

Biblioteka

WIADOMOŚCI
STATYSTYCZNYCH

Tadeusz Walczak

**OCHRONA DANYCH
W SYSTEMIE
INFORMACYJNYM
STATYSTYKI PUBLICZNEJ**

Tom 50

WARSZAWA 1996



GŁÓWNY URZĄD STATYSTYCZNY

PRZEDMOWA

OCHRONA DANYCH W SYSTEMIE INFORMACYJNYM STATYSTYKI PUBLICZNEJ

Warszawa, 1982

Kolegium Redakcyjne
Biblioteki Wiadomości Statystycznych

Rada Programowa Wiadomości Statystycznych
dr Stanisław Róg (przewodniczący), członkowie:
dr Stanisław Bartczak, *prof. dr hab. Czesław Domański*,
prof. dr hab. Jan Kordos, *mgr Tadeusz Persz*,
mgr Ryszard Wiśniewski, *prof. dr hab. Kazimierz Zajac*,
Stanisław Jońca (sekretarz)

Opracowanie redakcyjne
Stanisław Jońca

Redakcja techniczna
Zofia Wapniarek

Korekta
Teresa Chmielewska
Urszula Kazmierska
Izabela Krawczyk
Halina Kwiatkowska



Przedruk w całości lub w części oraz wykorzystanie danych
statystycznych w druku dozwolone wyłącznie z podaniem źródła



Zakład Wydawnictw Statystycznych, 00-925 Warszawa,
al. Niepodległości 208. Nakład 500 egz, ark. druk. 7,875. Papier kl. III, 70 g.
Oddano do składania w sierpniu 1996 r. Druk ukończono w listopadzie 1996 r.
Cena zł 11,00 (110 000,—).
Druk. ZWS, Warszawa. Zam. 284.
Informacje w sprawach zbytu publikacji — tel. 608-32-10.

PRZEDMOWA

Niniejszy 50 tom Biblioteki Wiadomości Statystycznych poświęcony jest problematyce ochrony danych osobowych i indywidualnych w statystyce publicznej oraz środkom i metodom zabezpieczającym tę ochronę.

Obowiązująca od 1 listopada 1995 r. ustawa o statystyce publicznej z 29 czerwca 1995 r. — w której zawarto nowe spojrzenie na problemy ochrony danych osobowych w statystyce, oparte na standardach międzynarodowych — stanowi, że dane osobowe i dane indywidualne uzyskiwane w spisach i innych badaniach statystycznych są chronione tajemnicą statystyczną. Oznacza to, że dane te mogą być wykorzystywane wyłącznie do opracowań, zestawień i analiz statystycznych. Udostępnianie lub wykorzystywanie danych indywidualnych i danych osobowych dla innych celów jest zabronione. Pracownicy służb statystyki publicznej oraz inne osoby wykonujące czynności na zlecenie lub na rzecz statystyki publicznej są zobowiązane do bezwzględnego przestrzegania tajemnicy statystycznej. Działaniu przeciw ochronie danych osobowych i indywidualnych oraz wykorzystywaniu danych statystycznych w celu osiągnięcia korzyści majątkowej lub osobistej (np. wykorzystywanie informacji indywidualnych o sytuacji spółek giełdowych) nadano rangę przestępstwa karnego.

Prezentowane opracowanie pt. *Ochrona danych w systemie informacyjnym statystyki publicznej* przedstawia źródła, przyczyny i miejsca powstawania zagrożeń utraty warunków do zachowania tajemnicy statystycznej, w tym w szczególności niebezpieczeństwa związane z warunkami pracy komputerów w sieciach i wielodostępem w procesie zbierania, opracowania danych i udostępniania informacji. Omówione zostały również niektóre metody zapobiegania identyfikacji i ujawniania danych źródłowych i danych w informacjach zagregowanych oraz podano opis próby automatyzacji procedur ochrony informacji przed ujawnieniem.

PREZES GŁÓWNEGO
URZĘDU STATYSTYCZNEGO

Tadeusz Toczyński

SPIS TREŚCI

AWOMONIA

	Str.
Przedmowa	3
Wprowadzenie	6
1. Potrzeba ochrony informacji w systemach informacyjnych	9
2. Definicje ważniejszych pojęć dotyczących ochrony danych	15
3. Prawne, etyczne i pragmatyczne przesłanki ochrony danych	18
3.1. Konwencje i Zalecenia Rady Europy	18
3.1.1. Zalecenie Rady Europy w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych	22
3.2. Podstawowe zasady statystyki oficjalnej i kodeks etyczny statystyków	26
3.3. Polska ustawa o statystyce publicznej	28
3.4. Przekazywanie poufnych danych do Eurostatu	35
3.5. Zapewnienie ochrony danych jako warunek uzyskiwania prawdziwych informacji	36
4. Miejsca i przyczyny zagrożeń dla bezpieczeństwa danych	43
4.1. Ważniejsze etapy badań i opracowań statystycznych	43
4.2. Zagrożenia dla bezpieczeństwa danych na wstępnych etapach realizacji badań	44
5. Ochrona danych na etapie rozpowszechniania informacji	51
5.1. Obowiązek ochrony danych a realizacja powszechnego dostępu do informacji	51
5.2. Wykrywanie zagrożeń identyfikacji w zbiorach danych	58
5.2.1. Naruszenie zasady wielkości progowej	60
5.2.2. Naruszenie zasady przeważającego udziału	61
5.2.3. Ujawnienie danych metodą odejmowania	62
5.2.4. Czy można całkowicie zapobiec ujawnieniu danych jednostkowych?	63

	Str.
6. Metody zapobiegania identyfikacji i ujawnieniu danych	65
6.1. Metody zapobiegania identyfikacji w przypadku udostępniania danych źródłowych (mikrodanych)	65
6.2. Redukcja danych	69
6.2.1. Usuwanie identyfikatorów bezpośrednich	69
6.2.2. Udostępnianie próby	70
6.2.3. Ograniczenie szczegółowości danych	70
6.2.4. Zmiana grupowania danych	71
6.2.5. Usuwanie	72
6.2.6. Modyfikacja danych	72
6.3. Metody zapobiegania identyfikacji danych w informacjach zagregowanych (w makrodanych)	78
6.3.1. Metody ograniczające szczegółowość informacji	82
6.3.2. Metody wprowadzające zakłócenia	86
6.3.3. Metody wprowadzania zakłóceń losowych do informacji zagregowanych	89
7. Próby automatyzacji procedur ochrony informacji przed ujawnieniem	91
8. Podsumowanie i wnioski	94
Literatura	98
Aneks	102
Konwencja (Rady Europy) z 28 stycznia 1981 r. dotycząca ochrony osób w związku z automatycznym przetwarzaniem danych osobowych, podpisana w Strasburgu	102
Zalecenie w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych	113

WPROWADZENIE

Od początku lat dziewięćdziesiątych rozpoczął się w Polsce szczególnie intensywny proces zmian w statystyce. Zmiany te, nazwane później *procesem transformacji statystyki*, objęły niemal wszystkie elementy tego systemu informacyjnego.

Zmiany w statystyce były konsekwencją dokonujących się w naszym kraju przekształceń ustrojowych, przy czym w odniesieniu do statystyki wynikały one przede wszystkim:

- z nowych potrzeb organów państwowych i demokratycznego społeczeństwa,
- z konieczności dostosowania systemu zbierania, opracowania i udostępniania informacji do potrzeb gospodarki rynkowej,
- z konieczności dostosowania metodologii badań statystycznych i zasad funkcjonowania systemu informacyjnego do standardów międzynarodowych.

Prawnym ukoronowaniem przekształceń, jakich dokonano w statystyce od początku lat dziewięćdziesiątych, było przyjęcie nowej ustawy o statystyce publicznej.¹⁾ Przepisy ustawy stworzyły jednocześnie prawne gwarancje trwałości wprowadzonych zmian.

Podstawowe regulacje prawne nowej ustawy dotyczą zwłaszcza:

- 1) roli statystyki w kraju o gospodarce rynkowej oraz w warunkach demokratycznego społeczeństwa;
- 2) zawodowej niezależności służb statystyki publicznej w dziedzinie prowadzenia badań statystycznych;
- 3) ochrony danych osobowych i indywidualnych zbieranych w procesie badań i wykorzystywania ich wyłącznie do celów statystycznych w formie uniemożliwiającej identyfikację danych odnoszących się do indywidualnych osób, jednostek gospodarczych lub innych respondentów.

Ten ostatni aspekt zasługuje na szczególne podkreślenie, stanowi on bowiem fundament, od trwałości którego zależy kompletność i rzetelność statystyki. Ścisłe przestrzeganie zasad ochrony informacji indywidualnych stanowi również przejaw szacunku ze strony statystyków dla poszanowania

¹⁾ Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej, Dz. U. Nr 58, poz. 439.

praw i wolności obywatelskich określonych w wielu konwencjach i zaleceniach międzynarodowych.

Przyjęte w polskiej ustawie o statystyce publicznej rozstrzygnięcia prawne dotyczące ochrony tajemnicy statystycznej są zgodne z obowiązującymi w tej dziedzinie standardami międzynarodowymi i odpowiadają tendencjom światowym w odniesieniu do upowszechniania informacji z zachowaniem praw jednostki do ochrony jej dóbr osobistych i prywatności.

Jak wynika jednak z praktyki wielu krajów, w których obowiązują podobne zasady przestrzegania tajemnicy statystycznej, jak również nasze pierwsze doświadczenia pracy służb statystycznych w warunkach nowej ustawy, w praktycznej realizacji przepisów o tajemnicy statystycznej i obowiązkach ochrony informacji statystycznej występuje szereg trudności, do rozwiązania których konieczne jest podejmowanie całego szeregu przedsięwzięć zapewniających zarówno pełną ochronę danych indywidualnych, jak i powszechny dostęp do informacji.

W przedstawianej Czytelnikom pracy autor podjął próbę omówienia najważniejszych problemów ochrony danych w statystyce publicznej, akcentując zwłaszcza te sprawy, które w świetle aktualnego rozwoju badań statystycznych i obowiązków służb statystycznych, w dziedzinie udostępniania informacji mają, zdaniem autora, najistotniejsze znaczenie.

Praca składa się z 8 rozdziałów i aneksu.

W rozdziale pierwszym akcentuje się znaczenie zapewnienia właściwego systemu ochrony wszystkich części składowych systemu informacyjnego, zwłaszcza tak obszernego, różnorodnego i mającego niezwykle szeroki rezonans społeczny systemu, jakim jest system informacyjny statystyki.

W rozdziale drugim przytoczono krótkie omówienie ważniejszych pojęć dotyczących ochrony danych. Wyjaśnienie tych pojęć ma istotne znaczenie dla jednolitego rozumienia treści całej pracy.

Rozdział trzeci poświęcony jest omówieniu prawnych, etycznych i pragmatycznych przesłanek ochrony danych. W rozdziale tym przedstawiono obowiązujące **zalecenia Rady Europy** dotyczące ochrony danych osobowych, **podstawowe zasady statystyki oficjalnej** oraz **zasady etyczne** sformułowane w kodeksie etyki zawodowej statystyków. Przedstawiono w nim również omówienie aspektów ochrony danych w świetle przepisów polskiej ustawy o statystyce publicznej. W rozdziale tym omówiono również konieczność przestrzegania zasad ochrony informacji, jako niezbędnego warunku zapewnienia sobie przez statystykę pełnego zaufania respondentów; tylko bowiem w takich warunkach statystycy mogą liczyć na uzyskanie zgody respondentów na uczestnictwo w badaniach oraz na uzyskanie pełnych i rzetelnych informacji. Tę przesłankę kierującą statystykami przy rozwiązywaniu zadań ochrony informacji autor nazywa **przesłanką pragmatyczną**.

W rozdziale czwartym omawia się najbardziej zagrożone miejsca oraz przyczyny powstawania zagrożeń dla bezpieczeństwa danych. W rozdziale tym omówiono podstawowe etapy badań i opracowań statystycznych ze

szczególным uwzględnieniem zagrożeń dla bezpieczeństwa danych, jakie występują na wstępnych etapach realizacji badań, zwłaszcza na etapie uzyskiwania i rejestracji danych źródłowych oraz na etapie kodowania cech (symbolizacji), tworzenia maszynowych nośników, przesyłania i kontroli danych.

Rozdział piąty poświęcony jest omówieniu zagrożeń ujawnienia informacji na etapie jej rozpowszechniania. Akcentuje się konieczność podejmowania przez statystyków wysiłków zapewniających pełną ochronę danych, ale bez nadmiernego ograniczania dostępu użytkownikom do informacji. Omówiono także podstawowe metody wykrywania zagrożeń ujawnienia danych w działalności informacyjnej służb statystycznych.

W rozdziale szóstym omówiono metody zapobiegania identyfikacji danych zarówno w przypadku udostępniania zbiorów źródłowych (tzw. mikro danych), jak i wyników opracowań (makro danych).

Ze względu na to, że żadna z tych metod nie była dotąd szerzej w Polsce stosowana oraz że **głównym celem pracy jest ogólne wprowadzenie do tego obszernego i jednocześnie wywołującego zastrzeżenia i wątpliwości zagadnienia**, autor ograniczył się jedynie do wyjaśnienia na czym polegają poszczególne metody, pomijając całe, nieraz dość złożone uzasadnienie matematyczne towarzyszące ich stosowaniu. Zagadnienia te, zdaniem autora, **powinny stanowić przedmiot dalszych badań opartych na doświadczeniach wynikających z praktycznego zastosowania każdej z tych metod w naszej praktyce statystycznej.**

W rozdziale siódmym przedstawiono krótkie omówienie prób automatyzacji kontroli danych przed niebezpieczeństwem ujawnienia danych podlegających ochronie oraz przed udostępnieniem, jak również komputeryzacji procedur zapobiegających ujawnieniu.

W rozdziale ósmym przedstawiono krótkie podsumowanie oraz niektóre wnioski wynikające, zdaniem autora, z treści pracy.

W aneksie zamieszczono mało w Polsce znany tekst **Konwencji Rady Europy** z 28 stycznia 1981 r. (tzw. Konwencji 108) dotyczącej ochrony osób w związku z automatycznym przetwarzaniem danych osobowych oraz tekst **Zalecenia Rady Europy** w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych. Nie jest to wprawdzie tekst ostateczny. Nie został on jeszcze przyjęty przez Komitet Ministrów Rady Europy, ale z uwagi na jego bardzo dużą aktualność dla bieżącej polityki informacyjnej służb statystycznych oraz na duże prawdopodobieństwo, że ostateczny tekst nie będzie odbiegał w istotny sposób od projektu z czerwca 1996, zapoznanie się z nim autor uznał za uzasadnione.

Autor przyjmie z wdzięcznością od Czytelników wszelkie uwagi i sugestie zmierzające do zmian i uzupełnień, które mogłyby wykorzystać w ewentualnej dalszej pracy nad tym tematem.

Autor

I. POTRZEBA OCHRONY INFORMACJI W SYSTEMACH INFORMACYJNYCH

Na temat informacji, jej roli i zadań we współczesnym społeczeństwie napisano i powiedziano już bardzo dużo. Często drugą połowę XX wieku nazywa się erą informacji.

Jeśli posłużyć się dla ilustracji tego stwierdzenia przykładem periodyzacji etapów rozwoju gospodarki i społeczeństwa amerykańskiego, to tam przyjmuje się, iż gospodarka i społeczeństwo przebyły dotąd 4 etapy rozwoju: **pierwszy etap** trwający od czasów kolonialnych do końca XIX w. — to etap rozwoju rolnictwa. Większość ludności pracowała na roli i rolnictwo stanowiło podstawę gospodarki.

Mniej więcej od początku XX wieku do końca drugiej wojny światowej, gospodarka wkroczyła w **drugi etap rozwoju** — okres rozwoju przemysłowego. To przemysł, zwłaszcza przemysł ciężki, stanowił siłę napędową całej gospodarki.

Po drugiej wojnie światowej, do połowy lat pięćdziesiątych, gospodarkę amerykańską zdominował sektor usług. W tym okresie, **trzecim etapie rozwoju**, ponad połowa aktywnej zawodowo ludności znalazła zatrudnienie właśnie w tym sektorze.

Pod koniec lat pięćdziesiątych gospodarka i społeczeństwo amerykańskie wkroczyły w kolejny — **czwarty etap rozwoju**. We wszystkich dziedzinach życia gospodarczego i społecznego coraz silniej odczuwa się oddziaływanie procesów informacyjnych: coraz to większą rolę odgrywają czynności zbierania, analizy i rozpowszechniania informacji i wiedzy.²⁾ Zauważa się wprawdzie, że czynności związane ze zbieraniem i przetwarzaniem informacji są elementem sektora usług, niemniej jednak specjaliści uważają, że działalność informacyjną należy traktować jako odrębny, czwarty etap rozwoju amerykańskiej gospodarki [Dun95, s. 5].

Trzeba jednakże pamiętać, że informacja nie stanowi wartości sama w sobie. Nie ma właściwości wyrażającej się w tym, że samo jej posiadanie czyni jej właściciela bogatszym. Wartość informacji lub, inaczej mówiąc, korzyść z jej posiadania zależy od umiejętności jej wykorzystania.

²⁾ Zgromadzona i przechowywana informacja często nazywana jest wiedzą. Wiedza — to zakumulowane informacje, a sama informacja — to przyrost wiedzy. Por. [Fla89, s. 105].

Informacja daje jej posiadaczowi potencjalną przewagę nad partnerem lub konkurentem, który tej informacji nie posiada. Czy przewaga ta zostanie rzeczywiście wykorzystana i z jakim skutkiem — to zależy nie tylko od samej informacji. Dlatego m.in. tak trudno oszacować rzeczywistą wartość informacji.

Informacja, a zwłaszcza jej prawidłowe wykorzystanie wywiera istotny wpływ nie tylko na rozwój gospodarki, wyrażający się we wskaźnikach wzrostu produktu krajowego brutto (PKB), ale i na cały szereg elementów rozwoju społecznego, takich jak: wzrost poziomu wiedzy, oświaty, uczestnictwa w kulturze, sytuacji jednostki w społeczeństwie, stylu życia itp.

Zdaniem amerykańskiego autora Marca Porata, informacje i usługi informacyjne stanowiły w 1967 r. 25% wartości produktu narodowego brutto USA [Dun95, s. 7].³⁾

Rzeczony rozwój działalności informacyjnej i umożliwienie szerokiego dostępu do informacji traktuje się również jako niezwykle ważny czynnik rozwoju społecznego, jako warunek rozwoju społeczeństwa w kierunku demokracji [Mar70, s. 320]. Nie można jednak nie zauważyć również, że w miarę rozpowszechniania się działalności związanej ze zbieraniem, przechowywaniem i zwielokrotnieniem się możliwości rozpowszechnienia informacji, w wyniku rozwoju baz i banków danych, rośnie również zaniepokojenie w związku z potencjalnym zagrożeniem nadużyć, powodowanych niewłaściwym wykorzystaniem informacji indywidualnych z uszczerbkiem dla jednostek, które te informacje dostarczyły w dobrej wierze.

Mniej więcej od początku lat pięćdziesiątych, a zwłaszcza po upowszechnieniu się komputerów o dużej mocy obliczeniowej i ogromnych możliwościach „pamiętania” nieograniczonych wprost ilości danych, zaczęły powstawać liczne systemy informacyjne wyspecjalizowane w zbieraniu, gromadzeniu, przechowywaniu, opracowywaniu i udostępnianiu informacji z określonych dziedzin. Jeśli na tego rodzaju systemy spojrzeć od strony użytkownika, to jego systemy te interesują przede wszystkim jako źródło możliwej do uzyskania informacji, która powinna być dokładna, wewnętrznie spójna, terminowa i dostosowana do jego specyficznych potrzeb.

Inne spojrzenie na system informacyjny musi obowiązywać organizację lub instytucję odpowiedzialną za projektowanie i bieżące funkcjonowanie systemu (gospodarza systemu informacyjnego). Gospodarz systemu informacyjnego musi nie tylko zaspokoić bieżące i potencjalne potrzeby użytkowników informacji, ale również zagwarantować bezawaryjne funkcjonowanie systemu i uchronienie go od najróżniejszych zagrożeń występujących w każdym systemie informacyjnym.

³⁾ Ten zaskakująco wysoki udział wartości usług informacyjnych w PNB USA zasługiwałby na dokładniejszy komentarz, którego cytowani autorzy nie podają. Chodzi prawdopodobnie o wszelkie usługi informacyjne, a więc usługi w dziedzinie informacji naukowo-technicznej, informacji finansowej, statystycznej, reklamy i marketingu itp. (T.W.)

W każdym systemie informacyjnym mamy do czynienia z nagromadzeniem bardzo cennych zasobów, których odtworzenie w przypadku uszkodzenia lub zniszczenia byłoby niezwykle kosztowne, a nierzadko — wręcz niemożliwe. Aspekt ten jest szczególnie ważny w przypadku tak obszernego i wszechstronnego systemu, jak **system informacyjny statystyki**.⁴⁾

Rozpatrując system informacyjny statystyki z punktu widzenia nagromadzonych w nim zasobów, można wyróżnić w nim zwłaszcza:

- **Dane źródłowe** uzyskane w wyniku badań własnych oraz w wyniku współpracy z innymi systemami informacyjnymi. Wszystkie te dane przechowywane są w różnych postaciach maszynowych nośników, tj. w komputerach, na taśmach i dyskach magnetycznych, na dyskietkach, a część z nich również na tradycyjnych nośnikach papierowych, tj. na sprawozdaniach, kwestionariuszach, ankietach, tablicach itp. Dane źródłowe ze względu na swoją masowość stanowią najbardziej kosztowny element zasobów informacyjnych systemu informacji statystycznej. Łączny koszt zebrania, kontroli i przeniesienia na maszynowe nośniki danych źródłowych można szacować na około 80% całości kosztów związanych ze zbieraniem i opracowaniem informacji statystycznej.

- **Dokumentację** niezbędną do każdego badania, w tym szczegółowe zasady metodologiczne poszczególnych badań — instrumentarium badań, zasady kontroli poprawności danych, algorytmy obliczania wskaźników pochodnych, tworzenia tablic wyników itp.

- **Komputerowe programy** przetwarzania i analiz statystycznych wraz z dokumentacją.

- **Rejestry** i inne wykazy jednostek, wykorzystywane zarówno w procesie przetwarzania danych, jak i w charakterze operatów losowania w badaniach reprezentacyjnych.

- **Klasyfikacje, nomenklatury i inne elementy metainformacji**. Zasoby te w coraz to większym zakresie utrzymywane są na nośnikach elektronicznych.

- **Sprzęt komputerowy, systemy łączności**, wyposażenie techniczne i biuro-
we środowiska, w którym funkcjonuje system informacyjny.

Warunkiem prawidłowego i bezawaryjnego funkcjonowania systemu informacji jest zapewnienie integralności i bezpieczeństwa wszystkich wymienionych zasobów systemu. Dopuszczenie do zniszczenia lub uszkodzenia któregokolwiek z wymienionych elementów prowadzi zawsze do zakłóceń w funkcjonowaniu systemu, może także doprowadzić do poważnych, a nawet do nie powetowanych strat. Wyobraźmy sobie, przykładowo, iż na skutek zaniedbań, nieszczęśliwego wypadku czy klęski żywiołowej uległyby zniszczeniu materiały masowego badania w rodzaju np. spisu

⁴⁾ Mówiąc o systemie informacyjnym statystyki mam na myśli wszystkie ogniwa organizacyjne tego systemu, tzn. służby statystyki publicznej wraz z ich zapleczem technicznym, informatycznym, telekomunikacyjnym, badawczym i obsługowym.

rolniczego, którego samo przygotowanie i przeprowadzenie kosztuje kilkadziesiąt milionów (nowych) złotych. Trudno sobie wyobrazić, aby było możliwe powtórne przeprowadzenie podobnego badania, uwzględniając nie tylko same koszty w wyrażeniu pieniężnym, ale również ogromny wysiłek tysięcy pracowników organizujących i przeprowadzających badanie oraz czas stracony przez respondentów na udzielanie odpowiedzi w czasie przeprowadzania badania.

Podobnie, trudno sobie wyobrazić zakłócenia i straty, jakie mogłyby spowodować zniszczenie lub zniekształcenie oprogramowania komputerowego, od którego zależy obecnie w decydującym stopniu poprawność i terminowość opracowań statystycznych i zachowanie zbiorów informacyjnych.

W każdym systemie informacyjnym musi dlatego obowiązywać cały szereg zabezpieczeń i ochrony wszystkich elementów systemu — we wszystkich cyklach jego życia.

Niestety, nie zawsze problemy te są doceniane. Można zaobserwować wiele przykładów funkcjonowania systemów informacyjnych na granicy ogromnego ryzyka i często tylko dzięki szczęśliwemu zbiegowi okoliczności udaje się uniknąć większych zakłóceń i strat.

Sygnalizowany wyżej problem konieczności ochrony zasobów systemu informacyjnego odnosi się do każdego systemu. W przypadku statystyki występuje niezwykle istotny dodatkowy element, a mianowicie **konieczność ochrony danych indywidualnych i osobowych** przed ich (świadomym lub nieświadomym) ujawnieniem na jakimkolwiek etapie zbierania, opracowania, przechowywania i udostępniania informacji.

Zagadnienia bezpieczeństwa i ochrony systemów informacyjnych nie są zagadnieniami nowymi. Począwszy od początku lat siedemdziesiątych ukazuje się bardzo wiele publikacji sygnalizujących najróżniejsze przypadki, zwłaszcza świadomego niszczenia informacji, a także przypadki tzw. szpiegostwa technologicznego oraz nadużyć finansowych popełnianych przez odpowiednie manipulowanie zbiorami danych i programami komputerowymi.

Zagadnienia ochrony danych stały się tematem pracy wielu naukowców, programistów i kierowników ośrodków obliczeniowych. Problemy te były i są nadal przedmiotem wielu konferencji i seminariów międzynarodowych.⁵⁾

Szczególnie dużo pisano na ten temat w latach siedemdziesiątych, tj. w okresie gwałtownego rozwoju komputeryzacji systemów informacyjnych. Jak podaje Antoni Bossowski, Instytut Badawczy Stanforda, zajmujący się zagadnieniami bezpieczeństwa i ochrony danych, zebrał do 1976 r. 148 różnych przypadków nadużyć komputerowych. Były wśród nich przypadki

⁵⁾ Na odnotowanie zasługują zwłaszcza 2 seminaria międzynarodowe zorganizowane przez Eurostat — pierwsze w Dublinie w 1992 r. oraz drugie w Luksemburgu w 1994 r.

poważnych defraudacji poprzez manipulowanie danymi gromadzonymi w komputerach, rozprowadzanie fałszywych polis ubezpieczeniowych itp. [Bos76].

Zainteresowanych Czytelników możemy odesłać również do cytowanej wyżej obszernej pracy Jamesa Martina i Adriana R. D. Normana [Mar70], w której omówiono wiele różnorodnych przypadków zarówno pomyłek, jak i oszustw, na które narażone są systemy informacyjne w warunkach funkcjonowania komputerów i które są szczególnie trudne do wykrycia, zwłaszcza jeśli są one dziełem wykwalifikowanych programistów, którzy wprowadzając odpowiednie modyfikacje do programów mogą przyczynić się do powstania mniejszych lub większych strat. Nadużycia komputerowe są również bardzo trudne do ścigania sądowego ze względu na to, że przestępca nie zostawia klasycznych śladów w postaci np. odcisków palców, własnoręcznie sfalszowanych dokumentów itp. Często nawet przestępca nie pozostawia śladu w postaci sfalszowanego programu, jeśli w porę zastąpi go programem poprawnym [Mar70, s. 369].

Sprawy zagrożeń związanych z upowszechnieniem się różnych systemów informacyjnych stały się ostatnio przedmiotem ożywionych dyskusji również w Polsce, zwłaszcza w związku z podjęciem prac nad projektem ustawy o ochronie danych osobowych. Wyrażane są obawy, że wobec ogromnej ilości dokumentów i kartotek zawierających dane osobowe⁶⁾, informacje w nich zawarte mogą być wykorzystane w celu zwiększenia ingerencji w ich sprawy osobiste ze szkodą dla obywateli, w celu zniesławienia, szantażu lub dyskredytacji.

Można przytoczyć również całkowicie świeży przykład przestępstwa komputerowego z naszej własnej polskiej praktyki. Jak donosi prasa pracownik Wielkopolskiego Banku Kredytowego, wykorzystując dostęp do sieci komputerowej, przelewał na założone przez siebie konta pieniądze, które za pośrednictwem WBK przekazywały różne firmy. Operacji tych pracownik dokonywał w piątki po południu, a w niedzieląki wszystko wracało na swoje miejsce. Tylko przez sobotę i niedzielę pieniądze procentowały na kontach oszusta. W ten sposób podejrzany wyprowadził z systemu bankowego 23 miliardy starych złotych.

W statystyce, z uwagi na inny rodzaj gromadzonych informacji, nie występuje z takim nasileniem groźba nadużyć popełnianych z chęci zysku, jak w przypadku informacji finansowych, w przedsiębiorstwach, w bankach itp., nie znaczy to jednak, że sprawy ochrony danych mają w statystyce mniejsze znaczenie. Przeciwnie, istnieje szereg powodów, dla których właśnie

⁶⁾ Można dla przykładu wymienić tylko: Powszechny Elektroniczny System Ewidencji Ludności (PESEL), księgi stanu cywilnego, dokumenty ewidencji ludności, ewidencję wydanych paszportów, ewidencję osób objętych powszechnym obowiązkiem wojskowym, ewidencję dzieci objętych obowiązkiem szkolnym, ewidencję kont bankowych, rejestry skazanych i rejestry osób ukaranych za wykroczenia, dokumentację medyczną, ewidencję wydanych i zatrzymanych praw jazdy, ewidencję osób objętych świadczeniami ZUS, rejestry policyjne, wykazy absolwentów szkół.

w odniesieniu do statystyki, bezpieczeństwo i ochrona mają szczególne znaczenie. Można zwrócić uwagę w szczególności na 3 takie powody:

po pierwsze, wielkość zasobów informacyjnych i związanych z nimi materiałów metodologicznych, zbiorów metadanych, oprogramowania itp., co powoduje, iż zaniedbania ochrony mogą doprowadzić do znacznych strat i zakłóceń w funkcjonowaniu systemu;

po drugie, w warunkach rozwijającej się żywiolowo działalności usługowo-konsultacyjnej zwiększa się zapotrzebowanie na różnego rodzaju informacje, z których część jest możliwa do uzyskania wyłącznie z systemu informacyjnego statystyki. Stwarza to potencjalne zagrożenie nielegalnego wykorzystywania dostępu do zasobów informacyjnych statystyki, w celu przechwytywania niektórych chronionych danych;

po trzecie, w przypadku statystyki obowiązują (o czym wspomiano wyżej) szczególne rygory zakazujące udostępniania informacji jednostkowych (indywidualnych i osobowych). Rygory te wynikają zarówno z międzynarodowych konwencji i zaleceń, z etycznych zasad obowiązujących statystyków, jak i z wyraźnych przepisów przyjętych w ustawie o statystyce publicznej z dnia 29 czerwca 1995 r.⁷⁾ Przestrzeganie rygorów w tej dziedzinie nie jest możliwe bez zapewnienia całego systemu ochrony.

W niniejszej pracy szczególnie dużo uwagi poświęcimy różnym aspektom bezpieczeństwa i ochrony danych indywidualnych i osobowych właśnie dlatego, że ten aspekt występuje ze szczególną ostrością w przypadku statystyki.

Ogólne problemy bezpieczeństwa danych oraz metody ochrony systemów informacyjnych — wspólne dla wszystkich systemów — zainteresowany Czytelnik może znaleźć w dość obszernej literaturze dostępnej również na naszym rynku wydawniczym.⁸⁾

⁷⁾ Dz. U. Nr 58, poz. 439.

⁸⁾ Patrz np. [Hof82], [Idz79], [Sok87].

2. DEFINICJE WAŻNIEJSZYCH POJĘĆ DOTYCZĄCYCH OCHRONY DANYCH

W omawianej pracy spotkamy się z szeregiem pojęć i określeń, które mogą nie być w pełni zrozumiałe dla Czytelnika, zwłaszcza że niektóre z tych pojęć mają często nieco inne znaczenie w potocznym języku. Z tego względu, dla poprawy komunikatywności pracy, zdefiniujemy niektóre ważniejsze pojęcia i określenia.

Bezpieczeństwo danych (informacji), bezpieczeństwo systemu informacyjnego. Pod tym pojęciem rozumiemy zespół warunków zapewniających wyeliminowanie lub maksymalne ograniczenie zagrożeń dla danych lub informacji przed przypadkowym lub zamierzonym zniszczeniem, ujawnieniem lub samowolną zmianą.

Bezpieczeństwo komputerowe oznacza wprowadzenie technologicznych, organizacyjnych, technicznych i administracyjnych procedur możliwych do zastosowania dla ochrony sprzętu, oprogramowania oraz zasobów informacyjnych.

Ochrona danych (informacji), ochrona systemu informacyjnego. Jest to pojęcie bliskoznaczne z pojęciem bezpieczeństwa. Ochrona oznacza zastosowanie przedsięwzięć, mających na celu zachowanie lub restytucję zasobów informacyjnych i materialnych systemu informacyjnego.

Prywatność, prawo do prywatności (*privacy, right to privacy*). Ściśle rzecz biorąc pojęcie to odnosi się do indywidualnych osób. Oznacza ono prawo do decydowania jakie informacje zainteresowana osoba jest skłonna udostępnić osobom trzecim oraz jakie informacje jest gotowa przyjąć od innych. Arwid Mednis, cytując A. Kopfa, zauważa, że w polskiej doktrynie definicja prywatności rozumiana jest dość szeroko, jako *prawo jednostki do życia własnym życiem układanym według własnej woli z ograniczeniem do niezbędnego minimum wszelkiej ingerencji zewnętrznej* [Med95]. Często jednak prywatność osób łączy się z pojęciem prywatności rodzin, gospodarstw domowych i nienaruszalności mieszkania.

Należy zauważyć, że przytoczona wyżej definicja prywatności wyraża subiektywny stosunek do prywatności poszczególnych osób, uzależniony od indywidualnych cech osobniczych, od rodzaju prowadzonej działalności, a także od warunków środowiska, tradycji, zwyczajów miejscowych itp. Przykładowo, niektóre osoby w określonych warunkach i sytuacjach mogą

uznać za informacje naruszające ich dobra osobiste, np. miejsce urodzenia, wiek czy stan cywilny. Z kolei inne osoby (np. dążące do uzyskania rozgłosu) mogą być zainteresowane w ujawnieniu i rozgłoszeniu nawet bardziej intymnych informacji. Dlatego tak trudno ustalić zakres cech osobowych, które nie wymagają ochrony. Z tego względu służby statystyczne, w trosce o pozyskanie pełnego zaufania opinii publicznej, muszą stosować bezwzględny zakaz ujawniania jakichkolwiek danych osobowych uzyskanych w toku badań statystycznych. Można zauważyć również, że zadaniem statystyki nie jest ochrona prywatności osób, lecz **zagwarantowanie poufności danych uzyskanych od tych osób** w procesie badań statystycznych lub otrzymanych z innych źródeł, np. ze źródeł administracyjnych.

Poufność. Pojęcie to odnosi się do danych i do informacji statystycznej. Poufność oznacza status, jaki został uzgodniony między podmiotem przekazującym dane i instytucją, organizacją lub osobą, które dane te otrzymują. Stopień poufności określa również zakres niezbędnej do zastosowania ochrony.

Integralność oznacza pełną zgodność danych znajdujących się w zasobach informacyjnych z zapisami w dokumentach źródłowych. O naruszeniu integralności danych mówimy, gdy dane te zostały zmienione lub zniszczone w sposób przypadkowy lub celowy.

Dane osobowe są to dane o osobach fizycznych, charakteryzujące ich cechy demograficzne, zawodowe, sytuację materialną, stan zdrowia, poglądy, przyzwyczajenia itp. Dane te pochodzą głównie z badań demograficznych, spisów oraz różnych badań ankietowych. O danych osobowych mówimy jedynie w przypadku gdy dotyczą one konkretnej, dającej się zidentyfikować osoby. W przypadku gdy charakteryzujemy jednostki anonimowe mamy do czynienia nie z danymi osobowymi, lecz z charakterystyką zbiorowości odznaczającej się określonymi cechami demograficznymi, społecznymi, zawodowymi itp.

Dane indywidualne są to dane o podmiotach gospodarczych, innych osobach prawnych i jednostkach organizacyjnych nie mających osobowości prawnej i ich działalności. W przypadku niektórych podmiotów (np. indywidualnych gospodarstw rolnych, osób fizycznych prowadzących działalność gospodarczą) dane indywidualne dotyczące działalności podmiotu zawierają również zwykle dane osobowe. Z tego faktu wynika szczególny obowiązek ochrony tych danych. Pod pojęciem danych indywidualnych rozumiemy wyłącznie takie dane, które dotyczą konkretnego zidentyfikowanego podmiotu lub podmiotu możliwego do identyfikacji na podstawie korelacji charakteryzujących go cech.

Dane jednostkowe. Zgodnie z postanowieniami art. 2 p. 9) ustawy o statystyce publicznej dane jednostkowe są to *dane dające się powiązać z konkretną osobą fizyczną lub dane indywidualne dające się powiązać z podmiotem gospodarczym albo z inną osobą prawną, bądź jednostką organizacyjną nie mającą osobowości prawnej.*

Należy zauważyć, że definicja ta może powodować w praktyce pewne trudności w jednoznacznym rozumieniu tego pojęcia. W rozpowszechnionej dotychczas terminologii pojęcie danych jednostkowych odnosi się do wszelkich danych wchodzących do systemu przetwarzania danych (niezależnie od tego, czy dają się one powiązać z konkretnymi jednostkami czy nie) i służących jako podstawa opracowania danych wynikowych, które w odróżnieniu od danych jednostkowych noszą nazwę danych zbiorczych. Z tego rodzaju „danymi jednostkowymi” mamy np. do czynienia w przypadku badań anonimowych, których stosowanie będzie się zapewne rozszerzać w związku z większym prawdopodobieństwem uzyskania bardziej rzetelnych i wiarygodnych informacji od respondentów.

W celu uniknięcia wspomnianej dwuznaczności, w niniejszej pracy pojęcie „dane jednostkowe” będziemy stosowali wyłącznie w rozumieniu zdefiniowanym w ustawie. Zbierane i przetwarzane dane uzyskiwane od podmiotów, niezależnie od tego, czy dają się one odnieść do konkretnego podmiotu lub osoby czy nie, będziemy nazywali danymi *źródłowymi*.

W terminologii anglojęzycznej występuje bliskoznaczne pojęcie „micro-data” lub „micro-statistics”, w odróżnieniu od „agregated data” lub „macro-statistics”. Termin micro-data oznacza zestaw pozycji jednostkowych (rekordów) o indywidualnych podmiotach badania, zawierających informacje na poziomie respondenta. Mogą to być dane dotyczące osób, gospodarstw domowych, przedsiębiorstw, zakładów, małych obszarów geograficznych itp. Każdemu podmiotowi odpowiada indywidualny wektor danych (zestaw danych), który może zawierać cechy ilościowo-wartościowe lub cechy jakościowe. W niniejszej pracy termin „dane źródłowe” oraz „mikrodane” będziemy używali zamiennie.



3. PRAWNE, ETYCZNE I PRAGMATYCZNE PRZESŁANKI OCHRONY DANYCH

Obowiązujące we wszystkich krajach demokratycznych przepisy prawne dotyczące statystyki publicznej (statystyki oficjalnej) przewidują, iż dane jednostkowe zebrane w toku badań statystycznych mogą być wykorzystywane wyłącznie w celu opracowania zestawień i analiz statystycznych i ich poufność musi być w pełni zagwarantowana, pod groźbą określonych w przepisach sankcji. Przyjęcie tego rodzaju rozwiązań prawnych wynika z kilku przyczyn zasługujących na nieco bliższe omówienie.

Ochrona danych jednostkowych, w tym zwłaszcza danych osobowych, traktowana jest jako ważny element gwarantujący prawo do prywatności.

