



IAEA



FAO



UNESCO



UNIDO

**INTERNATIONAL SYMPOSIUM ON THE FUTURE OF SCIENTIFIC,
TECHNOLOGICAL AND INDUSTRIAL INFORMATION SERVICES**

Leningrad, USSR, 28-31 May 1990

IAEA-SM-317/ 60

BIOTECHNOLOGY INFORMATION SERVICE OF THE GDR

Eleonore Poetzsch

**Scientific Information Centre of the Academy of
Sciences of the GDR, Berlin**

This is a preprint of a paper intended for presentation at a scientific meeting. Because of the provisional nature of its content and since changes of substance or detail may have to be made before publication, the preprint is made available on the understanding that it will not be cited in the literature or in any way be reproduced in its present form. The views expressed and the statements made remain the responsibility of the named author(s); the views do not necessarily reflect those of the government of the designating Member State(s) or of the designating organization(s). In particular, neither the IAEA nor any other organization or body sponsoring this meeting can be held responsible for any material reproduced in this preprint.

**We regret that
some of the pages
in the microfiche
copy of this report
may not be up to
the proper legibility
standards, even
though the best
possible copy was
used for preparing
the master fiche**

BIOTECHNOLOGY INFORMATION SERVICE OF THE GDR

The paper gives a survey of the biotechnology information in the GDR and describes the establishment of the **Biotechnology Information Service of the GDR (BioInfo GDR)**.

BioInfo GDR is a referral database and is to provide information on information sources available in the GDR, and on institutions working in the various fields of biotechnology in the GDR. An addition, some general problems of the building and use of databases are discussed.

1. INTRODUCTION

The availability of high-quality, up-to-date and comprehensive information is an important requirement in biotechnology and its applications in ever wider fields of medicine, ecology, agriculture and food industry. The last ten years have seen a considerable and at the same time necessary increase in the amount of available information on biotechnology. Information is becoming a key factor in biotechnological research, development and industry.

The results of the rapid development in biotechnology with the increasing growth of data on the one hand, and the possibilities of information technology on the other hand, are

very large databases

containing different kinds of information

- bibliographic information
- factographic information
- referral information
- full-text information

and

highly sophisticated software

for processing and use of these databases.

This development is connected with

- the organization of a comprehensive **information service** including
 - . various possibilities for the availability of databases (magnetic tape, floppy disk, CD-ROM);
 - . communication networks (online access, electronic mail, computer conferencing);
 - . special information processing services;
- the enhancement of **international cooperation** with the purpose of
 - . increase and coordination of the information exchange, and
 - . establishing of only a few, but powerful information centres for collection, processing, and exchange of data;
- high governmental and industrial **investment** to provide the preconditions necessary for the establishment of databases [1].

But the increasing amount of information possibilities on the one hand, and the multidisciplinary character of biotechnology with very different information needs of users in research and industry on the other hand, demand new ways of solving the information problems. One of these tools are **referral databases** which can serve as guides to information sources, research projects, products etc.

In the following the paper deals with the establishment of such a referral database: **Biotechnology Information Service of the GDR (BioInfo GDR)**

2. THE ESTABLISHMENT OF THE BIOTECHNOLOGY INFORMATION SERVICE OF THE GDR

2.1. Starting point

The GDR's situation in the field of biotechnology information is characterized by the fact that:

- the information possibilities are inadequate;

- the existing information possibilities are insufficiently known and the routes are troublesome;
- precisely in regard of information from or about the GDR there is a lack of knowledge.

Proceeding from this situation it was determined in March 1988 to design a "*Biotechnology Information Service of the GDR*" (*BioInfo GDR*) [2].

2.2. Institution in charge.

Academy of Sciences of the GDR
 Scientific Information Centre
 Schiffbauerdamm 19, DDR-1040 Berlin

2.3. Objective

The *BioInfo GDR* is to improve the information provision in the area of biotechnology and become a powerful information tool.

It will:

- help information centers and libraries to inform users of the existing information possibilities quickly and comprehensively;
- serve the scientist as an information guide which stewards him through the information landscape to the desired information.

2.4. Tasks

1. The *BioInfo GDR* is to provide information on the information sources in the fields of biotechnology available in the GDR (databases, special information materials, periodicals, software, etc.)

These files are to serve to answer the following questions:

Which information sources are available in a particular field of biotechnology in the GDR? Does a particular source of information exist in the GDR and where is it to be found? What software is available in the GDR for processing biotechnological data (e.g. sequences, structures)?

2. The **BioInfo GDR** is to provide information on institutions and companies working in the various fields of biotechnology in the GDR, including information concerning completed research projects and others which are in progress: products, technologies, possible services, culture collections etc.

This database "Who-What-Where" is to provide an answer to the following question:

Which research and industrial facilities in the GDR are working on what biotechnology relevant research projects and are in the position to supply which products or provide which services?

