

**GLÓWNY
URZĄD
STATYSTYCZNY**

**OŚRODEK BADAWCZO-ROZWOJOWY
SYSTEMU
PAŃSTWOWEJ INFORMACJI STATYSTYCZNEJ**

SERIA

SYSTEMY INFORMATYCZNE

SEMINARIUM

SPIS '77

**MATERIAŁY
INFORMACYJNE
I SZKOLENIOWE**

1/78

CZĘŚĆ

I

Problemy metodologii i organizacji centralnych systemów informatycznych

Semiotyka ekonomiczna w projektowaniu systemów centralnych — wybrane problemy

Systemy informatyczne a modelowanie procesów ekonomicznych w makroskali

**METODY PROJEKTOWANIA
I WDRAŻANIA CENTRALNYCH
SYSTEMÓW
INFORMATYCZNYCH**

materiały seminarium SPIS '77

redaktor naukowy Józef OLEŃSKI

**redaktor Andrzej BRATKOWSKI
Andrzej GALAS**

**Ośrodek Badawczo-Rozwojowy SPIS
Warszawa, Al. Niepodległości 208**

Druk

METODY PROJEKTOWANIA
I WDRZANIA CENTRALNYCH
SYSTEMÓW
INFORMATYCZNYCH

Wydawnictwo Naukowe PWN



191310

SPIS TREŚCI

	str.
Przedmowa	9
Wystąpienie Prezesa GUS	19

CZĘŚĆ I

PROBLEMY METODOLOGII I ORGANIZACJI CENTRALNYCH SYSTEMÓW INFORMATYCZNYCH

T. Walczak	Niektóre aspekty metodologii projektowania rządowych systemów informatycznych	27
J. Oleński	W sprawie metodyki projektowania i wdrażania centralnych systemów informatycznych	51
S. Abt	Problemy metodyki budowy banków informatycznych	66
J. Bankowski	Metodologia prac projektowych dla potrzeb SINTO	87
J. I. Czerniak	Informacyjne aspekty zautoantyzowanych systemów zarządzania w ZSRR	95
J. Eysmontt	Uwagi na temat zasad projektowania informatycznego systemu planowania centralnego	115
J. Kiśielnicki	Organizacyjne problemy projektowania wielkich systemów informatycznych	130
B. Warzecha	Analiza systemowa w projektowaniu	

centralnych systemów informatycznych /z doświadczeń projektowania SPIS-WBD/	145
---	-----

SEMIOTYKA EKONOMICZNA W PROJEKTOWANIU SYSTEMÓW CENTRALNYCH
/WYBRANE PROBLEMY/

J. Kowalczyk	Proces konstrukcji modelu jako podstawa budowy języka użytkownika baz danych SPIS	215
J. Oleński	Słownik w systemie informacji statystycznej	234
A. Paradowska	Katalog użytkownika i kodowanie informacji w bazie danych	251
Z. Peuker	Funkcje koordynacyjne podsystemu "Słownik" w systemie badań statystycznych /założenia ogólne/	269
B. Szybisz	Zasady oraz program opracowania słownika kategorii społeczno-ekonomicznych w systemie GENPLAN	279

SYSTEMY INFORMATYCZNE A MODELOWANIE PROCESÓW EKONOMICZNYCH
W MAKROSKALI

A. Dąbkowski	Modelowanie systemowe ogólnogospodarczych proporcji rozwojowych	297
J. Kolek	Makroekonomiczna analiza statystyczna na bazie zautomatyzowanych systemów informacyjnych SSRS	332
W. Maciejewski	Wybrane problemy skomputeryzowanych systemów modeli	355
Z. Pawłowski	O systemie prognostycznym w gospodarce narodowej	375

CZĘŚĆ II

	<u>str.</u>
Przedmowa	389
TECHNOLOGIA BANKÓW DANYCH - PROBLEMY PROJEKTOWANIA I WDRAŻANIA	
W. Jamontt	Problemy projektowania bazy danych statystycznych o jednostkach gospodarstwa uspołecznionej 399
W. Jamontt	Relacyjny model baz danych 424
W. Kamiński	System bazy danych a system wyszukiwania informacji 442
M. A. Kowalewski	Problemy dostępu bezpośredniego w systemach zarządzania bazą danych 465
M. Lasota	Bank danych statystycznych w technologii DBTG Codasyl /opis wdrożenia eksperymentalnego/ 487
S. Mężyk	Mechanizmy odwzorowań w systemach zarządzania bazą danych 516
J. Pasuła	Problemy obsługi jednoczesnego dostępu do zintegrowanej bazy danych .. 547
A. Scheber	SOFIS - system zarządzania wielkimi bazami danych 575
A. Stremowski	Odtwarzanie stanów bazy danych, metody i techniki 591
DOŚWIADCZENIA Z PRAKTYCZNYCH WDROŻEŃ CENTRALNYCH SYSTEMÓW INFORMATYCZNYCH	
K. Czerniewski	Zagadnienia projektowania rolniczego systemu informacyjnego 621
E. Działoszyński	Elementy systemu informatycznego o środowisku glebowym 627

J. Homa,	Wielowojewódzki bank danych /z doświadczeń	
G. Zienkiewicz	funkcjonowania zespołu wojewódzkich banków	
	danych SPIS-WBD w OE GUS - Katowice/	636
S. Kosewski	Wykorzystanie standardowych pakietów prze-	
	tworzenia danych w pracach nad wdrożeniem	
	systemów informacyjnych dla potrzeb	
	planowania przestrzennego	648
W. Leśnikowski	Refleksje wdrożeniowe na przykładzie	
	systemu identyfikacji terenowej	652
S. Małkus	Komunikat o uruchomieniu w OE GUS we	
	Wrocławiu systemu zdalnego dostępu do	
	zbiorów danych statystycznych na sprzęcie	
	teletransmisji produkcji krajowej	665
N. N. Pietrowa	Metodologiczne podstawy projektowania	
	informatycznego systemu zarządzania re-	
	sortem /na przykładzie Ministerstwa	
	Skupu Zbóż RFSRR/	670
H. Postaremczak	Automatyczny system oceny wyników dzia-	
	łalności przemysłu w warunkach zmodyfiko-	
	wanego systemu ekonomiczno-finansowego	
	/ASO/	696
K. Prussak	Doświadczenia w zakresie opracowania	
	oprogramowania własnego /użytkowego/ dla	
	potrzeb banków informacji statystycznej	729
M. Salata	System informatyczny statystyki handlu	
	zagranicznego	742
B. Sieroi	Systemy statystyki w zasilaniu banku	
	danych /z doświadczeń projektowania i	
	wdrażania SPIS - WBD/	770

