

Cena zł 10,00

Indeks 381306
PL ISSN 0043-518X

WIADOMOŚCI STATYSTYCZNE

GŁÓWNY
URZĄD
STATYSTYCZNY

POLSKIE
TOWARZYSTWO
STATYSTYCZNE

MIESIĘCZNIK
ROK XLVI
WARSZAWA
WRZESIEŃ 2001

9

w numerze m.in.:

JÓZEF PISKORZ

Program rozwoju informatyki w statystyce publicznej

DOROTA PEKASIEWICZ

Zastosowanie metod estymacji sekwencyjnej do badania rozmieszczenia sieci telekomunikacyjnej

PAWEŁ KACZOROWSKI, ALEKSANDRA ROGUT, TOMASZ TOKARSKI
Sektorowe zmiany strukturalne gospodarki w ujęciu regionalnym

DOROTA KWIATKOWSKA-CIOTUCHA, URSZULA ZAŁUSKA
Ranking rodzajów działalności produkcyjnej w Polsce w latach 1995—1999

RAFAŁ CZYŻYCKI, MARCIN HUNDERT, RAFAŁ KLÓSKA
Zróżnicowanie sytuacji społeczno-gospodarczej w woj. zachodniopomorskim



KOLEGIUM REDAKCYJNE:

prof. dr hab. Tadeusz Walczak (red. nac. tel. 608-32-89, e-mail: t.walczak@stat.gov.pl),
dr Stanisław Paradysz (zast. red. nac.), prof. dr hab. Józef Zegar (zast. red. nac. 826-14-28),
Anatol Kula (sekretarz redakcji 608-32-25, e-mail: a.kula@stat.gov.pl), mgr Jan Berger (608-32-63),
mgr Alina Głuchowska (608-32-40), mgr Wiesław Łagodziński (608-31-57), mgr Grażyna Marciniak
(608-33-54), dr Tomasz Pawlak (608-35-42), dr hab. Krystyna Pruska (0-42, 635-51-76),
mgr Lucyna Przybylska (0-32, 598-685), mgr Grażyna Rosiek (0-12, 656-31-55 w. 21),
prof. dr hab. Bogdan Stefanowicz (849-53-95), mgr inż. Henryka Wanke (608-34-62)

Redaktor techniczny — Zbigniew Karpiński

Korekta — Wydział Korekty ZWS

Skład i łamanie — Ewa Krawczyńska

REDAKCJA pok. 453, tel. 608-32-25, al. Niepodległości 208, 00-925 Warszawa

<http://www.stat.gov.pl/stale/pts/index.htm>

e-mail: e.grabowska@stat.gov.pl

RADA PROGRAMOWA:

dr Halina Dmochowska (przewodnicząca, tel. 608-34-32), mgr Magdalena Rychlik (sekretarz,
608-36-97), członkowie: dr Stanisław Bartzak, dr Marek Cierpiał-Wolan, prof. dr hab. Czesław
Domański, mgr Małgorzata Fronk, prof. dr hab. Jan Kordos, dr Teresa Śmiłowska, prof. dr hab.
Kazimierz Zając

ZAKŁAD WYDAWNICTW STATYSTYCZNYCH



al. Niepodległości 208, 00-925 Warszawa, tel. 608-31-45.

Informacje w sprawach nabywania czasopism tel. 608-32-10, 608-38-10

Indeks 381306

WARUNKI PRENUMERATY REALIZOWANEJ PRZEZ RUCH S.A.

Wpłaty na prenumeratę przyjmują:

- a) jednostki kolportażowe RUCH S.A. właściwe dla miejsca zamieszkania lub siedziby prenumerującego; dostawa egzemplarzy następuje w uzgodniony sposób;
- b) od osób zamieszkałych lub instytucji mających siedzibę w miejscowościach, w których nie ma jednostek kolportażowych RUCH S.A. istnieje możliwość dostawy prenumeraty „pod opaską”; wpłaty należy wносить na konto RUCH S.A., Oddział Krajowej Dystrybucji Prasy, PBK S.A. XIII O/Warszawa 370044-16551-2700-1-06 lub w kasach Oddziału Warszawa, ul. Towarowa 28 (poniedziałek—piątek, godz. 8.00—14.00); dostawa w takim przypadku odbywa się pocztą zwykłą w ramach opłaconej prenumeraty.

Cena prenumeraty ze zleceniem dostawy za granicą jest o 100% wyższa od krajowej. Wpłaty przyjmują kasy RUCH S.A. Oddział Krajowej Dystrybucji Prasy. Dostawa odbywa się pocztą zwykłą, z wyjątkiem dostawy pocztą lotniczą, której koszt w pełni pokrywa zamawiający.

Terminy przyjmowania wpłat na prenumeratę „Wiadomości Statystycznych”:

- do 05.12 — na I kwartał roku następnego lub na cały rok następny,
 - do 05.03 — na II kwartał roku bieżącego,
 - do 05.06 — na III kwartał roku bieżącego,
 - do 05.09 — na IV kwartał roku bieżącego.
-

NR 9 (484)

WIADOMOŚCI STATYSTYCZNE

CZASOPISMO GŁÓWNEGO URZĘDU STATYSTYCZNEGO
I POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO WRZESIEŃ 2001

Józef PISKORZ

Program rozwoju informatyki w statystyce publicznej

Spółczeństwo jutra to społeczeństwo informacyjne. W otoczeniu statystyki szybko wzrasta znaczenie informacji. Statystyka jest jednym z podstawowych systemów informowania społeczeństwa. Aby swoje zadania realizować efektywnie musi dążyć do maksymalnego wykorzystania narzędzi informatycznych do zbierania, opracowania i udostępniania informacji w celu ułatwienia i rozszerzenia dostępu użytkowników do informacji statystycznej, z której może skorzystać każdy użytkownik w miarę swoich potrzeb i zainteresowań.

Równoległym problemem są koszty badań. Statystyka ma wielu partnerów zarówno po stronie wejścia danych do systemu statystyki, jak i na wyjściu procesu badań statystycznych. Jediną drogą stopniowej obniżki kosztów jest maksymalne wykorzystanie narzędzi informatycznych i komunikacyjnych we wszystkich etapach realizacji badania — zbierania i przetwarzania danych statystycznych, jak i udostępniania informacji statystycznej jej użytkownikom.

Udostępnianie wyników badań statystycznych poprzez Internet jest jedną z głównych technik upowszechniania informacji w dobie społeczeństwa informacyjnego ze względu na maksymalny zasięg tego rozwiązania (np. w Norwegii liczbaostępów użytkowników do strony internetowej statystyki wynosi ponad 5 mln rocznie), jak również szybki dostęp do informacji z dowolnego miejsca w kraju.

INWESTYCJE W INFORMATYCE STATYSTYCZNEJ W OSTATNICH LATACH

Polska statystyka korzysta z narzędzi informatycznych od 35 lat. Opracowany w 1990 r. plan modernizacji informatyki, przewidujący zastosowanie tzw. systemów otwartych i systemów zarządzania bazami danych, rozwój centralnych banków danych, usprawnienie przesyłania danych i wyposażenie stanowisk statystyków w mikrokomputery, nie został zrealizowany w całości z powodu braku środków finansowych, ale codzienna praktyka potwierdziła słuszność przyjętego podejścia do informatyzacji badań statystycznych.

Ostatnimi programami inwestycyjnymi w informatyce statystycznej były programy wykonywane z funduszu PHARE: POLSTAT 2 i REJESTR ROLNY. Realizowane były one w latach 1995—1997. Łączne nakłady inwestycyjne w ramach tych programów wyniosły ok. 50 mln zł. W wyniku realizacji programów zbudowano sieć korporacyjną statystyki, umożliwiającą integrację techniczną jej systemu informatycznego. Do głównych elementów tej sieci należą: sieci lokalne urzędów statystycznych, sieć rozległa statystyki łącząca sieci lokalne z ośrodkiem centralnym oraz zestawy serwerów centralnych i serwerów urzędów statystycznych, a także stacje robocze stanowiące obecnie standardowe wyposażenie stanowisk pracy statystyków GUS i urzędów statystycznych. Programami tymi w niewielkim stopniu objęte zostały systemy aplikacyjne. Dotyczyło to głównie systemu informatycznego rejestru REGON, pracującego w trybie zbliżonym do systemów czasu rzeczywistego oraz udostępnianego w Internecie Banku Danych Regionalnych i systemu Centralnej Biblioteki Statystycznej im. S. Szulca.

Zbudowana infrastruktura stworzyła warunki do unowocześnienia i zwiększenia efektywności pewnej grupy systemów aplikacyjnych. Należą do niej m.in. system poczty elektronicznej statystyki, system prezentacji internetowej statystyki, w tym: GUS, urzędów statystycznych, CBS i innych jednostek organizacyjnych, udostępniane w Internecie bazy Banku Danych Regionalnych i DEMOGRAFIA, a także inne systemy o charakterze pomocniczym, wśród których wymienić należy system obsługi finansowej i kadrowej statystyki SOFTUS.

W latach 1998—1999 statystyka nie podejmowała nowych programów rozwojowych w infrastrukturze informatycznej, co spowodowało moralne zatarzenie się sprzętu komputerowego, przede wszystkim serwerów — zagraża to realizacji podjętych zadań aplikacyjnych. Przykładem niewydolności infrastruktury jest system informatyczny rejestru REGON, który wymagał, dwukrotnie, wielomiesięcznego przetwarzania zbiorów podmiotów gospodarczych do budowy systemu baz operacyjnych i użytkowych na poziomie centralnym i wojewódzkim (dla ok. 3 mln podmiotów nierolniczych, a następnie dla ok. 2 mln gospodarstw rolnych).

Ponadto podjęte prace nad nową generacją systemów aplikacyjnych wykazały, że zbudowana infrastruktura charakteryzuje się zbyt małą wydajnością i niezawodnością, aby realizować efektywnie zintegrowane systemy obsługujące badania statystyczne i zadania rejestrowe zlecone statystyce.

