ROMÂNIA

Acord de servicii de asistenţă tehnică rambursabile privind Consolidarea Sistemului Statistic Naţional (P167217)

REZULTAT Nr. 10c

Raport privind serviciile de asistență tehnică oferite Beneficiarului referitoare la setul de instrumente proiect ale recensământului statistic care utilizează metode multimodale pentru a promova protecția și securitatea datelor

Octombrie 2022

Revizuit noiembrie 2022

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

Proiect cofinanțat 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 părerile Directorilor executivi ai Băncii Mondiale sau ale guvernelor pe care aceștia le reprezintă. Banca Mondială nu garantează acuratețea datelor incluse în prezentul document și nu își asumă responsabilitatea pentru orice erori, omisiuni sau inconsecvențe în ceea ce privește informațiile sau răspunderea cu privire la utilizarea sau neutilizarea informațiilor, metodelor, proceselor sau concluziilor prezentate. Granițele, culorile, denumirile și alte informații afișate pe orice hartă din acest document nu implică nicio apreciere din partea Băncii Mondiale cu privire la statutul juridic al oricărui teritoriu sau aprobarea sau acceptarea acestor granițe.
Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului României.

Acest raport a fost transmis în luna octombrie 2022 și revizuit în luna noiembrie 2022, în cadrul Acordului de Servicii de Asistență Tehnică Rambursabile privind Consolidarea Sistemului Statistic Național (P167217), semnat între Institutul Național de Statistică din România și Banca Internațională pentru Reconstrucție și Dezvoltare la data de 17 septembrie 2019. Acesta corespunde Rezultatului nr. 10 din acordul menționat mai sus.

Cuprins

Lista figurilor

Listă de acronime

API	Interfață de programare a aplicației
BM	Banca Mondială
CAPI	Interviu personal asistat de calculator
CAWI	Interviu web asistat de calculator pe internet
CNP	Cod Numeric Personal
DMZ	Zonă demilitarizată
HTTPS	Hipertext Transfer Protocol Secure
INS	Institutul Național de Statistică
IT	Tehnologia Informației
RAS	Servicii de asistență tehnică rambursabile
RPL	Recensământul Populației și Locuințelor
SDC	Controlul Divulgării Datelor Statistice
SHA-1	Secure Hash Algorithm 1
STS	Serviciul de Telecomunicații Speciale
SuSo	Survey Solutions
VPN	Rețea privată virtuală

Introducere

Scopul acestui raport este de a prezenta ***setul de instrumente proiect ale recensământului statistic care utilizează metode multimodale pentru a promova protecția și securitatea datelor.*** Aceasta face parte din livrabilele din cadrul Acordului de Servicii de Asistență Tehnică Rambursabile (RAS) privind Consolidarea Sistemului Statistic Național (P167217). Proiectul este implementat de Institutul Național de Statistică (INS) cu sprijinul Băncii Mondiale.

Acest raport prezintă metodele de securitate aplicate și utilizate pentru evitarea riscurilor de securitate în etapele și procesele de colectare a datelor CAWI și CAPI (de exemplu, pre-inregistrare, autorecenzare, completare chestionar, transfer de date către persoane auto-recenzare/recenzori, transfer de date între recenzori, recenzori sefi, INS central) și pe medii dedicate (de exemplu, medii de colectare a datelor și mediu de prelucrare a datelor) care au susținut Recensământul Populației și Locuințelor (RPL), runda 2021.

Raportul are trei secțiuni. Prima secțiune oferă o privire de ansamblu asupra măsurilor fizice și logice de protecție și securitate ale sistemului de colectare a datelor RPL care au fost aplicate în timpul recensământului la nivel interorganizațional în cadrul organizațiilor participante (INS, STS, furnizori de servicii).

A doua secțiune se referă la metodele aplicate pentru securitatea datelor în timpul recensământului, prin configurarea securității în mediile IT ale STS și INS, prin funcțiile Survey Solutions - soluția de colectare a datelor utilizată în etapele CAWI și CAPI ale recensământului și la metodele specifice utilizate pentru preluarea datelor din sursele administrative în vederea pregătirii și implementării recensământului.

A treia secțiune prezintă câteva recomandări provenite din implementarea recensământului, cu privire la acțiunile necesare pentru protecția și securitatea datelor, în vederea viitoarelor sondaje ce vor fi implementate de INS.