W. Wchnachter	Założenia projektowo-programowe systemu elektronicznego przetwarzania danych Narodowego Spisu Powszechnego 1978 r.	786
---------------	--	-----

WYBRANE PROBLEMY DOSKONALENIA PROJEKTOWANIA SYSTEMÓW
CENTRALNYCH

K. Bernatowicz	Ewidencja systemów EPD - sytuacja i propozycje	801
J. I. Czerniak	Teoretyczne podstawy doskonalenia systemów informatycznych	814
J. Heronimek	Problemy spójności i współdziałania rządowego systemu informatycznego SPIS z resortowymi systemami informatycznymi	827
J. Ilczuk	SCERT - Metoda racjonalnego doboru zestawu komputerowego	840
T. Możdżyńska	Problemy koordynacji prac projektowych i oprogramowanie systemów przetwarzania danych w ośrodkach elektronicznych GUS	850

Doc. dr hab. Tadeusz WALCZAK

Wiceprezes GUS

NIEKTÓRE ASPEKTY METODOLOGII PROJEKTOWANIA
RZĄDOWYCH SYSTEMÓW INFORMATYCZNYCH

Rok temu - w końcu października 1976 r. odbyło się pierwsze spotkanie poświęcone omówieniu szeregu istotnych problemów związanych z zadaniami i funkcjonowaniem Systemu Państwowej Informacji Statystycznej oraz jego powiązaniem z innymi systemami informacyjnymi w kraju./1/

Głównym tematem tego seminarium było omówienie zadań informacji statystycznej z punktu widzenia potrzeb władz terenowych.

Bezpośrednia wymiana poglądów, jaka miała miejsce w czasie seminarium pomiędzy specjalistami reprezentującymi rządowe systemy informatyczne, niektóre systemy resortowe, ośrodki naukowe oraz usługowe ośrodki informatyczne pomogła wyjaśnić szereg istotnych wspólnych problemów oraz sprecyzować bliżej podstawowe kierunki prac nad dalszym rozwojem prac projektowych i wdrożeniowych w rządowych systemach informatycznych, w tym zwłaszcza w SPIS.

Obecne kolejne seminarium SPIS-77 poświęcone jest przede wszystkim omówieniu metod projektowania i wdrażania dużych systemów informatycznych na tle doświadczeń w projektowaniu systemów rządowych i resortowych.

Postawienie w charakterze głównego hasła seminarium problemu metodologii projektowania rządowych systemów informatycznych może zrodzić pytanie, czy oznacza to, że nie mamy dotychczas metodolo-

gii projektowania takich systemów? Czy istniejące metody projektowania, wypracowane głównie na bazie doświadczeń komputeryzacji w systemach zarządzania przedsiębiorstw, instytucji finansowych, handlowych itp. /zwanych ogólnie systemami obiektowymi/ są nieprzydatne w projektowaniu systemów rządowych?

Sądzę, że powinniśmy spróbować odpowiedzieć na to pytanie przed podjęciem prób definiowania i rozwiązywania szeregu szczegółowych problemów związanych z projektowaniem i wdrażaniem systemów rządowych.

Podjęwane przez nas prace nad projektowaniem rządowych systemów informatycznych mają w znacznej mierze charakter nowatorski, a nawet pionierski. Wynikiem końcowym tych prac i ich podstawowym celem jest stworzenie systemu informowania bardziej sprawnego i efektywnego, służącego zwiększeniu efektywności gospodarowania w skali całego kraju oraz lepszemu spełnianiu społecznych celów społeczeństwa socjalistycznego.

Jednocześnie prace te wymagają z konieczności bardzo dużego wysiłku i znacznych środków zarówno technicznych, jak i kadrowych, przy czym przyjęte rozwiązania w systemach rządowych rzutować będą nie tylko na sprawność i koszty funkcjonowania tych systemów, ale także w dużej mierze na ogólny koszt utrzymania systemów informacyjnych w ogromnej liczbie jednostek gospodarki uspołecznionej, które będą zasilać danymi źródłowymi systemy rządowe. Dlatego też trzeba aby projektanci systemów rządowych wykorzystywali w maksymalnym stopniu cały istniejący dorobek teoretyczny i doświadczenia praktyczne w projektowaniu i funkcjonowaniu systemów informatycznych, aby nowe zadania badawcze, specyficzne dla funkcjonowania systemów rządowych podejmowane były z pełną

świadomością wszystkich mocnych i słabych stron istniejących rozwiązań we wszystkich podstawowych systemach informacyjnych zarządzania istniejących w naszym kraju.

Teoria i metodologia projektowania systemów informatycznych rozwijała się równolegle z rozwojem środków technicznych przetwarzania danych. Sformułowane zostały zwłaszcza podstawowe etapy projektowania oraz metody realizacji poszczególnych etapów: analiza systemu, sformułowanie zadań projektowych, podział systemu na podsystemy, metody dokumentowania istniejącego systemu oraz metody opracowania dokumentacji nowego systemu. Sformułowane zostały także podstawowe zasady strategii wdrażania systemów informatycznych /2,3/.

Często można usłyszeć zastrzeżenia, że istniejące metody projektowania i wdrażania systemów, wyrosłe z doświadczeń uzyskanych w pracach nad systemami obiektowymi nie dają zadowalających rezultatów przy projektowaniu systemów rządowych oraz prowadzą do nieopłacalnych rozwiązań organizacyjnych i technologicznych, że trzeba wypracować odrębne metody analizy, projektowania i wdrażania^{1/}.

