

Włodzimierz Trzebny

Ośrodek Informacji Naukowej PAN

Poznań

## BAZY DANYCH DLA BIOTECHNOLOGII

### WPROWADZENIE

Na rozwój biotechnologii coraz większy wpływ będzie miała informacja. Obecnie biotechnologia jest dziedziną wiedzy w dużej mierze uzależnioną od przemysłu, który z kolei potrzebuje szybkiej i wyczerpującej informacji. Oprócz uzyskania niezbędnych wiadomości na temat literatury naukowej, równie ważnym elementem jest informacja dotycząca rozwiązań technicznych i technologicznych, patentów oraz informacji na temat wszystkich zagadnień związanych z produkcją i handlem.

Szybkie uzyskanie pożądaných informacji jest obecnie możliwe, dzięki opracowywaniu licznych baz danych. Opublikowany w roku 1985 przez Stowarzyszenie Bibliotek Amerykańskich (American Library Association) oraz wydawnictwo Elsevier Science Publishers - katalog, zawiera charakterystykę 2805 baz danych. W lipcu 1987 r. został opublikowany katalog baz danych, wydany przez Elsevier Science Publishers, w którym przedstawiono 795 baz danych dotyczących nauk przyrodniczych (life sciences), biomedycyny, ochrony zdrowia i pokrewnych dziedzin nauki w tym 148 baz danych dostępnych w trybie konwersacyjnym.

Niezależnie od możliwości dostępu do baz danych w trybie konwersacyjnym (on-line), z wielu baz danych można korzystać w trybie wsadowym (off-line) lub przy zastosowaniu systemu

CD-ROM (compact disc - read only memory czyli płyta kompaktowa - pamięć tylko do czytania).

Bazy danych można podzielić na trzy rodzaje:

- 1) ukierunkowane na jedną dziedzinę wiedzy,
- 2) transdyscyplinarne, tworzone w aspekcie jednej dziedziny wiedzy,
- 3) wielodzielzinowe.

Kolejno bazy danych mogą występować jako dokumentacyjne (zwane bibliograficznymi) oraz faktograficzne (skierowujące).

W niniejszym artykule opisano bazy ukierunkowane i transdyscyplinarne, zarówno dokumentacyjne jak i faktograficzne.

#### I. Bazy danych ukierunkowane dotyczące biotechnologii:

##### A. Bazy danych dokumentacyjne

###### ABSTRACTS IN BIOCOMMERCE

Producent - Biocommerce Data Ltd., Slough, Wielka Brytania.

Zakres tematyczny danych w bazie - wszelkie doniesienia mające charakter handlowy oraz przemysłowy w biotechnologii, handlowe aspekty nauki.

Zasięg czasowy - od 1979 r. do chwili obecnej.

Przybliżona wielkość bazy - 50 000 rekordów.

Częstotliwość aktualizacji - dwa razy w miesiącu.

Średnia wielkość uzupełnień bazy w skali rocznej - 6500 rekordów.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie on-line.

Organizacje udostępniające informacje - DIALOG

INFORMATION SERVICES INC., Palo Alto, CA, USA; DATA - STAR,  
Radio-Suisse Ltd., Bern, Szwajcaria.

Publikacje wydawane w oparciu o zbiory bazy - Abstracts  
in Biocommerce - wersja drukowana, 2 razy w miesiącu.

Język dokumentów źródłowych - angielski.

## BIOBUSINESS

Producent - Bio-Sciences Information Service (BIOSIS)  
Philadelphia, PA, USA.

Baza danych jest wynikiem wieloletnich badań i współpracy  
dwóch firm: Bio Sciences Information Service (BIOSIS) oraz  
Information Access Company (IAC).

Zakres tematyczny danych w bazie - inżynieria genetyczna,  
bioprzetwarzanie, enzymy, technologia leków na drodze  
rekombinacji DNA, farmaceutyki, badania, produkcja i  
testowanie nowych leków; technologia lekarska i  
oprzyrządowanie; rolnictwo: ekonomiczne implikacje produkcji  
zwierzęcej i roślinnej, zwalczanie szkodników; energia i  
ochrona środowiska, oczyszczanie ścieków, zanieczyszczenia,  
zarządzanie zasobami; żywność i napoje: przemysłowa  
mikrobiologia, technologia żywności, przetwórstwo i  
magazynowanie.