În completarea acestui raport, pot fi regăsite informații suplimentare în alte rapoarte transmise către INS ca parte a acestui proiect, în special următoarele:

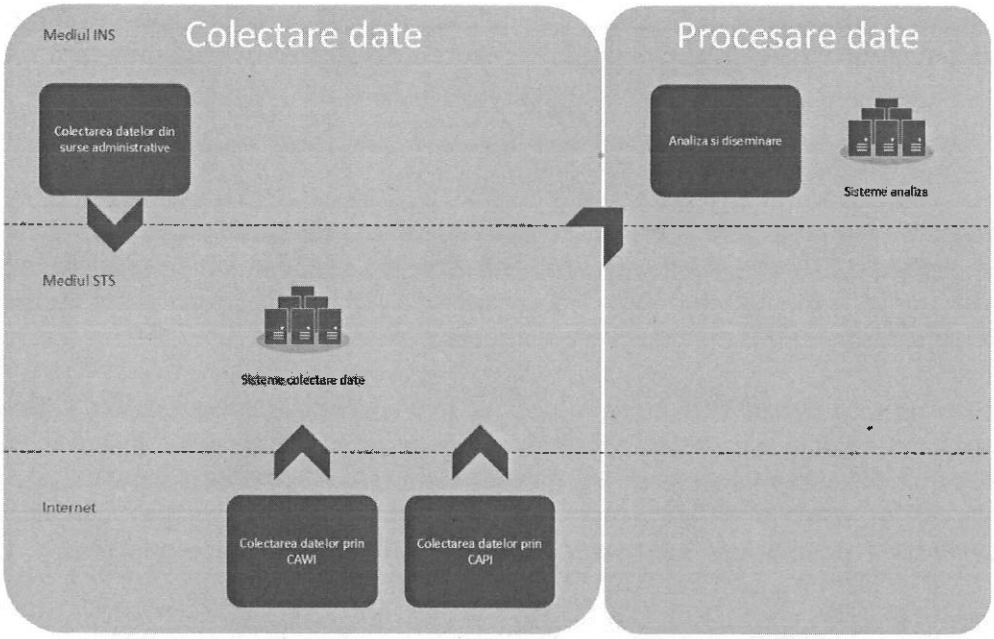
- Rezultatul nr. 4.1c: *Raport privind serviciile de asistență tehnică oferite Beneficiarului cu privire la Planul documentat pentru sistemul integrat pentru implementarea RPL2021 (detaliază modul în care se va realiza implementarea infrastructurii IT pentru RPL2021);*
- Rezultatul nr. 10a: *Raport privind recomandările oferite Beneficiarului pentru modul de efectuare a procesului de pilotare a RPL2021;* și
- Rezultatul nr. 7a*: Metodologia de evaluare și promovare continuă și în timp real a calității și acoperirii datelor colectate în cadrul RPL2021 și a protecției/securității datelor.*

1. Privire de ansamblu asupra protecției și securității datelor RPL

Protecția și securitatea datelor RPL se bazează pe o varietate de măsuri aplicate la nivel interorganizațional în cadrul organizațiilor participante (INS, STS, furnizori de servicii) și a vizat configurațiile sistemelor IT care transferă date sensibile de la punctele finale de introducere a datelor (cetățeni și recenzori) în vederea securizării stocării și prelucrării datelor.

O privire de ansamblu asupra fluxurilor de date între diferitele etape ale RPL (vezi Figura 1) expune procesele, mediile și zonele asociate cu riscuri potențiale pentru securitatea datelor în timpul colectării și procesării acestora.

Figura 1 – Fluxuri de date în timpul RPL



Riscurile de afectare a securității fizice și a securității logice sunt cele două categorii principale de riscuri potențiale la adresa componentelor și proceselor întregului sistem utilizat în timpul colectării și prelucrării datelor și, în consecință, acestea vor avea efect asupra securității și protecției datelor (vezi figura 2).