Opinia za, ma dość daleko idące konsekwencje praktyczne: sugeruje bowiem konieczność podjęcia rozległych prac badawczych nad metodologią projektowania rządowych systemów informatycznych jako wielkich systemów, Z tego względu uważam, że wymaga ona bliższej krytycznej oceny, ponieważ w tak ogólnym sformułowaniu może prowadzić do błędnych wniosków.

1/ pogląd ten znalazł między innymi wyraz w notatce informacyjnej na temat obecnego seminarium.

Wspólne cechy systemów rządowych i innych systemów
informatycznych

1. Przede wszystkim sędzę, że nie należy nie doceniać stopnia skomplikowania systemów obiektowych. Systemy zarządzania, a wraz z nimi i systemy informacyjne współczesnych przedsiębiorstw /systemy obiektowe/ są systemami niezwykle złożonymi. Stanowią one odzwierciedlenie złożonych stosunków wewnętrznych pomiędzy różnymi służbami przedsiębiorstwa, jak również wielostronnych stosunków zewnętrznych wynikających z powiązań kooperacyjnych /często międzynarodowych/, z hierarchicznych układów zarządzania, zobowiązań informacyjnych wobec systemów centralnych itp.

Pojęcie "wielki system" nie zostało dotąd jednoznacznie sprecyzowane. W ramach systemów zaliczanych do "wielkich", w zależności od punktu widzenia może istnieć wiele klas systemów. Według Aleksandra J. Lerner'a /4/ charakterystycznymi cechami wielkich systemów są:

- a/ możliwość wydzielenia pewnych części - podsystemów;
- b/ istnienie celu działania dla każdego podsystemu i możliwość oceny efektywności jego działania na podstawie funkcjonau, zależnego od oddziaływań sterujących, przyłożonych do danego podsystemu;
- c/ istnienie ogólnego celu działania WS i możliwość oceny efektywności jego działania na podstawie funkcjonau, zależnego od ruchu każdego podsystemu;
- d/ hierarchiczna struktura sterowania;

- e/ współdziałal w wielkim systemie ludzi, maszyn i oddziaływującego środowiska,
- f/ istnienie wielkiej liczby sprzężeń wewnątrz każdego podsystemu i między podsystemami,
- g/ konieczność organizacji rozgałęzionej sieci informacyjnej, zapewniającej funkcjonowanie systemu i optymalizację jego działania".

Jeżeli przyjmiemy przytoczoną wyżej charakterystykę wielkich systemów, to musimy stwierdzić, że systemy obiektowe, a przynajmniej te spośród nich które dotyczą wielkich przedsiębiorstw przemysłowych należą niewątpliwie do rzędu wielkich systemów.

Z tego względu, bez obawy wywołania nieporozumień i zamętu terminologicznego nie powinniśmy używać określenia "wielki system". W stosunku do systemów rządowych i innych systemów o zasięgu centralnym w sensie przeciwstawienia tego terminu systemom obiektowym, które w świetle przytoczonych wyżej kryteriów mają również wszystkie cechy właściwe wielkim systemom.

2. Zadania i cele systemów obiektowych wykazują szereg zbieżności z systemami rządowymi i resortowymi. Podobny jest również podstawowy zakres niezbędnych do przeprowadzenia prac projektowych przed zastosowaniem komputerów w obu tych klasach systemów.

3. Szereg szczegółowych rozwiązań w metodologii projektowania stosowanej w projektowaniu systemów obiektowych można zastosować w pełni również przy projektowaniu systemów rządowych. Mam na myśli zwłaszcza metody analizy istniejącego stanu zbierania i opracowania informacji, metody projektowania kodów, wzorów dokumentów źródłowych oraz maszynowych nośników danych, stosowanie

schematów blokowych lub tablic decyzyjnych w trakcie opracowania dokumentacji systemów itp.

4. Obie klasy systemów stawiają podobne wymagania z punktu widzenia rodzaju i konfiguracji sprzętu. Różnice jakie tu mogą wystąpić są bardziej różnicami ilościowymi wynikającymi z różnicy w skali wielkości tych klas systemów niż różnicami jakościowymi. Pomijam tutaj oczywiście problem jakości i niezawodności sprzętu, uważam bowiem, że istniejący w tym zakresie stan nie odpowiada wymogom ani systemów rządowych ani obiektowych i na dłuższą metę jest nie do tolerowania.

Opierając się na przedstawionych wyżej przesłankach, uważam, że potrzebna jest nam nie nowa teoria i metodologia projektowania systemów rządowych, lecz twórcza adaptacja istniejących metod projektowania oraz wykorzystanie coraz bogatszych doświadczeń funkcjonujących w naszym kraju systemów informatycznych.

Specyficzne cechy systemów rządowych

Jednocześnie istnieje potrzeba rozwinięcia prac badawczych i wdrożeniowych na tych odcinkach, które odznaczają się silną specyfiką w stosunku do innych funkcjonujących dotychczas systemów obiektowych.

Oprócz bowiem całego szeregu wspólnych cech i wspólnych problemów właściwych dla projektowania systemów rządowych i obiektowych, o których mówiłem wyżej, niektóre aspekty projektowania i wdrażania w obu klasach systemów różnią się w znacznym stopniu i wymagają zastosowania nieco innych metod, aby przyjęte rozwiązania były maksymalnie efektywne. Niektóre z tych różnic wywierających wpływ na metody projektowania i wdrażania systemów rządowych chciałbym krótko omówić opierając się na doświadczeniach projektowania

jednego z rządowych systemów informatycznych a mianowicie Systemu Państwowej Informacji Statystycznej /SPIS/.

Pomijam tutaj problem różnic w metodach projektowania i wdrażania pomiędzy poszczególnymi systemami rządowymi, jakie niewątpliwie mogą istnieć jako rezultat różnic w celach i zadaniach poszczególnych systemów rządowych i wywierających wpływ między innymi na treść i zawartość zasobów informacyjnych, algorytmów przetwarzania, zasady udostępniania informacji itp. Uważam, bowiem, że z punktu widzenia podstawowych zasad projektowania różnice te nie odgrywają istotnej roli i że korzyści uzyskane z połączenia wysiłków projektantów tych systemów nad rozwiązaniem wspólnych trudności i problemów będą większe aniżeli ewentualne trudności związane z odstępstwem od rozwiązań indywidualnych dostosowanych do specyfiki danego systemu rządowego.