3.1. Konwencje i Zalecenia Rady Europy

Utworzona w 1949 r. Rada Europy⁹⁾ przyjęła w 1950 r.¹⁰⁾ *Konwencję w sprawie praw człowieka i podstawowych swobód*. Była to jedna z pierwszych konwencji przyjętych przez tę organizację. W art. 8 tej Konwencji stwierdza się, że:

1. *Każdy ma prawo do prywatności swojego życia osobistego, rodzinnego i domowego oraz do prywatności korespondencji.*
2. *Władze publiczne nie mogą podejmować działań sprzecznych z tym prawem, z wyjątkiem przypadków przewidzianych w przepisach prawa i koniecznych w społeczeństwach demokratycznych w interesie bezpieczeństwa narodowego, bezpieczeństwa publicznego lub sytuacji ekonomicznej kraju*

⁹⁾ Rada Europy jest instytucją ponadnarodową, umożliwiającą rządów europejskich krajów demokratycznych współpracę zmierzającą do „osiągnięcia większej jedności swoich członków w celu zapewnienia realizacji ideałów i zasad stanowiących ich wspólne dziedzictwo i ułatwienie postępu gospodarczego i społecznego”. Do Rady Europy należy obecnie 35 krajów, w tym Polska. Radą kieruje Komitet Ministrów Spraw Zagranicznych. Funkcje doradcze spełnia Izba Parlamentarna oraz szereg międzyrządowych komitetów ekspertów, zakres działania których obejmuje niemal wszystkie dziedziny życia obywateli Europy, z wyjątkiem zagadnień obrony. Do najważniejszych spraw rozpatrywanych przez Radę Europy należą prawa człowieka, walka z narkotykami, terroryzmem, zapobieganie stosowaniu tortur, pomoc uchodźcom.

¹⁰⁾ 4 listopada w Rzymie.

oraz w celu zapobiegania zaburzeniom lub przestępstwom, w celu ochrony zdrowia lub moralności, lub w celu ochrony praw i swobód innych osób. Jednocześnie jednak w art. 10 teŝe Konwencji stwierdza się, że:

1. *Każdy ma prawo do swobody wyrażania opinii. Prawo to powinno obejmować swobodę posiadania oraz dostępu do bezstronnych informacji i idei bez ingerencji ze strony władz publicznych i niezależnie od granic (...).*
2. *Ze względu na to, że korzystanie z tych swobód wiąŝe się z określonymi obowiązkami i odpowiedzialnością, to może ono podlegać ograniczeniom lub karom przewidywanym prawem, niezbędnym w społeczeństwach demokratycznych, w interesach bezpieczeństwa narodowego, integralności terytorialnej lub bezpieczeństwa publicznego, w celu zapobiegania zaburzeniom lub przestępstwom, w celu ochrony zdrowia lub moralności lub w celu ochrony reputacji lub praw innych osób, w celu zapobieżenia ujawnieniu informacji uzyskanych w zaufaniu lub w celu zachowania autorytetu i bezstronności władz sądowych.¹¹⁾*

Jak łatwo zauwaŝyć, sformułowania art. 8 i 10 Konwencji Praw Człowieka mają znaczenie komplementarne, niemniej jednak praktyczne korzystanie z tych praw przez jedną osobę może w określonych przypadkach prowadzić do naruszenia praw innej osoby. Podobny dylemat istnieje w prawodawstwie statystycznym, w którym z jednej strony przewiduje się prawo do powszechnego dostępu do informacji, ale jednocześnie pełną ochronę informacji indywidualnych.

Od wejścia w życie wspomnianej Konwencji (od 1950 r.) nastąpił gwałtowny wzrost zakresu zbierania danych osobowych dla najróżniejszych celów, a rozszerzanie się zastosowań komputerów doprowadziło do sytuacji szczególnego zagrożenia ingerencją w sferę prywatności obywateli — w związku z gromadzeniem danych osobowych i możliwością ich szerokiego wykorzystywania w systemach informacyjnych — nie tylko przez władze publiczne, ale również przez osoby prywatne i przedsiębiorstwa.

W latach 1973 i 1974 Komitet Ministrów Rady Europy (RE) przyjął dwie rezolucje (odpowiednio nr 22 i 29) w sprawie ochrony danych osobowych w automatycznych bankach danych w sektorze prywatnym (rezolucja 22) i w sektorze publicznym (rezolucja 29). Już w czasie prac nad tymi rezolucjami stwierdzono potrzebę dalszych kroków w kierunku modyfikacji krajowych przepisów prawnych oraz wzmocnienia tych przepisów odpowiednimi uzgodnieniami międzynarodowymi. Podobne sugestie wysunęła w 1972 r. Europejska Konferencja Ministrów Sprawiedliwości.

W 1977 r. rozpoczęto prace nad nową Konwencją w sprawie ochrony interesów jednostki w związku z automatycznym przetwarzaniem danych osobowych. Konwencja ta, przyjęta w styczniu 1981 r., znana jest pod nazwą **Konwencji 108**.¹²⁾

¹¹⁾ Cyt. na podstawie [Aus94].

¹²⁾ Pełny tekst Konwencji zamieszczono w aneksie.

Stosownie do tej Konwencji członkowie RE powinni wydać krajowe przepisy prawne zapewniające wprowadzenie w życie zasad określonych w tej Konwencji z uwzględnieniem ochrony danych osobowych każdego obywatela zamieszkałego na terenie danego kraju. Zasady te odnoszą się do uczciwego i zgodnego z prawem zbierania danych, określenia celów zbierania danych, zasady proporcjonalności, jakości danych, poufności wrażliwych danych, informowania podmiotu o fakcie zbierania o nim danych, prawa dostępu do danych i do ich korygowania.

Konwencja zapewnia również swobodny przepływ danych osobowych między stronami, które podpisały i ratyfikowały Konwencję.¹³⁾ Przepływ ten nie może być hamowany pod pozorem ochrony danych, jeśli w danym kraju obowiązują przepisy ochrony, zgodne z Konwencją 108.

Konwencja ustanowiła **Komitet Konsultacyjny** składający się z przedstawicieli stron w Konwencji. Komitet odpowiada za interpretację przepisów oraz za ułatwienie ich wdrożenia.

Od przyjęcia Konwencji 108 znaczenie ochrony danych jeszcze wzrosło; z jednej strony nastąpił taki stopień komputeryzacji, że komputer osobisty stał się niemal powszechnie dostępny, a z drugiej strony, obywatele są coraz silniej uzależnieni od wielu usług publicznych: banków, usług kredytowych, ubezpieczeń społecznych i innych form pomocy społecznej, ubezpieczeń dobrowolnych, opieki lekarskiej, które to służby współdziałają z zautomatyzowanymi systemami administracyjnymi. W tych warunkach problemy ochrony danych stały się niezwykle ważne. Problemy te powinny być rozwiązywane przez każdy kraj, przez odpowiednie służby ochrony danych. W większości krajów funkcje te spełnia *Rzecznik Ochrony Danych (Data Protection Commissioner)*. Rzecznicy ci stali się w społeczeństwach demokratycznych częścią składową systemu kontrolnego, podobnie jak rzecznicy praw obywatelskich. Również Rada Europy posiada od 1989 r. własnego Rzecznika Ochrony Danych, który nadzoruje sprawy ochrony danych osobowych w Sekretariacie Generalnym Rady Europy.

Rosnący strumień danych osobowych od podmiotów do różnych organizacji zbierających dane, od organizacji zbierających dane do ośrodków przetwarzających, od ośrodków przetwarzających do innych ośrodków przetwarzających wymaga uregulowania, nawet jeśli na pierwszy

¹³⁾ Polska nie podpisała dotąd Konwencji 108 ze względu na to, że w naszym kraju nie przyjęto ustawy o ochronie danych osobowych, co traktowane jest jako warunek podpisania Konwencji.

Przepisy dotyczące ochrony danych w ustawie o statystyce publicznej regulują te problemy w dziedzinie badań statystycznych. Warunkiem przystąpienia do Konwencji 108 jest przyjęcie przez kraj rozwiązań prawnych regulujących **całokształt** problemów ochrony danych osobowych (danych medycznych, danych wykorzystywanych do badań naukowych, do celów reklamy i marketingu, ubezpieczeń społecznych, danych dotyczących zatrudnienia, danych wykorzystywanych przez policję i inne służby ochrony porządku publicznego, danych dotyczących skazanych itp.).

rzut oka ten przepływ informacji odbywa się w interesie podmiotów i dane zostały zebrane za ich zgodą.

Konwencja 108 zawiera ogólne zasady odnoszące się do każdej sytuacji. Zasady te w odniesieniu do różnych dziedzin, jak: medycyny, ubezpieczeń społecznych, ubezpieczeń dobrowolnych, bankowości, zatrudnienia, reklamy, statystyki, telekomunikacji, policji itd. muszą być dopiero opracowane. Opracowanie i wprowadzenie w życie takich zasad odbywa się za pośrednictwem zaleceń (rekomendacji).

Zalecenia w sprawie ochrony danych. W każdej wspomnianej wyżej dziedzinie dane powinny być zbierane i przetwarzane zgodnie z podstawowymi zasadami Konwencji 108, ale środki i sposoby mogą być różne. W niektórych dziedzinach metody mogą być bardziej elastyczne niż w innych, a szczególnie regulacje mogą być rozwinięte w różnym stopniu w odniesieniu do różnych zawodów. Tak więc, dla osiągnięcia większej harmonizacji zasad ochrony danych, Konwencja wymaga opracowania specyficznych zasad ochrony w każdej ze wspomnianych wyżej dziedzin.

Komitet Ministrów RE doszedł do wniosku, że lepszym rozwiązaniem od zmiany konwencji lub uzupełniania jej protokołami, będzie zastosowanie w tym celu innego narzędzia, a mianowicie opracowywania **zaleceń** (rekomendacji) pod adresem rządów. Zaletą zaleceń jest, że są one łatwiejsze do wprowadzenia; zamiast podpisywania i ratyfikowania przez każdy z krajów członkowskich, wymagają one jedynie jednogłośniego przyjęcia przez Komitet Ministrów. Procedura jest więc znacznie prostsza niż w przypadku zmiany konwencji, a przede wszystkim, mimo iż zalecenia nie są prawnie obowiązujące, to mogą być adresowane do wszystkich państw, niezależnie od tego, czy podpisały one Konwencję 108 czy nie. Zalecenia w tym przypadku nakładają na wszystkie kraje moralny obowiązek rozpatrzenia możliwości ich zastosowania.

W celu opracowania wspomnianych różnych zaleceń, co wymaga, obok doświadczeń i kwalifikacji prawniczych, również specjalistycznych kwalifikacji w dziedzinie, której zalecenia dotyczą, Komitet Ministrów RE utworzył w 1976 r. **Komitet Ekspertów do Spraw Ochrony Danych**, który ostatnio został przemianowany w **Grupę Projektową do Spraw Ochrony Danych** (*Project Group on Data Protection*). Grupa składa się z przedstawicieli szczebla kierowniczego 34 krajów członkowskich, którzy w swoich krajach są odpowiedzialni za ochronę danych osobowych. W toku opracowania zaleceń dziedzinowych przedstawiciele ci często zasięgają opinii specjalistów z danej dziedziny. Ponadto w takich międzynarodowych spotkaniach Grupy Projektowej uczestniczą obserwatorzy z europejskich organizacji zawodowych i stowarzyszeń zainteresowanych określonymi aspektami działania Grupy Projektowej. Oba ciała (Grupa Projektowa i Komitet Konsultacyjny) współpracują z sobą, unikając dublowania prac. Komitet Konsultacyjny czuwa nad przestrzeganiem i popularyzacją Konwencji 108, natomiast

Grupa Projektowa zajmuje się bardziej szczegółowymi zagadnieniami odnoszącymi się do poszczególnych dziedzin tematycznych.

W ciągu kilku lat od utworzenia Grupy Projektowej opracowano szereg zaleceń, w tym m.in.: Zalecenie nr R1 (1981) w sprawie zasad postępowania z danymi medycznymi w bankach danych (23 stycznia 1981), Zalecenie nr R10 (1983) w sprawie ochrony danych osobowych wykorzystywanych do badań naukowych i statystyki (23 września 1983), Zalecenie nr R20 (1985) o ochronie danych osobowych wykorzystywanych do celów bezpośredniego marketingu (25 października 1985), Zalecenie nr 1 (1986) o ochronie danych osobowych wykorzystywanych w ubezpieczeniach społecznych (23 stycznia 1986), Zalecenie nr R15 (1987) dotyczące wykorzystania danych osobowych w działalności policji (17 września 1987), Zalecenie nr R2 (1989) w sprawie ochrony danych osobowych wykorzystywanych w związku z zatrudnieniem (18 stycznia 1989), Zalecenie nr R10 (1991) w sprawie przekazywania stronom trzecim danych osobowych przechowywanych przez instytucje publiczne (9 września 1991).

Oprócz zaleceń, Grupa Projektowa opublikowała wyniki szeregu ważnych badań dotyczących problemów ochrony danych: *Nowe technologie: wyzwania do ochrony prywatności* (1989), *Ochrona danych oraz środki masowego przekazu* (1990), *Wprowadzenie i wykorzystanie identyfikatorów osobowych: problemy ochrony danych* (1991).¹⁴⁾

3.1.1. Zalecenia Rady Europy w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych

Sposób wykorzystania danych w badaniach statystycznych tym różni się od wykorzystywania danych do innych wymienionych wyżej celów, że dane uzyskane w badaniach statystycznych nie służą do podejmowania decyzji w stosunku do indywidualnych podmiotów. Statystyka dostarcza jedynie zagregowanych informacji (ilościowych lub jakościowych), które służą do prowadzenia analiz oraz do podejmowania decyzji, lecz nie w stosunku do indywidualnych osób, ale w zakresie obszarów tematycznych lub problemów gospodarczych i społecznych, których zebrane informacje dotyczą.

Jeśli tak spojrzymy na cele statystyki oraz jeśli uwzględnimy, że dostarcza ona jedynie anonimowych i zagregowanych wyników, to na pierwszy rzut oka może się wydawać, że działalność statystyczna oraz wyniki jakich ona dostarcza nie podpadają pod przepisy o ochronie danych. Jest to jednak

¹⁴⁾ Por.: Council of Europe, Working Group on Data Protection (CJ-PD-GT/13 — Statistics), *Draft Explanatory Memorandum of the Draft Recommendation on the protection of personal data collected and processed for statistical purposes*, Strasbourg, 17 November 1995 (*Rada Europy, Grupa Robocza do Spraw Ochrony Danych* (CJ-PD-GT/13 — Statystyka), *Projekt Memorandum objaśniającego dotyczącego projektu zaleceń w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych*, Strasbourg, 17 listopada 1995).

pogląd uproszczony. W procesie zbierania i przetwarzania danych statystycznych istnieje ryzyko dostępu do danych indywidualnych i wykorzystania ich do celów niezgodnych z celami badań statystycznych. Podobne ryzyko powstaje, gdy, mimo swojej anonimowości i zbiorczego charakteru, prezentowane wyniki (jeśli obejmują zwłaszcza niewielkie zbiorowości) umożliwiają identyfikację podmiotu na podstawie skorelowania różnych cech, charakteryzujących prezentowany agregat. Z tego względu zasady ochrony danych muszą odnosić się również do statystyki. W dziedzinie badań statystycznych istnieje jednocześnie szereg sytuacji specyficznych, które powodują, że niezależnie od ogólnych przepisów Konwencji 108 istnieje potrzeba bardziej szczegółowego opracowania szeregu aspektów ochrony danych dotyczących korzystania z danych osobowych w badaniach naukowych lub do celów statystycznych.

W listopadzie 1990 r. Komitet Ekspertów do Spraw Ochrony Danych doszedł do wniosku, że istnieje potrzeba bardziej precyzyjnego określenia zasad ochrony danych osobowych zbieranych do celów statystycznych, zwłaszcza danych spisów. Uzgodniono, że Zalecenie R(83)10 powinno być poddane ocenie z punktu widzenia problemów ochrony danych w całym procesie badań statystycznych. W tym celu utworzono Grupę Roboczą (Grupa nr 13) z udziałem ekspertów z Austrii, Danii, Francji, Grecji, Portugalii, Szwajcarii i Węgier. Grupa ta odbyła 7 spotkań w okresie grudzień 1991—styczeń 1995. Grupa pracowała pod przewodnictwem dra Jean-Philippe'a Waltera ze Szwajcarii. W spotkaniach Grupy brali udział obserwatorzy z Eurostatu i międzynarodowej Izby Gospodarczej. W marcu 1993 r. dokonano wymiany poglądów między Komitetem do Spraw Poufności Eurostatu i Grupą Roboczą RE. Grupa Robocza do Spraw Statystyki zakończyła swoją pracę w styczniu 1995 r. i przekazała projekt zaleceń w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych Grupie Projektowej do Spraw Ochrony Danych. Projekt ten był przedmiotem kilku posiedzeń Grupy Projektowej oraz posiedzeń Biura Grupy.¹⁵⁾

Wspomniany projekt jest aktualnie w trakcie omawiania przez Grupę Projektową, szereg sformułowań budzi wiele zastrzeżeń i prawdopodobnie będzie jeszcze podlegać modyfikacji, niemniej jednak, ze względu na duże znaczenie tego dokumentu dla naszych prac zmierzających do usprawnienia i doskonalenia systemu ochrony informacji indywidualnych, dokument ten zasługuje na nieco bliższe omówienie.¹⁶⁾ Projekt zaleceń podkreśla, że odnoszą się one zarówno do jednostkowych danych osobowych, jak i do wyników, jeśli na ich podstawie można zidentyfikować podmiot, którego dane te dotyczą.

¹⁵⁾ W pracach tych przedstawiciele GUS nie brali udziału.

¹⁶⁾ Pełny tekst projektu zamieszczono w aneksie do niniejszej pracy.

Wprawdzie zalecenia odnoszą się do warunków automatycznego zbierania i przetwarzania danych, niemniej jednak kraje członkowskie RE proszone są o zastosowanie analogicznych zaleceń w stosunku do ręcznych procedur zbierania i przetwarzania danych, jak również o niewprowadzanie takich procedur w celu uchylecia się od stosowania omawianych zaleceń. Dane osobowe mogą być zbierane i przetwarzane dla celów statystycznych wyłącznie jeśli: a) jest to określone w przepisach prawa krajowego, b) wynika to z innych prawnych zobowiązań lub c) prawo krajowe nie zabrania, a osoby badane lub ich prawni przedstawiciele wyrażają na to zgodę.

W celu ograniczenia zbierania identycznych danych od tych samych osób, dane zebrane w celach niestatystycznych mogą być w razie potrzeby wykorzystywane do celów statystycznych. W tym przypadku jednak należy zagwarantować, że dane te nie mogą być wykorzystywane do **podejmowania decyzji dotyczących badanych podmiotów**.

W tym miejscu należałoby nieco szerzej omówić zagadnienie przepływu danych między systemami administracyjnymi i systemem informacji statystycznej. Sprawy te wywołują nierzadko wątpliwości i zastrzeżenia. W większości krajów obowiązuje zasada, iż dane gromadzone w systemach administracyjnych (ewidencjach, rejestrach, dokumentach) **mogą być przekazywane służbom statystycznym** do ich wykorzystania do opracowań i analiz statystycznych. W naszej ustawie o statystyce publicznej, z dnia 29 czerwca 1995 r., w art. 13 ust. 3 i 4 wyraźnie stwierdza się, że:

organy administracji rządowej i samorządu terytorialnego, inne instytucje rządowe, organy prowadzące urzędowe rejestry i Narodowy Bank Polski nieodpłatnie przekazują służbom statystyki publicznej zgromadzone dane administracyjne w zakresie, formie i terminach określanych każdorazowo w programie badań statystycznych statystyki publicznej, w szczególności w formie wyciągów z rejestrów, kopii: zbiorów danych, zebranych deklaratywnych, dokumentów ewidencyjnych i innych formularzy urzędowych, wyników pomiarów, danych monitoringu środowiska oraz udostępniają dane z baz danych systemów informatycznych.

Niedopuszczalny i prawnie zabroniony jest natomiast przepływ danych jednostkowych w odwrotnym kierunku — od służb statystyki publicznej do systemów administracyjnych. Dotyczy to zarówno danych posiadanych przez służby statystyczne z własnych badań, jak i danych uzyskanych z innych systemów. Tak więc przepływ danych jest wyłącznie jednokierunkowy.

Dane osobowe mogą być zbierane w formie badań statystycznych na zasadzie obowiązku statystycznego wyłącznie w przypadkach przewidzianych przez prawo krajowe. Dane identyfikacyjne osób mogą być zbierane i przetwarzane wyłącznie w przypadkach, gdy jest to niezbędne dla osiągnięcia celów badań.

Jeśli w badaniu statystycznym zbierane są dane osobowe, to osoby badane muszą zostać poinformowane:

- 1) czy badanie jest obowiązkowe czy dobrowolne oraz w oparciu o jaką podstawę prawną dane są zbierane,
- 2) w jakim celu dane są zbierane i opracowywane,
- 3) kto jest upoważniony do zbierania i opracowywania danych (nazwisko i stanowisko służbowe osoby zbierającej dane oraz nazwa instytucji),
- 4) że zbierane dane podlegają ochronie oraz że będą wykorzystywane wyłącznie w celach statystycznych,
- 5) że zainteresowani mogą uzyskać w miarę potrzeby dodatkowe informacje.

Wspomniane wyżej w p. 5) informacje mogą dotyczyć zwłaszcza tego, w jakiej formie osoba badana może odmówić udziału w badaniu lub wycofać wyrażoną uprzednio zgodę na uczestnictwo w badaniu (w przypadku badań dobrowolnych), lub jakie sankcje jej grożą za odmowę udziału w badaniu (w przypadku badań obowiązkowych). Zainteresowane osoby powinny także uzyskać na żądanie dodatkowe informacje na temat tego, komu i w jakiej formie będą dostarczane wyniki badań lub dane (nie dające się zidentyfikować) oraz w jaki sposób zapewnia się ochronę informacji indywidualnych.

Służby statystyczne powinny informować o prowadzonych badaniach również wtedy, gdy przewiduje się zbieranie danych bez bezpośredniego udziału osób badanych, przykładowo, gdy dane osobowe są uzyskiwane ze źródeł (dokumentów, rejestrów) administracyjnych.

Jeśli zbieranie danych osobowych wymaga zatrudnienia ankierów lub innych osób, które będą uzyskiwać odpowiedzi od osób badanych, należy podjąć szczególne środki ostrożności w czasie doboru osób oraz instytucji oraz zapewnić takie formy zbierania danych, aby zagwarantować właściwe spełnienie celu badań. Jednocześnie należy zapewnić pełną poufność danych oraz ochronę prywatności, a osoba przeprowadzająca badanie powinna posiadać upoważnienie do przeprowadzania badania. Również wszystkie osoby biorące udział w przetwarzaniu danych osobowych do celów statystycznych powinny zapewnić poufność tych danych poprzez przedsięwzięcie odpowiednich środków technicznych i organizacyjnych. Są one zobowiązane zwłaszcza do podjęcia środków zapobiegających nieupoważnionemu dostępowi do danych, ich zmianie, przekazaniu lub innej formie nieupoważnionego przetwarzania. Jeśli dane muszą być przechowywane w postaci możliwej do identyfikacji, to należy zapewnić użycie wszelkich organizacyjnych i technicznych środków, w tym zwłaszcza środków automatycznych, aby zapobiec nielegalnej identyfikacji badanych podmiotów.

Jeśli badanie przewiduje zbieranie i przetwarzanie danych osobowych od osób, które z jakiegokolwiek powodu nie mogą podejmować decyzji we własnym imieniu, to wymagane jest uzyskanie zgody na uczestnictwo w badaniu od osób prawnie upoważnionych do działania w imieniu tych osób.

W części dotyczącej przechowywania danych zalecenia przewidują, że dane osobowe zawierające część identyfikacyjną powinny być zniszczone lub pozbawione informacji identyfikującej niezwłocznie po tym, jak część identyfikacyjna przestaje być konieczna w procesie opracowania statystycznego. Jeśli natomiast dane identyfikacyjne są niezbędne w celu przewidywanych badań w przyszłości (przykładowo, w przypadku badań panelowych), to dane identyfikacyjne powinny być oddzielone od pozostałych informacji i przechowywane oddzielnie.

Dane identyfikacyjne mogą być wykorzystywane również do tworzenia zbiorów adresowych, do prowadzenia badań statystycznych, jeśli prawo krajowe na to pozwala oraz jeśli podmioty zostały o tym poinformowane.

Informacje wynikowe mogą być publikowane lub udostępniane w inny sposób jedynie w przypadku, gdy na podstawie tych informacji nie jest możliwe zidentyfikowanie badanego podmiotu lub gdy podmiot wyraził na to zgodę, lub też, gdy środki przewidziane przez prawo krajowe gwarantują, że ujawnienie tych danych nie naruszy prywatności badanego podmiotu.

Korelowanie lub łączenie danych osobowych, zebranych w określonym celu w badaniach statystycznych, lub przetwarzanie danych osobowych zebranych w celach niestatystycznych jest dopuszczalne wyłącznie wtedy, gdy jest to przewidziane w prawie krajowym i gdy prawo to zapewnia niezbędne środki zachowania poufności i ochrony danych, w szczególności zapobiega ich przetwarzaniu i przekazywaniu do celów pozastatystycznych.

Zalecenie podkreśla również, że osoby i instytucje zajmujące się badaniami statystycznymi winne przyjąć i opublikować kodeks etyczny, zgodny z omawianymi zaleceniami i określającymi w szczególności:

- a) osoby i instytucje, które mają dostęp do danych osobowych,
- b) środki, które należy podejmować w celu ochrony, zapewnienia poufności i bezpieczeństwa tych danych oraz środki zapewniające przestrzeganie zasad etyki statystycznej,
- c) osoby uczestniczące w przetwarzaniu danych statystycznych.

W celu zapewnienia szerokiego dostępu do narzędzi informacyjnych oraz wiedzy technicznej uwzględniającej skuteczną ochronę danych osobowych zbieranych do celów statystycznych, kompetentne władze rządowe powinny ściśle współpracować w dziedzinie rozwoju tych narzędzi i wiedzy technicznej oraz uruchamiać programy współpracy międzynarodowej, wymiany doświadczeń, przekazywania wiedzy i pomocy technicznej.

3.2. Podstawowe zasady statystyki oficjalnej i kodeks etyczny statystyków

Drugim dokumentem o dużym międzynarodowym znaczeniu dla poufności i ochrony danych jednostkowych jest dokument pt. *Podstawowe zasady statystyki oficjalnej na obszarze Europejskiej Komisji Gospodarczej*, przyjęty

na 47 sesji EKG 15 kwietnia 1992 r.¹⁷⁾ Podstawowe zasady zostały również przyjęte w formie deklaracji w sprawie podstawowych zasad statystyki oficjalnej w 1994 r. przez Komisję Statystyczną ONZ i od tego czasu stanowią one dokument określający zasady funkcjonowania systemów informacji statystyki publicznej o zasięgu światowym.

Wśród 10 punktów sformułowanych w tych zasadach, jeden punkt (punkt 6) ustala, że *dane indywidualne zbierane przez organy statystyczne w celu sporządzania statystyk odnoszących się do osób fizycznych lub prawnych muszą być ściśle poufne i mogą być wykorzystywane wyłącznie do celów statystycznych*. Ten zapis ma szerszy zasięg znaczeniowy w porównaniu do zapisów ujmowanych w „zaleceniach” rozpatrywanych przez Radę Europy. Dokumenty RE, z racji zakresu zainteresowania tej organizacji, dotyczą wyłącznie danych osobowych. W przypadku „Podstawowych zasad...” chodzi zarówno o ochronę prawa do prywatności, jak i o ochronę danych ekonomicznych, których ujawnienie mogłoby przynieść określoną szkodę danej jednostce.

Bardzo ważnym dokumentem wywierającym wpływ na postawy i zachowania statystyków w całym procesie przygotowania i prowadzenia badań oraz upowszechniania informacji statystycznej jest *Deklaracja w sprawie etyki zawodowej statystyków*. Deklaracja stanowi swoisty kodeks postępowania statystyków. Dokument ten jest wspólnym dorobkiem statystyków wielu krajów. Został on przyjęty przez Międzynarodowy Instytut Statystyczny na jego 45 sesji plenarnej w 1985 r. w Amsterdamie [Wal95].

W odróżnieniu od omówionych wyżej dokumentów Rady Europy i „Podstawowych zasad”, które adresowane są przede wszystkim do rządów oraz do służb statystyki publicznej i odnoszą się do funkcjonowania statystyki jako systemu informacyjnego, „Deklaracja” określa zasady postępowania indywidualnych statystyków w czasie wykonywania przez nich funkcji zawodowych.

W części dotyczącej obowiązków statystyków wobec podmiotów badanych i ochrony danych jednostkowych w Deklaracji podkreślono, że statystycy prowadząc badania muszą mieć świadomość, że niemal każde badanie wiąże się z określoną uciążliwością dla jednostek wyznaczonych lub wybranych losowo do badań (osób, gospodarstw domowych, jednostek prowadzących działalność gospodarczą, instytucji). Stopień tej uciążliwości może być różny. Różne też mogą być odczucia i reakcje poszczególnych jednostek. Niektóre rodzaje pytań objętych programem badań mogą wywoływać negatywne reakcje respondentów oraz niechęć udzielania na nie odpowiedzi. Część danych o badanych jednostkach statystycy mogą uzyskiwać z innych źródeł bez wiedzy i zgody zainteresowanych podmiotów.¹⁸⁾

¹⁷⁾ Tekst „Podstawowych zasad...” opublikowały „Wiadomości Statystyczne” 1995, nr 2 oraz „Biblioteka Wiadomości Statystycznych”, tom 47, 1995.

¹⁸⁾ W tej sprawie projekt Zalecenia Rady Europy jest bardziej rygorystyczny. Zaleca on **informowanie podmiotów** o fakcie wykorzystywania danych osobowych ze źródeł niestatystycznych do celów statystycznych — por. p. 5.4 Zaleceń przytoczonych w aneksie.

Aby ograniczyć uciążliwość związaną ze zbieraniem danych statystycy powinni rozszerzać wykorzystywanie danych ze źródeł administracyjnych, pamiętając w każdym przypadku o konieczności zapewnienia ochrony danych jednostkowych i wykorzystywania zebranych danych wyłącznie do celów analityczno-statystycznych, a nie do celów administracyjnych.

Obowiązek ochrony danych indywidualnych zebranych w formie badań statystycznych oraz zakaz ich udostępniania komukolwiek i w jakiegokolwiek formie w Deklaracji akcentuje się szczególnie mocno. Zobowiązuje ona statystyków nie tylko do skutecznej ochrony i do nieudostępniania danych jednostkowych, ale i do podejmowania właściwych środków zapobiegających bezpośredniemu lub pośredniemu ujawnieniu takich danych.

3.3. Polska ustawa o statystyce publicznej

Zagadnienia ochrony danych i tajemnicy statystycznej akcentowane są w niemal wszystkich przepisach prawnych rangi ustawowej, regulujących funkcjonowanie statystyki w Polsce.

W pierwszej, wydanej po I wojnie światowej, *Ustawie z dnia 21 października 1919 r. o organizacji statystyki administracyjnej*¹⁹⁾ w art. 4 postanawia się, że *Zeznania, poczynione na podstawie obowiązku, nałożonego w myśl art. 2, mogą być używane tylko do celów statystycznych i nie mogą być udzielane władzom publicznym, ani też osobom prywatnym w innym celu. Za ścisłe przestrzeganie tego przepisu odpowiadają osobiście funkcjonariusze, biorący udział w dochodzeniu.*

Również w pierwszym przepisie prawnym, dotyczącym statystyki po II wojnie światowej, a mianowicie w *Dekrecie z dnia 31 lipca 1946 r. o organizacji statystyki państwowej i o Głównym Urzędzie Statystycznym*²⁰⁾ zawarto w art. 12 przepis ustalający, że *Zeznania indywidualne uzyskane w wyniku spisów powszechnych i innych badań statystycznych, wymagających składania zeznań przez prywatne osoby i instytucje, mogą być używane wyłącznie do zestawień statystycznych. Dane te nie mogą być udzielane ani władzom publicznym ani też instytucjom i osobom prywatnym w innym celu. Za ścisłe przestrzeganie tego przepisu osobiście odpowiadają osoby, biorące udział w badaniu statystycznym.*

W kolejnej ustawie z dnia 15 lutego 1962 r. o organizacji statystyki państwowej²¹⁾ przepis dotyczący tajemnicy statystycznej jeszcze bardziej zaostrożono. W art. 17 ustala się, że *Wykorzystywanie informacji i zeznań indywidualnych, uzyskanych w wyniku spisów powszechnych i innych badań statystycznych, dla celów innych niż zestawień statystycznych jest zabronione, a w art. 18 postanawia się, że Kto wykracza przeciwko zakazowi określone*

¹⁹⁾ Dz. U. Nr 85, poz. 464.

²⁰⁾ Dz. U. Nr 41, poz. 239.

²¹⁾ Dz. U. Nr 10, poz. 47.

w art. 17, podlega karze aresztu do trzech miesięcy lub grzywny do 4500 złotych.

Podobnie, w ustawie z dnia 26 lutego 1982 r. o statystyce państwowej²²⁾ w art. 24 postanawia się, że *wykorzystywanie informacji indywidualnych, uzyskanych w wyniku spisów powszechnych i innych badań statystycznych, dla celów innych niż dla zestawień i analiz statystycznych jest zabronione*. Należy zauważyć, że ustawa ta wprowadza również pojęcie ochrony publikowanych wyników badań, określając w art. 25, że *publikowanie wyników badań statystycznych oraz dostęp do informacji statystycznych podlegają ograniczeniom wynikającym z ochrony tajemnicy państwowej i gospodarczej oraz sfery praw osobistych obywateli*.

Obowiązująca obecnie, od początku listopada 1995 r., ustawa o statystyce publicznej, przygotowywana przez dłuższy okres, przy współudziale wielu znanych specjalistów statystyków z urzędów statystycznych za granicą²³⁾, przewiduje cały szereg przepisów gwarantujących ochronę danych jednostkowych, zarówno danych osobowych jak i danych indywidualnych. Jednocześnie ustawa zobowiązuje służby statystyczne do informowania podmiotów o prowadzonych badaniach oraz o ciążyącym na służbach statystycznych obowiązku zachowania poufności danych, tj. do zachowania tajemnicy statystycznej. Spośród przepisów ustawy odnoszących się do poufności danych jednostkowych i obowiązku ich ochrony należy zwrócić uwagę zwłaszcza na:

1) sformułowanie art. 10 przewidujące, że:

zbierane i gromadzone w badaniach statystycznych statystyki publicznej dane indywidualne i dane osobowe są poufne i podlegają szczególnej ochronie; dane te mogą być wykorzystywane wyłącznie do opracowań, zestawień i analiz statystycznych oraz do tworzenia przez służby statystyki publicznej operatu do badań statystycznych prowadzonych przez te służby; udostępnianie lub wykorzystywanie danych indywidualnych i danych osobowych dla innych niż podane celów jest zabronione (tajemnica statystyczna);

2) sformułowanie art. 38 określające, że:

1. Nie mogą być publikowane ani udostępniane dane jednostkowe uzyskane w badaniach statystycznych statystyki publicznej. 2. Nie

²²⁾ Dz. U. Nr 7, poz. 58.

²³⁾ Szereg ciekawych wniosków i propozycji, w tym również w dziedzinie gwarancji prawnych dotyczących poufności i ochrony danych indywidualnych i osobowych, wynika z obrad Konferencji Okrągłego Stołu nt. *Podstawowe zasady statystyk oficjalnych oraz ich prawnych aspektów w krajach w okresie przejściowym*. Konferencja odbyła się w Jachrance w dniach 27—29 września 1993 r., a jej organizatorami byli: Konferencja Statystyków Europejskich, Eurostat, Międzynarodowy Instytut Statystyczny, Polskie Towarzystwo Statystyczne oraz Główny Urząd Statystyczny. Obszerny wybór materiałów z tej konferencji opublikowano w 47 tomie „Biblioteki Wiadomości Statystycznych”, zob. [Pod95].

mogą być publikowane ani udostępniane uzyskane w badaniach statystycznych statystyki publicznej informacje statystyczne możliwe do powiązania i zidentyfikowania ich z konkretną osobą oraz dane indywidualne charakteryzujące wyniki ekonomiczne działalności podmiotów gospodarczych, w szczególności jeżeli na daną agregację składa się mniej niż trzy podmioty lub udział jednego podmiotu w określonym zestawieniu jest większy niż trzy czwarte całości;

3) zobowiązanie Prezesa GUS do

zapewnienia przechowywania zgromadzonych danych statystycznych w sposób gwarantujący przestrzeganie tajemnicy statystycznej.

Kiedy mówimy o konieczności zachowania tajemnicy statystycznej, zwłaszcza w odniesieniu do danych osobowych, zadawane są często pytania — przez jak długi okres tajemnica ta obowiązuje? Czy np. dane spisu ludności przeprowadzonego przez GUS w 1931 r. są nadal objęte tajemnicą statystyczną?

Jednoznacznego stwierdzenia w tej sprawie w ustawie brak, z czego można by wnioskować, że okres zachowania tajemnicy statystycznej jest bezterminowy. Przykładowo, w W. Brytanii dane jednostkowe spisu ludności są traktowane jako poufne przez okres 100 lat od daty przeprowadzenia spisu. Do celów naukowo-badawczych mogą być jednak wykorzystywane wcześniej, jeśli zostaną pozbawione informacji identyfikującej poszczególne gospodarstwa domowe i osoby [Fea94, s. 116].

Argumentem wzmacniającym obowiązki przestrzegania zachowania tajemnicy statystycznej i gwarancje ochrony danych gromadzonych w systemie informacyjnym statystyki są przepisy karne ustawy. Określają one (art. 54), że:

kto narusza tajemnicę statystyczną, podlega karze pozbawienia wolności do lat 3,

natomiast art. 55 wzmacnia jeszcze wysokość kar, gdy ujawnienie danych podlegających ochronie następuje w celu osiągnięcia korzyści majątkowych:

kto w celu osiągnięcia korzyści majątkowej lub osobistej wykorzystuje dane statystyczne, z którymi zapoznał się w związku z wykonywaniem pracy lub czynności na zlecenie prowadzącego badanie statystyki publicznej, podlega karze pozbawienia wolności do lat 5.

Analizując przepisy polskiej ustawy o statystyce można stwierdzić z zadowoleniem, że uwzględnia ona niemal wszystkie (poza jednym przypadkiem, który omówimy dalej) sformułowania i zalecenia międzynarodowe, a jej sformułowania są nawet bardziej rygorystyczne i ograniczające. I tak, przykładowo:

1. Omawiane wyżej projektowane zalecenia RE odnoszą się do zbierania i **automatycznego** przetwarzania danych osobowych, natomiast przepisy naszej ustawy odnoszą się do wszystkich przypadków zbierania i opracowania danych, niezależnie od stosowanych w tym celu środków technicznych.

2. W naszej ustawie wprowadza się dość rygorystyczne ograniczenia co do tematyki badanych cech osobowych. Służby statystyczne w swoich badaniach służących celom statystycznym i dla przygotowania prognoz demograficznych mogą zbierać wyłącznie następujące dane o osobach fizycznych (art. 35): a) imię i nazwisko, b) płeć, c) data i miejsce urodzenia, d) obywatelstwo, e) stan cywilny, f) data zawarcia i ustania małżeństwa, g) adres zamieszkania, h) data zgonu, i) identyfikator systemu ewidencji ludności. Inne niż określone wyżej dane, dające się zidentyfikować z konkretną osobą fizyczną, mogą być zbierane przez służby statystyki publicznej wyłącznie wtedy, gdy w programie badań statystycznych wskazano jako źródło danych dokument zawierający dane osobowe nie dające się w prosty sposób oddzielić przy ich przekazywaniu do celów statystycznych lub gdy identyfikacja jest niezbędna do opracowania wyników badania. Zbieranie jakichkolwiek danych w formie spisów powszechnych dopuszczalne jest wyłącznie na podstawie odrębnej ustawy.

Jeszcze bardziej rygorystyczne ograniczenia wprowadza art. 8 ustawy, w którym stwierdza, że w *badaniach statystycznych z udziałem osób fizycznych nie mogą być zbierane na zasadzie obowiązku informacje dotyczące rasy, wyznania, życia osobistego oraz poglądów filozoficznych i politycznych*. Zalecenie Rady Europy (p. 4.8) dopuszcza zbieranie podobnych danych (zwanymi danymi wrażliwymi) pod warunkiem, że są one niezbędne do spełnienia uzasadnionych celów badania i z odpowiednimi gwarancjami przewidzianymi w prawodawstwie krajowym.

3. W projekcie zaleceń Rady Europy (RE) przewiduje się możliwość udostępniania danych osobowych zebranych w celach statystycznych lub niestatystycznych, jeśli badany podmiot lub jego prawny przedstawiciel wyraził na to zgodę (p. 11.4 zaleceń). Nasza ustawa takiego wyjątku nie przewiduje.

W jednym punkcie występuje pewna różnica między naszą ustawą o statystyce publicznej a ustaleniami w projekcie zaleceń RE. W projekcie zaleceń w p. 4.1 w obecnym sformułowaniu stwierdza się, że *dane osobowe zebrane lub przetwarzane w celach statystycznych winny służyć wyłącznie do tych celów. W szczególności, nie mogą być wykorzystywane do podejmowania decyzji odnoszących się do danego podmiotu lub do korygowania zbiorów zawierających dane osobowe, które są przetwarzane w celach niestatystycznych*.

