

błędy skanowania i dopiero możliwe było przetwarzanie. Dzisiaj dzięki formularzowi elektronicznemu system nie wpuści błędnych danych, tylko oczekiwane, i to najczęściej z rozwijanego słownika. Dane praktycznie od razu nadają się do przetwarzania. Pierwsze wyniki będą w ciągu miesiąca od zakończenia spisu.

Prace nad spisem zaczęliście 4 lata temu.

W lutym 2007 roku prezes Józef Oleński ściągnął mnie do GUS, żeby przygotowywać spisy. Ja zaprosiłem do współpracy Agnieszkę Nowakowską i razem krok po kroku zbudowaliśmy zespół GIS-owy.

Poza państwa dwójką pracuje tu więcej osób, które przyszły z GUGiK.

W sumie pięć. Ale nasz zespół spisowy liczy 400 osób (w GUS i WUS-ach), głównie metodologów, bo geodetów mamy tylko sześcioro.

Dodatkowe wynagrodzenia dla tego 400-osobowego zespołu zostały zakwestionowane przez kontrolę kancelarii Rady Ministrów. Jak to się skończyło?

Afera była zupełnie nieuzasadniona. Spisy zawsze były regulowane specjalną ustawą, która gwarantowała dodatko-

we etaty i pieniądze. Tym razem dodatkowych etatów nie dostaliśmy, a ustawa spisowa weszła w życie, kiedy prace były już w toku. Jedynym sposobem dodatkowego wynagradzania ludzi w okresie dwóch poprzednich lat były zatem umowy zlecenia. Wszystko odbywało się zgodnie z prawem, co potwierdziła zakończona dwa tygodnie temu kontrola NIK.

Czy każdy obywatel musi zostać spisany?

Udzielenie rachmistrzowi rzetelnej informacji jest obowiązkowe pod karą grzywny do 5 tys. złotych. Mamy bardzo mało odmów, a w stosunku od opornych staramy się nie stosować sankcji, raczej wysyłamy do nich lidera, a nawet wójta czy prezydenta, i przekonujemy, co zwykle odnosi skutek. Zresztą podobnie postąpiliśmy z bezdomnymi. W piątek i sobotę [15 i 16 kwietnia – KPK] przeprowadziliśmy akcję, w której udało nam się spisać około 10 tys. bezdomnych. Nie uciekali przed nami, bo szliśmy do nich z osobami znanymi im z instytucji pomocowych albo ze straży miejskiej, której też ufają.

W przypadku bezdomnych trudno jednak mówić o adresie.

Zgadza się, ale spisując ich, zbieraliśmy współrzędne. Chodzi o to, by można było nałożyć na mapę miasta miejsca gromadzenia się bezdomnych.

Na czym polega cała filozofia spisu?

Sprowadza się ona do tego, że bierzemy kolejne rejestry, takie jak PESEL, dane z KRUS, ZUS, urzędów skarbowych (ale danych o zarobkach nie zbieramy), ARiMR, GUGiK, i składamy te warstwy informacyjne jedna na drugiej. Można to porównać do kostki złożonej z plastrów sera szwajcarskiego. Plastry, czyli nasze warstwy informacyjne, są dziurawe, bo nie wszystkie rejestry są kompletne. Ale idąc przez te warstwy, potrafimy znajdować najbardziej prawdopodobne wartości zmiennych i tworzymy dla każdego z nas tzw. master rekord, a po spisie – złoty rekord. Ankieta telefoniczna, wywiad rachmistrza i samospis internetowy to są kolejne warstwy. W tym zbiorze nie ma danych sensytywnych, bo pytamy o nie tylko w bezpośrednim wywiadzie, tylko w badaniu reprezentacyjnym i tylko za zgodą badanego.

Jakie to dane?

OKO NA SPIS

Mówi **AGNIESZKA NOWAKOWSKA**, naczelnik Wydziału Informacji Geoprzestrzennej w GUS oraz kierownik Centrum Zarządzania Spisem



FOT. JERZY PRZYWARA

KATARZYNA PAKUŁA-KWIECIŃSKA: Czy tych kilka osób siedzących tutaj przy komputerach to całe Centrum Zarządzania Spisem?

AGNIESZKA NOWAKOWSKA: Tak. Wszystkich dyspozytorów jest 16, pracujemy na zmiany od godziny 8 do 20. Główną pracę wykonują dyspozytorzy wojewódzcy, a my na dużym monitorze śledzimy postępy spisu w całej Polsce, np. w Mazowieckiem 44% rachmistrzów jest w terenie. Na sąsiednim wykresie są widoczne synchronizacje, z podziałem na prawidłowe i nielegalne nieprawidłowe.

Na bieżąco weryfikujecie także ankiety internetowe?