Rozległość tematyczna i złożoność zasobów informacyjnych

Patrząc z tego punktu widzenia na system rządowy na przykładzie SPIS musimy stwierdzić niezwykle obszerny zasięg tematyczny oraz dużą złożoność struktur wewnętrznych w zasobach informacyjnych tego systemu. W zasięgu zainteresowań systemu SPIS znajdują się niemal wszystkie przejawy działalności gospodarczej człowieka oraz zjawiska społeczne i kulturalne, które podlegają mierzeniu i które można wyrazić za pomocą odpowiedniego systemu wskaźników. System SPIS gromadzi więc dane dotyczące ludności i jej poziomu życia, zatrudnienia, dechodów i spożycia, świadczeń społecznych, zasobów dóbr trwałego użytku u ludności, usług dla ludności, cen, wyników działalności produkcyjnej w działach i gałęziach gospodarki narodowej, zasobów majątkowych w gospodarce narodowej, dane dotyczące rozwoju nauki, szkolnictwa, kultury i sztuki,

ochrony zdrowia, turystyki, sportu i wypoczynku, ochrony środowiska naturalnego człowieka itp.

Informacje z wymienionych dziedzin gromadzone i udostępniane są w najróżniejszych podziałach: gałęziowych, branżowych, organizacyjnych i terytorialnych oraz najczęściej za szereg okresów czasowych, co stanowi niezbędny warunek analizy kształtowania się zjawisk w czasie.

W tych warunkach, z punktu widzenia projektowania i eksploatacji systemów informatycznych trudno mówić o SPIS jako jednym systemie informatycznym. Mamy tutaj do czynienia raczej ze zbiorem systemów EPD różniących się między sobą źródłami zasilania, algorytmami przetwarzania, wymogami pod względem dostępności, zasadami i okresem przechowywania itp. W wielu przypadkach różni się również użytkownicy poszczególnych systemów.

Jeśli przyjąć jako kryterium wyodrębnienia systemu przetwarzania danych odrębność źródeł informacyjnych, odrębność algorytmów przetwarzania oraz odrębność wymagań użytkowników, to aktualnie tylko w ośrodku elektronicznym GUS w Warszawie realizującym podstawowe zadania w zakresie przetwarzania danych SPIS w eksploatacji, znajduje się około 130 systemów, a w stadium projektowania około 70 nowych systemów. Ponadto kilkadziesiąt systemów przetwarza się na minikomputerach oraz kilkadziesiąt systemów realizuje się w ośrodkach terenowych na potrzeby organów władzy i administracji terenowej.

Stawia to przed projektantami systemu rządowego SPIS szereg zadań specyficznych, nie występujących z taką siłą przy projektowaniu innych systemów, zwłaszcza systemów obiektowych. Jednym z tych zadań jest zapewnienie spójności informacyjnej pomiędzy za-

sobami informacji przetwarzanej w ramach różnych podsystemów oraz stworzenie przesłanek organizacji wspólnej bazy danych.

W tradycyjnych systemach przetwarzania danych statystycznych opartych w dużej mierze na doświadczeniach przetwarzania przy użyciu maszyn licząco-analitycznych oraz uwzględniających możliwości eksploatacyjne komputerów II generacji zbiory gromadzone były na taśmach magnetycznych. System przetwarzania danych zakładający partiowe gromadzenie i przetwarzanie danych zorientowany był całkowicie na sekwencyjną organizację zapisu danych. Niezbędna w takim systemie była znaczna redundancja informacji wprowadzana najczęściej celowo dla ułatwienia przetwarzania oraz ułatwienia dostępu danych w różnych układach.

Inną cechą tego systemu przetwarzania jest pełna zależność programów użytkowych od organizacji zbiorów powodująca konieczność przepracowywania i powtórnej komplikacji programów w każdym przypadku kiedy wprowadza się jakiegokolwiek modyfikacje w zbiorze.

Sekwencyjna organizacja zbiorów właściwa dla przetwarzania taśmowego coraz to mniej odpowiada rosnącym wymaganiom użytkowników. Chodzi przy tym nie tylko o utrudniony dostęp do zbiorów gromadzonych sekwencyjnie, ale również o trudności łączenia zbiorów informacyjnych pochodzących z różnych badań i niemożliwość osiągnięcia w ten sposób pełniejszej integracji zasobów informacyjnych.

Wdrożenie do eksploatacji maszyn III generacji wyposażonych w urządzenia pamięci o dostępie bezpośrednim o dużej pojemności pozwała wprowadzić bardziej doskonałą organizację gromadzenia danych opartą o koncepcję bazy danych.

Baza danych w pojęciu potocznym oznacza zasób informacyjny dowolnego systemu informacyjnego stanowiący materiał wyjściowy /tworzywo/ na podstawie którego system wydaje informację wynikową użytkownikom.

W terminologii projektowania systemów informatycznych pojęcie bazy danych oznacza szczególne podejście do organizacji gromadzenia i korzystania z zasobów informacyjnych różniące się od tradycyjnej organizacji sekwencyjnej zbiorów następującymi dwoma elementami:

- 1/ logiczne powiązanie wzajemne zapisów /rekordów/ należących do różnych zbiorów oraz możliwość ich wspólnego wielokrotnego opracowania za pomocą oprogramowania bazy danych zapewniającego łatwy dostęp do poszczególnych zapisów niezależnie od ich fizycznego rozmieszczenia w bazie.
- 2/ zastosowanie systemu programów zwanych systemem zarządzania bazą danych /SZBD/ sterującego operacjami wprowadzania danych do bazy, wyszukiwania i aktualizacji danych oraz spełniającego funkcje pośredniczące pomiędzy bazą danych i programami użytkowymi /5/.