Nasza ustawa o statystyce publicznej upoważnia służby statystyczne do wykorzystywania danych pochodzących z badań statystycznych do aktualizacji rejestrów podmiotów gospodarki narodowej. Mówi się o tym w art. 42 ust. 8: *Informacje objęte rejestrem podmiotów są aktualizowane na podstawie danych uzyskanych od podmiotu, danych uzyskanych w badaniach statystycznych oraz danych uzyskiwanych od organów rejestrowych i ewidencyjnych*.

Zgodnie z postanowieniem art. 41 ust. 3, zbierane i gromadzone w krajowych rejestrach urzędowych informacje **nie są danymi statystycznymi**

w rozumieniu ustawy i z wyjątkiem danych o liczbie pracujących oraz, w przypadku gospodarstw rolnych, o powierzchni ogólnej i użytków rolnych nie podlegają rygorom poufności i tajemnicy statystycznej. Ze sformułowań tych wynika, że wykorzystanie danych pochodzących z badań statystycznych do aktualizacji rejestru krajowego nie jest wykorzystaniem danych do celów statystycznych, jest więc niezgodne z ustaleniami p. 4.1 projektu zaleceń RE.

Dla ścisłości należy podkreślić, że niezgodność ta dotyczy nie wszystkich podmiotów ujętych w rejestrze, lecz tylko osób fizycznych prowadzących działalność gospodarczą oraz w przyszłości także indywidualnych gospodarstw rolnych, tylko bowiem w tych przypadkach w rejestrze mogą wystąpić dane osobowe, którymi zgodnie z definicją zamieszczoną w p. 1 zaleceń, są „wszelkie informacje odnoszące się do osób zidentyfikowanych lub dających się zidentyfikować”²⁴).

Należy podkreślić, że w rejestrze jednostek prowadzonym zgodnie z ustaleniami art. 42 ustawy o statystyce publicznej nie są rejestrowane dane uznawane zwykle za wrażliwe, dotyczą bowiem wyłącznie takich danych jak: 1) imię i nazwisko, adres oraz nr identyfikacyjny systemu ewidencji ludności, 2) wykaz jednostek lokalnych wyodrębnionych pod względem organizacyjnym, lokalnym i finansowym, 3) forma prawna i forma własności, 4) rodzaj wykonywanej działalności, 5) data powstania, rozpoczęcia, zawieszenia i zakończenia działalności, 6) liczba pracujących oraz, w przypadku gospodarstw rolnych, powierzchnia ogólna i powierzchnia użytków rolnych. Niemniej jednak niezgodność z „Zaleceniami” jest oczywista.

Główny Urząd Statystyczny wystąpił do Grupy Projektowej do Spraw Ochrony Danych, przygotowującej projekt zaleceń do przyjęcia przez Komitet Ministrów RE, z sugestią odpowiedniej modyfikacji p. 4.1 „Zaleceń”, polegającej na dodaniu zapisu, iż „dane osobowe pochodzące z badań statystycznych mogą być wykorzystywane do aktualizacji rejestrów prowadzonych przez służby statystyczne w zakresie przewidzianym przez ustawy obowiązujące w danym kraju”. Skuteczność tej interwencji trudno przewidzieć. W Grupie Projektowej, jak się wydaje, decydujący głos mają przedstawiciele krajowych służb ochrony danych, którzy bardzo niechętnie czynią jakiegokolwiek odstępstwa od ogólnie przyjętych zasad.²⁵) Wydaje się w związku z tym, że w celu uniknięcia potencjalnych konfliktów z zaleceniami RE, zwłaszcza po ich oficjalnym przyjęciu przez kraje, należałoby każdorazowo w przepisach prawnych wprowadzających badania

²⁴) Patrz „Zalecenie...” zamieszczone w aneksie.

²⁵) Występujący w dyskusji w podobnej sprawie, na jednym z posiedzeń Grupy, autor niniejszej pracy spotkał się ze stanowczą repliką, że Grupa nie może zbyt liberalnie podchodzić do wniosków statystyków w sprawie rozszerzania możliwości wykorzystywania danych, zwłaszcza jeśli wnioski te pochodzą od przedstawicieli krajów, które nie mają wystarczająco długich tradycji w dziedzinie ochrony danych.

statystyczne, w których zebrane informacje będą wykorzystywane również do aktualizacji rejestrów urzędowych, wyraźnie o tym celu poinformować, określając dokładnie, które dane będą wykorzystywane do aktualizacji rejestrów.

Należy również zauważyć, że w protokole z 9 posiedzenia Biura Grupy Projektowej do Spraw Ochrony Danych Rady Europy przyjęto stwierdzenie, że „uwzględniając uwagę eksperta z Polski, Biuro uzgodniło, aby w Memorandum wyjaśniającym (*Explanatory Memorandum*) wskazać, że korygowanie zbiorów utrzymywanych w celach statystycznych za pomocą danych osobowych zbieranych do innych celów statystycznych samo w sobie nie stanowi celu niezgodnego z tymi innymi celami”.²⁶) Uwaga ta (przyjęta w bardzo skomplikowanym sformułowaniu prawniczym) nie w pełni uwzględnia zgłaszane przez przedstawiciela GUS sugestie, stanowi jednak pewien krok naprzód, bowiem sformułowania poprzednie zabraniały wykorzystywania do korygowania zbiorów informacji statystycznej również danych pochodzących z innych badań statystycznych, co oznaczałoby przykładowo, że zbiorów danych pochodzących z badania budżetów rodzinnych nie można by korygować na podstawie danych pochodzących z badań prowadzonych w zakładach pracy lub z innych badań, np. ankietowych.

Pewne utrudnienia praktyczne mogą zaistnieć także w związku z p. 9 „Zaleceń”, dotyczącym zbierania danych, ustalającym, że: *w przypadku gdy w badaniach statystycznych zbierane i przetwarzane są dane identyfikacyjne, winne być one oddzielone i przechowywane oddzielnie od pozostałych danych osobowych, z wyjątkiem przypadków, gdy jest to jawnie niecelowe lub niepraktyczne.* W polskiej ustawie o statystyce publicznej, w art. 35 ust. 4 zakazuje się *gromadzenia w jednym zbiorze lub w bazie danych wszystkich zebranych przez służby statystyki publicznej danych osobowych dotyczących konkretnej osoby fizycznej, pochodzących z różnych badań statystycznych, które w połączeniu mogą służyć do charakterystyki lub oceny osoby.* Jednocześnie w ust. 5 tegoż artykułu ustala się, iż *opracowywanie danych osobowych i wprowadzanie ich do baz danych systemów informatycznych służb statystyki publicznej następuje z pominięciem danych dotyczących imienia i nazwiska oraz adresu zamieszkania.*

Przytoczone wyżej przepisy (art. 35 ustawy o statystyce publicznej) regulują nieco inne sytuacje niż wspomniany punkt 9 „Zaleceń” odnoszący się do zbierania danych, w którym zakazuje się przechowywania danych osobowych łącznie z ich identyfikatorami, a jeśli takie dane muszą być przechowywane przez dłuższy czas (np. w przypadku badań panelowych), to informacje identyfikujące powinny być przechowywane oddzielnie od pozostałych danych. Informacje identyfikujące mogą być również wykorzysta-

²⁶) *Report of the 9th Meeting of the Bureau, Strasbourg, 5—8 March 1996.*

tywane do tworzenia zbiorów adresowych do celów statystycznych (np. jako operat losowania lub wykaz kontrolny respondentów), ale tylko w przypadkach, gdy prawo krajowe na to pozwala oraz gdy podmioty są o tym poinformowane i wyrażają na to zgodę, i nie zgłaszają sprzeciwu. Autorzy „Zaleceń” zdają sobie sprawę, że wymagania wyrażenia świadomej zgody mogą być trudne do zrealizowania, biorąc pod uwagę cel i wielkość badania. W niektórych przypadkach zgodę podmiotu można domniemywać, jeśli podmiot został poinformowany o zbieraniu, przetwarzaniu i wykorzystywaniu danych o nim oraz jeśli nie zgłosił sprzeciwu. Ten sprzeciw może być wyrażony na piśmie lub w sposób domniemany, np. poprzez odmowę udziału w badaniu. Akt wyrażenia zgody nie może być domniemany, kiedy przepisy prawa wymagają wyrażenia zgody na piśmie, a więc przykładowo, gdy dane osobowe dotyczą tzw. danych wrażliwych, określonych w art. 6 Konwencji 108 (dane dotyczące pochodzenia rasowego, poglądów politycznych i religijnych oraz innych wierzeń, jak również dane osobowe dotyczące stanu zdrowia, życia płciowego lub popełnionych przestępstw).

W niektórych krajach (np. w Danii) statystyka oficjalna opiera się głównie na danych osobowych, które, zgodnie z przepisami prawa krajowego, są zbierane dla celów pozastatystycznych. Dane potrzebne do opracowań statystycznych są łączone na poziomie danych jednostkowych za pomocą nru identyfikacyjnego, który, zgodnie z definicją, jest traktowany jako dana identyfikująca. W tych przypadkach, ze względu na ułatwienie przetwarzania, bardziej właściwe jest przechowywanie identyfikatorów łącznie z pozostałymi danymi pod warunkiem, że służby statystyczne zapewnią właściwe środki ochrony danych oraz że prawo krajowe na to pozwala.

We wszystkich przypadkach, gdy w „Zaleceniu” istnieje sformułowanie „w przypadku, gdy prawo krajowe na to pozwala”, to w odniesieniu do krajów członków UE oznacza to również akty prawne wydawane przez Unię w zakresie jej kompetencji.

Autorzy „Zaleceń” przyznają również, że w wielu przypadkach wymagane oddzielne przechowywanie informacji identyfikujących od pozostałych danych może być trudne. Przykładowo, dotyczy to przypadków, gdy wszystkie te dane umieszczone zostały na jednym kwestionariuszu. Sugestie autorów „Zaleceń” zmierzają do takiego projektowania układu kwestionariuszy, aby oddzielenie to było możliwe (np. **przez stosowanie kwestionariuszy z wyodrębnioną częścią identyfikacyjną łatwą do oderwania**). Zwraca się również uwagę, że w niektórych przypadkach zachowanie pełnych kwestionariuszy, łącznie z częścią identyfikacyjną, może być uzasadnione np. celami badań historycznych, potrzebą ich wykorzystania w sprawach sądowych, w których stroną pozwaną byłyby służby statystyczne lub w przypadku potrzeby doprowadzania danych do porównywalności na skutek zmian w metodologii badań. W każdym

przypadku służby statystyczne zobowiązane są zapewnić wszelkie konieczne warunki ochrony tych danych przed nie upoważnionym dostępem i ujawnieniem.

3.4. Przekazywanie poufnych danych do Eurostatu

Omawiając prawne aspekty ochrony danych i zachowania tajemnicy statystycznej nie sposób pominąć omówienia sposobu postępowania przy przesyłaniu poufnych danych statystycznych z poszczególnych krajów do Urzędu Statystycznego UE — Eurostatu.

W miarę przybliżania się terminu wstąpienia Polski do UE, zakres przekazywanych danych przez GUS do Eurostatu będzie się rozszerzał; nie od rzeczy jest więc przedstawienie obowiązujących w tej sprawie przepisów i zasad postępowania.

Zagadnienie przekazywania danych podlegających ochronie, ze względu na przepisy tajemnicy statystycznej, reguluje rozporządzenie Rady z dnia 11 czerwca 1990 r.²⁷⁾

W preambule rozporządzenia stwierdza się, że Urząd Statystyczny EWG (obecnie Eurostat UE) zobowiązany jest zapewnić analogiczne gwarancje poufności danych, jak urzędy krajowe, a pracownicy zatrudnieni w urzędach Wspólnoty są zobowiązani do nieujawniania, uzyskanych w czasie ich zatrudnienia, informacji również po zaprzestaniu pracy. Takie ustalenie powinno w zasadniczy sposób rozwiązać trudności związane z przesyłaniem chronionych danych do Eurostatu.²⁸⁾

Zgodnie z definicją przytoczoną w omawianym rozporządzeniu, poufnymi danymi statystycznymi są dane uznane za poufne przez kraje członkowskie na podstawie krajowych przepisów prawnych lub praktycznych zasad postępowania dotyczących tajemnicy statystycznej. Poufnymi z samej definicji są informacje o prywatnym życiu osób fizycznych oraz ich rodzin. Wykorzystanie danych do celów statystycznych oznacza, iż dane mogą być wykorzystane wyłącznie do opracowania tablic statystycznych lub do analiz i że dane te nie mogą być wykorzystywane do celów administracyjnych, prawnych, podatkowych lub do kontroli prowadzonych wobec badanych jednostek.

Rozporządzenie Rady upoważnia władze krajowe do przekazywania danych poufnych do Eurostatu, z tym że powinno być ono dokonywane w taki sposób, aby badane jednostki nie mogły być zidentyfikowane

²⁷⁾ Council Regulation (Euratom, EEC) No 1588/90 of 11 June 1990 on the transmission of data subject to statistical confidentiality to Statistical Office of the European Communities, Official Journal No L 151/1.

²⁸⁾ Omawiany przepis prawny reguluje przekazywanie danych podlegających ochronie ze względu na przepisy o tajemnicy statystycznej, nie dotyczy natomiast przesyłania informacji podlegających ochronie na mocy specjalnych przepisów Wspólnoty lub przepisów krajowych.

bezpośrednio. Krajowe instytucje nie mogą być zwłaszcza zobowiązane do przekazywania danych odnoszących się do osobistego życia osób fizycznych, jeśli na podstawie przekazywanych informacji istniałaby możliwość identyfikacji tych osób w sposób bezpośredni lub pośredni.

Jednocześnie rozporządzenie zobowiązuje Komisję, a zwłaszcza Dyrektora Generalnego Urzędu Statystycznego Wspólnoty do podjęcia wszelkich niezbędnych środków prawnych, administracyjnych, technicznych i organizacyjnych, zapewniających zachowanie poufności przekazywanych danych z krajów członkowskich. Obowiązek ten dotyczy zarówno danych przekazywanych zgodnie z obowiązkami ciążącymi na krajach zgodnie z przepisami prawnymi Wspólnoty, jak i danych przekazywanych dobrowolnie przez kraje członkowskie do Wspólnoty.

Wewnątrz Eurostatu obowiązuje przestrzeganie odpowiednich reguł zapewniających ochronę gromadzonych tam zbiorów informacji. Za zachowanie integralności i ochronę każdego zbioru odpowiada wyznaczona osoba. Każdy zbiór musi być zarejestrowany, a każde udostępnienie informacji musi być odnotowane w odpowiednim dokumencie [Nan94].

Dostęp do danych poufnych Eurostatu mogą mieć wyłącznie jego pracownicy o statusie funkcjonariuszy (officials). Wszelkie operacje na zbiorach danych mogą być dokonywane wyłącznie w pomieszczeniach znajdujących się na obszarze chronionym i dane mogą być wykorzystywane wyłącznie do celów statystycznych. Komisja może zezwolić w szczególnych przypadkach na dostęp do tych danych innym pracownikom Urzędu Statystycznego UE oraz innym osobom fizycznym pracującym w Eurostacie na zasadach kontraktowych, ale również wyłącznie w celach opracowań statystycznych.

Posiadane przez Eurostat dane poufne mogą być rozpowszechniane wyłącznie w połączeniu z innymi danymi, w formie uniemożliwiającej bezpośrednią lub pośrednią identyfikację jednostek badanych.

Omawiane rozporządzenie powołało do życia Komitet do Spraw Tajemnicy Statystycznej, składający się z przedstawicieli wszystkich krajów członkowskich. Pracą Komitetu kieruje Dyrektor Generalny Eurostatu. Zadanie Komitetu polega na wypracowaniu przedsięwzięć zapewniających przestrzeganie przepisów dotyczących ochrony danych.

Z tego co powiedziano wyżej wynika, że przy przekazywaniu danych do Eurostatu przez służby statystyczne w Polsce, powinny być stosowane analogiczne zasady, jak przy udostępnianiu danych na terenie kraju. Dotyczy to zarówno danych dotyczących osób, jak i danych o działalności gospodarczej podmiotów.

3.5. Zapewnienie ochrony danych jako warunek uzyskiwania prawdziwych informacji

Obok przepisów prawnych nakładających na statystyków obowiązek zachowania tajemnicy statystycznej i zapewnienia bezpieczeństwa danych,

statystycy w całym procesie zbierania, opracowania, przechowywania danych i udostępniania informacji muszą kierować się również przesłankami mającymi charakter pragmatyczny. Chodzi mianowicie o znalezienie najbardziej skutecznych sposobów zapewnienia prawdziwej i rzetelnej informacji.

Stworzenie rzetelnego i obiektywnego systemu informacji statystycznej wymaga, oczywiście, zastosowania całego szeregu środków profesjonalnego prowadzenia badań, począwszy od określenia celów badań, opracowania poprawnej metodologii, zastosowania właściwej organizacji i środków opracowania informacji itp. Niemniej jednak najważniejsze i najtrudniejsze zadanie stojące przed statystykami w całym procesie zbierania i opracowania danych polega na uzyskaniu od badanych jednostek kompletnej, prawdziwej i zgodnej ze stanem faktycznym informacji jednostkowej.

Żadne, nawet najbardziej wymyślne i nowoczesne metody i technologie nie są w stanie zapewnić dobrej jakościowo informacji, jeśli dane źródłowe wpływające do systemu informacyjnego statystyki będą niekompletne, niedokładne lub celowo zniekształcone.

W nowej ustawie wprowadzono bardziej zaostrzone kary i zmianę kwalifikacji czynów związanych z przekazywaniem nieprawdziwych danych lub odmową wykonania obowiązków statystycznych. O ile w poprzednio obowiązującej ustawie czyny te były traktowane jako wykroczenia i podlegały orzecznictwu kolegów do spraw wykroczeń, to w świetle aktualnie obowiązującej ustawy czyny te traktowane są jako przestępstwa, a ich rozpatrywanie należy do kompetencji prokuratorów, a orzekanie do sądów powszechnych. Rola służb statystyki publicznej polega na zebraniu dowodów pozwalających na stwierdzenie podejrzenia popełnienia przestępstwa i skierowaniu sprawy do prokuratury [Pup95]. Tylko przypadki opóźnień w przekazywaniu danych służby statystyczne będą mogły kierować do kolegów do spraw wykroczeń, które w stosunku do sprawców orzekają kary grzywny.

W ramach prowadzenia uporczywej walki o prawdziwą i wiarygodną informację statystycy powinni oczywiście umiejętnie wykorzystywać nowe przepisy karne przewidziane w nowej ustawie, niemniej jednak trzeba zdawać sobie sprawę z tego, że są to przepisy dość trudne do wyegzekwowania, zwłaszcza w świetle ogromnego wzrostu liczebności badanych podmiotów i w związku z tym mają ograniczoną skuteczność.

W sprawach dyscypliny statystycznej badanych jednostek przepisy poprzedniej ustawy o statystyce państwowej²⁹⁾ dawały służbom statystycznym o wiele większe uprawnienia. Upoważniały one służby statystyczne do prowadzenia kontroli legalności, rzetelności, prawidłowości i terminowości sprawozdawczości statystycznej oraz do dokonywania ocen poprawności

²⁹⁾ Ustawa z dnia 26 lutego 1982 r. o statystyce państwowej, Dz. U. Nr 32, poz. 171.

systemów ewidencyjno-informacyjnych. Przy wykonywaniu czynności kontrolnych służby statystyczne były uprawnione do wglądu do sprawozdań, ewidencji podstawowej i innych dokumentów stanowiących materiał źródłowy danych podawanych w sprawozdaniach statystycznych (art. 22.1). Również ustawa z dnia 22 maja 1989 r. o zmianie ustawy o statystyce państwowej³⁰⁾ zachowywała prawo służb statystycznych do kontroli prawidłowości i terminowości wykonywania obowiązków sprawozdawczych (art. 22) oraz do kontroli rzetelności danych w przypadkach, gdy otrzymane dane liczbowe budziły uzasadnione podejrzenie co do ich zgodności ze stanem faktycznym (art. 23).

Przepisy obowiązującej obecnie ustawy z dnia 29 czerwca 1995 r. o statystyce publicznej nie dają upoważnień służbom statystycznym do prowadzenia kontroli poprawności informacji przekazywanych przez badane jednostki. Wynika to nie z chęci ograniczenia wpływu służb statystycznych na kompletność i wiarygodność uzyskiwanych danych, ale z zasadniczych zmian jakie zaszły w sytuacji społeczno-ekonomicznej, z konieczności ograniczenia ingerencji organów państwowych w wewnętrzne sprawy jednostek gospodarczych oraz, przede wszystkim, ze zmiany funkcji służb statystycznych, które

z dominującej poprzednio roli organu administracji państwowej w dziedzinie statystyki zostały przekształcone w służbę publiczną, powołaną do opracowywania i dostarczania informacji zapewniających rzetelne, obiektywne i systematyczne informowanie społeczeństwa, organów państwa i administracji oraz podmiotów gospodarczych o sytuacji ekonomicznej, demograficznej, społecznej oraz środowiska naturalnego (art. 3 ustawy).

Zadania służb statystycznych w dziedzinie zapewnienia uzyskiwania kompletnych i maksymalnie prawdziwych informacji od badanych podmiotów, w warunkach nowej ustawy o statystyce publicznej, nie zmniejszyły się, *muszą natomiast nastąpić zmiany w metodach działalności tych służb w całym systemie oddziaływania na użytkowników informacji oraz na respondentów*, czy będą to jednostki prowadzące działalność gospodarczą, czy gospodarstwa domowe, czy wreszcie indywidualni członkowie naszego społeczeństwa.

Jeśli zgodzimy się co do tego, że w obecnych warunkach ekonomicznych, społecznych i prawnych argumenty siły, którymi dysponowały służby statystyczne, utraciły swoją skuteczność (jeśli w ogóle były kiedykolwiek wystarczająco skuteczne), to *musimy wypracować inny system oddziaływania na otoczenie, w którym funkcjonuje statystyka, aby był on bardziej skuteczny i dający, zwłaszcza w dalszej perspektywie, lepsze gwarancje poprawy informacyjnego systemu statystyki.*

³⁰⁾ Dz. U. Nr 32, poz. 171.

W tym systemie oddziaływania na otoczenie, mającym na celu poprawę warunków uzyskiwania bardziej kompletnych i rzetelnych informacji, najbardziej istotne znaczenie mają dwa fundamentalne zadania:

- 1) zmiana sposobu postrzegania celów i zadań statystyki we współczesnym społeczeństwie oraz
- 2) zmiana stosunku potencjalnych respondentów do uczestnictwa w badaniach statystycznych.

To pierwsze zadanie ma bodaj największe znaczenie, ponieważ opinia o użyteczności i randze społecznej informacji statystycznej ma decydujący wpływ nie tylko na ogólne funkcjonowanie statystyki, ale również na postawy respondentów wobec badań statystycznych.

Brak należytego zrozumienia funkcji i znaczenia informacji statystycznej we współczesnym społeczeństwie występuje zresztą nie tylko w Polsce. Warto w związku z tym przypomnieć inicjatywę grupy naukowców Międzynarodowego Instytutu Statystycznego w sprawie bardziej aktywnego zaangażowania się w proces przeciwdziałania niekorzystnym tendencjom do niedoceniań znaczenia, a nawet degradacji znaczenia statystyki zarówno w dziedzinie gospodarki, jak i w instytucjach badawczych i na uczelniach. Na 49 Sesji MIS w 1991 r. we Florencji przedstawiono specjalny raport grupy naukowców w tej sprawie. Jego główna idea sprowadza się do stwierdzenia, iż

zarówno statystycy, jak i użytkownicy informacji muszą lepiej zrozumieć rolę statystyki we współczesnym zmieniającym się świecie. Rządy (organy administracji) muszą bardziej doceniać znaczenie rzetelnej i poprawnej metodologicznie informacji jako niezbędnej przesłanki właściwego zarządzania.

Kierownicy instytucji badawczych i organizacji gospodarczych muszą opanować podstawy wiedzy statystycznej, aby docenić rosnące jej znaczenie. Społeczeństwo musi zdobyć umiejętność czytania liczb, aby mogło wyciągać świadomie wnioski z informacji statystycznych, udostępnianych za pośrednictwem środków masowego przekazu.³¹⁾

Obok zrozumienia potrzeby upowszechnienia informacji statystycznej, jako jednego z ważnych czynników rozwoju społecznego, pokonanie niechęci respondentów do udziału w badaniach statystycznych i ugruntowanie w nich przekonania do konieczności udzielania szczerych i prawdziwych odpowiedzi na pytania zawarte w programach badań zależy od tego, czy prowadzącym badania statystykom uda się przekonać potencjalnych respondentów, że ich udział w badaniach nie może im przynieść żadnego uszczerbku ani w ich prawach do prywatności, ani w ich interesach zawodowych. Jakiegokolwiek, nawet pojedyncze, przypadki ujawnienia przez służby statystyczne informacji jednostkowych, mogą na długi okres

³¹⁾ Szerzej na ten temat patrz [Wal94].

poderwać zaufanie do tych służb, co natychmiast odbije się na zwiększonej liczbie odmów udzielania informacji, a w tzw. badaniach przymusowych, w których obowiązek udzielania odpowiedzi ustalany jest pod groźbą sankcji w odpowiednich przepisach prawnych, zwiększa się liczba odpowiedzi zawierających dane nierzetelne.

Cala działalność służb statystycznych zarówno w czasie przygotowania i prowadzenia badań, jak i w czasie udostępniania i popularyzacji wyników badań musi zmierzać w kierunku uzyskania przez te służby opinii instytucji zaufania publicznego, która uzyskując przekazywane jej dane wykorzysta je wyłącznie dla dobra wspólnego i nie dopuści do wykorzystania powierzonych jej danych przeciwko interesom indywidualnych podmiotów.

Na temat konieczności zachowania pełnej poufności indywidualnych danych uzyskiwanych zarówno w bezpośrednich badaniach statystycznych, jak i w drodze przyjmowania takich danych z innych systemów informacyjnych w formie tzw. danych administracyjnych istnieje bardzo obszerna literatura. Sprawie tej poświęcało wiele uwagi liczne grono znanych statystyków. Problem ten poruszany był zwłaszcza bardzo często w latach sześćdziesiątych i siedemdziesiątych, tj. w okresie początków rozwoju komputerowych baz i banków danych,³²⁾ kiedy to statystycy zaczęli dostrzegać niebezpieczeństwo ujawnienia danych indywidualnych gromadzonych w skomputeryzowanych systemach informacyjnych i płynące stąd zagrożenie dla możliwości uzyskiwania rzetelnych informacji od respondentów.

Literatura ta, głównie ze względu na barierę językową, jest mało w Polsce znana. W tym okresie, kiedy problemy te były traktowane w kręgach międzynarodowych jako problemy decydujące o kształtowaniu się stosunków zaufania między statystykami i społeczeństwem, warunkujących możliwość uzyskiwania wiarygodnych informacji jednostkowych, w naszym kraju ze względu na zasadniczo odmienną sytuację społeczno-polityczną i inne funkcje statystyki problemy te miały nieco inny charakter.

W okresie, gdy jedną z podstawowych funkcji statystyki (w dziedzinie statystyki gospodarczej) była funkcja kontroli wykonania planów, dane statystyczne stanowiły narzędzie bezpośredniego oddziaływania władz gospodarczych i politycznych na poszczególne przedsiębiorstwa. Z nie tak dawnej polskiej praktyki lat 1983—1990 pamiętamy publikowanie w prasie danych GUS, np. o „500 największych przedsiębiorstwach przemysłu przetwórczego”,³³⁾ „Przedsiębiorstwach budowlano-montażowych”,³⁴⁾ „Przedsiębiorstwach handlu zagranicznego”.³⁵⁾

³²⁾ Por. np. [Nor68], [Bac69], [Fel72].

³³⁾ Por. „Zarządzanie” 1984, nr 6.

³⁴⁾ Por. „Zarządzanie” 1985, nr 10.

³⁵⁾ Por. „Rynki Zagraniczne”, lata 1985—1990.

Również wojewódzkie urzędy statystyczne publikowały indywidualne dane o działalności największych na terenie danego województwa przedsiębiorstw. Warto zauważyć, że tego rodzaju publikacje spotykały się z bardzo pochlebnymi recenzjami w prasie fachowej i nie były kwestionowane przez same zainteresowane przedsiębiorstwa.³⁶⁾ Sytuacja w tej dziedzinie uległa jednak od tego czasu radykalnej zmianie. Wśród przedsiębiorstw, w owym czasie niemal wyłącznie państwowych, pojawiły się podmioty prywatne, w tym bardzo dużo osób fizycznych prowadzących działalność gospodarczą, zastrzyżyły się warunki konkurencji między przedsiębiorstwami, na skutek czego niektóre dane słusznie traktowane są jako „wrażliwe” z punktu widzenia interesów jednostek gospodarczych. Praktyka publikowania wspomnianych wyżej danych jednostkowych, z którą GUS ostatecznie zerwał w 1990 r., wytworzyła jednak w świadomości niektórych użytkowników przekonanie, że służby statystyczne powinny nadal opracowywać i dostarczać tego rodzaju dane oraz że w ich przekonaniu nie ma w tym nic niewłaściwego.

Do dziś można spotkać się z zarzutami, że nadal, mimo zapewnień i jednoznacznych przepisów prawnych, niektóre czasopisma publikują indywidualne dane charakteryzujące wyniki działalności przedsiębiorstw. Osoby formułujące takie zarzuty nie wiedzą zapewne, że dane te **nie pochodzą z badań statystycznych i że służby statystyczne nie mają z tym nic wspólnego**, natomiast publikowanie danych, jeśli nie są one uzyskiwane w formie badań statystycznych oraz jeśli podmioty, których dane te dotyczą wyrażają na to zgodę, nie jest zabronione.

Z drugiej strony, w świadomości niektórych potencjalnych respondentów do dnia dzisiejszego pozostał brak przekonania, czy podane lub przekazane w badaniach statystycznych dane będą rzeczywiście podlegać pełnej ochronie i nie zostaną ujawnione lub wykorzystane ze szkodą dla ich interesów. Dlatego tak duże znaczenie ma stałe i uporczywe wyjaśnianie celów prowadzenia badań statystycznych, z których jasno wynika, że:

statystyka zajmuje się badaniami masowych zjawisk i procesów, a dane jednostkowe, będące obiektem obserwacji wykorzystywane są wyłącznie do obliczania grupowych charakterystyk badanych populacji, co oznacza, że dane jednostkowe nie mogą występować w charakterze wyniku badania, nie mogą być nikomu udostępnione oraz nie mogą być wykorzystane do innych celów niż opracowania statystyczne. Nie mogą być więc wykorzystane dane jednostkowe, uzyskane od respondentów w formie badań statystycznych, do ich konfrontacji z danymi przekazywanymi władzom podatkowym, nie mogą stanowić dowodu w dochodzeniach prokuratorskich, w przewodach sądowych itp.

³⁶⁾ Należy podkreślić, że publikowanie danych jednostkowych dotyczyło wyłącznie danych indywidualnych o działalności gospodarczej przedsiębiorstw „uspołecznionych”. Organy statystyczne bardzo skrupulatnie zawsze przestrzegały natomiast zasad pełnej ochrony danych osobowych.

Z tego również względu urzędy statystyczne niechętnie publikują w swoich wydawnictwach indywidualne informacje pochodzące z innych źródeł niż badania statystyczne. Jak wiadomo, zgodnie z określonymi przepisami, np. spółki akcyjne publikują bardzo szczegółowe bilanse charakteryzujące znacznie bardziej szczegółowo sytuację ekonomiczną tych jednostek aniżeli dane pochodzące z badań statystycznych. Mimo iż zgodnie z obowiązującym prawem nic nie stoi na przeszkodzie, aby służby statystyczne publikowały dane opublikowane wcześniej gdzie indziej, na podstawie innych przepisów prawnych niż ustawa o statystyce, służby statystyczne, z zasady, tego unikają, aby nie narażać się, w opinii osób nie zorientowanych w tajnikach obowiązujących przepisów, na zarzut, że w publikacjach statystycznych ujawnia się jakiegokolwiek indywidualne dane.

W związku z kierowanymi dość często do służb statystycznych, ze strony organów administracji państwowej i samorządowej, postulatami, dotyczącymi udostępniania indywidualnych danych o przedsiębiorstwach, dla których organy te są organami założycielskimi, należy podkreślić także, że nawet jeśliby założyć (czysto teoretycznie), że takie dane służby statystyczne im udostępnią, to organom tym również nie wolno byłoby wykorzystywać tych danych do innych celów niż zbiorcze opracowania i analizy statystyczne i że danych tych nie wolno wykorzystywać do żadnych decyzji i ocen dotyczących zainteresowanych podmiotów. W tej sprawie przepisy ustawy o statystyce są jednoznaczne i obowiązują w jednakowy sposób zarówno służby statystyczne, jak i wszystkie inne instytucje i osoby prywatne.

4. MIEJSCA I PRZYCZYNY ZAGROŻEŃ DLA BEZPIECZEŃSTWA DANYCH

4.1. Ważniejsze etapy badań i opracowań statystycznych

Charakterystyczną cechą systemu zbierania i opracowania danych statystycznych, wpływającą w istotnym stopniu na ochronę i bezpieczeństwo tego systemu, jest jego rozległość tematyczna, różnorodność źródeł informacji, zróżnicowanie form badań i opracowań oraz duże rozproszenie miejsc powstawania i rejestracji danych.

Osoby odpowiedzialne za funkcjonowanie systemu informacji statystycznej, a zwłaszcza projektanci systemu muszą uwzględniać te aspekty, aby zapewnić najbardziej efektywne systemy zabezpieczeń przed zniszczeniem, uszkodzeniem i ujawnieniem podlegającej ochronie informacji.

Problemy ochrony informacji powinny stanowić część składową każdego projektu systemu badania na równi z innymi elementami, takimi jak: przygotowanie metodologii badania, opracowanie zasad kontroli i korekty oraz zasad przetwarzania i rozpowszechniania wyników. Zastosowanie konkretnych metod ochrony musi uwzględniać stopień zagrożenia występujący z różnym nasileniem na różnych etapach badań statystycznych.

Z punktu widzenia zapewnienia ochrony danych w systemach informacyjnych statystyki wyróżniono ważniejsze etapy w wykr. 1.

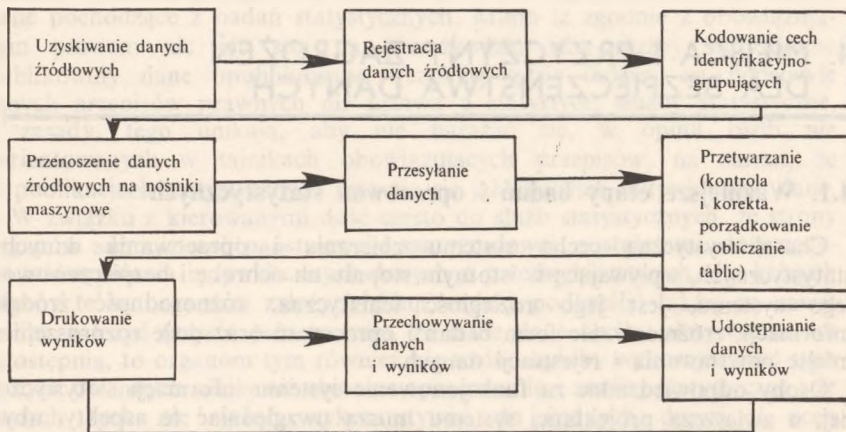
W znacznej większości tematów badań statystycznych proces realizacji badania zaczyna się od uzyskania danych źródłowych niezbędnych do realizacji danego celu badawczego określonego w dokumentacji projektowej badania. Dane uzyskuje się z bardzo różnorodnych źródeł, przy czym w miarę rozwoju metod badań i doskonalenia technologii informacyjnych obserwujemy w tej dziedzinie bardzo istotne zmiany.³⁷⁾

Na obecnym etapie rozwoju badań statystycznych w Polsce nośnikami danych statystycznych mogą być: sprawozdania, ankiety przedsiębiorstw, ankiety wypełniane przez gospodarstwa domowe i indywidualne osoby,

³⁷⁾ Są to problemy niezwykle istotne dla przyszłego rozwoju statystyki. Ich omówienie nie mieści się jednak w ramach tego opracowania.

różne kwestionariusze spisowe, kwestionariusze różnych badań reprezentacyjnych, dokumenty administracyjne, rejestry itp.

Wykres 1 WAŻNIEJSZE ETAPY FUNKCJONOWANIA SYSTEMU INFORMACYJNEGO STATYSTYKI



4.2. Zagrożenia dla bezpieczeństwa danych na wstępnych etapach realizacji badań

Tej różnorodności nośników danych towarzyszy bardzo duże zróżnicowanie sposobów zbierania i przesyłania danych z miejsc ich powstawania do miejsc przetwarzania. Ilustruje to wykres 2.

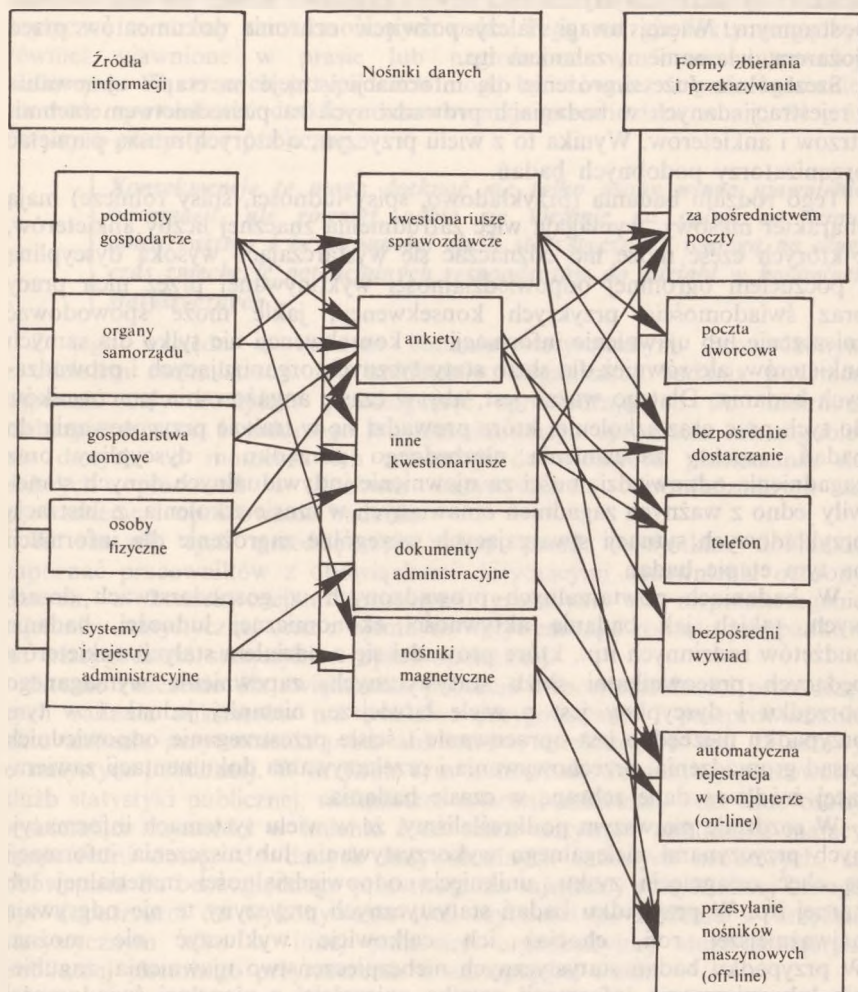
Źródłowe dane statystyczne mogą wpływać do systemu statystycznego za pośrednictwem poczty (także poczty dworcowej), mogą być dostarczane do służb statystycznych bezpośrednio przez jednostki badane (głównie z miejscowości, w których znajduje się siedziba służb statystycznych prowadzących badanie). Niektóre rodzaje danych, a zwłaszcza wyjaśnienia i ewentualne korekty mogą być przyjmowane przez telefon. Szereg danych zbiera się metodą bezpośrednich wywiadów prowadzonych przez ankieterów i rachmistrzów odwiedzających jednostki badane w miejscu ich pracy lub w mieszkaniu.

Nośniki maszynowe mogą być przesyłane bezpośrednio do komputerów poprzez linie transmisji danych lub sieci komputerowe (on line), bądź też metodą spedycji jako przesyłki pocztowe lub przesyłane w inny sposób.

Przedstawione wyżej w sposób schematyczny operacje uzyskiwania danych w badaniach statystycznych zasługują na szczególną uwagę z punktu widzenia bezpieczeństwa danych, zarówno jeśli chodzi o ochronę przed

zagubieniem lub zniszczeniem, jak i ze względu na zachowanie tajemnicy statystycznej.

