



ROMÂNIA

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

REZULTATUL nr. 7b

Raport privind serviciile de asistență tehnică oferite Beneficiarului, pentru

Recomandări, bune practici și îndrumări pentru elaborarea unei metodologii de personalizare a unei soluții IT pentru geo-referențierea exploatațiilor agricole în unități statistice de tip grid RGA2020

Decembrie 2021

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 elaborat de personalul Băncii Mondiale. Constatările, interpretările și concluziile exprimate în acest document nu reflectă neapărat 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 aprobarea sau acceptarea acestor granițe. Prezentul raport nu reprezintă neapărat poziția Uniunii Europene sau a Guvernului României.

Declarația privind drepturile de autor

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

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

Prezentul raport a fost transmis în luna decembrie 2021 și revizuit în luna noiembrie 2022, în cadrul Acordului privind Serviciile de Asistență Tehnică Rambursabile pentru 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 corespunde Rezultatului 7b din cadrul acordului mai sus menționat.

Cuprins

Abrevieri și acronime	7
Introducere	8
1. Metodologia aplicată pentru georeferențiere	9
Control cartografic (testat în Windows Edge și Google Chrome)	11
Tipuri de grid.....	11
ID-uri de grid.....	20
Formate pentru descărcare	22
Descărcarea grilei de referință.....	22
2. Pregătirea datelor și variabilele după tipul de fermă	23
Organizarea bazei de date RGA	23
Actualizarea bazei de date.....	25
3. Instalarea și configurarea aplicației	26
Descrierea codului sursă.....	26
Cerințe de sistem (dependențe hardware și software).....	26
Kit de instalare, procedura de instalare și parametrii de configurare.....	26
4. Recomandări și bune practici	27
5. Anexe	29
Anexa 1: Filtrarea observațiilor prin selecție din tabelul centralizator cu celulele de grid	29
Anexa 2: Descărcarea datelor	34
Datele în puncte.....	34
Datele privind numărul de celule	35
Fișierul raster.....	35
Anexa 3: Microdate RGA	37
Anexa 4: Variabile în georeferințele exploatațiilor agricole	41
Anexa 5: Script R pentru corespondențe de date și corecții de coordonate GPS corespunzătoare localizării exploatației	43
Anexa 6: Script R pentru transformarea fișierelor FOXCON în postgres(-postgis db)	50
Descriere	50
Încărcarea pachete și date.....	50
Curățare și Imputare	51
Extragerea I_KEY lipsă	54
Imbinarea datelor.....	55
Generarea unui obiect spatial din date	55
Scrierea in baza de date DB	56

Încărcarea gridului de referință	57
Cearea obiect spatial cu terra	57
Proiectarea datelor pe Lambert Azimuthal Equal Area (epsg 3035)	57
Crearea unui singur grid pentru toate datele	57
Sciere in baza de date.....	57

Mulțumiri

Prezentul raport a fost elaborat sub îndrumarea lui Michael Wild, statistician superior la Banca Mondială, cu sprijinul echipei locale de experți. Echipa dorește să își exprime mulțumirea față de reprezentanții Institutului Național de Statistică (INS), Florentina Gheorghe (Director General), Silvia Pisica (Director General și Manager de proiecte) și echipele acestora de specialiști pentru colaborarea lor constructivă.

Figuri

Figura 1: Ecranul de start al aplicației.....	10
Figura 2: Grid vid (gol)	11
Figura 3: Selectarea variabilelor disponibile pentru agregare.....	12
Figura 4: Înălțimea barei indică valoarea agregată.	13
Figura 5 - Diferite rezoluții pentru variabila selectată (10km, 5km, 1km)	14
Figura 6: Mesaj de avertizare care informează utilizatorul privind excluderea observațiilor la crearea gridului.....	15
Figura 7: Toate punctele.....	15
Figura 8: Punct de plasare a cursorului deasupra.....	16
Figura 9: Tabel lateral cu toate variabilele (în dreapta)	16
Figura 10: Datele privind distanța permit inspectarea unor distanțe mai mari	17
Figura 11: Datele privind distanța permit inspectarea unor distanțe mai mici.....	17
Figura 12: Controale dinamice ajustează în funcție de distanța minimă și maximă.....	18
Figura 13: Este posibilă crearea unor subseturi de unități în funcție de distanța lor maximă.....	18
Figura 14: Unități APIA și Non-APIA cu și fără grid.....	19
Figura 15: TOATE DATELE și celulă unică de grid.....	19
Figura 16: Cub de date pentru o singură caracteristică	20
Figura 17: Selecție de celule de grid prin care se identifică elemente singulare sau alte celule care încalcă limitele de separație.	21
Figura 18: Celulă de grid selectată pentru inspectare și urmărire ulterioare.....	21
Figura 19 - Comenzi pentru actualizarea bazei de date	25
Figura 21: Tabel cu numărul de celule de grid în dreapta hărții	29
Figura 22: Sortare după cea mai mare valoare	29
Figura 23: Sortare după cea mai mică valoare	30
Figura 23: Celulă de grid selectată.....	30
Figura 25: Selectare celulă de dimensiuni medii	31
Figura 26: Unități Apia din cadrul celulei de grid	31
Figura 27: Unități Non-Apia din cadrul celulei de grid	32
Figura 28: Selectare celulă de grid cu latura de 5 km	32
Figura 29: Selectare celulă de grid cu latura de 10 km	33
Figura 30: Încărcarea datelor pentru un anumit județ creează fișierul cu puncte și fișierele cu numărul de celule pentru descărcare.	34
Figura 31: Fișier .zip cu date în puncte și numărul de celule.....	35
Figura 32: Datele raster privind populația se pot descărca după ce devin vizibile pe hartă.	35
Figura 33: Fișier .zip cu date în puncte, număr de celule și date de tip raster.	35
Figura 34: Afișare QGIS a datelor raster pentru județe diferite și cu rezoluții diferite, toate pe o singură hartă.	36

Casete

Caseta 1 – Obiectivele instrumentului de distribuție în rețea a datelor RGA	9
---	---

Abrevieri și acronime

APIA	Agencia de Plăți și Intervenție pentru Agricultură
CASS	Sistem de topografiere asistat de calculator
CRS	Sistem de referință de coordonate
DPS	Prelucrare și stocare a datelor
DTS	Direcții/oficii teritoriale de statistică
CE	Comisia Europeană
UE	Uniunea Europeană
GA	Aplicație grid
RGA2020	Recensământul General Agricol 2020
GEOLOC	Aplicație informatică de localizare geografică
GIS	Sistem informațional geografic
GPS	Sistemul de Poziționare Globală
IFS	Statistici integrate privind fermele
INS	Institutul Național de Statistică
INSPIRE	Infrastructură pentru informații spațiale în Europa
IT	Tehnologia informației
RAS	Servicii de Asistență Tehnică Rambursabile
SDC	Controlul privind divulgarea informațiilor statistice

Introducere

Scopul prezentului raport este de a prezenta recomandări, bune practici și îndrumări privind dezvoltarea unei metodologii de personalizare a unei soluții IT de georeferențiere a exploatațiilor agricole în unități statistice de tip grid pentru Recensământul General Agricol 2020 (RGA2020).

Prezentul raport face parte din livrabilele din cadrul Acordului privind Serviciile de Asistență Tehnică Rambursabile (RAS) pentru Consolidarea Sistemului Statistic Național (proiect nr. P167217) și corespunde Rezultatului 7b: *Raport privind serviciile de asistență tehnică oferite Beneficiarului pentru Recomandări, bune practici și îndrumări în elaborarea unei metodologii de personalizare a unei soluții IT pentru geo-referențierea exploatațiilor agricole în unități statistice de tip grid pentru RGA2020*. Proiectul este implementat de către Institutul Național de Statistică, având sprijinul Băncii Mondiale.

Prezentul raport este structurat în patru secțiuni.

Prima secțiune oferă o descriere a metodologiei aplicate pentru georeferențierea fermelor recensate pe parcursul procesului de colectare a datelor din cadrul RGA2020, cu scopul de a afișa separat sau în comun datele de recensământ colectate APIA și Non-APIA, de a crea unități statistice de tip grid cu dimensiuni diferite de celule, de a îndeplini cerințele Directivei INSPIRE conform Regulamentului (UE) 2018/1091 privind Statistica Integrată a Fermei și de abrogare a Regulamentului (CE) nr.1166/2008 și (UE) 1337/201 și de a analiza abateri ale coordonatelor introduse față de cele inițiale colectate.

O descriere a organizării bazei de date RGA, respectiv sursele de date utilizate, structura microdatelor și procesul de pregătire (curățarea datelor) necesare în procesul de georeferențiere se regasesc în a doua secțiune..

A treia secțiune prezintă detalii privind codul sursă, cerințele de sistem – dependențe hardware și software, procedura de instalare și parametrii de configurare pentru instalarea și configurarea corectă a aplicației.

O listă cu recomandări specifice, pentru ca personalul INS să practice și să utilizeze în mod optim aplicația, este indicată în a patra secțiune.

Anexele raportului oferă utilizatorilor aplicației informații privind filtrarea observațiilor prin selecție din tabelul centralizator al celulelor de grid și procesul de descărcare a datelor, microdatele RGA și lista variabilelor din baza de date RGA utilizate în georeferențiere, scriptul R pentru corespondența datelor și corecții ale coordonatelor GPS, precum și scriptul R pentru transformarea fișierelor FOXCON în baza de date Postgres (postgis db).

1. Metodologia aplicată pentru georeferențiere

Georeferențierea exploatației agricole în unități statistice de tip grid reprezintă procesul de asociere între localizarea (GPS) exploatației agricole și codul de grid corespunzător, în conformitate cu regulamentele¹ specifice relevante pentru statisticile integrate ale fermelor, inclusiv GAC2020 și reglementările UE².

Soluția metodologică și IT aplicată în georeferențierea fermelor recensate pe parcursul procesului de colectare a datelor din cadrul RGA abordează fermele în mod diferit: fermele care beneficiază de subvenții și sunt incluse în lista APIA, care sunt identificate pe baza coordonatelor GIS din sursa APIA; și fermele mici, care nu sunt incluse în lista APIA, pentru care au fost colectate coordonate GIS în conformitate cu criteriile de localizare.

Box 1 – Obiectivele instrumentului de distribuție în rețea a datelor RGA

Sarcina principală a aplicației este de a produce grila și codurile de grilă pentru toate exploatațiile agricole.

Toate vizualizările sunt furnizate cu intenția de a sprijini personalul INS în observarea și examinarea datelor la nivel de județ și celulă grilă.

Codul sursă R, inclusiv documentația acestuia, este furnizat INS. Acest lucru oferă experților R din INS posibilitatea de a implementa ei înșiși orice modificări sau funcționalități ulterioare ca parte a procesului de consolidare a capacităților instituționale.

GA (Aplicația Grid) permite vizualizarea unor seturi vaste de date de recensământ, precum și analize spațiale corespunzătoare, fie ca puncte, fie ca seturi de date de grid. Obiectivul este de a sprijini activitatea statisticienilor prin informații GIS de bază. GA pentru România vizează următoarele aspecte relevante:

1. Afișarea separată, precum și în comun, a datelor de recensământ colectate pentru exploatațiile agricole APIA și Non-APIA.
2. Crearea de unități statistice de tip grid cu dimensiunile celulelor de 1km, 5km și 10 km.
3. Atribuirea unui COD de CELULĂ, după cum prevede directiva INSPIRE, pentru fiecare dintre unitățile din recensământ și pentru fiecare dintre diferitele dimensiuni de celulă de grid.
4. Generarea unităților de tip grid pentru o mai bună vizualizare și inspecție, utilă și în exportul datelor din grid și aplicabilă tuturor variabilelor numerice furnizate³.
5. Posibilitatea de a vizualiza abateri între coordonatele inițiale colectate și coordonatele introduse.

¹ D2.8.I.2 Specificații ale Datelor privind Sistemele de Carioaj Geografic – Orientări Tehnice; INSPIRE, 2014; Directiva 2007/2/CE a Parlamentului European și a Consiliului din 14 martie 2007 de instituire a unei infrastructuri pentru informații spațiale în Comunitatea Europeană (INSPIRE).

² Regulamentul UE 2018/1091 privind Statistica Integrată a Fermei și de abrogare a Regulamentului (CE) nr.1166/2008 și (UE) 1337/2011; Manualul Eurostat de statistică integrată a fermelor

³ Aplicația permite afișarea tuturor variabilelor corect furnizate după cum acestea devin disponibile prin conexiunea de încărcare/a bazei de date. Din motive de vizualizare, aceasta exclude și observații aflate mai jos (mai sus) de pragul de 5% sau 95%.

