

GŁÓWNY
URZĄD
STATYSTYCZNY

OŚRODEK BADAWCZO-ROZWOJOWY
SYSTEMU
PAŃSTWOWEJ INFORMACJI STATYSTYCZNEJ

1/80

SYSTEMY INFORMATYCZNE

seminarium

SPIS '79 BANKI DANYCH W CENTRALNYCH SYSTEMACH INFORMATYCZNYCH

część

I

Funkcje banku danych w Centralnym Systemie Informatycznym

Projektowanie banków danych — aspekt informacyjny

Warszawa

1980

1/80

**SYSTEMY
INFORMATYCZNE**

Materiały seminarium
SPIS '79

BANKI DANYCH W CENTRALNYCH
SYSTEMACH INFORMATYCZNYCH

Warszawa 1980



107599/1

**SYSTEMY
INFORMATYCZNE**

Redaktor naukowy: Józef Olefski

Redaktor: Andrzej Bratkowski

ORGANIZATORZY SEMINARIUM

**POLSKIE TOWARZYSTWO EKONOMICZNE
— ZARZĄD GŁÓWNY**
Komisja informatyki

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

RADA PROGRAMOWA SEMINARIUM SPIS'79

Przewodniczący: doc. dr hab. Jerzy Kisielnicki

Z-ca przewodniczącego dr Józef Olefiński

Członkowie: doc. dr Jerzy Eysymontt
dr Jan Iszkowski
mgr inż. Stanisław Jaskólski
mgr Jan Kwiatowski
mgr inż. Marek Lasota
dr Stefan Semczuk
dr Witold Staniszkis

**Sekretarz naukowy
seminarium:** mgr Henryk Dąbrowski

SPIS TREŚCI

Str.

Wystąpienie Prezesa GUS	XI
-----------------------------------	----

CZĘŚĆ I

FUNKCJE BANKU DANYCH W CENTRALNYM SYSTEMIE INFORMATYCZNYM

Tadeusz Walczak: Rola i funkcje banków danych w Centralnych Systemach Informatycznych	3
Mieczysław Adach: Bank danych a system informowania kierownictwa Centrali resortu przemysłu maszynowego	25
Andrzej Dąbkowski: Banki danych w modelowaniu uwarunkowań rozwoju społeczno-gospodarczego	39
Henryk Dąbrowski: Wybrane problemy Centralnego Banku Danych Statystycznych /na przykładzie Banku Danych Rozwój/	55
Ignacy Dziedziczak: Kooperacja informacyjna bazy danych księgowych z centralnymi bankami danych	79
Aleksandra Kurzyk, Adam Langer: Problemowe banki danych dla potrzeb administracji terenowej i kierownictwa Urzędu Wojewódzkiego w województwie gdańskim	86
Jacek Kuśnierz: Problemy projektowania banków danych wykorzystywanych w procesie dowodzenia	111
Teresa Moźdzysłowska: Miejsce i waga centralnych banków danych w SPIS	121
Bogusława Oskroba, Stanisław Paradysz: Bank danych statystycznych w świetle potrzeb użytkownika	132
Andrzej Tomasiak: Rola i zadania banku danych w rolniczym systemie informatycznym	146

Józef Wierzbołowski: Współpraca między systemami SPIS i GENPLAN na przykładzie podsystemu "Stosunki gospodarcze z zagranicą" realizowanego przez RCI Cenplan	163
Andrzej Zienkiewicz: Baza informacyjna Ministerstwa Nauki, Szkolnictwa Wyższego i Techniki oraz metody jej eksploatacji	191
PROJEKTOWANIE BANKÓW DANYCH - ASPEKT INFORMACYJNY	
Andrzej Bratkowski: Identyfikacja danych statystycznych w bankach danych dla potrzeb modelowania ekonometrycznego	207
Marek Cigáńnik: Systemy opisu i tezaursus	218
Jerzy Eysymontt: Język makrozarządzania jako podstawa projektowania bazy metadanych w centralnych systemach informatycznych	234
Tadeusz Jaegermann: Szczególne cechy banków danych w statystyce i wybrane metody ochrony danych w takich systemach	254
Zbigniew Jankowski, Stanisława Szwalek: Badania masowe jako źródło zasilania banków danych oraz baza do utworzenia operatu losowania prób do badań reprezentacyjnych	265
Jerzy Kisielnicki: Wybrane metody projektowania zawartości informacyjnej banków danych	278
Jan Kwiatowski: Bank danych oparty oparty o zasadę rozproszonej autoryzacji tematycznych baz danych jako narzędzie integracji informatycznego systemu planowania Centralnego	296
Peter Lisekovsky: Projektowanie banku danych w SIS	329
Edward Nawrocki: Wdózenie spójności językowej jako podstawowy warunek integracji informacyjnej w bankach danych	346
Józef Oleński: Projektowanie zintegrowanych baz danych w Centralnych Systemach Informatycznych	362

Dušan Soltes: Sformalizowany opis danych i tworzenie bazy metadanych ASIS	383
Vladimir Šima: Treść i metodyka sformalizowanego opisu wskaźników	397