W latach 1999—2000 zostały podjęte kluczowe decyzje dotyczące trzech poważnych programów rozwojowych w statystyce. Są one następujące:

- uchwalenie przez Sejm RP ustawy o przeprowadzeniu Narodowego Spisu Powszechnego (NSP) i Powszechnego Spisu Rolnego (PSR) zapewniającej przydzielenie niezbędnych środków inwestycyjnych, które umożliwią poprawę istniejącej infrastruktury informatycznej statystyki publicznej w skali nie mniejszej niż program POLSTAT 2;
- podjęcie przez Komitet Integracji Europejskiej Rady Ministrów w uzgodnieniu z odpowiednimi organami Unii Europejskiej decyzji o uruchomieniu:
 - programu „Statystyka Regionalna” z funduszu PHARE’2000, przewidującego realizację zadań rozwojowych w informatyce statystycznej na kwotę ok. 10 mln zł,
 - programu „Harmonizacja Statystyki” z funduszu PHARE’2001, przewidującego realizację zadań rozwojowych w informatyce statystycznej na kwotę ok. 14 mln zł.

ZAŁOŻENIA ROZWOJU INFORMATYKI STATYSTYCZNEJ

System informatyczny statystyki stanowi bazę narzędziową do wspomagania realizacji badań statystycznych i zadań zleconych statystyce. Stosowanie narzędzi informatycznych jest warunkiem koniecznym realizacji badań.

System informatyczny statystyki składa się z dwóch warstw:

- funkcjonalnej — obejmującej systemy aplikacyjne, realizujące zadania zbierania, przetwarzania, przechowywania i udostępniania danych statystycznych oraz systemy zabezpieczające funkcjonowanie statystyki,
- infrastrukturalnej — stanowiącej bazę sprzętową, systemową i narzędziową do budowy systemów aplikacyjnych.

Modernizacja systemu badań statystycznych, wynikająca z potrzeb dostosowania badań do rynkowego charakteru gospodarki i procesów społecznych, związanych z przekształceniami gospodarczymi, powoduje dążenie do:

- integracji systemu informacyjnego statystyki,
 - maksymalnego wykorzystania statystycznych danych źródłowych,
 - pozyskiwania danych z systemów informacyjnych administracji publicznej,
- co pociąga za sobą wzrost stopnia skomplikowania badań statystycznych oraz całej infrastruktury informatycznej statystyki.

Szczególną rolę w procesie rozwojowym systemu informatycznego statystyki spełniają badania masowe: NSP oraz PSR. Badania te generują ogromne zapotrzebowanie na rejestrację i redagowanie danych oraz na szybkie udostępnianie informacji wynikowych w różnorodnych przekrojach. Stwarzają one także możliwości modernizacji i rozbudowy infrastruktury informatycznej statystyki do poziomu wynikającego z potrzeb realizacji zadań spisowych.

Zadanie modernizacji i rozbudowy infrastruktury informatycznej polega na takim zaprojektowaniu inwestycji koniecznych dla potrzeb spisowych, aby były maksymalnie użyteczne dla wieloletnich potrzeb rozwoju badań statystycznych.

ZWIĘKSZENIE EFEKTYWNOŚCI REALIZACJI ZADAŃ STATYSTYKI

Efektywne udostępnianie danych jest uwarunkowane posiadaniem efektywnego systemu jej zbierania i przetwarzania.

Systemy informatyczne obsługujące badania statystyczne oraz systemy informatyczne rejestrów urzędowych oraz baz i banków danych funkcjonują obecnie na serwerach małej mocy, które w znacznym stopniu zestarzały się moralnie, a często są zbyt mało wydajne, aby w efektywny sposób realizować zadania statystyki w poszczególnych obszarach. Systemy te nie są również zintegrowane pod względem informacyjnym — mają najczęściej charakter wycinkowy dostosowany do obsługi jednego badania statystycznego — co wydłuża czas realizacji zadań wielotematycznych i zapytań o charakterze kompleksowym.

Wobec powyższego niezbędna jest modernizacja systemu informatycznego statystyki, umożliwiająca skrócenie czasu realizacji zadań informacyjnych i zmniejszenie ich kosztu przez wyeliminowanie powtarzalnej pracy ludzkiej. Modernizacja ta dotyczy musi poprawy infrastruktury informatycznej statystyki oraz integracji systemu informacyjnego statystyki.

Głównym elementem tej integracji jest zaprojektowanie i wdrożenie Wspólnej Bazy Opracowań (WBO) wraz z budową jej metabazy, która będzie spełniać wszelkie wymogi hurtowni danych w efektywny sposób dostępnej dla całej statystyki publicznej i wszystkich użytkowników informacji statystycznej.

Poprawa poziomu bezpieczeństwa funkcjonowania statystyki

Na bezpieczeństwo funkcjonowania statystyki składają się następujące elementy:

- 1) niezawodność działania systemów statystycznych, umożliwiająca uzyskiwanie wiarygodnych wyników w przewidzianych terminach oraz
- 2) ochrona danych oraz zasobów informacyjnych statystyki.

Osiągnięcie celu pierwszego wymaga budowy bezpiecznych rozwiązań sprzętowych i systemowych oraz kadrowych. Do zapewnienia płynnej realizacji zadań statystyki konieczne jest m.in. wprowadzanie nowoczesnych technologicznie rozwiązań sprzętowych.

Potrzeba podniesienia poziomu ochrony danych statystycznych wynika z ustawy o statystyce, jak również z ustaw o ochronie danych osobowych i informacji niejawnych. Generuje ona dodatkowe wymagania dotyczące ochrony obiektów statystyki i jej systemów komputerowych. Głównymi kierunkami działania będą tu elektroniczne systemy kontroli dostępu do obiektów i ich wydzielonych części, ograniczanie dostępu osób nieupoważnionych do miejsc przechowywania i przetwarzania danych, a także przemieszczanie funkcji przechowywania danych jednostkowych ze stacji roboczych na lepiej chronione serwery.

KIERUNKI PROCESU ROZWOJOWEGO

Do głównych priorytetów w nadchodzących latach należy zaliczyć poprawę efektywności działania, poprzez rozwój technologii informatycznych, w następujących obszarach statystyki :

- pozyskiwania danych statystycznych, w tym ze źródeł administracyjnych,
- przetwarzania danych statystycznych, w tym projektowania metabazy i budowy hurtowni danych,
- udostępniania danych statystycznych,
- systemów specjalistycznych, w tym rejestrów i systemu przetwarzania danych spisowych,
- systemów wspomagania organizacji prac statystycznych, administracyjnych i finansowych statystyki,
- bezpieczeństwa funkcjonowania statystyki, w tym ochrony danych statystycznych,
- szkolenia,
- infrastruktury informatycznej.

ROZWÓJ SYSTEMU POZYSKIWANIA DANYCH STATYSTYCZNYCH

Tradycyjnie polska statystyka publiczna wykorzystuje formularze i ankiety papierowe w celu gromadzenia danych źródłowych dla badań statystycznych. Formularze te po zebraniu są sprawdzane i kodowane, a następnie rejestrowane i redagowane w urzędach statystycznych i ich oddziałach. Technologia ta wymaga wielkich nakładów pracy ręcznej statystyków w urzędach statystycznych.

W warunkach rosnących zadań statystyki, przy ustabilizowanym budżecie i zatrudnieniu, niezbędne są działania restrukturyzacyjne prowadzące do oszczędności etatowych przy pracach prostych i rozwój stanowisk pracy o zadaniach analitycznych, metodologicznych oraz zapewniających bezpieczeństwo danych. Osiągnięcie tego celu jest możliwe w wyniku przedstawionych dalej działań.

Systemy OCR (Optical Charakter Recognition)

Zastosowanie systemu automatyzacji procesu optycznego odczytu dokumentów (OCR) umożliwi zwiększenie wydajności procesu rejestracji danych 2—3 razy, a także poprawę jakości danych. Automatyzacja tego procesu wymaga jednak inwestycji infrastrukturalnych oraz wymaga opracowania specjalistycznego oprogramowania do rejestracji i redagowania danych. Jest to więc technologia racjonalna dla badań o dużej liczbie formularzy do rejestracji.

Zakłada się stopniowe przeprojektowanie, z zastosowaniem technologii OCR, aplikacji obsługujących badania o dużej liczbie formularzy, szczególnie zbudowanych na zasadzie pytań ze zdeterminowanym zbiorem odpowiedzi. Proces ten obecnie oparty jest na zestawach OCR, użytkujących system rozpoznania znaków EBES firmy ELSAG. W zestawy takie zostały wyposażone wszystkie wojewódzkie urzędy statystyczne.

Ze względu na niewielką liczbę i wydajność zakupionych dotąd skanerów dla potrzeb obsługi NSP i PSR (w których wystąpi ok. 270 mln stron formularzy A4) muszą być zakupione nowe wydajniejsze skanery, które umożliwią uzyskanie wystarczającego poziomu wydajności i niezawodności technicznej systemów OCR dla potrzeb wykonania zadań spisowych. Potencjał ten będzie używany później do realizacji bieżących badań statystycznych.

Równoczesne pozyskanie dla potrzeb spisowych efektywnych rozwiązań aplikacyjnych do realizacji korekty graficznej oraz automatycznej symbolizacji danych słownikowych zostanie zapewnione przez ponad 20 systemów OCR, umożliwiających efektywną i niezawodną realizację zadań rejestracji danych ze sprawozdań, przy dużym ograniczeniu pracy ludzi.

Będzie to istotny krok w kierunku unowocześnienia metod wprowadzania danych w przyszłości dla wszystkich tych przedsięwzięć statystyki, które będą wymagały przetworzenia wielkich ilości dokumentów.

Pełna efektywność zastosowania systemów OCR w służbach statystycznych zostanie osiągnięta po wdrożeniu systemu METAFRAME, umożliwiającego przetwarzanie danych ze stacji roboczych w oddziałach, znajdujących się na serwerach urzędów statystycznych w województwach.