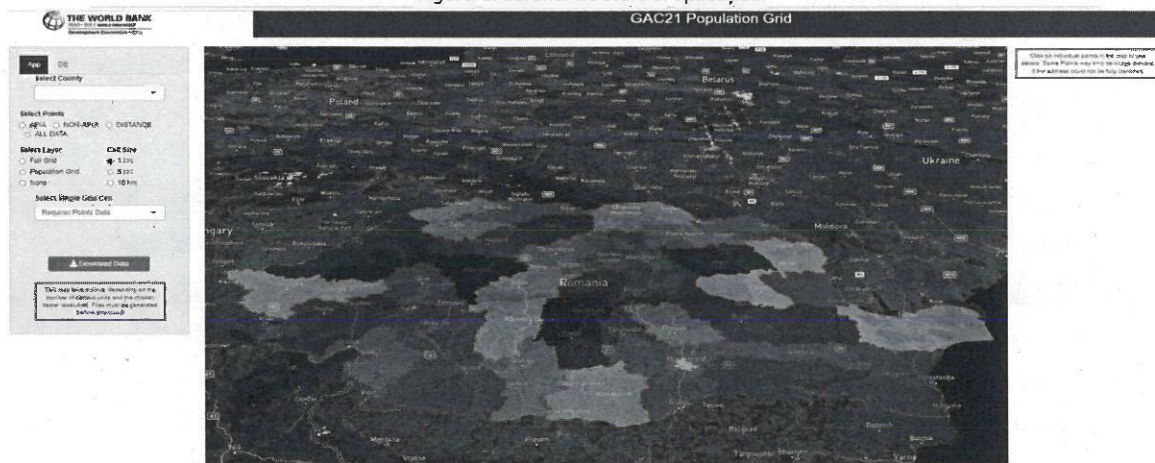
Aplicația finală rulează pe propriul sistem IT al INS și se va conecta la, și va utiliza propria sa bază de date PostGIS-Postgres⁴. Echipa BM a oferit asistență pentru instalarea pe sistemul integrat IT al INS și a oferit îndrumări personalului INS pentru utilizarea, instalarea și operarea corespunzătoare. Sistemul de referință de coordonate (CRS) este cel recomandat de directiva europeană INSPIRE⁵, care folosește o proiecție de suprafețe egale Proiecția azimutală echivalentă Lambert ETRS89, EPSG:3035. În funcție de CRS ales, codurile corespunzătoare de grid sunt create și atribuite fiecărei unități.

Aplicația prelucrează foarte multe date și depinde de disponibilitatea resurselor care au fost proiectate și parametrizate în scopurile RGA prin Output 4.1b., de numărul și tipul de procese de calcul rulate de sistemul informatic INS („Server de analiză”) atunci când aplicația este în uz și numărul de utilizatori concurenți (recomandat până la 3 care rulează simultan calculul datelor RGA pe grile) și, prin urmare, s-ar putea ca, uneori, aplicația să arate „înghețari” de ecran sau alte comportamente neașteptate. În astfel de cazuri, se recomandă reinițializarea (refresh) aplicației și începerea din nou a calculelor vizate.

Pe lângă faptul că permite INS să își îndeplinească obligația față de Eurostat conform cu Regulamentul UE 2018/1098, prin crearea unităților statistice de tip grid și atribuirea codurilor corespunzătoare de celulă pentru seturile de date, această aplicație permite și crearea unui raster cu dimensiunea celulei de 1000 (5000, 10.000) de metri⁶, care se poate utiliza pentru predicții privind acoperirea, optimizarea anchetelor agricole în perioadele intercenzitare sau în orice alt scop analitic determinat de activități viitoare.

După deschiderea aplicației în navigatorul preferat⁷, **prima selecție** care se va realiza este întotdeauna JUDEȚUL (COUNTY) de interes. Există o dublă motivație, o mai bună vizualizare pentru utilizator și, în al doilea rând, se reduce volumul de lucru și, ulterior, timpul de procesare. Selectarea județului este urmată de diferitele tipuri de operațiuni.

Figura 1: Ecranul de start al aplicației.



⁴ <https://postgis.net/>

⁵ Infrastructura pentru informații spațiale în Europe; <https://inspire.ec.europa.eu/>

⁶ 1000 m este o dimensiune destul de comună a celulei, utilizată în proiecții/vizualizări de populații, însă, întrucât aplicația este în R iar codul sursă este disponibil public, sunt posibile modificări pentru dimensiuni alternative de celule.

⁷ MS Edge, prin versiunea sa cea mai recentă, a oferit cele mai bune rezultate în materie de performanțe, Google Chrome funcționează bine, însă ar putea necesita mai multă memorie de lucru.

Control cartografic (testat în Windows Edge și Google Chrome)

Controalele cartografice se bazează pe motorul Web GL (care s-ar putea să nu ruleze pe versiuni mai vechi ale navigatorului sau pe tablete). Principalele controale sunt

- i. Clic dreapta/Mutare cursor → deplasarea hărții
- ii. Clic dreapta/Ctrl/Mutare cursor → schimbarea unghiului de vizualizare
- iii. Rotița de derulare +/- → apropiere în/îndepărtare din prim plan

Acestea trebuie să funcționeze, în mod normal, în toate navigatoarele web standard contemporane, însă au fost testate doar în Windows Edge⁸ și Google Chrome⁹ până în acest moment.

Tipuri de grid

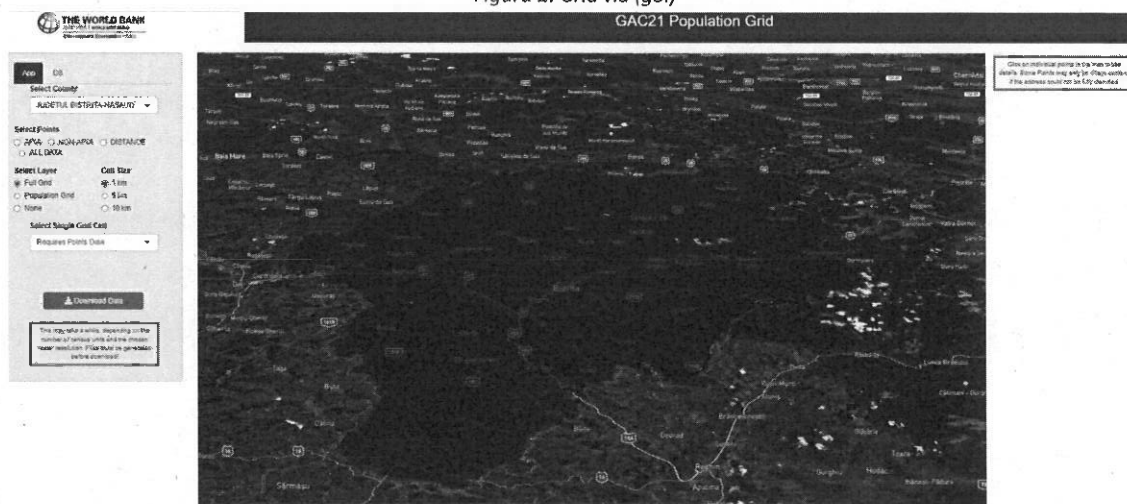
Aplicația permite (i) crearea unui grid vid care se va suprapune peste datele în puncte, precum și (ii) agregarea datelor în puncte cu diferite date statistice (i.e., total, medie, deviație standard) pentru o variabilă selectată, la nivel de celulă. Aceste valori ale populațiilor din grid se pot exporta și folosi ca date de intrare în alte scopuri, dar și pentru vizualizări accesibile public¹⁰.

1. GRID COMPLET

După selectarea județului în secțiunea SELECT COUNTY (*Selectați județul*), se selectează opțiunea Full Grid (*Grid complet*) în secțiunea SELECT LAYER (*Selectați strat*). Dimensiunea implicită a gridului este de 1 km și, dacă utilizatorul dorește să o schimbe, trebuie să facă acest lucru înainte de a selecta opțiunea SELECT LAYER (*Selectați strat*).

Opțiunea FULL GRID (*Grid complet*) permite utilizatorului să inspecteze vizual celule de grid care încalcă numărul minim de observații necesar unui set de date pentru uz public, precum și să inspecteze grupuri de observații, să verifice distanțe utilizând gridul drept referință etc. Când este creat gridul populației, se va aplica aceeași dimensiune de celulă precum în gridul vid produs în această secțiune.

Figura 2: Grid vid (gol)



⁸ Versiunea 94.0.992.50 (Versiune oficială) (64-bit)

⁹ Versiunea 95.0.4638.54 (Versiune oficială) (64-bit)

¹⁰ Un aspect foarte important în crearea seturilor de date pentru uz public/vizualizabile public este cel al Controlului privind Divulgarea Informațiilor Statistice, care va fi abordat, în mod ideal, atât cu sdcMicro, cât și cu sdcSpatial, ambele disponibile în R. Mai multe detalii sunt oferite în Anexă.

2. GRIDUL POPULAȚIEI (Statistic)

După selectarea județului, se selectează opțiunea Population Grid (*Grid populație*) în secțiunea SELECT LAYER (*Selectați strat*). Când este selectată această opțiune, va apărea un nou meniu derulant SELECT VARIABLE FOR AGGREGATION (*Selectați variabilă pentru agregare*). Acest meniu derulant arată toate denumirile variabilelor din setul de date identificate de sistem drept numerice, cu excepția coordonatelor principale¹¹.

Selectarea unei variabile creează un grid cu DIMENSIUNEA CELULEI și fie va calcula: (i) totalul, fie (ii) media, fie (iii) fie deviația standard, pentru toate valorile observate și valide ale variabilei selectate din toate observațiile din celula corespunzătoare de grid. Rularea calculațiilor depinde de dimensiunea celulei¹².

Figura 3: Selectarea variabilelor disponibile pentru agregare.

App DB

Select County

JUDETUL BISTRITA-NASAUD

Select Points

☒ APIA ☐ NON-APIA ☐ DISTANCE
☐ ALL DATA

Select Layer

☐ Full Grid ☒ Population Grid ☐ None

Cell Size

☐ 1 km ☐ 5 km ☒ 10 km

Aggregation Type

☒ Mean ☐ Total ☐ SD

Select variable for Aggregation

C1P2A

Select Single Grid Cell

Download Data

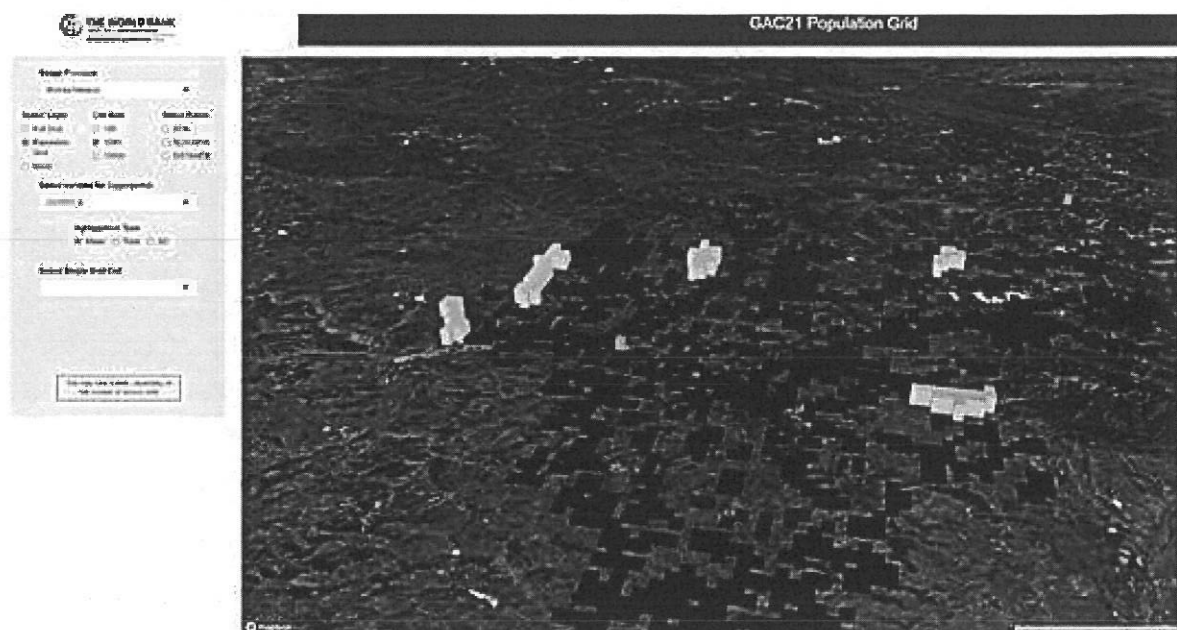
This may take a while, depending on the number of census units and the chosen raster resolution. Files must be generated before download!

Rezultatul final afișat și exportat conține doar valori din celule non-vide. Plasând cursorul deasupra, se poate afișa de asemenea valoarea celulei corespunzătoare. Înălțimea fiecărei bare constituie valoarea sa relativă, culoarea având strict rol de diferențiere.

¹¹ Adică, dacă o variabilă furnizată conține orice șir, fie și doar ca substituent, precum "###.###", sistemul nu va identifica variabila drept numerică și, ca urmare, nu o veți putea selecta în vederea procesării.

¹² Când gridul este recalculat, va apărea o mică fereastră jos în partea dreaptă și, de aceea, trebuie să schimbați valorile cu atenție deoarece procesul ar putea fi destul de îndelungat

Figura 4: Înălțimea barei indică valoarea agregată.