CZĘŚĆ II

PROJEKTOWANIE BANKU DANYCH - ASPEKT TECHNOLOGICZNY

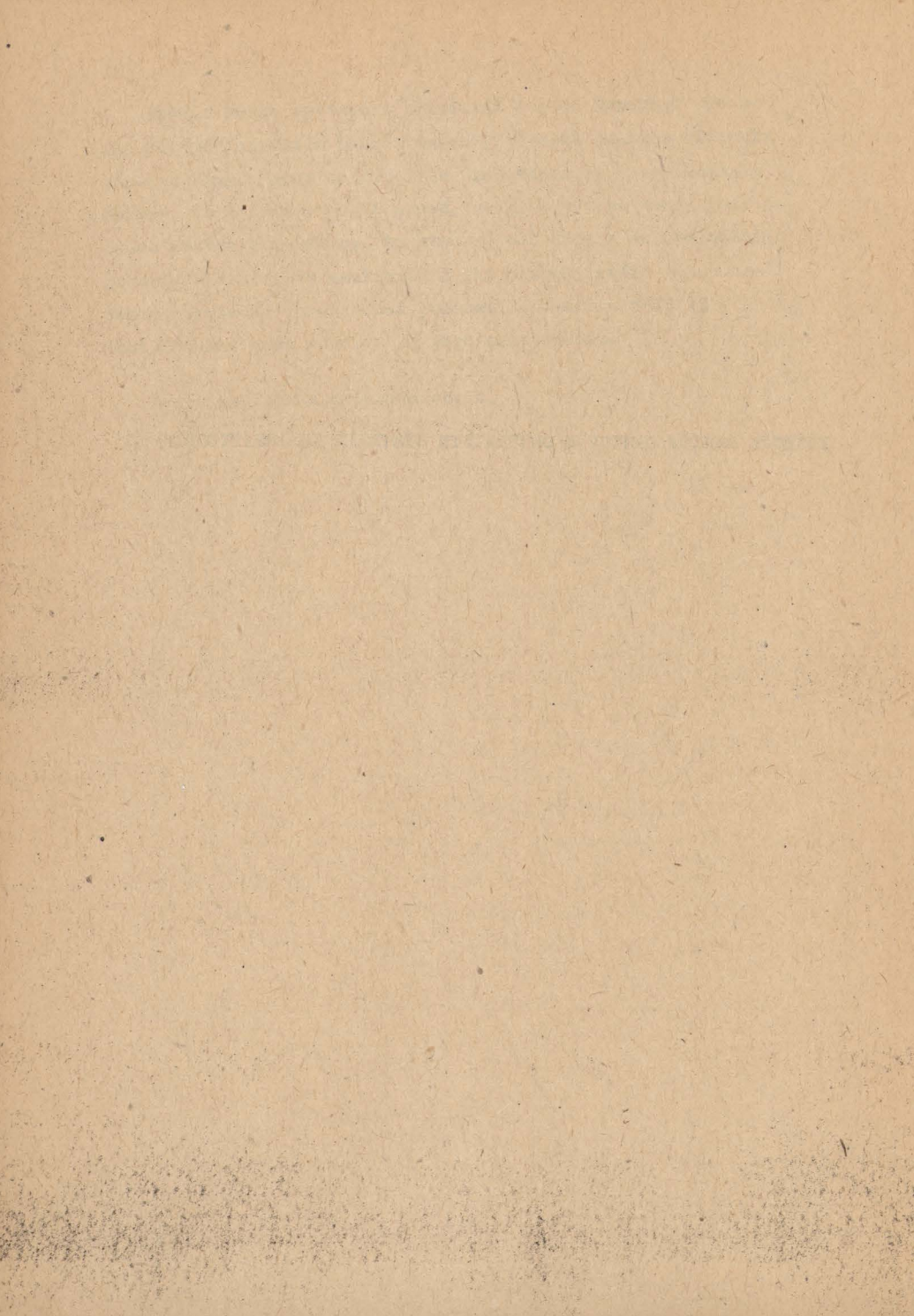
Bogumił Calica: Wybrane problemy struktury logicznej bazy danych	417
Marek Gaza: Projektowanie ciągów technologicznych z zastosowaniem standardowych elementów programowania	430
Witold Jamontt: Metainformacja a struktura bazy danych	442
Jerzy S.Korczak, Leszek A.Maciaszek: O redundancji w Systemach bazy danych	454
Andrzej Sokołowski: Zagadnienia ochrony informacji w bankach danych	469
Wanda Schiling: Problemy ochrony danych w systemach z bazami danych	481
Bolesław Szafranski, Stanisław Kozerański, Marek Niedzielski, Jan Pächolczyk: Integralność danych w bazie danych	497
Włodzimierz Zawadzki: System interaktywnej współpracy użytkowników z bazą danych	515

SYSTEMY ZARZĄDZANIA BAZAMI DANYCH DLA POTRZEB CENTRALNYCH SYSTEMÓW INFORMATYCZNYCH

Andrzej Brandt: Projektowanie fizycznej struktury danych dla bazy danych zarządzanej SZBD RODAN	527
Wiesław Dubczyński: Projektowanie procesów przetwarzania danych a metoda lokowania danych w bazie danych	542

Magdalena Dudziewicz, Jan Popiel: Mechanizmy wewnętrznego funkcjonowania sieciowej bazy danych	551
Tadeusz Gryc: Szczególne problemy i techniki ładowania bazy danych zarządzanej SZBD RODAN	573
Marek Lasota: Projektowanie struktury fizycznej bazy danych	594
Ewa Okuljar: Ładowanie i reorganizacja bazy danych w systemie TOTAL	613
J. Starosta, A. Macková: Problemy implementacji /teoretyczne i praktyczne/ Modułu Zbiorczego Wyszukiwania w SZBD SOFIS	627
Danuta Świątkowska-Królikowska: Projektowanie systemów przetwarzania przy użyciu SZBD TOTAL	654
PROBLEMY ORGANIZACYJNE, METODYCZNE I TECHNICZNE WDRAŻANIA I EKSPLOATACJI BANKÓW DANYCH W CSI	
Romuald Krzyżanowski: Wybrane problemy szkolenia użytkowników aktywnie współpracujących z Centralnymi Systemami Informatycznymi opartymi o bank danych . . .	667
Izabella Linkowska: Porównywalność danych w bankach podmiotowych	682
Teresa Mantorska: BA-DA-SEL /Bank Sumarycznych Danych Ludnościowych/	695
Bolesław Olechowski: Niektóre problemy metodologiczne i organizacyjne korzystania z banków podmiotowych SPIS	704
Danuta Skorupowska: Kryteria wyboru środków technicznych gromadzenia, przekazywania i udostępniania informacji w Centralnych Systemach Informatycznych . .	721
Wanda Skura: Problemy organizacyjne i koordynacyjne z oprogramowaniem bazy danych dla działu Budownictwo i Przemysł w systemie na III poziomie SPIS - na EMC Odra 1305	731

Agnieszka Wólczyńska: Pakiet wyszukiwania i wyprowadzenia informacji z banków danych Badabud i Badaprzem	745
METODYCZNE PROBLEMY PROJEKTOWANIA BANKÓW DANYCH W WARUNKACH OGRANICZONYCH ŚRODKÓW REALIZACJI /NA PRZYKŁADZIE WBD SPIS/	
Jan Chmielewski, Kazmierz Węglarz: Techniczna realizacja projektu regionalnej sieci teleinformatycznej	761
Jan Godula, Karol Mendrek: Projektowanie zdalnej aktualizacji i zmiennej algorytmizacji obliczeń w systemie banku danych	769
Jan Homa, Bolesław Sieroni: Organizacja wielowojewódzkich banków danych /z doświadczeń SPIS - WBD/ . . .	780
Irena Rotyńska, Lucjan Maśliński: Projektowanie banku danych w warunkach ograniczeń dostępnego sprzętu i oprogramowania	793
Bolesław Warzecha: Projektowanie banku danych w warunkach ograniczonych środków realizacji	804



ROLA I FUNKCJE BANKÓW DANYCH W CENTRALNYCH SYSTEMACH INFORMATYCZNYCH

Pojęcie i części składowe banku danych

Podstawowe idee banku danych sformułowane zostały pod koniec lat sześćdziesiątych i stanowiły wyraz poszukiwań nowych metod opracowania i udostępniania informacji, uwzględniających zarówno rosnące wymagania użytkowników informacji, jak i zwiększające się możliwości komputerów II i III generacji.