2.5. Contents

Because of the multidisciplinary character of biotechnology and the variety of questions directed to the biotechnology information service, the **basic fields of biotechnology** (molecular biology and genetic engineering, biochemistry and biophysics, microbiology and others) as well as the **fields of applied biotechnology** (medicine and pharmacy, chemistry and chemical engineering, agriculture and nutrition, ecology.etc.) are to be included.

2.6. Coverage

The **BioInfo GDR** contains only information about biotechnology in the GDR. For a comprehensive information provision we use the following possibilities:

1. Online-Access

- to the databases, which are offered through the international Online Services Data-Star, STN International, FIZ Technik, DIMDI.

Use is possible through information centers by the help of information brokers;

- to the databases which are offered through the centralized facilities of the GDR.

2. Offline-Use

- of international databases which are obtained on magnetic tape or CD-ROM

Bibliographic databases: MEDLARS, BIOTECHNOLOGY ABSTRACTS, VINITI-magnetic tape services (BIO, PCB) and others.

Factographic databases: GenBank, EMBL, PIR, SWISSPROT Protein Sequence Database, Protein Data Bank, and others;

- of databases established in the GDR

Bibliographic databases: Chemical Processing Technology, Food Industry, AGROSELEKT, UMWELT-INFORM, and others.

Factographic databases: EMBOPRO, TOXDABA, SPRE-SI, WIFODATA, European Barley Database, National Cancer Registry, and others.

The results of an analysis of the current situation regarding factographic and bibliographic databases relevant to basic and applied fields of biotechnology and available in the GDR have been compiled in two directories [3][4].

2.7. Pre-conditions

The pre-conditions necessary for the establishment of *BioInfo GDR* include

- data input

The acquisition of the necessary information is one of the main difficulties in the construction of the *BioInfo GDR*. At present we employ the following possibilities:

1. Distribution of questionnaires

- . Questionnaire on data collections, data banks and expert systems in biotechnology
- . Questionnaire on Biotechnology Information Service (File: Who-What-Where?)

2. Extraction of information from the literature

3. Extraction of information from other referral databases

- **controlled vocabulary**

For the contents analysis, indexing and retrieval of information is used the "Thesaurus Biotechnologie" [5] developed by the Institute of Biotechnological Research (FRG). In this way there is a linguistic compatibility in both German Biotechnology Information Systems BIKE, and AuBit;

- **investment**

The establishment of databases requires considerable investment and cannot be taken to be of secondary importance. The role of information will be determined by how a society answers the economic questions concerning the production of information, the distribution of information, and especially concerning the **development of an information infrastructure;**

- **hardware and software**

At present, *BioInfo GDR* is carried out by means of PC technique and a databank management system developed for *BioInfo GDR*;

- **a qualified staff including information management**

The interaction between biotechnology and information science demands a "hybrid" knowledge. At an international level, considerable attention is paid to information management.

2.8. Availability

1. In the printed form

for ready availability of information as reference books:

- Directory of factographic databases in the field of biotechnology available in the GDR [3]
- Directory of bibliographic databases in the field of biotechnology available in the GDR [4]
- Directory of information sources in the field of biotechnology published in the GDR [6]
- Directory of periodicals in the field of biotechnology available in the GDR (in preparation)

- Directory of institutions in the GDR: Who makes What in the GDR? (in preparation)
- Information on conferences, workshops, fairs in the field of biotechnology (in preparation)

2. In the form of a computer-aided information service

The advantages of the database consist in the fact, that it contains more information than the printed version and that it can answer multidimensional questions, typical for biotechnology, easily and quickly.

3. Online availability

In the future, we intend offering *BioInfo GDR* via online access.

2.9. Further Development

The aims of the further development of *BioInfo GDR* are

- to use its possibilities to identify trends in biotechnology,
- to support the planning of research,
- to provide a basis for the co-ordination of research and the application of research results in the field of biotechnology.

In this way *BioInfo GDR* could help to prevent parallel work in basic and applied research.

Another future aim will be to connect the both German biotechnology information services:

- BIKE (Biotechnologie-Informationen-Knoten für Europa), FRG,

and

- AuBit (Aukunftsdienst Biotechnologie), GDR.

A further extension would also include

- elements of a full-text system by setting up a file titled "Biotechnology in the GDR - News Service",
- new communication possibilities as mailbox service and telecon ferences.

At a later point of time we will investigate the necessity and possibility of including artificial intelligence methods. The aim could be to develop an expert system which would assist both information scientists and end-users in a referral situation.

3. SOME GENERAL PROBLEMS OF THE BUILDING AND USE OF DATABASES

1. Insufficient appreciation of "information" as a resource

Reasons

- an underdeveloped information awareness;
- we have not learned to use "information" as a resource comprehensively and to transform the provided information into knowledge.

2. The necessary pre-conditions for the building of databases are only insufficiently provided (especially valid for the GDR situation)

Reasons

- the information basis in qualitative and quantitative inadequate terms;
- the willingness to national and international co-operation does not meet the requirements;
- the effort in setting up databases is underrated and thus there is a lack of provision of the necessary manpower, financial resources and equipment;
- the information management is underdeveloped.

