

Chapter 32

Status of the Infrastructure of Information Resources Supporting U.S. Biotechnology

Ronald A. Rader

OMEC International, Inc., 727 15th Street, N.W., Washington, DC 20005

The infrastructure of information resources supporting U.S. biotechnology needs improvements in a number of areas. Existing resources do not meet the needs of many users and uses. Information resources need to be developed for safety assessment, international competition, public information, and other areas. Biotechnology-involved organizations need to initiate and improve their information centers, systems, resources, expertise and services. Information resources are a limiting factor affecting organizations' and national innovation, competitiveness, decision making, protection of intellectual property and obtaining patents, regulatory affairs, research, development, and commercialization in biotechnology.

The U.S. biotechnology information infrastructure is the sum of all the nation's readily available information resources, services, and professional expertise, both within organizations and publicly available. Although not contributing directly to the bottom line, readily accessible, high quality information resources, services and professionals have a great impact on organizations' and nations' capabilities, productivity, actions and reactions, and competitiveness.

Chemistry and toxicology/pharmacology are areas where a strong infrastructure of information resources exists in the U.S. Chemical, toxicology, and pharmacology information resources are available in abundance and tailored to diverse needs, often with high degrees of specialization and sophistication, such as in-depth subject indexing, registry and nomenclature systems, substructure searching, and structure and activity predictive systems.

The state of the infrastructure of information resources in biotechnology may be compared with that of toxicology and related life and chemical sciences as of about ten or more years ago (1). Prior to the mid-1970's, there were very few toxicology information resources available. Federal information activities mandated by the

0097-6156/88/0362-0375\$06.00/0
© 1988 American Chemical Society

Toxic Substances Control Act, the National Cancer Act, and other societal efforts to regulate and define chemical-related public and environmental health threats from the mid-1970's to the present have resulted in a very healthy array of information resources. Similarly, activities and resources in other life sciences, especially the biomedical sciences, have grown and a healthy array of information resources developed, many with federal support.

Yet, few and inadequate biotechnology-oriented information resources exist. More and better information resources are required to support the development of biotechnology into the \$40-100 billion industry it is predicted to be in the U.S. in the year 2000. In general, those biotechnology information resources which are available lack the sophistication required for many uses and users. There are a number of factors which have contributed to this situation, and a number of factors which should lead to development and availability of more and better biotechnology information resources.

Sources and Flow of Information in Biotechnology

A excellent overview of information sources and information flow in biotechnology has recently been published in second edition (2). Brief descriptions of most private sector information sources may be found here, along with much introductory and explanatory text. The biotechnology information marketplace is easily seen to be characterized by a very large number of primary sources of information, such as journals, meetings, conferences and proceedings, books, technical reports, and trade publications. Other sources of information for biotechnology include a number of often costly newsletters, consultants and experts. A large proportion of biotechnology-oriented information resources are devoted to commercial and competitive news and information reporting, as demonstrated by the large number of company directories available, and the commercial orientation of a large proportion of available specialized biotechnology abstracting and indexing services, as shown in Table I.

Table I. Some Major Biotechnology-specific Secondary Sources

Name	Orientation	Source
Abstracts in BioCommerce	Commercial	Britain
BioBusiness	Commercial	U.S.
Biotechnology Research Abstracts	Scientific	U.S.
Current Biotechnology Abstracts	Scientific	Britain
Derwent Biotechnology Abstracts	Scientific	Britain
Pascal Biotechnologies	Scientific	France
Telegen	Commercial	U.S.

Note: All but BioBusiness, a database, are available in publication and database form.

Biotechnology is a multidisciplinary, fragmented activity involving the interface of many scientific and commercial activities. The diverse sciences upon which biotechnology builds and to

which the biotechnologist needs information access includes many chemical and biological sciences and technologies, including microbiology, biochemistry, genetics, molecular biology, toxicology, pharmacology, and bioengineering. Also, biotechnologists require information about regulations, safety assessment, environmental effects, commerce, funding sources, patents, and other types of information. This fragmentation becomes even more apparent, when one realizes the diverse application areas of biotechnology, including industrial activities in pharmaceuticals and diagnostics, agriculture, food, energy, waste processing, and commodity and specialty chemicals. The fragmented nature of biotechnology, combined with its relative newness as a distinct activity, is a contributing factor to the lack of biotechnology information resources.

