typical for XML documents, the development of appropriate transaction models, devising new access methods (including methods of updating XML data), versioning and configuration management.

## 4. Conclusions

The amount of information we have at hand grows exponentially. The information is of varying quality and structure – from highly structured, "clean" data, appropriate for controlling devices or managing big enterprises, to irregular, imprecise and inconsistent data, distributed via Web sites of varied origin. Such situation creates new challenges for database systems, whose main task has always been to organize, store and make available data in a way most adequate for a given application. The main challenge can be formulated as follows: The priority is to develop database systems that would be able to collect, organize, store, analyse and make available all information resources of humankind in such a way that the information could be used on-line by anyone.

It is clear that this general goal is strongly related to Web technologies. Firstly, in the nearest future, most of our knowledge will be digitised (in the form of *digital libraries*) and made available globally through the Web. Secondly, the number of Internet users grows so fast that very soon most of the Earth's citizens (at least in the more developed areas) will become consumers of this knowledge, and consequently, clients of the knowledge repositories. Many of them will also become producers of knowledge, which will also require access to knowledge repositories. An ideal database system that we aim at should be able not only to respond to any queries formulated by any user of the global network, but also to anticipate users' queries and actively present useful information. As a matter of fact, we should strive to transform the huge storehouse of data called the Web into an integrated, intelligent, global information system based on an advanced and mature database technology.

### *References*


[1] Cattell R G G and Barry D K (Eds) 2000 *The Object Data Standard: ODMG 3.0*, Morgan Kaufmann Pub.

[2] Stonebraker M and Brown P 1999 *Object-Relational DBMSs: Tracking the Next Great Wave*, Morgan Kaufmann Pub.

[3] *Extensible Markup Language (XML) 1.0*, Second Edition, W3C Recommendation, October 2000, www.w3.org

[4] Hackathorn R D 1999 *Web Farming for the Data Warehouse*, Morgan Kaufmann Pub. Inc.

[5] Gulutzan P and Pelzer T 1999 *SQL-99 Complete, Really*, R&D Books

[6] Bernstein P, Brodie M, Ceri S, De Witt D, Franklin M, Garcia-Molina H, Gray J, Held J, Hellerstein J, Jagadish H V, Lesk M, Maier D, Naughton J, Pirahesh H, Stonebraker M and Ullman J 2000 *The Asilomar Report on Database Research*, www.acm.org/sigmod/record

[7] *OMG Unified Modeling Language Specification, Version 1.4*, September 2001, www.omg.org

[8] Florescu D, Levy A and Mendelzon A 1998 *ACM SIGMOD Record* **27** (3) 59

[9] *Document Object Model (DOM) Level 2 Specifications*, November 2000, www.w3.org

[10] *XML Path Language (XPath) Version 1.0*, W3C Recommendation, November 1999, www.w3.org


# Provision of Databases in the Poznan Supercomputing and Networking Center


Sławomir Niwiński[1], Iwona Pujanek[2] and Maciej Stroiński[1]

[1] *Poznan Supercomputing and Networking Center, Noskowskiego 10, 61-704 Poznan, Poland, {niwinski,stroins}@man.poznan.pl*

[2] *Poznan University of Technology, Main Library, Plac Skłodowskiej-Curie 5, 60-965 Poznan, Poland, iwona@ml.put.poznan.pl*


(Received 16 October 2002; revised manuscript received 13 January 2003)

**Abstract:** The article presents a concise report on the experience gained in the last three years in the scope of network provision of bibliographic databases by the Institute for Scientific Information (ISI) and full-text humanities and medical databases by EBSCO Publishing. The authors emphasise the importance and impact of the programme and the databases, co-financed by the State Committee for Scientific Research, on the initiation and continuation of organisational activities and efficient database access management. The paper contains a short review of the information resources presently available, including the titles of bibliographic and full-text databases, the scope of licences, subscription periods, and the volumes of archival resources. It provides statistics illustrating the distribution and extent of the bibliographic database usage by the scientific community, including, active institutional and individual, users and discusses the hardware and software used to provide the network database access services, the availability conditions, as well as the rules of license renewal and co-financing by the interested institutions. The report also deals with the access conditions and access abilities to the electronic versions of humanities, economics and medical databases offered by EBSCO Publishing. It is vital to show the

structure of the service, the scope of subjects available in the databases, title estimation and the access to html, as well as image source types. Moreover, the report focuses on the modes of access to single database and to several of them simultaneously, as well as on multi-aspect searching with the use of the author and subject indexes and keywords.