În plus, aplicația adaugă variabile DIST (care este distanța calculată). Cea din urmă este, de asemenea, disponibilă pentru agregare. După Agregare, utilizatorul obține o reprezentare 3D unde înălțimea și culoarea indică suma valorică din observațiile din limitele celulei de grid.

Gridul populației creat pentru vizualizare poate fi descărcat și într-un format **GeoTIFF**¹³. Gridurile populației fac obiectul mai multor utilizări, una dintre acestea fiind combinarea cu alte date geospațiale, furnizate în mod ideal într-un format similar, precum fotografii aeriene sau date de teledetecție. În plus, gridul oferă și un anumit grad de control privind divulgarea informațiilor statistice dacă agregarea respectă orientări stricte privind dimensiunile minime.

3. DIMENSIUNEA CELULEI

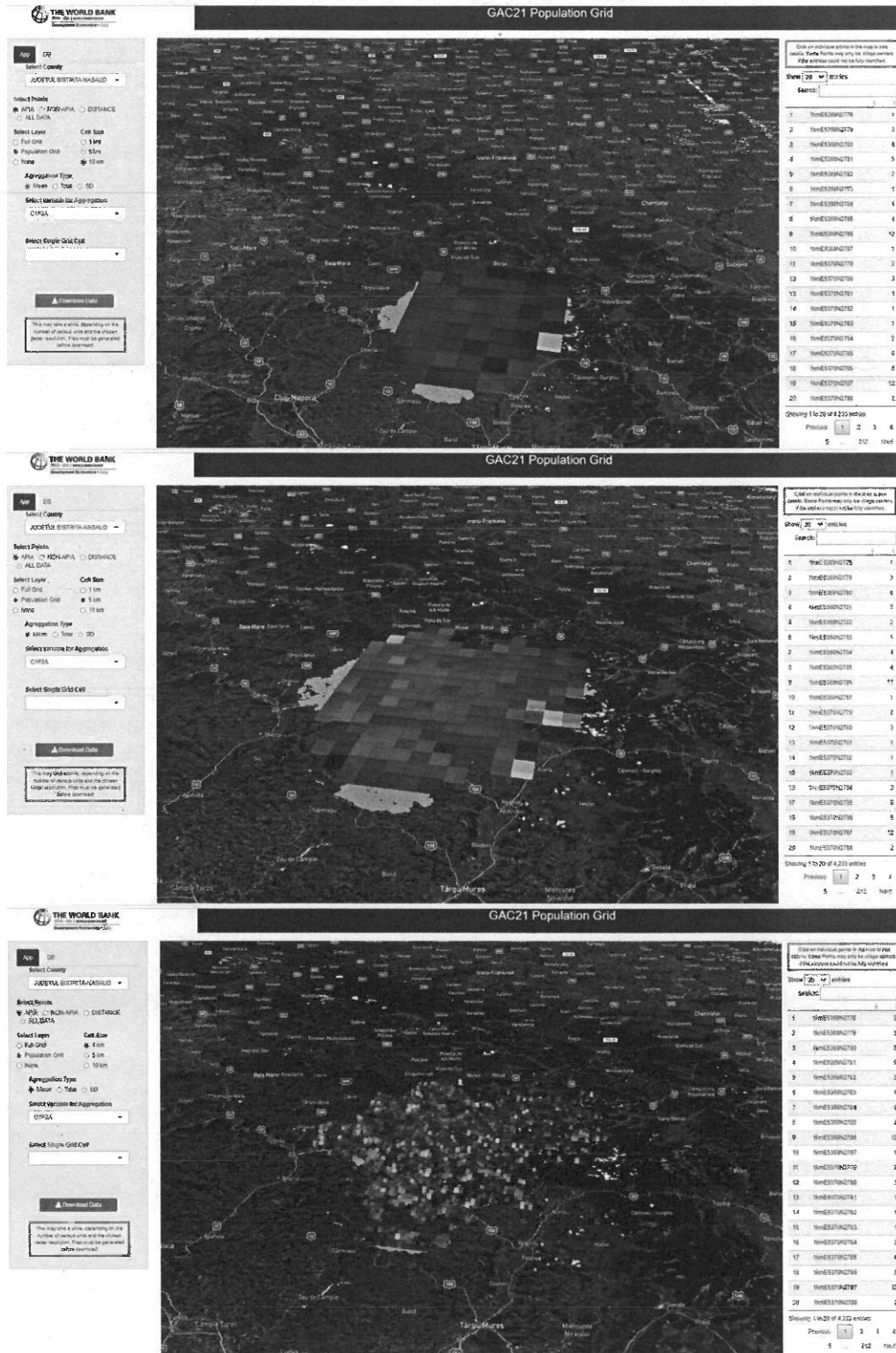
Dimensiunea celulei eset posibilă pentru 1000m 5000m sau 10000m. Dimensiunea celulei specifică dimensiunea celulelor FULL GRID (*Grid complet*) și a celulelor POPULATION GRID (*Grid populație*).

O dimensiune mai mică a celulei necesită un număr mai mare de celule pentru a acoperi suprafața, rezultând timpi mai îndelungați de procesare și cerințe de sistem mai ridicate. Prin urmare, se recomandă verificarea că există suficiente rezerve pentru resursele de calcul.

Figurile de mai jos arată diferitele rezoluții pentru variabila selectată C01P2A, pentru Bistrița-Năsăud.

¹³ GeoTIFF, Revizia 1.0 (loc.gov)

Figura 5 - Diferite rezoluții pentru variabila selectată (10km, 5km, 1km)

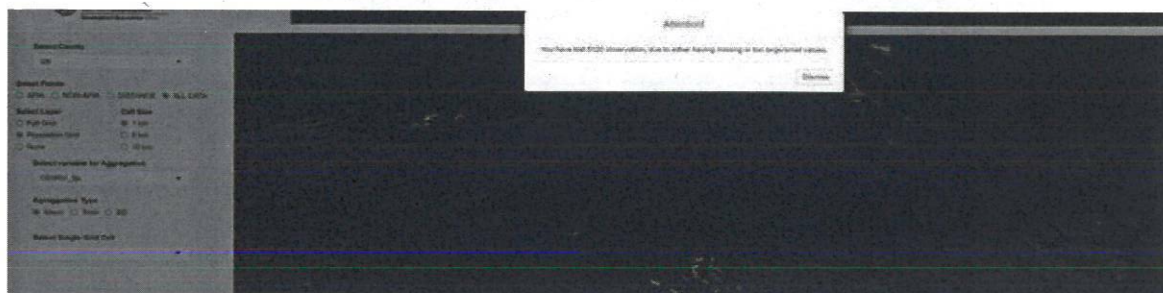


Atunci când este selectată rezoluția ar trebui luate în considerare mediile de procesare, după cum poate vedea mai sus în numărul de celule per rezoluție. De asemenea, fișierul raster descărcat va fi mult mai mare cu cât scade dimensiunea rezoluției.

Înălțimea fiecărei bare este valoarea statistică selectată corespunzătoare, adică medie, totală sau deviație standard a variabilei selectate. În acest fel, orice deviații normale sau extreme pot fi mai ușor reperate cu ochiul liber. Prin plasarea cursorului pe oricare dintre bare, se afișează înălțimea acestuia (adică, valoarea agregată, care va fi medie, totală sau deviație standard).

Întrucât aplicația permite selectarea și încărcarea oricăror variabile, în scopuri de vizualizare, unele valori extreme ar putea fi reduse strict pentru procedura de cartografiere. Aplicația va spune utilizatorului câte observații nu se află pe hartă.

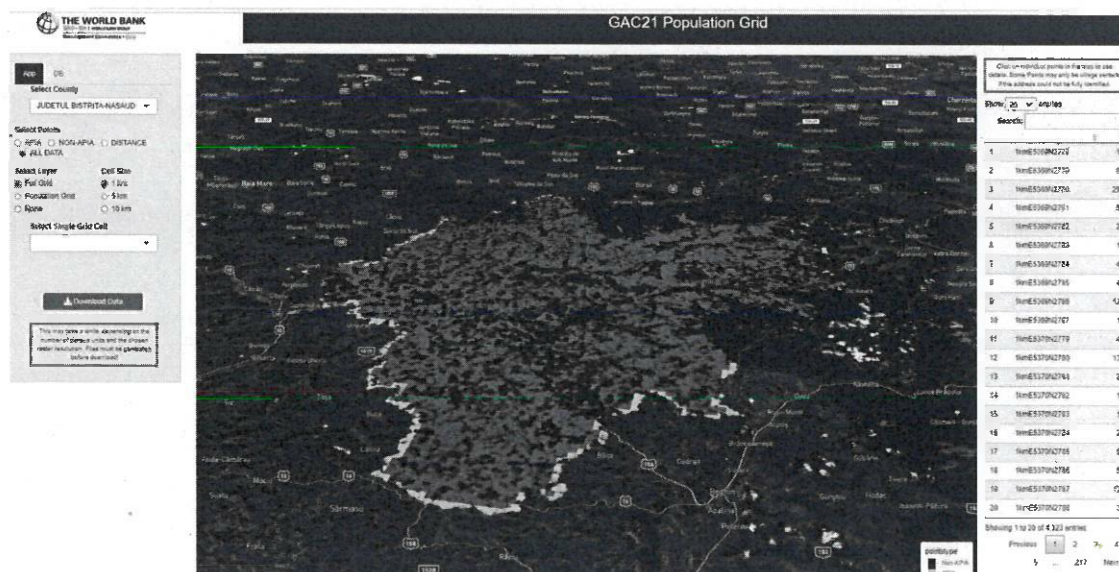
Figura 6: Mesaj de avertizare care informează utilizatorul privind excluderea observațiilor la crearea gridului.



4. SELECTAREA PUNCTELOR

Funcția vă permite să selectați și să afișați, în cele din urmă, datele colectate. Selecția va fi fie setul de date APIA, seturile de date NON-APIA, setul de date Apia cu DISTANCE (Distanța) între amplasamentul inițial și amplasamentul introdus, fie ALL DATA (Toate datele), prin combinarea seturilor de date APIA și Non-APIA.

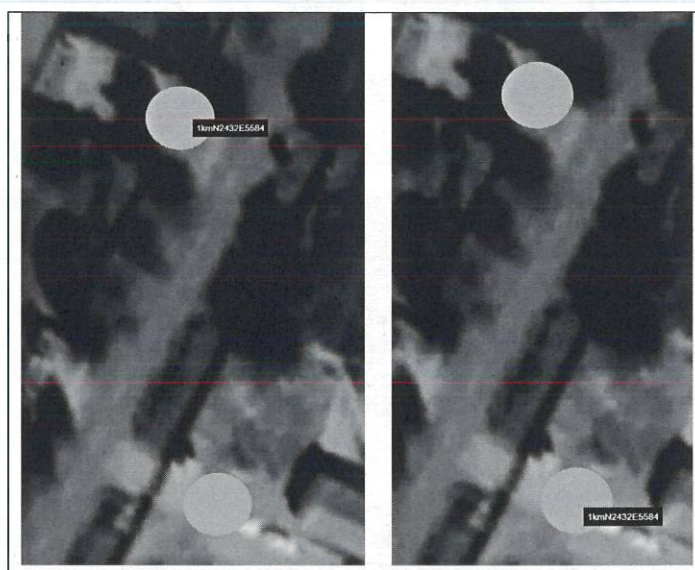
Figura 7: Toate punctele



Datele în materie de DISTANCE (*Distanța*) reprezintă ambele tipuri de date combinate și permit o comparație între amplasamentul introdus și cel efectiv de unde au fost colectate datele. Aceste date vor fi adăugate setului de variabile și afișate ca puncte, dar și într-o vizualizare de tip grid, facilitând reperarea oricăror deviații substanțiale după înălțimea celulei de grid.

Prin plasarea **cursorului deasupra**, se va afișa codul corespunzător de celulă atribuit respectivei unități de recensământ prin sistemul curent de coordonate după cum este furnizat de coordonatele sferice în grade zecimale, precum și distanța între datele inițiale și cele introduse.

Figura 8: Punct de plasare a cursorului deasupra



Cu un **click** pe orice punct se afișează în partea stângă, de asemenea, un tabel care conține toate variabilele pentru observația corespunzătoare. Acest lucru înseamnă că utilizatorul poate, de exemplu, să utilizezi cheia de interogare pentru a verifica detalii suplimentare în sistemul Survey Solutions.

Figura 9: Tabel lateral cu toate variabilele (în dreapta)



Datele din categoria DISTANCE (*Distanță*) permit comparații între coordonatele „vechi” și cele „noi”. Întrucât unele dintre coordonate au fost deja introduse, acest set de date prezintă coordonatele vechi și pe cele noi pe aceeași hartă și permite o comparație directă.