This often fragmentary interplay of disciplines and specialists is reflected in the numerous professional and trade organizations representing scientists and institutions involved in U.S. biotechnology. To date, neither professional nor trade associations in biotechnology have taken significant active roles in developing specialized information resources, a common situation for such organizations in many other fields.

Secondary sources, notably abstracting and indexing publications and bibliographic databases, are information resources routinely organizing and summarizing information about documents and their contents. Besides scanning of journals, these are usually the main means for keeping up with developments and for retrospective searching of the literature. Some major secondary publications and databases specifically oriented to biotechnology are shown in Table I. Examination of these and others reveals that they universally are spin-off or derivative (subset) products from major broader coverage scientific information services, and/or are primarily oriented to covering commercial news and activities in biotechnology.

Indexing and subject access in most secondary and other biotechnology and related information resources is rather primitive. Subject indexes either involve very simple and general classification schemes, employ keywords (no controlled indexing), or employ the classification and indexing schemes of their parent broader coverage resources. There has been little development of classification and indexing systems specifically for biotechnology information and access to the literature. The lack of classification and nomenclature schemes adversely affects the whole information infrastructure by making information retrieval and resources coordination more difficult and haphazard, and keeping information organization and exchange on very basic, simplistic level.

In the area of protein and nucleotide sequence databases, current resources are struggling to keep up with the published data, have reduced the amount of other information (annotations) recorded, capturing only the minimum data, and do not cover patents and commercial products. In fact, there are very few biotechnology information resources, whether bibliographic or other types, which seriously deal with biological technologies and products, rather than broad basic science or commercial activities.

Biotechnology Information Resources Marketplace

Biotechnology information resources have yet to successfully establish their niche in the U.S. marketplace for a number of reasons. Many biotechnology information resources find greater interest, market penetration, and their major markets in foreign countries, primarily Japan and Western Europe. Perhaps, the most important reasons for a sluggish U.S. demand and market for biotechnology information resources are: a general lack of recognition of the value of information resources as a strategic long- and short-term asset; lack of knowledge and exposure to specialized information resources among those most involved in biotechnology; and lack of highly visible national programs and activities.

U.S. organizations involved in biotechnology, including most biotechnology companies, are relatively weak in information resources, capabilities, and expertise. Biotechnology companies with a library/information center or an even partially dedicated information professional are a distinct minority. This situation occurs even in well-funded biotechnology companies with considerable research and development activity. In some cases, a local university library performs online searching and fulfills document requests on demand. For the most part, information handling is a haphazard and unorganized activity. Exceptions to this situation may be found in the established pharmaceutical and chemical firms becoming involved in biotechnology, most of which have information centers/libraries and information specialists thoroughly integrated into their research, development, marketing, and regulatory affairs efforts (3).

Biotechnology executives and researchers, when questioned about the need for information services and resources within their organization, very often reply that they are on the cutting-edge or forefront of their particular areas of research and development, go to all the right meetings, and keep in touch with the right people. Many fail to recognize the value and provide support for building and providing information resources, services, and expertise within their organization. Executives and researchers complain they have more information than they can assimilate, and mistake this for the information they may really need and that others in their organization should have long-term ready access to.

Biotechnology companies, on the whole, do not budget for information handling and organization, as do more established companies involved in pharmaceutical and chemical research and development. This is probably due to: their relatively recent entry into the commercialization and regulatory phases of product and process development; their not making profits, yet; and the history of most start-up companies' researchers and executives coming from academia or other biotechnology companies. Established chemical and pharmaceutical industries spend on the order of 2% of their research and development budget for library/information center and related resources and staff, but this is not observable in biotechnology companies.