Sieć ankietów statystycznych

Obsługa zbierania i rejestracji danych z badań ankietowych przebiega obecnie dwufazowo: ankieteryzy wypełniają w trakcie wywiadu formularze ankiet, które następnie podlegają rejestracji w urzędach statystycznych. Taka organizacja pracy powoduje konieczność przyjazdu ankietów do urzędów statystycznych w celu zarejestrowania danych na komputerach stacjonarnych oraz — w przypadku wykrycia przez system aplikacyjny błędnych danych — konieczność ponownego nawiązania kontaktu z osobami ankietowanymi: telefonicznego lub osobistego. Dane zbierane przez ankietów są danymi szczególnie wrażliwymi, gdyż zazwyczaj dotyczą osób. Obecny system pracy nie daje możliwości dostatecznego zabezpieczenia danych wielokrotnie przenoszonych przez ankietów na dokumentach papierowych w różnych warunkach oraz przechowywanych w domach, czyli poza obszarami przygotowanymi do ochrony danych jednostkowych w urzędach.

Wyposażenie techniczne ankietów

W celu skrócenia czasu oraz zmniejszenia kosztów zakłada się wyposażenie ankietów w rejestratory danych, umożliwiające wykonanie rejestracji danych ankietowych oraz niezbędnej kontroli i korekty tych danych w trakcie wywiadu, a następnie transmisję zarejestrowanych danych do serwera w urzędzie statystycznym. Wprowadzane dane będą kontrolowane co do zakresu i typu. Będzie też możliwość ich korygowania. Zmniejsza się dzięki temu ilość powrotów do punktów prowadzenia wywiadów i notowań w celu wyjaśnienia błędów. Wcześniejsze próby poszukiwania takich rozwiązań nie przyniosły rezultatu wobec wysokich kosztów zakupu notebooków oraz niestabilnego składu zespołów ankieterskich.

Obecna organizacja pracy ankieterów, zakładająca stabilność i specjalizację zespołów ankieterów wykonujących różne badania, sprzyja ponownemu podjęciu tego przedsięwzięcia. Równocześnie na rynku pojawiły się tańsze rejestratory danych (komputery typu hand-held i palmtop) oraz odpowiednie oprogramowanie w postaci uproszczonego systemu operacyjnego WINDOWS CE. Wdrożenie tego rozwiązania wymaga dokonania wyboru i zakupu najlepszych dla realizacji badań ankietowych rejestratorów danych z modemami do transmisji danych. Konieczne jest również przeprojektowanie aplikacji obsługujących badania ankietowe w taki sposób, by rejestracja i podstawowa kontrola danych była realizowana przez oprogramowanie rezydujące na zakupionych urządzeniach, zmniejszając w ten sposób nakład pracy ankieterów i odciążając od zadań rejestracji stacje robocze w urzędach statystycznych.

Rozwiązanie to umożliwia więc oszczędność pracy i kosztów poprzez:

- eliminację podwójnej rejestracji danych (na formularzu papierowym, na klawiaturze komputera),
- eliminację drukowania formularzy papierowych,
- eliminację powrotów do osób ankietowanych.

Aby efekty zrównoważyły koszty zastosowania tego rozwiązania liczba przeprowadzanych wywiadów musi być odpowiednio duża.

Opracowanie aplikacji

Zbudowanie opisanego systemu pracy ankieterów wymaga opracowania aplikacji dwóch rodzajów: rejestracji danych przez ankietera w trakcie wywiadu oraz przesłania danych do serwera urzędu statystycznego i włączenia ich do odpowiednich zbiorów. Aplikacje te powinny być przyjazne dla użytkownika dzięki przejrzystości i prostocie systemu konwersacji. Zespół projektujący te aplikacje musi być odpowiednio przeszkolony.

Organizacja współpracy ankieterów w sieci korporacyjnej

Sposób i terminy przekazywania danych przez ankieterów do urzędu statystycznego powinny być ściśle zdefiniowane przez obowiązujące harmonogramy realizacji badań oraz uwarunkowania techniczne (lub technologiczne). Dla badań powtarzanych cyklicznie ankieter ma obowiązek dostarczenia danych przed datą ustaloną harmonogramem. Nie ma więc potrzeby, aby ankieter nawiązywał komunikację z serwerem urzędu częściej niż raz w cyklu danego badania, chyba że planuje się analizę danych niekompletnych w trakcie ich zbierania.

Ankieter ma bezwzględny obowiązek ochrony danych jednostkowych, wobec czego dostęp do danych powinien być chroniony systemem haseł. Nakłada to dodatkowy obowiązek na ankietera, aby jego komputer, jak i hasła nie były dostępne dla osób postronnych.

Informatyczne szkolenie ankieterów

Bez względu na typ rozwiązania aplikacji dla ankieterów, konieczne będzie włączenie do cyklu szkoleń dla ankieterów szkolenia z obsługi komputera przenośnego (rejestratora danych), jak i samej aplikacji do przeprowadzania badania.

Personalizacja formularzy

Personalizacja formularzy polega na zastosowaniu formularza sprawozdawczego zawierającego dane organizacyjne oraz dane statystyczne za poprzedni okres sprawoz-

dawczy. Umożliwia to zmniejszenie pracochłonności sprawozdawców, którzy ograniczają swój nakład pracy przy wypełnianiu formularza wyłącznie do modyfikacji zmienionych w ostatnim okresie danych statystycznych. Ponadto, umożliwiając sprawozdawcy porównanie wpisywanych do formularza danych statystycznych z odpowiednimi danymi za okres poprzedni, personalizacja formularzy eliminuje znaczną część potencjalnych błędów sprawozdawcy — co w efekcie wpływa na poprawę jakości danych statystycznych.

Stosowanie formularzy personalizowanych wymaga poniesienia kosztów związanych ze zwiększoną objętością formularza i „inteligentnym” jego drukiem, tzn. drukiem o indywidualnej zawartości dla każdego sprawozdawcy. Taki druk nie jest możliwy na tradycyjnych maszynach offsetowych, lecz wymaga posiadania drukarek sterowanych komputerem w ilości odpowiedniej do liczby i objętości drukowanych formularzy oraz uwzględniającej rozkład druku tych formularzy w czasie. Minimalna liczba drukarek musi, ze względów niezawodnościowych, wynosić co najmniej dwie sztuki.

Alternatywą dla zakupu drukarek są usługi w wyspecjalizowanych drukarniach. Rozwiązanie takie wymaga jednak ustabilizowanego budżetu przeznaczonego na ten cel. Decyzja o podjęciu personalizacji każdego formularza musi być zatem poprzedzona wszechstronną analizą ekonomiczną, wartościującą wzrost jakości danych uzyskany w wyniku personalizacji i porównującą go ze wzrostem kosztu stosowania.

Formularze elektroniczne

Wykorzystanie do zbierania danych tzw. formularzy elektronicznych umożliwia zmniejszenie pracochłonności i kosztu realizacji procesu gromadzenia danych oraz poprawę jakości zbieranych danych w wyniku zastosowania w oprogramowaniu formularza systemu kontroli poprawności i spójności danych stosowanych obecnie w systemach ich rejestracji w urzędach statystycznych.

Praktycznie dostępne są dwie technologie zastosowania formularzy elektronicznych:

- zbieranie danych na dyskietkach; oprogramowanie formularza może być upowszechniane na dyskietkach lub na stronie www statystyki,
- zbieranie danych poprzez Internet.

Obecnie w fazie dynamicznego rozwoju komunikacji internetowej podstawową technologią rozwojową pozyskiwania danych od podmiotów gospodarczych powinna być technologia internetowa.

Zbieranie danych poprzez Internet

Rozwijający się handel internetowy, dla którego w najbliższych latach powstanie krajowa infrastruktura prawna oraz organizacyjna i techniczna, powoduje, że dla badań podmiotów gospodarczych najlepszym i najtańszym narzędziem do pozyskiwania i wprowadzania danych jest Internet. Możliwość taka w warstwie formalnej i praktycznej powstała w wyniku regulacji prawnych, zobowiązujących podmioty gospodarcze do składania tą drogą dokumentów ubezpieczeniowych do ZUS. Również Ministerstwo Finansów zapowiedziało stworzenie możliwości składania zeznań podatkowych przez Internet.

Aby udostępnić formularze elektroniczne dla szerokiej grupy sprawozdawców konieczne jest opracowanie internetowego standardu formularzy. Na przykładzie krajów, które wykorzystują tę technologię już obecnie wiadomo, że redukuje ona do 40% ilość błędów w danych źródłowych.

Podstawowym problemem w opracowaniu i wdrożeniu systemu zbierania danych przez Internet jest autoryzacja danych i ochrona ich poufności. Problem ten zostanie

rozwiązany w warstwie legislacyjnej przez ustawę o podpisie elektronicznym, która znajduje się w fazie prac sejmowych.

Doświadczenie uzyskane przy wdrożeniu systemu do badania Z-12 wykazało, że kwestia zaprojektowania prostego oprogramowania formularza elektronicznego jest sprawą niełatwą. Dlatego rozważenia wymaga możliwość przyjęcia oferty współpracy z firmami oferującymi tego rodzaju oprogramowanie, np. elektronicznych formularzy podatkowych lub utworzenie wyspecjalizowanego zespołu do projektowania takiego systemu.

Zbieranie sprawozdań na dyskietkach

Wspomniane już próby zbierania danych na nośnikach elektronicznych dla formularza Z-12 zostały tak przygotowane, aby respondent mógł pobrać oprogramowanie ze strony www statystyki i dostarczyć zarejestrowane i zweryfikowane za pomocą tego oprogramowania dane na dyskietkach. Rozszerzając zakres wykorzystywania programów do zbierania sprawozdań na dyskietkach należy dokonywać indywidualnej oceny efektywności tego rozwiązania, szczególnie w kontekście możliwości zbierania danych przez Internet.

Inne usprawnienia w standardowych metodach pozyskiwania danych

Do czasu uruchomienia systemu udostępniania programów rejestracji i kontroli danych źródłowych na stronie www statystyki możliwe i celowe jest jej wykorzystanie do udostępniania obrazów formularzy sprawozdawcom. Obraz taki może być wydrukowany i wypełniony ręcznie przez sprawozdawców. Ze względu na możliwe ogromne zróżnicowanie rodzajów papieru i jakości oraz kolorystyki druku, taki formularz musi być rejestrowany ręcznie. Efektem zastosowania tego rozwiązania jest możliwość wcześniejszego zapoznania się sprawozdawcy z formularzem oraz możliwość wysyłania do podmiotów gospodarczych po jednym egzemplarzu formularza. Rozwiązanie takie zostało już zastosowane przez statystykę (w formacie ACROBAT).