Figura 10: Datele privind distanța permit inspectarea unor distanțe mai mari

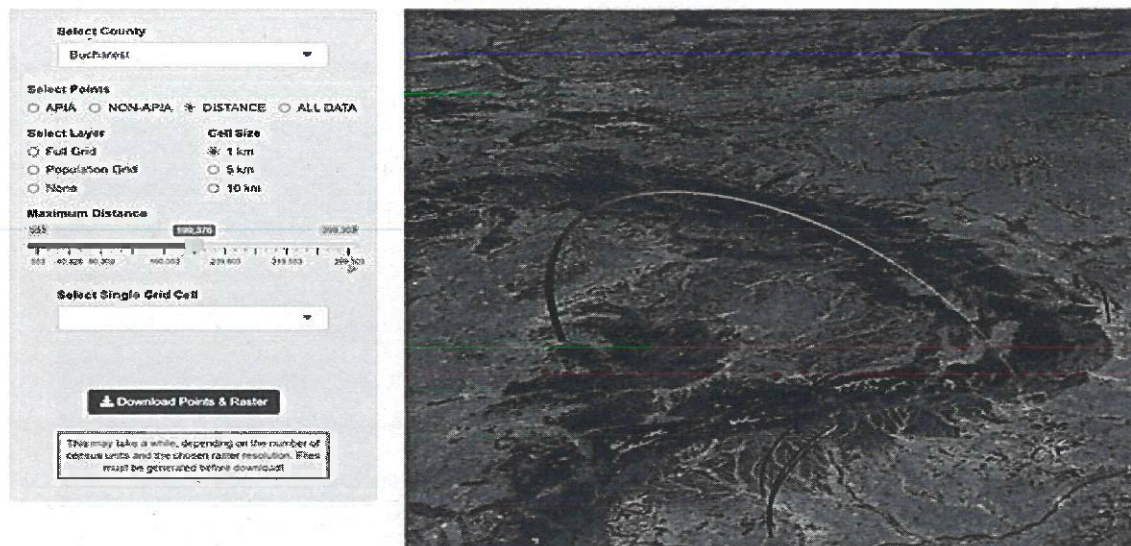


Figura 11: Datele privind distanța permit inspectarea unor distanțe mai mici



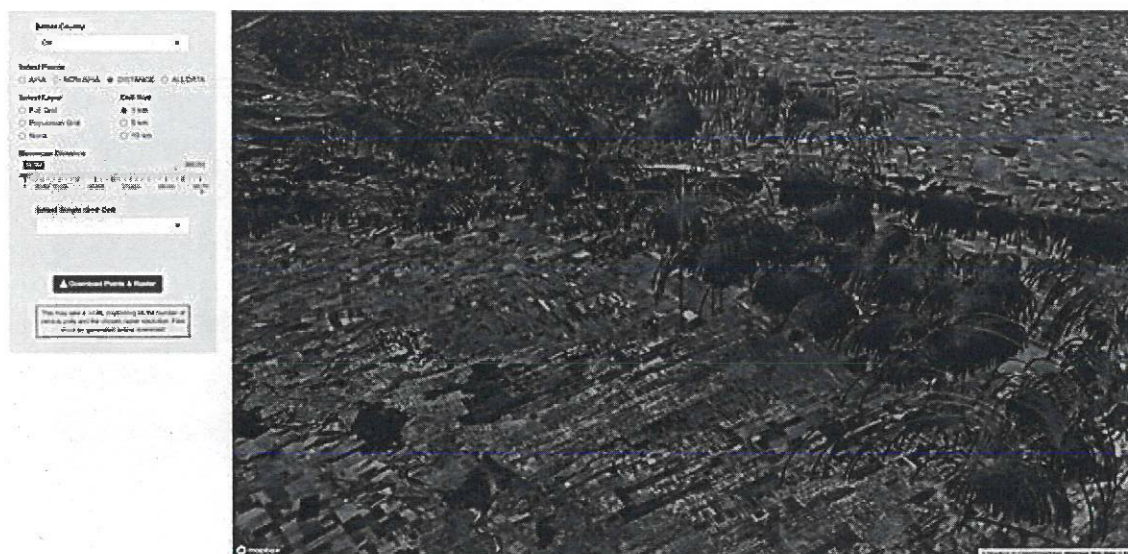
Pentru a facilita orice inspecție pe bază de distanță și a observa doar un subset, GA afișează și o bară de selectare a Maximum Distance (*Distanță maximă*) pentru ajustarea distanței maxime afișate. Este o funcție extrem de utilă, care aduce în prim plan multe unități și lasă doar câteva în plan îndepărtat.

Figura 12: Controale dinamice ajustează în funcție de distanța minimă și maximă.



Crearea unor subseturi de date referitoare la o anumită Maximum Distance (*Distanță maximă*) reîncarcă harta și permite o mai bună inspectare a acestor unități.

Figura 13: Este posibilă crearea unor subseturi de unități în funcție de distanța lor maximă.



Modul ALL DATA (*Toate datele*) permite vizualizarea directă a unităților APIA și Non-APIA. Este un aspect deosebit de util atunci când se dorește inspectarea distribuției într-o zonă geografică mai îngustă sau într-o celulă de grid, a se vedea figurile de mai jos.

Figura 14: Unități APIA și Non-APIA cu și fără grid

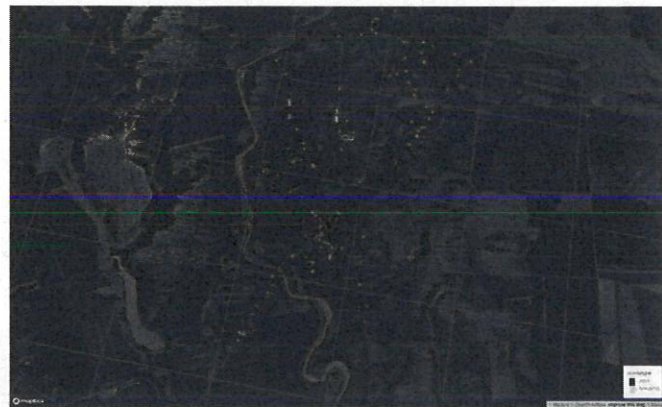


Figura 15: TOATE DATELE și celulă unică de grid

Basemap Drawing

Basemap

Basemap Points

☐ APIA ☐ Non-APIA ☐ DISTANCE ☐ ALL DATA

Basemap Layout

☐ Full Grid ☐ Grid Size

☐ Precision: 10m ☐ 100m

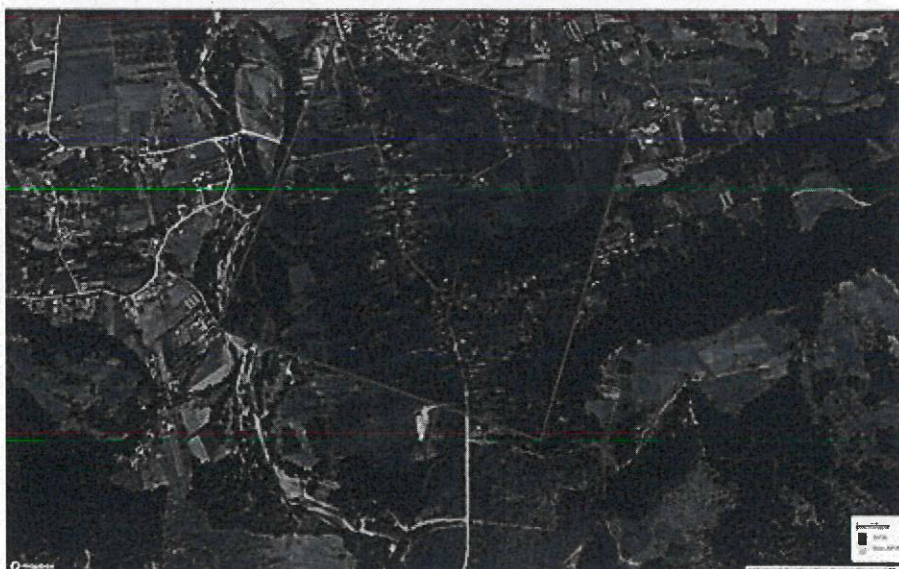
☐ 1000m ☐ 10000m

Basemap Single Grid Cell

Basemap Single Grid Cell

Download Points & Basemap

The data is for the purpose of the Basemap & Basemap Points. The data is for the purpose of the Basemap & Basemap Points. The data is for the purpose of the Basemap & Basemap Points.



ID-uri de grid

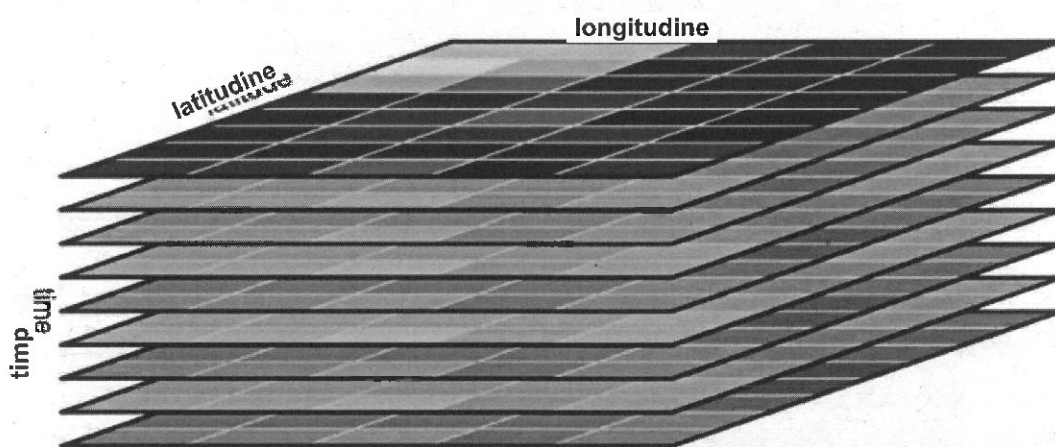
Identificatoarele de grid (ID-uri) sunt produse în conformitate cu directiva INSPIRE (conform cu Regulamentul UE 2018/1098) și permit organizarea (agregarea) unităților de recensământ la nivelul unei distanțe predefinite. Primul pas este, prin urmare, proiecția coordonatelor în spațiul de coordonate, respectiv CRS corespunzător. Acest spațiu de coordonate există pentru a facilita generarea unor statistici georeferențiate și constituie fundamentul Sistemului European Unificat de Codificare a gridurilor, având următoarele atribute:

- ușor de manipulat,
- ierarhic,
- bazat pe un Sistem European Unificat de Codificare a gridurilor,
- bazat pe unități cu suprafață egală,
- adoptă ETRS-LAEA

Deși raportarea valorilor și a codurilor pentru griduri de 1 km face parte din obligația INS față de Eurostat, este utilă folosirea în continuare a sistemului de griduri și pentru alte dimensiuni furnizate de către sistem. Se obține astfel o stocare eficientă a datelor georeferențiate.

Întrucât unitățile standard de recensământ bazate pe populații sunt considerate staționare, un astfel de cod de grid este creat o singură dată și poate fi apoi folosit în activități ulterioare de colectare a datelor, cum ar fi anchetele. Analiza și vizualizarea geospațiale sunt facilitate prin stocarea datelor în acest mod.

Figura 16: Cub de date pentru o singură caracteristică



Sursa: preluat din <https://keen-swartz-3146c4.netlify.app/datacube.html>

Diferitele straturi indică datele colectate în decursul timpului pentru diversele unități și agregate în respectivele valori de celulă de grid. Este un mod eficient de stocare a datelor spațio-temporale¹⁴. Codurile sunt astfel generate o singură dată și în conformitate cu gridul subiacent aferent de populație (adică, 1 km, 5 km sau 10 km).

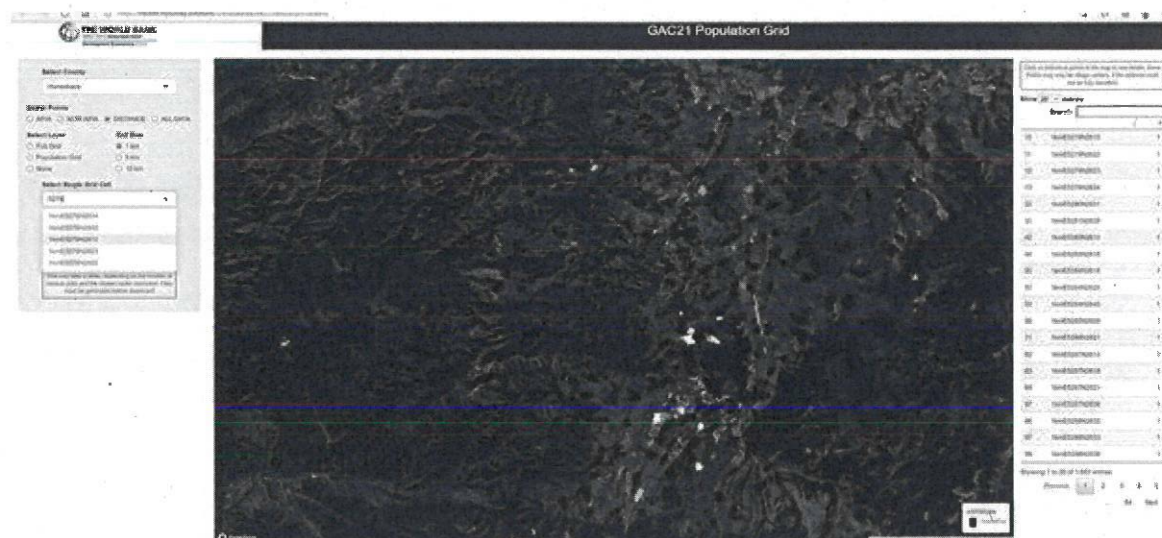
Activitățile ulterioare de procesare a datelor pot agrega observații tabelare după codul de grid și pot raporta imediat rezultatele la nivel de celulă, în locul procesării prealabile a informațiilor în datele

¹⁴ Spatial Data Science with Applications in R (Știința datelor spațiale cu aplicații în R), <https://keen-swartz-3146c4.netlify.app/>

spățiale respective. De asemenea, un format de acest gen al datelor facilitează și finalități analitice, concepte de eșantioane pentru anchete și predicții spațiale.