A situation of information rich vs. information poor may arise or presently exist in biotechnology. An elite of larger biotechnology and other companies may be better able to conduct cost-effective research, commercialization, regulatory affairs, obtain and defend

patents, and survive in the world marketplace. With the history of most significant biotechnology innovations and developments arising from small companies, universities, and research institutions, this may have broad strategic implications for these organizations and U.S. biotechnology.

Some biotechnology companies are finding that they need to develop their own information resources. Some major biotechnology companies have become information vendors through commercialization of initially in-house information resources. Examples include Abstracts in BioCommerce, originally developed by Celltech in Britain, the AGRIBUSINESS database developed by Pioneer Hybrid, and the BioScan corporate activities directory developed by Cetus Corporation. Many companies are finding that organized information is a marketing asset. Often, companies distribute extensive bibliographies, and some operate online electronic mail networks relating to their products. These trends will likely continue.

The very nature of biotechnology complicates information handling and the protection of inventions through patents. For example, there are many ways to define and characterize biotechnology-related organisms, their products and components, and processes. One can identify and describe organisms and their products based on sequences and structures of DNA/RNA and proteins, uses and applications, observable characteristics and appearance, metabolic activities, and other parameters. Terminology used in biotechnology is far from standardized, and may be purposefully ambiguous or unclear to broaden and obscure boundaries of patent coverage (4). Only a few of many patent and intellectual property issues in biotechnology have been resolved in the U.S. and foreign countries' courts.

International Competition in Biotechnology Information

The U.S. presently is the leader in most aspects of biotechnology, due primarily to the considerable basic biomedical and life sciences research efforts of the federal government and a strong entrepreneurial industrial sector of biotechnology start-up companies which have built upon this research (5). Similarly, many large and established chemical, pharmaceutical, biomedical, agricultural and other U.S. firms have become very involved in biotechnology research, development, and commercialization.

However, a number of foreign governments have targeted biotechnology as an important area where they are developing coordinated national efforts to challenge U.S. research and market preeminence. Development of information resources is formally recognized as an important component in these efforts.

The European Communities (Common Market) has sponsored the European Biotechnology Information Program (EBIP), recently accorded permanent funding status and renamed the Biotechnology Information Service, within the the British Library for several years (6). EBIP sponsors an annual meeting concerning biotechnology information, provides information services on demand, assists inquirors with information acquisition, and is actively analyzing and reviewing the information requirements of its member countries' research and commercial institutions. EBIP has sponsored studies, including assessments of the feasibility of a computerized information system

for European culture collections (collections of viable samples of microorganisms) and an information system on enzymes and enzyme engineering. The U.K. has recently implemented online access to its various culture collections' holdings.

The Japanese government has well established and coordinated industrial biotechnology research and development programs and research centers with a number of associated specialized information centers and activities. A branch of the Japanese government has recently outlined development plans for an integrated protein data network. These foreign government-supported efforts are too new to assess their impact on international competition, but are worthy of our attention. Also, most of the specialized secondary information resources shown in Table 1 and many others originate in Europe.

International competitiveness and encouragement of innovation are ever growing issues in the U.S. Information resources are not a solution to U.S. problems in these areas. However, information resources need to be recognized as a limiting factor for competitiveness and innovation at both the organizational and national level.

Federal Biotechnology Information Resources and Activities

The federal government is the single main organization responsible for and involved in biotechnology. Biotechnology originally developed from federally funded research, which remains the primary impetus for biotechnology research and development activity in the U.S. and the reason for acknowledged U.S. leadership in the field. Federal agencies spent over \$2 billion dollars for biotechnology and related research in Fiscal Year 1986 and this level of spending will likely be maintained (7-8). Despite major U.S. interests and investments in biotechnology, generally, the federal government has not initiated development of biotechnology information resources to support national needs and federal mandates.

OMEC International, Inc. has recently completed its Federal Biotechnology Information Network (FBIN) project with partial federal funding. This has resulted in publication of the Federal Biotechnology Information Resources Directory (9), describing over 470 federal biotechnology-relevant information resources, and the Federal Biotechnology Program Directory (10), describing over 470 biotechnology-relevant research, regulatory, technology transfer and other federal programs and activities. Together, these provide the first comprehensive description of the infrastructure of federal resources and programs supporting and affecting biotechnology, exclusive of facilities.