Figura 2 - Componentele potențial afectate de riscuri

Medii	Componentele de colectare și prelucrare a datelor recensământului expuse la riscuri de securitate fizică și logică
INS	➢ Aplicații și soluții utilizate pentru colectarea și transferul de date din surse administrative; ➢ Aplicații și soluții utilizate pentru analiza datelor în timpul colectării și raportării zilnice, investigarea informațiilor furnizate în timpul autorecenzării în cazurile de funcționare defectuoasă a sistemului; ➢ Validarea datelor colectate de recenzori; ➢ Tablete și date stocate de recenzori în timpul colectării datelor CAPI (presupunând că recenzorii ca parte a mediului de colectare a datelor INS);
STS	➢ Soluții hardware și software de bază IT care compun și asigură funcționarea sistemului de colectare a datelor (în centrele de date primare și secundare) ➢ Survey Solutions instalat și configurat pentru etapele CAWI și CAPI
Internet	➢ Survey Solution Interfață web pentru colectarea datelor CAWI (inclusiv preinregistrarea) ➢ Survey Solution Interfață web pentru colectarea și validarea datelor CAPI

În timpul recensământului, securitatea fizică a sistemului s-a bazat 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 de INS și/sau de BM.

Securitatea logică a sistemului avea diferite niveluri de protecție:

- firewall-urile rețelei STS protejau accesul la sistemele găzduite în centrele de date STS;
- toate dispozitivele aveau implementate firewall-uri gazde care permit doar traficul minim necesar pentru funcționalitatea sistemului;
- serverele Windows erau protejate de antivirusul nativ Windows Defender;
- accesul administrativ la sistem a fost protejat prin soluție VPN restricționată point-to-site, disponibilă doar personalului autorizat.

Securitatea fizică a infrastructurii IT găzduite de STS și dedicată colectării datelor (și utilizate în recensământul propriu-zis) era/este asigurată de procedurile și protocoalele organizatorice STS. Accesul fizic la acest mediu era/este permis doar personalului autorizat STS. Accesul logic era/este permis numai persoanelor care au un nivel de autorizare eligibil și este posibil doar prin autentificarea VPN cu doi factori.

Aplicațiile expuse web au fost supuse unei evaluări de securitate, efectuată de o echipă dedicată STS. În procesul de colectare a datelor pentru efectuarea recensămintelor, raportul de evaluare a fost prezentat INS de către echipa STS. Recomandările privind funcționalitatea aplicațiilor și configurarea aplicațiilor au fost aplicate corespunzător.

În ceea ce privește securitatea fizică a mediului IT al INS, sistemul a fost pregătit de către INS și furnizor și a aplicat aspecte ale configurației legate de securitate. INS a aplicat propriile proceduri administrative și a decis în consecință să permită accesul personalului responsabil cu vizualizarea sau validarea datelor. Acestea au fost completate de acțiunile de securitate logică pentru a asigura securitatea robustă a mediului IT (de exemplu, soluție VPN, configurații de sisteme etc., deja existente).

Procesul de colectare a datelor a inclus informații sensibile cu privire la CNP-ul persoanelor recenzate. Acest lucru a fost asigurat prin pseudonimizare, printr-o funcție HASH cu algoritm SHA-1.

2. Metode aplicate pentru securitatea datelor în timpul recensământului

2.1. Caracteristicile Survey Solutions de securitate a datelor

Survey Solutions este conceput în vederea utilizării în aplicații care implică manipularea datelor de identificare personală (PID), precum nume, adrese, numere de telefon, locații GPS și alte informații sensibile despre respondenți, cum ar fi venitul, consumul/cheltuielile, economiile etc. Accesul neautorizat la aceste date poate duce la riscuri în privința reputației pentru organizațiile care efectuează sondajul și la consecințe negative pentru persoanele ale căror date au fost utilizate abuziv.

Încercările de utilizare abuzivă a datelor pot proveni de la:

- persoanele implicate în sondaj (atacuri interne);
- persoane care nu sunt implicate în sondaj (atacuri de hackeri).

Survey Solutions reduce riscurile **atacurilor interne** prin limitarea accesului utilizatorilor la anumite informații. Unul dintre mecanismele eficiente de a face acest lucru este divizarea activității în mai multe fațete implementate de diferite echipe. Această divizare a activității este implementată ierarhic, fiecare echipă fiind responsabilă exclusiv de un segment dedicat, neavând astfel acces la alte segmente. În timp ce pentru unii utilizatori care coordonează munca echipelor este menținut accesul la toate datele, această compartimentare reduce posibilitatea de a căuta datele unei anumite persoane (cum ar fi datele unui politician celebru, sportiv sau alte persoane de interes).