Mimo upływu około 10 lat od tego czasu do dziś nie udało się specjalistom osiągnąć jednoznaczności w interpretacji tego pojęcia. Z jednej strony, niektórzy informatycy skłonni są zawężać pojęcie banku danych do wyspecjalizowanego oprogramowania zapewniającego manipulowanie zbiorami informacji, wyszukiwanie i udostępnianie informacji. Z drugiej strony, użytkownicy informacji banku danych skłonni są nazywać bankiem dowolny zbiór informacji charakteryzujący określone zasoby lub zjawiska. Często można więc słyszeć doniesienia o utworzeniu banku krwi, banku części zamiennych, banku informacji o poszczególnych dyscyplinach sportowych, banku zbędnych maszyn, banku pomysłów itd. itp.

Nie wdając się w szczegółowe dyskusje terminologiczne, które w danym przypadku nie mają istotnego znaczenia dla rozwinięcia tematu postawionego w tytule referatu, pojęciem banku danych określać będziemy pewien system gromadzenia, przechowywania i udostępniania informacji wzajemnie z sobą powiąza-

nych i zaspokajających potrzeby wielu użytkowników /por. 8 str. 75/.

Ponieważ nazwany wyżej system musi być zawsze realizowany w określonych warunkach technicznych i organizacyjnych, można zgodzić się również z inną definicją banku danych, określającą tym pojęciem "jednostkę wyposażoną w komputer, magazynowane dane i programy maszynowego wprowadzania danych, wprowadzanie zmian, odzyskiwanie oraz - w pewnych okolicznościach - dalsze ich przetwarzanie" /9, str. 17/. Jeszcze bardziej szeroką definicję banku danych można znaleźć w przedmowie do rosyjskiego tłumaczenia pracy /6/, w myśl której pod pojęciem banku danych należy rozumieć system organizacyjno-techniczny składający się z baz danych, środków technicznych i programowych niezbędnych do tworzenia i eksploatacji tych baz oraz zespołu specjalistów zapewniających funkcjonowanie systemu.

Z powyższych definicji wynika, że w ramach banku danych możemy wyróżnić następujące ważniejsze części składowe:

1/ Zasoby informacyjne wyselekcjonowane i zapisane w pamięci komputera w taki sposób, aby z ich pomocą można było wszechstronnie charakteryzować badane zjawiska i procesy stosownie do potrzeb użytkowników informacji. Zasoby informacyjne utworzone i przystosowane do wymogów ich użytkowania w banku danych nazywamy bazą danych. Specyficzne cechy bazy danych, jako elementu banku danych, różniące ją od zbiorów informacji tradycyjnych systemów informacyjnych, polegają na logicznym powiązaniu wzajemnym zapisów /rekordów/ należących do różnych zbiorów oraz na takim zorganizowaniu ich zapisu w pamięci komputera, aby

można było zapewnić łatwe ich wyszukiwanie i opracowywanie za pomocą programów komputerowych, niezależnie od fizycznego rozmieszczenia poszczególnych zapisów w bazie.

- 2/ System programów komputerowych zapewniających wprowadzanie danych do bazy, wyszukiwanie i aktualizację danych, przekształcanie postaci /formatu/ danych w bazie stosownie do potrzeb poszczególnych programów użytkowych oraz ochronę danych przed zniszczeniem i niepowołanym dostępem. Taki system programów nazywamy Systemem Zarządzania Bazą Danych /SZBD/.
- 3/ Katalogi i słowniki zawierające nazwy danych źródłowych i wynikowych; nazwy i symbole układów grupowania; zbiór jednostek miar, przeliczników, spis typowych tablic wynikowych itp.; dokładny opis danych oraz ich powiązań; typowe procedury przetwarzania. Szczególne znaczenie posiada szczegółowy opis danych zawartych w bazie zwany schematem bazy danych /postać danych, opis poszczególnych pól, opis lokalizacji zapisów/. Z tego opisu korzystają wszystkie programy składające się na SZBD, dzięki czemu można uzyskać niezależność programów od postaci danych. Niezależność ta oznacza, że zmiany w układzie danych nie pociągają za sobą konieczności modyfikowania programów, co wywołuje tyle trudności i konfliktów przy przetwarzaniu tradycyjnym z użyciem indywidualnych programów komputerowych.
- 4/ Komputer i urządzenia pomocnicze niezbędne do bieżącej eksploatacji banku.

Zadania banku danych z punktu widzenia użytkownika informacji

Zarówno specjaliści - informatycy jak i użytkownicy informacji traktują system informacyjny oparty o metodę banku danych jako poważny krok naprzód w porównaniu do tradycyjnego systemu przetwarzania sekwencyjnego wykorzystującego indywidualne zbiory informacji nie zawsze dostatecznie powiązane z sobą tematycznie i stosującego system indywidualnych programów uzależnionych całkowicie od organizacji i sposobu zapisu poszczególnych zbiorów.

Jeśli chodzi o użytkowników informacji statystycznej, to ich nadzieje na poprawę systemu informowania dzięki zastosowaniu banków danych opierają się na następujących przesłankach:

1. Metody projektowania baz danych dla banków danych zapewniają zgromadzenie w pamięci komputera, z możliwością stosunkowo łatwego dostępu, określonego zasobu danych, za pomocą których można charakteryzować bardziej wszechstronnie badane zjawiska, niż to było możliwe za pomocą zbiorów danych przetwarzanych według metod tradycyjnych.