From this project and other experience, a number of general conclusions may be reported regarding federal biotechnology information resources and activities:

- 1) There has been no significant development or discussion of new, needed information resources for biotechnology (with some exceptions noted below).
- 2) Most federal biotechnology-related information resources and programs are not specific for biotechnology. Rather, they support underlying or related basic research, or more generalized regulatory or other agency activities.

- 3) Existing biotechnology-related information resources, on the whole, are relatively stagnant, receiving little additional funding for qualitative or quantitative improvements.
- 4) There exist insufficient information resources to appropriately support biotechnology-related public health and environmental safety assessments. Information resources do not exist or are not readily available to assist persons in information gathering and assessment to evaluate the effects of releases of genetically engineered or other novel microorganisms and their products in the environment and marketplace.
- 5) Many agencies formerly active in chemical and biological information resources development and information dissemination are now significantly less active in these areas. This is most notable among the regulatory agencies. This general situation may be due to the political climate for deregulation. Many policy and programmatic decision-makers are not favorably disposed to information resources, recognizing that information resources are required and may be used to support development of regulations and spot potential and developing problems.

Major ongoing federal biotechnology-specific information resources and activities include: GENBANK and other nucleotide sequence database systems; the Protein Identification Resource (PIR) protein sequence database; the Microbial Strain Data Network (MSDN), a directory to culture collections' holdings; and the National Library of Medicine's biotechnology information research program and Biotechnology Information Resources Directory, to be an online database and published directory of worldwide information resources.

Biotechnology Safety and Oversight Information Resources

The lack of biotechnology information resources, accessible information, and infrastructure development is already having an adverse impact on U.S. biotechnology. This is most obvious in the related areas of regulation and oversight of research and premarket testing, safety and hazard assessment, information dissemination, and public (mis)perception and (mis)understanding of biotechnology-related hazards. New, innovative technologies, and especially biotechnology, require well-developed, comprehensive, coordinated, science-based regulations to establish public and industry confidence in regulatory and oversight actions and procedures. Currently, important regulatory and safety assessment are performed on a case-by-case basis by a handful of persons with experience and/or credentials in this area. The Biotechnology Sciences Coordinating Committee (BSCC) has been formed and a coordinated framework for federal regulation is being put in place. However, there are few, if any, information resources available to assist in assessments of novel biotechnology products and organisms or make this information available to the biotechnology community and general public.

The lack of biotechnology product and process safety-related information resources is likely to make itself more evident as more legal, regulatory, and safety-related delays, uncertainties, and misjudgements. Even at this early stage in the development of U.S. biotechnology, a number of procedurally based, obstructive lawsuits

have successfully diverted and delayed federal, academic, and industry testing and commercialization plans. Both small biotechnology companies and large, established chemical firms have made significant mistakes in the design of premarket testing strategy, protocols, and information provided (or not provided) to government agencies and the public.

Although the slowly advancing unresolved and uncoordinated nature of regulation and oversight within and among the federal and other government agencies is a major factor in regulatory and judicial delays and uncertainties, the general lack of organized and accessible information is surely a strong contributing factor. No fatal or other significant biotechnology-related accidents or adverse environmental modifications have occurred yet, but there are ample examples to be taken from the chemical industry of unidentified and misassessed hazards resulting in mishaps, public and environmental health hazards, and corporate liabilities. In partial response to this situation, OMEC International has recently published Biotechnology Regulations: Environmental Release Compendium, a compilation of U.S. federal, state, and local regulations, laws, and guidelines concerning releases in to the environment of genetically engineered microorganisms (11).