Wykres 2 ŹRÓDŁA, NOŚNIKI ORAZ FORMY ZBIERANIA DANYCH



Na tych początkowych etapach istnieje szczególnie duże niebezpieczeństwo zniszczenia, zagubienia oraz ujawnienia danych i naruszenia tajemnicy statystycznej. Więcej uwagi należy więc poświęcić na uporządkowanie

i unormowanie zasad przyjmowania dokumentów, ich rejestracji i przygotowania do przetwarzania, obiegu w czasie przetwarzania i przechowywania po zakończeniu przetwarzania, zapewniając niezbędną dyscyplinę i porządek w taki sposób, aby dostęp do dokumentów mieli wyłącznie zainteresowani pracownicy służb statystycznych i aby uniemożliwić dostęp do nich osobom postronnym. Więcej uwagi należy poświęcić ochronie dokumentów przed pożarem, włamaniem, zalaniem itp.

Szczególnie duże zagrożenie dla informacji istnieje na etapie ujmowania i rejestracji danych w badaniach prowadzonych za pośrednictwem rachmistrzów i ankieterów. Wynika to z wielu przyczyn, o których muszą pamiętać organizatorzy podobnych badań.

Tego rodzaju badania (przykładowo, spisy ludności, spisy rolnicze) mają charakter masowy, wymagają więc zatrudnienia znacznej liczby ankieterów, z których część może nie odznaczać się wystarczająco wysoką dyscypliną i poczuciem ogromnej odpowiedzialności wykonywanej przez nich pracy oraz świadomością przykrych konsekwencji, jakie może spowodować zniszczenie lub ujawnienie informacji — konsekwencji nie tylko dla samych ankieterów, ale również dla służb statystycznych organizujących i prowadzących badania. Dlatego ważne jest, aby w czasie angażowania pracowników do tych prac oraz szkolenia, które prowadzi się w trakcie przygotowania do badań, aspekty zapewnienia niezbędnego porządku i dyscypliny oraz zagadnienia odpowiedzialności za ujawnienie indywidualnych danych stanowiły jedno z ważnych zagadnień omawianych w czasie szkolenia, z ilustracją przykładowych sytuacji stwarzających szczególne zagrożenie dla informacji na tym etapie badań.

W badaniach powtarzalnych prowadzonych w gospodarstwach domowych, takich jak badanie aktywności ekonomicznej ludności, badanie budżetów rodzinnych itp., które prowadzi się z udziałem stałych ankieterów będących pracownikami służb statystycznych, zapewnienie wymaganego porządku i dyscypliny jest o wiele łatwiejsze, niemniej jednak i w tym przypadku niezbędne jest opracowanie i ścisłe przestrzeganie odpowiednich zasad gromadzenia, przechowywania i przekazywania dokumentacji zawierającej źródłowe dane zebrane w czasie badania.

W rozdziale pierwszym podkreślaliśmy, że w wielu systemach informacyjnych przyczynami nielegalnego wykorzystywania lub niszczenia informacji są chęć osiągnięcia zysku, uniknięcia odpowiedzialności materialnej lub karnej itp. W przypadku badań statystycznych przyczyny te nie odgrywają najważniejszej roli, chociaż ich całkowicie wykluczyć nie można. W przypadku badań statystycznych niebezpieczeństwo ujawnienia, zagubienia lub zniszczenia informacji wynika najczęściej z niepełnej świadomości konsekwencji tego czynu, niedbałości, skłonności do gadulstwa i plotkowania, chęci popisania się wiadomościami, niechęci lub wrogości do badanych jednostek czy osób należących do innych grup etnicznych lub narodowościowych itp.

Ujawnienie informacji w miejscu lub blisko miejsca ich zbierania jest szczególnie niepożądane również ze względu na to, że może ono być przyczyną określonych szkód moralnych lub materialnych dla osób badanych i znanych w danym środowisku. Ujawnienie informacji w środowisku lokalnym może również spowodować natychmiastową reakcję krytyczną ze strony zainteresowanych osób. Tego rodzaju fakty mogą być również ujawnione w prasie lub nagłościone w formie interwencji radiowych. W rażących przypadkach mogą być również ścigane sędownie, co może spowodować dotkliwe konsekwencje przewidziane w art. 54 i 55 ustawy o statystyce publicznej.

Konsekwencje te mogą dotknąć nie tylko osoby winne ujawnienia informacji, ale również odbić się ujemnie na opinii na temat obiektywizmu i bezstronności służb statystycznych i mogą na długi czas zniechęcić potencjalnych respondentów do udziału w badaniach statystycznych.

Z tego względu w każdym badaniu statystycznym, prowadzonym z udziałem rachmistrzów lub ankierów, organizator badania powinien zapewnić cały kompleks przedsięwzięć sprowadzających do minimum niebezpieczeństwo świadomego lub nieświadomego ujawnienia lub zagubienia danych od momentu ich rejestracji do momentu przekazania do dalszych opracowań w służbach statystycznych odpowiedzialnych za prowadzenie i opracowanie badania.

W ramach tych przedsięwzięć należy przede wszystkim dokładnie zapoznać pracowników z obowiązkami dotyczącymi zapewnienia ochrony danych, z konsekwencjami prawnymi grożącymi za nieprzestrzeganie tajemnicy statystycznej oraz z ujemnym wpływem tego rodzaju przypadków na wiarygodność statystyki w opinii społecznej.

Istotne znaczenie dla zwiększenia poczucia odpowiedzialności prowadzących badania ankierów ma właściwe przygotowanie i przeprowadzenie aktu złożenia przyrzeczenia przez ankierów, ustalonego w art. 12 ustawy o statystyce publicznej. W artykule tym stanowi się, że wszyscy pracownicy służb statystyki publicznej, rachmistrze spisowi, ankierzy oraz inne osoby wykonujące czynności w imieniu i na rzecz statystyki publicznej, mający bezpośredni dostęp do danych indywidualnych i danych osobowych, są obowiązani do bezwzględnego przestrzegania tajemnicy statystycznej i mogą być dopuszczeni do wykonywania tych czynności po złożeniu w urzędzie statystycznym albo w innej jednostce organizacyjnej służb statystyki publicznej pisemnego przyrzeczenia o następującej treści:

Przyrzekam, że będę wykonywać swoje prace na rzecz statystyki publicznej z całą rzetelnością, zgodnie z etyką zawodową statystyka, a poznane w czasie ich wykonywania dane jednostkowe zachowam w tajemnicy wobec osób trzecich.

Złożenie przyrzeczenia przez ankierów nie powinno sprowadzać się do formalnego przekazania do podpisu tekstu przyrzeczenia jako jednego z wielu papierków, które angażowany do pracy ankier podpisuje. Powinno się ono odbyć w formie, która pracownikowi utkwi na dłuższy czas w pamięci i zostanie w jego świadomości jako ważne zobowiązanie, któremu pracownik nie może się sprzeniewierzyć.

Przy tej okazji należałoby zasygnalizować potrzebę opracowania i przyjęcia w naszym kraju kodeksu lub deklaracji w sprawie etyki zawodu statystyka. Przywołane w przyrzeczeniu pojęcie etyki zawodowej ma na uwadze „Deklarację w sprawie etyki zawodowej statystyków” opracowaną i przyjętą przez Międzynarodowy Instytut Statystyczny (patrz p. 3.2). Nie wszyscy obywatele polscy mogą czuć się w pełni zobowiązani do przestrzegania postanowień tego dokumentu. W kilku krajach, niezależnie od Deklaracji MIS, opracowano własne „krajowe” kodeksy lub deklaracje.

Opracowanie podobnego „kodeksu” w Polsce, przykładowo, wspólnie przez GUS i Polskie Towarzystwo Statystyczne lub pod egidą Rady Statystyki, można by wykorzystać do przeprowadzenia szerszej kampanii informacyjnej na temat problemów funkcjonowania statystyki publicznej, potrzeby pogłębienia społecznego zaufania do statystyki oraz podniesienia rangi statystyki w społeczeństwie.

Potrzeba opracowania i przyjęcia polskiego kodeksu etyki zawodu statystyka (lub karty zasad etycznych statystyków) wynika również z projektu zaleceń Rady Europy, które w p. 15 (patrz aneks) stwierdzają, iż specjaliści statystycy, firmy oraz instytucje zajmujące się opracowaniami statystycznymi winni przyjąć i opublikować kodeks etyczny, który odpowiada regułom ustalonym w zaleceniu.

Kolejną ważną sprawą, zapewniającą bezpieczne zebranie i przekazanie materiałów źródłowych badania, jest dokładne określenie zasad postępowania z dokumentacją badania. Przykładem dobrze opracowanych zasad są przepisy „instrukcji w sprawie metodologii spisu ludności i mieszkań w 1995 r.”, określające dokładnie zarówno czynności i sposób postępowania rachmistrza w czasie przeprowadzenia spisu, jak i jego obowiązki w dziedzinie porządkowania, kontroli kompletności i sposobów przekazywania kompletnej dokumentacji badania upoważnionym służbom.

Pewną lukę w tych ustaleniach stanowi brak określenia sposobu postępowania z dokumentacją w trakcie badania. W przypadkach gdy badanie prowadzone jest przez kilka dni powinny być ściśle ustalone zasady przechowywania dokumentów badania, a zwłaszcza wypełnionych ankiet. Jeśli opracowanie takich zasad w sposób jednolity dla całego kraju jest niemożliwe lub niecelowe ze względu na różne warunki i sytuacje lokalne, instrukcja ogólna powinna zobowiązywać służby terenowe (urzędy statystyczne, władze lokalne) do określenia tych zasad.

Dla stworzenia właściwych warunków ochrony zbieranych informacji powinno się zwłaszcza określić sposób przechowywania dokumentacji w czasie badania, np. w wyznaczonych punktach lokalnych (urzędach gmin lub miast, urzędach statystycznych itp.) w taki sposób, aby uniemożliwić dostęp do dokumentów przez osoby postronne.

W zasadzie nie należy dopuszczać do przechowywania źródłowych materiałów badań w mieszkaniach prywatnych ankierów. W warunkach rodzinnych trudno zagwarantować pełną ochronę danych. W przeszłości zdarzały się (na szczęście sporadyczne) przypadki zniszczenia lub uszkodzenia kwestionariuszy spisu, a nawet pozostawienia ich w lokalach gastronomicznych, w środkach transportu publicznego itp. Niebezpieczeństwo powstawania takich zagrożeń należy akcentować w czasie angażowania i szkolenia ankierów.

Zakaz przechowywania źródłowych materiałów badań w mieszkaniach prywatnych ankierów o wiele trudniej egzekwować w przypadku badań trwających przez dłuższy czas, jak np. w badaniach budżetów rodzinnych czy badaniach wykorzystania czasu. W tych przypadkach jednak należy ustalić indywidualnie z ankierami w jaki sposób i gdzie będą przechowywane częściowo wypełnione kwestionariusze i w jaki sposób zagwarantowana będzie ich jak najlepsza ochrona. Aspekty te, niestety, są z zasady wstydliwie pomijane w dokumentacji związanej z badaniami statystycznymi.

Podobne zasady należy stosować w przypadku wykonywania przez służby statystyczne symbolizacji materiałów lub wprowadzania danych z materiałów źródłowych do mikrokomputerów poza lokalem służbowym. Ten system pracy będzie się upowszechniać, zwłaszcza po wyposażeniu ankierów w przenośne mikrokomputery (lap-topy). W wielu krajach rozpowszechnione są systemy prowadzenia badań metodą wywiadów przez telefon, badań z jednoczesną rejestracją danych na przenośnym komputerze (tzw. metoda CAPI).³⁸⁾ Względy ochrony danych nie mogą hamować wprowadzania nowoczesnych metod i techniki pracy,³⁹⁾ ale jednocześnie nie wolno statystyków od obowiązku zagwarantowania pełnej ochrony informacji. Jednostka odpowiedzialna za dane badanie musi opracować dla każdego tematu badania bardzo szczegółowe zasady postępowania na wszystkich etapach badania i opracowania danych, aby zapewnić maksymalną ochronę informacji i nie dopuścić do jej ujawnienia lub zniszczenia.

Mówiąc o zagrożeniach i konieczności ochrony danych na wstępnych etapach opracowań, nie sposób również pominąć sprawy uregulowania obiegu tzw. wydruków błędów, tj. wydruków komputerowych z jednostkowymi pozycjami z kwestionariuszy, które stosownie do założeń

³⁸⁾ CAPI — Computer-Aided Personal Interviewing.

³⁹⁾ W wielu krajach stosuje się już system pracy, zwany „biuro w domu”. W tej formie świadczenia pracy dopatruje się istotnych oszczędności kosztów utrzymania biura i czasu dojazdów do pracy.

automatycznej kontroli nie odpowiadają zadany parametrom. W wydrukach błędów znajdują się z zasady dane umożliwiające identyfikację indywidualnych jednostek, co oznacza, że przesyłanie i przechowywanie wydruków błędów powinno podlegać analogicznym rygorom, jak przechowywanie i przesyłanie wypełnionych kwestionariuszy.

Wprawdzie, dzięki postępowi w dziedzinie unowocześnienia informatycznych środków przetwarzania i rozszerzenia praktyki korygowania pozycji błędnych metodą bezpośredniego wprowadzania korekt do komputerów z eliminowaniem pracochłonnych i kosztownych wydruków, przypadki wykonywania i przesyłania zestawień błędów będą coraz rzadsze, niemniej jednak, w szeregu badań masowych „metoda wydruków” jest nadal stosowana, w związku z czym zgłoszona wyżej uwaga jest nadal aktualna.

5. OCHRONA DANYCH NA ETAPIE ROZPOWSZECHNIANIA INFORMACJI

5.1. Obowiązek ochrony danych a realizacja powszechnego dostępu do informacji

Sprawy ochrony danych na etapie rozpowszechniania informacji należy rozpatrywać w ścisłym powiązaniu z podstawowymi obowiązkami statystyki polegającymi na maksymalnym zaspokojeniu potrzeb informacyjnych społeczeństwa, organów władzy państwowej, administracji publicznej, organów samorządu, pracowników nauki i edukacji oraz zainteresowanych podmiotów gospodarczych.

We właściwym spełnianiu funkcji informacyjnej wyraża się sens istnienia statystyki w każdym kraju.

Zasoby informacyjne statystyki tworzone są w wyniku wysokich kosztów społecznych wyrażających się zarówno stosunkowo wysokimi wydatkami na badania statystyczne z budżetu państwowego, jak i dużymi nakładami pracy ponoszonymi przez respondentów (jednostki gospodarcze i obywateli) z tytułu uczestniczenia w badaniach statystycznych. Im lepiej i pełniej wykorzystywane są informacje zgromadzone przez służby statystyczne, tym większa jest efektywność nakładów ponoszonych na badania statystyczne.

Nieprzypadkowo funkcja rozpowszechniania informacji wymieniona jest na pierwszym miejscu „Podstawowych zasad statystyki oficjalnej na obszarze Europejskiej Komisji Gospodarczej”:

Statystyka oficjalna stanowi niezbędny element informacyjny społeczeństwa demokratycznego, dostarczając administracji publicznej, sektorowi gospodarczemu i społeczeństwu danych nt. sytuacji ekonomicznej, demograficznej, społecznej oraz środowiska naturalnego. Centralne organy statystyczne powinny zbierać dane statystyczne spełniające kryterium użyteczności i udostępniać je na zasadach bezstronności, mając na uwadze uszanowanie dostępu obywateli do informacji publicznej [Pod95].

Również ustawa o statystyce publicznej (art. 14) wyraźnie określa, że wyniki obliczeń, opracowań i analiz statystycznych dokonanych na

podstawie zebranych w badaniach statystycznych danych są powszechnie dostępne na zasadzie równoprawności, równorzędności i równoczesności.

Użytkownicy informacji statystycznej bardzo chętnie korzystają z danych statystyki publicznej. Doceniają ich, na ogół, dobrą jakość i możliwość uzyskania na dogodnych warunkach finansowych, zwłaszcza w porównaniu z kosztem ponownego zbierania i tworzenia zbiorów danych.

Służby statystyczne w kraju poświęcają, zwłaszcza w ostatnich latach, wiele uwagi sprawom upowszechniania informacji. Wyraża się to w aktywizacji polityki informacyjnej poprzez rozwój systemu informowania o opracowaniach i publikacjach GUS. W Głównym Urzędzie Statystycznym oraz w urzędach statystycznych utworzono specjalne służby informacyjne mające na celu usprawnienie informowania użytkowników o posiadanych przez statystykę informacjach oraz okazywanie pomocy w uzyskiwaniu poszukiwanych informacji. Systematycznie organizuje się spotkania z prasą, ważniejsze wyniki badań prezentowane są w telewizji. Prowadzi się szereg innych kampanii informacyjnych popularyzujących statystykę, jak przykładowo, konkursy na znajomość statystyki wśród młodzieży szkolnej itp. Niemniej jednak, skuteczność tych działań jest nadal niezadowalająca. Nadal nakłady publikacji statystycznych są rażąco niskie, nawet w porównaniu do sytuacji sprzed II wojny światowej.⁴⁰⁾

Sprawy upowszechnienia informacji statystycznych stanowiące jeden z ważnych elementów podnoszenia ogólnej wiedzy o gospodarce i sprawach społecznych, będące dla statystyków zadaniem najważniejszym z ważnych, nie mogą wchodzić w konflikt z wymaganiami pełnej ochrony danych indywidualnych i osobowych, o czym mówiliśmy obszerniej w poprzednich rozdziałach. Jednocześnie, wymogi ochrony informacji indywidualnych nie mogą być wykorzystywane jako argument do ograniczania udostępniania informacji.

W praktycznej działalności informacyjnej służb statystycznych występuje jednak szereg trudności w zachowaniu równowagi między postulatami ze strony użytkowników informacji a obowiązkami ochrony danych jednostkowych przed ich ujawnieniem. Trudności te można rozwiązać jedynie pod warunkiem zapewnienia niezbędnej równowagi między stopniem szczegółowości udostępnianej przez statystykę informacji a poziomem ochrony informacji indywidualnych przed ujawnieniem. Odnosi się to przede wszystkim do tych użytkowników, którzy posiadają własną techniczną bazę informatyczną i są zainteresowani w uzyskaniu danych do dalszego przetwarzania według własnych algorytmów ze względu na to, że układy analityczne prezentowane przez służby statystyczne nie odpowiadają ich potrzebom.

⁴⁰⁾ Być może dlatego ostatnio Zakład Wydawnictw Statystycznych zaprzestał podawania wysokości nakładu w wielu publikacjach.

Temat ten jest przedmiotem szerokiego zainteresowania zarówno w światowej literaturze statystycznej, jak i w praktycznej działalności służb statystycznych. Stephen E. Fienberg problem ten nazywa wprost „konfliktem między potrzebą dostępu do informacji statystycznej a wymogami tajemnicy” [Fie94].

Szczególne niebezpieczeństwo ujawnienia danych indywidualnych i danych osobowych istnieje w dwóch przypadkach:

- 1) w przypadku udostępniania przez służby statystyczne zbiorów danych źródłowych (mikrodanych) oraz
- 2) w przypadku udostępniania i publikowania tablic i zestawień zbiorczych (makrodanych) w szczegółowych grupowaniach według rodzajów działalności oraz w grupowaniach terytorialnych.

Nie dopuszczając do żadnych kompromisów grożących ujawnieniem tajemnicy statystycznej, statystycy nie mogą jednocześnie formalnie zastępować się tymi przepisami dla usprawiedliwienia nadmiernego, nieuzasadnionego ograniczania udostępniania informacji.

Jak to już podkreślałem kilkakrotnie, dane źródłowe pozyskiwane zarówno w drodze własnych badań, jak i z dokumentów i systemów administracyjnych stanowią dla służb statystycznych materiał, z którego budowane są opracowania i analizy statystyczne służące odpowiednim władzom i ośrodkom naukowo-badawczym do podejmowania decyzji i formułowania określonych programów w dziedzinie gospodarki i spraw społecznych.

Nie zawsze jednak metody, sposób ujęcia i szczegółowość opracowań i analiz zadowala w pełni wszystkich użytkowników. Częściowo bodaj tę rozbieżność między zakresem badań i analiz prowadzonych przez służby statystyczne a potrzebami użytkowników można zapewne zniwelować poprzez lepszą współpracę na etapie tworzenia programów badań i opracowań statystyki publicznej w taki sposób, aby program ten uwzględniał wymagania maksymalnej liczby użytkowników. Ten kierunek prac musi być kontynuowany, mimo znacznej uciążliwości i pracochłonności procedur uzgadniania postulatów różnych użytkowników i mimo ciągłych ograniczeń, na jakie napotykają służby statystyczne na drodze realizacji wynegocjowanych i przyjętych postulatów.

Jeszcze pełniej sygnalizowaną rozbieżność można usunąć poprzez znalezienie sposobów wykonywania przez służby statystyczne opracowań i analiz nie tylko stosownie do przyjętych rozwiązań w programie badań. Program ten, będący z konieczności wyrazem kompromisu między nie zawsze zgodnymi postulatami różnych użytkowników a możliwościami służb statystycznych, często nie zadowala niektórych użytkowników wywodzących się zwłaszcza z ośrodków naukowych, ośrodków konsultacyjnych, służb informacyjnych różnych organizacji itp. W tych przypadkach służby statystyczne, wykorzystując bogaty zasób danych, mogłyby na zlecenie zainteresowanych firm i instytucji realizować opracowania i analizy odpowiadające indywidual-

nemu zapotrzebowaniu poszczególnych użytkowników i według opracowanych przez nich algorytmów.

Postulat ten nie jest wprawdzie możliwy do natychmiastowej realizacji, ponieważ służby statystyki publicznej utrzymywane ze środków budżetowych nie mają w tej chwili wystarczających możliwości realizacji opracowań na indywidualne zapotrzebowania. Możliwości takie można by jednak stworzyć, rozwijając działalność agencji statystycznych, do tworzenia których upoważnia Prezesa GUS ustawa o statystyce publicznej (art. 29). Byłby to niewątpliwie nieporównanie mniej kosztowny dla użytkowników sposób uzyskiwania potrzebnych informacji, w porównaniu z alternatywą organizowania własnych badań.

Obok sygnalizowanych sposobów pełniejszego zaspokojenia postulatów wielu „nietypowych” użytkowników, nadal pozostanie niemały margines potrzeb, których nie są w stanie zaspokoić służby statystyki publicznej.

Doświadczenia wielu krajów wykazują, że ogromne bogactwo zasobów informacyjnych zgromadzone w służbach oficjalnej statystyki można by znacznie lepiej wykorzystać, gdyby znaleziono sposób legalnego udostępniania użytkownikom (głównie z ośrodków naukowych) zbiorów danych źródłowych (nieidentyfikowalnych).

Zbiory danych źródłowych niezbędne są zwłaszcza w przypadkach stosowania statystycznych metod analizy wielu zmiennych. Zapotrzebowanie na wykonywanie takich analiz będzie rosło. Naukowcy zwracają uwagę, że aby zrozumieć i wyjaśnić strukturę i dynamikę zmian zachodzących w społeczeństwach Europy, niezbędne jest rozszerzanie i pogłębianie badań w tej dziedzinie. Często podobne badania, które ograniczały się dawniej do zbiorowości poszczególnych krajów, rozszerzają się obecnie na całą integrującą się Europę, w tym na kraje Europy Wschodniej. Europa zaczyna być postrzegana jako nowy obiekt badań, zwiększa się zapotrzebowanie na analizy porównawcze różnych aspektów rozwoju społecznego różnych krajów europejskich [Mul94].

Takiemu wykorzystywaniu danych sprzyja rozwój badań prowadzonych w różnych krajach według wspólnie uzgodnionej metodologii, zapewniającej porównywalność danych zbieranych w różnych krajach. Do takich badań należą przykładowo prowadzone regularnie w większości krajów badania siły roboczej, warunków bytu, badania wykorzystania czasu itp.

Naukowcy podkreślają często, podobnie jak statystycy, że są zainteresowani w korzystaniu z mikrodanych nie w celach badania zidentyfikowanych indywidualnych jednostek, lecz w celach uzyskiwania grupowych charakterystyk określonej zbiorowości, do czego są im niezbędne charakterystyki poszczególnych jednostek i że, podobnie jak statystycy, są zdecydowani chronić poufność indywidualnych danych ze względu na to, że tylko takie postępowanie gwarantuje im otrzymanie wiarygodnych i nie zniekształconych danych.

Jeśli stwierdzamy, że nie ma zasadniczych różnic w celach wykorzystywania danych do opracowań i analiz wykonywanych przez statystyków i przez ośrodki naukowe, to należy poszukiwać takich rozwiązań, aby wspólne wykorzystanie zbiorów danych było możliwe bez naruszenia obowiązujących w tej dziedzinie bardzo restrykcyjnych przepisów prawnych oraz bez zagrożenia utraty zaufania do służb statystycznych ze strony respondentów.

W różnych krajach poszukuje się różnych rozwiązań tego dylematu. Jako uzasadnienie konieczności znalezienia kompromisowego wyjścia z sygnalizowanych trudności wysuwa się zwłaszcza następujące argumenty:

- 1) za zebrane i sprawdzone przez służby statystyczne dane społeczeństwo już zapłaciło ze środków budżetowych, dlatego maksymalne wykorzystanie tych danych do analiz leży w interesie podatników,
- 2) jeśli raz zebrane dane mogą być wykorzystywane bardziej wszechstronnie, to nie trzeba będzie ich ponownie zbierać. Są więc w takim rozwiązaniu zainteresowani sami respondenci,
- 3) umożliwienie wykorzystania danych przez placówki naukowe zwalnia je od konieczności odrębnego zbierania danych, ponoszenia kosztów i strat czasu na oczekiwanie opracowania wyników. Wykorzystanie więc danych zgromadzonych przez statystykę leży w interesie nauki,
- 4) wykorzystanie do analiz naukowych danych statystycznych mających opinię danych dokładnych i rzetelnych sprzyja uzyskaniu poprawnych i rzetelnych analiz, są więc w tym zainteresowani politycy odpowiedzialni za podejmowanie decyzji w konkretnych dziedzinach działalności,
- 5) ponieważ służby statystyczne, przeciążone bieżącymi obowiązkami w dziedzinie projektowania badań, zbierania, opracowania danych i udostępniania informacji, nie są w stanie w pełni wykorzystać całego bogactwa zgromadzonej informacji, to pogłębienie przez placówki naukowe analiz opartych na tych danych leży również w interesie służb statystycznych [Nob94].

Jedno z rozwiązań mierzących do pełniejszego wykorzystania do analiz danych źródłowych wprowadzono we Włoszech. We wrześniu 1989 r. na mocy dekretu-ustawy⁴¹⁾ utworzono tam *Krajowy System Statystyczny* (SISTAN), który zmierza do koordynacji działalności statystycznej w kraju. System ten obejmuje szeroką sieć organizacji publicznych, której głównym ogniwem jest Instytut Statystyczny — ISTAT (odpowiednik GUS). Obok ISTAT system SISTAN obejmuje centralne i lokalne organy administracji, Izbę Handlową, lokalne ośrodki zdrowia oraz ważniejsze instytucje publiczne. Wszystkie instytucje wchodzące w skład SISTAN zobowiązane są do utworzenia komórek statystycznych, które są odpowiedzialne za zbieranie i opracowanie informacji zgodnie z Narodowym Programem Statystycznym. Stworzenie takiego systemu umożliwiło swobodny przepływ

⁴¹⁾ Dekret z mocą ustawy nr 322 z dnia 6 września 1989 r.

informacji między wszystkimi ogniwami SISTAN, a tym samym umożliwiło pełniejsze wykorzystanie ogromnych ilości danych, których w poprzednim systemie nie można było wykorzystać.

Jednocześnie obowiązujące przepisy prawa formułują ściśle zasadę tajemnicy statystycznej, zobowiązując wszystkie ogniwa korzystające z mikrodanych do pełnej ich ochrony przed ujawnieniem. Przepisy te są znacznie ostrzejsze aniżeli przepisy dotyczące służb administracyjnych. O ile w świetle przepisów dotyczących dostępu do dokumentów administracyjnych „wszystko jest dozwolone co nie jest zabronione”, to w odniesieniu do tajemnicy statystycznej „wszystko jest zabronione co nie jest dozwolone” [Sim94].

W Holandii, w związku z ostrą krytyką ze strony środowisk naukowych bardzo restrykcyjnej polityki ograniczania przez Centralne Biuro Statystyczne (CBS) dostępu do zbiorów mikrodanych, próbą wyjścia naprzeciw potrzebom badań naukowych, z jednoczesnym pełnym zachowaniem obowiązków ochrony danych, było utworzenie, z inicjatywy Ministra Edukacji i Nauki i po szczegółowych uzgodnieniach z Komisją do Spraw Ochrony Danych, Naukowej Agencji Statystycznej, która na podstawie uzgodnień między Organizacją Badań Naukowych i CBS może otrzymywać każdego roku ściśle określone zbiory danych źródłowych.

Kontrakt w tej sprawie podpisano w kwietniu 1994 r. po długich i ożywionych dyskusjach. Zbiory udostępniane są odpłatnie, chociaż po cenach nie wygórowanych. W ramach wspomnianych uzgodnień CBS udostępnia niektóre zbiory, ale wyłącznie z badań społecznych, w których udział respondentów jest dobrowolny. Do takich tematów należą m.in. badania siły roboczej, panelowe badanie społeczno-gospodarcze, kohortowe badanie edukacji, badania budżetów rodzinnych, badanie planowania rodziny, badanie opinii konsumentów, badania stanu zdrowia, badania wypoczynku i rekreacji, badanie mobilności społecznej. Dane źródłowe ze statystyki gospodarczej zbierane na zasadzie obowiązku w żadnym przypadku nie mogą być udostępniane. Zgoda na udostępnianie danych nie obejmuje również niektórych tematów badań statystyki społecznej. Do takich tematów należy zwłaszcza statystyka zgonów [Nob96].⁴²⁾

W Australii, która szczyty się dużym społecznym zaufaniem do służb statystycznych, wyrażającym się m.in. wysokim wskaźnikiem udziału w badaniach, wynoszącym w miesięcznych badaniach siły roboczej prowadzonych w gospodarstwach domowych 96%, w miesięcznych badaniach gospodarstw domowych — 85—90%, wyodrębniono tzw. zbiory do użytku publicznego, zawierające szereg danych pochodzących z badań prowadzonych w gospodarstwach domowych oraz ze spisu ludności. Zbiory te udostępniane są każdemu, kto podpisze zobowiązanie do wykorzysta-

⁴²⁾ Poufność tych danych chroniona jest szczególnie starannie ze względu na niebezpieczeństwo ujawnienia tajemnicy lekarskiej. CBS odmówił nawet udostępnienia tych danych na prośbę (nie na polecenie) sądu w związku z procesem lekarzy w sprawie zarzutu o eutanazję.

tywania uzyskanych danych wyłącznie do celów statystycznych. W praktyce zbiory te są wykorzystywane wyłącznie przez agencje rządowe i instytucje naukowe. Przed podjęciem decyzji o włączeniu określonych danych do zbiorów publicznych, prowadzi się szczegółową analizę z punktu widzenia niebezpieczeństwa naruszenia tajemnicy statystycznej, zwłaszcza w przypadkach, gdy użytkownik takiego zbioru dysponuje zbiorami administracyjnymi, które mógłby wykorzystywać do łączenia ze zbiorem statystycznym. Zasady wykorzystania zbiorów publicznych nadzoruje specjalny komitet stanowiący ciało doradcze prezesa Urzędu Statystycznego.

Prowadząc aktywną i elastyczną politykę w dziedzinie udostępniania informacji służby statystyczne wspólnie ze środowiskami naukowymi poszukują takich rozwiązań, które uwzględniłyby zarówno wzrost potrzeb użytkowników, jak i ściśle przestrzeganie zasad ochrony danych.

Badania w tej dziedzinie koncentrują się na dwóch podstawowych problemach:

- 1) jak wykręcić i zapobiec przypadkom ujawnienia danych indywidualnych w publikacjach i w innych formach udostępniania wyników badań statystycznych oraz
- 2) jakie zastosować metody zapobiegania pośredniej identyfikacji danych w przypadkach udostępniania danych w bardzo szczegółowych układach lub w przypadkach udostępniania danych jednostkowych.

W kraju problem ewentualnego korzystania z mikrodanych przez środowiska naukowe, przy pełnym zachowaniu przepisów o tajemnicy statystycznej, nadal czeka na rozwiązanie. Aby znaleźć właściwe rozwiązania niezbędne są nie tylko przychylność i dobra wola statystyków, ale również bardziej aktywna współpraca środowisk naukowych, które

w dialogu ze statystykami muszą występować nie tylko jako wymagający odbiorcy tych danych, ale również muszą pomagać statystykom wypracować skuteczne metody wykrywania zagrożeń w zbiorach informacyjnych i zapobiegania ujawnieniu indywidualnych danych.

O metodach tych będzie mowa w dalszej części pracy.

W tym kontekście należy zwrócić uwagę na dość istotną różnicę w podejściu do problemów ochrony danych w naszym kraju i w krajach zachodnich. W Polsce, bardzo często, środowiska naukowe formułują pod adresem służb statystycznych krytyczne zarzuty, iż służby te traktują przepisy prawne o tajemnicy statystycznej zbyt formalnie, że utrudniając dostęp do mikrodanych nie stymulują rozwoju wielu badań naukowych, zwłaszcza w dziedzinie statystyki społecznej. Podobne zarzuty, aczkolwiek oparte o inne przesłanki, formułowane są często przez przedstawicieli organów administracji państwowej i samorządowej w odniesieniu do mikrodanych, charakteryzujących działalność gospodarczą jednostek. Służby statystyczne w Polsce w realizacji polityki informacyjnej, przewidującej pełną

ochronę danych osobowych i indywidualnych, nie znajdują wystarczającego poparcia środowisk naukowych i administracyjnych.

W odróżnieniu od sytuacji w Polsce, w wielu krajach badania w dziedzinie metod ochrony danych jednostkowych rozwijają się najsilniej nie w służbach statystycznych, ale w środowiskach naukowych, przy czym środowiska te na prowadzenie badań uzyskują istotne poparcie finansowe z funduszy państwowych. Tak, przykładowo, w Niemczech badania naukowe dotyczące określenia ryzyka ujawnienia mikrodanych i zapewnienia ich poufności prowadzone są w Uniwersytecie w Mannheim, a badania wspierane są funduszami z budżetu Ministerstwa Postępu i Technologii. Zespół badawczy, pracujący pod kierunkiem prof. Waltera Müllera, opublikował wyniki swoich badań m.in. w [Mül94] oraz [Mül95].

We Włoszech, wszechstronne badania metod nielegalnego ujawniania danych i sposobów ochrony jednostkowych danych prowadzone są we współpracy z ISTAT przez pracowników naukowych Wydziału Statystyki Uniwersytetu we Florencji oraz w Urbino, por. [Big94], [Bor94], [Buz94].

W Stanach Zjednoczonych oraz w Izraelu badania w tej dziedzinie prowadzone są przez prof. Stephena E. Fienberga na Uniwersytecie w Pittsburgu oraz prof. Udi E. Makova na Wydz. Statystyki Uniwersytetu w Hajfie. Ich podejście polega na zastosowaniu metod bayesowskich do ujawniania jednostkowych danych przez potencjalnego intruza, por. [Mak94].

5.2. Wykrywanie zagrożeń identyfikacji w zbiorach danych

Aby można było zapewnić w pełni przestrzeganie zasad ochrony danych służby statystyczne muszą zapewnić sprawne funkcjonowanie całego systemu przedsięwzięć uniemożliwiających nie tylko bezpośrednie (świadome lub nieświadome) ujawnienia danych indywidualnych lub osobowych, ale również zapobieżenie przypadkom, w których na podstawie danych zagregowanych lub danych jednostkowych pozbawionych cech identyfikujących użytkownik mógłby, przy zastosowaniu określonych procedur, doprowadzić do ujawnienia danych podlegających ochronie. Pojęcie „ujawnienia” rozumie się, na ogół, stosownie do definicji wprowadzonej przez Tore Daleniusa:

jeżeli udostępnienie pewnej informacji statystycznej daje możliwość ustalenia określonej wartości odnoszącej się do znanej jednostki w sposób bardziej dokładny niż to jest możliwe bez dostępu do tej informacji, to w tym przypadku mamy do czynienia z ujawnieniem [Dal77].

Ujawnienie informacji indywidualnych może być całkowite (lub dokładne) oraz częściowe (przybliżone). Całkowite ujawnienie występuje wtedy, gdy użytkownik statystyki może określić dokładnie cechy A_i dla indywidualnej

jednostki i znajdującej się w bazie danych. Ujawnienie częściowe lub przybliżone ma miejsce, jeśli użytkownik jest w stanie określić estymator \hat{A}_i dla indywidualnej jednostki i , której wariancja odpowiada założeniu: $Var(\hat{A}_i) > c^2$, gdzie c jest parametrem, który należy określić. Inaczej mówiąc z ujawnieniem częściowym mamy do czynienia, jeśli istnieje możliwość niepożądanie bliskiej oceny indywidualnej jednostki [Sch93, s. 2].

Ilustracją metody identyfikacji informacji jednostkowej na podstawie zespołu cech charakteryzujących tę jednostkę (identyfikacja pośrednia) jest przykład przytoczony w cytowanej już pracy Lance J. Hoffmana [Hof77]. W rozdziale wspomnianej pracy, poświęconym ochronie informacji w statystycznych bankach danych, Autor podkreśla, że nawet jeśli system zaprojektowany został w taki sposób, że będzie udostępniał wyłącznie dane wynikowe i blokował udostępnianie pozycji (rekordów) jednostkowych, to mimo braku danych identyfikujących można wyodrębnić dane dotyczące określonej osoby, jeśli osoba zainteresowana w ujawnieniu tych danych będzie dysponowała określonym zasobem informacji, z czego nie zawsze zdajemy sobie sprawę. Często można spotkać się z uproszczonym poglądem, że jeśli ze zbioru danych jednostkowych usuniemy informacje adresowe oraz inne informacje identyfikujące, to zapewnimy poufność danych w zbiorze danych statystycznych.

Jako jeden z możliwych sposobów postępowania potencjalnego intruza,⁴³⁾ zmierzającego do wyodrębnienia ze zbioru danych dossier dotyczącego określonej osoby, cytowany autor przytacza możliwy sposób postępowania zmierzający do identyfikacji danych dotyczących osoby b. prezydenta Stanów Zjednoczonych Geralda Forda.

Przypuśmy, iż chcemy ze zbioru jednostkowego zawierającego dane osobowe (w tym dane dotyczące wyznania) uzyskać informacje dotyczące wyznania G. Forda. Dla zawężenia pola badania, wykorzystując informacje zawarte w ogólnie dostępnej publikacji "Who is who?" określamy, że:

- 1) urodził się w stanie Omaha,
- 2) urodził się w 1913 r.,
- 3) jest członkiem klubu Phi Delta Phi.

Zwracając się do systemu o liczbę osób posiadających cechy 1, 2, 3 możemy ustalić, czy otrzymamy w wyniku liczbę 1, co oznacza, że cechy te identyfikują jednoznacznie G. Forda. Jeśli na pytanie o liczbę osób w zbiorze odpowiadającym wymienionym cechom otrzymamy liczbę większą od 1, to korzystając z dostępnych źródeł wprowadzamy dalsze znane nam cechy, np. 4) mieszkał w Aleksandrii w stanie Virginia oraz 5) ukończył Uniwersytet Michigan w 1935 r. Jeśli w trakcie kolejnych zapytań uzyskamy w wyniku wyłącznie jedną osobę, możemy być pewni, że osobą tą jest G. Ford. Tą metodą możemy także ujawnić szereg innych wrażliwych

⁴³⁾ Intruzem (ang. intruder lub invader) nazywamy osobę zainteresowaną w identyfikacji badanego podmiotu i w uzyskaniu o nim informacji, która nie jest powszechnie dostępna.

informacji dotyczących interesującej nas osoby, np. czy była ona karana. Możemy w tym przypadku zażądać podania liczby osób karanych spośród osób posiadających cechy 1, 2, 3, 4, 5. Jeśli na to zapytanie otrzymamy odpowiedź = 0, to będzie to oznaczać, że zidentyfikowana osoba nie była karana [Hof77, s. 125].⁴⁴⁾

Stosując podobną metodę postępowania możemy w szeregu przypadków doprowadzić do identyfikacji w bardziej przystających do naszej praktyki sytuacjach, np. w zbiorach danych z badań budżetów rodzinnych. Przykładowo, wykorzystując pewne dostępne informacje z systemów administracyjnych (z systemu PESEL, z wydziałów komunikacji), z publikacji prasowych lub z własnej znajomości lokalnych realiów, można na podstawie korelacji cech (np. wieku, zawodu, stanu rodzinnego, posiadania samochodu itp.) doprowadzić do identyfikacji poszczególnych osób, zwłaszcza osób odznaczających się stosunkowo rzadkimi cechami.