Pozyskiwanie danych ze źródeł administracyjnych

Podjęte zostały działania nad wykorzystaniem, jako źródłowych danych, statystycznych zbiorów danych z rejestrów i systemów informacyjnych administracji publicznej, przede wszystkim rządowej. Istotna dla statystyki możliwość wykorzystania tych danych wystąpi po uzyskaniu przez główne systemy informacyjne administracji rządowej (POLTAX, KSI ZUS, KRUS, inne) pełnej operacyjności działania.

Z punktu widzenia zasilenia statystyki danymi administracyjnymi, największym problemem jest brak spójności systemów informacyjnych statystyki i źródeł administracyjnych. W integracji systemów informacyjnych statystyki i źródeł administracyjnych istotna jest spójność metodologiczna, techniczna i technologiczna (sprzętu, oprogramowania, możliwości elektronicznej wymiany danych pomiędzy systemami informatycznymi z uwzględnieniem odpowiedniego poziomu zabezpieczeń).

Główne problemy informatyczne związane z wykorzystaniem w statystyce zbiorów danych administracyjnych związane są z koniecznością:

- budowy systemów dostępu i wymiany danych między administracyjnymi źródłami danych a systemem statystyki publicznej,
- zmian systemów aplikacyjnych obsługujących badania, dla:
 - włączania do zbioru statystycznego danych administracyjnych o różnym stopniu agregacji,

— uzupełniania pozostałej części potrzebnych danych z ograniczonych co do zakresu przedmiotowego i/lub podmiotowego zasobów administracyjnych.

Tryb i zakres przebudowy tych aplikacji musi być indywidualnie opracowany dla konkretnego zakresu przedmiotowego i podmiotowego danych administracyjnych dostępnych od określonego terminu dla statystyki i musi wynikać ze szczegółowych rozwiązań metodologicznych przyjętych w badaniu statystycznym.

ROZWÓJ SYSTEMU PRZETWARZANIA DANYCH STATYSTYCZNYCH

Statystyka eksploatuje ok. 200 systemów aplikacyjnych obsługujących realizację bieżących badań statystycznych. Ok. 140 spośród nich to systemy eksploatowane na dwóch poziomach organizacyjnych: centralnym i wojewódzkim. Są to systemy o różnym stopniu złożoności, wykonywane za pomocą różnych narzędzi programistycznych (od najstarszych do najnowszych) i pracujące w różnym środowisku systemowym.

Dynamiczne zmiany w metodologii i organizacji badań statystycznych, wynikające z procesu doskonalenia badań oraz ich dostosowania do standardów UE, powodują konieczność modernizacji systemów aplikacyjnych, obsługujących przetwarzanie danych statystycznych. Zmiany w metodologii realizacji badań powodują konieczność modyfikacji obsługujących je systemów informatycznych. Proces ten pochłania główną część mocy projektowych informatyki statystycznej. Stosuje się tu zasadę kompleksowego przeprojektowania systemu z zastosowaniem nowych narzędzi projektowania, w przypadku gdy efektywność funkcjonowania starej aplikacji jest niewystarczająca lub gdy obsługę informatyczną badania przejmuje nowa jednostka projektowa. Nowe narzędzia projektowania umożliwiają przeniesienie aplikacji ze środowiska znakowego do środowiska graficznego. Przenoszenie aplikacji do środowiska graficznego przyjęte zostało jako zasada postępowania w modernizacji systemów aplikacyjnych obsługi badań statystycznych.

Duża intensywność zmian w latach ubiegłych i realizacja tych zmian przez wyspecjalizowanych merytorycznie projektantów powodowały, że dokumentacja projektowa i użytkowa systemów aplikacyjnych obsługujących realizację badań jest na ogół bardzo uproszczona i nie zawsze aktualna. Powoduje to czasem zakłócenia przy przetwarzaniu danych w przypadku zmian kadrowych czy organizacyjnych. Wobec przewidywanego zmniejszenia dynamiki zmian tego rodzaju w systemach aplikacyjnych w nadchodzących latach niezbędne jest dokonanie analizy stanu dokumentacji projektowej i użytkowej oraz jej doprowadzenie do niezbędnego poziomu kompletności i szczegółowości.

Modyfikacja systemów aplikacyjnych obsługi badań w aspekcie współpracy z WBO

Istotnym etapem modernizacji systemów aplikacyjnych obsługi badań statystycznych będzie ich dostosowanie do zasilania WBO. Dostosowanie to obejmie zaprojektowanie modułów zasilania tej bazy danymi z poszczególnych systemów w standardzie ustalonym przez projekt lub w niezbędnych przypadkach przeprojektowanie całych systemów, jeśli efektywność zasilania struktur danych WBO przez posiadane systemy aplikacyjne badań okaże się niewystarczająca lub niezgodna logicznie z metabazą WBO.

Poprawa efektywności funkcjonowania systemów

Głównymi kierunkami poprawy efektywności pracy eksploatowanych w statystyce aplikacji obsługujących badania statystyczne są:

- unowocześnienie systemów wymagających w eksploatacji aplikacji dużych nakładów pracy ludzkiej na rzecz rozwiązań współczesnych, charakteryzujących się większą ergonomicznością i automatyzujących pracę ręczną, np.: zastosowanie technik OCR do rejestracji i redagowania danych, zastosowanie technik automatycznej symbolizacji danych o charakterze słownikowym, przechodzenie ze środowiska znakowego do środowiska graficznego, umożliwiającego szersze wspomaganie pracy operatora,
- budowa aplikacji w architekturze klient—serwer; zamierza się stosować wielowarstwową strukturę tego typu aplikacji. Wyróżniana jest tu warstwa klienta — zwykle jest to tzw. cienki klient, czyli najczęściej po prostu wywołanie ze stacji roboczej aplikacji na serwerze aplikacyjnym lub wywołanie przeglądarki internetowej (jeśli jest to aplikacja typu Intranet/ Internet). Serwer aplikacji (lub serwer www w tym drugim przypadku) jest więc tu drugą warstwą. Mogą wystąpić także kolejne warstwy: serwer relacyjnej bazy danych (lub grupa takich serwerów) i serwer komunikacyjny (lub grupa serwerów tego typu).

Rozwiązania te umożliwiają podniesienie efektywności przetwarzania danych na skutek wzrostu efektywności pracy pracowników statystyki oraz wzrostu wykorzystania zasobów posiadanej infrastruktury w wyniku zastosowania narzędzi umożliwiających rozpraszanie procesu przetwarzania (np. systemu METAFRAME). Ich zastosowanie wymaga jednak znacznych nakładów związanych z zakupem oprogramowania systemowego i narzędziowego oraz szkoleniem pracowników i przeprojektowaniem aplikacji obsługujących badania statystyczne.

Projekt WBO

Największym zamierzeniem rozwojowym systemu przetwarzania danych statystycznych jest projekt WBO. Składać się ona będzie z systemu baz danych, w tym baz danych hurtowni i systemu metabazy oraz systemów zasilających ją danymi z poszczególnych badań i systemów udostępniania danych. Obsługiwać będzie wszystkie etapy opracowań statystycznych: redagowanie i przetwarzanie danych oraz analizę i udostępnianie wyników badań.

System metabazy zawierający Bazę Jednostek Statystycznych, Bazę Klasyfikacji, Słownik Pojęć Statystyczno-Ekonomicznych i inne systemy uzupełniające będzie wspomagał etap projektowania i przygotowania oraz realizacji badań.

System baz danych statystycznych WBO zintegruje dane dotyczące tych samych podmiotów z różnych badań, umożliwi dokonywanie oceny jakości zebranych danych i automatycznej ich korekty oraz imputacji brakujących danych o podmiotach — na podstawie danych z poprzednich okresów i danych o podmiotach podobnych. Realizacja tak wielkiego przedsięwzięcia wymaga wielkich nakładów finansowych, pracy ludzi i dysponowania wsparciem wyspecjalizowanych wykonawców zewnętrznych. Zakłada się etapową budowę WBO.

Obecnie dysponujemy zatwierdzoną koncepcją merytoryczną WBO, określającą jej zakres i funkcje oraz podstawowe etapy budowy, z punktu widzenia potrzeb merytorycznych. Ponadto opracowany został plan organizacyjny realizacji tego przedsięwzięcia.

Prace wykonawcze obecnie dotyczą metabazy:

- zaprojektowany został system Słownik, którego wdrażanie znajduje się w fazie kompletowania bazy danych zawierającej pojęcia statystyczno-ekonomiczne,
- zaprojektowany został system Bazy Jednostek Statystycznych, stanowiący bazę operatorów do badań podmiotów gospodarki narodowej, którego wdrażanie znajduje się w fazie „czyszczenia” danych identyfikacyjnych i opisujących te podmioty,
- trwają prace przygotowawcze do przeprojektowania prototypu systemu Klasyfikacje.

W 2001 r. rozpoczynają się prace projektowe nad hurtownią danych z NSP i PSR, która stanowić będzie pierwszy podsystem hurtowni danych statystycznych WBO, rozbudowywanej po zakończeniu spisów o zbiory danych z pozostałych badań statystycznych. W szczególności w ramach tego projektu powinny być podjęte możliwie najszybciej moduły hurtowni niezbędne do zaspokojenia potrzeb programów rządowych oraz wynikających ze zobowiązań międzynarodowych.

Integracja organizacyjna procesów przetwarzania

Obecnie procesy przetwarzania danych na poziomie krajowym realizowane są w Centralnym Ośrodku Informatyki Statystycznej i kilkunastu US. Stan taki wynika ze stosowanej powszechnie w statystyce zasady integracji funkcji projektowania systemu i przetwarzania danych na szczeblu krajowym w jednym ośrodku informatycznym i z dawnej struktury informatyki statystycznej.

Zmiany organizacyjne, kadrowe i technologiczne, które nastąpiły w ostatnich latach, powodują potrzebę modyfikacji takiego modelu pracy służb informatycznych statystyki. Większość urzędów statystycznych nie jest w stanie obecnie realizować w sposób niezawodny i na wysokim poziomie jakościowym zadań projektowania nowych, znacznie bardziej skomplikowanych niż dawniej systemów informatycznych, koncentrując swoje niezbyt wielkie zasoby na wykonaniu rosnących zadań informatycznych poziomu regionalnego.