Zastosowanie w projektowaniu systemów informatycznych metody bazy danych odnosi się zarówno do systemów obiektowych jak i systemów rządowych. Specyfika wdrożenia tej koncepcji w SPIS polega na tym, że:

- 1/ system zarządzania bazą danych statystycznych musi zapewnić budowę nowych modeli oraz umożliwić prowadzenie analizy zjawisk metodami, które nie są znane lub zdefiniowane w momencie zakładania bazy:

- 2/ baza danych statystycznych jest bardzo obszerna oraz występuje potrzeba częstego dostępu do dużej liczby danych w celu uzyskiwania danych zbiorczych;
- 3/ specyficzną cechą danych statystycznych jest ich ciągłe gromadzenie w czasie w celu tworzenia szeregów czasowych. Mimo, że najczęściej korzysta się z danych za ostatni okres, to również dane za poprzednie okresy muszą być przechowywane w takiej formie aby dostęp do nich był możliwy;
- 4/ z bazy danych statystycznych mogą korzystać specjaliści z różnych dziedzin, a więc zarówno statystycy, jak i branżyści, planiści, ekonometrycy, informatycy. Dla wszystkich więc powinna być zrozumiała struktura i zawartość bazy danych /6/.

Budowa systemu informatycznego w oparciu o koncepcję bazy danych jest jedną z najbardziej rozpowszechnionych tendencji w informatyce lat 70.

Za podstawową zaletę bazy danych uważa się stworzenie niezależności danych od programów i umożliwienie różnym użytkownikom i różnym programom dostępu do tych samych zbiorów.

Również w systemach informacji statystycznej we wszystkich krajach idea ta rozwija się dynamicznie i przybiera coraz to bardziej realne kształty. Istnieje jednak jednocześnie szereg spraw wymagających od informatyków bardziej precyzyjnych rozwiązań i rozstrzygnięć. Należą do nich między innymi:

Problem zasięgu i organizacji baz danych

Uwzględniając rozmiary i stopień skomplikowania zasobów informacyjnych SPIS wydaje się, że byłoby nierealne oczekiwanie, że wszystkie dane statystyczne będą stamowić, przynajmniej w naj-

bliższym dziesięcioleciu, część składową jednej bazy danych. Taka baza była by praktycznie niemożliwa do utrzymania i eksploatacji. Bardziej realne jest wydaje się uznanie celowości istnienia wielu baz danych, a ich integracja powinna polegać na stosowaniu wspólnych rozwiązań metodologicznych dla wszystkich baz danych. Potrzebna jest nam więc logiczna, a nie fizyczna integracja danych, aby użytkownik wiedział jak informacja jest dostępna, gdzie jest ona przechowywana oraz aby miał on w razie potrzeby zapewniony dostęp do informacji.

Jeśli dopuszczamy istnienie wielu baz danych, to powstaje pytanie według jakich kryteriów powinno następować wydzielanie poszczególnych baz danych. Mogą tutaj istnieć różne podejścia: branżowe, organizacyjne, problemowe itp. Podjęte zostały już w tej sprawie pewne kierunkowe decyzje, które są realizowane w projektowaniu baz danych SPIS, tym niemniej sprawy te wymagają dalszych dyskusji i uzasadnień.

Problemem otwartym jest także, czy wszystkie dane gromadzone w SPIS /a odnosi się to również prawdopodobnie do innych systemów rządowych/ powinny być włączone do bazy danych i opracowywane zgodnie z regułami zarządzania bazą danych? Chodzi o to, że oprócz niewątpliwych zalet systemu opartego o bazę danych /integracja informacji, zmniejszenie refundacji, lepsza centralna kontrola informacji, łatwiejszy dostęp do danych/ ze stosowaniem tego systemu wiąże się również szereg trudności. Ocenia się, że systemy oparte o bazę danych bardziej obciążają sprzęt w porównaniu do tradycyjnych systemów sekwencyjnych; zajmują zwłaszcza znacznie więcej pamięci operacyjnej i pamięci dyskowej komputerów. W systemach opartych o bazę danych znacznie większe znaczenie ma także niezawodność sprzętu i oprogramowania. Wymagane są również

wyższe kwalifikacje personelu projektującego i eksploatującego system informatyczny /7/.

Wydaje się, że odpowiedzi na to pytanie można udzielić jedynie w wyniku analizy sposobu wykorzystania zbiorów gromadzonych w systemie pamiętając, że bardziej celowe jest zachowanie sekwencyjnej organizacji zbiorów w tych przypadkach, gdzie mamy do czynienia z badaniami jednorazowymi prowadzonymi na ściśle określone zapotrzebowanie użytkowników, których potrzeby informacyjne można zaspokoić za pomocą odpowiedniego programu publikacji lub zestawień komputerowych i w których nie ma konieczności częstego dostępu bezpośredniego do danych.

Sądzę również, że z formułowaniem bardziej konkretnych wniosków w sprawie włączenia poszczególnych zbiorów do bazy danych powinniśmy poczekać do czasu uzyskania konkretnych doświadczeń w eksploatacji bazy danych o ograniczonym zasięgu, w przeciwnym bowiem przypadku nasze argumenty w tej sprawie będą często emocjonalne.

Problem wyboru systemu zarządzania bazą danych

Wdrażając do SPIS metodę baz danych musimy dokonać wyboru SZBD. Nie jest to problem łatwy, bowiem jak wynika z analiz prowadzonych przez ekspertów urzędów statystycznych kilku krajów /6/ nie ma w tej dziedzinie ściśle określonych preferencji. Urzędy statystyczne wykorzystują zarówno dostępne na rynku uniwersalne SZBD, jak i opracowane we własnym zakresie przez zespoły programistów niektórych urzędów statystycznych.

Spośród systemów uniwersalnych najbardziej znany jest system TOTAL stosowany przez Biuro Spisów USA. Jest on jednak używany nie do celów statystycznych, lecz w systemie wewnętrznego zarzą-

dzania pracą Biura: obliczania płac, w księgowości, ewidencji kadrowej itp. Urząd Statystyczny Kanady również zakupił ten system w celu stopniowego zapoznawania pracowników urzędu z technologią bazy danych. Trudność we wdrażaniu systemu TOTAL polega na tym, że wymaga on stosowania w modelach danych dość złożonej struktury sieciowej, która nie jest łatwa do zrozumienia przez użytkowników informacji.