Przy przetwarzaniu tradycyjnym, wzorowanym w znacznym stopniu na technologii przetwarzania za pomocą stosowanych poprzednio przez wiele lat maszyn licząco-analitycznych, poszczególne zbiory informacji wprowadzane do komputera i zapisywane najczęściej na taśmach magnetycznych obejmowały z zasady ściśle określony, stosunkowo wąski wycinek działalności charakteryzowany za pomocą kategorii ekonomicznych ujętych w poszczególnych sprawozdaniach. Jeśli posłużyć się przykła-

dem z dziedziny przemysłu, to przy tradycyjnym przetwarzaniu w skład odrębnych zbiorów wchodziły, przykładowo, dane dotyczące:

- a/ sprzedaży wyrobów i usług, produkcji globalnej, przeciętnego zatrudnienia, funduszu płac,
- b/ produkcji dodanej, czystej i kosztów,
- c/ akumulacji, strat i zysków,
- d/ zapasów materiałów, wyrobów i przedmiotów nietrwałych,
- e/ środków trwałych,
- f/ stanu i wykorzystania obrabiarek,
- g/ nakładów finansowych na rozwój techniki i efektów postępu technicznego.

Taka organizacja zbiorów /zwana potocznie zbiorami formularzowymi/ odznacza się 2 podstawowymi mankamentami: po pierwsze prowadzi do pewnego nadmiaru informacji w związku z tym, że w każdym zbiorze trzeba umieścić pewne informacje powtarzające się, niezbędne m.in. do identyfikacji i porządkowania zbiorów, na przykład identyfikator przedsiębiorstwa, symbole przynależności organizacyjnej, branżowej, terytorialnej itp.; po drugie /i to wydaje się istotniejsze/ przy takiej organizacji zbiorów bardzo trudno uzyskać informacje wynikowe zawierające wzajemnie skorelowane dane pochodzące z różnych zbiorów. Typowa organizacja przetwarzania zbiorów formularzowych polega na odrębnym porządkowaniu i opracowywaniu każdego zbioru oraz sporządzaniu żądanych tablic wynikowych, a analizę związków między poszczególnymi cechami zawartymi w różnych zbiorach i w różnych tablicach pozostawia się, w tym przypadku, użytkownikowi.

Utworzenie wspólnej bazy danych w ramach prac nad bankiem danych, przez umieszczenie w jednej bazie wszystkich niezbędnych informacji jednostkowych, podnosi na zasadniczo wyższy poziom wartość poznawczą zasobów informacyjnych i pozwala uzyskać o wiele bogatsze i bardziej różnorodne informacje wynikowe umożliwiające prowadzenie znacznie głębszych i bardziej wszechstronnych analiz badanych zjawisk społeczno-gospodarczych.

Oceniając krytycznie tradycyjną organizację przetwarzania danych opartą o tematyczne zbiory formularzowe trzeba jednakże gwoili ścisłości stwierdzić, że tworzenie odrębnych zbiorów z danymi zawartymi na 1 określonym wzorze dokumentu źródłowego /formularza statystycznego/ i tak oznaczało ogromny postęp w integracji tematycznej zbiorów w stosunku do organizacji przetwarzania za pomocą maszyn licząco-analitycznych. Przy przetwarzaniu danych na maszynach licząco-analitycznych jako odrębne zbiory traktowane były informacje zawarte w każdym dziale sprawozdania, w związku z czym izolacja tematyczna pomiędzy poszczególnymi zbiorami informacji była o wiele większa.

Wprowadzenie komputerów do przetwarzania danych stanowiło pierwszy krok do integracji zbiorów, a prowadzone następnie eksperymenty wspólnego opracowywania informacji wynikowych na podstawie różnych zbiorów zawierających wzajemnie powiązane dane niejako w sposób naturalny prowadziły do wdrożenia koncepcji wspólnej bazy i banków danych.

2. Zasoby informacyjne gromadzone w organach statystycznych wielokrotnie przewyższają możliwości i celowość ich udo-

udostępniania w formie drukowanych i rozpowszechnianych szeroko publikacji. Dążenie do szerokiego publikowania wszystkich wyników badań statystycznych powoduje nadmierny wzrost kosztów prac wydawniczych oraz niedopuszczalny wzrost zużycia papieru i prowadzi ponadto do trudnego do kontroli zalewu informacjami użytkowników, którzy w tych warunkach muszą sami dokonywać selekcji informacji oraz powtórnego ich przetwarzania.

System informacyjny oparty o metody banku danych umożliwia, niezależnie od drukowania wymaganych informacji w formie tablic nadających się do bezpośredniego powielania, również udostępnianie wybranych informacji przez bezpośredni kontakt użytkownika z zasobami informacyjnymi banku za pośrednictwem urządzeń końcowych zainstalowanych u użytkownika i połączonych liniami transmisji danych z komputerem.

Bank danych pozwala więc realizować bardziej nowoczesne formy udostępniania informacji, które od dawna są przedmiotem szczególnego zainteresowania ze strony projektantów państwowych systemów informatycznych. Sprowadzają się one do szerokiego publikowania podstawowych informacji o rozwoju społeczno-gospodarczym, do udostępniania bardziej szczegółowych lub wybranych informacji grupie zainteresowanych odbiorców w formie opracowań /tablic, analiz, notatek/ wydawanych w niewielkiej liczbie egzemplarzy oraz do umożliwienia pewnej grupie podstawowych odbiorców korzystania na bieżąco z wszelkich dodatkowych informacji nie występujących w opracowaniach publikowanych.

3. Oprogramowanie banku danych umożliwia korzystanie nie tylko z danych gromadzonych w jego zasobach informacyjnych ale pozwala również wykonywać na bazie tych danych szereg dodatkowych obliczeń i analiz z wykorzystaniem metod matematycznych. Możliwości te zwiększają znacznie wartość poznawczą informacji i ich użyteczność jako instrumentu planowania i zarządzania.