Z powyższych względów w praktyce występuje duże niebezpieczeństwo ujawnienia danych jednostkowych nie tylko w przypadku zbiorów zawierających zmienne identyfikujące, ale również w przypadku, gdy w zbiorze lub w tablicach przedstawiających dane w szczegółowych układach analitycznych znajduje się niewielka liczba podmiotów odznaczających się rzadkimi cechami. Wymaga to zwrócenia większej uwagi na ten aspekt każdorazowo, zanim podejmiemy decyzję o udostępnieniu określonego opracowania lub publikacji.

Pilną do załatwienia sprawą jest więc opracowanie odpowiednich metod (w tym komputerowych) kontroli zbiorów informacji wynikowych, ujawniających niebezpieczeństwo udostępnienia takich danych, na podstawie których zainteresowany w identyfikacji i często dysponujący dodatkowo innymi informacjami intruz, mógłby doprowadzić do ujawnienia indywidualnych danych i naruszenia przepisów o tajemnicy statystycznej.

Do opracowania i ogłoszenia takich metod zobowiązuje Prezesa GUS art. 38 ustawy o statystyce publicznej.

5.2.1. *Naruszenie zasady wielkości progowej*

Jeszcze trudniej zapobiec identyfikacji danych indywidualnych w statystyce gospodarczej, gromadzącej wyniki działalności jednostek gospodarczych. W tym przypadku poufne są nie tylko wszelkie dane indywidualne uzyskiwane od respondentów, ale za poufne uznaje się również wszystkie te pozycje tablicy, które zawierają dane o mniejszej od m liczbie jednostek. Liczbę m , zwaną *wielkością progową*, ustala się w zależności od wymaganego stopnia poufności. W różnych krajach Unii Europejskiej wielkość progowa

⁴⁴⁾ Powyższe rozważanie oparte jest na założeniu, że potencjalny intruz nie ma dostępu do zbioru informacji jednostkowych, który jest prawnie chroniony, nie może więc uzyskać pełnej informacji o interesującej go osobie, może natomiast otrzymać odpowiedź na pytanie o liczbę osób odznaczających się określonymi cechami.

m waha się od 3 do 5. Oznacza to, że każda pozycja tablicy musi być traktowana jako poufna jeśli liczba w tej pozycji obejmuje dane o liczbie jednostek mniejszej od m [Sch93, s. 3].

Przyjęcie jako minimalnej wielkości progowej $m=3$ opiera się na założeniu, że gdyby ta wielkość $m=2$, to jedna z tych jednostek miałaby możliwość zidentyfikowania drugiej jednostki oraz ujawnienia wszystkich wielkości zawartych w tablicy przez odjęcie od wielkości występujących w tablicy własnych danych jednostki intruza. Można wprawdzie założyć, że również w przypadku, gdy $m=3$ do identyfikacji mogłoby dojść, gdyby dwa podmioty z trzech domówiły się między sobą i dokonały odpowiednich obliczeń na podstawie danych zawartych w tablicy. Dlatego wielkość progową m w koniecznych przypadkach zwiększa się do 5.

Wprawdzie w ustawie o statystyce publicznej pojęcie wielkości progowej zostało wprowadzone w związku z potrzebą zapobieżenia (a właściwie utrudnienia) identyfikacji jednostki gospodarczej, to jednak w wielu krajach pojęcie to stosuje się również do danych osobowych. Chodzi o to, aby w przypadku udostępnienia danych źródłowych (mikrodanych) w zbiorze nie znalazły się podmioty, w których występuje tzw. rzadka kombinacja cech lub inaczej mówiąc, aby w zbiorze liczba podmiotów o określonej kombinacji cech (np. płeć×wiek×stan cywilny×zawód) nie była mniejsza od m . W tym przypadku m można nazwać umownie wielkością progową. Wielkości progowe w tym przypadku różniące się zwykle w zależności od wielkości obszaru ze względu na to, że na małym obszarze (np. miejscowości) o wiele łatwiej zidentyfikować osobę o danej kombinacji cech aniżeli na terenie np. województwa. Jeśli liczba podmiotów w określonej kombinacji cech jest mniejsza od przyjętej wielkości progowej m , to takie podmioty noszą nazwę rzadkich i podlegają odpowiednim procedurom ochrony. Podmioty o liczbie większej od przyjętej wielkości progowej noszą nazwę powszechnych i nie podlegają szczególnej ochronie [Vri94].

5.2.2. *Naruszenie zasady przeważającego udziału*

Innym powodem uznawania szczegółowych pozycji tablicy za poufne, ze względu na niebezpieczeństwo ujawnienia informacji indywidualnych, jest tzw. „zasada przeważającego udziału”. Zastosowanie tej zasady oznacza, że pozycja tablicy traktowana jest jako poufna, jeśli udział największych podmiotów n w wielkości danej pozycji tablicy jest większy od k procent. W rozstrzygnięciach prawnych różnych krajów Unii Europejskiej obowiązują różne wielkości zarówno dla n , jak i k . Najczęściej za n przyjmuje się liczbę 2, a za k wielkość 85%. Oznacza to, że pozycja tablicy uznawana jest za poufną i nie może być udostępniana, jeśli dwa największe podmioty mają w danej pozycji udział większy niż 85% [Sch93, s. 3].

W przypadku prezentacji wyników badań w układach szczegółowych (np. według rodzaju działalności, w szczegółowych układach terytorialnych)

należy przed udostępnieniem danych przekonać się, czy w jakiejś pozycji tablicy nie występują wielkości podlegające wymienionym wyżej ograniczeniom, tj., czy nie została naruszona zasada progowa lub zasada przeważającego udziału.

5.2.3. Ujawnienie danych metodą odejmowania

W niektórych przypadkach dane indywidualne mogą zostać ujawnione nawet w przypadkach pewnego przestrzegania zasady nieprzekraczania wielkości progowej i zasady przeważającego udziału. Może to nastąpić w niektórych przypadkach, gdy tablica zawiera obok pozycji liczbowych również sumy ogólne, sumy dla wierszy i rubryk lub sumy pośrednie obejmujące kilka rubryk lub wierszy. W niektórych tego rodzaju przypadkach można doprowadzić do ujawnienia poufnych danych poprzez odjęcie od sum tablicy odpowiednich sum cząstkowych.

Przykładowo, jeśli w tablicy występują dane dotyczące transportu ogółem oraz sumy cząstkowe obejmujące transport lądowy i wodny, to poprzez odjęcie sum cząstkowych od ogółem można ujawnić dane dotyczące transportu lotniczego. Gdyby tablica zawierała dane charakteryzujące wyniki ekonomiczne działalności badanych podmiotów gospodarczych (w tym przypadku w transporcie), to stosując metodę odejmowania można doprowadzić do ujawnienia danych podlegających ochronie i w konsekwencji do naruszenia tajemnicy statystycznej.

Podobne sytuacje mogą zaistnieć w szczegółowych publikacjach (najczęściej obejmujących wyniki badań masowych), zawierających tablice w układach hierarchicznych: w podziale geograficznym, w układzie rodzajów działalności itp.

Wymienione wyżej zagrożenia i konieczne środki ostrożności odnoszą się zwykle do wyników badań pełnych, tj., kiedy dane zawarte w tablicach oparte są na danych obejmujących całą populację. Jeżeli dane obejmują jedynie wybrane losowo do badania jednostki, to, na ogół, uważa się, że można nie stosować zwłaszcza zasady progowej, ponieważ potencjalny intruz nie jest w stanie dowiedzieć się, jakie jednostki zostały wylosowane do próby. W przypadku przesyłania danych do Eurostatu, decyzję, czy należy przestrzegać zasady progowej w przypadku badań reprezentacyjnych podejmuje kraj członkowski [Sch93, s. 3].

W Polsce, zgodnie z postanowieniami w art. 38, ust. 2 ustawy o statystyce publicznej, nie mogą być publikowane ani udostępniane dane charakteryzujące wyniki ekonomiczne działalności podmiotów gospodarczych, jeżeli na daną agregację składają się mniej niż 3 podmioty⁴⁵⁾ lub udział jednego podmiotu w określonym zestawieniu jest większy niż trzy czwarte całości.⁴⁶⁾

⁴⁵⁾ $m = 3$.

⁴⁶⁾ $k = 75\%$.

5.2.4. Czy można całkowicie zapobiec ujawnieniu danych jednostkowych?

Cytowany Tore Dalenius opracował wyrażone w sposób sformalizowany twierdzenie, że ujawnienie danych jednostkowych (indywidualnych i osobowych) (P_r) nastąpi, jeśli:

$$P_r(D_k = 1|S,E) > P_r(D_k = 1|E)$$

gdzie:

D_k stanowi cechę, którą może być liczba jednostek lub wielkość liczbową dotycząca danej jednostki, a D stanowi wartość cechy D odnoszącą się do k -tej jednostki — O_k . W przypadku charakterystyk liczbowych indywidualne dane D_k mogą nie odpowiadać ściśle określonej wielkości, lecz mieścić się w pewnych granicach,

S oznacza udostępniane z systemu dane,

E to informacja znana lub dostępna dla użytkownika.

Dla uproszczenia Autor przyjmuje, że $D_k = 1$, jeśli O_k posiada określoną cechę, a w przeciwnym przypadku $D_k = 0$ (cyt. za [Fie94, s. 120]).

Ponieważ niemal każdy przypadek udostępnienia danych dostarcza pewnych informacji o D_k , to całkowite uniknięcie ujawnienia jest praktycznie niemożliwe i mimo że ustawodawstwo dotyczące tajemnicy statystycznej często pisane jest przy milczącym założeniu zerowego ryzyka ujawnienia, to jest oczywiste, że jeśli udostępniamy jakąkolwiek informację, nie da się całkowicie uniknąć ryzyka ujawnienia informacji jednostkowej. W praktyce chodzi więc o ograniczenie ujawnienia [Fie94, s. 121].

W aktach prawnych Unii Europejskiej dotyczących ochrony danych (będących nadal w sferze projektów) wyraźnie podkreśla się względność pojęcia ryzyka ujawnienia. W projekcie art. 14a projektu Dyrektywy w sprawie ochrony danych postanawia się, iż zasada ochrony odnosi się do dowolnej informacji dotyczącej osób zidentyfikowanych lub dających się zidentyfikować, przy czym możliwość identyfikacji należy rozpatrywać w kategoriach zdroworoządkowych, ponieważ teoretycznie możliwości identyfikacji nie można wykluczyć; należy ją jednak rozpatrywać w kontekście proporcjonalności nakładów na identyfikację i ewentualnych korzyści z ujawnienia zidentyfikowanych danych [Eng94].

Największe ryzyko ujawnienia informacji powstaje w przypadku udostępniania danych w bardzo szczegółowych przekrojach terytorialnych. J. Steinberg cytuje prosty przykład, iż zainteresowany użytkownik może łatwo wydedukować informacje o dochodach konkretnego lekarza z tablicy przedstawiającej dochody według zawodów i przychodni lekarskich według miast [Ste83].

Opracowaniem i doskonaleniem metod wykrywania zagrożeń ujawnienia danych jednostkowych w udostępnianych zbiorach informacji zajmuje się obecnie wielu specjalistów zarówno w urzędach statystycznych, jak i w szeregu placówek uniwersyteckich w USA, w Kanadzie, a w Europie zwłaszcza w Holandii, w Niemczech, we Francji, we Włoszech i w Finlan-

dii.⁴⁷⁾ Prace te opierające się na analizie konkretnych zbiorów informacji zarówno ze statystyki gospodarczej, jak i statystyki społecznej (ludność, gospodarstwa domowe) mają na celu ujawnienie niebezpieczeństwa i niedopuszczenie do publikowania danych w szczególowości stwarzającej zagrożenie identyfikacji danych jednostkowych. Wszystkie te opracowania, wykorzystujące często dość złożone modele matematyczne, opierają się na założeniu, że potencjalny intruz dysponuje własnymi dodatkowymi źródłami informacji, które może wykorzystać jako informację pomocniczą ułatwiającą identyfikację.

Trudność opracowania uniwersalnych metod wykrywania zagrożeń ujawnienia wynika z bardzo dużej różnorodności danych statystycznych oraz z trudności przewidzenia jakimi dodatkowymi informacjami dysponuje potencjalny intruz. Z tego wynika również trudność bezpośredniego wykorzystania w polskich warunkach wypracowanych w innych krajach metod, ponieważ muszą one uwzględniać rzeczywiste cechy badanych w naszym kraju zbiorowości oraz łatwość dostępu ze strony potencjalnych intruzów do informacji dodatkowych (tzw. informacji a priori) możliwych do wykorzystania do identyfikacji danych udostępnianych przez statystykę.

⁴⁷⁾ Por. np. [Yeo94, Laa94, Mak94, Big94].

6. METODY ZAPOBIEGANIA IDENTYFIKACJI I UJAWNIENIU DANYCH

6.1. Metody zapobiegania identyfikacji w przypadku udostępniania danych źródłowych (mikrodanych)

Prowadzenie przez służby statystyczne aktywnej polityki informacyjnej związane jest z koniecznością ciągłej kontroli, czy udostępniane przez nie informacje nie zostaną zidentyfikowane z podmiotem, który dane te przekazał służbom statystycznym w procesie prowadzonych przez nie badań w przekonaniu, że nie zostaną one nikomu ujawnione, zgodnie z przepisami o tajemnicy statystycznej.

Konkretne metody kontroli i zapobiegania zagrożeniom identyfikacji danych zależą od tematu badania, metody badania, wielkości zbioru, liczby cech badanych, dostępności innych pozastatystycznych źródeł pokrewnych danych itp.

Metody zapobiegania identyfikacji różnią się jednak przede wszystkim w zależności od tego, czy udostępniane są dane źródłowe (mikrodane)⁴⁸⁾, czy wyniki opracowań, ujmowane najczęściej w tablicach statystycznych (makrodane).

Dane źródłowe odnoszące się do poszczególnych respondentów lub podmiotów badań przechowywane są przeważnie w zbiorach danych w postaci zapisów (rekordów) składających się z szeregu zmiennych zwanych **wektorami danych** [Sch93, s. 48].

Przykładowo, w badaniu warunków życia ludności wektor danych może zawierać następujące zmienne (część z nich wyrażona w formie symboli cyfrowych lub literowych): miejsce zamieszkania, nr gospodarstwa domowego, płeć, data urodzenia, stan cywilny, poziom wykształcenia, główne i dodatkowe źródło utrzymania, aktywność zawodowa i bezrobocie, inwalidztwo, wydatki na żywność według grup towarów, wydatki

⁴⁸⁾ Zgodnie z definicją pojęć przytoczoną w rozdziale 2 pojęcie to oznacza dane dotyczące poziomu respondenta. Według powszechnie stosowanej dotąd terminologii dane te określano pojęciem „dane jednostkowe”. W omawianym tutaj kontekście unikam pojęcia „dane jednostkowe”, bowiem określenie to w ustawie o statystyce publicznej zdefiniowano nieco w innym znaczeniu — por. definicje przytoczone w rozdziale 2.

na usługi, warunki mieszkaniowe, wyposażenie gospodarstwa domowego, oszczędności itp.

W ramach każdego wektora danych możemy wyróżnić dwa rodzaje zmiennych: zmienne identyfikujące i zmienne charakteryzujące. Zmienne charakteryzujące, tj. informujące o właściwościach jednostek badanych zbiorowości nazywamy również cechami statystycznymi oraz, ponieważ bardzo często według tych cech grupujemy badane zbiorowości, to, cechy te można również nazwać cechami grupującymi.

W ramach zmiennych identyfikujących możemy wyróżnić **identyfikatory bezpośrednie** i **identyfikatory pośrednie** lub quasi-identyfikatory. Identyfikatorami bezpośrednimi nazywamy takie zmienne identyfikujące, na podstawie których można jednoznacznie zidentyfikować badany podmiot. Takim identyfikatorem jest np. imię i nazwisko, data urodzenia i adres respondenta.⁴⁹⁾ W przypadku gdy nie są znane identyfikatory bezpośrednie, badany podmiot może być w wielu przypadkach zidentyfikowany na podstawie cech statystycznych (zmiennych charakteryzujących), zwłaszcza jeśli osoba zainteresowana w identyfikacji dysponuje pewnymi informacjami dodatkowymi. W tym przypadku cechy te spełniają funkcje identyfikatorów pośrednich.

Przytoczone określenia możemy zilustrować następującym przykładem kilku (fikcyjnych) danych jednostkowych:

TABL. 1

Lp.	Nazwa zmiennej	Jednostka 1	Jednostka 2	Jednostka 3
1	Imię i nazwisko	Jan Smólski	Józef Ton	Anna Rola
2	Adres	Sanok, ul. Jas- na 6	Krosno, ul. Li- powa 15	Rabka 56
3	Płeć	mężczyzna	mężczyzna	kobieta
4	Wiek (liczba lat)	37	52	90
5	Stan cywilny	rozwidziony	żonaty	wdowa
6	Zawód	tokarz	art. malarz	emeryt. lekarz
7	Wykształcenie	średnie	wyższe	wyższe
8	Źródło utrzymania	praca	bezrobotny	emerytura
9	Liczba dzieci	0	7	0
10	Dochód mies. w zł	900	500	300
11	Stan oszczędności ogółem w zł	2000	6000	12000
12	w tym lokaty w banku	1800	6000	10000

Przytoczony przykład zbioru jednostkowego podlega całkowitej ochronie ze względu na tajemnicę statystyczną, bowiem każda jednostka badana

⁴⁹⁾ Ścisłej mówiąc badanego podmiotu lub badanej jednostki, ponieważ w przypadku np. statystyki zgonów niezręcznie byłoby nazywać badaną jednostkę respondentem.

może być zidentyfikowana na podstawie identyfikatora bezpośredniego, tj. imienia i nazwiska oraz adresu. Nie oznacza to jednak, że usunięcie ze zbioru wiersza 1 i 2 spowoduje, że zbiór jako anonimowy można bezpiecznie udostępniać bez obawy ujawnienia informacji osobowych wszystkich lub niektórych znajdujących się w zbiorze jednostek. Ujawnienie w zbiorze zawierającym mikro dane ma miejsce, jeśli zainteresowana w ujawnieniu osoba (intruz) jest w stanie zidentyfikować interesującą ją jednostkę i uzyskać informacje dotyczące zmiennych charakteryzujących daną jednostkę. Musimy przy tym brać pod uwagę, że intruz może dysponować własnymi informacjami z innych dostępnych źródeł. Informacje te, w połączeniu z niektórymi cechami zawartymi w zbiorze można wykorzystać jako identyfikatory pośrednie i doprowadzić do identyfikacji i ujawnienia.

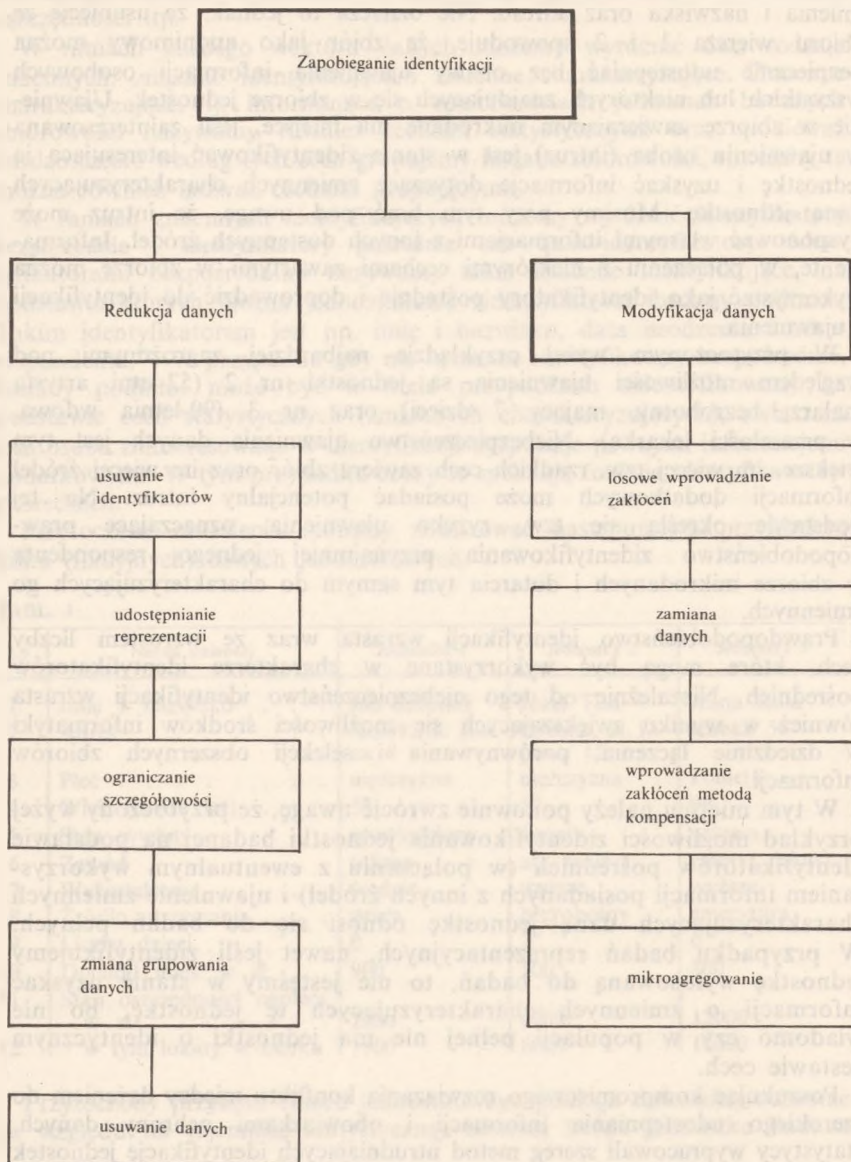
W przytoczonym wyżej przykładzie najbardziej zagrożonymi pod względem możliwości ujawnienia są jednostki nr 2 (52-letni artysta malarz, bezrobotny, mający 7 dzieci) oraz nr 3 (90-letnia wdowa, w przeszłości lekarka). Niebezpieczeństwo ujawnienia danych jest tym większe, im więcej tzw. rzadkich cech zawiera zbiór oraz im więcej źródeł informacji dodatkowych może posiadać potencjalny intruz. Na tej podstawie określa się tzw. ryzyko ujawnienia, oznaczające prawdopodobieństwo zidentyfikowania przynajmniej jednego respondenta w zbiorze mikro danych i dotarcia tym samym do charakteryzujących go zmiennych.

Prawdopodobieństwo identyfikacji wzrasta wraz ze wzrostem liczby cech, które mogą być wykorzystane w charakterze identyfikatorów pośrednich. Niezależnie od tego niebezpieczeństwo identyfikacji wzrasta również w wyniku zwiększających się możliwości środków informatyki w dziedzinie łączenia, porównywania i selekcji obszernych zbiorów informacji.

W tym miejscu należy ponownie zwrócić uwagę, że przytoczony wyżej przykład możliwości zidentyfikowania jednostki badanej na podstawie identyfikatorów pośrednich (w połączeniu z ewentualnym wykorzystaniem informacji posiadanych z innych źródeł) i ujawnienie zmiennych charakteryzujących daną jednostkę odnosi się do badań pełnych. W przypadku badań reprezentacyjnych, nawet jeśli zidentyfikujemy jednostkę wylosowaną do badań, to nie jesteśmy w stanie uzyskać informacji o zmiennych charakteryzujących tę jednostkę, bo nie wiadomo czy w populacji pełnej nie ma jednostki o identycznym zestawie cech.

Poszukując kompromisowego rozwiązania konfliktu między dążeniem do szerokiego udostępniania informacji i obowiązkami ochrony danych, statystycy wypracowali szereg metod utrudniających identyfikację jednostek przez potencjalnych intruzów. W sposób schematyczny metody te przedstawia wykr. 3.

Wykres 3 METODY ZAPOBIEGANIA IDENTYFIKACJI, UJAWNIEŃ DANYCH JEDNOSTKOWYCH



Większość z tych metod ogranicza wartość poznawczą informacji, a nawet w szeregu przypadków powoduje ich zniekształcenie. Użytkownicy, którzy nierzadko krytykują tego rodzaju metody, muszą jednak wybrać czy wolą uzyskać informacje nieco uboższe, częściowo zniekształcone lub nawet „spreparowane”, czy nie uzyskać ich w ogóle. Jednocześnie żadna z tych metod zastosowana oddzielnie nie stanowi wystarczającego zabezpieczenia przed ujawnieniem pośrednim. Dlatego są one stosowane zwykle w połączeniu, a więc przykładowo, obok usuwania stosuje się metodę modyfikacji.

Dla uniknięcia nieporozumień należy również podkreślić, że omawiane w tym punkcie metody zapobiegania identyfikacji stosuje się głównie do danych badań społecznych, w których jednostkami badania są pojedyncze osoby, gospodarstwa domowe, rodziny itp. Metod tych nie stosuje się, na ogół, do danych jednostkowych charakteryzujących działalność gospodarzących podmiotów ze względu na to, że danych o działalności podmiotów gospodarczych, uzyskanych w drodze badań statystycznych lub z systemów administracyjnych, służby statystyczne z zasady nie udostępniają. Są one bowiem o wiele łatwiejsze do zidentyfikowania, zwłaszcza w przypadku dużej liczby badanych cech. Wprawdzie w niektórych krajach (m.in. w Holandii) indywidualne dane mogą być publikowane (lub udostępniane w inny sposób) w przypadkach, gdy zainteresowane podmioty wyrażą na to zgodę, ale w tym przypadku omawiane w niniejszym punkcie metody nie mają zastosowania.

6.2. Redukcja danych

Pojęciem tym określamy szereg metod, które zmierzają do identyfikacji danych podlegających ochronie poprzez zmniejszenie liczby zmiennych zawartych w wektorze danych lub przez ograniczenie w inny sposób dostępności do pełnego zbioru danych. Redukcja danych może wyrażać się: w usuwaniu identyfikatorów, w udostępnianiu reprezentacji (prób) zamiast zbioru pełnego, ograniczeniu szczegółowości, w zmianie grupowań danych oraz w usuwaniu niektórych cech (zmiennych).

6.2.1. Usuwanie identyfikatorów bezpośrednich

Jest to metoda najprostsza. Polega ona na usunięciu ze zbioru identyfikatorów bezpośrednich. Metoda ta nosi również nazwę *anonimizacji* lub *deidentyfikacji*.

Po usunięciu identyfikatorów bezpośrednich ewentualna identyfikacja, jak wskazywano wyżej, jest możliwa jedynie na podstawie identyfikatorów pośrednich, do czego jednak niezbędne są dodatkowe informacje, które niekoniecznie zawsze są dostępne dla potencjalnego intruza.

Deidentyfikacja stanowi konieczny, chociaż zwykle niewystarczający zabieg przeprowadzany przed udostępnieniem zbioru mikrodanych.

6.2.2. Udostępnianie próby

Metoda ta może być stosowana zarówno w przypadku, gdy zbiór mikrodanych pochodzi z badania pełnego obejmującego wszystkie jednostki populacji (np. spisu), jak i w przypadku badania reprezentacyjnego.

Udostępniając reprezentację mikrodanych, tj. pewną ilość n rekordów wybranych losowo z pełnej liczby rekordów N zmniejszamy poważnie ryzyko identyfikacji, ponieważ zidentyfikowanie jednostki⁵⁰⁾ w zbiorze n nie musi oznaczać jej zidentyfikowania w zbiorze N . Założenie to można zresztą w każdym przypadku stosunkowo łatwo sprawdzić metodą prostego komputerowego porównania zbiorów n i $N-n$.

W przypadku, gdy omawianą metodę stosujemy do zbioru reprezentacyjnego (np. badania gospodarstw domowych), z pełnego zbioru reprezentacyjnego N , wybieramy próbkę n . W tym przypadku jeszcze bardziej maleje ryzyko identyfikacji i ujawnienia informacji jednostkowej.

6.2.3. Ograniczenie szczegółowości danych

Jak wspomniano wyżej, ryzyko identyfikacji pośredniej wzrasta wraz ze wzrostem liczby zmiennych, które można wykorzystać do identyfikacji pośredniej badanej jednostki. Omawiana w tym punkcie metoda może być realizowana w dwojaki sposób:

- 1) poprzez usunięcie ze zbioru niektórych cech, które mogą służyć za podstawę do identyfikacji pośredniej (w połączeniu z dodatkowymi informacjami posiadanymi przez potencjalnego intruza),
- 2) poprzez zmniejszenie liczby wariantów cech rejestrowanych w ramach poszczególnych cech.

Ilustrując omawianą metodę przykładem danych zamieszczonych w przytoczonej tabl. 1, można by znacznie utrudnić pośrednią identyfikację, np. przez usunięcie ze zbioru takich cech, jak „płeć” i „wiek” (sposób pierwszy) lub zmniejszając liczbę wariantów cech, np. podając zamiast konkretnej liczby lat — tylko grupy wieku, np. 0—10, 11—18, 19—65, powyżej 65 (sposób drugi). Innym przykładem zastosowania drugiego sposobu ograniczenia szczegółowości danych byłoby ograniczenie charakterystyki stanu cywilnego do 2 wariantów (żonaty/nieżonaty, zamężna/niezamężna) lub zamiana konkretnego zawodu na grupę zawodów (przykładowo, według klasyfikacji zawodów i specjalności ISCO — tylko 10 tzw. grup „wielkich”).

Szczególnie niebezpieczne z punktu widzenia identyfikacji pośredniej są niektóre cechy mające warianty występujące rzadko, np. bardzo rzadko spotykane zawody, niektóre zawody uznawane za „męskie”, a występujące

⁵⁰⁾ Mowa wyłącznie o identyfikacji pośredniej, ponieważ identyfikatory bezpośrednie usuwa się ze zbioru w każdym przypadku przed udostępnieniem zbioru.

u kobiet (lub odwrotnie), duża liczba dzieci w rodzinie, niektóre zawody w korelacji z wiekiem, liczba dzieci w korelacji z wiekiem itp.

Przed udostępnieniem zbioru należy każdorazowo sprawdzić, czy nie występują w nim przypadki rzadkie, których liczba znajduje się poniżej poziomu uznanego za bezpieczny. Jest to szczególnie ważne w przypadkach udostępniania zbiorów dla małych obszarów — gmin, miast, a nawet województw. Liczbę przypadków rzadkich należy zbadać, przyjmując za podstawę zarówno poszczególne cechy, jak i ich skorelowanie z innymi cechami.

6.2.4. Zmiana grupowania danych

Inna metoda ograniczenia szczegółowości danych stosowana jest w przypadkach, gdy dane jednostkowe zawierają informacje wyrażone w wielkościach bezwzględnych zgrupowanych według klas wielkości, przykładowo:

TABL. 2

Miesięczny dochód w złotych	Liczba osób
0— 100	1
101— 500	2
501—1000	20
1001—1500	50
1501—2000	12
2001—3000	3
3001—5000	1

Jeśli w zbiorze liczba badanych jednostek w skrajnych grupach (najniższej i najwyższej) jest niewielka, co zwiększa niebezpieczeństwo identyfikacji pośredniej, można wprowadzić zmiany w grupowaniu, ograniczając zwłaszcza zakres klas w najniższych i najwyższych grupach dochodów, przykładowo:

TABL. 3

Miesięczny dochód w złotych	Liczba osób
Do 500	3
501—1000	20
1001—1500	50
1501—2000	12
Powyżej 2000	4

Podobny sposób zmniejszenia liczby grup w celu zmniejszenia niebezpieczeństwa identyfikacji można zastosować w przypadku, gdy dane jednostkowe zawierają ujęte w klasy informacje o wysokości wydatków na osobę, wysokości wydatków na cele kulturalne, na alkohol itp.

6.2.5. Usuwanie

Metodę tę stosuje się do przypadków, gdy dane jednostkowe zawierają cechy wyjątkowe i nietypowe lub też kombinacje takich nietypowych cech. Usunięcie ze zbioru takich przypadków jest niezbędne w celu utrudnienia identyfikacji badanego podmiotu za pomocą identyfikacji pośredniej.

Przykładem skrajnego przypadku są osoby o wieku powyżej 100 lat, a ilustracją rzadkich i nietypowych lub nawet wyjątkowych sytuacji — niektóre kombinacje płci, wieku, stanu cywilnego, dochodu itp., przykładowo: kobieta — kapitan żeglugi wielkiej, emeryt posiadający bardzo wysokie dochody, osoba w wieku powyżej 60 lat mająca małoletnie dzieci, osoba dojeżdżająca do pracy powyżej 300 km itp.

Mówiąc o sytuacjach nietypowych zakładamy, że występują one jako przypadki prawdziwe i że nie wynikają one z błędów rejestracji lub przetwarzania, zwykle bowiem tego rodzaju nietypowe kombinacje cech uwzględnia się w założeniach do automatycznej kontroli i pozostawia w zbiorze po dokładnym sprawdzeniu. Metodę usuwania stosuje się zwykle w dwóch odmianach.

Po pierwsze, cechę lub kombinację cech uznanych za wyjątkowe zastępuje się informacją „brak danych”. Użytkownik otrzymujący zbiór może być poinformowany, że tego rodzaju zapis nie wynika z niedokonałości badania, lecz z konieczności ukrycia cech wyjątkowych, z tym, że użytkownik nie będzie wiedział dokładnie na czym ta „wyjątkowość” polegała, nie będzie więc w stanie cech tych wykorzystać do identyfikacji pośredniej.

Po drugie, usuwa się ze zbioru całe rekordy zawierające wyjątkowe cechy lub kombinacje cech wyjątkowych. Tę odmianę stosuje się zwłaszcza wtedy, gdy badaniem są objęte m.in. osoby powszechnie znane, identyfikacja których na podstawie identyfikatorów pośrednich byłaby stosunkowo łatwa [Sch93, s. 59].

6.2.6. Modyfikacja danych

Metoda ta sprowadza się do dokonania takich zmian w treści danych źródłowych, aby uniemożliwić identyfikację poszczególnych badanych jednostek.

Modyfikacji danych należy dokonywać w taki sposób, aby zapobiegając niebezpieczeństwu identyfikacji zachować jednocześnie maksymalne podobieństwo do oryginalnego zbioru danych, tj., aby zminimalizować straty w wartości poznawczej modyfikowanego zbioru. Modyfikację stosuje się najczęściej w dwóch wariantach:

- wprowadzanie zakłóceń przypadkowych,
- zamiana danych.

Wprowadzanie zakłóceń przypadkowych (losowych)⁵¹⁾ polega na wprowadzeniu do zbioru danych jednostkowych informacji „zakłócającej” lub tzw. „szumu informacyjnego”.

Ochrona danych przed ujawnieniem polega w tym przypadku na tym, że oryginalną wartość cechy X zamienia się cechą X' , która stanowi w tym przypadku wartość zakłócającą.

Celem takiego postępowania jest utrudnienie potencjalnemu intruzowi uzyskania rzeczywistych charakterystyk respondenta i zmniejszenie ryzyka jego identyfikacji. Rzeczywiste zmniejszenie ryzyka identyfikacji zależy od wielkości wprowadzonego do cechy X zakłócenia e . Zakłócenie e ma charakter zmiennej losowej. Im większe jest to zakłócenie, tzn. im większa jest wariancja $V(e)$ tego zakłócenia, tym silniej zostaną zakłócone pierwotne dane w wyniku zamiany X przez X' . Istnieje zatem odwrotna relacja między zakłóceniem i ochroną z jednej strony oraz między zakłóceniem i użytecznością danych — z drugiej. Im większa jest wariancja zakłócenia (szumu) i tym samym zniekształcenie danych, tym lepiej chroniona jest poufność respondenta, ponieważ zniekształcenie to zmniejsza możliwość identyfikacji za pomocą identyfikatorów pośrednich, ale jednocześnie maleje użyteczność zniekształconych w ten sposób danych do dalszych analiz [Sch93, s. 62].

Wprowadzanie zakłóceń e do cech X badanych jednostek nie może być dokonywane w sposób mechaniczny. Zanim dokonamy wyboru odpowiedniego algorytmu modyfikacji w zbiorze, musi być on poddany szczegółowej analizie, zwłaszcza z punktu widzenia skali zmienności poszczególnych cech w wektorze zmiennych danego respondenta oraz stopnia zmienności danej cechy u różnych respondentów w modyfikowanym zbiorze. Dlatego dla każdej zmiennej wielkość zakłócenia e musi być określana niezależnie.⁵²⁾

Należy zauważyć, że metoda wprowadzania zakłóceń losowych może być stosowana wyłącznie w odniesieniu do modyfikacji zmiennych (cech) mierzalnych (ilościowych). Jej zastosowanie do modyfikacji cech niemierzalnych (jakościowych) mogłoby spowodować zasadnicze zniekształcenie struktury zbioru lub nawet wprowadzenie do zbioru sprzecznych pod względem logicznym wartości cech. Przykładowo, próba modyfikowania tą metodą cechy „płeć” o symbolach: 1 — „mężczyźni” i 2 — „kobiety”, przy wprowadzeniu w charakterze wielkości zakłócających $e = -1, 0, +1$, otrzymalibyśmy: przy $e = 0$ symbol płci by się nie zmienił — nie wprowadzilibyśmy więc ani modyfikacji, ani ochrony informacji jednostkowej; przy $e = -1$, pozycje dotyczące kobiet zamienilibyśmy na mężczyzn, natomiast w pozycjach dotyczących mężczyzn (płeć = 1) otrzymalibyśmy nie

⁵¹⁾ Ang.: *random perturbation*.

⁵²⁾ Metody modyfikacji zbiorów danych jednostkowych poprzez wprowadzanie zakłóceń są przedmiotem wielu badań specjalistycznych. Czytelników zainteresowanych w pogłębianiu wiedzy na ten temat możemy odesłać do specjalistycznych pozycji literatury, m.in.: [Cox86, Gre90].

istniejący symbol 0. Podobnie, przy $e = +1$, pozycje dotyczące mężczyzn zamienilibyśmy na symbol 2 (kobiety), natomiast dla pozycji „kobiety” (symbol=2) otrzymalibyśmy nie istniejący symbol 3.

Z powyższych względów, omawiana w tym punkcie metoda nie może być stosowana do modyfikacji cech jakościowych, a inne sposoby ochrony pozycji jednostkowych muszą w tym przypadku wynikać z analizy rodzaju i liczby cech jakościowych w poszczególnych pozycjach zbioru i ryzyka identyfikacji. Jeśli dane jednostkowe nie zawierają dużej liczby cech jakościowych oraz nie występują wśród nich cechy rzadkie i nietypowe, można pozostawić je w zbiorze bez modyfikowania. Jeśli natomiast cech tych jest wiele oraz są wśród nich cechy o wartościach nietypowych, co mogłoby ułatwić identyfikację pośrednią, wtedy należy w stosunku do tych cech zastosować inną metodę ochrony, tzn. jedną z metod omawianych poprzednio względnie metodę zamiany danych, o której będzie mowa niżej.

Zamiana danych.⁵³⁾ Modyfikując zbiór danych jednostkowych metodą zamiany danych, traktujemy ten zbiór jako macierz, w której mamy do czynienia z liczbą n wierszy dla n jednostek badanych oraz z liczbą m zmiennych charakteryzujących każdą badaną jednostkę. Dane jednostkowe mają w tym przypadku postać macierzy o rozmiarach $n \times m$.

Metodę zamiany danych rozpatrzmy na przykładzie macierzy zawierającej 8 pozycji jednostkowych (8 rekordów) pozbawionych już identyfikatorów bezpośrednich, przy czym każda pozycja jednostkowa zawiera 4 zmienne:

X — płeć z 2 wariantami: mężczyzna — 1, kobieta — 2,

Y — stan cywilny z 4 wariantami: wolny — 1, żonaty/zamężna — 2, wdowiec/wdowa — 3, rozwiedziony/rozwiedziona — 4,

Z — wykształcenie z 4 wariantami: bez wykształcenia — 0, podstawowe — 1, średnie — 2, wyższe — 3,

N — liczba dzieci z k wariantami: 0, 1, 2, 3, ..., k .

Przyjmijmy, że macierz pierwotna (oryginalna) ma następującą postać (macierz M_0):

TABL. 4

Płeć	Stan cywilny	Wykształcenie	Liczba dzieci
1	1	2	0
2	2	3	4
1	4	1	2
2	3	0	0
1	2	2	5
2	1	1	3
1	2	3	1
2	1	1	6

⁵³⁾ Ang.: data swapping. Metoda ta w literaturze nazywana jest również metodą transformacji wielowymiarowej.

W pierwszym wierszu macierzy znajduje się informacja o respondencie płci męskiej, kawalerze, ze średnim wykształceniem i liczbie dzieci = 0; w wierszu drugim: kobieta, zamężna, z wyższym wykształceniem, posiadająca 4 dzieci itd.