**Keywords:** bibliographic databases, full-text databases, network access, EIFL Programme

## 1. Introduction

The activity of the Poznan Supercomputing and Networking Center (PSNC) is concentrated on five main areas:

- operating the POL-34/622 Polish National Scientific and Education Network as well as international connections,
- operating the Poznan Metropolitan Area Network – POZMAN,
- an HPC Centre,
- an System and Network Security Centre,
- R&D on new generation networks, grids and portals.

Since 1996 PSNC has maintained and provided users with access to bibliographic databases [1], and since 2000, to electronic versions of the full-text databases of EBSCO Publishing.

Literature databases containing basic data concerning publications in the various fields of science and general knowledge are a significant source of information used in scientific work. Apart from the basic bibliographic data on publications, these databases more and more frequently contain abstracts created by authors and publishing houses, as well as full-text articles, and also faithful images of publications with graphical elements and photographs. Information run in databases is usually organised in a specific way, *e.g.* as subject indexes. A vital criterion of database usefulness is the availability of proper searching mechanisms, both simple and advanced, one- and multi-criterial, but easy-to-use, intuitive and user-friendly.

Such access is offered by EBSCO Publishing within the Electronic Information for Libraries – EIFL Direct Project. It provides online access to literature databases, available at *http://search.epnet.com/*.

## 2. Bibliographic databases

### *2.1. Starting the database access service*

At the end of 1996, the Poznan Supercomputing and Networking Center obtained network licenses for the following bibliographic databases for the scientific community:

1. **SCI/CDEA** (Science Citation Index, Compact Disc Edition with Abstracts);
2. **CCCD** (Current Contents on CD with Abstracts, Six Editions), containing 6 series: Engineering, Computing & Technology, Life Sciences, Clinical Medicine, Social & Behavioral Sciences, Agriculture, Biology & Environmental Sciences, Physical, Chemical & Earth Sciences;
3. **A&HCI/CDE** (Arts and Humanities Citation Index).

All databases have been purchased from one vendor – the Institute for Scientific Information (ISI) from Philadelphia. The licenses for the aforementioned databases are of network and regional character and, when purchased, referred to the whole of the Poznan scientific community, *i.e.* universities, Polish Academy of Sciences institutes and R&D institutions. Purchasing a one-year license of any database means that it remains the property of PSNC (and the whole community) also after that period and will continue to be available in the following years [2, 3].

On purchasing the license to use the bibliographic databases, a database server and access Info Ware CD/HD UltraNet Software were installed. The system makes it possible to reload information recorded on CD-ROMs to disc memory and make it available via a computer network, so that over 100 users can be managed at the same time, to simultaneously store and provide several databases on-line, to provide management and access via several network protocols (TCP/IP, IPX, NetBios) concurrently.

### *2.2. Financing*

Three one-year licenses for bibliographic databases: Current Contents (CC), Science Citation Index (SCI) and Arts and Humanities Citation Index (AHCI) with the subscription period starting from 1997 were purchased for the funds granted in 1996 by the State Committee for Scientific Research. The subsidy covered 90% of the license price, with PSNC covering the remainder. The first licenses were purchased on preferential conditions within an ISI grant. To renew the license in 1998, a similar amount was granted by the Polish Foundation of Science Dissemination, which made possible the purchase of licenses for SCI and CC bases at an extra charge

from PSNC (approximately 13% of the license value). The subsidy from the State Committee for Scientific Research (approximately 35% higher than in 1997) for the renewal of the SCI license made it possible to purchase all three databases (SCI, CC, AHCI) for the year 1999 and, additionally, the SCI database license for the year 1996 (used for 2 years at no cost) as well as the AHCI database license for the year 1998.

Starting from 2000, the State Committee for Scientific Research has subsidised only the SCI database license, up to 25% of its value. The Poznan scientific community have decided to pay the remaining 75% of the SCI database license value and to fully finance the Current Contents database on their own. The amount of the fee paid by the institutions interested in the access to database titles is based the total annual usage time. PSNC does not charge users for the costs of maintaining the service.

### *2.3. Users*

In December 1999, the following scientific and R&D institutions of the Poznan community were actively using all the available databases: Academy of Economics, University of Medical Science, Agriculture University, Academy of Physical Education, Institute of Bioorganic Chemistry of the Polish Academy of Sciences (PAS), Institute of Dendrology (PAS), Institute of Energetics, Institute of Molecular Physics (PAS), Department of Human Genetics (PAS), Institute of Plant Genetics (PAS), Institute of Plant Protection, Institute of Natural Fibres, Poznan University of Technology, Adam Mickiewicz University, Industrial Institute of Agricultural Machines, and R&D Center of Machines and Special Devices.

### *2.4. Provision of bibliographic databases to the Poznan scientific community*

The server of the bibliographic databases is available at *baza1.man.poznan.pl*.