3. The interfaces between the research and the information are insufficiently developed

Reasons on the side of information

- the "classical" understanding of documentation continues to prevail;
- the necessary pre-conditions for a comprehensive information service do not exist.

Reasons on the side of research

- a lack of possibilities as a consequence of inadequate pre-conditions or insufficient data quality, but also unwillingness to feed research findings into databases and thus make them more widely available;
- occasionally a dangerous **ignorance** is exhibited with regard to information.

4. The guarantee of the quality of the database

concerning

- the actuality of the database

determined by the time it takes to make new information available on a database, and by the possibility of rapid access to the database;

- the reliability and accuracy of data

connected with the necessity

- . to integrate evaluation algorithms,
- . to expand the syntactic and semantic representation possibilities,
- . to formulate requirement characteristics;

- the complexity of data and databases

connected with the necessity

- . to crosslink different (kinds of) information which are part of different (kinds of) databases,

- . to combine different forms of information representation,
- . to extend the numeric data by means of supplementary, descriptive information,
- . to use standardized or easily translatable formats and they must be interconnected.

4. CONCLUSIONS FOR THE FUTURE OF BIOTECHNOLOGY INFORMATION IN THE GDR

1. Formation of a new information awareness and behaviour

- in the conviction that information comprises an indispensable component of all fields of science and technology;
- combined with the understanding that in many fields there are no alternatives to the use of international databases.

2. Creation of the necessary pre-conditions

- for the use of the large international databases for which there are no national equivalents;
- for the setting up of specific databases, which contain information on research findings and are available either as independent databases or as data input for international databases.

3. Development of a highly sophisticated software

With the large databases we have a huge information potential at our disposal, but we are not able to utilize it to a high degree. A highly sophisticated software is required in order to make the knowledge contained implicitly in these large databases explicitly available, especially artificial intelligence methods should be introduced in order to improve the processing of different kinds of information.

In the future, various forms of representation patterns will have to be combined in order to be able

- to adequately and completely represent objects' properties, relations und interrelations;
- to transfer information into knowledge.

4. Importance of factographic information

Factographic databases derive their importance from the following:

- Factographic databases provide a **research instrument** which
 - . as a consequence of its indispensability for certain types of problem solving, has led to the formation of a new attitude toward information on the part of scientists;
 - . arises at the interface between subject area and information science whereby the scientist increasingly assumes the role of producer and user of information;
 - . bears witness to the increasing influence of information technologies on the research process [7].
- Factographic databases have to provide information required for Computer Aided Design, e. g. protein design, and as a part of the **knowledge bases** of expert systems [8].

5. Participation in the international information exchange

The increasing amount of information and the considerable expenditure involved cause the gathering, processing and distribution of factographic data to be possible only with **international scientific co-operation and collaboration by means of modern information technology**. Maintenance of international databases is particularly important to provide timely access to information from research conducted around the world.

Proceeding from the fact that significant databases in the field of biotechnology like nucleotide sequence, protein sequence, hybridoma and

culture collections databases etc. can only be created through **international co-operation**, it is **imperative** that all countries participate in the international information exchange.

I do not know if there is a sensible reason why a "world" sequence database, a "world" culture collections database and "world" hybridoma databases should not be set up.

At present time, the basis for an international co-operation is so good than never before. **We should use this chance!**

REFERENCES

- [1] FOETZSCH, E., Databases in molecular biology - the international situation, *Studia Bio physica* 129 2-3 (1989) 109-116.
- [2] FOETZSCH, E., Vorstellungen zum Aufbau des Auskunftsdienstes Informationen zur Biotechnologie, *Informatik*, 35 5 (1989) 174-176.
- [3] FOETZSCH, E., Verzeichnis faktographischer Datensammlungen und Datenbasen zu Grundlagen- und Anwendungsgebieten der Biotechnologie, *Sci. Inf. Centre*, Berlin (1988).
- [4] EBNER, A., Verzeichnis bibliographischer Datensammlungen und Datenbasen zu Grundlagen- und Anwendungsgebieten der Biotechnologie, *Sci. Inf. Centre*, Berlin (1989).
- [5] *Biotechnologie-Thesaurus*, Inst. f. Biotechn. Res., Brunswick (1988).
- [6] LUECKE, E.M., MUENNICH, B., FOETZSCH, E., Verzeichnis von Informationsmitteln zu Grundlagen- und Anwendungsgebieten der Biotechnologie, *Sci. Inf. Centre*, Berlin (1989).
- [7] FOETZSCH, E., Stand und Entwicklung faktographischer Informationssysteme auf dem Gebiet der Biotechnologie, *Informatik*, 34 2(1987) 73-76.
- [8] FOETZSCH, E., "Forderung von CAD und Expertensystemen an die Entwicklung faktographischer Datenbasen auf dem Gebiet der Biotechnologie", *Wissensvermittlung-Informatik (Proc. Int. Symp. Erfurt, 1987)* 178 - 191.