Instytut Statystyczny /INSEE/ Francji zastosował u siebie 2 uniwersalne systemy - system IMS i SOCRATE, lecz podobnie jak Biuro Spisów USA również nie do celów statystycznych. Kilka innych urzędów statystycznych również wdrożyło lub wdraża dostępne uniwersalne SZBD. Można tu wymienić zwłaszcza:

- Węgierski GUS wdrażający system MARK IV opracowany przez firmę Informatics Incorporated. System może opracowywać zbiory o strukturze drzewopodobnej o maksymalnej liczbie 9 poziomów. System ten stosowany jest w statystyce przemysłu.
- Urzędy statystyczne CSRS wprowadzają eksperymentalnie SZBD CDC MARS III. System ten pozwala przetwarzać zbiory o prostej strukturze, ale nie pozwala na bezpośrednie jednoczesne przetwarzanie kilku zbiorów.
- Norweski Urząd Statystyczny stosuje standardowy SZBD TDS dla prowadzenia rejestru przedsiębiorstw. System umożliwia aktualizację zbioru metodą przetwarzania transakcji - za pomocą monitorów ekranowych.
- Centralny Urząd Statystyczny W. Brytanii prowadzi prace wdrożeniowe systemu DMS 1100 na maszynie UNIVAC 1106.

Oprócz systemów uniwersalnych, opracowywanych przez producentów komputerów lub wyspecjalizowane firmy produkujące oprogramowanie kilka urzędów statystycznych opracowało własne SZBD.

- Urząd Statystyczny Kanady opracował w okresie 1968-70 2 specjalistyczne systemy zwane CANSIM i STATPAK. Oba systemy są nadal stosowane i ocena ich jest pozytywna.
System CANSIM zarządza bazą danych zawierających szeregi czasowe, przypominające nasz bank ROZWÓJ.
- Centrum Badawczo-Naukowe w Bratysławie opracowało system programowania znany pod nazwą ISIS. System obejmuje programy odczytu, kontroli i korekty danych wejściowych, zapisu danych w pamięci zewnętrznej lub w bazie danych, selekcji danych według zadanych kryteriów, wykonywania obliczeń oraz drukowania wyników w formie wymaganej przez użytkownika. Jedną z podstawowych cech tego systemu jest modułowość, która umożliwia przetwarzanie dwóch strumieni danych z zastosowaniem lub bez zastosowania modułu zarządzania bazą danych.
- Francuski Instytut Badań Statystycznych i Ekonomicznych /INSEE/ opracował kompleks pakietów programowych znanych pod nazwą LEDA i BIBLOS. Przeznaczone są one do wykonywania wybranych operacji przetwarzania danych statystycznych.
Urząd Statystyczny NRD opracował oprogramowanie bazy danych znane pod nazwą Gromadzenie Danych Statystycznych /Data Storage of Statistics/. Składa się ona z szeregu modułów, które mogą być wykorzystywane niezależnie od siebie.
- Urząd Statystyczny Szwecji opracował kilka systemów zarządzania bazą danych. Należą do nich m.in. system ARKDABA - system

dialogowy służący do opracowania zbiorów jednostkowych, system TSDB - system wyszukiwawczy dla analizy zbiorów danych zbiorczych. System ten jest podobny do kanadyjskiego systemu CANSIM, system RSDB - system dialogowy dla przetwarzania danych statystycznych i planowania regionalnego - możliwy do zastosowania zarówno dla zbiorów danych jednostkowych, jak i sumarycznych.

Jak wynika z przytoczonych wyżej danych w dziedzinie SZBD w statystyce brak jest dotąd jakiegoś bardziej uniwersalnego rozwiązania, które by mogło wykazać swoją przewagę nad innymi. Niektóre urzędy stosują systemy ogólnodostępne opracowane przez producentów sprzętu lub przez wyspecjalizowane firmy softwarowe bo nie są w stanie ponieść ogromnych kosztów opracowania własnego systemu oprogramowania oraz nie są w stanie zatrudnić odpowiedniej wysoko-kwalifikowanej kadry projektantów i programistów. W innych urzędach statystycznych dostępne na rynku systemy oprogramowania stosuje się jako rozwiązania eksperymentalne - głównie w celu zapoznania kadry statystyków z techniką bazy danych oraz w celu ukierunkowania prac badawczych w tym zakresie.

Wynika z tego, że mimo wielu wysiłków oraz znacznych kosztów poniesionych na stworzenie odpowiednich systemów oprogramowania bazy danych - systemy te są jeszcze w początkowym stadium rozwoju. Ogólnie panuje pogląd, że dostępne systemy opracowane przez producentów komputerów nie odpowiadają specyficznym wymogom opracowań statystycznych głównie z następujących powodów:

- 1/ Stosując uniwersalny system zarządzania bazą danych projektant bazy danych nie ma swobody wyboru struktury danych lecz musi przyjąć strukturę którą narzuca dany SZBD.
- 2/ Struktura danych przewidziana w systemach uniwersalnych jest zbyt złożona i zbyt odbiega od aktualnych potrzeb statystyki.

- 3/ Urzędy statystyczne mają do czynienia z bardzo dużymi rozmiarami bazy danych, zorganizowanymi w bardzo obszernych zbiorach. Zastosowanie uniwersalnych SZBD ze złożoną strukturą zapisu danych w pamięci okazuje się mało efektywne z punktu widzenia czasu przetwarzania.
- 4/ Istniejące systemy zarządzania bazą danych nie spełniają wszystkich funkcji wymaganych w statystyce.

Problem wyboru SZBD staje się również problemem pilnym dla SPIS. Już w tej chwili brak takiego systemu hamuje rozwój niektórych prac projektowo-wdrożeniowych, zwłaszcza w zakresie niektórych podsystemów o dużym znaczeniu integracyjnym w rodzaju: REGON, SŁOWNIK, czy banku danych ROZWÓJ.