Riscurile de manipulare a datelor sunt reduse și mai mult prin conceptul **rolurilor de utilizator**, care sunt emise fiecărui utilizator care este autorizat în sistem. Rolurile sunt atribuite fiecărui cont în parte și doar o singură dată (în timpul creării contului). Rolul determină ce funcționalitate a programului este disponibilă utilizatorului care deține acest cont. De exemplu, posibilitatea de a aproba interviurile este limitată conturilor pentru rolurile de *Supervizor* și *Coordonator*, dar accesul la răspunsurile la întrebări individuale este structurat astfel încât utilizatorii *Supervizori* și *Coordonator* să poată citi răspunsurile, dar nu să le și modifice, în timp ce utilizatorii cu tipul de cont *Recenzor* sunt împuterniciți atât pentru operațiuni de citire, cât și de editare.

Atribuirea rolurilor este privilegiul utilizatorului *Administrator*, care este, de asemenea, împuternicit să creeze toate conturile celorlalți utilizatori, să le reseteze parolele și să efectueze alte sarcini de gestionare a contului. Patru persoane dețin drepturi de *Administrator*.

Produsul exportului de date de la Survey Solutions este o arhivă a bazei de date care conține toate informațiile obținute ca parte a colectării datelor. Acest produs are cea mai mare valoare, deoarece conține toate răspunsurile la toate întrebările (inclusiv PID). Scurgerea potențială a acestui fișier este una dintre amenințările cu cea mai mare severitate. Doar utilizatorii cu roluri de *Coordonator* și *Administrator* au posibilitatea de a iniția procesul de export de date și de a obține produsul acestuia (această acțiune poate fi efectuată și folosind conturi API robotizate, care sunt opționale în Survey Solutions). Această **arhivă descărcabilă poate fi astfel protejată suplimentar cu o parolă principală** pe care o poate atribui *Administratorul* serverului de date Survey Solutions (separat pentru fiecare spațiu de lucru). Utilizarea parolei pe arhiva descărcabilă reduce și mai mult numărul utilizatorilor care au acces la date doar la Administratorul însuși și utilizatorii cărora le încredințează parola principală (de obicei, un administrator de rezervă).

Protejarea parolei contului intră în responsabilitatea utilizatorului. Pentru a se asigura că parola nu este cunoscută de niciun alt utilizator din sistem, Survey Solutions impune **schimbarea parolei atunci când este utilizată pentru prima dată cu o nouă parolă, care este cunoscută doar de acel utilizator.** Această acțiune are loc înainte ca orice activitate să fie efectuată din acel cont, iar acțiunea de schimbare a parolei este înregistrată.

Pentru a crește și mai mult securitatea parolelor, este posibil un al doilea mecanism de autentificare, în care utilizatorii pot activa selectiv **autentificarea cu doi factori (2FA)** pentru conturile lor. În acest caz, pentru a se conecta la server, deținătorul contului trebuie să furnizeze parola corectă și, de asemenea, un simbol corect, care este emis de o aplicație de autentificare care rulează pe telefonul sau PC-ul utilizatorului. Sunt disponibile diverse alternative și implementări open source ale unor astfel de aplicații de autentificare folosind algoritmul TOTP (o singură parolă în timp real; en. *time-based one time-password*) de la furnizorii importanți de software (Microsoft, Google etc.), (2FA nu este implementată pe tablete).

Utilizatorii care au nevoie de resetarea parolei sau dezactivarea 2FA trebuie să contacteze *Administratorul* serverului.

Politica referitoare la parole care se aplică tuturor parolelor din Survey Solutions (lungime minimă a parolei și set de caractere) este configurabilă în afara software-ului în fișierul de configurare. În mod implicit, necesită parole cu lungimea de 10-12 caractere, cu cel puțin o literă minusculă, o majusculă și o cifră. Aceeași politică se aplică parolelor pe care utilizatorii le specifică ca fiind proprii atunci când le schimbă.