Întrucât una dintre verificările relevante în materie de limite de separație este cea a numărului de unități pe celulă de grid, aplicația afișează un tabel cu coduri ale celulelor de grid, alături de numărul acestora, număr care poate fi ordonat (respectiv, de la cel mai mic la cel mai mare) sau în care se poate căuta.

Figura 17: Selecție de celule de grid prin care se identifică elemente singulare sau alte celule care încalcă limitele de separație.



Valoarea poate fi apoi utilizată în selecția derulantă a celulelor de grid, în partea stângă, când este selectată o astfel de celulă de grid pentru identificare.

Figura 18: Celulă de grid selectată pentru inspectare și urmărire ulterioare.



Formate pentru descărcare

Formatele curente de descărcare se regăsesc într-un fișier .zip care conține.

- a) un fișier .csv care conține date despre unități și codurile de grid atribuite acestora (două fișiere .csv în arhiva zip)
- b) un fișier .tif în conformitate cu convenția GeoTIFF 1.1, 2019 (inclusiv CRS): fișierul tiff apare numai dacă variabilele sunt selectate pentru agregare și depinde de selecție (vezi Figura 29 și Figura 31)

<https://earthdata.nasa.gov/esdis/eso/standards-and-references/geotiff>

Mențiuni suplimentare:

- Gridul rezultat poate fi exportat și utilizat în alte scopuri analitice (ex. descărcare griduri pe județ).
- Aplicația utilizează deck.gl, un software cu sursă deschisă (detalii pe [website](https://deck.gl/)).
- Aplicația necesită fie o cheie de licență Mapbox (<https://www.mapbox.com/>), care este chiar gratuită în condițiile anumitor restricții, fie propriile dumneavoastră hărți de referință, ambele opțiuni fiind posibile. Când sunt furnizate hărți proprii de referință, acestea trebuie să provină de pe un server cartografic precum Maptiler, <https://github.com/maptiler/tiles-server-gl>

Descarcarea grilei de referință

Grila de referință poate fi descărcată direct din secțiunea DB și produce fișierul grid, livrat ca fișier de formă ESRI și așa cum este utilizat de EUROSTAT. Conține grila pentru toată țara ca un singur fișier de formă

2. Pregătirea datelor și variabilele după tipul de fermă

Organizarea bazei de date RGA

Organizarea bazei de date potrivit chestionarului implementat în Survey Solutions, care include informațiile necesare în pregătirea georeferențierii, constă din:

1. „Identification Data” (*Date de identificare*), unde se găsesc variabilele chestionarului SuSo aferente datelor de identificare ale exploatației agricole;
2. „Areas, Livestock, Animals housing, and other” (*Suprafețe, inventar viu, ferme/adăposturi de animale și altele*), unde sunt conținute variabilele chestionarului SuSo referitor la suprafețele, animalele și alte date ale exploatației agricole;
3. „Labor Force” (*Forța de muncă*), unde sunt conținute variabilele chestionarului SuSo referitor la forța de muncă a exploatației agricole;
4. „Sales” (*Vânzări*), unde se găsesc variabilele chestionarului SuSo referitor la vânzarea producției proprii a exploatației agricole.

Datele colectate din RGA au fost exportate din Survey Solutions și apoi transformate în microdate ale RGA, în conformitate cu descrierea elaborată împreună cu Echipa INS. Tabelul cu corespondențe între variabile este detaliat în **Anexa 3**.

Microdatele sunt structurate în 5 secțiuni, fiecare divizată în 42 de fișiere, câte unul pentru fiecare județ din România (*nn* - coduri ale județelor), după cum urmează:

- Fișiere *IDnn* - conțin variabilele aferente informațiilor generale privind exploatația agricolă;
- Fișiere *ALTnn* - conțin variabilele referitoare la adăposturi pentru animale; managementul îngrășămintelor naturale și altele; agricultură ecologică; activități lucrative și altele; dezvoltare rurală și altele;
- Fișiere *FMnn* - conțin variabilele pentru forța de muncă;;
- Fișiere *SAnn* - conțin variabilele pentru destinația terenurilor; inventarul viu;
- Fișiere *VZnn* - conțin variabilele pentru vânzări.

Variabilele cheie care leagă aceste 5 tipuri de fișiere, care corespund celor 5 secțiuni descrise, sunt:

- *i_key/interview_key*
- Numărul folderului (*C1P1a/ HLD_ID*), combinat cu numărul chestionarului din acel folder, *C1P2a/ HLD_ID2*, (*Hartă + Formular*). Primele 2 cifre ale hărții reprezintă codul SIRUTA al județului, chiar dacă ar putea exista un raport de redundanță cu informațiile *Sirjud/C01P01_3*.

Pregătirea datelor pentru georeferențierea exploatațiilor agricole pe grid constă în combinarea informațiilor din 4 surse de date, în funcție de acuratețea datelor GPS, după cum urmează:

- microdate din fișiere *IDnn* (3225908 de înregistrări), ce reprezintă datele colectate în cadrul RGA, organizate ca microdate;
- fișiere *APIA* cu forme (*ipa_2020_changed_centroids.zip* - 6046848 de înregistrări), ce reprezintă o sursă de date administrative, furnizate de APIA, care conțin localizarea GPS a fiecărei parcele a exploatației agricole determinat cu mare acuratețe (punctul central al parcelelor);

- zone APIA (APIA_2020_suprafete_sept_2021.xlsx - 1001440 de înregistrări), ce reprezintă o sursă de date administrative, furnizate de APIA, folosite în identificarea Codului APIA pentru persoanele care nu au un Cod APIA în datele colectate RGA, însă au cod ISU, C1P311 (C01P04_1d) sau C1P321 (C01P04_2b);
- puncte centrale ale localităților ("Sate_centroide_latlon.csv") (12959 de înregistrări), reprezentând baza de date cu localizarea GPS a punctelor centrale pentru fiecare localitate din România, utilizată când sunt înlocuite valorile NA aferente localizării GPS pentru exploatațiile agricole, colectate cu erori sau necollectate și care nu au putut fi corectate cu date din alte surse.

Valorile colectate pentru variabilele C1P2B_1 (GPS_Latitude) și C1P2B_2 (GPS_Longitude) sunt verificate în raport cu limitele geografice ale României, Nord, Sud, Est și West (N = 48.265083, S = 43.621100, E = 29.7, V = 20.261759). Valorile non-valide au fost modificate în NA.

Datele C1P2B_1 (GPS_Latitude) și C1P2B_2 (GPS_Longitude) au fost corectate pe baza datelor APIA, întrucât coordonatele exploatației trebuie să corespundă cu locația exploatației, nu cu locul unde s-a făcut interviul (fișiere APIA cu forme), cu respectarea măsurilor de mai jos, în ordinea prezentată:

- pentru toate codurile APIA (FARM_ID) cu corelare „unul la unul” cu datele colectate (C1P1B), coordonatele colectate au fost înlocuite cu coordonatele APIA care corespund corect cu locația exploatațiilor (38236 înregistrări);
- pentru toate codurile APIA (FARM_ID) cu corelare „unul la mai multe”, unul din datele colectate (C1P1B), multe în datele APIA, coordonatele colectate au fost înlocuite cu coordonatele APIA (627520 de înregistrări), pe baza exploatației agricole cu cea mai mare suprafață de teren, care de asemenea corecte corespund locației exploatației *max(area_decla)*;
- pentru toate codurile APIA (FARM_ID) cu corelare „multe la multe”, multe din datele colectate (C1P1B), multe în APIA, coordonatele colectate au fost înlocuite cu coordonatele APIA (67612 de înregistrări), realizându-se corelare cu toate înregistrările unde SIRCOM == sirsup_cod, pe cea mai întinsă suprafață (*max(area_decla)*);
- pentru toate codurile APIA (FARM_ID) cu corelare „multe la multe”, multe în datele colectate (C1P1B), multe în APIA, coordonatele colectate au fost înlocuite cu coordonatele APIA (129 de înregistrări), fiind identificat codul APIA prin codul ISU din fișierul "APIA_2020_suprafete_sept_2021.csv" (cheie=CNP_CUI APIA CODE = IDF; cheie=C1P311 APIA CODE = C1P1B);
- înlocuiți NA (deoarece coordonatele lipseau din diferite motive) din C1P2B_1 și C1P2B_2 cu punctele centrale ale localității din fișierul "Sate_centroide_latlon.csv" file, prin cheia=SIRUTA_SUP (C1P2B_2=xcoord, C1P2B_1=ycoord). (18188 de înregistrări).

Variabilele bazei de date pentru georeferințiere exploatații agricole sunt prezentate în **Anexa 4**. Scriptul R pentru corelarea datelor și rectificarea coordonatelor GPS este prezentat în **Anexa 5**.

Actualizarea bazei de date

Scopul actualizării bazei de date este de a sprijini modificări ulterioare ale acestora în ceea ce privește conținutul sau variabilele obținute din alte anchete (structurale) sau pentru analiza suplimentară a datelor.

Deoarece datele sunt disponibile la INS în fișierele de export FoxPro, datele sunt preprocesate pentru a îndeplini standardele de date contemporane. Acest lucru se realizează prin scriptul prezentat și documentat în **Anexa 6**.

Figura 19 - Comenzi pentru actualizarea bazei de date

App DB

Update Database

Update Database with new data

Updating the database will run a script in the background, which reads the data from the provided FoxPro database and stores it in a PostgreSQL database. This is only required, if the FoxPro database export has been updated. For more details see the corresponding chapter in the report

Download Grid

Download Reference Grid

This allows you to download the national reference grid, including population counts in a single file.

MapBox Key

Provide MapBox Key

Your mapbox key

The background maps require a mapbox key (<https://www.mapbox.com/>). Without the key, the map visualization won't work. However you can also process the data, without any visualization. For more details see the corresponding chapter in the report

Dacă se adaugă date noi, acestea trebuie să fie plasate în directorul original de pe server și, ulterior, trebuie activat butonul Actualizare bază de date cu date noi. Important de remarcat în acest sens este că schema bazei de date nu trebuie să se modifice. Aceasta înseamnă că numele și formatul variabilelor trebuie să fie aceleași chiar și în cazul oricăror date noi.

3. Instalarea și configurarea aplicației

Descrierea codului sursă

Codul sursă conține configurația standard a aplicației R Shiny, constând din:

- fișier **app.R** care conține atât UI cât și fișierul Server,
- **helpers**: un director de funcții helper
- **date**: un director de fișiere de date statice
- **www**: un director de resurse web, cum ar fi logo-ul și foile de stil CSS.

Cerințe de sistem (dependențe hardware și software)

Configurația standard, așa cum este recomandată în Rezultatul 4.1b: „Raport privind serviciile de consiliere furnizate Beneficiarului cu privire la Recomandările pentru elaborarea documentației tehnice în vederea organizării achiziției unui sistem IT integrat, inclusiv hardware (inclusiv tablete) și software (licențe), inclusiv stocarea datelor, protecția și securitatea datelor pentru desfășurarea activităților desfășurate în cadrul RAS (GIS, PHC2021, GAC2020, perioade inter-recensământ, SICCA)” este suficientă. Nu sunt necesare cerințe hardware suplimentare.

Pentru a rula orice aplicație, este necesară o instalare R validă, așa cum este descris aici: <https://www.digitalocean.com/community/tutorials/how-to-install-r-on-ubuntu-18-04>

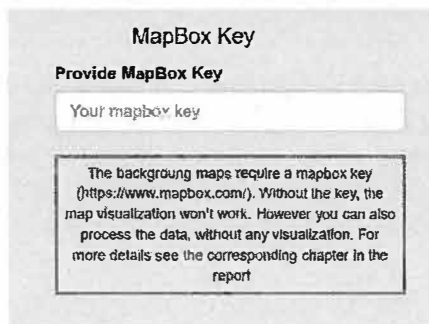
Precum și o instalare validă cu surse deschise Shiny, așa cum este descris aici: <https://www.rstudio.com/products/shiny/download-server/ubuntu/>

Pentru toate, configurația implicită este suficientă.

Kit de instalare, procedura de instalare și parametrii de configurare

Întrucât componentele hardware și software ale sistemului integrat INS sunt disponibile, Banca Mondială a efectuat instalarea și configurarea tuturor componentelor software (programe, licențe, scripturi etc.) necesare pentru buna funcționare a aplicației și acestea sunt disponibile pe INS. „server de analiză” - link xx.x.xxx.xx/rga-grid/.

