mult decât suficiente pentru a gestiona RPL.

3.4. Backup

Datorită arhitecturii de înaltă disponibilitate, backupul configurațiilor serverului nu este necesar, deoarece fiecare configurație de nod are backup într-un alt nod. Deci, backupul sistemului se reduce la backupul datelor.

Postgresql a fost configurat pentru backup online, astfel încât este posibilă o recuperare consecventă la un moment dat pentru a recupera din erorile logice. Jurnalele de arhivă sunt configurate pe fiecare nod MASTER din ambele centre de date.

3.5. Securitate

Securitatea fizică a sistemului se bazează pe controalele de securitate implementate în centrele de date STS și pe procedurile și reglementările interne ale INS și BM pentru utilizatorii care accesează sistemul de la terminalele găzduite fie de INS, fie de BM.

Securitatea logică a sistemului are diferite niveluri de protecție:

- Firewallurile de rețea STS protejează accesul la sistemele găzduite în centrele de date STS. Accesul la interfața web publică a sistemului este limitat la IP-uri naționale.
- Toate dispozitivele au implementat firewalluri-gazdă care permit doar traficul minim necesar pentru funcționalitatea sistemului.
- Serverele Windows sunt protejate de antivirusul nativ Windows Defender.
- Accesul administrativ la sistem este protejat de o soluție VPN restricționată point-to-site, disponibilă exclusiv personalului autorizat.
- Accesul la aplicația web este posibil numai pentru utilizatorii autentificați. Numele de utilizator și parola sunt obligatorii pentru accesarea funcționalităților sistemului, iar utilizatorii privilegiați sunt protejați suplimentar prin autentificare în doi pași. De asemenea, se folosește un mecanism captcha (serviciu Google Recaptcha) pentru a refuza accesul programatic.

Securitatea generală a sistemului a fost evaluată de STS conform bunelor practici. Pentru o protecție continuă împotriva vulnerabilităților comune, corecțiile de securitate pentru componentele sistemului au fost aplicate în mod regulat.

3.6. Gestiunea jurnalelor

Toate jurnalele de sistem au fost centralizate de STS. Acest lucru permite setarea alertelor pentru a anunța persoanele responsabile de anumite evenimente legate de securitatea, disponibilitatea și fiabilitatea sistemului.

3.7. Recuperare în urma dezastrelor

Mecanismul de recuperare în caz de dezastru se bazează pe comutarea DNS și pe replicarea bazei de date. Primul nod de pe site-ul secundar este replicat continuu de la primul nod de pe site-ul primar folosind replicarea în flux, astfel încât baza de date de pe site-ul secundar să fie actualizată într-o manieră consecventă.

Atunci când este declarată o situație de dezastru de către persoanele responsabile, se declanșează procedura de recuperare în caz de dezastru, care va avea următoarele etape principale:

1. Se comută conexiunea la rețea de la site-ul primar la cel secundar. Această acțiune va folosi actualizarea înregistrărilor DNS, astfel încât modificările să fie aplicate conform TTL convenit. În acest caz, procesul va dura o oră. Site-ul secundar pentru sistemele de echilibrare a sarcinii va publica inițial o pagină HTML de notificare privind întreținerea.
2. Între timp, o procedură manuală va promova primul nod de baze de date din al doilea centru de date pentru a avansa la rangul de MASTER. Va fi configurată replicarea cu nodul secundar din al doilea centru de date, iar când totul este în ordine, vor fi pornite celelalte servicii.
3. Pagina de notificare privind întreținerea va fi înlocuită cu o redirecționare către serviciile Survey Solutions, astfel încât utilizatorii să poată accesa un mediu nou.

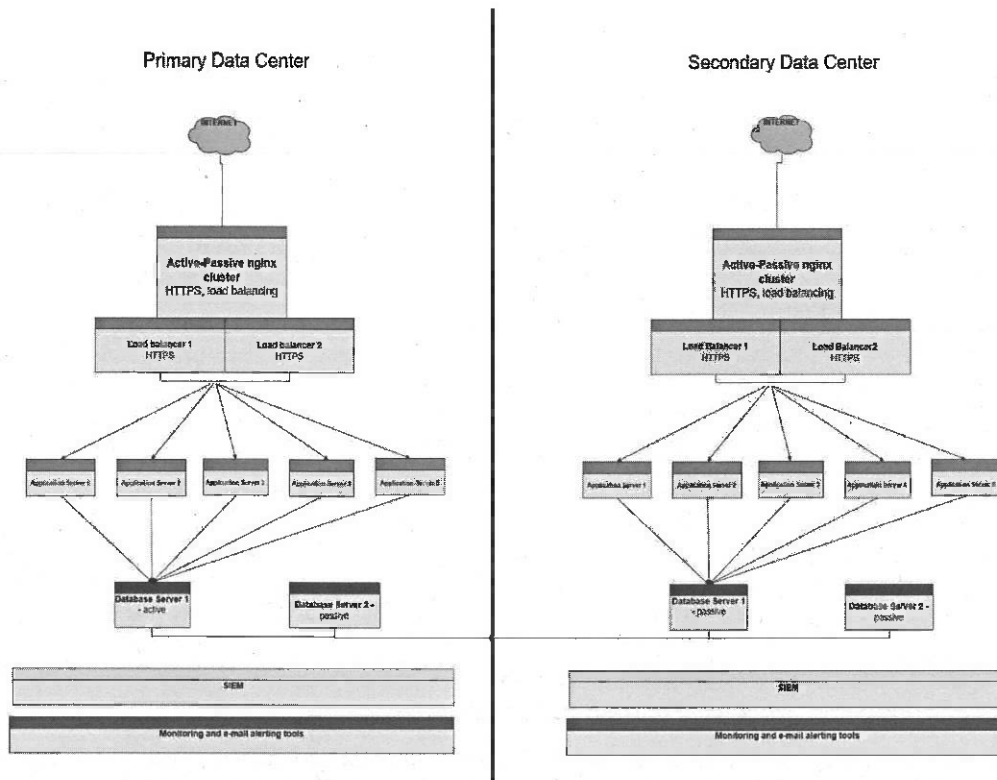
Este foarte importantă menținerea configurației în centrul de date secundar în aceeași stare ca în centrul de date principal, de așa natură încât de fiecare dată când configurația se modifică în centrul de date principal, aceasta să fie replicată în centrul de date secundar. Următoarele resurse au fost utilizate în centrul de date secundar pentru a acoperi capacitatea centrului de date primar:

Tabelul 2 – Resurse ale centrului secundar de date RPL

Denumire	Spațiu necesar	SO invitat	Memorie	Procesoare	Adresă IP
INS.BM-DR-APP01	1000 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-DR-APP02	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-DR-APP03	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-DR-APP04	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-DR-APP05	82.09 GB	Microsoft Windows Server 2019 (64-bit)	96 GB	24	
INS.BM-DR-DB01	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-DR-DB02	1050 GB	Ubuntu Linux 20.04 (64-bit)	256 GB	48	
INS.BM-DR-LB01	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.BM-DR-LB02	58.08 GB	Ubuntu Linux 20.04 (64-bit)	32 GB	16	
INS.VM-DR-SL	98.09 GB	Ubuntu Linux 20.04 (64-bit)	48 GB	16	

De asemenea, o stocare NFS dedicată asigură un spațiu de 2T pentru primul nod al INS.BM-DR-DB01.

Figura 8 – Configurarea inițială a Survey Solutions pentru RPL



Dispozitive virtuale (Obiectiv de testare)

- Sistem de echilibrare a sarcinii (Load Balancer)
 - 2x - 15vcpu, 32GB RAM
 - Linux Server (Ubuntu 20.04 LTS)
 - NGINX Load Balancer
- Servere de aplicații (Application Servers)
 - 5x - 24vcpu, 96GB RAM
 - Configurate cu Windows Server 2019
 - Survey Solutions HQ
- Database Server
 - 2x - 46vcpu, 256GB RAM
 - Windows Server ori Linux (Ubuntu 20.04 LTS)
 - PostgreSQL Server (version 12.4)

Dispozitiv de testare a sarcinii (sursa sarcinii)

- 3x 16vcpu, 48GB RAM
- Linux OS (Ubuntu 20.04 LTS)
- Docker

Rețea(le):

DMZ – 2 x Load Balancers

Privată – 5 x servere de aplicații, 1 x server de baze de date, 1 x testare a sarcinii

Porturi între rețele:

Internet - Load Balancers 80, 443 (port 80 va fi redirecționat către 443 după instalarea certificatului SSL)

DMZ – Internet 80, 443

Testare a sarcinii – dispozitive de echilibrare a sarcinii – 443, 9100, 8080

Dispozitive de echilibrare a sarcinii – servere de aplicații – 9700

Privată(e) – Internet 80, 443

Configurația detaliată a sistemului este prezentată în **Anexa 1 – Sistemul informatic de colectare a datelor RPL – Raportul conform cu execuția.**

Notă: Aceasta este configurația în sistemul gestionat de STS - informațiile din Anexa 1 vor fi utilizate pentru configurarea sistem IT integrat al INS în realizarea viitoarelor anchete structurale.

4. Monitorizare și raportare pe durata colectării datelor RPL

4.1. Monitorizare și raportare zilnică

Procesul de monitorizare și raportare pentru colectarea datelor a reprezentat o sarcină zilnică realizată pentru întreaga perioadă a RPL (15 martie – 31 iulie 2022), sub forma unui efort comun al echipelor INS și BM. Soluțiile furnizate au constat în scripturi create pentru generarea unor rapoarte de monitorizare a colectării datelor în cadrul fazei CAWI (15 martie – 27 mai) și al fazei CAPI (31 mai – 31 iulie); de asemenea, a colectării pe județe, UAT-uri, ARA și recenzori.