W bazie danych dokumentowanych jest przeszło 1100  
czasopism dotyczących nauk przyrodniczych oraz wszystkich  
dziedzin zarządzania przemysłem: artykuły, komunikaty  
badawcze, nowe produkty, artykuły przeglądowe, patenty i  
sprawozdania z kongresów i konferencji.

Zasięg czasowy - od 1985 r., do chwili obecnej.

Wielkość bazy - 30 000 rekordów (1985).

Częstotliwość aktualizacji - co miesiąc.

Średnia wielkość uzupełnień bazy w skali rocznej - 30 000 rekordów.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie on-line.

Typ rekordu - bibliograficzny.

Elementy rekordu - numer akcesyjny, oryginalny tytuł dokumentu w języku angielskim, autorzy, pełny tytuł czasopisma (publikacji), dane dotyczące źródła, język publikacji, kod pojęciowy (concept codes) - pięciocyfrowy kod indeksowy stosowany do szerokiej identyfikacji zakresu przedmiotu. Poza tym w około 65% rekordach podaje się dodatkowe informacje - streszczenia i deskryptory.

Organizacje udostępniające informacje - DIALOG INFORMATION SERVICES INC., Palo Alto, CA, USA; DATA-STAR, Radio Suisse Ltd., Bern, Szwajcaria; MEAD DATA CENTRAL, Dayton, OH, USA.

Materiały instruktażowe - Bio Business Search Guide.

## BIOTECHNOLOGY ABSTRACTS

Producent - Derwent Publication Ltd., Wielka Brytania.

Zakres tematyczny danych w bazie - wszystkie dziedziny biotechnologii począwszy od manipulacji genetycznej poprzez inżynierię i fermentację biochemiczną do procesów produkcyjnych, przemysłowego zastosowania mikroorganizmów, enzymów, kultur komórkowych, hodowli roślin, hybrydyzacji komórek i usuwania ścieków przemysłowych.

W bazie danych dokumentowanych jest 1000 najważniejszych w świecie czasopism naukowych i technicznych, sprawozdań z konferencji i raportów z zebrań, ponadto patenty publikowane przez 15 największych urzędów patentowych.

Zasięg czasowy - od 1982 r. do chwili obecnej.

Wielkość bazy - 55 000 rekordów (styczeń 1987).

Częstotliwość aktualizacji - co miesiąc.

Średnia wielkość uzupełnień bazy w skali rocznej - 13 000 rekordów.

Nośnik - od stycznia 1987 r. taśmy magnetyczne, przysposobalnie od 1988 r. w systemie CD-ROM.

Możliwość wyszukiwania informacji - w trybie on-line.

Typ rekordu - bibliograficzny.

Elementy rekordu - numer akcesyjny, oryginalny, tytuł dokumentu w języku angielskim, język publikacji, autorzy i ich adresy, dane dotyczące źródła, ewentualnie dane patentu, słowa kluczowe, streszczenia (do 200 słów).

Organizacja udostępniająca informacje - SDC - ORBIT Search Service, CA, USA.

Materiały instruktażowe - On-line Users Manual; Thesaurus.

Publikacje wydawane w oparciu o zbiory bazy - Biotechnology Abstracts - wersja drukowana, miesięcznik od roku 1982.

#### CAMBRIDGE SCIENTIFIC ABSTRACTS LIFE SCIENCES COLLECTION

Producent - Cambridge Scientific Abstracts Life Sciences Collection.

Zakres tematyczny - biochemia, aminokwasy, peptydy i białka, membrany biologiczne, kwasy nukleinowe, biotechnologie, tkanki zwapniałe, chemorecepcja, ekologia, genetyka, immunologia, mikologia, przemysłowa i stosowana mikrobiologia, toksykologia, wirusologia. W bazie danych dokumentów jest ponad 500000 reaktywnych i jądrowych czasopiśmie, książek, monografiach, sprawozdaniach z konferencji, międzynarodowych patentów i danych statystycznych.

Zasięg czasowy - od 1982 r. do chwili obecnej.

Wielkość bazy - ponad 500 000 rekordów (1986 r.).

Częstotliwość aktualizacji - kwartalnie.

Średnia wielkość uzupełnień bazy w skali rocznej - ponad 98 000 rekordów.

Nośnik - CD-ROM. Możliwość wyszukiwania informacji - w systemie CD-ROM.

Typ rekordu - bibliograficzny.