Pentru hărți este necesar un abonament gratuit Mapbox. Un abonament valid poate fi obținut aici: <https://www.mapbox.com/>. După obținerea cheii Mapbox, aceasta va fi inclusă în spațiul indicat în aplicație (vezi mai jos). Cu toate acestea, aplicația funcționează și fără o cheie mapbox. În acest caz, aplicația permite producerea diferitelor grile, dar nu va afișa nicio vizualizare.



4. Recomandări și bune practici

Recomandările prezentate mai jos sunt în conformitate și încorporează cele mai bune practici pentru dezvoltarea, întreținerea și exploatarea aplicațiilor care gestionează o cantitate mare de date într-un proces de georeferențiere. Recomandările se concentrează pe date, pe aspecte organizaționale și legate de capabilități. Scopul aplicării acestora este utilizarea optimă a aplicației cu datele colectate GAC2020. Cu toate acestea, recomandările ar trebui luate în considerare și pentru viitor, atunci când se utilizează aplicația în alt context.

- Atribuirea coordonatelor GPS corecte face parte din colectarea datelor (sau din exercițiul de creare a cadrului/structurii datelor anterior recensământului) și organizarea acestora, pentru care s-au formulat, de asemenea, unele recomandări, cum ar fi, înregistrarea coordonatelor doar din amplasamentul exploatației, preîncărcarea coordonatelor APIA etc. Aplicația gestionează datele începând cu această etapă.
- Este necesară furnizarea datelor în aceeași manieră precum cea evidențiată de scriptul R din Anexa 5 în legătură cu toate etapele transformării acestora.
- Crearea cadrului structurii datelor se va realiza cu ajutorul unor soluții software dedicate întrucât acest lucru va facilita procesarea ulterioară. MS Excel și soluții similare nu sunt recomandate pentru o sarcină atât de importantă. Și sistemul Survey Solutions poate fi luat în calcul pentru astfel de sarcini interne.
- O cooperare mai strânsă cu producătorii de date administrative este utilă pentru a permite accesul la datele lor administrative (vezi lipsa surselor de date administrative de la ANCPI - Agenția de Cadastru și Publicitate Imobiliară) și o mai bună cunoaștere a calității datelor lor (vezi APIA ca sursă a datelor pentru georeferința exploatațiilor agricole).
- O infrastructură armonizată de date (spațiale) va aborda mai multe probleme întâlnite pe durata postprocesării. Aceasta necesită și o infrastructură IT integrată.
- Organizarea în unități de tip grid se poate realiza ulterior cu orice coordonate GPS, fără a fi necesare coordonate specifice, ca element de reținut în acest context. În funcție de mediul local de date, recomandăm stocarea localizării în proiecția locală geospațială (CRS) și transformarea sa în coordonate sferice (zecimale) doar când este cazul. Se facilitează în acest fel stocarea (zecimală vs. număr întreg), precum și eventuale calcule pe bază de spațiu, precum distanța.
- Se poate verifica dacă respectivele coordonate colectate se află în cadrul poligonului localității și apoi ajustate în mod corespunzător. O procedură de acest gen poate fi automatizată în activități viitoare.
- Dacă sunt preîncărcate coordonatele (exploatației agricole), Survey Solutions permite georeferențierea chestionarului. În acest fel, INS se poate asigura că coordonatele de deținere sunt cel puțin în poligonul localității.
- În cel mai recent set de date au fost 72.705 înregistrări fără cheie de interviu și rânduri complet goale. așa cum este imputat din sursa administrativă de date APIA. Deoarece nu au fost colectate, cheia interviului nu este disponibilă. Pentru a putea folosi în continuare aceste înregistrări în aplicații, a fost generată o cheie de interviu sintetică.
- Lipsa de răspuns pentru o variabilă (item) ar trebui să urmeze proceduri de imputare clare și transparente. Se recomandă o metodologie și o documentație aplicabilă pentru procedurile de

imputare și ar trebui să existe la începutul recensământului sau anchetei pentru producerea de statistici.

- Geolocalizările în timpul culegerii datelor în teren trebuie efectuate la locul exploatației agricole, și nu la locul în care se desfășoară interviul. Imputarea locației ar trebui, de asemenea, să respecte reguli clare și transparente.
- Pentru a avea consistență în variabila I_KEY, ca bună practică și recomandare pentru lucrări viitoare, se generează o cheie de interviu pentru toate înregistrările fără cheie (NA). Orice corecție trebuie să aibă loc asupra datelor sursă. Practicile de curățare a datelor implementate ar trebui să evite adăugarea ID_variables atunci când acestea lipsesc.

O recomandare se referă la abilitățile și capacitățile specialiștilor care lucrează cu aplicația în vederea dezvoltărilor viitoare, pe măsură ce nevoile evoluează, și pregătirea unei baze de date adecvate. În acest sens, este foarte importantă integrarea cunoștințelor permanente despre limbajele de programare R și R Studio, Shiny și GIS în cadrul echipei/direcției INS care gestionează statistica agriculturii, ca sarcină și practică permanentă. Acest lucru ar putea fi implementat și pentru alte departamente de statistică ale INS, deoarece aceasta este o practică aplicată în instituțiile similare la nivelul UE și a fost observată în timpul schimburilor de experiență recente (în Polonia și în Țările de Jos).

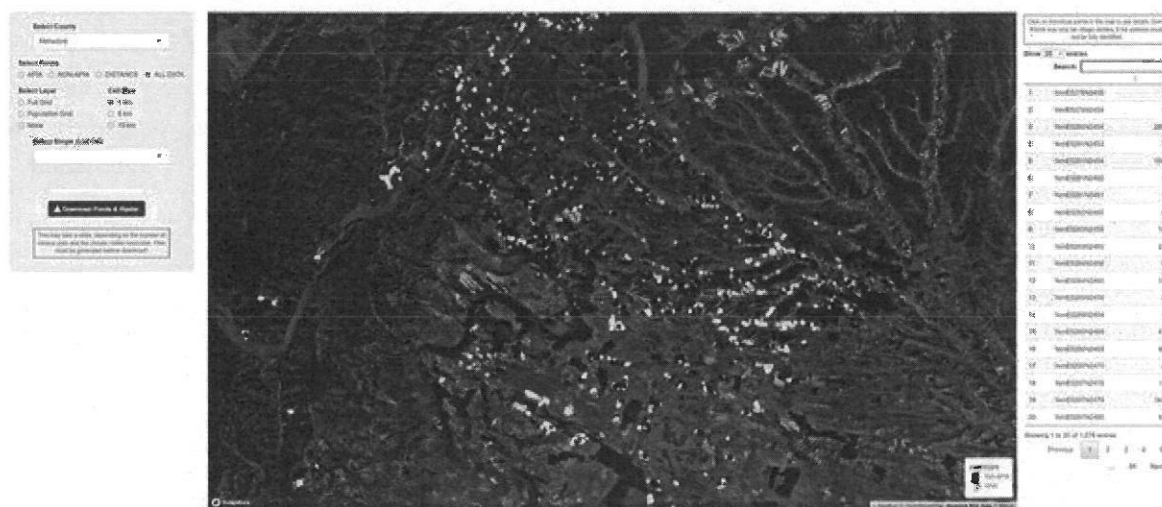
Dobândirea de capacități și abilități ar putea fi realizată prin traininguri și practică la locul de muncă (lista cursurilor de formare este furnizată în „*Elaborarea unui plan multianual de nevoi de formare (2021-2024) pentru INS, inclusiv BTS, pe baza nevoilor de formare*” din Rezultatul 12 în cadrul aceluiași Acord RAS) fie prin angajarea de personal certificat cu experiență în statistică și IT. Această abordare ar putea evita lipsa de specialiști sau volumul mare de muncă cu care se confruntă INS în cadrul departamentului IT.

5. Anexe

Anexa 1: Filtrarea observațiilor prin selecție din tabelul centralizator cu celulele de grid

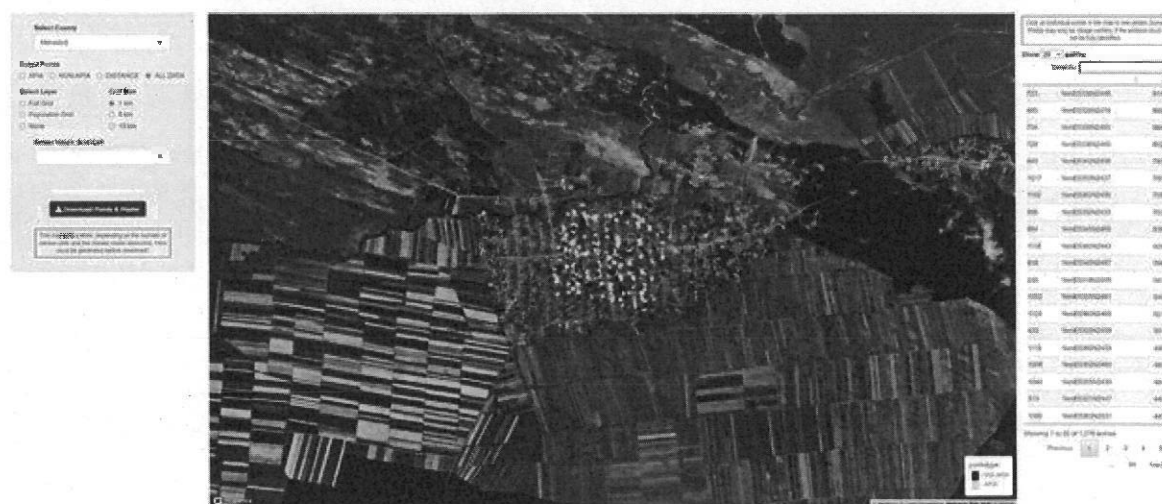
Select Single Grid Cell (*Selectați o singură celulă de grid*) permite deja utilizatorului să selecteze celule individuale de grid după codul celei. Pentru a permite de asemenea o inspecție mai detaliată a totalurilor de unități per celulă de grid și a identifica elemente singulare, dar și alte numere de unități per celulă, aplicația permite utilizatorului și să selecteze orice rând din tabelul din dreapta, care arată numărul de celule de grid și este creat după ce este selectată opțiunea Select Points (*Selectați puncte*).

Figura 20: Tabel cu numărul de celule de grid în dreapta hărții



Tabelul cu numărul de celule poate fi sortat după cel mai mare număr de unități per celulă ...

Figura 21: Sortare după cea mai mare valoare



Sau după cel mai mic număr de unități per celulă.

Figura 22: Sortare după cea mai mică valoare



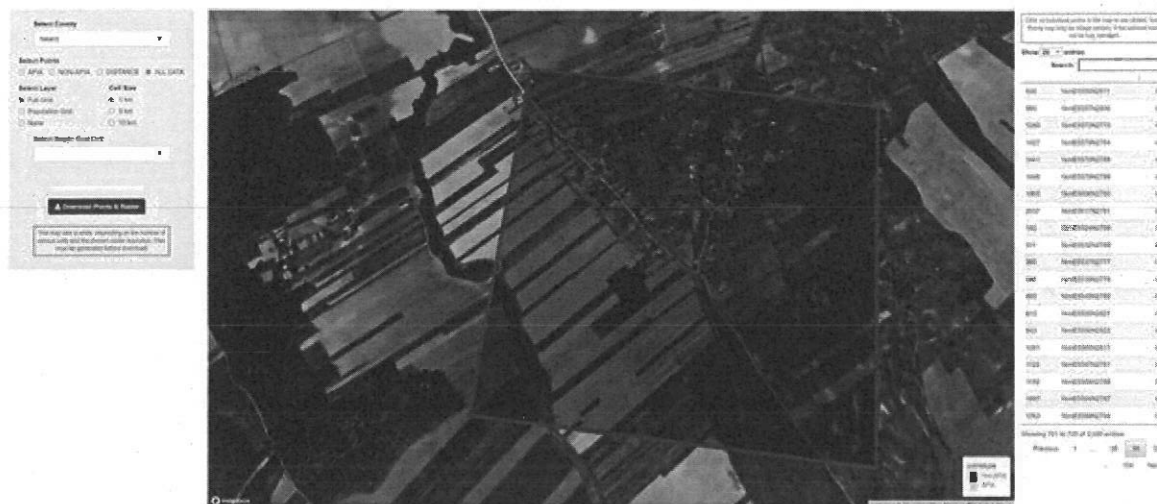
De asemenea, poate fi adăugată celula corespunzătoare de grid.

Figura 23: Celulă de grid selectată



În plus, se poate selecta oricare celulă de pe oricare pagină a tabelului.

Figura 24: Selectare celulă de dimensiuni medii



Chiar și setul de date în puncte poate fi mutat pe diferite straturi Select Points (*Selectați puncte*) care ilustrează date APIA ...

Figura 25: Unități Apia din cadrul celei de grid



Sau unități Non-APIA pentru aceeași celulă de grid.