Niektóre problemy związane z tworzeniem i funkcjonowaniem banków danych

Jak wspomniano wyżej koncepcja banku danych nie jest wynalazkiem nowym, mówi się o niej już od 10 lat. Również w założeniach Systemu Państwowej Informacji Statystycznej /SPIS/ formułowanych w latach 1971-1972 za podstawę gromadzenia i udostępniania informacji przyjęto metody banku danych /10, 12/. Mimo niewątpliwych zalet metod właściwych dla banków danych w porównaniu z tradycyjnymi systemami EPD banki danych nie stały się dominującą formą funkcjonowania systemów informacyjnych. Jeśli chodzi o SPIS można nawet powiedzieć więcej: nadal w tym systemie tradycyjne metody EPD przeważają. Niemnie jednak doświadczenia zdobyte w toku projektowania i użytkowania kilku pierwszych banków danych pozwalają dzisiaj w sposób bardziej konkretny i wszechstronny formułować propozycje i wnioski w tej sprawie.

Przede wszystkim należy podkreślić, że uzyskane doświadczenia potwierdzają w pełni główne tezy informatyków o tym, że system informacyjny oparty o metodę banku danych podnosi na znacznie wyższy poziom jakość obsługi informacyjnej użytkowni-

ków informacji i poważnie zwiększa użyteczność systemu jako instrumentu zarządzania i analiz ekonomicznych. Świadczą o tym między innymi doświadczenia kilkuletniej eksploatacji wojewódzkich banków danych /11/. Oznacza to, że w toku dalszych prac nad doskonaleniem rządowych systemów informatycznych problemom banków danych nadal musimy poświęcać nieśląbną uwagę.

Z tego względu, bardzo ważne jest zdanie sobie również sprawy z szeregu trudności i problemów jakie napotkano w trakcie tworzenia pierwszych banków danych, które nie pozwoliły uzyskać wyników w takim tempie jak zakładano - aby można było w:spólnymi siłami opracować sposoby przewyciężenia trudności i dokonania kolejnego kroku naprzód.

Trudności i problemy jakie napotkano w toku tworzenia pierwszych banków danych miały charakter zarówno informatyczny /sprzęt, oprogramowanie, transmisja danych/ jak i merytoryczno-informacyjny stanowiący przejaw słabości i mankamentów tradycyjnego systemu informacji statystycznej, które uwidoczniły się szczególnie jaskrawie w zetknięciu z nowoczesnymi metodami projektowania banków danych. Sądzę, że ta druga grupa problemów, rozwiązanie których zależy od zgodnej współpracy wielu specjalistów: statystyków, informatyków oraz zewnętrznych użytkowników informacji, zasługuje na szczególną uwagę.

1. Pierwszy problem wymagający rozstrzygnięcia na samym początku prac nad bankiem danych sprowadza się do określenia zawartości informacyjnej wspólnej bazy danych. Nie trzeba uzasadniać, że od zawartości bazy danych w decydującym stopniu zależy użyteczność informacyjna banku. Przyjęcie zbyt ubogiego za-

sobu bazy danych do banku może poważnie ograniczyć jego użyteczność informacyjną, natomiast umieszczenie w banku zbyt obszernych zbiorów może utrudnić lub wręcz uniemożliwić jego realizację na posiadanym sprzęcie.

Określenie zawartości informacyjnej sprowadza się do ustalenia:

- wykazu wskaźników /kategorii ekonomicznych/ podlegających włączeniu do bazy,
- stopnia ich agregacji /dane jednostkowe wg przedsiębiorstw lub zakładów, dane zbiorcze wg zjednoczeń, województw itp./,
- częstotliwości zbierania i aktualizacji /dane miesięczne, kwartalne, roczne/,
- okresu gromadzenia /za 1 okres sprawozdawczy, za 1 rok, szeregi wieloletnie/.

Ustalenie zawartości informacyjnej bazy danych przy projektowaniu banku nie jest sprawą prostą i jako żywo przypomina nierozwiązany dotąd problem określenia potrzeb informacyjnych użytkownika przy projektowaniu dowolnego systemu informacyjnego.

W wyniku szeregu dyskusji konfrontowanych z doświadczeniami użytkowania pierwszych banków danych przyjęto, że w SPIS bazy danych projektowane będą jako bazy hierarchiczne kilku-poziomowe /7/:

- Poziom 1 najobszerniejszy, zawierający wszystkie dane zbierane przez organy statystyczne i opracowywane przy użyciu komputerów. W ich skład wchodzi zarówno dane zbierane w formie sprawozdawczości jak i dane z wszelkich innych badań /spisów, badań ankietowych itp./. Zbiory poziomu 1 gro-

madzone są i przetwarzane z zastosowaniem głównie tradycyjnych zasad sekwencyjnego przetwarzania partiowego i służą za podstawę opracowania podstawowej masy tablic wynikowych zawierających informacje udostępniane przez GUS metodami tradycyjnymi. Zbiory poziomu 1 rozpatruje się także, jako podstawowe źródło zasilania baz danych w bankach danych, a więc jako potencjalny zasób informacji do automatycznego zasilania wyższych hierarchicznie baz danych.

- Poziom 2 powstający w wyniku przeniesienia i połączenie w jeden zintegrowany zbiór poszczególnych zbiorów informacji poziomu 1. Baza danych poziomu 2 zawierać będzie wszelkie niezbędne do wszechstronnej analizy dane zawarte w zbiorach poziomu 1, za okres 1 roku chociaż łączna objętość bazy poziomu 2 będzie znacznie mniejsza od sumy zbiorów poziomu 1, z uwagi na to, że w toku tworzenia bazy zintegrowanej nastąpi usunięcie redundancji informacji oraz wyeliminowanie danych nie mających istotnego znaczenia dla analizy zjawisk.

Poziom 2 tworzony będzie zgodnie z wymogami programów zarządzania bazą danych, które będą zastosowane przy tworzeniu i eksploatacji banku. Zakłada się, że dane wprowadzone do bazy poziomu 2 nie powinny w zasadzie zawierać błędów z uwagi na to, że były one sprawdzane szczegółowo w ramach obszernych programów kontroli automatycznej realizowanych przy tworzeniu zbiorów na poziomie 1. Dodatkowa kontrola jest jednak niezbędna w celu usunięcia ewentualnych niezgodności ujawnionych w toku łącznej kontroli pomiędzy cechami w ramach zintegrowanych zbiorów w bazie poziomu 2.