Poza sprawną i niezawodną realizacją zadań przetwarzania danych szczebla wojewódzkiego urzędy statystyczne muszą zapewnić sprawne funkcjonowanie rozległej sieci regionalnej, składającej się z ośrodka głównego w swojej siedzibie, ośrodków w oddziałach o pełnej infrastrukturze informatycznej i ośrodków wspomagających w małych oddziałach.

W sieci regionalnej realizowane będą w sposób kompletny wszystkie zadania statutowe urzędów statystycznych i ich dyrektor nadzorował będzie całą infrastrukturę urzędu oraz ponosił pełną odpowiedzialność za sprawne administrowanie wykonaniem zadań. W ten sposób dokonywana będzie integracja techniczna i organizacyjna urzędów statystycznych, odpowiednio do obecnej struktury organizacyjnej statystyki w terenie.

W tej sytuacji rozpoczęto koncentrację zadań projektowania i przetwarzania na szczeblu krajowym w COIS i jego zakładach w Radomiu i Łodzi. Przewiduje się stopniowe przenoszenie pozostałych centralnych zadań informatycznych do COIS. Tempo tego przekształcenia zależy od możliwości utrzymania w najsilniejszych urzędach statystycznych zespołów projektowych o niezbędnych możliwościach wykonawczych. W szczególności rozważenia wymaga celowość utrzymania centralnych zadań informatycznych w urzędach statystycznych, w których zlokalizowane zostały ważne centralne zadania statystyczne.

ROZWÓJ SYSTEMU UDOŚTĘPNIANIA DANYCH STATYSTYCZNYCH

Udostępnianie danych statystycznych realizowane jest w wielu różnorodnych formach. Tradycyjnej technice udostępniania wyników badań statystycznych, jaką są publikacje statystyczne sprzedawane lub udostępniane przez CBS i biblioteki urzędów statystycznych, towarzyszą obecnie formy związane z potrzebami współczesnego społeczeństwa informacyjnego oraz obecnymi możliwościami technicznymi. Należą do nich udostępnianie informacji przez Internet oraz udostępnianie, zamawianych przez użytkowników informacji statystycznej, specjalnych zestawień analitycznych na komputerowych nośnikach danych. Różnorodne techniki udostępniania informacji statystycznych są

stosowane w informatoriach statystycznych, które łączą funkcję podręcznej biblioteki z możliwością przeglądania zasobów informacji statystycznej z baz statystycznych za pomocą komputera.

Informatoria

Wysoki poziom usług osiągnęło Centralne Informatorium Statystyczne w GUS. Zarówno warunki lokalowe, jak i poziom informatyzacji znajdują się na zadowalającym poziomie. Doskonali się jego działalność w wybranych szczegółowych zagadnieniach. Dotyczy to przede wszystkim podniesienia poziomu usług informacyjnych, świadczonych telefonicznie — co zamierzamy uzyskać w wyniku wykorzystania dodatkowych funkcji nowej wersji oprogramowania centrali telefonicznej GUS.

Szerszym problemem jest poprawa poziomu usług informatoriów w urzędach statystycznych. Zakłada się tu doprowadzenie do jednolitego standardu usług poprzez jednolite wyposażenie w sprzęt informatyczny i dostępne na nim zbiory danych statystycznych. Celowe jest także udostępnianie na stanowisku komputerowym informatorium dostępu do katalogu biblioteki urzędu statystycznego oraz katalogu CBS. Ponadto konieczne jest zapewnienie w niezbędnych przypadkach wsparcia pracowników informatoriów urzędów statystycznych przez bardziej wyspecjalizowanych fachowców CIS. Sprzyjać temu będzie zamierzone przeniesienie poczty głosowej na łącza sieci rozległej statystyki w relacjach między GUS i urzędami statystycznymi.

System wydawniczy

System wydawniczy statystyki zapewnia opracowanie i wydawanie publikacji na dwóch szczeblach struktury organizacyjnej statystyki: centralnym i regionalnym.

Głównymi kierunkami rozwoju systemu wydawniczego w aspekcie jego informatyzacji są:

- automatyzacja przygotowania publikacji statystycznych,
- rozszerzenie zakresu tzw. druku inteligentnego,
- wzrost liczby publikacji na płytach CD.

Pierwszy kierunek działania zmierza do zwiększenia efektywności przygotowania publikacji do druku poprzez maksymalne wykorzystanie projektu publikacji, opracowanego przez autorów publikacji. Umożliwia to oszczędność pracy ludzkiej i skrócenie czasu wydania publikacji poprzez eliminację składu publikacji w drukarni. Uzyskanie tego efektu jest możliwe w wyniku zastosowania odpowiednich, jednolitych, standardów narzędziowych przygotowania publikacji. W przypadku publikacji o charakterze roboczym standard taki, np. MS Word, umożliwia wprost wydrukowanie publikacji przygotowanej w departamencie autorskim. Publikacje o wymaganym wysokim poziomie edytorskim po przygotowaniu przez autorów w ustalonym standardzie wyjściowym mogą być wykorzystane w całości przez przyjęty w drukarni standard edytorski. Pracownicy poligrafii wykorzystują jedynie dodatkowe funkcje własnego standardu do podniesienia jakości edytorskiej publikacji. Taki system pracy wymaga kompatybilności obu standardów „statystycznego” i „poligraficznego” — eliminuje natomiast pracochłonne i czasochłonne powtórne ręczne składanie publikacji z użyciem specyficznych narzędzi poligraficznych, stosowanych dotąd w statystyce.

Celem podjęcia drugiego kierunku rozwoju systemu wydawniczego jest zastąpienie druku małych nakładów publikacji za pomocą maszyn offsetowych drukiem na skomputeryzowanych drukarkach lub drukarko-kopiarkach. Zmiana ta umożliwi elastyczne dopasowanie nakładu publikacji do potrzeb użytkowników eliminując druk „na skład”

— ponieważ możliwe jest szybkie dodrukowanie potrzebnej liczby egzemplarzy poprzez uruchomienie programu druku. W przypadku małych nakładów publikacji jest to również druk najtańszy.

Zakłada się wyposażenie urzędów statystycznych w sprzęt i oprogramowanie umożliwiające wydawanie podstawowych publikacji we własnym zakresie, ewentualnie przy uzupełniających usługach ZWS. Najważniejsze publikacje urzędów statystycznych będą wydawane w całości w ZWS.

Znaczne obniżenie kosztów wydania publikacji na płytach CD, jakie ma miejsce w ostatnim okresie, oraz gwałtowny wzrost liczby komputerów biurowych i domowych powodują, że szybko rośnie zapotrzebowanie na publikacje na takim nośniku. Obecnie wydajemy na płytach CD tylko podstawowe roczniki statystyczne, ale w najbliższym okresie, poczynając od publikacji spisowych, możliwe będzie znaczne rozszerzenie tej oferty.

Biblioteki

System informatyczny CBS im. S. Szulca został rozbudowany poprzez wymianę jego serwera oraz zainstalowanie nowszej wersji systemu ALEPH — co spowodowało zwiększenie wydajności i funkcjonalności całego systemu. Dalsza modernizacja systemu CBS dokonana zostanie w ramach programu PHARE'2000, obejmując m.in. implementację modułów systemu ALEPH dla bibliotek w informatoriach urzędów statystycznych i ich połączenie z CBS, za pośrednictwem którego informatoria urzędów statystycznych uzyskają także dostęp do sieci bibliotek w kraju i na świecie.

Podstawowym zadaniem systemu bibliotecznego będzie istotne zwiększenie zakresu posiadanych opisów bibliograficznych własnych zasobów bibliotecznych. Wykonanie tej pracy jest zadaniem czasochłonnym i kosztochłonnym, przede wszystkim dla CBS posiadającej wielkie zasoby biblioteczne.

Udostępnianie informacji statystycznej w Internecie

Internet jest obecnie uznawany za najbardziej efektywny i przyszłościowy system upowszechniania informacji, w tym także informacji statystycznej. W Norwegii liczba kontaktów internetowych ze stroną www urzędu statystycznego tego kraju jest większa niż liczba ludności. Strona www GUS nie jest jeszcze tak często odwiedzana, ale liczba kontaktów ciągle rośnie. Obecnie udostępniamy w Internecie informację o polskiej statystyce publicznej, jej zadaniach i strukturze, dane bieżące oraz trzy bazy danych: BDR, DEMOGRAFIĘ i bazę SDDS.

Dalszy rozwój tego systemu udostępniania informacji statystycznej powinien objąć:

- informacje z metabazy danych statystycznych — klasyfikacje statystyczne, słownik pojęć statystyczno-ekonomicznych,
- kolejne bazy danych statystycznych, w tym cieszącą się wielkim zainteresowaniem bazę danych handlu zagranicznego HANZA,
- wyciągi z rejestru podmiotów gospodarczych REGON,
- rejestr podziału terytorialnego kraju TERYT.

Udostępniane w Internecie bazy danych muszą zawierać bardzo dokładny i prosty opis tych danych, czyli metabazę umożliwiającą jednoznaczna i poprawną interpretację pozyskiwanych przez użytkowników automatycznie danych.

Ważnym problemem do rozwiązania, związanym z udostępnianiem danych przez Internet, jest sposób poboru opłaty za udostępniane dane. Stosowany system musi być bardzo łatwy w użyciu, aby był efektywny i przyjazny dla indywidualnych użytkowników informacji, także doraźnie korzystających z zasobów informacyjnych statystyki.

Hurtownia danych

Najbardziej efektywnym narzędziem przygotowania danych do udostępniania jest system hurtowni danych, jako główna część WBO.

W założeniach do projektu hurtowni danych statystycznych przyjęto, że wybrany system hurtowni danych musi dysponować na wyjściu efektywnymi narzędziami udostępniania danych: do generowania raportów, do generowania publikacji, do podawania danych w Internecie oraz narzędzia typu OLAP (dynamiczne generowanie odpowiedzi na zapytania), a także narzędzia typu data mining (do pogłębionej analizy zawartych w bazach hurtowni różnorodnych danych). Przewiduje się etapowy zakup tych narzędzi, zależnie od dysponowanych funduszy i możliwości rzeczywistego ich zastosowania w praktyce statystycznej.