Figura 26: Unități Non-APIA din cadrul celei de grid



De asemenea, și cu siguranță, este posibilă selectarea din diferite griduri de referință, fie cu celule de 1 km, după cum se poate vedea mai sus în, fie cu celule de 5 km, după cum se ilustrează mai jos,

Figura 27: Selectare celulă de grid cu latura de 5 km



precum și cu celule cu latura de 10 km.

Figura 28: Selectare celulă de grid cu latura de 10 km



În acest mod este posibilă o monitorizare (spațială) mai detaliată **pe parcursul și ulterior** colectării datelor. Toate resursele create, adică celule de grid, pot fi exportate, dacă se dorește acest lucru, și transmise către CASS Survey Solutions spre utilizare în alte activități de colectare a datelor¹⁵. Acest lucru ar putea fi necesar, de exemplu, în cazul anchetelor post-recenzare sau al altor activități similare. Unele procese pot fi chiar automatizate, însă necesită o definire clară a sarcinilor de automatizare. Mai multe detalii pe această temă sunt disponibile în secțiunea finală a metodologiei PHC21.

¹⁵ Momentan, exportăm doar fișierele descrise în manualul principal, însă poate fi exportat de asemenea orice rezultat produs intern.

Anexa 2: Descărcarea datelor

În mod obișnuit, pasul final în cadrul unei transformări spațiale, după cum este pus la dispoziție în această de aplicație, este de a transmite datele către următoarea etapă de procesare. Într-un mediu integrat de procesare a datelor, rezultatul acestei etape de procesare ar putea fi sub formă de:

- stocare într-o bază de date sau
- procesare suplimentară pentru
 - verificări de date,
 - verificări de acoperire,
 - alte scopuri de monitorizare.

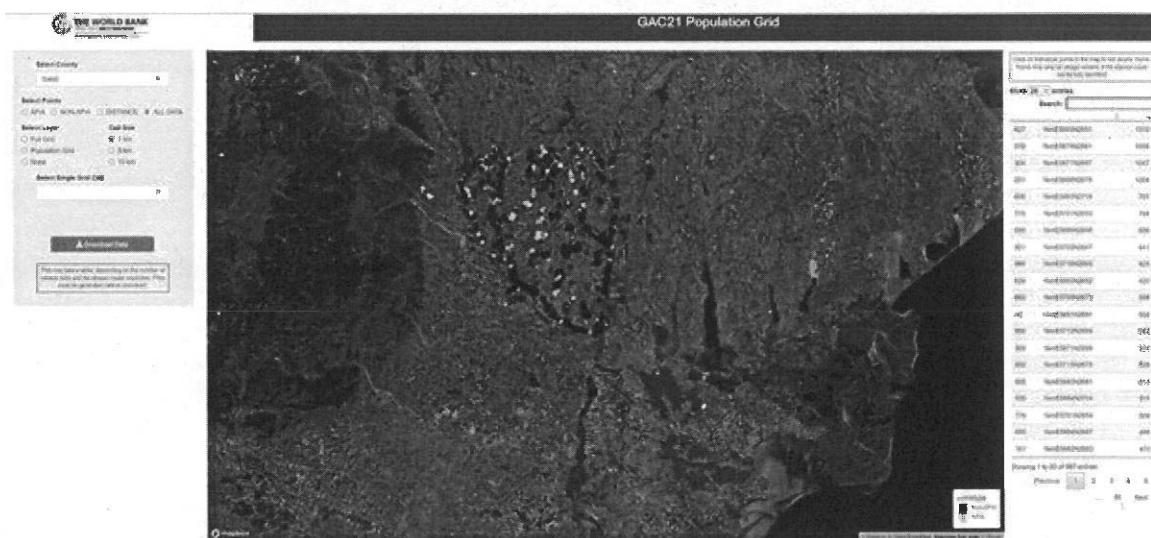
În cazul în care este inclusă, în oricare dintre etapele ulterioare de procesare, **colectarea datelor primare** (respectiv, re-interogarea datelor CATI pentru a verifica unele răspunsuri), prin internet, telefonic sau în persoană, pot fi exportate fișiere necesare de suport, precum date de identificare sau chiar limite de celulă, posibile chiar și la rezoluții mici de până la 100 m¹⁶.

În funcție de zona selectată, rezoluție și datele de intrare, crearea fișierului și procesul de compresie ar putea fi relativ îndelungate.

Datele în puncte

Primul set de date disponibil spre descărcare, după selectarea județului în câmpul Select County (*Selectați județul*), este cel al datelor în puncte.

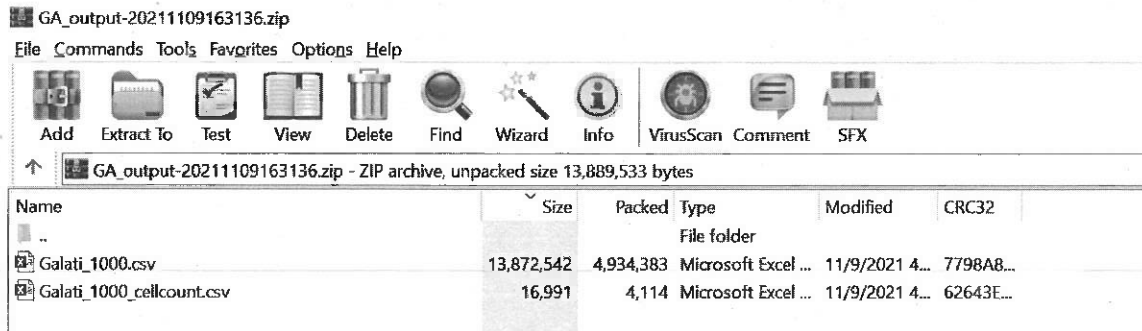
Figura 29: Încărcarea datelor pentru un anumit județ creează fișierul cu puncte și fișierele cu numărul de celule pentru descărcare.



Deschiderea fișierului .zip oferă următoarea vizualizare a directorului:

¹⁶ Dimensiunea de 100 m se utilizează într-o serie de aplicații de învățare automatizată axate pe date privind populația din grid. Astfel, ar putea fi posibilă folosirea acestor seturi de date disponibile public în corelare sau, de asemenea, împreună cu alte date raster privind populația.

Figura 30: Fișier .zip cu date în puncte și numărul de celule.



Datele privind numărul de celule

Acest fișier conține doar aceleași date precum cele afișate în partea din dreapta a aplicației, mai exact numărul unităților de recensământ, pe celule de grid, **strict pentru rezoluția selectată**.

Dacă este necesar să descărcați datele, precum și codurile în diferite rezoluții, veți efectua o descărcare separată de fiecare dată când schimbați oricare dintre rezoluții. Datele în puncte și datele cu numărul de celule sunt întotdeauna incluse după ce au fost selectate datele pentru un anumit județ.

Fișierul raster

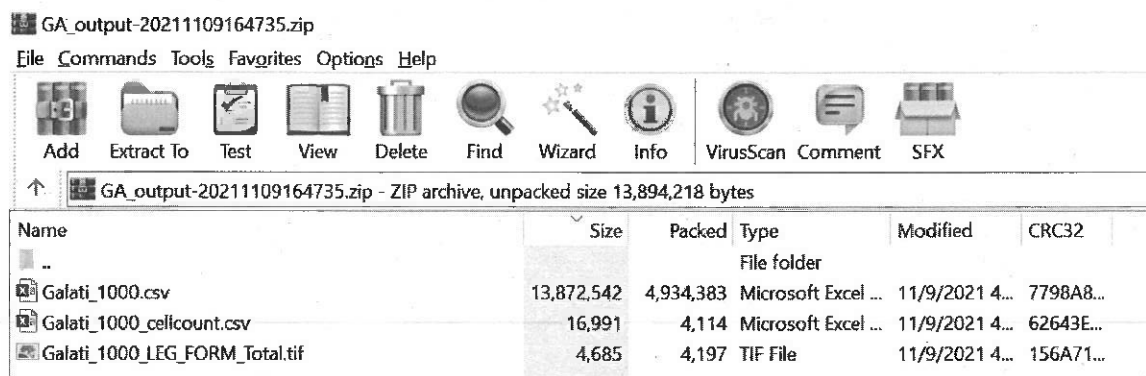
Fișierul raster este inclus doar dacă, din opțiunile disponibile în câmpul Select Layer (*Selectați strat*)) este selectată opțiunea **Population Grid (Grid populație)**.

Figura 31: Datele raster privind populația se pot descărca după ce devin vizibile pe hartă.



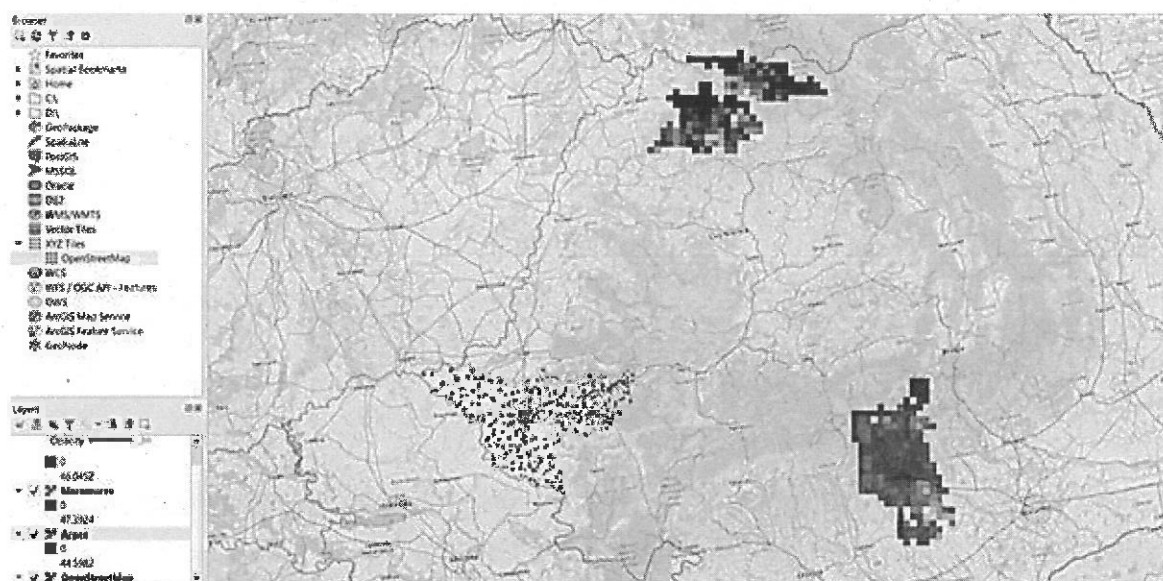
Stratul inclus va conține Parametrul [var] selectat, agregat în funcție de tipul [tip] specific de agregare și la rezoluția [res] dorită.

Figura 32: Fișier .zip cu date în puncte, număr de celule și date de tip raster.



După descărcarea și dezarhivarea fișierului, puteți afișa rasterul și în aplicația informatică GIS preferată. Pentru fiecare județ, puteți descărca fișierul raster la diferite rezoluții, însă toate compatibile cu structura de codificare INSPIRE.

Figura 33: Afișare QGIS a datelor raster pentru județe diferite și cu rezoluții diferite, toate pe o singură hartă.



Dacă este necesar să descărcați datele și codurile în diferite rezoluții, trebuie să inițiați o descărcare separată de fiecare dată când modificați oricare dintre parametrii de intrare, cum ar fi rezoluția, datele (variabilele) selectate sau tipul selectat de agregare.

Anexa 3: Microdate RGA

Fișiere IDnn - conțin variabilele aferente informațiilor generale privind exploatarea agricolă.