Using the databases run on the server is possible after installing client software with access to the Internet. The software is free and available from the ftp server at
*ftp://ftp.man.poznan.pl/pub/windows/ultranet/*.

Also, a www service containing detailed information on client software installation was prepared and made available at
*http://www.man.poznan.pl/software/databases/*.

## 3. EBSCO Publishing literature databases

### *3.1. Subject characteristics, service structure and periodical rank*

EBSCO Publishing offers access to two sets of literature databases: EBSCOhostWeb and EBSCO Medline, containing a total of 18 literature databases referring to: socio-economic sciences, the humanities, education, technology, business, information technology, bio-physics, bio-chemistry and medicine. The databases are of bibliographic-abstract character and comprise publications from over 8 000 periodicals, including over 6 000 in full text (in html, pdf, and xml) and represent such publishers as: Springer, Swets & Zietlinger Publishers, American Institute of Physics, Scandinavian University Press International Division, Blackwell Publishers, Taylor & Francis, B.C. Decker. Inc., Academy of Management, University of California Press, Arnold Publishers, Harvard Business School Publishing, Industrial & Labor Relations Review, Lawrence Erlbaum Associates, and many others. A detailed subject characteristics of the currently available databases follows [4]:

- **Academic Search Premier:** Full text of more than 3 460 scholarly publications covering academic areas of study including social sciences, humanities, education, computer sciences, engineering, language and linguistics, arts and literature, medical sciences and ethnic studies;
- **Business Source Premier:** Provides full text of over 2 800 scholarly business journals covering management, economics, finance, accounting, international business, *etc.*;
- **MasterFILE Premier:** Full text of over 1 900 periodicals, general reference, business, health;
- **Newspaper Source:** Full text of regional U.S. newspapers, international newspapers, newswires, newspaper columns, indexing and abstracts for national newspapers;
- **USP DI Volume II, Advice for the Patient:** Patient-oriented drug information in lay language;
- **Regional Business News:** Full text newswire database that incorporates business wires from all over the world, including A&G Information, Africa News Service, Inter Press Service, Resource News International, South American Business, M2 Communications, PR

Newswire, Business News Wire, Canadian Corporate News;

- **Health Source-Nursing/Academic Edition:** Nearly 580 full text scholarly journals focusing on many medical disciplines. Also featured are abstracts and indexing for over 615 journals;
- **Medline:** Authoritative medical information on medicine, nursing, dentistry, veterinary medicine, the health care system, pre-clinical sciences, *etc.* Created by the National Library of Medicine, allows to search abstracts from over 4 600 current biomedical journals;
- **ERIC:** Contains citations and abstracts from over 980 educational and education-related journals, as well as full text of more than 2 200 digests;
- **Health Source-Consumer Edition:** Health topics including the medical sciences, food sciences and nutrition, childcare, sports medicine and general health;
- **INSPEC:** Bibliographic information from the world's leading scientific and technical literature on physics, engineering, electronics, computers, and information technology;
- **Econlit:** Source of references to economic literature, includes journal articles, essays, research papers, books, dissertations, book reviews and working papers on accounting, consumer economics, monetary policy, labor, marketing, demographics, modeling, economic theory;
- **GeoRef:** Geoscience database on mineralogy and crystallography;
- **PsycINFO:** Citations and summaries of journal articles, book chapters, books, dissertations and technical reports, all in the field of psychology;
- **CINAHL:** Current nursing and allied health journals and publications;
- **Cochrane Database of Systematic Reviews (CDSR):** Reviews of the effects of healthcare;
- **Database of Abstracts of Reviews of Effectiveness:** Abstracts of published research reviews on the effects of health care from around the world;
- **Cochrane Controlled Trials Register:** Bibliographic listing of controlled trials in health care.

Over 2 000 titles from the full text periodicals are journals of high scientific rank, with high Journal Impact Factor. The EBSCOhostWeb service software provides users with access and the possibility to search one or several databases simultaneously.

### *3.2. Database organization, identification of information*

An important element of data structure organisation in databases is a clear and detailed record of data identifying each literature item. It is vital to unambiguously and quickly identify the record position in a database and to construct effective algorithms of reviewing its content and rendering it accessible. It should be noted, that in the case of EBSCO Publishing databases such unambiguous data identification common for all databases has not been introduced or made possible. In some databases, the record structure is, however, the same or very similar and, therefore, simultaneous searching of such databases is really effective. The basic information identifying a publication on a database is a bibliographic record created for the sake of identification needs.

In some databases, the record may be more detailed and contain, *e.g.* abstracts by the publication's authors. The record concerning a publication source is composed of a number of elements and contains the following information: the title of the periodical, year, volume, issue number, the number of the first page of the issue, the number of pages of the original, the number and type of graphical elements (pictures, diagrams, charts *etc.*). The richer the content of the bibliographic record, the easier the identification of a publication on the database.