The NRC Committee on Biotechnology Nomenclature and Information Organization

A workshop sponsored by the National Library of Medicine (NLM) of the National Research Council Committee (NRC) on Biotechnology Nomenclature and Information Organization was held in May 1986 (12). Various subcommittees examined the state and relevance of chemical and biological nomenclature, the organization of biotechnology information, and developed a number of recommendations. Major recommendations included:

- 1) All federal agencies involved in biotechnology should continue current and initiate new programs and activities in biotechnology information. This could involve the establishment of information centers of excellence in biotechnology which might develop and provide information resources, conduct research related to biotechnology information, and provide referral services.
- 2) The NLM should catalyze national and international efforts to coordinate and develop standardized subject vocabularies (for terminology and subject indexing schemes) for biotechnology disciplines, and a uniform nomenclature in the form of registries for organisms, clones, genetic elements, and other biotechnology materials and products.
- 3) The NLM should establish a "database of databases" for biotechnology and expand its role as an information resource center. This would involve expansion of the DIRLINE database, NLM's online directory of biomedical and other information resources. Work in this area will be initiated this Fall.
- 4) NLM should develop a cross-referencing system and a thesaurus (subject classification scheme) for biotechnology information resources. A cross-referencing system would work in tandem with the "database of databases" to facilitate use of common data elements, compatibilities, and data sharing among databases, and

also aid searchers in identifying and locating sources of desired types and forms of data and information.

- 5) The NLM should facilitate networking among database systems and establish "transparent" interfaces among them.

The report emphasized that the federal government needs to recognize the importance of biotechnology information as a national resource vital to science, technology, commerce and other national interests. The Committee recognized the need for deficit and federal budget reduction, but reported that the economic advantages of developing, processing, and disseminating biotechnology information far outweigh the costs. Biotechnology deserves a high standard of information resources and federal involvement in these, much as other developing technologies have a federally-sponsored common denominator of information resources.

The Committee reported that vocabulary in biotechnology is suboptimal. This includes the terminology used by scientists, such as the fabricated terms used for transposable genetic elements, the undeveloped or nonexistent nomenclatures for biotechnology products and processes, and the biological and chemical nomenclatures now in use. Registries need to be developed for clones, genetic elements, and other materials used in biotechnology to provide unique and unambiguous identifiers and descriptions. Biological nomenclature currently provides taxonomic descriptions of whole organisms and does not extend to their components or below the species level, which is the level at which biotechnology functions. Similarly, chemical nomenclature is not oriented to complex macro- and multi-molecular biological materials. These nomenclatures break down when applied to recombinant organisms, cell lines, genetic elements, modified proteins, antibodies and other biotechnology materials.

Congressional Activities

New programs and significant reorientations of funding and priorities within federal agencies are difficult without Congressional mandates or other high-level directives. As discussed above, much of the U.S. infrastructure of information resources in the chemical and related life sciences may be traced to laws passed in the mid-1970's. Congressional actions are likely to be required to initiate similar activity in biotechnology information.

Rep. Pepper has introduced the National Biotechnology Information Act (H.R. 393) in Congress. The bill would establish a National Center for Biotechnology Information within the National Library of Medicine and provide additional funding of \$10 million/year. The bill does not contain much detail about specified programs and activities. It is primarily oriented to the molecular biology and biomedical research communities and National Institutes of Health (NIH) activities. Activities mentioned in the bill and supporting materials include nucleotide and protein sequence databases, development of information resources for gene mapping, and the coordination and integration of computer-based information resources. Resources and programs for safety assessment, international competitiveness, public information, technology transfer, regulatory coordination, classification schemes and registries are not specifically addressed. It will be interesting to follow the evolution of this

bill and the level of effort to be directed to the critical applied and technological information needs of biotechnology. Those concerned with biotechnology information should take note of this bill and participate in its formulation and debate.

Recommendations

The author endorses the NRC Committee's recommendations, especially the first calling for recognition of biotechnology information as a national asset and establishment of information centers of excellence. The NRC Committee had a distinct biomedical orientation, properly reflecting the interests of its sponsor and the predominance of biomedically-oriented biotechnology within the federal and private sectors. Many of the same findings also apply to critical biotechnology information needs for agriculture, commerce, energy, and defense.