Sistemul Survey Solutions are, de asemenea, apărare încorporată împotriva accesului neautorizat (**atacuri de hacker**). Punctul de intrare principal în orice sistem protejat cu parolă este pagina de autentificare, care evaluează utilizatorii ca fiind autentificați sau nevalizi. Această pagină poate fi supusă teoretic unui atac de forță brută (sondarea unui număr mare de combinații de nume de utilizator și parole). Pentru a descuraja astfel de atacuri, software-ul reacționează la încercările de conectare nevalide: folosind numărători ale încercărilor de conectare (specifice fiecărui cont) atunci când este detectat un anumit număr de încercări de conectare nevalide, utilizatorului i se prezintă un **test CAPTCHA**. În cazul în care contul este atacat de un robot, aceasta este o modalitate eficientă de a-l împiedica să continue atacurile de forță brută. În cazul în care combinațiile sunt încercate de o persoană, care poate rezolva testul CAPTCHA, atunci persoana respectivă va avea un număr suplimentar (configurabil) de încercări înainte de blocarea contului pentru orice încercare de conectare pentru o perioadă fixă de timp (configurabilă). O combinație de CAPTCHA și de blocări ale contului permite prevenirea eficientă a atacurilor de refuzare a serviciului (DOS), în cazul în care un utilizator rău intenționat care cunoaște un nume de autentificare valid nu poate dezactiva în mod eficient acel cont introducând combinații nevalide de autentificare + parolă.

Nu există CAPTCHA sau blocarea contului pe tabletă (aplicațiile Recenzorului și Supervizorului).

În timpul sesiunilor CAWI de introducere a datelor, comunicarea datelor din browserul utilizatorului către serverul de date este protejată de un **certificat SSL**, care permite schimbul securizat de informații folosind HTTPS. În timp ce Survey Solutions poate funcționa din punct de vedere tehnic prin HTTP obișnuite, o astfel de utilizare este descurajată, deoarece lasă traficul de schimb neprotejat de atacurile de interceptare.

Datele comunicate în sesiunile CAPI sunt protejate de aceeași tehnologie (comunicare securizată prin HTTPS).

Datele stocate pe tabletă (în timpul sesiunilor de interviu și, de asemenea, pentru interviurile amânate) sunt securizate cu următoarele două mecanisme:

1. răspunsurile la orice întrebări sunt criptate;
2. datele sunt stocate într-o „stocare privată", care este o stocare izolată specifică aplicației în sistemul de operare Android.

Pachetele de depanare de date care pot fi trimise către server de către personal folosind aplicații pentru tabletă conțin și date în formă criptată, unde acele date reflectă răspunsurile respondenților. Aceste pachete pot fi decriptate fără probleme pe serverul către care sunt trimise de către administratorul serverului. Acest lucru se întâmplă automat și fără a fi nevoie de acțiuni speciale din partea utilizatorului.

Pentru a preveni descărcarea tuturor datelor de pe tabletă, **centralizatorul aplicației Android** conține semnalizatoare care indică faptul că aplicația nu ar trebui să fie instalabilă pe un suport amovibil sau să se facă copii de rezervă folosind facilitățile de transfer și backup încorporate ale aplicației Android.

Pentru a preveni transmiterea datelor de pe alte dispozitive, fiecare cont de recenzor este conectat la un singur dispozitiv fizic urmărit de numărul său unic intern. În timp ce recenzorii pot trece la un alt dispozitiv (re-conectare) în cazul în care unul este pierdut, furat sau deteriorat fizic, acel dispozitiv nu va avea permisiunea să înainteze datele sau să primească răspuns de la server în cazul în care cineva care le găsește încearcă să facă acest lucru.

Software-ul Survey Solutions face ca fiecare server să pară unic prin generarea unei semnături aleatorii în timpul procesului de instalare. Deși aceasta nu este o caracteristică de securitate în sine, ajută la prevenirea greșelilor accidentale ale recenzorilor de a trimite datele spre un server eronat.

Codul sursă al software-ului (atât versiunea mobilă, cât și aplicația de pe server) este disponibil în întregime pe un site public care permite experților să investigheze liber posibili vectori de atac și să transmită mesaje dezvoltatorilor și utilizatorilor deopotrivă.

Măsurile suplimentare care cresc securitatea întregului sistem atunci când sunt implementate sunt:

1) Asigurarea că autentificarea este activată pe dispozitivele recenzorilor (ecranul de blocare a sistemului de operare Android este protejat cu un PIN sau o parolă sau altă metodă similară). Acest lucru va proteja suplimentar datele confidențiale, prin crearea unui alt nivel de protecție.
2) Criptarea bazei de date pe server, de exemplu, prin plasarea acesteia pe volumele de disc protejate cu sistemul de operare integrat sau instrumente de criptare a datelor terțe.
3) Deoarece baza de date în sine nu este criptată în camera serverului, trebuie să existe o verificare a personalului de securitate fizică și de întreținere pentru a preveni orice acces neautorizat la nivel fizic.