### *3.3. Access mechanisms, searching criteria, work modes*

The interface provides a searching form in one of three available searching modes. The layout of each searching form depends on the context in which it was required. The layout is strictly defined if the search concerns one database only. The content of the form is defined mainly by fields identifying the bibliographic record. The following searching modes are possible: Basic Search, Guided Search and Expert Search. The Basic Search Mode is the simplest and requires neither experience in searching modes and strategies nor knowledge of the data structures. However, it is frequently used by experts, especially when they are beginning to learn a new subject, in order to narrow down the search scope. The other two modes of different complexity are chosen by more advanced

users, as they help narrow down the searching scope more effectively and quickly to the expected literature list.

### *3.4. EIFL Direct Programme in Poland*

On 15 May 2000, the Poznan Foundation of Scientific Libraries and Adam Mickiewicz University Library signed an agreement with the Open Society Institute (OSI) – Electronic Publishing Development Programe in Budapest (a branch of the George Soros Foundation) concerning co-ordination of the Electronic Information for Libraries (EIFL) Programme in Poland. The agreement assumes:

a) creating an open National Consortium of Libraries interested in access to electronic databases, mainly referring to the humanities, economic and medical sciences,
b) collecting funds of 50% of the license value, constituting the Consortium's own costs of the project, from all the Consortium participants,
c) acting to maintain and develop the Consortium,
d) subsidizing the project by the Soros Foundation.

On 31 August 2000, the Institute of Bioorganic Chemistry (PAS) – Poznan Supercomputing and Networking Center signed an agreement with the Poznan Foundation of Scientific Libraries and the Adam Mickiewicz University in order to create a National Consortium of Libraries. The consortium is supposed to provide authorised libraries in Poland with access to the electronic versions of EBSCO Publishing databases. PSNC has hoped to receive from the State Committee for Scientific Research a subsidy constituting 50% of the license value, and provided technology and organisation for the project realisation. Within the Consortium, PSNC is responsible for the installation, updating and maintenance of the local database copy (the Polish archive of EBSCO Publishing databases), as well as the development, introduction and maintainance of software providing access to the databases during the realisation period.

Information concerning the possibilities of network access to the electronic versions of EBSCO Publishing databases is available from company pages at *http://search.global.epnet.com* or *http://search.epnet.com/*, as well as from the pages of the Poznan Foundation of Scientific Libraries (*http://www.pfsl.poznan.pl/*).

### *3.5. Polish archive of EBSCO Publishing databases*

Copies of two largest databases, Academic Search Premier and Business Source Premier, are available in the Polish archive of EBSCO Publishing databases. The copies do not contain Full Text Images. The Polish database archive is available for the participants of the Polish Consortium of Libraries at *http://ebsco.man.poznan.pl/*.

A Polish language graphic interface has been created, and a search mode similar to the Guided (formerly Advanced) Mode available in the EBSCO Publishing service has been implemented for the needs of the archive.

## 4. Conclusions

The process of organising new community services, their maintenance and further development is obviously complicated, expensive and time-consuming. It was undoubtedly the significant subsidy from the State Committee for Scientific Research in the first years of disseminating access to the bibliographic databases [5] and the electronic versions of EBSCO Publishing databases [6] that has greatly contributed to the success of the aforementioned undertaking. The current databases are mostly co-financed or fully financed by the interested institutions.

### *References*

[1] Niwiński S and Stroiński M 1997 *Proc. "INFOBAZY'97 – Databases for Science"*, TASK Computer Center in Gdansk, Poland, pp. 348–352
[2] Niwiński S and Pujanek I 2001 *Advanced Methods in Accessing Bibliographic Database SCIENCE CITATION INDEX on CD with Abstracts*, Second Edition, Poznan Supercomputing and Networking Center, Report no. RA-002/2001
[3] Niwiński S and Pujanek I 2001 *Advanced Methods in Accessing Bibliographic Database CURRENT CONTENTS on CD with Abstracts*, Second Edition, Poznan Supercomputing and Networking Center, Report no. RA-001/2001
[4] Niwiński S 2001 *Polish Graphical Interface for the EBSCO Publishing Databases Archive Project and implementation*, Poznan Supercomputing and Networking Center, Report no. RA-005/2001
[5] Niwiński S and Stroiński M 2002 *Proc. "INFOBAZY'2002 – Databases for Science"*, TASK Computer Center in Gdansk, Poland, pp. 159–162
[6] Pujanek I and Niwiński S 2002 *Proc. "INFOBAZY 2002 – Databases for Science"*, TASK Computer Center in Gdansk, Poland, pp. 151–158