Ankieta, która wraca jako zamknięta, wymaga tylko akceptacji dyspozytora. Jeśli jest niezakończona, dyspozytor kieruje ją do rachmistrza albo do ankietera statystycznego. Dyspozytorzy mają dostępne w aplikacji wskaźniki wypełnienia ankiety, które ułatwiają im podjęcie decyzji.

To czym konkretnie tutaj się zajmujecie, skoro wszystko dzieje się „samo”?

Od 1 kwietnia w systemie pojawiło się 1581 zgłoszeń problemów od liderów gminnych i dyspozytorów wojewódzkich, które rozwiązujemy. To są najróżniejsze sprawy, np. rachmistrz idzie do mieszkania, w którym star-

sza osoba nie jest w stanie odpowiedzieć na pytania. I co wtedy zrobić? Albo ankieta przesłana internetem nie jest wypełniona. Czy skierować tam rachmistrza? Są też bardziej złożone problemy związane np. z oprogramowaniem.

Cała spisowa marszruta rachmistrza jest do odtworzenia?

Oczywiście. Na bieżąco śledzimy synchronizacje, czyli połączenia z bazą. Większość rachmistrzów zaczyna spisywanie dopiero po godz. 15. Nie wymagamy od nich pełnej dyspozycyjności. Sami regulują sobie godziny pracy.

Kim są rachmistrze na co dzień?

Sporo jest studentów, bezrobotnych, pracowników gmin. Chętni musieli przejść przez sito naszych szkoleń i sprawdzianów. Często zgłaszali się ci, którzy pracowali przy spisie rolnym. Naboru dokonywały gminy, my włączaliśmy się na etapie szkoleń.

Ile osób może spisać jeden rachmistrz?

Obecnie przeprowadzane jest badanie reprezentacyjne – na rachmistrza przypada 100-150 wylosowanych mieszkań. W każdym mieszkaniu jest inna liczba osób, które podlegają spisowi. Wynagrodzenie zależy od liczby spisanych osób.

Liderzy gminni są pracownikami statystyki?

Niepełnosprawność, dzietność kobiet, wyznaczenie i związki partnerskie. Nie ma obawy, że zaczniemy te dane profilować i wykorzystywać w sposób niedozwolony, możemy je przedstawić dopiero po zregulowaniu. Istotą spisu jest możliwość przeprowadzenia później analiz. Jeśli wiemy, że w danej dzielnicy zamieszkuje populacja składająca się z określonej liczby kobiet i mężczyzn, w wieku produkcyjnym lub nie, pracujących lub bezrobotnych, to możemy wywnioskować, czy warto tu zbudować fabrykę, linię tramwajową, czy będzie potrzebne przedszkole, czy raczej dom starców. Tego oczekują społeczności lokalne. I choć większość tych danych mają w swoich rejestrach, to dopiero jak je poskładamy, można prowadzić analizy.

Wy, statystycy, bardzo dużo o nas wiecie.

Dużo i jesteśmy jedyną instytucją, która ma do tego prawo. Jednak zakres informacji, które możemy zbierać, jest ściśle ograniczony przepisami. Ponadto obowiązują nas określone zasady, bo generalnie istnieje zakaz profilowania osób, czyli łączenia danych z różnych zakresów. Nawiasem mówiąc, mamy właśnie

kontrolę GIODO, która sprawdza wszystkie procedury. Dane identyfikowalne trzymamy tylko tak długo, jak to jest niezbędne do złożenia danych z różnych rejestrów, a korzystamy z 28 źródeł administracyjnych i pozaadministracyjnych. Najczęściej identyfikacja odbywa się po numerze PESEL, ale jak go nie ma albo jest błędny, to wykorzystujemy imię, nazwisko i adres. Po złożeniu wszystkich danych część identyfikacyjną odrzucamy. Nazywa się to anonimizacją albo depersonalizacją. Przejście do danych identyfikacyjnych możemy trzymać dwa lata, później jest ono niszczone.

Dlaczego?

To jest właśnie gwarancja tajemnicy statystycznej. Zresztą sami chcemy to zrobić jak najszybciej, bo zdarza się, że różne instytucje, jak prokuratura, starają się dojść do tych danych, mimo iż nie mamy prawa ich nikomu udostępnić. To jest często okupowane procesami sądowymi, szczególnie za czasów ministra Ziobry prokuratura domagała się od nas danych konkretnych osób. Oczywiście nie udostępniliśmy ich.

A co w przypadku złamania tajemnicy?

Każdemu z nas, kto udzieliłby jakiegokolwiek informacji dotyczącej danej jednostkowej lub osobowej, grozi kara 5 lat pozbawienia wolności. Dane jednostkowe to są nawet takie dane, które już nie zawierają imienia i nazwiska, ale pozwalają zidentyfikować osobę. Na przykład nie wolno nam powiedzieć nic o działalności szewskiej w miejscowości, w której jest jeden szewc, a nawet dwóch. Dopiero jak jest ich trzech, dane możemy udostępnić.