SYSTEMY SPECJALISTYCZNE

Zaliczono tu systemy, które nie realizują bezpośrednio zadań związanych z realizacją bieżących badań statystycznych, lecz ich zadaniem jest obsługa NSP i PSR oraz zadań zleconych statystyce, jak: prowadzenie rejestrów podmiotów gospodarczych i podziału terytorialnego kraju lub wspomagających organizację badań statystycznych BJS czy też udostępniania danych statystycznych (centralne bazy danych statystycznych, bazy publiczne, geograficzne systemy informatyczne).

Oto ich krótka charakterystyka oraz zamierzenia rozwojowe.

System informatyczny spisów

NSP i PSR stanowią główne zadanie statystyki w ciągu dekady. Jest to zadanie o ogromnej skali i wielkich kosztach realizacji. Budowa systemu informatycznego spisów jest przedsięwzięciem o charakterze doraźnym. System musi być zbudowany w ściśle określonym terminie, działać efektywnie w okresie ok. jednego roku, a następnie zostanie zrekonfigurowany dla potrzeb realizacji bieżących badań statystycznych. Budowane dla potrzeb realizacji spisu zasoby infrastrukturalne muszą być tak dobierane, aby wykonane oprogramowanie i kupowany sprzęt w możliwie maksymalnym stopniu mógł być użytkowany w statystyce po zakończeniu prac spisowych.

System informatyczny spisów zapewnić musi kompletną realizację zadań spisowych, w następujących fazach:

- przygotowania spisów,
- rejestracji i redagowania danych spisowych,
- udostępniania danych wynikowych spisów.

Faza przygotowania spisów wymaga wykonania dwóch podstawowych systemów:

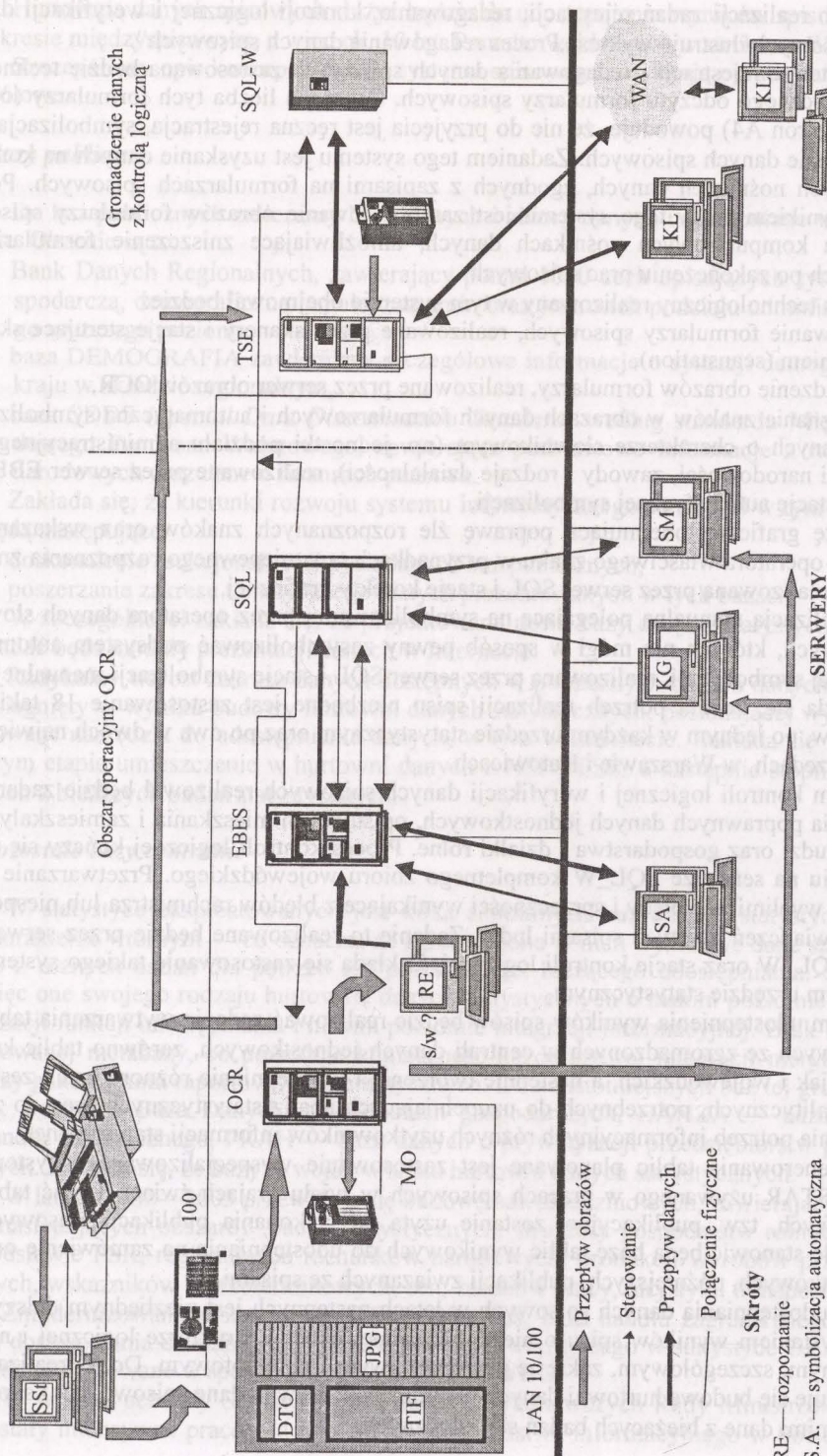
- przygotowania danych adresowych dla ok. 12 mln mieszkań i budynków oraz ok. 3 mln gospodarstw w działek rolnych,
- obsługi administracyjnej zatrudnienia i ubezpieczenia oraz rozliczenia finansowego ok. 170 tys. rachmistrzów spisowych.

Zakłada się zastosowanie każdego z tych systemów w każdym urzędzie statystycznym.

Faza rejestracji i redagowania danych spisowych wymaga zbudowania trzech podstawowych systemów:

- rejestracji i symbolizacji danych z formularzy spisowych,
- kontroli logicznej i weryfikacji zarejestrowanych danych,
- gromadzenia danych wynikowych spisów.

PROCES REDAGOWANIA DANYCH SPISOWYCH



- RE — rozpoznanie
 S.A. — symbolizacja automatyczna
 KG — korekta graficzna
 SM — symbolizacja manualna
 KL — kontrola logiczna
- OCR — Serwer NT do rozpoznania znaków w systemach OCR
 EBES — Serwer NT do symbolizacji automatycznej
 SQL — Serwer NT do korekty graficznej i symbolizacji manualnej
- TSE — Serwer NT do zarządzania obrazami formularzy spisowych
 SQL_W — Serwer NT do kontroli logicznej i weryfikacji danych

Sposób realizacji zadań rejestracji, redagowania, kontroli logicznej i weryfikacji danych spisowych ilustruje wykres „Proces redagowania danych spisowych”.

W systemie rejestracji i redagowania danych spisowych zastosowana będzie technologia optycznego odczytu formularzy spisowych. Ogromna liczba tych formularzy (ok. 270 mln stron A4) powoduje, że nie do przyjęcia jest ręczna rejestracja, symbolizacja i redagowanie danych spisowych. Zadaniem tego systemu jest uzyskanie danych na komputerowych nośnikach danych, zgodnych z zapisami na formularzach spisowych. Ponadto wynikiem pracy tego systemu jest zarchiwizowanie obrazów formularzy spisowych na komputerowych nośnikach danych, umożliwiające zniszczenie formularzy spisowych po zakończeniu prac spisowych.

Proces technologiczny realizowany w tym systemie obejmował będzie:

- skanowanie formularzy spisowych, realizowane przez skanery i stacje sterujące skanowaniem (scanstation),
- gromadzenie obrazów formularzy, realizowane przez serwer obrazów OCR,
- rozpoznanie znaków w obrazach danych formularzowych i automatyczna symbolizacja danych o charakterze słownikowym (np. jednostki podziału administracyjnego, kraje i narodowości, zawody i rodzaje działalności), realizowane przez serwer EBES oraz stację automatycznej symbolizacji,
- korektę graficzną obejmującą poprawę źle rozpoznanych znaków oraz wskazanie przez operatora właściwego znaku w przypadkach tzw. niepewnego rozpoznania znaków, realizowaną przez serwer SQL i stację korekty graficznej,
- symbolizacją manualną polegającą na symbolizowaniu przez operatora danych słownikowych, których nie mógł w sposób pewny zasymbolizować podsystem automatycznej symbolizacji, realizowaną przez serwer SQL i stację symbolizacji manualnej.

Zakłada się, że dla potrzeb realizacji spisu niezbędne jest zastosowanie 18 takich systemów, po jednym w każdym urzędzie statystycznym oraz po dwa w dwóch największych urzędach, w Warszawie i Katowicach.

System kontroli logicznej i weryfikacji danych spisowych realizował będzie zadanie uzyskania poprawnych danych jednostkowych, opisujących mieszkania i zamieszkałych w nich ludzi oraz gospodarstwa i działki rolne. Proces kontroli logicznej kończy się po uzyskaniu na serwerze SQL_W kompletnego zbioru wojewódzkiego. Przetwarzanie w tej fazie wyeliminuje błędy i sprzeczności wynikające z błędów rachmistrza lub niespójnych oświadczeń objętych spisami ludzi. Zadanie to realizowane będzie przez serwery TSE i SQL_W oraz stacje kontroli logicznej. Zakłada się zastosowanie takiego systemu w każdym urzędzie statystycznym.

System udostępnienia wyników spisów będzie realizować zadanie wytwarzania tablic wynikowych ze zgromadzonych w centrali danych jednostkowych, zarówno tablic krajowych jak i wojewódzkich, a następnie tworzenie i udostępnianie różnorodnych zestawień analitycznych, potrzebnych do uzupełniających analiz statystycznych oraz do zaspokojenia potrzeb informacyjnych różnych użytkowników informacji statystycznej.