Metoda zamiany danych zmierzająca do uniemożliwienia lub poważnego utrudnienia identyfikacji pośredniej i ujawnienia danych polega na przekształceniu macierzy oryginalnej na nową macierz, w której dokonano zamiany niektórych cech między poszczególnymi wierszami macierzy. Dokonując tej zamiany należy zadbać o to, aby w tablicach wynikowych opracowanych na podstawie przekształconego zbioru, zbiorcze wyniki były maksymalnie zbliżone do wyników uzyskanych na podstawie macierzy oryginalnej.

Metoda zamiany danych została opracowana początkowo w celach modyfikacji cech jakościowych. Obecnie stosowana jest również do modyfikowania cech ilościowych (mierzalnych).

Opierając się na danych zamieszczonych w macierzy M zawierającej dane oryginalne o 8 jednostkach badanych możemy obliczyć:

TABL. 5. LICZBA OSÓB WEDŁUG WYKSZTAŁCENIA I STANU CYWILNEGO

Wyszczególnienie	Liczba osób	Stan cywilny			
		wolny	żonaty/ zamężna	wdowiec/ wdowa	rozwidziony/a
Razem	8	3	3	1	1
mężczyźni	3	2	0	1	0
kobiety	5	1	3	0	1

Modyfikując zbiór metodą zamiany symbolu stanu cywilnego możemy utworzyć nową macierz M_1 .

TABL. 6 MACIERZ M

Płeć	Stan cywilny	Wykształcenie	Liczba dzieci
1	2	2	0
2	1	3	4
1	3	1	2
2	4	0	0
1	1	2	5
2	2	1	3
2	1	3	1
2	2	1	6

Obliczając obecnie na podstawie zmodyfikowanej macierzy analogiczną tablicę o liczbie osób według płci i stanu cywilnego otrzymamy (patrz tabl. 7):

TABL. 7

Wyszczególnienie	Liczba osób	Stan cywilny			
		wolny	żonaty/ /zameżna	wdowiec/ /wdowa	rozwidziony/a
Razem	8	3	3	1	1
mężczyźni	3	1	1	1	0
kobiety	5	2	2	0	1

Po zmodyfikowaniu zbioru metodą zamiany otrzymaliśmy nieco zniekształconą strukturę zbioru według płci i stanu cywilnego. Oznacza to, że każdorazowo przed udostępnieniem zbioru po dokonaniu zamiany w poszczególnych jego pozycjach należy przeprowadzić analizę, czy z punktu widzenia celów poznawczych, jakim ma służyć udostępniany zbiór, wprowadzone modyfikacje są dopuszczalne, a jeśli nie — to należy poszukiwać innych sposobów zamiany. Rzeczywiste algorytmy modyfikacji zbiorów metodą zamiany są dość skomplikowane. Nie jesteśmy w stanie w tej pracy ich szerzej omówić. Zainteresowanych Czytelników możemy odesłać m.in. do [Dal82], [Rei84].

Wprowadzanie zakłóceń metodą kompensacji. Metoda ta polega na modyfikowaniu danych jednostkowych w przypadkach, gdy w zbiorze występuje zbyt mała liczba jednostek odznaczających się identyczną kombinacją określonych cech jakościowych, co mogłoby ułatwić identyfikację badanej jednostki. Jako minimalną liczbę identycznych kombinacji cech jakościowych (częstość) przyjmuje się liczbę progową 3.

W zależności od konkretnej liczby jednostek o rzadkich kombinacjach cech w zbiorze metodę kompensacji stosuje się w kilku wariantach.

Wariant 1. Jeśli w zbiorze częstość występowania identycznych kombinacji cech jakościowych wyniesie 2 oraz 1, to pozycję o częstości 1 dodaje się do pozycji o częstości 2, a pozycji o częstości 1 nadaje się wartość 0.

Wariant 2. Jeśli w zbiorze występują trzy pozycje o unikalnej kombinacji cech jakościowych, to liczebność pierwszej zamienia się na 3, a dwóch pozostałych na 0.

Wariant 3. Jeśli w zbiorze występuje wyłącznie jedna pozycja o określonej kombinacji cech, to do pozycji tej dodaje się pozycję o największej liczebności, a pozycję o liczebności 1 zamienia się na 0.

Wariant 4. Jeśli w zbiorze występują 2 przypadki o identycznej kombinacji cech (liczebność tych pozycji = 2), to odejmuje się jeden rekord od pozycji o największej częstości występowania i dodaje się go do rekordów o częstości występowania 2.

Wariant 5. Jeśli częstość występowania określonej kombinacji cech wyniesie 2 (dwie identyczne pozycje w zbiorze) oraz kilka pozycji o częstości 3, to pozycjom o częstości występowania 2 nadaje się wartość 0 oraz zwiększa się o 2 pozycje o częstości występowania 3.

Metodę zakłóceń metodą kompensacji można stosować nie tylko w stosunku do cech jakościowych, ale i do cech ilościowych (mierzalnych).

Omawiany sposób wprowadzania zakłóceń metodą kompensacji można zilustrować następującym przykładem:

Oryginalny zbiór danych składający się z 50 pozycji i zawierający 4 cechy jakościowe ($M1-M4$) oraz jedną cechę ilościową W , po zagregowaniu według wartości cech $M1-M4$ ma postać:

TABL. 8

Nr poz.	M1	M2	M3	M4	Częstość	Dochód ogółem	Dochód przeciętny
1	1	1	1	1	10	6000	600
2	1	1	1	2	5	3600	720
3	1	1	1	3	3	3000	1000
4	1	1	2	1	1	850	850
5	1	1	2	2	1	800	800
6	1	2	1	1	2	1800	900
7	1	2	1	2	15	14000	934
8	2	1	1	1	1	900	900

W celu zmniejszenia niebezpieczeństwa identyfikacji zbioru metodą identyfikacji pośredniej na skutek małej liczby pozycji odznaczających się rzadką kombinacją cech $M1-M4$:

- 1) pozycje unikalne nr 4, 5 i 8 zmodyfikujemy według wariantu 2, zwiększając częstość pozycji nr 4 na 3 i wprowadzając do poz. 5 i 8 częstość 0;
- 2) pozycję nr 6 możemy zmodyfikować według wariantu 4, zmniejszając pozycję nr 7 o 1 oraz zwiększając pozycję 6 o 1.

Po wprowadzeniu tych modyfikacji, zmodyfikowany zbiór będzie miał postać:

TABL. 9

Nr poz.	M1	M2	M3	M4	Częstość	Dochód ogółem	Dochód przeciętny
1	1	1	1	1	10	6000	600
2	1	1	1	2	5	3600	720
3	1	1	1	3	3	3000	1000
4	1	1	2	1	3	2550	850
5	1	1	2	2	0	x	x
6	1	2	1	1	3	2734	911
7	1	2	1	2	14	13066	933
8	2	1	1	1	0	x	x

Usunęliśmy w ten sposób ze zbioru wszystkie pozycje, w których występowały rzadkie kombinacje cech, a których obecność w zbiorze mogłaby ułatwić ujawnienie danych indywidualnych metodą identyfikacji pośredniej.

Omawiany sposób wprowadzania zakłóceń metodą kompensacji został opracowany przez służby statystyczne dużych miast w Niemczech na początku lat dziewięćdziesiątych dla umożliwienia wykorzystania danych spisu ludności do celów badań naukowych z jednoczesnym zapewnieniem pełnej ochrony danych indywidualnych [App92].

Mikroagregowanie. Metoda ta polega na tworzeniu ze zbioru danych, zawierającego wyłącznie pojedyncze pozycje danych, zbioru danych częściowo zagregowanych w taki sposób, aby poszczególne pozycje odpowiadały kryteriom progowym oraz kryterium przeważającego udziału, o których była mowa w p. 5.2. Po przeprowadzeniu modyfikacji zbioru metodą mikroagregowania, zamiast zbioru jednostkowego udostępnia się zbiór częściowo zagregowany według minimalnej wielkości progowej m (w Polsce $m=3$) oraz zasady przeważającego udziału k (w Polsce $k=75\%$).

Uzyskane w ten sposób „mikroagregaty” spełniają dla użytkownika rolę fikcyjnych pozycji jednostkowych zwanych w literaturze danymi quasi-jednostkowymi (ang. pivot). Tworząc takie quasi-jednostkowe pozycje należy dążyć do tworzenia maksymalnie jednorodnych grup danych, aby uzyskiwane wyniki analizy takiego „spreparowanego” zbioru były maksymalnie zbliżone do wyników analizy zbioru oryginalnego.

Z matematycznego punktu widzenia problem ten polega na minimalizacji „wariancji międzygrupowych”. Im mniejsza będzie wariancja wewnątrz grup, tym bardziej jednorodne będą dane grupowe i tym mniejsze straty informacji powodowane operacją mikroagregowania.

Bardziej szczegółowe informacje na ten temat zainteresowany Czytelnik może znaleźć w [Def92] oraz [Anw93].

6.3. Metody zapobiegania identyfikacji danych w informacjach zagregowanych (w makrodanych)

Zgodnie z tym, co przedstawiono w rozdziale 5, udostępnianie przez statystykę zbiorów danych źródłowych (mikrodanych) po ich całkowitej anonimizacji wynika z dążenia statystyków do maksymalnego zaspokojenia specyficznych potrzeb informacyjnych niektórych grup użytkowników, którzy mają tradycyjnie dobrą współpracę ze służbami statystycznymi i w której statystycy są szczególnie zainteresowani. Do takich użytkowników należą przede wszystkim pracownicy naukowcy oraz instytucje naukowo-badawcze.

Ostatnio coraz częściej w uzyskiwaniu mikrodanych zainteresowane są również różne instytucje naukowe i badawcze z zagranicy. Ten ostatni rodzaj zapotrzebowania będzie wzrastać w miarę rozszerzania się analiz związanych z procesami integracji społeczno-gospodarczej w Europie i w świecie.

Przypomnieć również należy, że kiedy mówiliśmy o udostępnianiu zbiorów mikrodanych, zawsze mieliśmy na myśli zbiory mikrodanych

pochodzące z badań społecznych, w których podmiotami badania są osoby, gospodarstwa domowe lub rodziny. Mikrodane charakteryzujące działalność gospodarczą podmiotów są traktowane jako całkowicie poufne i nie są z zasady udostępniane nikomu, nie mają więc do nich zastosowania metody zapobiegania identyfikacji, o których była mowa w p. 6.2.

Inaczej przedstawia się sprawa w przypadku udostępniania wyników badań statystycznych zarówno w dziedzinie badań społecznych, jak i badań działalności gospodarczej.

Z jednej strony, statystycy pragną jak najlepiej wywiązać się z ciężących na nich zadań, jak najpełniejszego udostępniania wyników badań, z drugiej jednak strony muszą bezwzględnie zapobiegać zagrożeniom identyfikacji i ujawnienia danych w udostępnianych wynikach.

Główne zagrożenie stwarza w tej dziedzinie naruszenie wielkości progowej oraz zasady przeważającego udziału (patrz p. 5.2).

Największe niebezpieczeństwo ujawnienia danych w wynikach badań powstaje w następujących trzech przypadkach:

- 1) w przypadku udostępniania wyników badań jednostek gospodarczych w dziedzinach o dużej koncentracji działalności (transport, łączność, niektóre grupy i klasy działalności przemysłowej),
- 2) w przypadku szczegółowych wyników badań według rodzaju działalności (klas i podklas EKD),
- 3) w przypadku opracowań w szczegółowych układach terytorialnych, zwłaszcza w kombinacji z rodzajami działalności.

Przepisy art. 38, ust. 2 ustawy o statystyce publicznej o ograniczeniach w publikacjach i udostępnianiu danych charakteryzujących wyniki ekonomiczne działalności podmiotów gospodarczych (patrz p. 3.3) oraz wykładnia „Komisji do spraw ochrony tajemnicy statystycznej w zakresie udostępniania danych statystycznych” stanowią, że danymi charakteryzującymi wyniki ekonomiczne działalności podmiotów gospodarczych są dane wartościowe⁵⁴) powinny zachęcić (a właściwie zmusić) statystyków do poszukiwania takich rozwiązań, aby przestrzegając ściśle przepisów ustawy, nie dopuścić jednocześnie do istotnego zaburzenia publikowanych informacji.

Sygnalizowany dylemat nie jest dylematem specyficznym dla polskiej statystyki. Ponieważ nasze rozwiązania prawne dotyczące statystyki, a zwłaszcza przepisy odnoszące się do ochrony danych, wzorowane są w dużym stopniu na rozwiązaniach stosowanych w innych krajach demokratycznych o gospodarce rynkowej, to również w trakcie realizacji tych przepisów napotykamy na podobne trudności jak statystycy wielu krajów.

Podstawowym obowiązkiem statystyków przed udostępnieniem informacji jest sprawdzenie, czy w jakiegokolwiek pozycji tablicy wynikowej nie wystę-

⁵⁴) Ustalenia Komisji do spraw ochrony tajemnicy statystycznej w zakresie udostępniania danych statystycznych, Warszawa 1996.05.15.

pują informacje naruszające zasadę wielkości progowej lub zasadę przeważającego udziału. Ten obowiązek ciąży w jednakowym stopniu na statystykach we wszystkich krajach i nie budzi, na ogół, większych wątpliwości i zastrzeżeń. Problemy i wątpliwości oraz sposoby ich rozwiązywania zaczynają się natomiast od momentu, kiedy takie pozycje w tablicach rzeczywiście występują, w związku z czym statystycy muszą podjąć decyzję, czy przestrzegając formalnie litery przepisów prawnych mają zrezygnować całkowicie z publikacji szczegółowych tablic, czy *poszukiwać sposobów takiej modyfikacji układu i treści tablic lub zbiorów danych będących podstawą opracowania tablic, aby nie naruszając przepisów o tajemnicy statystycznej, nie rezygnować z publikowania szczegółowych danych, nawet za cenę wprowadzenia do treści tablic pewnych zmian, które wprawdzie nieco obniżają wartość poznawczą publikowanych informacji, ale które nie pozbawiają użytkownika całkowicie oczekiwanej przez niego informacji.*

W literaturze statystycznej oraz w praktycznej działalności wielu urzędów statystycznych można znaleźć szereg metod zapobiegania naruszeniu tajemnicy statystycznej w praktyce publikowania informacji w szczegółowym układzie. Metody te charakteryzuje schemat pokazany na wykr. 4.

Z wykresu 4 wynika, że w dążeniu statystyków do znalezienia kompromisu między obowiązkiem bezwzględного przestrzegania przepisów o tajemnicy statystycznej a chęcią maksymalnego zaspokojenia potrzeb użytkowników informacji, wypracowano szereg metod zapobiegania ujawnieniu danych w publikowanych tablicach. Metody te podzielić można na dwie podstawowe grupy:

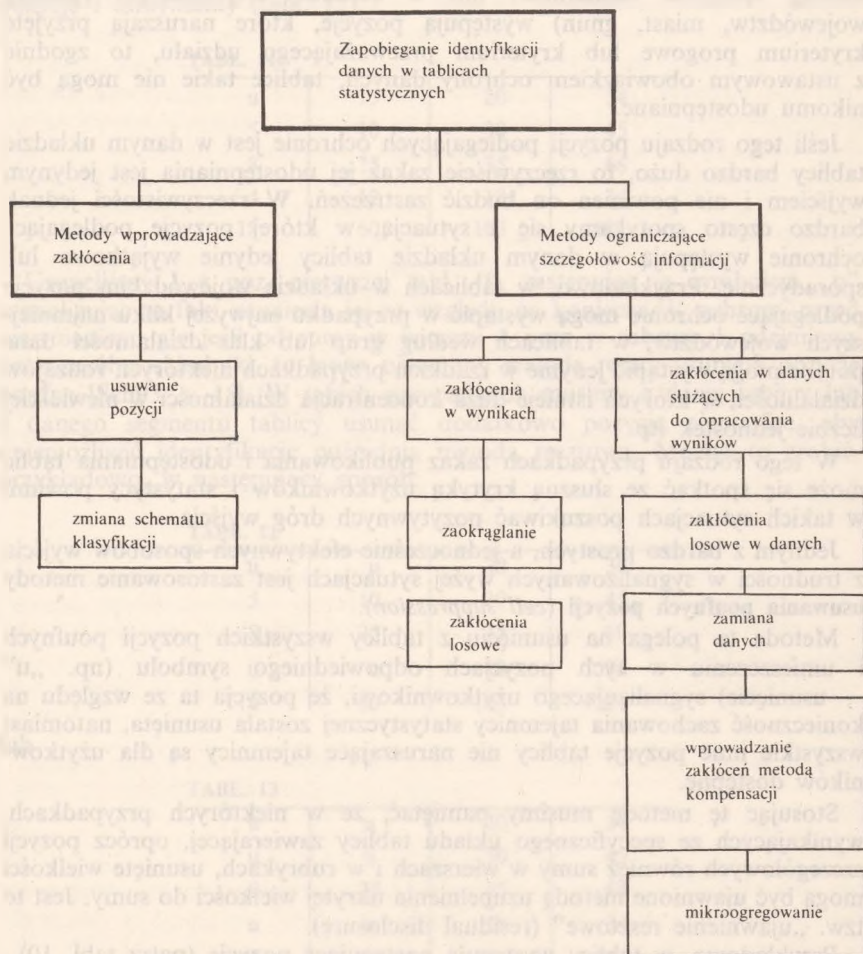
- metody ograniczania szczegółowości informacji,
- metody wprowadzania zakłóceń.

Obie wymienione grupy metod zmierzających do zapobieżenia identyfikacji mikrodanych mogą być realizowane zarówno w odniesieniu do zagregowanych danych (do poszczególnych pozycji tablic), jak i do danych źródłowych, na podstawie których tablice te są opracowywane. W tym drugim przypadku zasady zapobiegania identyfikacji są identyczne jak przy bezpośrednim udostępnianiu mikrodanych użytkownikom, o czym była mowa obszerniej w poprzednich punktach.

Przedstawiając w niniejszej pracy omówienie stosowanych metod na podstawie dostępnych opracowań, należy podkreślić, że na obecnym etapie rozwoju naszej statystyki, wobec całkowitego braku doświadczeń w tej dziedzinie, trudno byłoby zasugerować, które z omawianych metod zasługują na preferencje w poszczególnych dziedzinach naszej statystyki. Wątpliwe, czy w ogóle można mówić o bezwzględnych zaletach jednej metody nad inną. Wszystko zależy od rodzaju danych, jakimi dysponujemy w poszczególnych tematach badań, wymaganego poziomu zachowania

poufności oraz rzeczywistych potrzeb informacyjnych użytkowników w każdym temacie badań i opracowań. Podstawowym celem prezentacji tych metod jest chęć zapoznania Czytelników z kierunkami prac w tej dziedzinie na świecie oraz zasygnalizowanie konieczności podjęcia szerszych badań w oparciu o nasze własne zasoby informacyjne i potrzeby naszych najważniejszych użytkowników.

Wykres 4 METODY ZAPOBIEGANIA UJAWNIANIA DANYCH W TABLICACH STATYSTYCZNYCH



6.3.1. Metody ograniczające szczegółowość informacji

Ze schematu na wykr. 4 wynika, że metoda ograniczania szczegółowości informacji w tablicach może być stosowana w dwóch wariantach:

- metodą usuwania pozycji podlegających ochronie oraz
- metodą zmian w grupowaniu cech.

Metoda usuwania. Jeśli z analizy treści opracowania wynika, że w tablicach opracowanych w jakimś bardziej szczegółowym przekroju według rodzajów działalności lub w układzie terytorialnym (według województw, miast, gmin) występują pozycje, które naruszają przyjęte kryterium progowe lub kryterium przeważającego udziału, to zgodnie z ustawowym obowiązkiem ochrony danych, tablice takie nie mogą być nikomu udostępniane.

Jeśli tego rodzaju pozycji podlegających ochronie jest w danym układzie tablicy bardzo dużo, to rzeczywiście zakaz jej udostępniania jest jedynym wyjściem i nie powinien on budzić zastrzeżeń. W rzeczywistości jednak bardzo często spotykamy się z sytuacją, w której pozycje podlegające ochronie występują w danym układzie tablicy jedynie wyjątkowo lub sporadycznie. Przykładowo, w tablicach w układzie wojewódzkim pozycje podlegające ochronie mogą wystąpić w przypadku najwyżej kilku najmniejszych województw, w tablicach według grup lub klas działalności dane poufne mogą wystąpić jedynie w rzadkich przypadkach niektórych rodzajów działalności, w których istnieje duża koncentracja działalności w niewielkiej liczbie jednostek itp.

W tego rodzaju przypadkach zakaz publikowania i udostępniania tablic może się spotkać ze słuszną krytyką użytkowników i statystycy powinni w takich sytuacjach poszukiwać pozytywnych dróg wyjścia.

Jednym z bardzo prostych, a jednocześnie efektywnych sposobów wyjścia z trudności w sygnalizowanych wyżej sytuacjach jest zastosowanie metody **usuwania poufnych pozycji** (*cell suppression*).

Metoda ta polega na usunięciu z tablicy wszystkich pozycji poufnych i umieszczeniu w tych pozycjach odpowiedniego symbolu (np. „u” — usunięcie) sygnalizującego użytkownikowi, że pozycja ta ze względu na konieczność zachowania tajemnicy statystycznej została usunięta, natomiast wszystkie inne pozycje tablicy nie naruszające tajemnicy są dla użytkowników dostępne.

Stosując tę metodę musimy pamiętać, że w niektórych przypadkach, wynikających ze specyficznego układu tablicy zawierającej, oprócz pozycji szczegółowych również sumy w wierszach i w rubrykach, usunięte wielkości mogą być ujawnione metodą uzupełnienia ukrytej wielkości do sumy. Jest to tzw. „ujawnienie resztowe” (*residual disclosure*).

Przykładowo, w tablicy występują następujące pozycje (patrz tabl. 10).

TABL. 10

1	13	20	34
5	10	30	45
8	25	35	68
4	12	18	34
18	60	103	181

Pozycja pierwsza (rubr. 1, wiersz 1) tabl. 10 nie może być publikowana ze względu na niebezpieczeństwo ujawnienia pośredniego. Usuniemy więc tę pozycję i otrzymamy (tabl. 11):

TABL. 11

u	13	20	34
5	10	30	45
8	25	35	68
4	12	18	34
18	60	103	181

Usunęliśmy 1 z poz. pierwszej tabl. 10, zastępując ją symbolem „u” sygnalizującym fakt usunięcia jej ze względu na konieczność ochrony przed ujawnieniem, ale jeśli od sumy w wierszu 1 oraz w rubryce 1 odejmiemy poszczególne składniki, to łatwo ujawnimy metodą reszty usuniętą pozycję poufną [Sch93, s. 15]. W takich przypadkach musimy z danej tablicy lub z danego segmentu tablicy usunąć dodatkowo pozycje niepoufne, aby uniemożliwić identyfikację pośrednią metodą resztową. Można to zrobić, przykładowo, w następujący sposób:

TABL. 12

u	u	20	34
5	10	30	45
8	25	35	68
u	u	18	34
18	60	103	181

lub

TABL. 13

u	u	20	34
u	u	30	45
8	25	35	68
u	u	18	34
18	60	103	181

Dodatkowe usuwanie pozycji jawnych w celu uniemożliwienia identyfikacji pozycji poufnych metodą resztową nosi nazwę **usuwania uzupełniającego** (complementary suppression) [Sch93, s. 15], [Cox86, s. 139].

Usuwanie uzupełniające nie może być przeprowadzane w sposób mechaniczny. Najczęściej zaleca się zastosowanie takiej metody usuwania, aby zminimalizować ogólną liczbę usuwanych pozycji oraz zminimalizować ogólną wartość usuniętych pozycji. W zamieszczonym wyżej przykładzie, stosując metodę usuwania uzupełniającego w obu przypadkach (tabl. 12 i 13) usunęliśmy 4 pozycje elementarne tablicy, natomiast jeśli chodzi o ogólną wartość, to w tabl. 12 usunięta wartość wynosi 14, natomiast w tabl. 13 usunięta wartość wyniesie 16.

Wielu autorów uważa, że przy stosowaniu usuwania uzupełniającego większe znaczenie ma minimalizowanie liczby usuwanych pozycji, a nie ich wartości, natomiast usuwając dane wartościowe należy preferować metodę zachowywania wartości skrajnych i usuwania raczej pozycji o wartościach zbliżonych do średniej lub mediany [Sch93, s. 16].

Wprowadzanie zmian w klasyfikacji. Metoda ta polega na wprowadzaniu zwiększonej agregacji do tablic i tym samym na zmniejszeniu prawdopodobieństwa pozycji poufnych.

Metoda ta ma zastosowanie zarówno do pozycji zawierających liczebności badanych podmiotów, jak i dane wartościowe. Omawiana metoda może być stosowana w dwóch wariantach:

- a) metodą zmniejszania szczegółowości tablicy,
- b) metodą zmiany w grupowaniu cech [Dal88].

Zmniejszanie szczegółowości tablic. W przypadku gdy ze względu na bardzo szczegółowy układ tablicy występuje w niej **znaczna liczba pozycji** podlegających ochronie byłoby nieracjonalne stosowanie metody usuwania. Tablica taka byłaby mało użyteczna, zwłaszcza gdyby obok usuwania pierwotnego musielibyśmy zastosować usuwanie uzupełniające.

W takich przypadkach bardziej racjonalną metodą uniknięcia ujawnienia pozycji podlegających ochronie jest **metoda zmniejszania szczegółowości** układu tablicy. Polega ona na zmniejszeniu liczby wierszy i rubryk tablicy, co w wielu przypadkach pozwala zlikwidować lub znacznie ograniczyć liczbę pozycji poufnych.

Jeśli, przykładowo, mamy tablicę, w której występuje 5 cech: *U*, *W*, *X*, *Y* i *Z*:

TABL. 14

Wyszczególnienie	<i>U</i>	<i>W</i>	<i>X</i>	<i>Y</i>	<i>Z</i>	Razem
Razem	5	8	10	12	16	51
mężczyźni	2	7	5	7	9	30
kobiety	3	1	5	5	7	21

i w rubr. *U* i *W* występują małe liczebności jednostek podlegające ochronie, można połączyć rubryki z cechami *U* i *W*, w których występują te niewielkie liczebności i opracować tablicę mniej szczegółową — patrz tabl. 15.

Jeśli pokazany w tabl. 14 stopień zmniejszenia szczegółowości tablicy nie wystarcza, gdyby bowiem nadal występowały w niej małe liczebności pozycji stwarzające niebezpieczeństwo ujawnienia podmiotów metodą identyfikacji pośredniej, to niestety trzeba by jeszcze bardziej zmniejszyć rozmiar i szczegółowość tablicy.

TABL. 15

Wyszczególnienie	U + W	X	Y	Z	Razem
Razem	13	10	12	16	51
mężczyźni	9	5	7	9	30
kobiety	4	5	5	7	21

Zmniejszenie szczegółowości tablicy może mieć również miejsce w przypadku prezentowania w tablicy wskaźników wartościowych. Przykładowo, jeśli tablica obrazuje wartość produkcji według klas EKD w układzie regionalnym i, na skutek dużego rozdrobnienia wielkości liczbowych oraz koncentracji produkcji w niewielkiej liczbie jednostek, wystąpi bardzo dużo pozycji podlegających ochronie, to można zmniejszyć szczegółowość zarówno w podziale terytorialnym, jak i w układzie klasyfikacji, np. zamiast układu wojewódzkiego opracować dane według grup województw, a zamiast 4-cyfrowych klas EKD wyodrębnić jedynie 3-cyfrowe grupy rodzaju działalności.

Jeśli w wyniku wspomnianych wyżej zabiegów zmniejszenia szczegółowości układu tablicy nadal pozostaną w niej **pojedyncze** pozycje poufne, to można je utajnić metodą **usuwania**.

Metoda zmian w grupowaniu cech. Metodę tę rozpatrzmy na przykładzie zamieszczonym w pracy Daniela Schackis [Sch93, s. 22].

Przypuśćmy, że mamy zbiór źródłowy zawierający następujące dane — patrz tabl. 16.

TABL. 16

Wiek	<12	12	13	14	15	16	17	18	19	20	20 i więcej	Razem
Razem	25	5	4	8	8	5	6	6	8	5	20	100
nie karani	23	3	3	7	7	3	4	4	7	4	15	80
karani	2	2	1	1	1	2	2	2	1	1	5	20

Na podstawie zbioru źródłowego przedstawionego w tabl. 16 opracujemy tablicę zawierającą informacje zagregowane:

TABL. 17

Wiek	<12	12—15	16—19	20 i więcej	Razem
Razem	25	25	25	25	100
nie karani	23	20	18	19	80
karani	2	5	7	6	20

Z analizy poszczególnych pozycji zamieszczonych w tabl. 17 wynika, że w wierszu „karani”, w rubryce poniżej 12 lat występuje pozycja zawierająca liczebność 2 podlegająca ochronie ze względu na duże niebezpieczeństwo identyfikacji pośredniej. Musimy więc dokonać zmiany w grupowaniu cech, jak to pokazano w tabl. 18.

TABL. 18

Wiek	Poniżej 13	13—16	17—20	21 i więcej	Razem
Razem	30	25	25	20	100
nie karani	26	20	19	15	80
karani	4	5	6	5	20

W ilustrowanym przykładzie tabl. 18 szczegółowość tablicy nie została zmieniona, tj. zachowaliśmy identyczną liczbę wierszy i rubryk. Dokonałiśmy jedynie zmian w grupowaniu cech, co pozwoliło uniknąć ujawnienia pozycji poufnej. Ponieważ zmiany w grupowaniu są nieznaczne, to można sądzić, że nie spowodują one istotnego zubożenia wartości informacyjnej tej tablicy.

Jeśli z analizy konkretnego układu danych w tablicach wyniknie taka konieczność, to oczywiście obie omówione wyżej metody: zmniejszania szczegółowości układu tablicy i dokonywanie zmian w grupowaniu cech można zastosować jednocześnie.

6.3.2. Metody wprowadzające zakłócenia

Zaokrąglenie. Metoda zaokrąglenia danych w szczegółowych pozycjach tablicy jest jedną z metod wprowadzania zakłóceń do danych wynikowych w celu utrudnienia identyfikacji pośredniej, bez wprowadzania jakichkolwiek modyfikacji do zbioru źródłowego, na podstawie którego została opracowana tablica. W niektórych przypadkach metoda ta uznawana jest za lepszą, w porównaniu do metody usuwania.

Metoda ta polega na zaokrągleniu każdej liczebności w pozycjach tablicy wymagających ochrony do najbliższej liczby całkowitej przyjętej w charakterze liczby bazowej do zaokrąglenia.

Teoretycznie, w charakterze liczby bazowej do zaokrąglenia, można przyjąć dowolną liczbę większą od wielkości progowej, poniżej której dane tablicy traktowane są jako poufne.⁵⁵⁾ Im większą liczbę przyjmiemy

⁵⁵⁾ Jak wiadomo, w Polsce zgodnie z art. 38 ustawy o statystyce publicznej, w przypadku gdy dane charakteryzują działalność gospodarczą jednostek, wielkość progowa wynosi 3. Dla innych danych brak dokładnego ustalenia wielkości progowej, zależy to bowiem od liczby i rodzaju cech charakteryzujących badane podmioty. Można wspomnieć przy okazji, że Biuro Spisów USA przy publikowaniu wyników spisów ludności stosuje wielkość progową = 5 dla gospodarstw domowych oraz 15 dla osób [Cox86, s. 139].

w charakterze liczby bazowej, tym lepiej będzie chroniona poufna liczebność w tablicy, ale tym większe zniekształcenie wprowadzimy do danych tablicy. Specjaliści amerykańscy sugerują, iż najlepiej w charakterze liczby bazowej przyjąć liczbę 5 [Cox86, s. 141], [Dal88].

Zastosowanie tej metody rozpatrzmy na prostym przykładzie zaczerpniętym z pracy Daniela Schackis [Sch93].

Przyjmijmy, że mamy tablicę lub segment tablicy, w której występują następujące liczebności jednostek:

TABL. 19

0	4	0	15	20
15	10	10	20	55
2	10	10	3	25
2	6	15	12	35
20	30	35	50	135

Występujące w tabl. 19 małe liczebności podlegające ochronie wydrukowano pogrubioną czcionką. Pewne wątpliwości może budzić niepogrubienie liczebności 0 występującej w wierszu 1 rubr 3. Jak twierdzi L. Cox w [Cox86, s. 140] w określonych przypadkach liczebność zerowa może pomóc intruzowi w identyfikacji innych jednostek.

Przyjmując za bazową liczbę do zaokrągleń liczbę 5, zaokrąglimy wszystkie liczby do najbliższego 0 lub 5, tj. liczebności kończące się na cyfrę 1 i 2 zaokrąglimy w dół i zamienimy na liczby kończące się na 0, a wszystkie liczebności kończące się na 3 lub 4 — zaokrąglimy w górę i zamienimy na liczby kończące się na 5. Liczebności kończące się na 0 lub 5 pozostają bez zmian. Otrzymamy w ten sposób „zakłócony” zbiór pokazany w tabl. 20.

TABL. 20

0	5	0	15	20
15	10	10	20	55
0	10	10	5	25
0	5	15	10	35
20	30	35	50	135

W pokazanym w tabl. 20 przykładzie, zaokrągleń dokonaliśmy wyłącznie w pozycjach tablicy obrazujących liczebności jednostek charakteryzujących się określonymi cechami. Nie dokonaliśmy modyfikacji w pozycjach sumarycznych według wierszy i rubryk. Na temat tego, czy modyfikować metodą zaokrągleń również pozycje sumaryczne nie ma wśród specjalistów

jednolitego poglądu. Częściej jednak zaleca się pozostawianie pozycji sumarycznych bez modyfikowania. Powoduje to oczywiście brak arytmetycznej zgodności w tablicach, do której jesteśmy w Polsce tak przyzwyczajeni. Jednocześnie jednak, użytkownicy otrzymują w tym przypadku sumaryczne dane bez zniekształceń, co w wielu przypadkach może być więcej warte aniżeli pełna zgodność arytmetyczna uzyskana w wyniku zniekształcenia informacji.

Dzięki zmodyfikowaniu tablicy metodą zaokrągleń usunęliśmy z niej pozycje podlegające ochronie, ale wprowadziliśmy pewne zniekształcenia rzeczywistych informacji przedstawionych w tablicy.

Pokazana na zamieszczonym wyżej uproszczonym przykładzie metoda zaokrąglania nosi nazwę **zaokrąglania tradycyjnego** (*conventional rounding*). Wyraża się ona w tym, że zawsze stosuje jednakowy algorytm zaokrąglania w górę lub w dół według schematu pokazanego wyżej. Zaletą tej metody jest łatwość jej realizacji pod względem obliczeniowym oraz to, że nie powoduje ona dużych strat informacji. Za mankament tej metody uważa się to, że w rezultacie zastosowania jednakowych i sztywnych reguł zaokrąglania, może w niektórych przypadkach umożliwić „odtworzenie” pierwotnych wielkości i zrekonstruować tablicę do jej pierwotnej postaci umożliwiającej ujawnienie danych podlegających ochronie [Cox86, s. 141]. Aby uniknąć tego mankamentu stosuje się inne warianty metody zaokrąglania, a mianowicie **zaokrąglanie losowe** oraz **zaokrąglanie sterowane**.

Zaokrąglanie losowe. Stosując metodę zaokrąglania tradycyjnego zawsze ostatnie cyfry liczebności zaokrąglaliśmy w dół do 0 — w przypadku, gdy liczby kończą się na 1 lub 2 oraz zawsze w górę (do 5), gdy równają się one 3 lub 4. Stosując metodę **zaokrąglania losowego** (*random rounding*) liczebności zaokrąglane są w dół z prawdopodobieństwem równym p oraz w górę z prawdopodobieństwem równym $1-p$, zapewniającym, że zaokrąglanie będzie nieobciążone.

Korzystając z przykładu zamieszczonego w [Sch93, s. 30]⁵⁶⁾ procedura zaokrąglania losowego realizowana jest według następującego schematu:

Wyszczególnienie	Gdy ostatnia cyfra liczebności wynosi					
	0	1	2	3	4	5
Zaokrągla się na 0 z prawdopodobieństwem $p =$. . .	1	4/5	3/5	2/5	1/5	0
Zaokrągla się na 5 z prawdopodobieństwem $1-p =$. . .	0	1/5	2/5	3/5	4/5	1

Zastosowanie tej procedury daje lepsze zabezpieczenie danych tablicy przed ujawnieniem, w porównaniu z zaokrąglaniem tradycyjnym dokonywa-

⁵⁶⁾ Przykład ten oparty jest na schemacie zaproponowanym przez [Nar72].

nym według stałego algorytmu. Ujemną cechą zaokrąglania losowego jest to, że powoduje ono większe straty informacji, w porównaniu z zaokrąglaniem tradycyjnym, ponieważ liczebności w tablicy nie zawsze są zaokrąglane do najbliższej wielkości liczby przyjętej za liczbę bazową (w naszym przykładzie liczbę 5).

Zaokrąglanie sterowane. Metoda została opracowana przez I. Fellegi'ego [Fel75]. Umożliwia ona modyfikowanie pozycji tablicy według każdorazowo określanego przez użytkownika algorytmu w taki sposób, aby zminimalizować straty informacji oraz zapewnić zgodność sum części składowych tablicy z obliczonymi sumami w rubrykach i wierszach. Zainteresowanych Czytelników w szerszym zapoznaniu się z tymi metodami wykorzystującymi m. in. metody sieciowe oraz metody programowania liniowego odsyłam do specjalistycznej literatury, zwłaszcza [Cox83], [Cau85], [Fag88].

Niezależnie od przyjętego wariantu zaokrąglania, metoda ta krytykowana jest za to, że powoduje ona duże straty informacji. Utrudnia również dalsze przetwarzanie danych na podstawie „zaokrąglonych” tablic.

6.3.3. Metody wprowadzania zakłóceń losowych do informacji zagregowanych

O modyfikowaniu zbiorów metodą wprowadzania zakłóceń losowych mówiliśmy w p. 6.1.6 w odniesieniu do zapobiegania ujawnieniu zbiorów danych źródłowych (mikrodanych).

Znacznie szerzej metoda ta jest stosowana w przypadku udostępniania wyników badań w układach szczegółowych zagrażających ujawnieniu danych podlegających ochronie. W omawianym przypadku informacje zakłócające (szum informacyjny) wprowadza się do określonych wielkości w tablicach, pozostawiając w nienaruszonym stanie zbiór z danymi źródłowymi. Można sobie również wyobrazić opracowanie tablic na podstawie zbioru źródłowego, do którego wprowadzono uprzednio informacje zakłócające. W tym przypadku jednak przed udostępnieniem tablicy należy również sprawdzić, czy nie zawiera ona pozycji podlegających ochronie.

Informacje zakłócające mogą być wprowadzane w tablicach zarówno w pozycjach, w których występują liczebności badanych jednostek (jeśli naruszają one zasadę wielkości progowej), jak i w pozycjach wartościowych (np. wartość produkcji, wielkość kosztów), jeżeli zamieszczona w określonej pozycji wielkość narusza zasadę przeważającego udziału.

Wielkość zakłócająca e stanowi zmienną losową, rozkład której musi być dobrany w dostosowaniu do specyfiki danego opracowania. Zwykle opracowuje się stałe zbiory wielkości zakłócających e wraz ze związanym z nimi prawdopodobieństwem $p(e)$ w taki sposób, aby wielkość zakłócająca

miała oczekiwaną wartość $E(e) = \sum [p(e) \times e] = 0$ oraz znaną wariancję $Var(e) = b^2$ [Sch93, s. 38].

Wielkość wariancji należy wybierać w drodze kompromisu między maksymalną pewnością ochrony danej pozycji tablicy a wielkością zniekształcenia informacji w rezultacie wprowadzonego zakłócenia.

Czytelników zainteresowanych w uzyskaniu szerszej informacji na temat praktycznych aspektów zastosowania tej metody odsyłam do [Sch93] oraz [Cox86].

7. PRÓBY AUTOMATYZACJI PROCEDUR OCHRONY INFORMACJI PRZED UJAWNIENIEM

Z przedstawionego w rozdziałach 5 i 6 omówienia metod ochrony informacji w procesie jej udostępniania oraz metod zapobiegania identyfikacji i ujawnienia wynika, że zadania jakim w tej dziedzinie statystycy muszą sprostać nie są łatwe.

Każdorazowo przed udostępnieniem zarówno zbioru danych źródłowych (mikrodanych), jak i opracowania wynikowego (tablicy) statystycy muszą:

- a) najpierw dokładnie sprawdzić czy w zbiorze lub w tablicy nie występuje liczebność mniejsza od określonych wielkości progowych, które mogłyby umożliwić identyfikację pośrednią⁵⁷⁾ podmiotu lub, czy zawarte w zbiorze lub w tablicy informacje nie grożą naruszeniem zasady przeważającego udziału;
- b) podjąć realizację procedur zapobiegających identyfikacji i ujawnieniu danych podlegających ochronie.