Elementy rekordu - pełne dane bibliograficzne i streszczenie.

Organizacja udostępniająca informacje - Microinfo Ltd., Alton, Wielka Brytania.

Publikacje wydawane w oparciu o zbiory bazy - wersja drukowana 18 czasopism wydawanych przez Cambridge Scientific Abstracts.

#### CURRENT BIOTECHNOLOGY ABSTRACTS

Producent - The Royal Society of Chemistry, The University Nottingham, Wielka Brytania.

Zakres tematyczny - manipulacja genetyczna, przeciwciała monoklonalne, enzymologia, białka pojedynczej komórki, technika fermentacji, farmaceutyki, wytwarzanie energii, rolnictwo, przemysł chemiczny, żywność, nowości z przemysłu, problemy prawne, nowości dotyczące bezpieczeństwa.

Dokumenty dla bazy są wybierane z ponad 200 czasopism źródłowych, patentów i książek.

Zasięg czasowy - od 1983 r. do chwili obecnej.

Wielkość bazy - ponad 11 000 rekordów (1985).

Częstotliwość aktualizacji - co miesiąc.

Średnia wielkość uzupełnień bazy w skali rocznej - 5000

rekordów.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie on-line.

Typ rekordu - bibliograficzny.

Elementy rekordu - numer akcesyjny, tytuł dokumentu, autorzy i ich adresy, pełny tytuł dokumentu, dane dotyczące źródła, język publikacji, deskryptory, streszczenie.

Organizacja udostępniająca informacje - PERGAMON INFOLINE Ltd., London, Wielka Brytania.

RSWT. (Fizyko-chemiczna biologia i biotechnologia)

Producent - Vsechsojuznyj Institut Naucznoej i Techniczeskoj Informacji, Moskwa, ZSRR.

Zakres tematyczny, chemia bioorganiczna, biologia molekularna, biofizyka, biochemia, immunologia molekularna i komórkowa, biotechnologia, ogólne problemy fizyko-chemicznej biologii i biotechnologii.

Dokumenty do bazy są wybierane z artykułów z czasopism (89,1%), artykułów z prac zbiorowych (9,1%), patentów (1,0%), książek (0,5%), zdeponowanych rękopisów (0,3%), napływających z 130 krajów świata w 66 językach.

Zasięg czasowy - od 1982 r. do chwili obecnej.

Wielkość bazy - 115 000 rekordów (V.1985 r.).

Częstotliwość aktualizacji - co miesiąc.

Średnia wielkość uzupełnień bazy w skali rocznej - ok. 40 000 rekordów.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie on-line.

Typ rekordu - typ BK (pełne opisy bibliograficzne + słowa kluczowe).

Organizacja udostępniająca informacje - Vsechsojuznyj  
Institut Naucznoy i Technicheskoy Informacji, Moskwa, ZSRR.

Publikacje wydane w oparciu o zbiory bazy - RSWT  
Fizyko-chemiczna biologia i biotechnologia, wersja  
drukowana.

Język bazy - podwójny alfabet (cyrylica i łaciński),  
opisy dokumentów i słowa kluczowe podane w języku rosyjskim,  
tytuły dokumentów w językach angielskim, francuskim i  
niemieckim występują w wersji oryginalnej, a pozostałe  
tłumaczone są na język rosyjski.

Materiały instruktażowe - poradnik dla użytkownika, wykaz  
komend, zalecenia do formułowania i korekty profili, wykaz  
słów kluczowych, słowniki deskryptorów, rubrykatory.

#### TELEGEN

Producent - Environment Information Center Inc, New York,  
N.Y., USA.

Zakres tematyczny - inżynieria genetyczna i  
biotechnologia. Dokumenty pochodzą z około 7000 źródeł z  
całego świata z naukowych i technicznych czasopism,  
sprawozdań z kongresów i konferencji, raportów towarzystw  
handlowych, rządowych oraz niezależnych laboratoriów.

Zasięg czasowy - od 1973 r. do chwili obecnej.

Wielkość bazy - 8 000 rekordów (1985).

Częstotliwość aktualizacji - co miesiąc.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie on-line.

Typ rekordu - bibliograficzny.

Organizacja udostępniająca informacje - DIALOG  
INFORMATION SERVICES Inc., Palo Alto, CA, USA.



Publikacje wydawane w oparciu o zbiory bazy - Telegen Reporter, wersja drukowana.