Principalele aspecte care trebuie reținute sunt:

- Pentru a proteja datele în tranzit, Survey Solutions utilizează protocolul HTTPS, care este standardul industrial pentru astfel de sarcini. Este utilizat în ambele etape de colectare a datelor, respectiv CAWI și CAPI.

- Pentru a accesa orice date colectate, utilizatorii trebuie întotdeauna să furnizeze datele de conectare și parola.
- Parolele sunt stocate în baza de date a Survey Solutions folosind coduri tip „hash", astfel încât să nu se poată inversa nici în cazul primirii unei valori tip „hash" a parolei.
- Autentificarea Supervizorului și a Coordonatorului au, de asemenea, capacitatea de a activa autentificarea cu doi factori pentru a adăuga un nivel suplimentar de protecție.
- Pagina de autentificare conține protecție împotriva atacului de forță brută prin utilizarea mecanismului de blocare și necesită completarea captcha în cazul mai multor încercări eșuate de conectare.
- Datele în repaus de pe tablete sunt stocate în spațiul de stocare izolat al aplicației Android și prezintă criptare suplimentară care este aplicată tuturor datelor colectate până când sunt trimise la modulul Coordonator al Survey Solutions.

2.2. Securitatea datelor în timpul recensământului

Metodele aplicate pentru securizarea datelor în timpul preluării din sursele administrative sunt:

- Informațiile din sursele administrative sunt stocate în rețeaua internă a INS și nu sunt expuse public.
- CNP-urile sunt stocate drept coduri tip „hash", astfel încât portalul de autorecenzare să poată prelua o singură înregistrare din sursă după ce utilizatorul furnizează CNP-ul său real.
- După completarea formularului, utilizatorul primește un e-mail cu un link unic. Linkul este valabil până la finalizarea interviului aferent. După finalizare, numai INS are acces la datele colectate folosind acreditările Administratorului sau Coordonatorului.
- O altă configurație de securitate aplicată etapei de autorecenzare a fost mascarea CNP-ului în notificările de autorecenzare prin e-mail. Această configurație a fost aplicată la nivelul șablonului aplicației de autorecenzare. Mascarea CNP-ului constă în ascunderea a 5 cifre, de la cifra 6 până la cifra 10, după cum urmează: 12345xxxx123.

Datele colectate în timpul etapei CAWI sunt disponibile respondenților până la finalizarea interviului. După finalizare, numai INS are acces la datele colectate utilizând acreditările de Supervizor sau de Coordonator.

Datele colectate în timpul etapei CAPI, datele de pe tablete, astfel cum sunt descrise în secțiunea 2.1, sunt stocate într-o formă criptată până când sunt trimise Coordonatorului. După ce sunt trimise, răspunsurile la interviu sunt șterse de pe tablete și stocate doar în baza de date centrală a Survey Solutions.

În ceea ce privește securitatea microdatelor, doar contul de Coordonator are acces la fișierul de date complete colectate. Supervizorii au acces doar la interviurile care au fost finalizate de echipa lor. După ce interviul este aprobat de către Supervizor, acesta devine vizibil numai pentru utilizatorii cu rol de Coordonator.

Pentru securitatea datelor agregate, INS utilizează protocoale interne sau de parteneriat (de exemplu, STS) pentru transferul datelor din sistemele de colectare a datelor în mediul IT propriu (după caz) și aplică regulile în consecință.

Procesul de colectare a datelor a fost monitorizat cu instrumentul Survey Solutions, disponibil drept componentă a platformei. În etapa CAWI a RPL, s-au rulat scripturile de validare pentru a determina completitudinea datelor colectate în urma autorecenzării. Zilnic, pe baza acestor validări, a fost creat

un set de date pentru prezentarea informațiilor despre autorecenzare. În timpul acestui proces, CNP-urile persoanelor au fost și sunt pseudonimizate printr-o funcție HASH cu algoritm SHA-1.

În etapa CAPI, pentru monitorizarea zilnică a activităților recenzorilor și agregarea datelor la nivel de județe, au fost realizate rapoarte prin scripturi R rulând pe un server de analiză, pe baza datelor exportate din Survey Solution. Scripturile de validare rulează zilnic pe datele colectate prin metoda CAPI, pentru a determina erorile și a aproba chestionarele validate.