Procesul de monitorizare și raportare a utilizat datele obținute de la aplicația Survey Solutions, mai exact exportul de date privind activitățile înregistrate pe durata fazei CAWI, cu referire la chestionare generate, completate și transmise/validate și la activitățile realizate de ARA (asistenți de autorecenzare), precum și pe durata fazei CAPI, cu referire la activitățile realizate de recenzori prin completarea chestionarelor cu o zi înainte, la chestionarele deschise și la cele completate de către fiecare membru al personalului de recenzie cu roluri în Survey Solutions, precum și pe UAT-uri și pe județe.

Rapoartele CAWI generate de scripturile disponibile pe „*Serverul de Prelucrare a Datelor*”, la adresa /home/calex/RPL CAWI stat/daily stats v3.R, au fost structurate sub formă de:

- Raport zilnic de monitorizare a colectării datelor pe județe, ce conține numărul total de chestionare completate în zona urbană și în zona rurală, cu sau fără suportul ARA;
- Raport zilnic de monitorizare a colectării datelor pe localități (Unitate administrativ-teritorială), ce conține numărul total de chestionare completate;
- Raport zilnic de monitorizare a colectării datelor pe ARA, ce conține numărul total de chestionare validate sau nevalidate completate, județul și UAT în care a fost înregistrată activitatea asistentului; și
- Raport zilnic de monitorizare a tipurilor de erori și mesaje transmise respondenților pe durata procesului de autoînregistrare;

Procesul de monitorizare zilnică a furnizat datele pentru comunicările privind colectarea zilnică a datelor RPL, pe durata fazei CAWI, furnizate de către echipa INS, agregate pe săptămâni, cu referire la domiciliu (urban/rural) și metoda de autorecenzare (cu sau fără ARA).

Figura 9 - Exemplu de raport zilnic de date CAWI colectate (pe județe)

g_jud	g_jud_code	chestionare	urban	rural	ara	non_ara
ALBA	10	177877	109211	68666	107800	70077
ARAD	29	249046	147412	101634	153441	95605
BACAU	47	257419	128585	128834	156150	101269
BIHOR	56	373203	189769	183434	205204	167999

Figura 10 - Exemplu de raport zilnic de date CAWI colectate (pe UAT-uri)

g_jud_code	g_muni	g_muni_code	g_sat	g_sat_code	chestionare
10	BLANDIANA	3397	ACMARIU	3413	222
10	BLANDIANA	3397	BLANDIANA	3404	340
10	BLANDIANA	3397	POIENI	3431	2
10	BLANDIANA	3397	RACATAU	3440	22
10	BUCIUM	3459	BUCIUM	3468	75
10	BUCIUM	3459	BUCIUM-SAT	3495	96

Figura 11 - Exemplu de raport zilnic de date CAWI colectate (pe ARA)

jud	uat	codr	nr_chest_valide	nr_chest_invalide	cod_ara_corect
AB	MUNICIPIUL_ALBA_IULIA	001017 0001	395	8	1
AB	MUNICIPIUL_ALBA_IULIA	001017 0002	549	4	1
AB	MUNICIPIUL_ALBA_IULIA	001017 0003	1021	4	1

Figura 12 - Exemplu de raport zilnic cu tipuri de erori/mesaje înregistrate în timpul CAWI

interview_key	interview_id	roster	id1	variable	type	message_number	message
41-74-13-67	0872abc356014e8d	NA	NA	AA_SLR	2	2	Atenție! Anul 2021 nu poate fi mai mic decât 2021 !
38-43-83-92	bc3baa6a643b456f	NA	NA	AA_SLR	2	1	Atenție! Anul 2021 nu poate fi mai mic decât anul de naștere
83-59-52-06	ac723fe6f5044ccf9	NA	NA	AA_SLR	2	1	Atenție! Anul 2021 nu poate fi mai mic decât anul de naștere
04-64-09-99	0fee9b1a588f4a00	NA	NA	AA_SLR	2	1	Atenție! Anul 2021 nu poate fi mai mic decât anul de naștere

Rapoartele CAPI generate de scripturile disponibile pe „*Serverul de Prelucrare a Datelor*”, la adresa /home/calex/RPL CAPI stat/monitorizare date colectate Judete.Rmd, au fost structurate sub formă de:

- Un raport zilnic cuprinzător de monitorizare a colectării datelor, pe țară și pe județe, ce conține numărul total de chestionare completate în fiecare județ și de către fiecare recenzor/supervizor, precum și etapele în care se află chestionarele (completat, (re)alocat, respins/aprobat de Supervizor; respins/aprobat de INS central;

Figura 13 - Exemplu de raport zilnic de date CAPI colectate (țară, județ, recenzor)

County	ID Supervisor (SUP)	ID enumerator (INT)	Total no. of questionnaires assigned to enumerator	Total no. of questionnaires on server	Total no. of enumerated persons	out of which, by stages					
						Completed	Questionnaires re(allocated)	Rejected by supervisor	Approved by supervisor	Rejected by Headquarter	Approved by Headquarter
AB	SUP_AB_01	INT_10170001	283	450	336	0	0	0	316	0	134
AB	SUP_AB_01	INT_10170002	343	506	346	0	0	0	237	0	269
VS	SUP_VS_N1	P_167320_0005	2	0	0	0	0	0	0	0	0
		Total COUNTRY	6026618	7809721	8812645	73191	82	7076	3258433	68	4470871

Figura 14 - Exemplu de raport zilnic de date CAPI colectate (județ, recenzor)

County	ID Supervisor (SUP)	ID enumerator (INT)	Total no. of questionnaires assigned to enumerator	Total no. of questionnaires on server	Total no. of enumerated persons	out of which, by stages					
						Completed	Questionnaires re(allocated)	Rejected by supervisor	Approved by supervisor	Rejected by Headquarter	Approved by Headquarter
AB	SUP_AB_01	INT_10170001	283	450	336	0	0	0	316	0	134
AB	SUP_AB_01	INT_10170002	343	506	346	0	0	0	237	0	269
AB	SUP_AB_01	INT_10170003	324	383	515	0	0	0	187	0	196
AB	SUP_AB_01	INT_10170004	322	498	238	0	0	0	210	0	288

În cadrul aceluiași proces de monitorizare a activităților RPL și a rezultatelor specifice, pe recenzori sau pe durate de completare a chestionarului, analiza paradelor poate oferi informații valoroase. În acest sens, se recomandă folosirea unei aplicații cu scop special disponibile pentru personalul INS (nu necesită date de autentificare), la adresa phc_censustools de pe *Serverul de Prelucrare a Datelor*.

În urma procesului de monitorizare și a distribuirii sarcinilor, informațiile generate au fost consolidate în rapoarte zilnice care au permis echipei INS de coordonare a RPL să informeze și să facă recomandări către DTS, recenzori, supervizori, coordonatori locali și observatori în legătură cu situația colectării datelor și nivelul de acoperire.

Machetele rapoartelor sunt disponibile ca arhive în format electronic care însoțesc prezentul raport ca **Anexa 2 - Exemple de rapoarte de monitorizare zilnică (CAWI, CAPI)**.

4.2. Monitorizarea și raportarea activității personalului de recenzare în vederea plății

Procesul de raportare pe marginea activității personalului de recenzare a cuprins două dimensiuni: una axată pe raportarea activității, cealaltă având drept obiectiv calculul sumelor brute de plătit personalului de recenzare.

Monitorizarea colectării datelor (folosind exporturi zilnice din Survey Solutions) a fost mijlocul de coordonare a activităților la nivel de recenzori, supervizori, coordonatori, precum și în procesele administrative referitoare la remunerarea activității desfășurate de aceștia. Informațiile privind numărul de chestionare completate, situația completării pe secțiuni – unitate locativă, persoană, clădire – a chestionarelor a asigurat informațiile de bază pentru monitorizarea și coordonarea recenziilor și, ulterior, pentru plata efortului depus de aceștia (a se vedea exemplul de raport pe recenzori și pe județe/UAT-uri în figurile de mai jos).

Figura 15 - Exemplu de raport de plată pentru activitatea recenzorului

Payrol statement - enumerator period May 31– July 31, 2022)								
Situația de plată - pentru recenzori (perioada 31 mai – 31 iulie 2022)								
ATU name		ATU SIRUTA						
Denumire UAT SLATINA		SIRUTA UAT		125347				
ATU category (locality type 1 - usual; type 2 - isolated or with territorial dispersion)					Enumerator ID			
Tip localitate (localitate1- obișnuită; 2-izolată sau dispersată teritorial)					1			
					Cod recenzor (user recenzor)			
					INT_123456789			
(Date) Data	(Password) Parola	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR DWELLING (LC) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA LOCUINTEI (LC) COMPLICATE	(Rate (LC) Tarif LC (Lei)	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR PERSON (P) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA PERSOANEI (P) COMPLICATE	Rate (P) Tarif P (Lei)	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR BUILDING (SC) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA SPAȚIULUI COLECTIV DE LOCUIT (SC) COMPLICATE	Rate (SC) Tarif SC (Lei)	Total ammount for payment Sumă totală de plată (Lei)
A	B	1	2 = 1 x 3.5Lei	3	4 = 3 x 7Lei	5	6 = 5 * 3.5Lei	7

Figura 16 - Exemplu de raport de plată pentru activitatea recenzorului, pe județe și UAT-uri