Duża trudność, a zwłaszcza bardzo duża pracochłonność realizacji tych zadań oraz świadomość odpowiedzialności za ewentualne dopuszczenie do ujawnienia danych i naruszenie tajemnicy statystycznej może prowadzić do nadmiernego ograniczenia działalności informacyjnej służb statystycznych, zwłaszcza w przekrojach terytorialnych, w których na skutek dużego rozdrobienia informacji istnieje szczególnie duże niebezpieczeństwo ujawnienia.

Dlatego tak ważne jest poszukiwanie metod i sposobów takich rozwiązań, aby znaleźć racjonalne wyjście gwarantujące zarówno pełne przestrzeganie zasad tajemnicy statystycznej, jak i prowadzenie aktywnej działalności w dziedzinie udostępniania informacji.

Ponieważ zrealizowanie obu wymienionych wyżej w p. a) i b) zadań jest bardzo pracochłonne, zwłaszcza w przypadku dużych zbiorów i szczegółowego programu opracowań, to statystycy od dawna poszukują rozwiązania, aby oba te zadania można było wykonać przy użyciu komputerów. Zadanie

⁵⁷⁾ Nie wspominam tu o identyfikacji bezpośredniej, ponieważ identyfikatory bezpośrednie zawsze powinny być usunięte ze zbioru po tym, jak przestały być one niezbędne w procesie kontroli i opracowania zbioru.

najważniejsze — to opracowanie odpowiedniego oprogramowania, które zapewniłoby:

- 1) odczytanie dowolnego zbioru źródłowego lub zbioru tablic, niezależnie od przyjętej w konkretnym przypadku organizacji zbioru i systemu zapisu;
- 2) umożliwienie wprowadzenia do programu zmiennych parametrów kontroli zbioru w zależności od rodzaju zbioru i wymaganego stopnia ochrony;
- 3) ujawnienie wszystkich przypadków grożących niebezpieczeństwem ujawnienia danych;
- 4) zrealizowanie dowolnej procedury modyfikacji zbioru lub tablicy minimalizującej zagrożenie ujawnienia.

Najbardziej znaną próbą opracowania oprogramowania, które mogłoby zrealizować przynajmniej niektóre z wymienionych wymagań jest pakiet ARGUS opracowany przez specjalistów z CBS Holandii.

Opracowanie tego oprogramowania właśnie w Holandii nie jest przypadkowe. Urząd ten zasłużył się posiadaniem wysoko kwalifikowanego zespołu informatyków-programistów. Wyrazem tego jest m.in. opracowanie cenionego wśród statystyków wielu krajów i znanego również w Polsce pakietu BLAISE przeznaczonego do projektowania i opracowania wyników badań statystycznych. Urząd ten poświęca również dużo uwagi opracowaniu takich rozwiązań, które zapewniając pełną ochronę informacji, nie prowadząby jednocześnie do nadmiernego ograniczania działalności informacyjnej służb statystycznych.

Pakiet jest przeznaczony do automatycznej kontroli zbioru jednostkowego (zbioru mikrodanych) i wykrywania tzw. rzadkich kombinacji cech, tj. takich kombinacji cech, które występują w zbiorze mniejszą liczbę razy od uznanej za dopuszczalną. Jeśli określona kombinacja cech występuje w zbiorze liczbę razy mniejszą od liczby dopuszczalnej, tj. od określonej wielkości progowej n , to taka kombinacja cech uznawana jest za rzadką i niebezpieczną z punktu widzenia możliwości identyfikacji pośredniej i musi być poddana określonej „obróbce” zapobiegającej ujawnieniu.

W aktualnej wersji pakiet ARGUS zapewnia analizę kombinacji do 5 cech, chociaż praktycznie w Holandii przy analizie zbiorów analizuje się częstość występowania jedynie kombinacji 3 cech (przykładowo, miejsce zamieszkania x płeć x zawód). Wiąże się to z tym, że w Holandii analizie poddaje się jedynie zbiory, które nie zawierają tzw. cech wrażliwych, których nie udostępnia się w ogóle [Waa95].

Pierwsza wersja pakietu ARGUS została opracowana w 1992 r. W 1995 r. opracowano czwartą kolejną wersję oznaczoną ARGUS 1.1. Ma ona nadal status prototypu. Pakiet zapewnia kontrolę zbiorów źródłowych przed ujawnieniem na skutek występowania w zbiorze nietypowych i rzadkich kombinacji cech, a w przypadku ich wykrycia zapewnia automatyczną realizację procedury zapobiegającej ujawnieniu.

Zarówno wielkości progowe liczebności cech, jak i procedury zapobiegania ujawnieniu określają sami statystycy w zależności od rodzaju zbioru (tj. stopnia wrażliwości cech zawartych w zbiorze) oraz oceny rzeczywistego niebezpieczeństwa identyfikacji pośredniej (np. w zależności od rozeznania co do dostępności dodatkowych informacji ułatwiających identyfikację).

Obecna wersja ARGUS umożliwia automatyczne wykrywanie kombinacji cech, których liczebność w zbiorze jest mniejsza od wartości progowej. Pozwala również zastosować odpowiednią procedurę zapobiegania identyfikacji. Procedurami tymi mogą być ograniczenie szczegółowości poprzez grupowanie kodów lub/oraz usuwanie (patrz p. 6.1). Włączenie w pierwszej kolejności do pakietu tych dwóch procedur tłumaczy się tym, że CBS Holandii preferuje te dwie metody zapobiegania ujawnieniu jako metody powodujące stosunkowo najmniejsze zniekształcenie zbiorów.

Kontrola zbiorów danych jednostkowych za pomocą programu ARGUS przebiega w dwóch etapach. W pierwszym etapie zmienne, które powodują powstawanie rzadkich kombinacji cech (np. rzadkie zawody w korelacji z płcią lub miejscem zamieszkania) są przekodowywane (np. poprzez łączenie kilku zawodów), a następnie, w drugim etapie, pozostające w zbiorze (zakłada się, że już nieliczne) kombinacje rzadkich cech usuwa się ze zbioru (patrz p. 6.1.5)

Program ARGUS wymaga użycia mikrokomputera co najmniej z procesorem 386, pracującego pod DOS oraz co najmniej 4 MB RAM [Waa95].

W 1994 r. pakiet ARGUS uzyskał poparcie Eurostatu oraz kilku krajów UE, które wyraziły chęć współpracy z CBS Holandii w kierunku dalszego rozwoju tego pakietu. Wyraża się również nadzieję, że wspólne stosowanie i rozwój tego pakietu sprzyjać będzie ujednoczeniu metod kontroli danych przed ujawnieniem i metod ochrony danych. Współpraca większej liczby urzędów statystycznych nad dalszym rozwojem pakietu pozwoli także poprawić jego funkcjonalność, tj. uwzględnić w nim również inne metody zapobiegania ujawnieniu poza metodami rozpowszechnionymi w praktyce holenderskich służb statystycznych. Istotnym argumentem na korzyść tego podejścia jest również znaczne ułatwienie szkolenia statystyków w ramach programu szkolenia europejskich statystyków (TES) [Waa94]. Są to sprawy, o których powinni pamiętać również polscy statystycy.

8. PODSUMOWANIE I WNIOSKI

Przedstawiona uwadze Czytelników praca stanowi pierwszą w polskiej literaturze statystycznej próbę przedstawienia szeregu problemów z jakimi mają do czynienia statystycy w czasie realizacji swoich zadań informacyjnych wobec władz i społeczeństwa. Pragną oni, z jednej strony, zaspokoić jak najlepiej zapotrzebowanie na informacje, a jednocześnie są zdecydowani chronić w pełni powierzone im informacje jednostkowe, co w stosunku do niektórych użytkowników oraz w stosunku do niektórych informacji nie zawsze realizuje się w sposób bezkonfliktowy.

Zagadnienia ochrony danych, w tym zwłaszcza danych osobowych nie są w polskiej statystyce niczym nowym. Świadczą o tym zarówno rozwiązania prawne, poczynwszy od pierwszej ustawy o organizacji statystyki administracyjnej z 1919 r. po obecną ustawę o statystyce publicznej, jak i wieloletnia praktyka służb statystycznych od ich utworzenia, w 1918 r., po dzień dzisiejszy.

Obecnie występuje jednak cały szereg nowych elementów związanych zarówno z aktualnymi warunkami i potrzebami krajowymi, jak i z konsekwencjami zacieśniającej się współpracy naszego kraju ze społecznością międzynarodową.

Sytuacja ta zmusza nas do nowego spojrzenia na szereg aspektów naszej statystyki, w tym zwłaszcza na zagadnienia bezpieczeństwa i ochrony danych.

Nie pretendując do wyczerpania tego skomplikowanego tematu, pragnę na zakończenie poddać pod rozwałę kilka spraw, które, jak się wydaje, można wyprowadzić logicznie z prezentowanej Czytelnikom pracy, a które zasługują na szczególną uwagę.

Zamieszczone niżej uwagi i sugestie nie wyrażają oczywiście oficjalnego stanowiska Głównego Urzędu Statystycznego. Są jedynie wynikiem własnych obserwacji autora sformułowanych z myślą o poszukiwaniu bardziej skutecznych metod gospodarowania zasobami informacyjnymi w nowych warunkach, w jakich pracuje i będzie pracować statystyka, tj. w warunkach, gdy coraz to więcej badań statystycznych musi być prowadzonych nie na podstawie przymusu administracyjnego, ale w oparciu o zrozumienie i zaufanie respondentów, którym w zamian za opartą na dobrej woli i zaufaniu współpracę ze służbami statystycznymi, należą się pełne gwarancje, że

przekazywane przez nich dane będą w pełni chronione i wykorzystywane wyłącznie w celach, którym służą badania statystyczne.

1. Sprawy ochrony danych muszą być traktowane jako jeden z elementów składowych procesu badań statystycznych, obejmujących projektowanie, zbieranie, przesyłanie, przetwarzanie, przechowywanie i udostępnianie danych. Oznacza to, że dla każdego tematu badania, uwzględniając jego cechy szczególne, zagadnienia bezpieczeństwa i ochrony danych muszą znajdować wyraźne uregulowanie, w tym zwłaszcza zagadnienia obiegu dokumentów w trakcie rejestracji, przesyłania i kontroli. Tych zagadnień nie da się rozwiązać w sposób jednolity, niezależny od tematu badań, w każdym badaniu istnieją bowiem różne warunki i różne zagrożenia. Inaczej, przykładowo, trzeba zapewnić obieg dokumentów w badaniach, które wymagają specjalistycznych metod kontroli (np. kwestionariusze z ochrony środowiska, statystyczne karty zgonów, karty wypadków przy pracy), a inaczej w przypadku spisu ludności. Sprawy ochrony danych muszą także znaleźć odzwierciedlenie w kalkulacji kosztów badań.

2. Sprawy bezpieczeństwa danych i zapobiegania przypadkom ujawnienia danych podlegających ochronie muszą zyskać odpowiednio wysoką rangę w całokształcie działalności informacyjnej służb statystycznych. Niezbędne jest zwłaszcza opracowanie metod kontroli wszystkich tablic przed ich udostępnieniem lub opublikowaniem z punktu widzenia niebezpieczeństwa pośredniej identyfikacji danych. Dotyczy to zwłaszcza tablic opracowywanych w bardziej szczegółowych układach według rodzaju działalności oraz w układach terytorialnych. Niezbędne jest również podjęcie (nie prowadzonych dotąd) prób zapobiegania identyfikacji danych, stosując początkowo najprostsze procedury, np. poprzez ograniczenie szczegółowości metodą grupowania kodów lub przez usuwanie.

3. Opracowania wymagają metody komputerowej kontroli danych zagregowanych z punktu widzenia zagrożeń ujawnienia danych ze względu na niebezpieczeństwo naruszenia wielkości progowych oraz zasady przeważającego udziału. Cały wysiłek statystyków i informatyków był dotąd skierowany na kontrolę i korygowanie błędów w danych jednostkowych. W tej dziedzinie mamy w naszej statystyce, zasługujący na podkreślenie, dorobek. Nie wypracowano natomiast dotąd automatycznych metod kontroli danych zagregowanych występujących w tablicach statystycznych, a stanowi to niezbędny warunek uniknięcia ujawnienia danych na etapie udostępniania.

4. Nowych uregulowań i rozwiązań wymagają sprawy przechowywania dokumentów badań statystycznych (sprawozdań, ankiet, dokumentów jednostkowych, takich jak: zgłoszenie urodzenia, karta statystyczna do karty zgonów, karta statystyczna zawarcia małżeństwa, tablic kontrolnych, tablic roboczych itp.). Chodzi nie tylko o pełne ich zabezpieczenie przed nieupoważnionym dostępem, ale również o realizację zaleceń Rady Europy, które przewidują, że „dane osobowe zbierane do celów statystycznych winny

być pozbawione części identyfikacyjnej niezwłocznie po zakończeniu operacji zbierania, kontroli i łączenia..." (por. p. 7 bis 1 „Zaleceń” zamieszczonych w aneksie). Zadanie to można zrealizować w dwojaki sposób": a) przez takie zaprojektowanie dokumentów statystycznych, aby po zakończeniu operacji, o których mowa wyżej, można było oddzielić od nich część identyfikacyjną (np. przez oderwanie perforowanej części dokumentu) względnie b) przez zmianę zasad przechowywania dokumentów, przewidując ich niszczenie niezwłocznie po zakończeniu opracowania. Oba te rozwiązania mogłyby także uwolnić służby statystyczne od zdarzających się sporadycznie przypadków nacisków na udostępnienie dokumentów badań statystycznych do różnych celów administracyjnych. Odrębnym problemem powstającym przy tej okazji jest ewentualne zachowanie niektórych dokumentów statystycznych (np. kwestionariuszy spisów powszechnych) dla celów archiwalno-historycznych. Można by ewentualnie rozważyć zachowanie dla tych celów minimalnej reprezentacji dokumentów.

5. Opracowując kompleksowy system ochrony danych, o którym mowa w p. 1, należy zwrócić szczególną uwagę na zapewnienie bezpieczeństwa danych znajdujących się w komputerowych systemach przechowywania i przetwarzania, tj. na ochronę danych na nośnikach maszynowych oraz na ochronę informacji w sieciach komputerowych, w systemach wielodostępnych itp.

6. Ogólny wzrost zainteresowania problemami bezpieczeństwa i ochrony informacji jednostkowych, który da się zapewne jeszcze silniej odczuć po przyjęciu przez Parlament ustawy o ochronie danych, uzasadnia podjęcie badań naukowych, dotyczących metod wykrywania zagrożeń ujawnienia danych przez potencjalnych intruzów z uwzględnieniem zwłaszcza możliwości wykorzystywania przez nich różnych dostępnych danych zarówno z systemu statystyki publicznej, jak i z innych źródeł. W ramach tych prac niezbędne jest opracowanie najbardziej skutecznych metod zapobiegania ujawnienia danych, opartych na analizie treści poszczególnych zbiorów informacji i rzeczywistych zagrożeń. Prace te, jak się wydaje, muszą być podjęte przez pracowników naukowych przy ścisłej współpracy z pracownikami służb statystycznych. Ze względu na ogromne znaczenie tych prac dla przyszłości statystyki udałoby się prawdopodobnie uzyskać finansową pomoc na ich prowadzenie ze środków państwowych przeznaczonych na badania naukowe.

7. Sprawy bezpieczeństwa i ochrony danych zasługują na szersze ich prezentowanie opinii publicznej. Stanowią one jedną z podstawowych przesłanek zdobycia zaufania społecznego do statystyki i łagodzenia niechęci do udziału w badaniach statystycznych. Dobrą okazją do szerszego zaprezentowania opinii publicznej podstawowych problemów statystyki oraz jej roli we współczesnym zmieniającym się świecie mogłoby być opracowanie i przyjęcie deklaracji w sprawie etyki zawodu statystyka.

8. Na zakończenie sygnalizuję potrzebę pogłębienia prac nad uściśleniem interpretacji niektórych przepisów ustawy o statystyce publicznej. Niewykluczone, że prace te mogą doprowadzić do uznania konieczności przygotowania pewnych zmian w tej ustawie. Z taką propozycją można by ewentualnie wystąpić po przyjęciu przez Komitet Ministrów Rady Europy „Zalecenia w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych”, aby doprowadzić do zgodności brzmienie art. 42, ust. 2 ustawy z tym zaleceniem (por. p. 3.3). Przygotowując ewentualne propozycje zmian do ustawy należałoby przeanalizować dokładniej sformułowania dotyczące ochrony danych jednostkowych. W części dotyczącej ochrony danych osobowych przepisy są bardzo jasne i precyzyjne i nie wymagają zmian. Pewne trudności nasuwa możliwość różnej interpretacji pojęcia „danych indywidualnych charakteryzujących wyniki działalności podmiotów gospodarczych” (art. 38, ust. 2)⁵⁸). Przy tej okazji należałoby także rozważyć takie sformułowanie postanowień ustawy, aby było możliwe publikowanie danych indywidualnych o działalności **niektórych** podmiotów gospodarczych **za ich zgodą**. Mam tu na myśli **nieliczne** podmioty o zasięgu ogólnokrajowym lub podmioty bardzo nieliczne, jak np. PKP, Polska Poczta, pojedyncze przedsiębiorstwa o zasięgu krajowym, parki narodowe itp.

⁵⁸) Nie wyjaśniają tego również do końca „Ustalenia Komisji do Spraw Ochrony Tajemnicy Statystycznej”, o której mowa w p. 6.3. niniejszej pracy. Wątpliwości dotyczą zarówno interpretacji pojęcia „wyniki ekonomicznej działalności gospodarczej”, jak i samo pojęcie „podmiotu gospodarczego”.

LITERATURA

- [Anw93] M. N. Anwar: *The Small Aggregates Method*, Eurostat, internal paper, 1993.
- [App92] G. Appel, S. Kinzel, D. Nölte: "SAFE" — *A Generally Usable Program System for the Annonimisation of Individual Data in Official Statistics*, Statistisches Landesamt Berlin, 1992.
- [Aus94] E. J. Ausems: *About the Council's Recommendations, Second International Seminar on Statistical Confidentiality*, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Bac69] R. Bachi, R. Baron: *Confidentiality problems related to data-banks*, Bulletin of the ISI, Proceedings of the 37-th Session, Vol. XLIII, book 1, Invited papers, London, 1969.
- [Big94] Luigi Biggeri, Laura Grassini: *Analysis of the potential disclosure risk: some empirical evidencies from the Italian Household Multipurpose Survey Data*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Bor94] S. Borra, C. Ceccarelli, G. Coccia, F. Crescenzi: *Standard Microdata Files on 1991 Italian Population Census*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Bos76] Antoni Bossowski: *Problemy ochrony danych komputerowych*, Wiadomości Statystyczne, 1976, nr 12.
- [Buz94] Lucia Buzzigoli, Antonio Giusti: *Merging External Files for Statistical Disclosure Purposes*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Cau85] B. D. Causey, L. H. Cox, L. R. Ernst: *Application of Transportation Theory to Statistical Problems*, Journal of the American Statistical Association, 80, 1985.
- [Cox83] L. H. Cox, L. R. Ernst: *Controlled Rounding, Some Recent Advances in Theory, Computation of Application of Network Flow Models*, University of Toronto Press, 1983.
- [Cox86] L. Cox, S. -K. Mc Donald, D. Nelson: *Confidentiality Issues at the United States Bureau of the Census*, Journal of Official Statistics, Statistics Sweden, Vol. 2, No 2, 1986.
- [Dal77] T. Dalenius: *Towards a methodology for statistical disclosure control*, Statistik Tidskrift, 1977, No 5, s. 429—444.
- [Dal82] T. Dalenius, S. P. Reiss, *Data Swapping: A Technique for Disclosure*

Control, Journal of Statistical Planning and Inference, North-Holland Publishing Comp., 1982, No 6.

- [Dal88] Tore Dalenius: *Controlling Invasion of Privacy in Surveys*, Department of Development and Research, Statistical Research Unit, Statistics Sweden, 1988.
- [Def92] D. Defays, Ph. Nanopoulos: *Panels of Enterprises and Confidentiality — The Small Aggregates Method*, Eurostat, internal paper, 1992.
- [Dek95] *Deklaracja w sprawie etyki zawodowej statystyków*, Wiadomości Statystyczne, 1995, nr 8.
- [Dun95] Joseph Duncan, Andrew C. Gross: *Statistics for the 21 st century. Proposals for improving statistics for better decision making*, The Dun and Bradstreet Corp. 1995.
- [Eng94] Christian Engelage: *The Draft Data Protection Directive — No reason to panic*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Fag88] J. Fagan, G. Greenberg, B. Hemming: *Controlled Rounding of Three Dimensional Tables*, Statistical Research Division Report Series, Census/SDR/RR-88/02, Statistical Division, Bureau of the Census, Washington D.C., 1988.
- [Fea94] John Feather: *The Information Society. A Study of Continuity and Change*, Library Association Publishing Ltd, London, 1994.
- [Fel72] I. P. Fellegi: *On Question of Statistical Confidentiality*, Journal of the American Statistical Association, March 1972, Vol. 67, No 337.
- [Fel75] I. Fellegi: *Controlled Random Rounding*, Survey Methodology, Statistics Canada, Ottawa, Ontario, Vol. 1, 1975.
- [Fie94] Stephen E. Fienberg: *Conflicts between the Needs for Access to Statistical Information and Demands for Confidentiality*, Journal of Official Statistics, Statistics Sweden, Vol. 10, No 2, 1994.
- [Fla89] Wiesław Flakiewicz, Józef Oleński: *Cybernetyka Ekonomiczna*, PWE, Warszawa, 1989.
- [Gre90] Brian Greenberg: *Disclosure Avoidance Research at the Census Bureau*, Report presented at the 1990 Annual Research Conference of the US Bureau of the Census, Arlington, Virginia, March 18—21, 1990.
- [Hof77] Lance J. Hoffman: *Modern methods for computer security and privacy*, Prentice Hall, Inc., 1977.
- [Hof82] L. J. Hoffman: *Poufność w systemach informatycznych*, WNT, Warszawa 1982.
- [Idz79] Andrzej Z. Idźkiewicz: *Ochrona informacji w procesie przetwarzania*, PWE, Warszawa 1979.
- [Laa94] Seppo Laaksonen: *Handling longitudinal micro data files of enterprise surveys*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Mak94] Stephen Fienberg, Udi E. Makov, Ashish P. Sanil: *A Bayesian approach to data disclosure: optimal intruder behaviour for continuous data*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.
- [Mar70] James Martin, Adrian R. D. Norman: *The computerized society*, Prentice-Hall, Inc., Englewood Cliffs, New York, 1970.

- [Med95] Arwid Mednis: *Prawna ochrona danych osobowych*, Wyd. Naukowe Scholar, Warszawa, 1995.
- [Mül94] Walter Müller, Heike Wirth: *Research needs for European Microdata and Data Confidentiality*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.
- [Mül95] Walter Müller, Uwe Blien, Heike Wirth: *Identification Risks of Microdata: Evidence from Experimental Studies*, Sociological Methods and Research, 1995.
- [Nan94] Photis Nanopoulos, Leonidas Kioussis: *Orientation of Future Work in Eurostat on Statistical Confidentiality Issues*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.
- [Nar72] M. S. Nargundkar, W. Saveland: *Random Rounding: A Means of Preventing Disclosure of Information about Individual Respondents in Aggregate Data*, Statistics Canada, 1972.
- [Nob94] Joris Nobel: *Data confidentiality and data access — practical and legal issues in the Netherlands*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.
- [Nor68] Svein Nordbotten: *Confidential treatment of data, and the use and classification of data*, Statistik Tidskrift, 6, No 5, 381–389, Stockholm, 1968.
- [Pod95] *Podstawowe zasady statystyk oficjalnych oraz ich prawnych i etycznych aspektów w krajach w okresie przejściowym*, Biblioteka Wiadomości Statystycznych, tom 47, Warszawa, 1995.
- [Pup95] Melania Pupkiewicz: *Nowe rozwiązania prawne ustawy o statystyce publicznej*, Wiadomości Statystyczne, 1995, nr 11.
- [Rei84] S. P. Reiss: *Practical Data-Swapping: First Steps*, *ACM Transactions on Database Systems*, Vol. 9, No 1, March 1984.
- [Sch93] Daniela Schackis: *Manual on Disclosure Control Methods*, Eurostat, 1993 (maszynopis).
- [Sim94] Maria Rosaria Simeone: *Rules Governing Statistical Confidentiality in Europe. Impact on Italian Statistical Regulations*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.
- [Sok87] Andrzej Sokolowski: *Ochrona informacji komputerowych*, Wyd. MON, Warszawa 1987.
- [Ste83] J. Steinberg: *Social Research Use of Archival Records: Procedural Solutions to Privacy Problems*. Chapter 13 in *Solutions to Ethical and Legal Problems in Social Research*, R. F. Boruch and J. S. Cecil, eds., New York: Academic Press, 249–261.
- [Vri94] R. E. de Vries, A. G. de Waal, L.C.R.J. Willenborg: *Distinguishing Rare from Common Characteristics in Microdata Sets*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.
- [Waa94] A. G. de Waal, L.C.R.J. Willenborg: *Development of ARGUS: Past, Present, Future*, Second International Seminar on Statistical Confidentiality, Luxembourg 28–29 November 1994, Pre-proceedings, Eurostat 1994.

- [Waa95] A. G. de Waal, A. J. Pieters: *ARGUS Users Guide*, Statistics Netherlands, Division Research and Development Department of Statistical Methods, August 1995.
- [Wal94] Tadeusz Walczak: *Informatyka w tematyce obrad 49 Sesji Międzynarodowego Instytutu Statystycznego*, Wiadomości Statystyczne 1994, nr 3.
- [Wal95] Tadeusz Walczak: *O etyce zawodu statystyka*, Wiadomości Statystyczne, 1995, nr 8.
- [Yeo94] Douglas Yeo, Dale Robertson: *Disclosure Control Issues at Statistics Canada*, Second International Seminar on Statistical Confidentiality, Luxembourg 28—29 November 1994, Pre-proceedings, Eurostat 1994.

ANEKS

KONWENCJA

z 28 stycznia 1981 r. dotycząca ochrony osób w związku z automatycznym przetwarzaniem danych osobowych, podpisana w Strasburgu

(Ogłoszona D.n. 85—1203, 15 list. 1985, — JO 20 list. 1985. Weszła w życie 1.10.1985)¹⁾

Preambuła

Państwa członkowskie Rady Europy, sygnatariusze niniejszej Konwencji. Zważywszy, że celem Rady Europy jest osiągnięcie większej jedności jej członków w poszanowaniu zwłaszcza prymatu prawa oraz praw człowieka i podstawowych wolności;

Zważywszy, że pożądane jest poszerzenie ochrony praw i podstawowych wolności każdego człowieka, w szczególności prawa do poszanowania prywatności, biorąc pod uwagę zwiększający się przepływ przez granice danych osobowych przetwarzanych automatycznie;

Potwierdzając jednocześnie ich zobowiązanie na rzecz swobodnego przepływu informacji bez względu na istniejące granice;

Uznając konieczność pogodzenia podstawowych wartości poszanowania prywatności ze swobodnym obiegiem informacji między narodami;

Uzgodniły, co następuje:

Rozdział I

Postanowienia ogólne

Artykuł 1

Przedmiot i cel

Niniejsza Konwencja ma na celu zagwarantowanie, na terytorium każdej ze stron, każdej osobie fizycznej, niezależnie od jej narodowości i miejsca

¹⁾ Przedruk tekstu zamieszczonego w pracy [Med95].

zamieszkania, poszanowanie jej praw i podstawowych wolności, w szczególności jej prawa do prywatności, w związku z automatycznym przetwarzaniem dotyczących jej danych osobowych („ochrona danych”).

Artykuł 2

Definicje

Przy stosowaniu niniejszej Konwencji:

- a) wyrażenie „dane osobowe” oznacza każdą informację dotyczącą osoby fizycznej o ustalonej tożsamości albo dającej się zidentyfikować („osoba zainteresowana”);
- b) wyrażenie „zbiór zautomatyzowany” oznacza ogół informacji automatycznie przetwarzanych;
- c) wyrażenie „automatyczne przetwarzanie” oznacza następujące operacje wykonywane w całości lub części przy pomocy metod zautomatyzowanych: rejestracja danych, stosowanie do tych danych operacji logicznych i/albo arytmetycznych, ich modyfikowanie, usuwanie, wybieranie lub rozpowszechnianie;
- d) wyrażenie „zarządzający kartoteką” oznacza osobę fizyczną lub prawną, władzę publiczną, służbę lub każdy organ właściwy, w myśl prawa miejscowego, dla podejmowania decyzji dotyczących celu jakiego ma służyć zautomatyzowana kartoteka kategorii danych osobowych jakie mają być rejestrowane oraz operacji jakim będą one poddawane.

Artykuł 3

Zakres stosowania

1. Strony zobowiązują się do stosowania niniejszej Konwencji do kartotek i do automatycznego przetwarzania danych osobowych w sektorze publicznym i prywatnym.

2. Każde państwo może, przy podpisaniu niniejszej Konwencji lub przy składaniu dokumentów ratyfikacji, przyjęciu, potwierdzeniu albo przystąpieniu, lub w każdym późniejszym terminie, powiadomić w oświadczeniu złożonym Sekretarzowi Generalnemu Rady Europy:

- a) że nie będzie stosować niniejszej Konwencji do niektórych kategorii zautomatyzowanych kartotek danych osobowych, których lista zostanie złożona. Jednakże na liście tej nie może umieścić zautomatyzowanych kartotek danych osobowych, które w myśl jego prawa wewnętrznego podlegają przepisom o ochronie danych. Będzie więc musiało zmodyfikować tę listę poprzez złożenie nowego oświadczenia, jeśli dodatkowe

kategorii zautomatyzowanych kartotek danych osobowych zostaną objęte przepisami o ochronie danych;

- b) że będzie stosować niniejszą Konwencję również do innych informacji dotyczących ugrupowań, stowarzyszeń, fundacji, spółek, korporacji i wszelkich innych organizacji gromadzących bezpośrednio lub pośrednio osoby fizyczne i posiadających lub nie osobowość prawną;
- c) że będzie stosować niniejszą Konwencję również do kartotek danych osobowych nie objętych automatycznym przetwarzaniem.

3. Państwo, które poszerzyło zakres stosowania niniejszej Konwencji w myśl oświadczenia wymienionego w ustępach 2 b) lub c) może zaznaczyć w tym oświadczeniu, że wymienione poszerzenie stosuje się jedynie do niektórych kategorii danych kartotek osobowych, których wykaz zostanie złożony.

4. Strona, która wyłączyła niektóre kategorie kartotek danych zautomatyzowanych w myśl oświadczenia, o którym mowa w ustępie 2a, nie może żądać stosowania niniejszej Konwencji do takich kategorii przez Stronę, która ich nie wyłączyła.

5. Podobnie Strona, która nie poszerzyła stosowania Konwencji tak, jak to przewidują paragrafy 2b) i c), nie może powoływać się na jej zastosowanie w tym zakresie wobec Strony, która takiego poszerzenia dokonała.

6. Oświadczenia, o których mowa w paragrafie 2 wywołują skutek w chwili wejścia w życie Konwencji w odniesieniu do państwa, które je złożyło przy podpisaniu lub przy składaniu dokumentów ratyfikacji, przyjęciu, potwierdzeniu albo przystąpieniu, albo w trzy miesiące po otrzymaniu oświadczeń Sekretarza Generalnego Rady Europy, jeśli zostały wycofane w całości lub w jego części drogą notyfikacji skierowanej do Sekretarza Generalnego. Wycofanie oświadczenia wywołuje skutek w trzy miesiące od chwili otrzymania notyfikacji.

Rozdział II

Podstawowe zasady ochrony danych

Artykuł 4

Zobowiązania Stron

1. Każda Strona zastosuje niezbędne środki dla wprowadzenia w swoim prawie wewnętrznym podstawowych zasad ochrony danych, wymienionych w niniejszym rozdziale.

2. Środki te powinny być zastosowane najpóźniej w chwili wejścia w życie niniejszej Konwencji w odniesieniu do danej Strony.

Artykuł 5

Jakość danych

Dane osobowe poddawane automatycznemu przetwarzaniu powinny być:

- a) uzyskane i przetwarzane rzetelnie i zgodnie z prawem;
- b) rejestrowane dla określonych i zasadnych celów oraz nie powinny być wykorzystywane w sposób niezgodny z tymi celami;
- c) stosowne, rzeczowe oraz nie wykraczające poza potrzeby wynikające z celów, dla których zostały zarejestrowane;
- d) dokładne i, w razie potrzeby, aktualizowane;
- e) przechowywane w sposób pozwalający na identyfikację osób zainteresowanych przez okres nie dłuższy niż ten, który jest niezbędny dla realizacji celów, dla których zostały zarejestrowane.

Artykuł 6

Szczególne kategorie danych

Danych osobowych ujawniających przynależność rasową, poglądy polityczne, przekonania religijne lub inne oraz danych osobowych dotyczących zdrowia lub życia seksualnego nie można przetwarzać automatycznie, chyba że prawo wewnętrzne przewiduje odpowiednie gwarancje. Przepis niniejszy stosuje się również do danych osobowych dotyczących wyroków karnych skazujących.

Artykuł 7

Bezpieczeństwo danych

Należy podjąć odpowiednie środki bezpieczeństwa w celu ochrony danych osobowych zarejestrowanych w zautomatyzowanych kartotekach chroniące przed zniszczeniem przypadkowym lub dokonany bez zezwolenia oraz przed przypadkową utratą, oraz przed udostępnieniem, zmianą lub rozpowszechnieniem bez zezwolenia.

Artykuł 8

Gwarancje dodatkowe dla osoby zainteresowanej

Każda osoba powinna mieć zapewnioną możliwość:

- a) posiadania wiadomości o istnieniu zautomatyzowanej kartoteki danych osobowych, o jej zasadniczych celach oraz o tożsamości i zwykłym miejscu zamieszkania lub siedzibie zarządzającego kartoteką;

- b) otrzymywania w rozsądnych odstępach czasu, bez nadmiernych kosztów i w niezbyt długim terminie, potwierdzenia, że w kartotece znajdują się lub nie znajdują jej dane osobowe oraz uzyskiwania tych danych w czytelnej formie;
- c) uzyskania, w razie potrzeby, sprostowania lub usunięcia tych danych, jeśli zostały przetworzone z naruszeniem przepisów prawa wewnętrznego wprowadzających w życie wymienione w art. 5 i 6 niniejszej Konwencji;
- d) złożenia zażalenia w razie nieuwzględnienia jej prośby o potwierdzenie albo udostępnienie, sprostowanie lub usunięcie danych w myśl pkt. b) i c) niniejszego artykułu.

Artykuł 9

Wyjątki i ograniczenia

1. Wyjątki od przepisów art. 5, 6 i 8 niniejszej Konwencji dopuszczalne są jedynie w granicach określonych w niniejszym artykule.

2. Odstąpienie od stosowania art. 5, 6 i 8 niniejszej Konwencji jest możliwe, jeśli przewidywane jest ustawą umawiającej się Strony jako środek konieczny w społeczeństwie demokratycznym:

- a) dla ochrony państwa, bezpieczeństwa publicznego, interesów walutowych państwa lub zwalczania przestępczości;
- b) dla ochrony osoby zainteresowanej oraz praw i wolności innych osób.

3. Ustawa może przewidywać ograniczenia realizacji praw, o których mowa w art. 8, paragraf b, c i d, gdy chodzi o zautomatyzowane kartoteki danych osobowych wykorzystywane dla celów statystycznych lub naukowych, jeśli zachodzi oczywisty brak ryzyka naruszania prywatności zainteresowanych osób.

Artykuł 10

Sankcje i zażalenia

Każda ze stron zobowiązuje się do wprowadzenia odpowiednich sankcji i zażeń dotyczących naruszeń przepisów prawa wewnętrznego wprowadzających w życie podstawowe zasady ochrony danych, wymienione w niniejszym artykule.

Artykuł 11

Ochrona poszerzona

Żaden z przepisów niniejszego rozdziału nie może być interpretowany jako ograniczający lub naruszający prawo każdej ze Stron do przyznania

osobom zainteresowanym ochrony szerszej niż przewidziana w niniejszej Konwencji.

Rozdział III

Przepływ danych ponad granicami

Artykuł 12

Przepływ danych osobowych ponad granicami a prawo wewnętrzne

1. Poniższe przepisy dotyczą przekazu poprzez granice państwowe, niezależnie od nośnika, danych osobowych automatycznie przetwarzanych lub zbieranych w celu takiego przetwarzania.

2. Umawiająca się Strona nie może zabronić ani uzależnić od specjalnego zezwolenia, wyłącznie w celu ochrony prywatności, przepływu danych osobowych poprzez granice terytorium innej Strony.

3. Jednakże każda ze Stron może odstąpić od stosowania przepisów pkt. 2.;

- a) o tyle, o ile jej ustawodawstwo reguluje odrębnie sprawy danych osobowych lub kartotek zautomatyzowanych takich danych ze względu na rodzaj tych danych lub kartotek, chyba że regulacja drugiej Strony zapewnia ochronę równoważną;
- b) jeśli przekazu dokonuje się z terytorium Strony na terytorium Państwa nie będącego Stroną, poprzez terytorium innej Strony, w celu niedopuszczenia do omińnięcia ustawodawstwa Strony wysyłającej.

Rozdział IV

Pomoc wzajemna

Artykuł 13

Współpraca między Stronami

1. Strony zobowiązują się do udzielania sobie wzajemnej pomocy przy wprowadzaniu w życie niniejszej Konwencji.

2. W tym celu:

- a) każda ze Stron wyznacza jeden lub więcej organów, których nazwę i adres podaje Sekretarzowi Generalnemu Rady Europy;
- b) każda ze Stron, która wyznaczyła kilka organów wskazuje w doniesieniu, o którym mowa w poprzednim ustępie, kompetencje każdego z tych organów.

3. Na życzenie organu wyznaczonego przez jedną ze Stron, organ wyznaczony przez inną Stronę;

- a) udzieli informacji dotyczących prawa i praktyki administracyjnej w zakresie ochrony danych;
- b) podejmie, zgodnie ze swym prawem wewnętrznym i wyłącznie w celu ochrony życia prywatnego, wszelkie odpowiednie środki w celu udzielenia informacji dotyczących określonego przetwarzania danych dokonanego na jego terytorium, z wyjątkiem danych osobowych ulegających temu przetworzeniu.

Artykuł 14

Udzielenie pomocy osobom zainteresowanym zamieszkałym za granicą

1. Każda ze Stron udziela pomocy każdej osobie mieszkającej za granicą w realizacji praw przewidzianych przez jej prawo wewnętrzne wprowadzające w życie zasady wymienione w art. 8 niniejszej Konwencji.

2. Jeśli taka osoba mieszka na terytorium innej Strony, powinna ona mieć prawo do zgłoszenia swego wniosku za pośrednictwem organu wyznaczonego przez tę Stronę.

3. Wniosek o udzielenie pomocy powinien zawierać wszelkie niezbędne dane, a w szczególności:

- a) nazwisko, adres i wszelkie inne właściwe dane dotyczące tożsamości wnioskodawcy;
- b) wskazanie kartoteki danych osobowych, której wniosek dotyczy, albo zarządzającego tą kartoteką;
- c) cel wniosku.

Artykuł 15

Gwarancje dotyczące pomocy udzielanej przez wyznaczone organy

1. Organ wyznaczony przez jedną ze Stron, który otrzymał informacje od organu wyznaczonego przez inną Stronę bądź jako poparcie wniosku o udzielenie pomocy, bądź jako odpowiedź na swój wniosek, może wykorzystać tę informację wyłącznie w celu wskazanym we wniosku.

2. Każda ze Stron będzie czuwać nad tym, aby osoby należące do wyznaczonego organu albo występujące w jego imieniu były związane obowiązkiem zachowania tajemnicy lub poufności tych informacji.

3. W każdym przypadku wyznaczony organ nie będzie upoważniony do składania wniosku o udzielenie pomocy, w myśl art. 14, pkt 2, w imieniu osoby zainteresowanej mieszkającej za granicą, z własnej inicjatywy i bez wyraźnego zezwolenia tej osoby.

Artykuł 16

Odmowa udzielenia pomocy

Wyznaczony organ, który otrzymał wniosek o udzielenie pomocy w myśl art. 13 lub 14 niniejszej Konwencji może odmówić uwzględnienia wniosku jedynie w następujących przypadkach:

- a) jeśli wniosek nie mieści się w zakresie kompetencji, dotyczących ochrony danych, organów uprawnionych do udzielenia odpowiedzi;
- b) jeśli wniosek nie jest zgodny z przepisami niniejszej Konwencji;
- c) jeśli uwzględnienie wniosku byłoby sprzeczne z suwerennością, bezpieczeństwem lub porządkiem publicznym Strony, która wyznaczyła organ lub z prawami i podstawowymi wolnościami osób podlegających jurysdykcji tej Strony.