Podział zasobów informacyjnych na poziom 1 i 2 ma w pewnej mierze charakter przejściowy i ma swoje źródło wyłącznie w ograniczonych możliwościach sprzętu informatycznego i niedoskonałościach oprogramowania. W przyszłości podział ten powinien zanikać w ten sposób, że poszczególne zbiory tematyczne w miarę ich przenoszenia na maszynowe nośniki i oczyszczania z ewentualnych błędów powinny być od razu wprowadzane do bazy zintegrowanej /obecny poziom 2/ i służyć do opracowywania w systemie banku danych wszelkich niezbędnych informacji.

Takie rozwiązanie nie jest wyłącznie propozycją teoretyczną, zastosowano je bowiem kilka lat temu w katowickim banku danych i jest stosowane obecnie jako rozwiązanie typowe WED. Obejmuje ono jednak wyłącznie ograniczone pod względem rozmiarów zasoby informacji operatywnej. Szersze wnioskowanie tworzenia zintegrowanej bazy danych /poziom 2/ w aktualnych warunkach sprzętowych byłoby absolutnie nierealne. Były by bowiem niezbędne w tym celu komputery co najmniej 10 krotnie szybsze i posiadające co najmniej 20 krotnie większą pojemność pamięci o dostępie bezpośrednim od największych komputerów eksploatowanych obecnie w SPIS.

Wydaje się również obecnie wielce prawdopodobne, że również w dalszej przyszłości niektóre rodzaje badań wygodniej i bardziej racjonalnie będzie opracowywać według tradycyjnej technologii przetwarzania wsadowego /na poziomie 1/ aniżeli według technologii banku danych. Dotyczy to zwłaszcza niektórych badań obejmujących stosunkowo niewielki zakres cech ale bardzo dużą liczbę obiektów. Do takich badań należą zwłaszcza powszechne spisy ludności, spisy kadrowe, badania społeczne

prorowadzone metodami ankietowymi itp. W tych badaniach najprawdopodobniej indywidualna organizacja przetwarzania dostosowana do specyfiki każdego badania oraz indywidualne oprogramowanie zapewnią znacznie wyższą efektywność przetwarzania aniżeli uniwersalne oprogramowanie baz danych. Trzeba będzie natomiast zapewnić ścisłe powiązanie pomiędzy tymi zbiorami poziomu 1 oraz wspólną bazą danych. Powiązanie to powinno być dwukierunkowe:

- częściowo zagregowane wyniki badań jednorazowych mogą zasilać wspólną bazę danych dla wzbogacenia zasobów informacyjnych bazy;
 - dokładna znajomość zasobów informacyjnych może pozwolić na pełniejsze wykorzystanie posiadanych już informacji oraz na ograniczenie tematyki badań uzupełniających do niezbędnego minimum.
- Poziom 3 tworzony jest przez wyselekcjonowanie z baz poziomu 1 i 2 najistotniejszych kategorii /wskaźników/, które tworzyć będą zintegrowaną bazę z danymi za okresy wieloletnie umożliwiającymi prowadzenie analizy zjawisk za dłuższe okresy czasu. Bazę danych trzeciego poziomu posiada eksploatowany od paru lat bank danych o działalności przedsiębiorstw budowlano-montażowych "Badabud" /1/ oraz znajdujący się w fazie wdrażania bank o przemysłowych jednostkach gospodarki społecznej "Badaprzem". W obu wspomnianych bankach przewiduje się gromadzenie informacji za 7 lat - pięć lat ostatnich oraz dwa lata bazowe kończące dwa poprzednie okresy pięcioletnie. W "Badabudzie" dane gromadzi się od 1974 r. a w "Badaprzemie" - od 1975 r. Liczbę kategorii ekonomicz-

nych umieszczonych w bazie 3 ograniczają poważnie istniejące możliwości sprzętowe zwłaszcza w przypadku "Badaprzemu" gromadzącego informacje dla około 5000 jednostek. Z tego względu liczbę informacji w tym banku trzeba było ograniczyć do około 250 /w "Badabudzie" - o ponad 1400/. Mimo to eksploatacja takiego banku zajmuje około połowy zasobów największego komputera pracującego aktualnie w SPIS, ograniczając poważnie jego możliwości wykonywania innych prac w systemie wieloprogramowym.

- Poziom 4 stanowi baza danych banku "Rozwój" zawierającego syntetyczne dane /aktualnie około 9500 informacji/ charakteryzujące rozwój kraju za okres do 20 lat - od 1960 r. /3/. Z uwagi na to, że prace nad zaprojektowaniem i uruchomieniem tego banku zaczęły się znacznie wcześniej niż nad bazami poziomu 2 i 3 to zasilanie poziomu 4 odbywa się w sposób ręczny. Stanowi to jeden z istotnych mankamentów tego banku. Prowadzone aktualnie prace nad automatyzacją zasilania banku "Rozwój" bezpośrednio z baz poziomu 1 oraz 3 doprowadzą praktycznie do likwidacji tego 4 poziomu bazy danych i jego scalenie z poziomem 3.

2. Drugi problem który należało rozwiązać w początkowym etapie projektowania banków danych dotyczył decyzji czy do bazy danych należy wprowadzać dane podstawowe /kategorie pierwotne/ czy również dane pochodne /wtórne/ poddające się wyliczeniu według określonych algorytmów na podstawie danych pierwotnych. Początkowo przyjęto alternatywę pierwszą kierując się dążeniem oszczędności miejsca w pamięci komputera i wy-

chodząc z założenia, że przy obecnej prędkości obliczeniowej komputerów uzyskiwanie kategorii pochodnych w trakcie korzystania z banku nie powinno stanowić większych trudności. Życie jednak skorygowało ten pogląd. Okazało się, że często wykorzystywane wskaźniki pochodne oraz wskaźniki uzyskiwane w wyniku realizacji bardzo złożonych algorytmów powinny być obliczone zawczasu na etapie tworzenia bazy danych i przechowywane w bazie razem z kategoriami podstawowymi. Takie podejście powoduje wprawdzie pewne zwiększenie rozmiarów bazy danych, ale jednocześnie znacznie ułatwia korzystanie z banku oraz pozwala uprościć język użytkownika, w którym korzysta on z zasobów informacyjnych banku.

Jako przykład ilustrujący to twierdzenie można przytoczyć, iż w pierwszych wersjach wojewódzkich banków danych gromadzących informacje jednostkowe według przedsiębiorstw i zakładów brak było informacji zbiorczych dla całego województwa, projektanci wychodzili bowiem z założenia, że użytkownicy interesują się najczęściej właśnie danymi jednostkowymi, natomiast informacje zbiorcze można uzyskać przed sumowaniem odpowiednich cech z wszystkich jednostek wchodzących w skład banku.