Aprobarea de către scripturi a chestionarelor identificate ca fiind corecte s-a făcut prin apeluri API. Regulile de validare au fost stabilite de echipa INS și implementate în RStudio pe serverul de analiză. În timpul activităților de monitorizare, nu s-a făcut analiza CNP-urilor, iar algoritmii au verificat numărul minim de caracteristici ale unităților statistice colectate în RPL.

Pentru anonimizarea datelor colectate și pentru pregătirea setului de date de utilizare publică pentru distribuire, se aplică o procedură SDC, astfel cum este prezentată de Eurostat[1]. SDC face parte din instrumentele luate în considerare pentru protecția și securitatea datelor și va fi aplicată de INS urmând conceptele descrise de echipa BM în cursul formării din decembrie 2021 (sesiune de formare de 5 (cinci) zile privind SDC pentru personalul INS, în cadrul Rezultatului 12) și metodele alese de INS ca parte a asistenței tehnice primite de la BM în cursul lunii iunie 2022 în cadrul Rezultatului 10b: *Raport privind serviciile de asistență tehnică oferite Beneficiarului pentru Asistență tehnică și recomandări referitoare la cele mai bune practici pentru SDC privind formarea personalului INS, confidențialitatea datelor, modalități de securizare a micro-datelor și a datelor agregate.*

2.3. Configurarea pentru securitate a mediului IT

2.3.1. Mediul IT găzduit de STS (colectare de date)

Mediul STS utilizat pentru colectarea datelor a fost un mediu dedicat pentru etapa de colectare a datelor RPL. Acesta a constat într-o DMZ care găzduiește doi echilibratori de sarcină care transmit cererile web de la cetățeni și recenzori către serverele de aplicații. Un grup de cinci servere de aplicații au efectuat aceste solicitări și le-au distribuit între aplicații (SuSo, autorecenzare). Datele au fost stocate într-o bază de date Postgresql 12, constând dintr-un cluster de două noduri în configurație activ-pasivă. Aceeași structură a fost replicată într-un centru de date STS secundar dedicat, găzduit în altă locație.

În afară de infrastructura serverului, mediul IT dedicat a fost protejat de accesul extern utilizând configurația standard de firewall STS (standardele și procedurile interne STS nu permit accesul terților la acele configurații și sunt denumite „standard") și cu acces VPN cu 2 factori pentru personalul autorizat.

Accesul la aplicații prin echilibratorul de sarcină s-a făcut folosind conexiunile SSL. Certificatele SSL au fost instalate pe ambii echilibratori de sarcină, iar accesul de pe internet a fost posibil doar prin utilizarea HTTPS.

[1] https://ec.europa.eu/eurostat/cros/system/files/SDC_Handbook.pdf

În timpul etapelor CAWI și CAPI ale RPL, au fost aplicate diferite configurații de securitate la niveluri echilibratorilor de sarcină (sunt disponibili în **Anexa 1** – Sistemul de colectare a datelor RPL – Raportul conform cu execuția).

În etapa de autorecenzare (CAWI), din cauza cerințelor INS conform cărora portalul de autorecenzare (autorecenzare.insse.ro) trebuia să fie accesibil și de pe adresele de internet din afara României, au fost aplicate condiții de securitate suplimentare pentru configurația echilibratoarelor de sarcină, limitând accesul la SuSo doar pentru finalizarea interviurilor. Limitarea a fost aplicată motorului NGINX folosind o expresie obișnuită care interzice accesul la alte funcționalități ale SuSo din locații externe (internet):

```
((\/[Ww]eb[Ii]nterview\/|\/js\/|\/fonts\/|\/CompanyLogo\/|\/[Ii]nterview\/|\/i
nterview$|\/WebInterviewResources\/|\/css\/|\/img\/|\/locale\/|\/pdf$))
```

Această expresie obișnuită este folosită pentru filtrarea cererilor doar pentru a permite traseele SuSo în raport cu finalizarea interviurilor.

În etapa CAPI, au fost aplicate la nivel național restricții de acces IP-geoblocare, iar aplicația de autorecenzare a fost dezactivată.

Toate serverele infrastructurii găzduite de STS au fost configurate să folosească și firewall-uri gazdă (UFW pe dispozitivele Ubuntu 20.04 și Windows Defender Firewall pe dispozitivele Windows 2016). Erau deschise doar porturile necesare pentru funcționarea normală a aplicațiilor. Pentru detalii, vă rugăm să consultați **Anexa 1** – Sistemul de colectare a datelor RPL – Raportul conform cu execuția.