VARIABILĂ	COD	Valoare sau interval așteptat (ă)	Variabilă aplicată	Tip	Valoare / exemple colectată(e)	Categorii/ Nomenclaturi/ Structura clasificării	SuSo	Nu
INFORMAȚII GENERALE PRIVIND EXPLOATAȚIA AGRICOLĂ								
LOCALIZAREA EXPLOATAȚIEI AGRICOLE								
Cheia interviului/chestionarului	i_key	text	interview_key	c	94-95-87-87		GAC20_Romani a9.tab	
Numărul mapei	C1P1a	valoare numerică	HLD_ID	i	280098		GAC20_Romani a9.tab	
Cod identificare APIA	C1P1b	text sau blank	C01P01_1b	c	999		GAC20_Romani a9.tab	
Numărul chestionarului în cadrul mapei	C1P2a	valoare numerică	HLD_ID2	i	163		GAC20_Romani a9.tab	
GPS_Lat	C1P2b_1	coord.GPS sau blank	visitLocation__Latitude	d	47.78160932		visit_summary.tab	
GPS_Long	C1P2b_2	coord.GPS sau blank	visitLocation__Longitude	d	24.76437437		visit_summary.tab	
Județul	C1P3	text						
Comuna/ Orașul/ Municipiul	C1P4	text						
Localitatea componentă (satul)	C1P5a	text						
Cod SIRUTA județ	Sirjud	cod Siruta	C01P01_3	i	38/Cod Siruta Județ	Options-in-question-1.3. County.txt	GAC20_Romani a9.tab	
Cod SIRUTA comuna	Sircom	cod Siruta	C01P01_4	i	14450/ Cod Siruta Oraș	Options-in-question-1.4. Municipality ~ Commune_Town.txt	GAC20_Romani a9.tab	
Cod SIRUTA sat	Sirsat	cod Siruta	C01P01_5a	i	14478/Cod Siruta Sat	Options-in-question-1.5a. Component locality (village).txt	GAC20_Romani a9.tab	

Strada	C1P6a	text sau blank	C01P01_6a	c	Bercheza/999		GAC20_Romani a9.tab	
Nr.	C1P6b	text sau blank	C01P01_6b	c	43/999		GAC20_Romani a9.tab	
Bloc	C1P6c	text sau blank	C01P01_6c	c	A3/999		GAC20_Romani a9.tab	
Scara	C1P6d	text sau blank	C01P01_6d	c	B/999		GAC20_Romani a9.tab	
Etaj	C1P6e	text sau blank	C01P01_6e	c	2/999		GAC20_Romani a9.tab	
Apart.	C1P6f	text sau blank	C01P01_6f	c	14/999		GAC20_Romani a9.tab	
Sector	C1P6g	text sau blank	C01P01_6g	c	1/999		GAC20_Romani a9.tab	
Telefon	C1P6h	text sau blank	C01P01_6h	c	0211234125/999		GAC20_Romani a9.tab	
Fax	C1P6i	text sau blank	C01P01_6j	c	0211234125/999		GAC20_Romani a9.tab	
Adresa e-mail	C1P6j	text sau blank	C01P01_6i	c	sample@site.com/999		GAC20_Romani a9.tab	
STATUT JURIDIC AL EXPLOATAȚIEI AGRICOLE								
Exploatații agricole fără personalitate juridică								
Exploatație agricolă individuală	C1P2s	valoare numerică	LEG_FORM	i	1		GAC20_Romani a9.tab	1
Persoană fizică autorizată, întreprindere individuală	C1P2s	valoare numerică	LEG_FORM	i	2		GAC20_Romani a9.tab	2
Întreprindere familială	C1P2s	valoare numerică	LEG_FORM	i	3		GAC20_Romani a9.tab	3
Exploatații agricole cu personalitate juridică								
Regie autonomă	C1P2s	valoare numerică	LEG_FORM	i	4		GAC20_Romani a9.tab	4
Societate/ asociație agricolă (Legea nr. 36/ 1991)	C1P2s	valoare numerică	LEG_FORM	i	5		GAC20_Romani a9.tab	5
Societate comercială cu capital majoritar privat (Legea nr. 31/ 1990)	C1P2s	valoare numerică	LEG_FORM	i	6		GAC20_Romani a9.tab	6
Societate comercială cu capital majoritar de stat (Legea nr. 31/ 1990)	C1P2s	valoare numerică	LEG_FORM	i	7		GAC20_Romani a9.tab	7
Institut/ stațiune de cercetare, unitate școlară cu profil agricol	C1P2s	valoare numerică	LEG_FORM	i	8		GAC20_Romani a9.tab	8

Primărie	C1P2s	valoare numerică	LEG_FORM	i	9		GAC20_Romani a9.tab	9
Alte instituții publice	C1P2s	valoare numerică	LEG_FORM	i	10		GAC20_Romani a9.tab	10
Unitate cooperatistă	C1P2s	valoare numerică	LEG_FORM	i	11		GAC20_Romani a9.tab	11
Alte tipuri (fundăție, așezământ religios, școală etc.)	C1P2s	valoare numerică	LEG_FORM	i	12		GAC20_Romani a9.tab	12
Exploatația agricolă cu personalitate juridică face parte dintr-un grup de exploatații agricole, altul decât cele menționate	C1P23	1 sau 2	C01P03b	i	1(Da) / 2 (Nu)		GAC20_Romani a9.tab	
DATE DE IDENTIFICARE								
Exploatații agricole fără personalitate juridică								
Numele, inițiala tatălui și prenumele capului exploatației agricole (utilizatorul)	C1P31 a	text	C01P04_1a	c	text		GAC20_Romani a9.tab	
Cod numeric personal (CNP) ISU	C1P31 1	text sau blank	C01P04_1d	c	13 cifre / 999		GAC20_Romani a9.tab	
Cod unic de identificare (CUI)/ Cod fiscal	C1P31 2	text sau blank	C01P04_1e	c	n cifre / 999		GAC20_Romani a9.tab	
Exploatații agricole cu personalitate juridică								
Denumirea exploatației agricole cu personalitate juridică	C1P32	text sau blank	C01P04_2a	c	text		GAC20_Romani a9.tab	
Cod unic de identificare (CUI)/ Cod fiscal	C1P32 1	text sau blank	C01P04_2b	c	text		GAC20_Romani a9.tab	
Adresa capului exploatației agricole fără personalitate juridică (utilizatorul) sau a sediului social al exploatației agricole cu personalitate juridică								

Localitatea componentă (satul) pentru capul exploatației agricole fără personalitate juridică	C1P33 a	text sau blank	C01P04_3a	c	text		GAC20_Romani a9.tab	
Strada	C1P33 c	text sau blank	C01P04_3b	c	text		GAC20_Romani a9.tab	
Nr.	C1P33 d	text sau blank	C01P04_3c	c	text		GAC20_Romani a9.tab	
Bloc	C1P33 e	text sau blank	C01P04_3d	c	text		GAC20_Romani a9.tab	
Scara	C1P33 f	text sau blank	C01P04_3e	c	text		GAC20_Romani a9.tab	
Etaaj	C1P33 g	text sau blank	C01P04_3f	c	text		GAC20_Romani a9.tab	
Apart.	C1P33 h	text sau blank	C01P04_3g	c	text		GAC20_Romani a9.tab	
Sector	C1P33 i	text sau blank	C01P04_3h	c	text		GAC20_Romani a9.tab	
Telefon	C1P33 j	text sau blank	C01P04_3i	c	text		GAC20_Romani a9.tab	
Fax	C1P33 k	text sau blank	C01P04_3k	c	text		GAC20_Romani a9.tab	
Adresa e-mail	C1P33 l	text sau blank	C01P04_3j	c	text		GAC20_Romani a9.tab	
COD DE COMPLETITUDINE								
Interviu complet	C13P4 5	valoare numeric a sau zero	C13P45	i	1-complet / 2-partial		visit_summary.tab	
Exploatație agricolă desființată	C01P0 2b	valoare numeric a sau zero	C01P02b	i	2		visit_summary.tab	1
Exploatație agricolă temporar fără activitate	C01P0 2b	valoare numeric a sau zero	C01P02b	i	3		visit_summary.tab	2
Interviu refuzat	C01P0 2b	valoare numeric a sau zero	C01P02b	i	1		visit_summary.tab	3
Alte situații (exploatație agricolă neidentificată, necontactată etc.)	C01P0 2b	valoare numeric a sau zero	C01P02b	i	4		visit_summary.tab	4

Anexa 4: Variabile în georeferințele exploatațiilor agricole

Microdate (fișiere IDnn)

Denumirile variabilelor *visitLocation* *Latitude* și *visitLocation* *Longitude*, din chestionarul Survey Solutions, în cadrul microdatelor RGA, sunt *C1P2B_1* pentru GPS *Latitude* și *C1P2B_2* pentru GPS *Longitude*.

Denumirile variabilelor sunt: "I_KEY" "C1P1B" "C1P311" "C1P312" "C1P321" "C1P3" "C1P4" "C1P5A" "SIRJUD" "SIRCOM" "SIRSAT" "C1P6A" "C1P6B" "C1P6C" "C1P6D" "C1P6E" "C1P6F" "C1P6G" "C1P2S" "C1P2B_1" "C1P2B_2"

I_KEY	C1P1B	C1P311
00-00-14-41:	1 RO :	42 afc97ea131fd7e2695a98ef34013608f97f34e1d: 16336
00-00-22-72:	1 RO123456789:	31 b9d0f181ffb5bd226d5690dc753aa2d0fbd88174: 18
00-00-26-73:	1 ROSIOARA :	14 dbb8e594cd7045fae7b28e6a595293675765e450: 12
00-00-29-23:	1 RO539322636:	13 da39a3ee5e6b4b0d3255bfef95601890afd80709: 8
00-00-38-70:	1 RO010690458:	11 83a62f7c422b23a8676758af971278862f0ad7b9: 8
00-00-51-93:	1 (Other) :	797216 (Other) :2960549
(Other) :3225902	NA's :2428581	NA's : 248977

C1P312	C1P321	C1P3	C1P4
0 :2918207	1590120 : 38	JUDETUL ARGES : 152487	POPESTI : 6788
- : 1618	14818116: 15	JUDETUL SUCEAVA: 151165	CALINESTI : 6134
99 : 563	999 : 15	JUDETUL DOLJ : 127020	VANATORI : 5979
990 : 348	41228341: 10	JUDETUL PRAHOVA: 124649	ORAS BORSA : 5713
09 : 335	6064801 : 10	JUDETUL BACAU : 122322	MUNICIPIUL ORADEA: 5573
(Other) : 55856	(Other) : 31457	(Other) :2545053	(Other) :3184282
NA's : 248981	NA's :3194363	NA's : 3212	NA's : 11439

C1P5A	SIRJUD	SIRCOM	SIRSAT	C1P6A
POIANA : 8923	38 : 152487	106746 : 5713	26573 : 5571	0 :1619354
BORSA : 6434	332 : 151165	26564 : 5573	106755 : 5208	Principala: 100882
SLOBOZIA: 6421	163 : 127020	15108 : 4702	76709 : 3847	PRINCIPALA: 75538
POPESTI : 6168	298 : 124649	32394 : 4592	72016 : 3576	principala: 18835
LUNCA : 6096	47 : 122322	13668 : 4420	54984 : 3404	Bisericii : 8695
(Other) :3181360	154 : 117401	(Other):3197696	(Other):3201057	(Other) :1399223
NA's : 10506	(Other):2430864	NA's : 3212	NA's : 3245	NA's : 3381

C1P6B	C1P6C	C1P6D	C1P6E	C1P6F
0 : 778668	0 :3192412	0 :3199352	0 :3201717	0 :3189811
1 : 45913	- : 1974	A : 5384	1 : 4238	- : 1975
2 : 43669	1 : 1194	B : 3278	2 : 3836	2 : 1694
3 : 40263	99 : 957	1 : 2362	3 : 2960	1 : 1670
4 : 39412	2 : 934	- : 2017	- : 2043	4 : 1573
(Other):2274574	(Other): 24999	(Other): 10074	(Other): 7668	(Other): 25740
NA's : 3409	NA's : 3438	NA's : 3441	NA's : 3446	NA's : 3445

C1P6G	C1P2S	C1P2B_1	C1P2B_2
0 :3207968	Min. : 0.000	Min. : -29.80	Min. : -115.26
- : 2028	1st Qu.: 1.000	1st Qu.: 44.90	1st Qu.: 23.66
1 : 796	Median : 1.000	Median : 45.86	Median : 25.13
990 : 698	Mean : 1.014	Mean : 45.60	Mean : 24.91
99 : 691	3rd Qu.: 1.000	3rd Qu.: 46.91	3rd Qu.: 26.57
(Other): 10279	Max. :12.000	Max. : 57.01	Max. : 127.41
NA's : 3448			

Fișiere APIA cu forme

Datele provin prin extragere din fișierul APIA, denumit "ipa_2020_changed_centroids.zip". Denumirile variabilelor sunt: "grid" "farm_id" "sirsup_cod" "bloc_nr" "parcel_nr" "crop_nr" "cat_use" "crop_code" "area_decla" "agro_env" "crop_code_" "area_dec_1" "m_13_anc" "long" "lat"

Suprafețe APIA

Fișierul "APIA_2020_suprafete_sept_2021.xlsx" a fost utilizat pentru identificarea Codului APIA în cazul persoanelor pentru care nu există un Cod APIA în datele colectate pentru RGA, însă care au un cod ISU *CIP311* (C01P04_1d) sau un cod fiscal *CIP321* (C01P04_2b). Denumirile variabilelor sunt: "IDF" "NUME" "CNP_CUI" "TARA" "TIP_F" "ADREXPL" "JUD_DOM" "LOCAL_DOM" "SECTOR" "SAT" "COD_POSTAL" "STRADA" "NR_STRADA" "NR_CLADIRE" "NR_SCARA" "NR_APART" "SIRUTA" "LOCAL" "SUP_UTILIZATA" "SUP_SOL"

Fișierul final generat pe bază de corelare a datelor și rectificări, care va fi utilizat în aplicații de georeferențiere cu titlu de date colectate din RGA și rectificate în baza datelor APIA, are următoarea structură:

```
interview__id
interview__key
C01P01_1b
GEO_LCT
C01P01_3
C01P01_4
C01P01_5a
C01P01_6a
C01P01_6b
C01P01_6c
C01P01_6d
C01P01_6e
C01P01_6f
C01P01_6g
LEG_FORM
visitLocation__Latitude
visitLocation__Longitude
visitLocation__Accuracy
visitLocation__Altitude
visitLocation__Timestamp
Latitude_APIA
Longitude_APIA
flag
```



Competența face diferență!

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