3. Największe trudności z jakimi zetknięto się przy projektowaniu banków danych dotyczą zapewnienia spójności wewnętrznej informacji wprowadzanych do wspólnej bazy oraz utrzymanie tej bazy w warunkach porównywalności przez dłuższe okresy czasu.

Brak pełnej spójności pomiędzy informacjami nie powstał oczywiście przy projektowaniu banków danych. Istniał on zawsze i był krytykowany od dawna jako jedna z istotnych ujemnych cech istniejącego systemu informacji statystycznej /por. np. 5/.



Prace nad utworzeniem wspólnej bazy danych pozwoliły jedynie wykazać z całą bezwzględnością przypadki tej niespójności i skoncentrować uwagę statystyków na konieczność ich usunięcia.

Należy przy tym podkreślić, że nie wszystkie przypadki różnic pomiędzy ujęciami danych powodujących ich niespójność są możliwe do usunięcia jedynie przez doskonalenie samej metodologii badań statystycznych. Niektóre z nich mają swoje przyczyny poza statystyką, inne uzasadnia się dążeniem do maksymalnego uproszczenia obowiązków sprawozdawczych nakładanych na jednostki gospodarki społecznej, jeszcze inne powstają na skutek ustalenia zbyt krótkich terminów egzekwowania pewnej części sprawozdań. Jako przykład ilustrujący trudności tworzenia wspólnej bazy danych z powodu braku pełnej spójności pomiędzy danymi zawartymi w różnych zbiorach poziomu I można przytoczyć:

- Różnice zakresowe w ujęciach poszczególnych kategorii ekonomicznych w sprawozdawczości finansowej i rzeczowej. Przykładowo w niektórych jednostkach prowadzących zróżnicowaną gospodarczo działalność w NPSG wydziela się zadania dla poszczególnych działów gospodarki narodowej /przemysł, handel, transport/. Odpowiednio do tego zbiera się w sprawozdawczości dane obrazujące wyniki działalności według analogicznego układu rzeczowego. W planie finansowym natomiast, a w ślad za tym również w sprawozdawczości finansowej cała działalność tych jednostek ujmowana jest łącznie. W tym stanie rzeczy niektóre kategorie ekonomiczne /wartość środków trwałych koszty działalności, wynik finansowy itp./ nie dają się podzielić według takich samych kryteriów jak wyniki działalności rzeczowej, w związku z czym celowość ich wprowadzania do wspólnej bazy danych staje się problematyczna, natomiast

ich brak poważnie ogranicza możliwości łącznej analizy nakładów i wyników działalności tych jednostek. Podobny problem występuje w przypadku gospodarstw pomocniczych jednostek budżetowych, Gospodarstwa te /np. warsztaty szkolne, pomocnicze gospodarstwa rolne/ sporządzają sprawozdawczość rzeczową według uproszczonych zasad oraz nie sporządzają bieżącej sprawozdawczości finansowej. Sprawozdawczość finansowa obejmuje natomiast całokształt działalności jednostek budżetowych łącznie z ich gospodarstwami pomocniczymi.

- Różnice pomiędzy zakresem ujęcia danych zbiorczych w ramach bieżącej sprawozdawczości oraz w formie jednorazowych badań. Dotyczy to przykładowo danych okresowej sprawozdawczości ze stanu i ruchu zatrudnionych oraz danych zbieranych w formie spisów kadrowych obejmujących imienny wykaz pracowników wraz z odpowiednimi charakteryzującymi ich cechami. W celu maksymalnego odciążenia od tego pracochłonnego badania niektórych małych jednostek nie obejmowano ich obowiązkiem spisów kadrowych. Nie było to zresztą konieczne z punktu widzenia celów tego rodzaju badań. Przy tworzeniu wspólnych baz danych obejmujących informacje z obu źródeł badań podobne różnice w zakresie podmiotowym danych stwarzają szereg dodatkowych trudności.
- Różnice dotyczące momentu ujmowania pewnych zjawisk powodujące trudności wprowadzania do wspólnej bazy danych pochodzących z różnych źródeł. Przykładowo dane GUS o realizacji obrotów handlu zagranicznego dotyczą faktycznie zrealizowanych operacji eksportowych i importowych, natomiast dane bankowe o płatnościach obejmują faktyczne wpływy i płatności

z tytułu obrotów handlu zagranicznego. Podobnie dane dotyczące funduszu płac zbierane przez GUS dotyczą naliczonego funduszu płac w danym okresie, natomiast dane bankowe obejmują faktyczne wypłaty wykonane w danym okresie.

- Różnice klasyfikacyjne powodujące przykładowo trudności w łączeniu danych dotyczących produkcji wyrobów, dostaw na rynek i spożycia oraz nie dość konsekwentne stosowanie przyjętych klasyfikacji. Ilustracją braku konsekwencji w stosowaniu przyjętych klasyfikacji jest fakt, iż na przykład działalność transportu miejskiego /przewozy, zatrudnienie i płace, środki itp./ odnoszona jest zgodnie z obowiązującą w naszym kraju klasyfikacją gospodarki narodowej do działu "gospodarka komunalna", natomiast przy obliczeniach dochodu narodowego komunikację miejską zalicza się do działu "transport i łączność" - stosownie do praktyki stosowanej w większości krajów.

4. Doświadczenia uzyskane w toku projektowania banków danych a zwłaszcza w czasie praktycznego użytkowania pierwszych banków postawiły w nowym świetle problem ochrony danych przed świadomym zniszczeniem oraz przed nieupoważnionym ich wykorzystaniem. Przy tradycyjnym przetwarzaniu danych wszelkie informacje wynikowe uzyskiwane na podstawie zbiorów danych opracowywane są według indywidualnych programów komputerowych sporządzanych stosownie do dokładnie określonych wymagań użytkownika, którym jest dla ośrodka komputerowego właściwa komórka urzędu statystycznego. Wszystkię informacje wynikowe w formie tablic komputerowych przekazywane są również bezpośrednio temu