Do generowania tablic planowane jest zastosowanie wyspecjalizowanego systemu SUPERSTAR używanego w pracach spisowych w wielu krajach świata. Część tablic wynikowych, tzw. publikacyjne, zostanie użyta do wykonania publikacji spisowych. Pozostałe stanowiąc będą bazę tablic wynikowych do udostępniania na zamówienie oraz do dodatkowych, późniejszych, publikacji związanych ze spisami.

Faza udostępniania danych spisowych w latach następnych jest niezbędnym dalszym wykorzystaniem wyników spisu o niezdefiniowanej obecnie strukturze logicznej i nieokreślonym, szczegółowym, zakresie przedmiotowym i podmiotowym. Do jej realizacji przewiduje się budowę hurtowni danych obejmującej zarówno dane spisowe, jak i skojarzone z nimi dane z bieżących badań statystycznych.

Hurtownia danych spisowych służyć będzie do udostępniania wyników spisu w całym okresie międzyspisowym, a więc ok. 10 lat. Ponadto wejdzie ona w strukturę WBO.

Szczegółowy opis koncepcji całego systemu spisowego zawierają opracowania szczegółowe.

Bazy publiczne

Do baz publicznych zaliczamy bazy danych i zbiory danych udostępniane w Internecie. Obecnie są to:

- Bank Danych Regionalnych, zawierający ponad 1000 cech opisujących sytuację gospodarczą, demograficzną i społeczną gmin oraz jednostek podziału administracyjnego wyższego poziomu,
- baza DEMOGRAFIA zawierająca szczegółowe informacje o sytuacji demograficznej kraju w układzie regionalnym,
- baza SDDS (*Special Data Dissemination Standard*) według standardu Międzynarodowego Funduszu Walutowego, zawierająca podstawowe informacje z rachunków narodowych oraz dane o finansach państwa.

Zakłada się, że kierunki rozwoju systemu informatycznego statystyki w tym obszarze będą następujące:

- doskonalenie funkcjonalności posiadanych baz publicznych,
- poszerzanie zakresu tematycznego danych prezentowanych w tych bazach.

W szczególności zakłada się, że projektowane nowe bazy danych statystycznych posiadać będą moduły prezentacji danych w Internecie.

Radykalny wzrost zakresu danych dostępnych w publicznych bazach danych zostanie osiągnięty w wyniku budowy hurtowni danych statystycznych, posiadającej wyspecjalizowane narzędzia do udostępniania danych, w tym w Internecie. Zakłada się w pierwszym etapie umieszczenie w hurtowni danych z NSP i PSR, a następnie stopniowo danych z bieżących badań statystycznych.

Pozostałe bazy centralne

W statystyce eksploatowanych jest kilka centralnych baz danych statystycznych o charakterze wtórnym — co oznacza, że zgromadzono w nich wynikowe dane statystyczne z różnych badań dla potrzeb ich późniejszego, bieżącego udostępniania. Stanowią więc one swojego rodzaju hurtownię danych statystycznych o niskim poziomie automatyzacji funkcji udostępniania i niskim poziomie integracji informacyjnej. Brak im zintegrowanej metabazy, co powoduje konieczność każdorazowej integracji merytorycznej przy generowaniu raportów wielotematycznych. Do najistotniejszych baz tej grupy należą: Centralna Baza Danych o podmiotach gospodarczych, HANZA — baza danych handlu zagranicznego, PRAH — baza danych o prywatyzacji przedsiębiorstw państwowych. Zakłada się, że bazy te wejdą w skład hurtowni danych statystycznych.

W latach 2002—2005 przewiduje się budowę baz dziedzinowych, zawierających dane z następujących obszarów badań statystycznych: struktury gospodarstw rolnych — na podstawie PSR, regionalnych rachunków narodowych, produkcji wyrobów przemysłowych, wskaźników krótkoterminowych, cen, zasobów pracy, turystyki, transportu.

Zmodernizowane zostaną: baza DEMOGRAFIA, baza handlu zagranicznego w celu jej dostosowania do standardu INTRASTAT, obowiązującego w statystyce UE. Bazy te będą projektowane w sposób spójny z projektem WBO.

W wyniku uchwały Sejmu RP oraz decyzji wykonawczych Rady Ministrów podjęte zostały intensywne prace nad rozwojem społeczeństwa informacyjnego w Polsce. Prze-

widuje się, że w wyniku realizacji tych prac powinna powstać m.in. baza danych, charakteryzująca poziom rozwoju społeczeństwa informacyjnego i powinna to być baza publiczna, dostępna w Internecie.

Rejestry

Statystyka publiczna jest zobowiązana do prowadzenia, poza systemem badań statystyki publicznej, dwóch rejestrów urzędowych:

- rejestru podmiotów gospodarki narodowej REGON (Krajowy Rejestr Urzędowy Podmiotów Gospodarki Narodowej),
- rejestru podziału terytorialnego kraju TERYT.

Z punktu widzenia systemu informatycznego statystyki rejestry te są obecnie najbardziej złożonymi systemami informatycznymi, utrzymywanymi przez statystykę. Presja społeczna na sprawne działanie tych rejestrów, wynikająca z ich administracyjnego charakteru oraz wielka liczność ich baz danych i wielka dynamika zmian wymaga poświęcania na rozwój i utrzymanie obu rejestrów znacznych nakładów sił i środków z budżetu statystyki.

REGON

Urzędowy charakter rejestru REGON oznacza, że jego aktualizacja może być oparta wyłącznie na źródłach urzędowych (niestatystycznych). Rejestr zawiera informacje o ok. dwóch milionach podmiotów gospodarczych, prowadzących działalność nierolniczą oraz o ponad dwóch milionach indywidualnych gospodarstw rolnych.

Dynamika obsługi każdej z grup podmiotów jest zróżnicowana:

- opisy podmiotów działalności nierolniczej muszą być aktualizowane na podstawie ich wniosków w czasie rzeczywistym i dynamice wynikającej z ich zgłoszeń do urzędu statystycznego — czas rejestracji podmiotu lub aktualizacji istniejącego opisu, mimo dwutygodniowego okresu dopuszczalnego wprowadzenia wniosku do rejestru, staje się wtedy krytyczny ze względu na respektowane przez statystykę istniejące oczekiwania społeczne,
- opisy gospodarstw rolnych są przede wszystkim aktualizowane raz w roku na podstawie danych uzyskanych z gmin oraz w przypadku zgłoszeń ich właścicieli do urzędów statystycznych; zgłoszenia takie występowały sporadycznie; w końcu 2000 r. na skutek wprowadzenia VAT w rolnictwie ich intensywność nieco wzrosła.

REGON funkcjonuje na dwóch poziomach przetwarzania danych:

- urząd statystyczny, gdzie odbywa się rejestracja podmiotów gospodarczych i aktualizacja danych oraz udostępnianie informacji o zasięgu lokalnym,
- COIS, gdzie odbywa się weryfikacja zarejestrowanych w urzędzie statystycznym wniosków, prowadzenie bazy centralnej i udostępnianie informacji o zasięgu ogólnokrajowym.

Zmodernizowany w ostatnich latach system informatyczny zapewnia wymagany poziom jakości funkcjonowania rejestru. Wobec jego wielkiej liczności modernizacji wymaga natomiast baza infrastrukturalna systemu REGON. Zakres tej modernizacji obejmuje:

- zwiększenie mocy serwerów w urzędach statystycznych i centrali,
- zastosowanie nowszej wersji szbd SQL SERWER v. 7 o wyższej wydajności działania.

Ponadto zamierza się dokonać restrukturyzacji baz operacyjnych systemu REGON w urzędach statystycznych, polegającej na włączeniu baz operacyjnych dużych oddziałów do baz urzędów statystycznych, tak by zakresy jego baz operacyjnych obejmowały cały obszar województwa.

Skupienie baz wojewódzkich w siedzibie urzędów statystycznych umożliwi również podniesienie poziomu ochrony danych o podmiotach zawartych w tych bazach oraz, wobec projektowanej budowy łączy backupowych między centralą i siedzibami urzędów, zwiększy poziom niezawodności funkcjonowania systemu.

Zakłada się, że w perspektywie kilku lat funkcje REGON-u, jako rejestru urzędowego podmiotów gospodarczych, powinny zostać przejęte przez Krajowy Rejestr Sądowy — wobec czego statystyka nie powinna ponosić zbyt dużych nakładów na modernizację jego infrastruktury, ograniczając się do zapewnienia niezbędnego poziomu efektywności i niezawodności jego funkcjonowania.

TERYT

Rejestr TERYT ma również charakter urzędowy i obejmuje katalogi identyfikatorów i nazw jednostek administracyjnego i statystycznego podziału terytorialnego kraju, identyfikatorów i nazw miejscowości, identyfikatorów i nazw ulic oraz bazę danych opisujących budynki mieszkalne i mieszkania. Rejestr ten służy przede wszystkim do organizacji spisów powszechnych i reprezentacyjnych badań statystycznych z dziedziny statystyki społecznej. Katalogi tego rejestru, podobnie jak baza danych o mieszkaniach, są bardzo liczne. Liczebność poszczególnych katalogów rejestru jest następująca:

- rejonów statystycznych ok. 32 tys.,
- obwodów spisowych ok. 180 tys.,
- katalogu miejscowości ok. 100 tys.,
- katalogu ulic ok. 35 tys. nazw ulic i ok. 160 tys. wystąpień ulic w różnych miejscowościach,
- liczebność bazy danych mieszkaniowych wynosi ok. 12 mln pozycji.

TERYT funkcjonuje na dwóch poziomach w zakresie utrzymywania katalogów podziału statystycznego oraz katalogów ulic i miejscowości. W zakresie utrzymywania bazy mieszkaniowej stosuje się rozwiązanie mieszane:

- bazy danych o zasięgu lokalnym utrzymywane są w urzędach statystycznych i ich dużych oddziałach,
- udostępnianie informacji realizowane jest na poziomie regionalnym i centralnym poprzez łączenie zbiorów wojewódzkich w Łodzi, która, w odniesieniu do systemu informatycznego rejestru, pełni funkcję ośrodków projektowego i przetwarzania.