Payrol statement - enumerators period May 31- July 31, 2022									
Situația de plată - pentru recenzori (perioada 31 mai – 31 iulie 2022)									
County	JUDET								
Judet									
(ATU name) Denumire UAT	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR DWELLING (LC) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA LOCUINTEI (LC) COMPLICATE	(Rate (LC) Tarif LC (Lei)	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR PERSON (P) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA PERSOANEI (P) COMPLICATE	Rate (P) Tarif P (Lei)	(From valid questionnaires eligible for payment.) Din chestionarele validate eligibile pentru plată: (TOTAL NO. OF SECTIONS FOR BUILDING (SC) ENUMERATION COMPLETED) TOTAL SECȚIUNI PENTRU RECENZAREA SPAȚIULUI COLECTIV DE LOCUIT (SC) COMPLICATE	Rate (SC) Tarif SC (Lei)	Total ammount for payment Sumă totală de plată (Lei)	ATU category Tip localitate	ATU Siruta UAT
6	1	2 = 1 x Tarif LC Lei	3	4 = 3 x Tarif P Lei	5	6 = 5 * Tarif SC Lei	7	8	9

Calculul sumei totale remunerate unui recenzor și sumele totale distribuite către UAT-uri și județe prin alocări de la Bugetul de Stat s-a realizat prin scripturile care sunt disponibile pe „*Serverul de Prelucrare a Datelor*”, la adresele [/home/calex/RPL_CAPI_stat/situatii_plata_CAPI_v5.R](#); [/home/calex/RPL_CAPI_stat/situatii_plata_CAPI_zile_v2.R](#)

Modelele de rapoarte sunt prezentate în **Anexa 3 - Exemple de rapoarte de plăți aferente CAPI** și sunt furnizate ca arhive în format electronic.

5. Procesele de validare a datelor colectate

Scopul procesului de validare a fost de a susține nivelul de pregătire al datelor recensământului în vederea prezentării rezultatelor preliminarilor și transmiterii lor către Eurostat, precum și de a disemina rezultatele către alți utilizatori (naționali / internaționali): publicații (tabele predefinite) și tabele generate în urma unor solicitări ad-hoc. Această secțiune prezintă informații despre proces și instrumentele de validare a datelor colectate.

Prelucrarea datelor recensământului a constat în validarea datelor, rectificarea erorilor, deduplicare, imputarea unor date lipsă sau incomplete, calculul indicatorilor statistici incluși în producția rezultatelor recensământului. De asemenea, toate activitățile au urmărit prelucrarea datelor cu caracter personal într-o manieră menită să le asigure securitatea corespunzătoare. Procesul de colectare a datelor recensământului presupune o serie de pași prin care se asigură integritatea și acuratețea datelor: Colectarea Datelor, Prelucrarea Datelor, Editarea Datelor, Imputarea Datelor, Analiza Divulgării Datelor (SDC), Controlul Calității Datelor, Publicarea Datelor (Diseminare).

Utilizarea metodelor CAWI și CAPI pentru colectarea datelor în recensământ a oferit avantaje pentru corectitudinea datelor, în special datorită regulilor de validare incluse în elaborarea chestionarului și a validării în timp real din faza de colectare a datelor. De asemenea, recenzorii au fost instruiți să respecte proceduri stricte pentru a se asigura că informațiile colectate sunt exacte și complete.

5.1. Prelucrarea datelor pe durata colectării datelor CAWI, CAPI

În timpul și ulterior colectării datelor, INS a prelucrat datele pentru a se asigura că acestea sunt exacte și complete. Activitatea a inclus identificarea erorilor și a neconcordanțelor și asigurarea faptului că toate datele sunt în formatul corect.

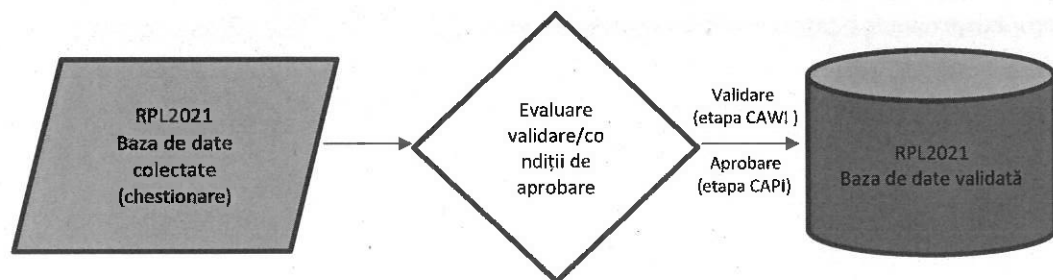
Reducerea riscului de erori de intrare (greșeli de ortografie) în timpul completării chestionarelor s-a realizat prin reguli avansate de navigare și validare în cadrul fazei de elaborare a chestionarului. Capacitățile platformei de colectare a datelor a oferit posibilitatea respectării condițiilor de validare din chestionar, atât pentru metoda CAWI, cât și pentru metoda CAPI.

Diferența dintre conceptul chestionarului CAWI și cel al chestionarului CAPI a constat în afișarea informațiilor de clarificare a întrebărilor și răspunsurilor, fiind adaptate persoanelor care se autorecenzau, în cazul CAWI, și recenzorilor, în cazul CAPI.

În cadrul chestionarelor CAWI și CAPI, au fost implementate reguli de validare pentru greșeli de ortografie, pentru intercorelări între răspunsuri și condiții logice referitoare la fluxul de lucru din chestionar. O metodă de a reduce numărul greșelilor de ortografie în CAWI a fost utilizarea pe cât posibil a listelor în locul căsuțelor pentru inserare de text liber.

Fluxurile de prelucrare a datelor au fost similare, însă procesul CAWI s-a încheiat cu validarea chestionarului întrucât procesul CAPI promise validarea și se încheiase cu aprobarea chestionarului (a se vedea figura de mai jos).

Figura 17 – Fluxuri automate pentru validare (faza CAWI) / aprobare (faza CAPI)



În etapa CAWI a recensământului au fost implementate regulile de validare automată, urmărindu-se caracterul complet al datelor colectate. Zilnic, pe baza datelor colectate disponibile, algoritmul automat de validare s-a aplicat pe toate chestionarele complete/completate (disponibile pe „*Serverul de Prelucrare a Datelor*” la adresa /home/calex/RPL CAWI stat/validari v3.R și însoțind în format electronic acest report, **Anexa 4: Scripturi de validare și aprobare a colectării datelor RPL (CAWI, CAPI).**

După această validare, pe chestionarelor pentru persoane, a fost generată o listă a persoanelor care au finalizat autorecenzarea cu succes și trimisă platformei web, care a emis dovada autorecenzării. Algoritmul automat a fost elaborat de către echipa de recenzie a INS și implementat ca scripturi R. A fost implementată o metodă de identificare a chestionarelor multiple completate de o singură persoană, fiind păstrat în baza de date chestionarul cel mai recent completat corect și integral.

În etapa CAPI au fost implementate reguli de validare automată pentru identificarea erorilor și a caracterului complet al datelor colectate. Regulile de validare sunt mai stricte decât cele din etapa CAWI deoarece recenzorii sunt instruiți să completeze chestionarul și să determine gradul de corectitudine al datelor colectate. Un proces automat pentru aprobarea chestionarelor corecte este implementat sub formă de script R, pe serverul analitic, și este rulat zilnic pe toate datele colectate (disponibil pe „*Serverul de Prelucrare a Datelor*” la adresele /home/calex/RPL CAPI stat/validari CAPI v2.R și /home/calex/RPL CAPI stat/aprobări automate RPL v1.10.R, însoțind în format electronic acest raport, **Anexa 4 - Scripturi de validare și aprobare a colectării datelor RPL).**

Pe baza rezultatelor aplicării algoritmilor în validarea și aprobarea chestionarelor, au fost întocmite zilnic rapoarte de stare privind chestionarele, recenzorii și supervizorii.

5.2. Prelucrarea datelor ulterior colectării datelor

Fazele descrise mai jos sunt componente ale prelucrării post-recensământ a datelor.

Editarea datelor - echipa INS pentru pregătirea și prelucrarea datelor folosit o varietate de proceduri de editare pentru a identifica și corecta erorile din date, proceduri care au inclus identificarea neconcordanțelor, a valorilor lipsă și a valorilor din afara intervalului admis.

Imputarea datelor - în cazul datelor lipsă, echipa INS a folosit surse administrative pentru completarea datelor și metode de imputare statistică pentru a estima valori. Această abordare a facilitat asigurarea faptului că datele finale sunt cât mai complete și exacte posibil.

Analiza divulgării datelor - echipa INS a examinat datele pentru a se asigura că acestea nu divulgă eventuale informații despre respondenți individuali. În acest fel se protejează confidențialitatea respondenților cu privire la datele diseminate. Metode de Control privind divulgarea informațiilor statistice sunt aplicate microdatelor brute pentru a pregăti microdatele finale în vederea diseminării. Metoda propusă pentru SDC, schimbul de înregistrări (en. ‘*record swapping*’), a rezultat din asistența tehnică oferită și din cursurile de formare în acest domeniu la care au participat experții INS.

Controlul calității datelor - echipa INS a efectuat controale cuprinzătoare ale calității la nivelul datelor pentru a asigura acuratețea și caracterul complet al acestora. De asemenea, echipa a efectuat evaluări independente pentru a evalua calitatea datelor.

Publicarea datelor (diseminare) - ca ultimă fază, echipa INS face publice datele. De asemenea, echipa se asigură că pune la dispoziție documentația și metadatele necesare pentru a-i ajuta pe utilizatori să înțeleagă și să utilizeze corect datele.