GIS wykorzystujecie do prowadzenia analiz, ale także do przygotowania spisu i zarządzania nim.

Przygotowanie spisu polegało na opracowaniu wykazów adresowo-mieszkańcowych, współrzędnych naszych mieszkań, a w konsekwencji nas, jeśli pozostajemy w relacji do jakiegoś mieszkania (tych relacji możemy ustawić wiele do wielu). W Polsce zidentyfikowaliśmy 6,048 mln budynków mieszkalnych. Bazę przygotowaliśmy jeszcze przed spisem i zweryfikowaliśmy te dane przy pomocy naszych rachmistrzów oraz gmin i jest to jedyna w kraju porządna baza adresowa ze współ-

Nie, to wybrani pracownicy gmin, których szkoliliśmy już od listopada. W Warszawie jest ich 70. Lider gminny ma dedykowaną aplikację, widzi mapę i marszrutę rachmistrza, ale nie ma dostępu do ankiet. **Lider wykonuje tę pracę w ramach służbowych obowiązków?**

Jest to praca dodatkowa, ale jeśli lider chce się dobrze wywiązywać z obowiązków spisowych, to niewiele czasu zostaje mu na te służbowe. Wynagrodzenie otrzymuje w dotychczasowej wysokości plus dodatek spisowy.

Jak wybieracie osoby do ankiet reprezentacyjnej?

Metodolodzy przygotowali algorytmy do losowania mieszkań. Algorytmy te uwzględniały, jakie osoby mają zamieszkiwać wylosowane mieszkania, tak aby w próbie znaleźli się zarówno emeryci, renciści, osoby pracujące czy rodziny wielodzietne. Ponadto określona została również próba dla każdej gminy, bo 20% to jest średnia dla Polski. W dużych aglomeracjach,

np. w Warszawie, próba jest mniejsza.

A skąd wiecie, że w danym mieszkaniu mieszka emeryt?

Ze źródeł administracyjnych: z NFZ, ZUS, KRUS.

Jak przechowywane są dane spisowe?

W urzędzeniach mobilnych, na których jest wypełniana ankieta, dane są szyfrowane, a następnie przesyłane do Operacyjnej Bazy Mikrodanych, w której są przechowywane. W Radomiu, gdzie mamy ośrodek przetwarzania danych, jest centrum zapasowe.

Jak 30 czerwca sprawdzicie, czy wszyscy zostali spisani?

Na bieżąco monitorujemy status każdego punktu. Osoby identyfikujemy przez PESEL i adres. Jeżeli chodzi o badanie reprezentacyjne, sprawdzamy, czy wszystkie wylosowane mieszkania zostały spisane i zatwierdzone przez dyspozytorów. Z kolei w badaniu pełnym monitorujemy przebieg spisu, biorąc pod uwagę wypełnienie ankiety dla osób objętych spisem.

Dla 90% populacji objętej badaniem pełnym większość danych pozyskaliśmy z rejestrów administracyjnych.

Czy przy spisie przydały się pani doświadczenia z pracy w GUGiK?

Na pewno pomogły mi właściwie identyfikować materiały źródłowe dostępne w PZGiK. No i od początku wiedziałam, że EGIB nie wystarczy pobrać, bo dane otrzymamy w SWDE, z którego nie będzie łatwo skorzystać. Zdawałam sobie sprawę z jakości tych danych. Musieliśmy przekonwertować je do SQL. Następnie okazało się, że dane są niepełne – często brakuje np. roku wybudowania budynku. Ten rejestr sprawił nam najwięcej problemów. Natomiast z adresami problem był we wszystkich rejestrach, bo nie ma w nich obowiązkowych słowników. Łódź pisana jest na kilkaset sposobów. To wszystko trzeba było wyczyścić.

Jakie dane przyjęliście jako wzorcowe?

Główną warstwą był TE-RYT, który zawiera słowniki,

i z nim porównywaliśmy inne rejestry. Ale i tak nie udało się zachować pełnej spójności. Na przykład budynek 4a i 4A to ten sam czy nie? Czasami udawało się to rozstrzygnąć po PESEL-ach mieszkańców.

I to wszystko było robione ręcznie?!

Wykorzystane zostały najnowocześniejsze rozwiązania – procesy ETL-owe, ale i tak kosztowało nas to ponad rok pracy. Najprościej można to wyjaśnić tak, że przy czyszczeniu pierwszego zbioru były zapisane reguły, co zmienić na co. Każdy następny był już czyszczony regułami z tego poprzedniego, najwyżej dodawano następne reguły.

Czyli na następny spis dane macie już wyczyszczone.

Powoli przymierzamy się do tego, żeby korzystać bezpośrednio z rejestrów. Dane z tych rejestrów, które się sprawdziły, chcielibyśmy pobierać co roku. Być może kiedyś również spis budynków przeprowadzimy tylko na podstawie danych z EGIB. ■