Przewiduje się — w ramach programu PHARE — modernizację systemu informatycznego rejestru w latach 2002—2003, obejmującą restrukturyzację baz wojewódzkich oraz zwiększenie wydajności pracy i poziomu automatyzacji realizacji funkcji rejestru poprzez modernizację infrastruktury sprzętowej i systemowej oraz zastosowanie nowych narzędzi do opracowania aplikacji.

Zasadniczą zmianą ma być zastosowanie technologii GIS do wprowadzenia i utrzymywania podziałów statystycznych. W efekcie tego działania spodziewana jest znaczna redukcja kosztów bieżącej aktualizacji map podziału kraju w stosunku do aktualnie utrzymywanego zasobu map papierowych.

Modernizacja realizowana będzie przez wykonawcę zewnętrznego z udziałem pracowników urzędu w Łodzi i tamtejszego zakładu COIS. Nowy system rejestru eksploatowany będzie przez zakład COIS w Łodzi.

Baza Jednostek Statystycznych

Baza Jednostek Statystycznych (BJS) jest rejestrem statystycznym podmiotów gospodarki narodowej. Podstawowym źródłem jej zasilania jest rejestr urzędowy REGON.

Źródłami uzupełniającymi, weryfikującymi zawartość rejestru, będą źródła administracyjne: rejestry publiczne, system podatkowy i system ubezpieczeniowy oraz badania statystyczne.

Obecny stan BJS wykorzystuje dane z REGON i wybranych badań statystycznych — ma on charakter prototypu zarówno w warstwie aplikacji, jak i bazy sprzętowo-systemowej.

Rozwój BJS dotyczył będzie wdrożenia zasilania bazy danymi z pozostałych badań statystycznych, a równolegle, w miarę możliwości pozyskiwania danych z rejestrów administracyjnych oraz systemu podatkowego i ubezpieczeniowego, także z tych źródeł. Równocześnie rozwijana będzie funkcjonalność bazy w module generowania kartotek do badań. BJS będzie jednym z podstawowych modułów WBO.

BJS będzie eksploatowana na serwerze o odpowiedniej wydajności w systemie wielo-dostępnym, umożliwiającym równoczesną pracę wielu użytkowników z GUS i urzędów statystycznych.

Geograficzne systemy informatyczne w statystyce

Celem budowy geograficznego systemu informacyjnego (GIS) w statystyce jest prezentacja danych statystycznych i ich udostępnianie w postaci ilustracji kartograficznych. Stosowanie GIS umożliwi integrację działań przy przeprowadzaniu badań statystycznych oraz współpracę z innymi organami administracji publicznej i organizacjami międzynarodowymi w obszarze systemów informacji przestrzennej. GIS jako narzędzie wspomagające statystykę może być użyte na każdym etapie przeprowadzania badania statystycznego.

Obecnie statystyka dysponuje pojedynczymi stanowiskami komputerowymi wyposażonymi w narzędzia do prezentacji danych statystycznych, a ich użycie sprowadza się do tworzenia mapek tematycznych wzbogacających publikacje urzędu. Brak jest bazy geograficznej na potrzeby statystyki, umożliwiającej analizę przestrzenną i jednolitą prezentację danych statystycznych na mapie.

W ramach budowy GIS w statystyce publicznej należy zorganizować infrastrukturę potrzebną do zbierania, zarządzania i używania wspólnych danych oraz dokonać standaryzacji tych działań. W tym celu został opracowany projekt GISTAT, który zakłada integrację danych geograficznych i statystycznych w jednym środowisku i umożliwia użytkownikom (dzięki systemowi informatycznemu) korzystanie poprzez sieć ze wspólnych zasobów bazy.

Projekt systemu GIS w statystyce przewiduje:

- ustalenie zespołów funkcjonalnych użytkowników systemu,
- budowę centralnej bazy i baz tematycznych,
- udostępnianie wybranych zasobów bazy użytkownikom zewnętrznym przez Internet,
- udział w budowie europejskiej sieci „GIS dla statystyków”.

Prace nad zastosowaniem GIS zintensyfikowane zostały w roku 2000, kiedy opracowana została koncepcja systemu. Będą one kontynuowane w ramach projektu rozwoju statystyki regionalnej finansowanej z programu PHARE'2000. Przewiduje się, że pierwszy etap wdrożenia systemu GIS, obejmujący zastosowanie map cyfrowych do prezentacji danych w rejestrze TERYT, bazie BDR oraz wybranych badań statystycznych, zostanie zakończony w 2003 r.

W celu zasilania bazy informacji geograficznej statystyki nawiązana została współpraca z Głównym Urzędem Geodezji i Kartografii, organizującym międzyresortowy zespół roboczy, który pracował będzie nad rozwojem systemów informacji przestrzennej

kraju oraz z GISCO, wydziałem Eurostatu zarządzającym europejską bazą geograficzną, instytucjami odpowiedzialnymi za dostarczanie informacji geograficznej i jej międzynarodową wymianę.

* * *

Przedstawiony program rozwoju systemu informatycznego statystyki został zatwierdzony przez Kolegium GUS w czerwcu br. Jest on kontynuacją strategii przyjętej w 1996 r. w opracowaniu „Kierunki modernizacji bazy informatycznej w sieci GUS na lata 1995—2000”¹⁾ i rozwinięciem przygotowanej w zeszłym roku „Strategii rozwoju systemu informatycznego statystyki w latach 2000—2005”, opartej na ustaleniach kierunkowych dokumentu o charakterze roboczym „Program rozwoju statystyki publicznej”.

Materiał przedstawia wizję rozwoju informatyki statystycznej w warstwie organizacyjnej i użytkowej, odnosząc się do potrzeb statystyki i jej struktury organizacyjnej i funkcjonalnej. Ta warstwa pojęciowa jest zrozumiała dla statystyków i informatyków, może więc być podstawą działania wszystkich grup pracowników statystyki.

Skrótowa postać koncepcji umożliwi efektywną dyskusję nad sposobami jej realizacji, a więc planowania i organizacji prac wykonawczych oraz ich finansowania.

Rozwinięcie koncepcji poszczególnych systemów informatycznych, które zasygnalizowane zostały w opracowaniu, znajduje się w materiałach szczegółowych opisujących te systemy.

Zakłada się, że przedstawiony program będzie corocznie aktualizowany w wyniku zmian czynników zewnętrznych, ponieważ:

- w ciągu roku mogą zmienić się, postrzegane jako najbardziej istotne lub pilne, potrzeby statystyki,
- zmienia się dostępność nowych technologii informatycznych, w jej aspekcie funkcjonalnym i ekonomicznym,
- w cyklu rocznym ustalana jest wielkość funduszy na rozwój informatyki — co przesądza o podjęciu poszczególnych prac rozwojowych.

Zakłada się, że realizacja programu rozwoju odbywać się będzie poprzez opracowanie i realizację kolejnych planów rocznych. Plan roczny opracowywany będzie po ustaleniu budżetu na dany rok. Do planu włączane będą zadania programu przypadające na dany rok i znajdujące pokrycie w zaplanowanych na ten cel środkach budżetowych.

Warunkiem skutecznej realizacji podjętych programów jest zapewnienie stabilności zatrudnienia i utrzymania prorozwojowej postawy kadry informatycznej — co z kolei uzależnione jest od możliwości zapewnienia tej kadrze wynagrodzeń zbliżonych do poziomu takiego, jak w otoczeniu statystyki²⁾.

dr inż. Józef Piskorz — wiceprezes GUS

¹⁾ Patrz „Wiadomości Statystyczne” nr 7 (446) z 1998 r.

²⁾ W następnym wydaniu „WS” zamieścimy artykuł Józefa Piskorza, traktujący o rozwoju infrastruktury informatycznej, narzędziach informatycznych wspomagających organizację pracy statystyków i systemie szkolenia oraz bezpieczeństwa statystyki.

SUMMARY

GUS Council in June 2001 approved the Programme of Development of Computing System of Statistics. This article presents a vision of development of statistical computing structure on the organizational and using level, considering the needs of statistics and present organizational and functional structure.

It is necessary to update this Programme each year in order to indicate an influence of changes of outside terms: needs of statistics, development of new computing technologies, budget possibilities. Execution of the Programme will be conducted by elaboration and realization of following annual plans. Activities for the year that can be financed within the planned budget financial means will be included into the Programme.

Author describes general developing activities in statistical computing system during previous years, targets and directions of development of statistical computing system and characterizes detailed future plans concerning the development of basic functional systems of statistics: statistical data collection including OCR technique and providing statistical interviewers with data registering device, collection of data via Internet; statistical data processing including creation of integrated system of statistical databases and data dissemination covering dissemination of data via Internet.

Moreover this article presents future plans concerning the specialistic systems of statistics including computing system of censuses, computing systems of REGON and TERYT administrative registers and statistical register of national business entities BJS (Statistical Units Database), systems of public databases, geographical information systems.

РЕЗЮМЕ

Обсуждаемая программа развития информатической системы статистики была утверждена Коллегией ЦСУ в июне с.г. Статья представляет картину развития статистической информатики в организационном и прикладном планах в соотношении с потребностями статистики и ее настоящей организационной и функциональной структурой.

Предполагается, что программа будет каждый год актуализироваться с целью учесть влияние изменений внешних факторов: нужд статистики, развития информатических технологий, финансовых возможностей. Программа развития будет реализовываться путем разработки и выполнения очередных годовых планов. План будет учитывать предусмотренные на данный год задания программы, на которые отведены бюджетные средства.

Автором обсуждаются общие действия по развитию статистической информатики за последние годы и, следовательно, цели и направления развития статистической информатики. Более подробно характеризуются планы по развитию основных функциональных систем статистики: сбора статистических данных, в т.ч. применения техники оптического чтения документов и оснащения статистических анкетеров регистраторами данных, сбора данных по Интернету; обработки статистических данных, в т.ч. создания интегрированной системы баз статистических данных а также распространения результатов обследований, в т.ч. с использованием Интернета.

Кроме того в статье представляются направления развития специальных систем статистики, к которым относится м.пр. информатическая система переписей, информатические системы официальных регистров REGON и TERYT а также статистического регистра субъектов народного хозяйства BJS, системы публичных баз, географические информационные системы.