6. Editarea și producția de date în scopul diseminării

Diseminarea datelor RPL cuprinde trei soluții specifice: hipercuburi de date, care se publică în mod obligatoriu pe portalul Eurostat, distribuția pe griduri a datelor RPL, ce permite diseminarea unui număr de 21 de indicatori (13+8) pe griduri (INSPIRE) de 1 km² pentru întreg teritoriul României, și diseminarea tabelor dinamice cu rezultate RPL. Deoarece echipa INS gestionează hipercuburile de date (pregătind mediul IT și instalând aplicațiile necesare furnizate de Eurostat cu scopul de a prelua date din antrepozitul de date RPL organizat în mediul IT al INS), această secțiune prezintă distribuția pe griduri și tablele dinamice dezvoltate în cadrul proiectului.

6.1. Distribuția pe griduri a datelor RPL

Gridurile de populație reprezintă un instrument puternic în înțelegerea și analiza caracteristicilor și tiparelor populațiilor umane. Ele permit cercetătorilor să studieze interacțiunile între activitățile umane și mediul înconjurător și să înțeleagă cauzele de la baza fenomenelor care depășesc granițe administrative². De exemplu, gridurile de populație sunt utile când se studiază efectele inundațiilor, ale navei și ale extinderii urbane asupra comunităților. Folosind griduri de populație pentru a analiza datele recensământului, cercetătorii pot înțelege mai bine detalii și nuanțe din datele demografice, comportamente și necesităților diferitelor populații și pot folosi aceste informații pentru a informa politici și procese decizionale. În plus, gridurile de populație pot fi un instrument util în evaluarea distribuției și disponibilității resurselor, precum locuințele, sănătatea și educația, și în identificarea domeniilor care ar putea necesita suport sporit sau intervenții. Pe scurt, transformarea datelor recensământului în griduri de populație poate oferi numeroase perspective ale dinamicii și complexității societății umane și ne poate ajuta să înțelegem mai bine și să gestionăm mai eficient provocările comunităților noastre.

Eurostat diseminează griduri de populație în cadrul eforturilor sale de a furniza date detaliate și exacte privind aspectele demografice, comportamentele și necesitățile populației europene. Aceste griduri de populație sunt create folosind date din recensămintele de populație și locuințe ale Uniunii Europene și urmăresc să contureze un tablou detaliat și coerent al populației la nivel de celule individuale, acoperind în mod obișnuit o suprafață de 1 km².

Gridurile de populație diseminate de Eurostat folosesc o varietate de variabile pentru a descrie caracteristicile populației din fiecare celulă, inclusiv caracteristici demografice precum vârsta, sexul și naționalitatea, precum și informații referitoare la locuință, educație, locul de muncă și alți factori socioeconomi. Aceste variabile sunt folosite pentru a crea profiluri detaliate ale populației din fiecare celulă, permițând cercetătorilor să înțeleagă caracteristicile și necesitățile diferitelor populații la nivel granular.

SCR utilizat de gridurile de populație ale Eurostat este proiecția ETRS89-LAEA (EPSG:3035), o variantă a Sistemului de Referință Terestru European (ETRS89) optimizat pentru utilizare în Europa. Această proiecție este utilizată pe scară largă în Europa pentru cartografiere și analiză spațială, fiind deosebit de bine adaptată aplicațiilor care presupun cartografiere de amploare și vizualizarea datelor dincolo de granițe naționale sau regionale (detalii în caseta cu informații).

Setul de instrumente creat pentru INS în sprijinul transformării datelor recensământului din România în date privind gridurile de populație conține două componente:

- i. Un script R care transformă datele în unități statistice de tip grid și le stochează într-o bază de date PostGIS-PostgreSQL.
- ii. O aplicație Shiny bazată pe script R pentru a vizualiza diferitele griduri de populație.

² https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_grids

Proiecția ETRS89-LAEA este o metodă utilizată pentru a cartografia localizări de pe suprafața Pământului sub forma unui plan plat. Este frecvent utilizată în Europa și deosebit de bine adaptată pentru cartografierea unor zone vaste la scări mici până la medii. În termeni simpli, proiecția este un mod de a lua suprafața curbată a Pământului și de a o aplatiza în forma unui plan bidimensional, cum ar fi o hartă. Proiecția ETRS89-LAEA este un tip specific de proiecție cunoscut pentru capacitatea sa de a păstra formele corecte ale suprafețelor întinse, minimizând în același timp distorsiunea dimensiunilor. Proiecția LAEA este cunoscută și pentru capacitatea sa de a păstra exactitatea măsurătorilor direcționale și angulare.

Script R pentru transformare

Pentru a face datele disponibile sub formă de griduri de populație, conform cerințelor Eurostat³, precum și pentru a minimiza timpii de procesare pentru vizualizare în aplicația Shiny, datele sunt pre-procesate într-un grid de populație ce conține 21 de variabile, dintre care 13 variabile sunt cele impuse de Eurostat, în timp ce 8 sunt considerate de către INS relevante în contextul României.

Cele 13 variabile impuse de Eurostat sunt:

Tabelul 3 – Categoriile temei „recensământ” în Eurostat

	Categoriile tematice ale recensământului pentru defalcare pe gridul de 1 km ² de referință	STAT. G.
0.	SEX.0.: Populație totală	0.
1.	SEX.1.: Bărbat	1.
2.	SEX.2.: Femeie	2.
3.	AGE.G.1.: Sub 15 ani	3.
4.	AGE.G.2.: Între 15 și 64 de ani	4.
5.	AGE.G.3.: 65 de ani și peste	5.
6.	CAS.L.1.1.: Persoane angajate ¹¹	6.
7.	POB.L.1.: Locul nașterii în țara de raportare	7.
8.	POB.L.2.1.: Locul nașterii într-un alt Stat Membru al UE	8.
9.	POB.L.2.2.: Locul nașterii în altă parte	9.
10.	ROY.1.: Domiciliul permanent cu un an înainte de recensământ și neschimbat	10.
11.	ROY.2.1.: Domiciliul permanent cu un an înainte de recensământ: mutare în granițele țării de raportare	11.
12.	ROY.2.2.: Domiciliul permanent cu un an înainte de recensământ: mutare din afara granițelor țării de raportare	12.

³ Eurostat. Legislația UE privind recensămintele populației și ale locuințelor din 2021: Note explicative. Luxembourg: Oficiul pentru Publicații al Uniunii Europene; 2019, <https://ec.europa.eu/eurostat/documents/3859598/9670557/KS-GQ-18-010-EN-N.pdf/c3df7fcb-f134-4398-94c8-4be0b7ec0494>; datele privind categoria „persoane angajate” vor fi transmise, în măsura în care acest lucru este posibil, în funcție de disponibilitatea acestora în Statul Membru care transmite

Cele 8 variabile suplimentare stabilite de INS sunt:

Tabelul 4 – Categoriile suplimentare (8) ale temei „recensământ” impuse de INS

	Categoriile tematice ale recensământului pentru defalcare pe gridul de 1 km ² de referință	STAT. G.
13.	EDU.1: Populație cu nivel scăzut al educației	0.
14.	EDU.2: Populație cu nivel mediu al educației	1.
15.	EDU.3: Populație cu nivel ridicat al educației	2.
16.	EDU.1.1.: Analfabetism	3.
17.	INA.1: Populație inactivă din punct de vedere economic	4.
18.	INA.1.1.: Populație inactivă din punct de vedere economic – pensionari	5.
19.	INA.1.2.: Populație inactivă din punct de vedere economic – elevi, studenți	6.
20.	INA.1.3.: Populație inactivă din punct de vedere economic – alte categorii inactive	7.

Importantă pentru prelucrarea fără obstacole a acestor variabile este calitatea datelor de intrare, care este subliniată prin următoarele detalii:

- un singur fișier .csv,
- un singur ID pentru fiecare unitate,
- cu toate celor 21 de variabile,
- plus latitudinea și longitudinea prezente,
- rezultă un fișier cu exact 23 de coloane,
- fără nicio valoare lipsă la nivelul variabilelor de interes,
- fără coordonate lipsă,
- fără coordonate din afara țării,
- cu toate variabilele numerice cu excepția ID,
- și cu SDC (spațial) aplicat.

Toate variabilele utilizate în acest proces necesită transformare pentru a reprezenta date contorizabile, însemnând că variabila conține 1 dacă persoana aparține unei anumite grupe de populație și 0 în caz contrar.

În mod oficial, cele de mai sus sunt scrise ca funcție de indicator:

$$1_{group}(x) = [x \in group]$$

Ulterior, datele curățate și transformate vor fi transformate într-un obiect spațial, iar coordonatele sale geografice din notația de latitudine-longitudine (EPSG:4326) vor fi transformate în proiecția ETRS89-LAEA descrisă mai sus (EPSG:3035) în vederea creării codurilor de grid.

Aplicația Shiny

Pentru vizualizarea datelor produse în etapa anterioară a fost dezvoltată o aplicație Shiny, pusă apoi la dispoziția INS pe serverul său de analiză.

Figura 18 – Aplicație pe bază de R Shiny pentru vizualizarea gridurilor de populație.



Aplicația dispune de următoarele funcționalități:

- Selectarea județului (provinciei).
- Selectarea stratului (variabilei) de interes.
- Vizualizarea gridului selectat de populație.
- Mutarea cursorului pe o singură celulă de grid afișează o fereastră pop-up cu informațiile corespunzătoare ale celulei de grid.
- Descărcarea gridului selectat de populație ca fișier .pdf..

Aplicația poate fi accesată prin următoarea adresă URL: 10.2.106.10/phc_grid/⁴. Pentru a crea și vizualiza hărțile interactive, se utilizează pachetul R mapdeck⁵ (*pachet de hărți*), pentru care se pot vedea detalii în caseta corespunzătoare cu informații.

De ce pachetul R mapdeck?