## B. Bazy danych faktograficzne

Poniżej zestawiono najważniejsze ukierunkowane, faktograficzne bazy danych.

### AIDS

Baza danych produkowana przez Bureau of Hygiene and Tropical Diseases, London, Wielka Brytania. Baza dostępna w trybie on-line przez organizację DATA-STAR, Radio Suisse Ltd., Bern, Szwajcaria.

### ATCC

Baza produkowana przez American Type Culture Collection, Rockville, MA, USA.

Zakres tematyczny - opisy kultur komórek, pierwotniaków, glonów i grzybów.

Dostęp do bazy przez sieć CODATA (Committee on Data for Science and Technology).

### CFISM

Francuska baza danych dotycząca szczepów mikroorganizmów, produkowana przez INRA, Institut National Agronomique, Paris, Francja.

## ENZIDEX

Baza produkowana przez Biocatalysts Ltd., Pontypridd.,  
Wielka Brytania.

Zakres tematyczny - enzymy.  
25 000 szczepów grzybów,  
1000 szczepów drożdży, 1000 szczepów Actinomyces. Dostęp

## GENBANK

Producent - National Institutes of Health, Bethesda, MD.,  
USA.

Zakres tematyczny - sekwencje DNA.  
Biology Laboratory,  
Heidelberg, RFN.

## Hybridoma Databank

Producent - Universite de Nice / CODATA, Nice, Francja.

## In Vitro Conservation Database

Producent - International Board for Plant Genetic  
Resources, Roma, Italia.

Zakres tematyczny - baza danych technik kultur tkankowych  
roślin.

## MICIS

Producent - Laboratory of the Government Chemist, London,  
Wielka Brytania.

Zakres tematyczny - baza dotyczy serwisu informacyjnego  
mikroorganizmów.

## MIRDAB

Producent - Elsevier Science Publishers, Amsterdam,  
Holandia.

Zakres tematyczny - zasoby mikrobiologiczne.

NCC

Producent - Centraalbureau voor Schimmel-Cultures, Baarn, Holandia.

Zakres tematyczny - kolekcja 26 000 szczepów grzybów, 3500 szczepów drożdży, 1000 szczepów Actinomycetes. Dostęp do bazy w trybie on-line.

Nucleic Acid Data Library

Producent - European Molecular Biology Laboratory, Heidelberg, RFN.

Zakres tematyczny - dane dotyczące kwasów nukleinowych.

PG - Trans

Producent - Computer Science Unit, Institut Pasteur, Paris, Francja.

Zakres tematyczny - bank danych sekwencji białek.

PHIND (Pharmaceutical and Healthcare News Database)

Producent - Datastar Marketing, London, Wielka Brytania.

Zakres tematyczny - zawiera pełne teksty czasopism SCRIP (przemysł farmaceutyczny), CLINICS (diagnostyka), ANIMAL PHARM (zdrowie zwierząt) oraz AGROW (środki chemiczne stosowane w rolnictwie). Baza dostępna w trybie on-line.

SWISS - PROT

Producent - Department de Biochimie Medical, Universite de Geneve, Geneve, Szwajcaria.

Zakres tematyczny - dane dotyczące sekwencji białek.

## II. Bazy danych transdyscyplinarne

### Bazy danych dokumentacyjne

#### AGRIS

Producent - Food and Agriculture Organisation of the United Nation (FAO), Roma, Italia.

Zakres tematyczny - rolnictwo i gospodarka żywnościowa.

Dokumentowane są artykuły z czasopism książki, materiały konferencyjne, patenty, normy, dysertacje, sprawozdania i inne dokumenty rządowe.

Zasięg czasowy - od 1975 r. do chwili obecnej.

Przybliżona wielkość bazy - 1 120 000 rekordów (XII 1984)

Częstotliwość aktualizacji - co miesiąc.

Średnia wielkość uzupełnień bazy w skali rocznej - 120 000 rekordów.

Nośnik - taśmy magnetyczne. Możliwość wyszukiwania informacji - w trybie wsadowym (off-line) oraz w trybie on-line.

Typ rekordu - bibliograficzny.

Organizacja udostępniająca informacje - DIALOG INFORMATION SERVICES Inc., Palo Alto, CA. USA.

#### BIOSIS PREVIEWES

Producent - Bio-Sciences Information Service (BIOSIS) Philadelphia, PA, USA.