La începutul RPL, toate serverele care funcționau în sisteme au fost actualizate la cele mai recente actualizări de securitate.

Aplicațiile găzduite pe nodurile serverelor de aplicații au fost supuse unei evaluări de securitate efectuate de către STS, iar rezultatele evaluării au fost aplicate în consecință - pentru referință, a se vedea fișierele distribuite de STS către INS – *Raport de evaluare a vulnerabilităților aplicația web CAPI rplcapi.insse.ro, 14.02.2022* și *Raport de evaluare a vulnerabilităților aplicația web autorecenzare.insse.ro, 20.02.2022.*

Mediul găzduit de STS utilizat pentru colectarea datelor a fost conectat printr-un tunel VPN securizat cu infrastructura găzduită de INS – sarcină implementată de INS prin proceduri proprii. Singura conexiune permisă între mediul IT al INS și mediul IT al STS este o conexiune https (cu VPN criptată), respectiv între serverul de analiză și echilibratorii de sarcină, pentru ca operatorii de prelucrare a datelor să primească date de la SuSo (acces API pentru exporturi și validări de date).

2.3.2. Mediul IT găzduit de INS (prelucrarea datelor)

Infrastructura INS nou achiziționată constă într-un mediu cluster VMWare, care permite crearea de dispozitive virtuale. Singura componentă care comunică cu infrastructura de colectare a datelor găzduită de STS este serverul de analiză care se bazează pe un dispozitiv Ubuntu de înaltă capacitate de calcul cu instrumentele aferente (R-Studio, server Shiny), necesare pentru manipularea datelor după și în timpul procesului de colectare a datelor.

Infrastructura INS nou achiziționată este protejată de o VPN, permițând doar personalului autorizat să acceseze componentele sale. INS gestionează accesul VPN în conformitate cu procedurile interne.

Serverul de analiză este configurat cu conturi individuale pentru a urmări accesul fiecărei persoane și există spații partajate folosite pentru a păstra datele colectate și prelucrate cu permisiunile necesare, pentru a permite doar persoanelor responsabile să le citească sau să le manipuleze. Conturile individuale ale serverului de analiză sunt gestionate de INS conform procedurilor interne.

3. Recomandări pentru protecția și securitatea datelor viitoarelor sondaje

Pentru viitoarele recensăminte ale populației, ar trebui să fie pus un accent deosebit pe testarea securității aplicației SuSo și a componentei de autorecenzare. Toate versiunile selectate ale SuSo și ale autorecenzării ar trebui să facă obiectul evaluărilor de securitate efectuate de entități independente.

Programarea testelor de securitate pentru următoarele recensăminte ar trebui să fie planificată cu atenție, cu suficient timp în prealabil, pentru a permite echipelor responsabile (SuSo, autorecenzare) să aplice recomandările din evaluarea securității în noile versiuni sau ramuri și să soluționeze de manieră corespunzătoare potențialele probleme de securitate.

În ceea ce privește RPL, runda 2021, componentele de colectare a datelor au fost găzduite de STS. Aceeași abordare este recomandată pentru recensămintele viitoare. În cazul în care INS intenționează să utilizeze propria infrastructură pentru colectarea datelor, trebuie să ia în considerare provocările generate de această intenție care constau în, dar nu se limitează la:

- existența unui centru de date secundar pentru scenarii de recuperare în caz de dezastru;
- implementarea procedurilor pentru activarea asistenței 24x7 pentru răspunsul la incidente;
- dezvoltarea capacităților tehnice ale unei echipe cu cunoștințe adecvate pentru a suporta configurații de înaltă disponibilitate, configurații de securitate (inclusiv, dar fără a se limita la sistemele de prevenire a intruziunilor și de detectare a intruziunilor); și
- provocări de performanță legate de sisteme, aplicații și baze de date.

În cele din urmă, fiecare problemă de securitate și protecție a datelor ar trebui să se concentreze pe fluxul de recensământ convenit. Fluxul în sine ar trebui analizat din perspectiva securității și protecției datelor, pentru a evita orice încălcare care ar putea compromite securitatea generală a sistemului și protecția datelor.

4. Anexe

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

Competența face diferența!

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