Również prace nad organizacją banków danych opartych na zasadzie podmiotowej - takich jak bank danych o przedsiębiorstwach budowlano-montażowych, o przedsiębiorstwach przemysłowych, system informacji o rolnictwie itp. powinny być orientowane na użycie określonego SZBD.

Informatycy powinni w tej sprawie zająć stanowisko. Możliwych jest tutaj kilka rozwiązań. Zbadania i uzasadnienia wymaga zwłaszcza celowość:

- a/ nabycia jednego z dostępnych na światowym rynku softwar'owych systemów w rodzaju TOTAL, IDMS;
- b/ uzyskanie najbardziej odpowiedniego SZBD opracowanego i wdrożonego bodaj częściowo w jakimś urzędzie statystycznym za granicą;
- c/ podjęcie prób wdrożenia i ewentualnego zakupu opracowanych w naszym kraju SZBD w rodzaju systemu SEZAM na komputery Odra lub systemu RODAN na komputery jednolitego systemu;

d/ podjęcie własnych prac projektowych i programowych nad opracowaniem oryginalnego systemu uwzględniającego specyficzne warunki systemu rządowego SPIS i jego wyposażenie techniczne.

Od specjalistów powinniśmy otrzymać propozycje optymalne z punktu widzenia łatwości i kosztów wdrożenia oraz wygody eksploatacji.

Niezależnie od generalnej sprawy wyboru SZBD odpowiadającego wymogom SPIS istnieje potrzeba doskonalenia cząstkowych rozwiązań programowych, zwłaszcza opracowania standardowych programów naliczania i drukowania tablic, automatycznej kontroli i korekty danych, wprowadzania danych z różnych nośników itp.

Standaryzacja oprogramowania w warunkach systemu rządowego SPIS jest szczególnie niezbędna z uwagi na fakt, iż w systemie tym, w odróżnieniu od systemów obiektowych, znaczna część opracowań ulega systematycznym zmianom i trudności w operatywnej modyfikacji programów często uzbrudniają lub nawet uniemożliwiają spełnienie szeregu doraźnych wymagań użytkowników.

Należy oczekiwać, że część z tych trudności programowych można będzie rozwiązać przez wdrożenie SZBD. Tym niemniej, z uwagi na to, że jeszcze przez dłuższy czas szereg tematów opracowań realizowanych będzie poza systemem bazy danych, oraz że niektóre specyficzne moduły mogą być rozwiązane w bardziej racjonalny sposób przy pomocy własnych standardowych modułów programowych, oba te kierunki prac, a mianowicie prace wdrożeniowe w zakresie SZBD i w zakresie wdrożenia standardowych programów powinny być prowadzone równolegle.

Stworzenie systemu informacji o zasobach informacyjnych

W rozległych i tematycznie różnorodnych systemach informacyjnych skomplikowaną staje się nie tylko procedura uzyskiwania danych /co ma ułatwić cały system oprogramowania, łącznie z SZED/, lecz również informowanie użytkowników o tym, co konkretnie, za jaki okres, z jaką szczegółowością, w jakim układzie itp. zawarte jest w zasobach informacyjnych systemu.

W dotychczasowym systemie przetwarzania danych statystycznych bezpośrednimi użytkownikami skomputeryzowanego systemu informacyjnego byli wyłącznie pracownicy organów statystycznych /GUS i jego organów terenowych/. Problem informowania użytkowników o zasobach informacyjnych nie sprawiał trudności. Bezpośredni użytkownik systemu, będący jednocześnie zleceniodawcą, dysponentem i odbiorcą informacji sam dokładnie wiedział gdzie i jaka informacja się znajduje w zasobach informacyjnych komputera, jakie są metody badania, jak oceniać wiarygodność danych, w jakim zakresie i z jaką szczegółowością można poszczególne dane porównywać między sobą itp. Użytkownik ten jednocześnie nie odczuwał potrzeby bliższego zapoznania się z organizacją zapisu danych na nośnikach maszynowych, miejscem ich przechowywania itp., bowiem w przypadku potrzeby kontaktował się z odpowiednim pracownikiem ośrodka informacyjnego i składał odpowiednie zamówienie wykonania interesujących go zestawień, określając w sposób możliwie dokładny zakres swoich potrzeb i postać zestawień wynikowych, których wykonania oczekiwał. Pozostałe czynności związane z wyszukaniem odpowiedniego zbioru, wyszukaniem lub opracowaniem programu, dokonaniem obliczeń komputerowych, wydrukowaniem i dostarczeniem tablic wykonywane były przez odpowiednich pracowników ośrodka obliczeniowego.

Taki system współpracy, przeważający zresztą nadal obecnie, oprócz swoich niewątpliwych zalet, z których najważniejszą jest prostota i efektywność ma jednak szereg wad, które zaczynają się zwłaszcza ujawniać w konfrontacji z zwiększającymi się wymaganiami informacyjnymi oraz z rozszerzającymi się formami udostępniania informacji. System ten bowiem wychodzi z założenia, że końcowy użytkownik informacji /organy planowania i zarządzania, instytucje naukowe i inne/ zawsze kontraktują się z użytkownikami bezpośrednimi /statystykami/, którzy przekazują z kolei odpowiednie postulaty informatykom.

Po drugie system ten zakłada, że użytkownik bezpośredni szczególnie zna zawartość informacyjnych zasobów zgromadzonych w systemach skomputeryzowanych.

W miarę postępu komputeryzacji i rozwoju systemów informatycznych oba wymienione wyżej założenia są praktycznie trudne do spełnienia.

Z jednej strony użytkownicy bezpośredni /statystycy/ nie są w stanie pamiętać wszystkich szczegółów dotyczących zasobów informacyjnych, ich szczegółowości, zakresu, ujęć metodologicznych systemów cen, porównywalności z poprzednimi okresami itp.

Z drugiej strony rozwój i doskonalenie sprzętu stwarza użytkownikom możliwość bezpośredniego dostępu do skomputeryzowanych zbiorów informacji i sprawia, że użytkownik końcowy staje się często jednocześnie użytkownikiem bezpośrednim, który potrafi wprawdzie określić ogólnie jakie informacje wynikowe są mu potrzebne, nie jest jednak w stanie znać szczegółów co do treści i sposobu ujęcia źródłowych danych, na podstawie których mają powstać wymagane informacje wynikowe.