Zakres tematyczny - dane w bazie obejmują wszystkie dziedziny nauk biologicznych i badań medycznych. Dokumentowanych jest ponad 9000 czasopism oraz wydawnictw zwartych, książek i monografii, patentów, komunikatów z badań, sprawozdań z konferencji, kongresów i sympozjów.

Zasięg czasowy - od 1969 r. do chwili obecnej.

Wielkość bazy - ponad 5 000 000 rekordów (1986)

Częstotliwość aktualizacji - 2 razy w miesiącu.

Średnia wielkość uzupełnień bazy w skali roku - ok. 250 000 rekordów.

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie wsadowym (off-line) oraz w trybie on-line.

Typ rekordu - bibliograficzny.

Elementy rekordu - pięciodzaniowy unikalny identyfikator dokumentu (Coden), pełny tytuł dokumentu (i skrót tytułu), pełny opis bibliograficzny, deskryptory, pięciodzaniowe kody CROSS, kody taksonomiczne, adres pierwszego autora z listy. Każdy element rekordu może być wykorzystany jako element wyszukiwawczy.

Organizacje udostępniające informacje - DIALOG INFORMATION SERVICES Inc., Palo Alto, CA, USA; DATA-STAR, Radio Suisse Ltd., Bern, Szwajcaria.

Materiały instruktażowe - "BIOSIS Search Book", "Master Index", "Concept Code", "Biosystematic Code".

Publikacje wydawane w oparciu o zbiory danych - "Biological Abstracts" oraz "Biological Abstracts / RRM" (Reportes, Reviews, Meetings) w wersji drukowanej.

#### CA SEARCH

Producent - Chemical Abstracts Service, Columbus, OH., USA.

Zakres tematyczny danych w bazie - obejmuje wszystkie dziedziny biochemii, chemii nieorganicznej, organicznej i fizycznej, chemii stosowanej, chemii wielkocząsteczkowej,

inżynierii chemicznej. *on Service - materiały informacyjne.*

Dokumentowane są artykuły z około 14 000 czasopism naukowych i technicznych, patenty, raporty techniczne, dysertacje, sprawozdania z konferencji, kongresów i sympozjów, książek.

Zakres czasowy - od 1967 r. do chwili obecnej.

Wielkość bazy - ponad 6 600 000 rekordów (XII.1984).

Częstotliwość aktualizacji - 2 razy w miesiącu. *Books and*

Średnia wielkość uzupełnień bazy w skali roku - 460 000 rekordów. *Information Network Europe - materiały*

Nośnik - taśmy magnetyczne.

Możliwość wyszukiwania informacji - w trybie wsadowym (off-line) oraz w trybie on-line. *Informacyjne*

Typ rekordu - bibliograficzny. *8. 11.*

Elementy rekordu - pełny opis bibliograficzny, dokumenty źródłowe są podane w ponad 50 językach, opisy dokumentów w języku angielskim. Ponadto podana jest charakterystyka treści w postaci słów kluczowych. Rekord zawiera numer sekcji i podsekcji oraz numer abstraktu publikowanego w wersji drukowanej Chemical Abstracts.

Organizacja udostępniająca informacje - DIALOG INFORMATION SERVICES Inc., Palo Alto, CA. USA.

Publikacje wydawane w oparciu o zbiory danych - "Chemical Abstracts" w wersji drukowanej 2 razy w miesiącu.

Literatura:

Artowicz E., Zagadnienia Informacji Naukowej 1986, nr 2, s. 137.

ASSISTENT - Bazy danych WINITI na taśmach magnetycznych, Moskwa 1985.

BioCommerce Data Ltd., - materiały informacyjne.

Biosciences Information Service - materiały informacyjne.

Chemical Abstracts Service - materiały informacyjne.

Derwent Publications Ltd., - materiały informacyjne.

DIALOG - Database Catalog 1985.

EBIP News, nr 11, September 1986.

EBIP News, nr 13, March 1987.

Elsevier Science Publishers, New and Forthcoming Books and Journals, July 1987.

Microbial Information Network Europe - materiały informacyjne.

Microinfo Ltd., - materiały informacyjne.

Pergamon - Infoline - materiały informacyjne.

Popowska H., APID 1987, nr 1, s. 11.

Wakeford R., Trends in Biotechnology 1987, vol. 5, s. 120.