Vizualizarea seturilor de date cu poligoane spațiale poate fi o provocare, deoarece poligoanele se pot suprapune sau pot fi prea numeroase pentru a fi efectiv afișate. Se poate ajunge astfel la vizualizări aglomerate și ilizibile. De asemenea, seturi mari de date ar putea necesita prea multe resurse pentru a fi încărcate și randate într-un navigator web sau o aplicație cartografică.

Pachetul R mapdeck poate facilita vizualizarea unor seturi mari de date ce conțin poligoane spațiale prin oferirea unui set de instrumente pentru crearea hărților interactive și online. Acesta folosește biblioteca Mapbox GL JavaScript și Deck.gl pentru a randa hărți, permițând randarea rapidă și eficientă a seturilor mari de date. Pachetul oferă și o serie de caracteristici care pot ajuta la reducerea aglomerației și spori lizibilitatea vizualizării, precum capacitatea de a filtra date, a alocă coduri de culori poligoanelor în funcție de diferite atribute și a ajusta transparența poligoanelor. De asemenea, pachetul permite utilizatorilor să personalizeze aspectul vizual al hărții printr-o gamă largă de opțiuni stilistice.

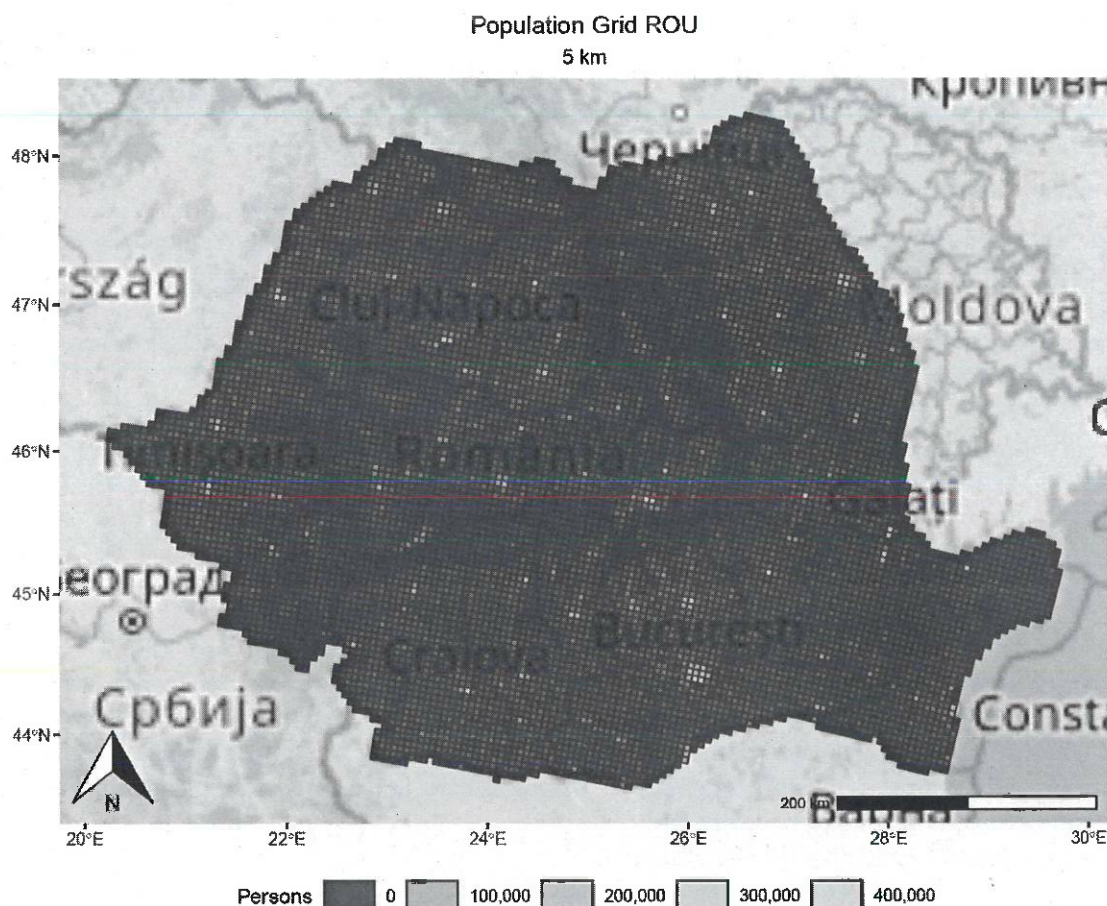
Pentru a folosi pachetul R mapdeck este necesar un abonament gratuit Mapbox. Un abonament valid se poate obține aici: <https://www.mapbox.com/>. Pentru detalii despre Deck.gl, accesați: <https://deck.gl/#/>.

⁴ Această adresă URL se bazează pe configurația internă de rețea a INS de la data întocmirii raportului. Orice modificare viitoare a acestei configurații poate afecta și adresa IP a gazdei.

⁵ <https://symbolixau.github.io/mapdeck/index.html>

După selectarea județului și a stratului corespunzător de populație, harta va fi afișată în forma ilustrată de **Error! Reference source not found.** După ce harta se încarcă, este posibilă descărcarea unui fișier .pdf al hărții afișate pe ecran. Aceasta este prezentată în **Error! Reference source not found.2.**

Figura 19 – Hartă în format .pdf disponibilă pentru descărcare.



Codul sursă este distribuit în plus față de fișierele standard ale aplicației R Shiny, cuprinzând:

- fișierul **ui.R**, care conține codul UI,
- **server.R**, care conține codul pe parte de server,
- **style.css**: stilurile CSS pentru UI
- **helpers** (*asistenți*): un director cu funcții ale asistenților
- **www**: pentru logouri etc.

Aceste fișiere vor fi copiate într-un director (denumit în prezent **phc-grid**) de la adresa **/srv/shiny-server** pentru a putea fi rulată aplicația. Toate fișierele sunt furnizate ca atare, însă, întrucât INS are acces la codul sursă, precum și la o echipă de programatori R competenți, INS are libertatea de a modifica acel cod în oricare formă preferată.

6.2. Diseminarea (diseminare tabelată) a datelor RPL

Diseminarea tabelată reprezintă componenta cheie pe care INS o va folosi pentru a asigura disponibilitatea datelor recensământului către publicul larg. Într-o anumită măsură, interfața de diseminare va înlocui diseminarea tabelată utilizată la recensământul din 2011, când toate tabelele au fost generate în prealabil. Abordarea propusă asigură un nivel de flexibilitate în ce privește exportul datelor și asigurarea instrumentelor adecvate pentru solicitări terțe de date. În plus, interfața va oferi o vizualizare cartografică minimă a datelor, specifică anumitor limitări tehnice (descrise mai jos). Interfața de diseminare este creată cu ajutorul limbajului de programare R și al cadrului {shiny} și a fost testată exclusiv pentru uz intern.

Cadru general și logică

Interfața de diseminare utilizează interfața Eurostat drept punct de plecare și structurează datele pe bază de teme, subteme și niveluri de dezagregare. Arborele de navigare al datelor pornește cu 8 teme de bază. Lista temelor și a subtemelor a fost întocmită de către INS și a ținut cont de necesitățile naționale de diseminare și de categoriile de date, de cerințele Eurostat și cerințele tehnice. Lista include următoarele teme:

1. CARACTERISTICI DEMOGRAFICE
2. CARACTERISTICI ETNO-CULTURALE
3. CARACTERISTICI EDUCAȚIONALE
4. MOBILITATEA POPULAȚIEI REZIDENTE
5. CARACTERISTICI ECONOMICE
6. GOSPODĂRII
7. CLĂDIRI REZIDENȚIALE
8. CONDIȚII DE LOCUIT ȘI DE TRAI

Fiecare temă este împărțită în subteme, rezultând un total de 145 de subteme, inclusiv o serie de dezagregări teritoriale și la nivel de indicatori. De exemplu, subtema „Populația rezidentă pe sexe și grupe de vârstă” este inclusă în tema „Caracteristici demografice”. De asemenea, subtema include dezagregări la diverse niveluri teritoriale și pentru fiecare indicator, precum și tabelările lor încrucișate, conform descrierii de mai jos:

1. CARACTERISTICI DEMOGRAFICE
 - a. Populația rezidentă pe sexe și grupe de vârstă
 - i. Niveluri teritoriale disponibile
 1. Național
 2. Macroregiuni
 3. Regiuni de dezvoltare
 4. Județe
 5. Localități
 - ii. Dezagregare a indicatorilor
 1. Sex (Total / Bărbați / Femei)
 2. Grupe de vârstă (Total, 0-4, 5-9, 10-14, 15-19, 29-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+)

Baza de date rezultată este deosebit de complexă și include un număr ridicat de variabile. Drept urmare, cadrul general este clădit pe un proces de selecție ierarhică care mai întâi selectează sursa principală de date, filtrează și alege variabilele de grupare, filtrează indicatorii necesari și apoi generează un cadru de date preprocesate folosit în scopuri de diseminare și vizualizare.

Cerințe tehnice și componente software utilizate

Similar distribuției pe griduri și a aplicării gridurilor pentru RPL, descrise mai sus, platforma de diseminare va folosi aceeași infrastructură și aceeași configurație descrise în Rezultatul 4.1b, fără a fi necesare configurații suplimentare. Platforma necesită rularea unor instalări de R și {shiny} (versiune cu sursă deschisă). Serverul RStudio nu este în mod obligatoriu necesar pentru implementare, însă este recomandat pentru a permite modificări și actualizări rapide.