Powstaje w związku z tym nowy problem o istotnym znaczeniu, a mianowicie problem stworzenia odpowiedniego systemu informacji o zasobach informacyjnych gromadzonych w systemie. System taki nazywany jest często systemem informacji o informacji lub meta-informacją /8/.

System ten powinien obejmować:

- a/ informacje o podstawowych cechach zbiorów danych np. typ, komórke odpowiedzialną za utrzymanie zbioru, datę powstania zbioru, warunki aktualizacji, stopień poufności, stopień wiarygodności itp.,
- b/ informację o logicznej strukturze danych oraz o ich znaczeniu semantycznym,
- c/ informację o sposobie fizycznego przedstawienia danych w urządzeniach pamięci,
- d/ katalog standardowych procedur lub programów,
- e/ wewnętrzną statystykę na temat korzystania z bazy danych oraz z systemu zarządzania bazą danych.

Fizyczną formą przedstawienia wspomnianej wyżej meta informacji są tzw. katalogi i słowniki.

Szczegółowa treść, układ i sposób prowadzenia słowników i katalogów nie zostały dotychczas jednoznacznie sprecyzowane. W wielu krajach, w których prace te są prowadzone czynione są próby poszukiwania własnych rozwiązań odpowiadające potrzebom informacyjnym poszczególnych urzędów. Celem tych prac jest stworzenie skomputeryzowanych systemów informacji i ułatwienie korzystania z zasobów informacyjnych oraz ułatwienie i usprawnienie projektowania nowych badań w oparciu o pełniejsze rozeznanie wszystkich posiadanych zasobów informacyjnych.

System meta informacji rozpatrywany jest również jako ważne narzędzie zapewniające wewnętrzną spójność systemu informacji jak również spójność pomiędzy systemem informacji statystycznej i innymi centralnymi systemami informacyjnymi.

Prace nad katalogami i słownikami i ich komputeryzacją należą do jednych z najbardziej aktualnych kierunków prac nad doskonaleniem systemów informacji statystycznej. Z tego względu powinny być przedmiotem naszego szczególnego zainteresowania.

Jednocześnie jednak musimy mieć świadomość, że zaprojektowanie i wdrożenie systemu komputerowego prowadzenia słowników i katalogów oznacza w rzeczywistości zorganizowanie nowego systemu meta informacji opartego głównie o tzw. przetwarzanie tekstów /word processing/, które jak wykazuje niewielka jeszcze praktyka w tym zakresie jest z punktu widzenia użytkowego i z punktu widzenia oprogramowania nie mniej skomplikowane niż przetwarzanie danych liczbowych.

Ten aspekt metodologii projektowania systemów rządowych jako stosunkowo nowy wymagać będzie dalszych szczegółowych analiz. Nie chodzi przy tym o doczekanie czy systemy meta informacji są niezbędne, to bowiem wydaje się być oczywiste. Istnieje natomiast potrzeba opracowania dokładnej koncepcji takiego systemu, jego zaprojektowanie, wybranie odpowiedniego systemu oprogramowania i stopniowe wdrażanie.

Projektując system meta informacji trzeba zwrócić szczególną uwagę na to, że nie może on być, przynajmniej w pierwszym etapie zbyt szczegółowy i zbyt złożony. Już obecnie można się spotkać z zastrzeżeniami bezpośrednich użytkowników systemu SPIS, iż projektowane systemy są zbyt skomplikowane, trudne do zrozumienia, mało elastyczne i bardziej zorientowane na właściwości systemów

komputerowych niż na codzienne potrzeby użytkownika. Obawy te wyraził dobitnie Prezes Centralnego Urzędu Statystycznego Wielkiej Brytanii prof. C. Moser, stwierdzając, że widzi on realne niebezpieczeństwo, że statystycy zostaną zahipnotyzowani możliwościami informatyki i że pozwolą się wciągać w budowę coraz to bardziej złożonych systemów informacyjnych, podczas gdy najbardziej potrzebne są proste systemy oraz lepsze wykorzystanie aktualnie gromadzonych danych /9/.

Aby uniknąć tego niebezpieczeństwa, którego realność moim zdaniem istnieje również u nas konieczne jest pełniejsze zaangażowanie bezpośrednich użytkowników informacji w procesy projektowania systemów rządowych i zapewnienie im większego wpływu na rozwiązywanie podstawowych aspektów użytkowania tych systemów. Konieczne jest także unikanie za wszelką cenę skomplikowanych rozwiązań informatycznych we wszystkich tych przypadkach gdy podobne wyniki można uzyskać przy pomocy prostych metod.

Bibliografia

1. M. Lipiec: Seminarium SPIS-76. "Wiadomości Statystyczne, 1977 nr 1 str. 32-34.
2. Zbigniew Gackowski: Projektowanie systemów informacyjnych zarządzania, WNT, Warszawa 1974.
3. Teresa Możdżyńska: Podstawowe zasady projektowania systemów EPD, Wyd. GUS, ZMIAOS, zeszyt nr 6, Warszawa, 1971.
4. Aleksander J. Lerner: Zadania teorii sterowania wielkimi systemami, w zbiorze: Problemy sterowania wielkimi systemami, Zakład Narodowy imienia Ossolińskich, PAN, Wrocław - Warszawa - Kraków, 1964.
5. M.J. Turner: The impact of Data Bases on **Statistical Agencies**, CES/WP-9/130.

6. Report on the use and future need for Data Base Management in National Statistical Services, CES/WP.9/150.
7. George Schussel: When not to use a Data Base, Datamation Nov. 1975, str. 82-91.
8. Bo Sundgren: Meta-Information in Statistical Agencies, CES/SEM. 9/5.
9. Claus Moser: The environment in which Statistical Offices will work in ten years time, Seminar on Statistical Services in ten years time. Washington D.C. 21-25.III.1977, CES/Sem. 8/3.