Besides the Committees recommendations, and the general requirement that biotechnology organizations upgrade their information resources, the author suggests prompt federal and private sector attention to:

- 1) Establishment of series of information centers collecting, translating, organizing, and assessing foreign biotechnology scientific and commercial information and developments;
- 2) Extensions of indexing and classification schemes used by established information resources, especially abstracting and indexing services, to better cover biotechnology;
- 3) Execution of user needs surveys, market studies, and assessments of available options and priorities in U.S. biotechnology information resources development;
- 4) Assessment by the federal government of the cost-effectiveness and appropriate means to assist the development, improvement and public release of private and nonprofit sector information resources;
- 5) Support for development and implementation of biotechnology information resources within the National Agricultural Library (NAL), Department of Commerce, and Department of Energy to parallel and keep up with the development of biomedically-oriented information resources;
- 6) Establishment of at least one information center and bibliographic and factual databases concerning the safety, risk assessment, and regulatory affairs of biotechnology products, processes, and materials; and implementation of an emergency response-capable information center and online database for biotechnology and industrial microbiology.
- 7) Development of knowledge-based and expert systems, and other information resources to supplement U.S. manpower and educational deficiencies and needs in bioprocessing, fermentation, and other areas of relative foreign dominance in biotechnology (5, 13); and
- 8) Establishment of federal and private sector clearinghouses to facilitate public access to biotechnology and related information.

In summary, biotechnology is a relatively new, diverse, major scientific and commercial activity in the U.S. and throughout the world. The infrastructure of information resources supporting U.S. biotech-

nology needs improvements on a number of levels, requiring efforts by all involved organizations - the federal government, the biotechnology and information industries, and research institutions. Greatest needs are for establishment and expansion of information collection, organization, and services within biotechnology-intensive organizations, especially U.S. biotechnology companies and research institutions, and the recognition and coordinated action of federal agencies to promptly address biotechnology information resource needs and problems. Federal implementation of safety, regulatory, and international information resources is required to protect the considerable U.S. investment in biotechnology. Although many author recommendations concentrate on the federal role and activities, the private sector needs to become involved in all aspects of these activities to assure understanding of biotechnology as a diverse technological and commercial, and not just as a biomedical research-oriented activity.

Literature Cited

1. Kissman, H. M. and Wexler, P., "Toxicology Information Systems: A Historical Perspective," Journal of Chemical Information and Computer Sciences, 25(3), pp. 212-217, Aug. 1985.
2. Crafts-Lighty, A., Information Sources in Biotechnology, 2nd edit., Nature Press, New York, 1986.
3. Brown, H. D., "A Drug is Born: Its Information Facets in Pharmaceutical Research and Development," J. Chem. Info. and Comp. Sci., vol. 25, pp. 218-224, 1985.
4. Meyers, N., "Biotechnology Patents: Don't Say Just What You Mean," Nature, vol. 324, p. 504, Dec. 11, 1986.
5. Office of Technology Assessment, Commercial Biotechnology: An International Analysis, OTA-BA-218, Government Printing Office, January 19, 1986.
6. Cantley, M., "Bio-Informatics in Europe: Foundations and Visions," Swiss Biotech., vol. 2, no. 4, pp 7-10, 13-14, April, 1984.
7. Office of Technology Assessment, Public Funding of Biotechnology Research and Training, (in press), workshop held Sept. 9, 1986, Washington, DC.
8. Perpich, J. G., "A Federal Strategy for International Industrial Competitiveness," Bio/Technology, vol. 4, pp. 522-525, June, 1986.
9. OMEC International, Federal Biotechnology Information Resources Directory, Washington, DC, 1987.
10. OMEC Int., Federal Biotechnology Programs Directory, Washington, DC, 1987.
11. Strauss, H. S., Biotechnology Regulations: Environmental Release Compendium, OMEC Int., Washington, DC, 1987.
12. Committee on Biotechnology Nomenclature and Information Organization, National Research Council, Biotechnology Nomenclature and Information Organization, National Academy Press, 1986.
13. Zaborsky, O. R., and Zubris, D. K., Biotechnology Engineers: Status Report 1985, OMEC Int., Washington, D.C., 1985.

RECEIVED July 30, 1987