În urma instalării, aplicația shiny trebuie să fie mutată către directorul de pe web. În cazul unei instalări normale, aceste directoare de aplicații ce conțin toate fișierele și datele ar fi localizate la adresa `/srv/shiny-server/`. Însă, având în vedere fișierele de mari dimensiuni utilizate de echipa INS și configurația resurselor pentru această instalare specifică, directorul de server a fost mutat la adresa `/arhiva1/srv/shiny-server/`⁶. Copierea și lipirea directoarelor la adresa `/arhiva1/srv/shiny-server/phc_diseminare/` va face aplicația disponibilă pe *Serverul de Prelucrare a Datelor*, în rețeaua locală, la adresa: http://xx.xxx.xx/phc_diseminare/.

Conceptul soluției și fluxul de proces

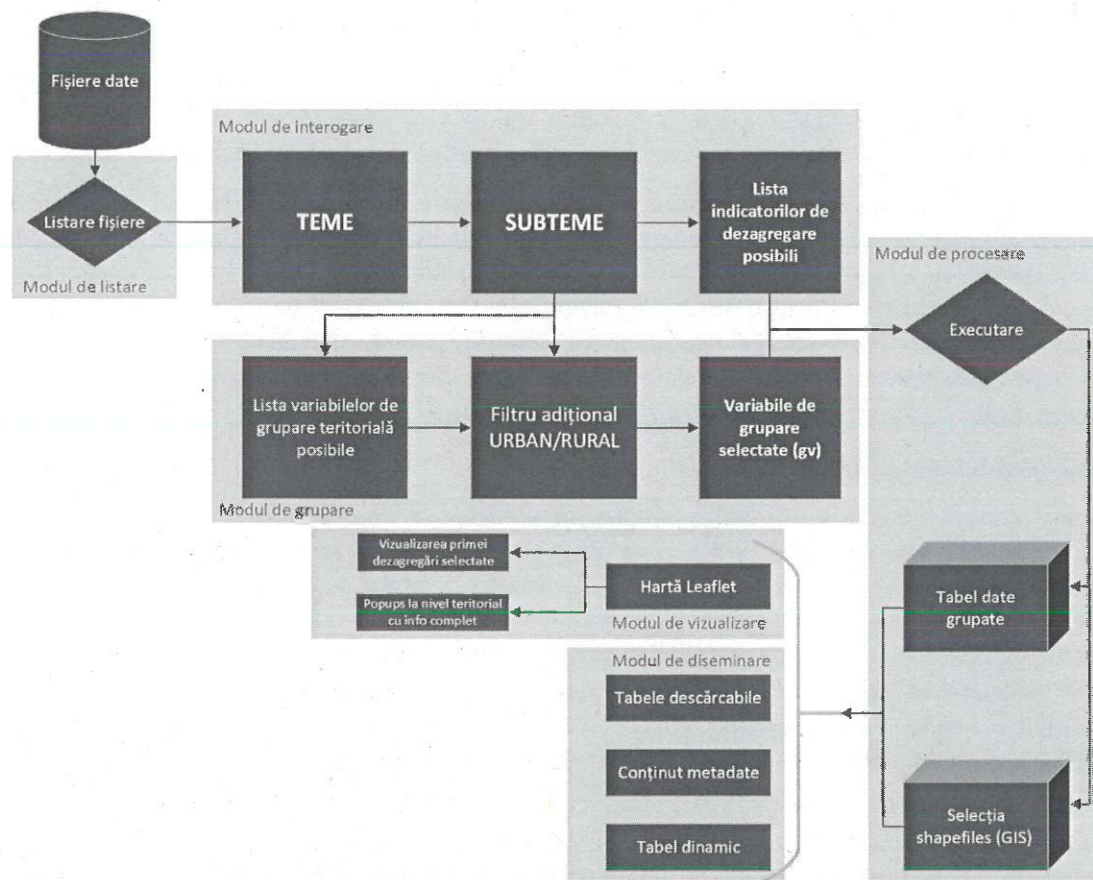
Soluția propusă integrează mai multe componente (definite ca module) care se interconectează și creează un sistem de interogare de fișiere ce poate genera mai mulți indicatori pe baza unui set de sugestii standardizate. Componenta principală de date a platformei de diseminare este colecția de tabele generate în prealabil de INS, cu toate calculele incluse și SDC aplicat la cel mai mic nivel de unitate administrativ-teritorială disponibil pentru fiecare subtemă. Datele sunt încărcate pe platformă, trecând prin fiecare modul pentru a ajunge în etapa de vizualizare și diseminare. Principalele module includ modulul de listare, modulul de interogare, modulul de grupare, modulul de prelucrare, modulele de vizualizare și diseminare. Procesul general este descris în diagrama de mai jos.

Platforma de diseminare listează fișierele de date din directorul sursă cu ocazia primei încărcări. În continuare, aceasta creează o listă de nume corelate cu un set predefinit de teme și subteme, generând temele și subtemele disponibile în date. Mai departe, pe baza selecției de subteme, modulul de interogare generează indicatorii disponibili, nivelurile de dezagregare disponibile și variabilele de grupare disponibile. La final, pe baza combinației de indicatori, variabile de dezagregare și de grupare, utilizatorul transmite interogarea către modulul de prelucrare. Acest modul rulează interogarea și generează două seturi de date: un cadru de date grupate bazat pe selecțiile utilizatorului și fișierele necesare shapefile pentru randarea datelor.

⁶

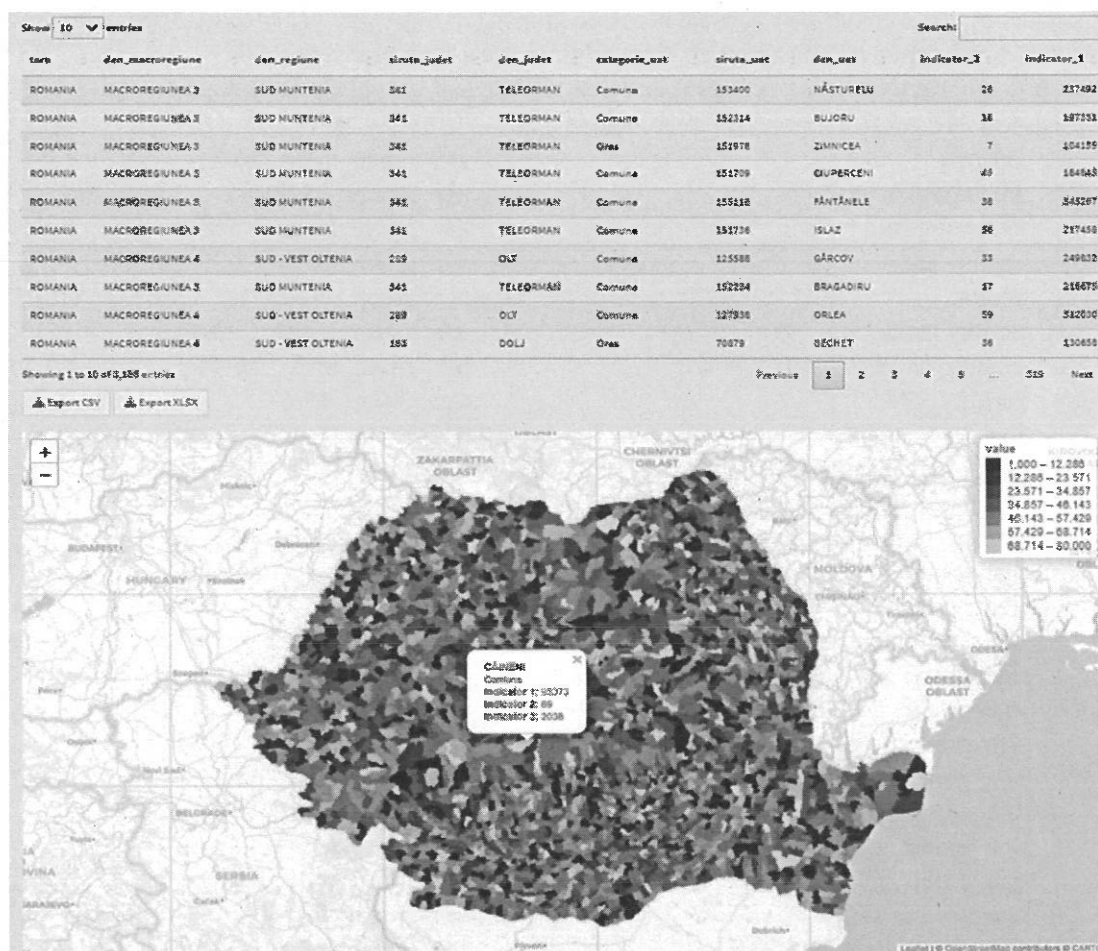
Recomandăm INS să utilizeze instalarea implicită conform recomandărilor Rstudio (cu denumirea nouă Posit) și documentației R.

Figura 20 – Fluxul procesului de diseminare



Pornind de la datele generate, modulul de vizualizare randează indicatorul selectat pe hartă folosind cadrul Leaflet. Întrucât nu pot fi randati simultan indicatori multipli, harta randează doar primul indicator din selecție. Harta folosește OpenStreet Map pentru imaginile de fundal, fiind cea mai puțin complexă soluție, ce nu necesită configurare suplimentară. De asemenea, pentru fiecare UAT/formă, harta va randa o fereastră pop-up cu sinteza selecției de indicatori. Harta permite selectarea fiecărei forme, apropierea în prim plan și panoramare. Odată cu harta, platforma generează un tabel ce conține datele generate de selecția utilizatorului și linkuri de descărcare pentru formatele CSV și XLSX de date, precum și o opțiune de salvare a hărții în format .png sau .pdf format (a se vedea exemplul ilustrativ de mai jos).