Artykuł 17

Koszty i postępowanie

1. Pomoc jaką Strony sobie przyznają wzajemnie w myśl art. 13 oraz pomoc jakiej udzielają zainteresowanym osobom mieszkającym za granicą w myśl art. 14, nie pociąga za sobą innych kosztów i opłat niż usługi rzeczoznawców i tłumaczy. Te koszty i opłaty obciążają Stronę, która wyznaczyła organ, który zgłosił wniosek o udzielenie pomocy.

2. Osoba zainteresowana nie może być zobowiązana do pokrywania, w związku z zabiegami podjętymi w jej imieniu na terytorium innej Strony, innych kosztów i opłat niż te, jakie są wymagane od osób zamieszkałych na terytorium tej Strony.

3. Inne warunki dotyczące udzielenia pomocy, w szczególności zaś jej form i procedury oraz używanych języków zostaną ustalone bezpośrednio przez zainteresowane Strony.

Rozdział V

Komitet Doradczy

Artykuł 18

Skład Komitetu

1. Po wejściu w życie niniejszej Konwencji zostaje powołany Komitet Doradczy.

2. Każda ze Stron wyznacza w tym Komitecie swego przedstawiciela i jego zastępcę. Państwo będące członkiem Rady Europy, która nie jest Stroną Konwencji może być reprezentowane w Komitecie przez obserwatora.

3. Komitet Doradczy może, mocą jednomyślnie podjętej uchwały, zaprosić każde państwo, które nie jest członkiem Rady Europy ani Stroną Konwencji, aby na jednym z posiedzeń Komitetu było reprezentowane przez obserwatora.

Artykuł 19

Funkcje Komitetu

Komitet Doradczy:

- a) może zgłaszać propozycje mające na celu ułatwienie lub usprawnienie stosowania Konwencji;
- b) może składać wnioski o nowelizację niniejszej Konwencji, zgodnie z art. 21;
- c) wydaje opinię o każdym wniosku o nowelizację niniejszej Konwencji, który jest mu przedkładany stosownie do art. 21, pkt. 3;
- d) może, na wniosek jednej ze Stron, wydawać opinię w każdej sprawie dotyczącej stosowania niniejszej Konwencji.

Artykuł 20

Procedura

1. Komitet Doradczy jest zwoływany przez Sekretarza Generalnego Rady Europy. Komitet odbywa swe pierwsze posiedzenie w ciągu dwunastu miesięcy od wejścia w życie niniejszej Konwencji. Następnie zwoływany jest co najmniej raz na dwa lata, a w każdym razie na wniosek jednej trzeciej przedstawicieli Stron.

2. Większość przedstawicieli Stron stanowi *quorum* niezbędne dla odbycia posiedzenia Komitetu.

3. Po każdym posiedzeniu, Komitet doradczy przedstawia ministrom Rady Europy sprawozdanie ze swych prac i z funkcjonowania Konwencji.

4. Z zastrzeżeniem stosowania przepisów niniejszej Konwencji, Komitet Doradczy uchwała swój regulamin wewnętrzny.

Rozdział VI

Nowelizacje

Artykuł 21

Nowelizacje

1. Wnioski o nowelizację niniejszej Konwencji mogą wnosić: jedna ze Stron, Komitet Ministrów Rady Europy lub Komitet Doradczy.

2. Sekretarz Generalny Rady Europy przekazuje wniosek o nowelizację Konwencji państwom członkowskim Rady Europy oraz każdemu państwu nie będącemu jej członkiem, które przystąpiło lub zostało zaproszone do przystąpienia do niniejszej Konwencji stosownie do przepisów art. 23.

3. Poza tym wniosek o nowelizację zgłoszony przez Stronę lub Komitet Ministrów jest przekazywany do Komitetu Doradczego, który przedstawia Komitetowi Ministrów swą opinię na temat tego wniosku.

4. Komitet Ministrów rozpatruje wniosek oraz opinię Komitetu Doradczego i może zatwierdzić proponowaną nowelizację.

5. Tekst zatwierdzony przez Komitet Ministrów stosownie do pkt. 4 niniejszego artykułu jest przekazywany Stronom do akceptacji.

6. Nowelizacja zatwierdzona stosownie do pkt. 4 niniejszego artykułu wchodzi w życie trzydziestego dnia od chwili, gdy wszystkie Strony powiadomią Sekretarza Generalnego o ich akceptacji.

Rozdział VII

Postanowienia końcowe

Artykuł 22

Wejście w życie

1. Niniejsza Konwencja jest otwarta do podpisu dla wszystkich państw członków Rady Europy. Podlega ona ratyfikacji, przyjęciu lub potwierdzeniu. Dokumenty ratyfikacji, przyjęcia lub potwierdzenia składa się Sekretarzowi Generalnemu Rady Europy.

2. Niniejsza Konwencja wejdzie w życie pierwszego dnia miesiąca następującego po upływie trzech miesięcy od chwili, w której pięć państw członków Rady Europy wyrazi zgodę na związanie niniejszą Konwencją stosownie do postanowień poprzednich ustępów.

3. W odniesieniu do każdego państwa, które wyrazi zgodę na związanie Konwencją w dalszej kolejności, Konwencja wchodzi w życie pierwszego dnia miesiąca następującego po upływie trzech miesięcy od dnia złożenia dokumentów ratyfikacji, przyjęcia lub potwierdzenia.

Artykuł 23

Przystąpienie państw nieczłonkowskich

1. Po wejściu w życie niniejszej Konwencji Komitet Ministrów Rady Europy może zaprosić każde państwo nie będące członkiem Rady Europy do przystąpienia do niniejszej Konwencji drogą decyzji podjętej większością

głosów, przewidzianą w art. 20 do Statutu Rady Europy oraz jednomyślnie przez umawiające się państwa mające prawo zasiadania w Komitecie.

2. W odniesieniu do każdego państwa przystępującego, Konwencja wejdzie w życie pierwszego dnia miesiąca następującego po upływie trzech miesięcy od chwili złożenia dokumentów ratyfikacyjnych u Sekretarza Generalnego Rady Europy.

Artykuł 24

Klauzula terytorialna

1. Każde państwo może, przy podpisaniu niniejszej Konwencji lub przy składaniu dokumentów ratyfikacji, przyjęcia, potwierdzenia lub przystąpienia, wskazać terytorium lub terytoria, na których niniejsza Konwencja będzie stosowana.

2. Każde państwo może w każdym późniejszym czasie, w oświadczeniu skierowanym do Sekretarza Generalnego Rady Europy, rozszerzyć stosowanie Konwencji na inne terytorium wskazane w oświadczeniu. W odniesieniu do tego terytorium, Konwencja wejdzie w życie pierwszego dnia miesiąca następującego po upływie trzech miesięcy od chwili otrzymania oświadczenia przez Sekretarza Generalnego.

3. Każde oświadczenie dokonane na podstawie dwóch poprzednich punktów może być wycofane, odnoście terytorium wskazanego w oświadczeniu, drogą notyfikacji skierowanej do Sekretarza Generalnego. Wycofanie staje się skuteczne pierwszego dnia miesiąca następującego po upływie sześciu miesięcy od dnia otrzymania notyfikacji przez Sekretarza Generalnego.

Artykuł 25

Zastrzeżenia do przepisów niniejszej Konwencji nie są dopuszczalne.

Artykuł 26

Wypowiedzenie

1. Każda ze Stron może wypowiedzieć w każdym czasie niniejszą Konwencję, w notyfikacji skierowanej do Sekretarza Generalnego Rady Europy.

2. Wypowiedzenie staje się skuteczne pierwszego dnia miesiąca następującego po upływie sześciu miesięcy od chwili otrzymania notyfikacji przez Sekretarza Generalnego.

Artykuł 27

Notyfikacje

Sekretarz Generalny Rady Europy notyfikuje państwom członkowskim Rady oraz każdemu państwu, które przystąpiło do niniejszej Konwencji:

- a) jej podpisanie;
- b) złożenie dokumentów ratyfikacji, przyjęcia, potwierdzenia lub przystąpienia;
- c) datę wejścia w życie niniejszej Konwencji stosownie do art. 22, 23 i 24;
- d) każdy inny akt, notyfikację lub wiadomość dotyczącą niniejszej Konwencji.

W dowód czego, niżej podpisani, należycie umocowani, podpisali niniejszą Konwencję.

Sporządzono w Strasburgu, dnia 28 stycznia 1981, w językach angielskim i francuskim, przy czym oba teksty są jednakowo autentyczne, w jednym egzemplarzu, który zostanie złożony w archiwum Rady Europy. Sekretarz Generalny przekaze uwierzytelnione podpisy każdemu z państw członków Rady Europy oraz każdemu państwu zaproszonemu do przystąpienia do niniejszej Konwencji.

ZALECENIE

w sprawie ochrony danych osobowych zbieranych i przetwarzanych do celów statystycznych²⁾

Preambuła

Komitet Ministrów, w ramach przepisów artykułu 15 b Statutu Rady Europy.

1. Zważywszy, że celem Rady Europy jest osiągnięcie większej jedności jej członków;
2. Mając świadomość potrzeb w dziedzinie rzetelnej statystyki zarówno sektora publicznego, jak i prywatnego do analiz i lepszego zrozumienia współczesnego społeczeństwa oraz do określenia polityki i strategii w zarządzaniu praktycznie wszystkimi aspektami codziennego życia;
3. Uznając, że opracowanie rzetelnej statystyki zależy w znacznym stopniu od zbierania możliwie najbardziej szczegółowych informacji z zastosowaniem zwiększających swą efektywność technologii informatycznych;
4. Mając świadomość tego faktu, że informacja może dotyczyć zidentyfikowanych lub dających się zidentyfikować osób („dane osobowe”);

²⁾ Projekt z czerwca 1996 r., tłumaczenie z angielskiego T. W.

5. Mając świadomość potrzeb rozwoju metod umożliwiających gwarancje anonimowości podmiotów, których dane dotyczą;
6. Uwzględniając troskę międzynarodowej społeczności statystyków o ochronę danych osobowych oraz opracowanie międzynarodowych zaleceń uwzględniających zasady etyki zawodowej statystyków;
7. Uwzględniając również podstawowe zasady statystyki oficjalnej przyjęte przez społeczność międzynarodową w ramach Narodów Zjednoczonych;
8. Biorąc pod uwagę współczesny rozwój narodowych i ponadnarodowych standardów prawnych zarówno w dziedzinie działalności statystycznej, jak i w dziedzinie ochrony danych;
9. Nawiazując do ogólnych zasad w sprawie ochrony danych przyjętych w Konwencji dotyczącej ochrony osób, w związku z automatycznym przetwarzaniem danych (Strasburg 1981, ETS 108);
10. Nawiazując również do odstępstw dopuszczonych na korzyść działalności statystycznej w odniesieniu do korzystania z niektórych praw przez podmioty danych ustalonych w Konwencji;
11. Zauważając, że odstępstwa tego rodzaju są również przewidziane w istniejącym prawodawstwie szeregu krajów członkowskich lub w prawodawstwie, będącym w trakcie przygotowania dotyczącym ochrony danych;
12. Mając na uwadze, że należy zapewnić równowagę między potrzebami badań i statystyki z jednej strony i koniecznością ochrony jednostki, szczególnie kiedy mamy do czynienia z automatycznym przetwarzaniem danych, z drugiej;
13. Mając świadomość potrzeby określenia właściwych procedur zmierzających do pogodzenia interesów różnych zainteresowanych stron;
14. Mając świadomość, że postęp jaki nastąpił w metodach statystycznych i technologii przetwarzania informacji od 1983 r. spowodował konieczność dokonania rewizji postanowień zawartych w Zaleceniu nr R(83)10 w sprawie ochrony danych osobowych wykorzystywanych do badań naukowych i statystyki.

Zaleca, aby rządy krajów członkowskich:

1. Podjęły kroki zapewniające, aby zasady zawarte w załączniku do niniejszego Zalecenia znalazły odzwierciedlenie w ich przepisach prawnych i w praktyce;
2. Zapewniły szerokie rozpowszechnienie zasad zawartych w załączniku do niniejszego Zalecenia wśród osób, władz publicznych oraz instytucji zajmujących się zbieraniem i przetwarzaniem danych osobowych do celów statystycznych zarówno w sektorze publicznym, jak i prywatnym, jak również wśród władz odpowiedzialnych za ochronę danych;
3. Zachęciły te osoby, władze publiczne i instytucje, jeśli tego dotychczas nie uczyniły, do wprowadzenia zasad etycznych opartych o zasady sformułowane w załączniku do niniejszego Zalecenia.

Niniejsze Zalecenie zastępuje Zalecenie R(83)10 w sprawie ochrony danych osobowych wykorzystywanych do badań naukowych i statystycznych, o ile Zalecenie to dotyczy zbierania i automatycznego przetwarzania danych osobowych do celów statystycznych.

Załącznik do Zalecenia

1. Definicje

Przy stosowaniu niniejszego Zalecenia:

Wyrażenie **dane osobowe** oznacza dowolną informację odnoszącą się do zidentyfikowanej lub dającej się zidentyfikować osoby fizycznej (podmiot danych). Osoby fizycznej nie należy traktować jako „dającej się zidentyfikować”, jeśli identyfikacja wymaga nieracjonalnie dużego nakładu czasu i pracochłonności. Jeśli osoba fizyczna nie daje się zidentyfikować, dane uważa się za anonimowe.

Dane identyfikacyjne obejmują te dane osobowe, które umożliwiają bezpośrednią identyfikację podmiotu danych i które są niezbędne w celu zbierania, kontroli i łączenia danych, lecz nie są następnie wykorzystywane do opracowania wyników.

Dane wrażliwe oznaczają dane osobowe ujawniające pochodzenie rasowe, poglądy polityczne, wierzenia religijne lub inne dane osobowe dotyczące stanu zdrowia, życia płciowego lub wyroków karnych skazujących, a także inne dane określane jako wrażliwe w prawie krajowym.

Przetwarzanie oznacza dowolną operację lub zespół operacji wykonywanych częściowo lub całkowicie z pomocą procesów automatycznych w odniesieniu do danych osobowych, takie jak: przechowywanie, konserwacja, adaptacja lub dokonywanie zmian, selekcja, przeglądanie, wykorzystywanie, przesyłanie, łączenie i kasowanie lub niszczenie.

Przesyłanie odnosi się do czynności, które powodują, że dane osobowe stają się dostępne dla osób trzecich, niezależnie od rodzaju zastosowanych sposobów lub nośników.

Do celów statystycznych odnosi się do wszelkich operacji zbierania i przetwarzania danych osobowych niezbędnych do badań statystycznych i opracowania wyników. Czynności te wykluczają jakiegokolwiek wykorzystanie uzyskanych informacji do podejmowania decyzji lub przedsięwzięć dotyczących określonej osoby.

Wyrażenie „wyniki badań statystycznych” oznacza informacje, które uzyskano za pomocą przetwarzania danych osobowych w celu scharakteryzowania zbiorowych zjawisk w badanej populacji.

Zarządzający oznacza osobę fizyczną lub prawną, władzę publiczną lub każdy inny organ, który sam lub we współpracy z innymi określa cele, sposoby i w szczególności organizację zbierania i przetwarzania danych osobowych.

2. Zakres stosowania

2.1. Niniejsze Zalecenie odnosi się do zbierania i automatycznego przetwarzania danych osobowych do celów statystycznych. Odnosi się ono również do wyników badań statystycznych, jeśli umożliwiają one identyfikację podmiotów.

2.2. Zaleca się krajom członkowskim rozciągnąć zastosowanie niniejszego Zalecenia na nieautomatyczne przetwarzanie danych osobowych do celów statystycznych.

2.3. Nie należy przetrzymywać danych osobowych w sposób nie zautomatyzowany w celu ominięcia zasad ustalonych w niniejszym Zaleceniu.

2.4. Kraje członkowskie mogą również rozszerzyć zastosowanie zasad określonych w niniejszym Zaleceniu do zbierania i przetwarzania danych odnoszących się do grup osób, stowarzyszeń, fundacji, przedsiębiorstw, korporacji oraz wszelkich innych organów składających się bezpośrednio lub pośrednio z osób, niezależnie od tego czy posiadają one czy nie osobowość prawną.

3. Poszanowanie prywatności

3.1. Poszanowanie praw i podstawowych swobód, w szczególności prawa do prywatności, winno być zagwarantowane, jeśli dane osobowe są zbierane i przetwarzane do celów statystycznych oraz

- a) jeśli dane te są przechowywane do celów wykorzystania w przyszłości;
- b) jeśli wyniki badań statystycznych są rozpowszechniane³⁾ oraz
- c) jeśli w celu zapewnienia lepszej reprezentatywności danych statystycznych lub z przyczyn poufności, dane osobowe muszą być modyfikowane.

3.2. Osoby, które w związku z prowadzoną działalnością statystyczną uzyskują dostęp do danych osobowych, winny być zobowiązane do zachowania tajemnicy zawodowej przez prawo krajowe lub praktykę krajową.

3.3. Dane osobowe zbierane i przetwarzane do celów statystycznych winny być pozbawione identyfikatorów, jak tylko nie są już one niezbędne w postaci zidentyfikowanej.

4. Ogólne warunki legalnego zbierania i przetwarzania danych do celów statystycznych

Cel

4.1. Dane osobowe zbierane lub przetwarzane do celów statystycznych winny służyć wyłącznie do tych celów. W szczególności, nie mogą być one

³⁾ Podkreślić w Memorandum wyjaśniającym, że rozpowszechnianie obejmuje również publikacje.

wykorzystywane do podejmowania decyzji lub do oceny w odniesieniu do podmiotu, lub do uzupełnienia, lub korekty zbiorów zawierających dane osobowe, które są przetwarzane dla celów niestatystycznych.

4.2. Przetwarzanie do celów statystycznych danych osobowych zbieranych do celów niestatystycznych nie jest niezgodne z celami, dla których dane były początkowo zebrane, jeśli są podjęte właściwe zabezpieczenia⁴⁾), zwłaszcza przed wykorzystaniem danych do podejmowania decyzji w odniesieniu do podmiotu.

Legalność

4.3. Dane osobowe mogą być zbierane i przetwarzane do celów statystycznych, jeżeli:

- a) jest to przewidziane w prawie krajowym lub
- b) prawo krajowe na to pozwala, pod warunkiem że:
 - i) podmiot lub jego/jej prawny przedstawiciel wyraził/wyraziła zgodę stosownie do p. 5 lub
 - i) (bis) podmiot został poinformowany o zbieraniu lub przetwarzaniu dotyczących go danych i nie zgłosił sprzeciwu oraz, o ile przetwarzanie nie dotyczy danych wrażliwych
 - ii) warunki, w których dane są zbierane oraz cel badania mają taki charakter, że dają możliwość udzielania odpowiedzi w imieniu i z upoważnienia innych osób, stosownie do p. 6 i o ile w sposób oczywisty nie istnieje ryzyko naruszenia prywatności tych osób.

4.4. W celu uniknięcia powtórnego zbierania danych, dane osobowe zebrane w celach niestatystycznych mogą być przetwarzane do celów statystycznych, jeśli jest to niezbędne:

- a) do wykonania zadania realizowanego w interesie publicznym lub wykonania obowiązku służbowego;
- b) w uprawnionym interesie zarządzającego danymi, z wyjątkiem przypadków, gdy prawa i podstawowe wolności podmiotu mają pierwszeństwo wobec tych interesów.

Dane zebrane w jednym celu statystycznym mogą być również przetworzone do innych celów statystycznych, przy zachowaniu warunków opisanych wyżej.

4.5. Dane osobowe mogą być zbierane i przetwarzane na zasadzie obowiązku do celów statystycznych wyłącznie, jeśli jest to wymagane przez prawo krajowe.

4.6. Dane osobowe lub zbiory danych osobowych mogą być łączone lub dobierane do celów statystycznych, jeśli prawo krajowe zapewnia właściwe środki ochrony zapobiegające ich przetwarzaniu lub przekazywaniu do celów niestatystycznych.

⁴⁾ W Memorandum wyjaśniającym należy wyjaśnić jakie rodzaje zabezpieczeń winny być podjęte.

Proporcjonalność

4.7. Tylko te dane osobowe winny być zbierane i przetwarzane, które są niezbędne do osiągnięcia celów statystycznych. W szczególności dane identyfikacyjne winny być zbierane i przetwarzane tylko wtedy, kiedy jest to niezbędne.

Dane wrażliwe

4.8. Jeśli dane wrażliwe podlegają przetwarzaniu do celów statystycznych, muszą być one zbierane w formie, w której podmioty nie są możliwe do zidentyfikowania.

Jeśli przetwarzanie danych wrażliwych prowadzone w określonych, prawnie uzasadnionych, celach statystycznych wymaga identyfikacji podmiotów badania, to prawo krajowe winno zapewnić właściwe środki ochrony, w tym specyficzne środki zapewniające oddzielenie danych identyfikujących, począwszy od etapu zbierania danych, chyba że jest to w sposób oczywisty nieracjonalne lub nie praktykowane.

5. Informacja

Pierwotne zbieranie danych

5.1. Jeśli do celów statystycznych zbierane są dane osobowe, badane osoby winny być poinformowane o:

- a) obowiązkowym lub dobrowolnym charakterze udzielania odpowiedzi oraz o podstawie prawnej określającej zbieranie danych;
- b) celu lub celach zbierania i przetwarzania;
- c) nazwisku i statusie prawnym osoby lub organu odpowiedzialnego za zbieranie i/lub przetwarzanie;
- d) tym, że dane będą utrzymywane w tajemnicy i wykorzystywane wyłącznie do celów statystycznych;
- e) możliwości uzyskania dalszych informacji na żądanie.
Na żądanie oraz/lub stosownie do metod i sposobów określonych w prawie krajowym, podmioty winny być poinformowane również o:
 - f) sposobach dobrowolnych badań, w których może być wyrażona odmowa lub cofnięta zgoda oraz, w przypadku badań obowiązkowych, jakie sankcje grożą za odmowę udziału w badaniu;
 - g) w stosownych przypadkach — warunkach korzystania z prawa dostępu oraz z prawa do sprostowania danych;
 - h) rodzajach osób lub organów, którym dane osobowe mogą być przekazane;
 - i) gwarancjach zapewniających poufność i ochronę danych osobowych;
 - j) rodzajach zbieranych i przetwarzanych danych.

5.2. Jeśli podmioty nie podlegają obserwacji metodą bezpośrednią winny być informowane o prowadzeniu badań, chyba że jest to w oczywisty sposób nieuzasadnione lub niepraktyczne. Winny one być w stanie poinformować się odpowiednio za pomocą elementów wyszczególnionych w p. 5.1.

5.3. Badane osoby, niezależnie od tego, czy są one podmiotami danych czy nie, winny zostać poinformowane najpóźniej w czasie zbierania danych. Sposób i zakres tej informacji winien być właściwy oraz dostosowany do warunków.

Jeśli jest to niezbędne dla osiągnięcia właściwych celów badania w opinii jego podmiotów oraz istoty badania, przedstawienie informacji lub jej części może być opóźnione. Informacja musi być jednakże udostępniona zanim dane osobowe zostaną przetworzone, chyba że konieczność nieujawnienia części lub całości informacji w sposób oczywisty istnieje nadal lub, jeśli przekazanie tej informacji jest jawnie nierozsądne lub niepraktyczne.

Wtórne zbieranie danych

5.4. Informacja o przetwarzaniu lub przesyłaniu do celów statystycznych danych osobowych zebranych do celów niestatystycznych powinna zostać odpowiednio upowszechniona. Podmioty badań winny mieć możliwość uzyskania we właściwy sposób informacji, o których mowa w p. 5.1, chyba że:

- a) dostarczenie tej informacji jest niemożliwe lub wymaga nieproporcjonalnie dużego wysiłku lub
- b) przetwarzanie lub przesyłanie danych do celów statystycznych jest wyraźnie określone w prawie krajowym.

Osoby pozbawione zdolności do działań prawnych

5.5. Jeśli podmiot danych jest osobą pozbawioną całkowicie zdolności do działań prawnych, niezdolną do podejmowania świadomych decyzji oraz prawo krajowe nie pozwala tej osobie działać w jej własnym imieniu, informacja winna być przekazana osobie prawnie upoważnionej do występowania w interesie podmiotu.

Jeśli osoba pozbawiona zdolności do działań prawnych jest w stanie rozumieć, winna być ona poinformowana przed tym, zanim dotyczące jej dane będą zbierane lub przetwarzane.

6. Zgoda

6.1. Jeśli wymagana jest zgoda podmiotu, to winna być ona wyrażona w sposób wolny i świadomy.

Podmiot winien mieć możliwość wycofania swojej zgody na udział w badaniu, o ile zgodnie z p. 8 dane identyfikacyjne nie zostały oddzielone od pozostałych zebranych danych lub wstrzymania zgody w dowolnym czasie i bez retroaktywnych skutków jego współpracy w badaniu trwającym przez dłuższy okres.

6.2. Jeśli wymagana jest zgoda podmiotu na zbieranie lub przetwarzanie wrażliwych danych, winna być ona wyrażona w sposób nieprzymuszony i świadomy, o ile nie zaszkodzi to osiągnięciu prawnie uzasadnionych celów badania, które odpowiada ważnym interesom publicznym.

6.3. Jeśli przewiduje się przetwarzanie do celów statystycznych danych osobowych od osób pozbawionych zdolności do działań prawnych, niezdolnych do podejmowania swobodnych decyzji oraz jeśli prawo krajowe nie pozwala podmiotowi występować w swoim imieniu, należy uzyskać zgodę osoby uznanej za prawnie upoważnioną do działań w imieniu podmiotu lub władzy, lub innej osoby, lub organu przewidzianego prawem.

Jeśli, stosownie do wymienionego wyżej p. 5.5, osoba pozbawiona zdolności do działań prawnych została poinformowana o zamiarze zbierania i przetwarzania jej danych osobowych, jej życzenie może być wzięte pod uwagę, o ile prawo krajowe nie przewiduje inaczej.

7. Prawo do dostępu i do prostowania

7.1. W celu ochrony swoich interesów każda osoba może otrzymać dotyczące jej dane osobowe przechowywane przez zarządzającego danymi, a w razie potrzeby może je sprostować.

7.2. Prawo to może być jednak ograniczone zgodnie z prawem krajowym, jeśli dane osobowe są przetwarzane wyłącznie w celach statystycznych i istnieją właściwe i specyficzne środki zapobiegające jakiegokolwiek identyfikacji przez osoby trzecie na podstawie danych indywidualnych lub wyników statystycznych.

7.bis. Usuwanie identyfikatorów

7.bis 1. Dane osobowe zbierane do celów statystycznych winny być pozbawione części identyfikacyjnej niezwłocznie po zakończeniu operacji zbierania, kontroli i łączenia, z wyjątkiem przypadków, gdy:

- 1) dane identyfikacyjne są nadal niezbędne do celów statystycznych, a środki ostrożności przewidziane w punkcie 9.1 są zachowane lub
- 2) sama istota przetwarzania danych statystycznych wymaga rozpoczęcia innych operacji przetwarzania przed usunięciem identyfikatorów i o ile środki ostrożności przewidziane w punktach 14.1—14.3 są zachowane.

8. Pierwotne zbieranie danych osobowych do celów statystycznych

8.1. Zbieranie danych osobowych winno być rzetelne, zwłaszcza jeśli chodzi o informację o osobach indywidualnych oraz o ich swobodę udzielania odpowiedzi.

8.2. Dane osobowe winny być uzyskiwane od podmiotu lub, w zależności od rodzaju badania, od członka jego gospodarstwa domowego. Dane osobowe mogą być zbierane od innej osoby niż podmiot badany lub od kogoś spoza jego gospodarstwa, lub od osoby prawnej, takiej jak przedsiębiorstwo lub instytucja publiczna, jeśli prawo krajowe to przewiduje oraz zawiera odpowiednie gwarancje lub jeśli nie istnieje żadne ryzyko naruszenia praw i podstawowych wolności badanego podmiotu.

8.3. Zbieranie danych osobowych do celów statystycznych bez przeprowadzania bezpośredniego wywiadu nie może zawierać danych identyfikacyjnych lub być łączone z danymi identyfikacyjnymi, z wyjątkiem przypadków, gdy prawo krajowe zawiera odpowiednie gwarancje oraz.

a) przewiduje zbieranie danych z danymi identyfikacyjnymi lub

b) pozwala łączyć zbierane dane z danymi identyfikacyjnymi, w celu tworzenia prób do badań reprezentacyjnych.

8.4. Dane dotyczące respondentów odmawiających udzielenia odpowiedzi, mające znaczenie dla planowania lub przeprowadzania badania, względnie informacje na temat przyczyn braku odpowiedzi, mogą być wykorzystane wyłącznie w celu zapewnienia reprezentatywności badania.

8.5. W przypadku gdy zbieranie danych osobowych wymaga zatrudnienia ankierów lub innych osób, które będą musiały znać treść odpowiedzi, należy podjąć szczególne starania dotyczące doboru osób oraz organizacji i metod badania, aby zapewnić przestrzeganie celów badania, poufności danych oraz ochrony prywatności.

8.6. Zarządzający winien podjąć odpowiednie środki umożliwiające badanym osobom przekonanie się co do uprawnień osoby zbierającej dane.

9. Dane identyfikacyjne

9.1. W przypadku gdy w badaniach statystycznych zbierane i przetwarzane są dane identyfikacyjne, winny być one oddzielone i przechowywane oddzielnie od pozostałych danych osobowych, z wyjątkiem przypadków, gdy jest to jawnie niecelowe lub niepraktyczne.

9.2. Dane identyfikacyjne mogą być jednakże wykorzystywane do tworzenia zbiorów adresowych do celów statystycznych, jeśli zezwala na to prawo krajowe, jeśli podmiot został o tym poinformowany oraz nie zgłosił sprzeciwu lub jeśli dane pochodzą z powszechnie dostępnego źródła.

10. Przechowywanie danych

10.1. Z wyjątkiem przypadków, gdy dane osobowe zostały pozbawione identyfikatorów lub gdy prawo krajowe zezwala na przechowywanie takich danych, dane osobowe zbierane i przetwarzane do celów statystycznych

winy być zniszczone lub skasowane, jeśli nie są już niezbędne do tych celów.

W szczególności dane identyfikacyjne winny być zniszczone lub skasowane niezwłocznie po tym, jak przestały być niezbędne:

- a) w celu zbierania, kontroli i łączenia danych,
- b) dla zapewnienia reprezentatywności badania lub
- c) w przypadku powtarzania badania z udziałem tych samych osób.

11. Przekazywanie danych

11.1. Dane osobowe zbierane do celów statystycznych nie mogą być przekazywane do celów niestatystycznych.

11.2. Dane osobowe przetwarzane do określonego celu statystycznego mogą być przekazywane do innych celów statystycznych, o ile są one wyraźnie określone i mają krótkotrwały charakter.

11.3. Jeśli odpowiednie zabezpieczenia, w sprawie przechowywania danych, nie są zapewnione przez prawo krajowe, przekazywanie o którym mowa w pkt. 11.2. winno odbywać się na podstawie dokumentu określającego prawa i obowiązki stron. Jeśli dane te są przekazywane, zarządzający winien w szczególności:

- a) zastrzec, że strona trzecia może te dane przekazywać wyłącznie za jego wyraźną zgodą;
- b) zastrzec, że strona trzecia podejmie odpowiednie środki bezpieczeństwa, zgodnie z p. 14.1. do 14.3. niniejszego Zalecenia;
- c) zapewnić, aby jakakolwiek publikacja wyników statystycznych uzyskanych przez tę stronę była zgodna z p.13 niniejszego Zalecenia.

11.4. Ponadto, dane wrażliwe mogą być przekazywane wyłącznie w przypadkach przewidzianych przez prawo krajowe lub gdy podmioty lub ich przedstawiciele prawni wyraziły na to formalną zgodę i o ile prawo krajowe się temu nie sprzeciwia.

12. Transgraniczny przepływ danych

12.1. Reguły niniejszego Zalecenia winny być stosowane w odniesieniu do transgranicznego przesyłania danych osobowych do celów statystycznych.

12.2. Transgraniczne przesyłanie danych osobowych do celów statystycznych do kraju, który ratyfikował Konwencję ETS 108⁵⁾ oraz który posiada przepisy prawne zapewniające co najmniej równoważną ochronę, nie powinno podlegać szczególnym warunkom odnośnie ochrony prywatności, praw i podstawowych wolności osób.

12.3. Nie należy wprowadzać ograniczeń w transgranicznym przesyłaniu danych osobowych do celów statystycznych do kraju, który nie ratyfikował

⁵⁾ Konwencja dotycząca ochrony osób w związku z automatycznym przetwarzaniem danych osobowych — patrz aneks.

Konwencji ETS 108, ale zapewnia poziom ochrony odpowiadający regułom tej Konwencji oraz niniejszego Zalecenia.

12.4. O ile prawo krajowe nie przewiduje inaczej, transgraniczne przesyłanie danych osobowych do celów statystycznych do kraju, który nie posiada prawnych zabezpieczeń zgodnych z regułami Konwencji ETS 108 oraz z niniejszym Zaleceniem, w zasadzie nie powinno mieć miejsca, chyba że:

- a) podjęto środki, w tym środki kontraktowe niezbędne do przestrzegania reguł Konwencji oraz niniejszego Zalecenia lub
- b) podmiot danych wyraził na to formalną zgodę.

13. Wyniki badań statystycznych

13.1. Wyniki badań statystycznych mogą być publikowane lub udostępniane Stronom trzecim, jeśli podjęto środki zapewniające, że podmioty danych nie mogą być zidentyfikowane na podstawie tych wyników, o ile rozpowszechnienie lub opublikowanie bezspornie nie przedstawia ryzyka naruszenia prywatności podmiotów.

14. Bezpieczeństwo danych osobowych

14.1. Zarządzający winni podjąć niezbędne środki zaradcze dla zapewnienia poufności danych osobowych. Są oni zobowiązani w szczególności podjąć środki zaradcze przeciwko nie upoważnionemu dostępowi, wprowadzeniu zmian, przekazaniu lub jakiegokolwiek innej formie nie upoważnionego przetwarzania danych.

14.2. Jeśli dane muszą być zachowywane w formie zidentyfikowanej, należy zastosować środki organizacyjne i techniczne, w szczególności środki automatyczne w celu zapobieżenia nie upoważnionemu zidentyfikowaniu podmiotu.

14.3. Należy podejmować środki zapobiegające reidentyfikacji podmiotów oraz wykorzystywaniu do celów pozastatystycznych danych osobowych zbieranych do celów statystycznych.

14.4. Specjaliści statystycy, firmy oraz instytucje zajmujące się opracowaniami statystycznymi winni rozwijać metody i procedury zapewniające anonimowość badanych podmiotów.

15. Kodeks etyczny

15.1. Specjaliści statystycy, firmy oraz instytucje zajmujące się opracowaniami statystycznymi winni przyjąć i opublikować kodeks etyczny, który odpowiada regułom ustalonym w niniejszym Zaleceniu, obejmujący informacje w szczególności o:

- a) innych kategoriach osób i instytucji, które mają dostęp do danych osobowych;

- b) środkach jakie należy podejmować w celu ochrony, zapewnienia poufności i bezpieczeństwa danych oraz o środkach zapewniających przestrzeganie etyki statystycznej;
- c) zarządzających procesem przetwarzania danych statystycznych.

16. Rozwój techniczny, współpraca i pomoc

W celu zapewnienia szerokiego dostępu do narzędzi informacyjnych oraz do wiedzy technicznej, właściwej dla efektywnej ochrony danych osobowych zbieranych do celów statystycznych, kompetentne służby rządowe winny ściśle współpracować w dziedzinie rozwoju tych narzędzi i wiedzy technicznej oraz uruchomić międzynarodowy program współpracy, wymiany doświadczeń, transferu wiedzy oraz pomocy technicznej.

Dotychczas w serii
BIBLIOTEKA WIADOMOŚCI STATYSTYCZNYCH
ukazały się następujące pozycje:

1967 r.

- Tom 1. Wybrane problemy statystyki w Polsce
- Tom 2. Zagadnienia statystyki rolniczej
- Tom 3. Statystyka regionalna. Aktualny stan i problemy rozwoju
- Tom 4. Problemy demograficzne Polski Ludowej

1968 r.

- Tom 5. Bilanse gospodarki narodowej

1969 r.

- Tom 6. Problemy demograficzne Ziem Zachodnich i Północnych PRL
- Tom 7. Zastosowanie metod matematycznych w statystyce
- Tom 8. Problemy demograficzne Kielecczyzny
- Tom 9. Mierniki rozwoju regionów

1970 r.

- Tom 10. Założenia programowe i organizacyjne Narodowego Spisu Powszechnego w 1970 r.
- Tom 11. Wybrane problemy prognoz statystycznych
- Tom 12. Aktualne problemy statystyki regionalnej w krajach europejskich
- Tom 13. Teoretyczne i metodologiczne problemy statystyki społecznej

1971 r.

- Tom 14. Badania statystyczne metodą reprezentacyjną w krajach socjalistycznych
- Tom 15. Wybrane problemy metodologiczne badań reprezentacyjnych
- Tom 16. Rola i zadania statystyki państwowej w planowaniu i zarządzaniu gospodarką narodową w krajach RWPG

1972 r.

- Tom 17. Aktualne problemy statystyki państwowej
- Tom 18. Eksperymentalne badania budżetów rodzinnych metodą rotacyjną

1973 r.

- Tom 19. Problemy demograficzne województwa zielonogórskiego
- Tom 20. Podstawowe problemy rozwoju statystyki regionalnej
- Tom 21. Stan i perspektywy rozwoju statystyki w Polsce
- Tom 22. Rozwój regionalny Polski w świetle badania dochodu narodowego

1974 r.

- Tom 23. Problemy mierników poziomu życia ludności
- Tom 24. Aktualne problemy demograficzne kraju

1976 r.

- Tom 25. Statystyka i ekonometria w Polsce Ludowej
- Tom 26. Zagadnienia metodologiczne statystyki społeczno-demograficznej

1978 r.

- Tom 27. Rola młodzieży w życiu społeczno-gospodarczym kraju
Tom 28. Narodowy Spis Powszechny 7 grudnia 1978 r.
Metodologia i organizacja
Tom 29. Metodologia badań reprezentacyjnych w GUS
Prace Komisji Matematycznej

1979 r.

- Tom 30. Narodowy Spis Powszechny 1978 jako źródło informacji o migracjach
Tom 31. Statystyka i ekonometria w Polsce Ludowej (wyd. drugie — zmienione)

1981 r.

- Tom 32. Tematyka i organizacja spisów powszechnych w Polsce

1983 r.

- Tom 33. Problemy demograficzne województwa lubelskiego

1987 r.

- Tom 34. Problemy integracji statystycznych badań gospodarstw domowych
Tom 35. Dokładność danych w badaniach społecznych (autor Jan Kordos)

1989 r.

- Tom 36. Problemy badań statystycznych metodą reprezentacyjną
Tom 37. Badanie budżetu czasu ludności

1990 r.

- Tom 38. Statystyczne metody badania cen

1991 r.

- Tom 39. Zarys metody reprezentacyjnej (autor Ryszard Zasępa)
Tom 40. Jakość życia i warunki bytu
Tom 41. Wskaźniki społeczne

1994 r.

- Tom 42. Rozwój myśli i instytucji statystycznych na ziemiach polskich
Tom 43. Wyzwania polskiej statystyki
Tom 44. Rozwój metodologii badań statystycznych w Polsce
Tom 45. Statystyka regionalna
Tom 46. Obchody jubileuszowe 200-lecia statystyki polskiej i 75-lecia Głównego Urzędu Statystycznego
Tom 47. Podstawowe zasady statystyk oficjalnych oraz ich prawnych i etycznych aspektów w krajach w okresie przejściowym

1995 r.

- Tom 48. Przepisy prawne w statystyce
Tom 49. 130 lat statystyki Warszawy 1864—1994

1996 r.

- Tom 50. Ochrona danych w systemie informacyjnym statystyki publicznej
(autor Tadeusz Walczak)



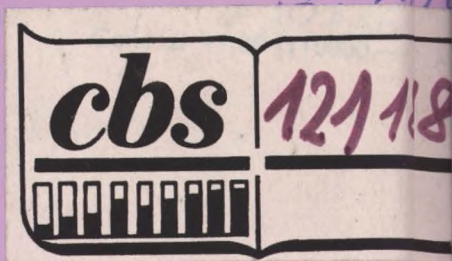
1875
28 APR

1875

11, -

A/573

4/227/9



CENTRALNA BIBLIOTEKA
STATYSTYCZNA - GUS



104063