samemu użytkownikowi. W tych warunkach użytkownik informacji odpowiada za jej przekazanie wyłącznie upoważnionym do tego odbiorcom /końcowym użytkownikom/ i może w sposób bezpośredni ten system udostępniania kontrolować. Inaczej przedstawia się sprawa w warunkach funkcjonowania banku danych. U podstaw samej idei banku leży założenie znacznego ułatwienia dostępu do informacji, likwidacji trudności i biurokratycznych przeszkód jakie stoją na drodze w swobodnym przepływie informacji z systemu informacyjnego do użytkownika. Rzecz jednak w tym, że to ułatwienie i usuwanie przeszkód musi dotyczyć wyłącznie użytkowników upoważnionych, natomiast przed wszelkimi próbami nieupoważnionego uzyskania informacji system musi być skutecznie chroniony. Jest to szczególnie trudne do realizacji w tych wariantach banku danych które przewidują techniczne możliwości bezpośredniego korzystania z informacji przez użytkowników końcowych za pośrednictwem terminali połączonych z komputerem. Jest to problem obszerny i poruszany szeroko w literaturze /por. np. 2,4/. W tym miejscu pragnę jedynie podkreślić, że zwłaszcza w dużych systemach państwowych problemowi temu muszą poświęcić projektanci systemu znacznie więcej uwagi niż dotychczas.

Również problem ochrony danych przed uszkodzeniem i zniszczeniem zasługuje na zwiększoną uwagę przy oparciu systemu informacyjnego o metody banku danych. Nie znaczy to oczywiście, że jest to problem nowy. Przy komputerowym przetwarzaniu danych następuje ogromna, niespotykana dotąd koncentracja informacji na niewielkich pod względem objętości, nietrwałych i łatwych do uszkodzenia nośnikach magnetycznych. Uszkodzenie tych

nośników może spowodować ogromne, trudne do odtworzenia straty, a czasem nawet wręcz bezpowrotne zniszczenie informacji. Niebezpieczeństwo to istnieje zarówno przy tradycyjnej organizacji EPD jak i w warunkach banków danych. Niemniej jednak jest ono znacznie groźniejsze w warunkach banku danych. Przy przetwarzaniu tradycyjnym dokonuje się na ogół bardzo szczegółowych opracowań wynikowych, w rezultacie których podstawowy zakres informacji znajduje swój wyraz w obszernych tablicach przechowywanych poza ośrodkiem obliczeniowym. Możliwe jest dzięki temu odtworzenie bodaj części informacji w sytuacjach krytycznych. W systemie banku danych w wielu przypadkach zasoby informacyjne są nie do odtworzenia i straty spowodowane zniszczeniem danych mogłyby się okazać nieobliczalne.

Uwagi końcowe

Kilkuletni okres prac nad bankami danych w ramach SPIS przyniósł szereg cennych doświadczeń, które musimy w pełni wykorzystywać w dalszych pracach nad doskonaleniem tego systemu. Wydaje się że z doświadczeń tych wyciągnąć trzeba przede wszystkim dwa podstawowe wnioski:

po pierwsze

Prowadzenie prac nad tworzeniem banków danych zarówno na szczeblu centralnym, jak i terenowym zależy w decydującym stopniu od aktywnego zaangażowania się w tych pracach bezpośrednich użytkowników informacji tzn. pracowników organów statystycznych. Ilość trudnych problemów jakie napotykamy w toku projektowania i wdrażania banków jest znacznie większa po stronie informacyjno-metodologicznej, aniżeli po stronie informa-

tycznej. Dotyczą one głównie przygotowania spójnego zbioru informacji do bazy danych, opracowania metod utrzymania tego zbioru w układzie porównywalności w dłuższych okresach czasu mimo szeregu zmian w układach organizacyjnych, w metodach planowania i zarządzania, zmianach cen, zmianach klasyfikacji itp. oraz zapewnienia efektywnego nadzoru i kontroli bieżącej nad systemem aktualizacji zbiorów i udostępniania informacji. Doświadczenia ostatnich lat wykazały, że w najbliższym czasie nie ma co oczekiwać automatycznego doprowadzania bazy danych do porównywalności. Komputery mogłyby wprowadzić zadanie to wykonać, kto jednak opracuje algorytm?

po drugie

Technologia banków danych stawia znacznie większe wymagania wobec sprzętu komputerowego niż to przewidywano początkowo. Już w tej chwili jednoczesna eksploatacja wszystkich oprogramowanych banków danych jest niemożliwa z uwagi na ograniczone możliwości sprzętu komputerowego, a jest to przecież dopiero pierwszy krok w kierunku banków danych. Aby nie dopuścić w tych warunkach do zahamowania prac potrzebne jest nam zarówno poważne unowocześnienie sprzętu jak i bardziej nowoczesna koncepcja technologiczna uwzględniająca między innymi bardziej oszczędne gospodarowanie zasobami komputera.

Literatura

1. Antoni Bączkowski: Bank danych statystycznych o działalności uspołecznionych przedsiębiorstw budowlano-montażowych "Badabud", Wiadomości Statystyczne 1979 nr 1.
2. Antoni Bossowski: Problemy ochrony danych komputerowych, Wiadomości Statystyczne 1976 nr 12.

3. Henryk Dąbrowski: Bank danych statystycznych ROZWÓJ i kierunki budowy języka użytkownika, Informatyka 1979, nr 7.
4. Andrzej Z. Idzikiewicz: Ochrona informacji w procesie przetwarzania, Warszawa, PWE 1979.
5. Stanisław Kuziński: Rola statystyki w systemie planowania i zarządzania, Nowe Drogi 1972, nr 11.
6. Dż. Martin: Organizacja baz danych w wycisłitielnych sistemach, Wyd.Mir, Moskwa 1978
7. Bolesław Olechowski: Problemy budowy systemu informacji (jednostkach gospodarki uspołecznionej, Wiadomości Statystyczne 1977, nr 8.
8. Zygmunt Ryznar: Bank danych w przedsiębiorstwach przemysłowych, PWE Warszawa 1978.
9. System banków informacji w NRF, CINETE, Materiały Informacyjne MI/1973
10. Tadeusz Walczak: System SPIS, Telewizyjny Kurs Informatyki, Poradnik Zawodowy D/29, Warszawa, OBRI, 1975.
11. Bolesław Warzecha: System typowy - Wojewódzkie Banki Danych SPIS, Wiadomości Statystyczne, 1979 nr 6.
12. Założenia Systemu Państwowej Informacji Statystycznej, Wiadomości Statystyczne, 1973 nr 7, 8 i 9.