Figura 21 – Format de date diseminate salvat



Cerințe și condiții în materie de date

Fișierele cu tabele de date constituie baza platformei de diseminare. Dat fiind gradul ridicat de variație și complexitate al temelor și subtemelor obligatorii, fișierele generate de INS (145 de fișiere, câte unul pentru fiecare subtemă) ca bază a platformei de diseminare trebuie să fie standardizate și să urmeze aceeași schemă de fiecare dată. În plus, INS trebuie să furnizeze și să actualizeze permanent fișierele shapefile necesare, inclusiv denumirile și codurile pentru toate nivelurile de UAT, toate regiunile și macroregiunile.

Lista minimă a coloanelor care vor fi incluse într-un fișier cu tabel de date este următoarea:

- country (*țară*)
- macroregion_code (*cod macroregiune*) (NUTS)
- macroregion (*macroregiune*)
- region_code (*cod regiune*) (NUTS)
- region (NA dacă nu este cazul)
- county_code (*cod județ*) (SIRUTA, NA dacă nu este cazul)
- county_name (*denumire județ*) (NA dacă nu este cazul)
- locality_code (*cod localitate*) (SIRUTA_SUP, NA dacă nu este cazul)
- locality_name (*denumire localitate*) (NA dacă nu este cazul)

- locality_category (*categorie localitate*) (NA dacă nu este cazul)

De asemenea, toate coloanele relevante necesare grupării/dezagregării vor fi incluse pentru fiecare subtemă, și vor fi furnizate instrucțiuni clare privind datele lipsă și nivelurile 0.

În scop demonstrativ, pentru exemplul de temă utilizat mai sus, „Populația rezidentă pe sexe și grupe de vârstă”, coloanele prevăzute vor include:

- | | |
|--|--------------------------|
| – country | – age_group_male_30-34 |
| – macroregion_code | – age_group_35-39 |
| – macroregion | – age_group_female_35-39 |
| – region_code | – age_group_male_35-39 |
| – region | – age_group_40-44 |
| – county_code | – age_group_female_40-44 |
| – county_name | – age_group_male_40-44 |
| – locality_code | – age_group_45-49 |
| – locality_category | – age_group_female_45-49 |
| – gender_total (<i>total sexe</i>) | – age_group_male_45-49 |
| – gender_male (<i>sex masculin</i>) | – age_group_50-54 |
| – gender_female (<i>sex feminin</i>) | – age_group_female_50-54 |
| – age_group_total (<i>total grupe de vârstă</i>) | – age_group_male_50-54 |
| – age_group_0-4 | – age_group_55-59 |
| – age_group_female_0-4 | – age_group_female_55-59 |
| – age_group_male_0-4 | – age_group_male_55-59 |
| – age_group_5-9 | – age_group_60-64 |
| – age_group_female_5-9 | – age_group_female_60-64 |
| – age_group_male_5-9 | – age_group_male_60-64 |
| – age_group_10-14 | – age_group_65-69 |
| – age_group_female_10-14 | – age_group_female_65-69 |
| – age_group_male_10-14 | – age_group_male_65-69 |
| – age_group_15-19 | – age_group_70-74 |
| – age_group_female_15-19 | – age_group_female_70-74 |
| – age_group_male_15-19 | – age_group_male_70-74 |
| – age_group_25-29 | – age_group_75-79 |
| – age_group_female_25-29 | – age_group_female_75-79 |
| – age_group_male_25-29 | – age_group_male_75-79 |
| – age_group_29-24 | – age_group_80-84 |
| – age_group_female_29-24 | – age_group_female_80-84 |
| – age_group_male_29-24 | – age_group_male_80-84 |
| – age_group_30-34 | – age_group_85+ |
| – age_group_female_30-34 | – age_group_female_85+ |
| | – age_group_male_85+ |

Soluția prezentată aici se bazează pe o listă de tabele statice stocate în același director cu fișierele de aplicație shiny. Tabelele pot fi actualizate manual, iar aplicația va include fișiere noi la reîncărcare. O altă opțiune este de a stoca tabelele de date într-o bază de date PostgreSQL. Singurele schimbări în acest caz includ modificarea modului de listare al platformei astfel încât acesta să se conecteze la baza de date și să listeze tabelele disponibile, precum și modificarea modului de prelucrare astfel încât acesta să importe geometrii din baza de date spațială.

6.3. Cerințe tehnice pentru aplicații de diseminare a RPL

Configurația standard conform recomandărilor din Rezultatul 4.1b: „*Recomandări privind elaborarea documentației tehnice cu scopul de a organiza achiziția unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocare de date, protecție și securitate a datelor pentru rularea activităților dezvoltate în baza RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA)*” este suficientă. Nu sunt necesare cerințe hardware suplimentare.

Pentru a rula aplicațiile, este necesară o instalare validă R⁷ conform descrierii de mai jos:

<https://www.digitalocean.com/community/tutorials/how-to-install-r-on-ubuntu-18-04>

<https://www.digitalocean.com/community/tutorials/how-to-install-nginx-on-ubuntu-20-04>

<https://www.digitalocean.com/community/tutorials/how-to-set-up-shiny-server-on-ubuntu-20-04>.

De asemenea, o instalare shiny validă cu sursă deschisă, conform descrierii de aici de mai jos:

<https://www.rstudio.com/products/shiny/download-server/ubuntu/>

Pentru toate acestea este suficientă configurația implicită. Aplicațiile au fost dezvoltate și testate strict pentru uz intern, însă INS le poate folosi public dacă dorește. Cu toate acestea, INS își asumă răspunderea deplină și unică pentru eventuale modificări necesare ale configurației IT.

Cerințe preinstalate pe scurt

- Server pe care rulează Ubuntu OS
- Mediu cu software R
- Shiny Server cu sursă deschisă
- Pachete R suplimentare: tidyverse, data.table, shiny, shinydashboard, shinyjs, rmarkdown, knitr, tinytex, sf, scales, leaflet, DT, shinyFilters, mapdeck

O nouă instalare a aceleiași configurații ar trebui să fie posibilă, cu condiția ca utilizatorul să înțeleagă noțiuni de bază ale sistemului de operare Ubuntu, despre linia de comandă și limbajul de programare R. Pașii de bază⁸ pentru replicarea acestei instalări sunt:

- Instalarea Ubuntu pe noul server și configurarea sistemului pentru a opera în rețeaua INS
- Instalarea R din CRAN
- Instalarea Shiny Server

Alte resurse necesare

S-a recomandat a se efectua și pregăti prelucrarea datelor RPL cu ajutorul aplicațiilor bazate pe limbajul R de programare statistică. Este important, și devine obligatoriu, pentru INS să își extindă și să își dezvolte un departament axat pe consolidarea capacităților legate de mediul R. Birourile statistice care folosesc aceste instrumente dispun de un mic departament, în structura IT, dedicat R.

În cadrul INS ar trebui să existe echipe mixte cuprinzând personal cu competențe de programare și dezvoltare în R și statisticieni în diverse domenii de interes (social, economie etc.) pentru a continua să dezvolte și să implementeze soluții R și să le utilizeze în producția statistică. RStudio Server și Shiny Server ar putea fi instalate pentru a fi folosite periodic de către fiecare utilizator în activitatea sa. Utilizatorii frecvenți ai aplicațiilor pot urma cursuri (precum cele propuse prin Planul Multianual de Formare⁹) și, alături de departamentul IT, ar putea extinde sau optimiza aplicațiile propuse.

⁷ <https://cran.r-project.org/doc/manuals/r-release/R-admin.html>

⁸ Documentația integrală și descrierea acestui proces sunt disponibile aici:

[Cum să obții propriile versiuni de RStudio Server și Shiny Server cu DigitalOcean \(deanattali.com\)](#)

⁹ Planul Multianual de Formare pentru personalul INS a fost elaborat de echipa BM în cadrul Rezultatului nr. 12 din același Acord RAS și conține cursuri IT obligatorii sau specifice IT pentru statisticieni, inclusiv cel în limbajul de programare R.

7. RPL2021 - concluzii privind implementarea efectivă

Valorile globale privind participarea la recensământ (datele colectate) au confirmat faptul că „toată lumea contează” pentru recensământ și că eforturile sunt răsplătite când sunt depășite provocările unui recensământ informatizat în premieră implementat în România, o țară cu 19,02 milioane de locuitori, 51% din populație fiind situată în zone rurale. Un sistem integrat solid și fiabil de colectare a datelor RPL a fost implementat și operat pentru a elimina dificultățile și a asigura un proces fluid de colectare a datelor atât în faza CAWI, cât și în cea CAPI, pentru toți respondenții, fie că aceștia erau locuitori în zone urbane sau rurale, erau cu sau fără educație digitală, indiferent de etnia acestora și dacă erau vorbitori nativi de limbă română sau de alte limbi materne înregistrate în țară (de exemplu, chestionare electronice au fost puse la dispoziție tuturor celor 53 de grupuri etnice înregistrate și în cele 17 limbi utilizate în întreaga țară).

Sistemul IT integrat și sistemul de suport al infrastructurii de date pentru recensământ și-a îndeplinit obiectivele: mai întâi, de a sprijini toate aspectele procesului de recensământ de la început până la final, de la colectarea datelor până la diseminarea datelor; în al doilea rând, dat fiind calendarul încărcat parcurs de recensământ și limitele de timp ale diferitelor etape din procesul recensământului, acesta a oferit fiabilitate și stabilitate; în al treilea rând, acesta a asigurat securitatea și confidențialitatea datelor colectate; și, în al patrulea rând, sistemul și-a demonstrat economia de costuri ținând cont că infrastructura IT necesară unui proiect vast precum recensământul ar putea să nu fie în mod continuu necesară. Timpul, efortul și resursele pentru disponibilitatea, digitalizarea, validarea și imputarea datelor au fost considerabil reduse de la 24 de luni, precum în 2011 pentru RPL anterior implementat de România cu colectare de date prin Interviu cu Înregistrarea Răspunsurilor pe Hârtie (PAPI), la cel mult 4 luni în prezent. Cu toate acestea, infrastructura IT a RPL a fost, este și va rămâne o componentă a sistemului IT integrat al INS, cel puțin din perspectiva anchetelor intercenzitare.

Din perspectiva proiectului CONRENA, RPL2021 a concentrat o asistență tehnică substanțială acordată prin intermediul a cincisprezece (15) rezultate, dintre care:

- zece (10) au reprezentat asistență directă și dedicată RPL:
 - o Rezultatul 3a: *Raport privind serviciile de asistență oferite Beneficiarului privind Nota cu recomandări legate de termenii de referință pentru campania de promovare a RPL2021*
 - o Rezultatul 3b: *Raport privind serviciile de asistență oferite Beneficiarului privind Nota privind planificarea, managementul și implementarea RPL2021*
 - o Rezultatul 3c: *Raport privind serviciile de asistență oferite Beneficiarului privind Notele legate de legislația revizuită pentru RPL2021 (patru (4) note provizorii cu recomandări, un (1) plan de implementare), inclusiv raportul privind două (2) ateliere pe tema legislației RPL2021*
 - o Rezultatul 4.1c: *Plan documentat pentru sistemul integrat pentru implementarea RPL2021 (detalii privind modul în care va fi realizată implementarea infrastructurii IT pentru RPL 2021);*
 - o Rezultatul 7a: *Raport privind serviciile de asistență oferite Beneficiarului privind Metodologia de evaluare și promovare continuă și în timp real a calității și acoperirii datelor colectate în cadrul RPL2021, precum și a protecției/securității datelor;*
 - o Rezultatul 7c: *Raport privind serviciile de asistență oferite Beneficiarului privind Recomandările și bunele practici de implementare a sistemului de management al datelor pentru datele geospațiale RPL2021 și sectorizarea efectivă;*

- Rezultatul 10a: *Raport privind Recomandările oferite Beneficiarului privind modul în care se desfășoară procesul de pilotare a RPL2021;*
- Rezultatul 10b: *Raport privind serviciile de asistență oferite Beneficiarului privind Asistența tehnică și recomandările de bune practici pentru SDC, confidențialitatea datelor, modalități de a securiza microdate și date agregate; și*
- Rezultatul 10c: *Raport privind serviciile de asistență oferite Beneficiarului privind Setul de instrumente statistice provizorii pentru recensământ folosind metode multimodale de promovare a protecției și securității datelor; și raportul propriu-zis*
- Rezultatul 13: *Raport privind serviciile de asistență oferite Beneficiarului privind elaborarea Recomandărilor și orientărilor referitoare la planul documentat pentru sistemul integrat dedicat implementării RPL2021 (inclusiv gestionarea utilizatorilor, criptarea ID-urilor, recuperarea în urma dezastrelor, menținerea disponibilității sistemului în caz de avarie)*
- iar cinci (5) au oferit suport încrucișat pentru RPL și alte componente ale proiectului CONRENA (RGA, perioada intercenzitară, SICCA).
 - Rezultatul 1: *Raport privind serviciile de asistență oferite Beneficiarului privind Analiza capacității de operare NSS pentru producția de Statistici Oficiale și îndeplinirea recensământului general agricol și a reconcilierii populației și a locuințelor*
 - Rezultatul nr. 4.1b: *Recomandări privind elaborarea documentației tehnice cu scopul de a organiza achiziția unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocare de date, protecție și securitate a datelor pentru rularea activităților dezvoltate în baza RAS (GIS, RPL2021; RGA2020, perioade intercenzitare, SICCA)*
 - Rezultatul nr. 4.2a: *Recomandări și orientări oferite Beneficiarului pentru rularea sistemului IT integrat cu componentele sale, hardware (inclusiv tablete) și software (licențe), inclusiv stocare de date, protecție și securitate a datelor pentru rularea activităților dezvoltate în baza RAS (GIS, RPL2021, RGA2020, perioade intercenzitare, SICCA);*
 - Rezultatul 6e: *Raport privind serviciile de asistență oferite Beneficiarului privind Raportul referitor la două (2) ateliere de cinci zile pentru statisticienii INS cu privire la metodologii și instrumente de transformare a chestionarelor în e-chestionare statistice inteligente pentru perioada intercenzitară*
 - Rezultatul 12: *Raport privind serviciile de asistență oferite Beneficiarului privind Dezvoltarea unui plan multianual al necesarului de formare (2021- 2024) pentru INS, inclusiv STD-uri în funcție de necesarul de formare; efectuarea: unui (1) curs de formare de două (2) zile pentru specialiștii INS și părțile interesate NSS relevante, unui număr de trei (3) cursuri de formare de trei (3) zile pe metoda de lucru pentru dezvoltarea Relațiilor Europene și a unui (1) curs de formare de cinci (5) zile pe tema SDC pentru personalul INS; recomandări pentru a identifica instituții adecvate în care s-ar putea desfășura vizite de studiu pe teme de statistică; două (2) vizite de studiu; și o conferință pe tema bunelor practici la nivelul UE*

RPL, etapa 2021, a fost pregătit și implementat în cadrul unui efort comun al INS și al altor autorități publice, pe durata ultimilor cinci ani și a demonstrat importanța colaborării continue cu părțile interesate, a planificării solide și coerente, precum și avantajele inovației aduse de metode, instrumente și aplicații pentru producția de statistici.

8. Anexe

Anexa 1 - Sistemul informatic de colectare a datelor RPL – Raport conform cu execuția

Anexa 2 - Modele de rapoarte de monitorizare zilnică (CAWI, CAPI)

Anexa 3 - Modele de rapoarte de plăți aferente CAPI

Anexa 4 - Scripturi de validare și aprobare a colectării datelor RPL (CAWI, CAPI)

Toate anexele sunt furnizate ca arhive în format electronic și atașate la prezentul raport.

Anexa 1 - Sistemul informatic de colectare a datelor RPL – Raport conform cu execuția

Anexa este furnizată ca arhivă în format electronic și conține fișierele enumerate mai jos:

<u>Summary of PHC as-build system architecture</u>	1
<u>Overall architecture</u>	1
<u>Load balancers</u>	2
<u>Database servers</u>	2
<u>Application servers</u>	2
<u>Monitoring server</u>	3
<u>High availability</u>	3
<u>Performance</u>	5
<u>Backup</u>	5
<u>Security</u>	5
<u>Data protection</u>	6
<u>Log management</u>	7
<u>Disaster recovery</u>	7
<u>Detailed configuration</u>	9
<u>Load balancers configuration</u>	9
<u>Basic system setup and NGINX installation</u>	9
<u>Monitoring configuration</u>	16
<u>High availability configuration</u>	16
<u>Security configuration</u>	20
<u>Log collection configuration</u>	20
<u>Database servers configuration</u>	36
<u>Basic system setup and Postgresql installation</u>	36
<u>Monitoring configuration</u>	40
<u>Replication and high availability configuration</u>	41
<u>Security configuration</u>	42
<u>Log collection configuration</u>	43
<u>Application servers configuration</u>	60
<u>Survey Solutions HQ installation</u>	60
<u>RabbitMQ installation</u>	66
<u>Self-registration component installation</u>	70
<u>Monitoring configuration</u>	80
<u>Security configuration</u>	80
<u>Log collection configuration</u>	86
<u>Load testing and monitoring server configuration</u>	93
<u>Basic system setup and Docker installation</u>	93
<u>Monitoring configuration</u>	94
<u>Log collection configuration</u>	177
<u>Maintenance procedures</u>	193
<u>Initialization procedures</u>	193
<u>High-availability procedures</u>	193
<u>Failback procedure for database servers</u>	193
<u>Failback procedure for application servers</u>	194
<u>Disaster recovery procedure</u>	196
<u>Capacity procedures</u>	197
<u>Extend disk space procedure for Ubuntu machines</u>	197

Anexa 2 – Modele de rapoarte de monitorizare zilnică (CAWI, CAPI)

Anexa este furnizată ca arhivă în format electronic și conține fișierele enumerate mai jos:

 11_CAWI_ara_20220527_v3.csv		 11_CAWI_ARA_sumar_20220527_v3...		 11_CAWI_daily_stats_v3.R	
 11_CAWI_situatie_erori_RPL_202205...		 11_CAWI_situatie_pe_judete_RPL_20...		 11_CAWI_situatie_pe_localitati_RPL_...	
 12_CAPI_00_RO_RPL2021_situatie_c...		 12_CAPI_AB_RPL2021_situatie_chesti...		 12_CAPI_model_tabel_raportari_jude...	
 12_CAPI_monitorizare_date_colectat...		 12_CAPI_RPL_reports_functions.R			

Anexa 3 - Modele de rapoarte de plăți aferente CAPI

Anexa este furnizată ca arhivă în format electronic și conține fișierele enumerate mai jos:



Anexa 4 - Scripturi de validare și aprobare a colectării datelor RPL (CAWI, CAPI).

Anexa este furnizată ca arhivă în format electronic și conține fișierele enumerate mai jos:

 41_CAWI_validari_v3.R



 42_CAPI_aprobari_automate_RPL_v1...



 42_CAPI_validari_CAPI_v2.R





Competența face diferența!
Proiect selectat în cadrul Programului Operațional Capacitate Administrativă, co-finanțat de Uniunea Europeană din
Fondul Social European

