

Technical profile

Software features Search function

The speed and flexibility of the BioRS™ Integration and Retrieval System greatly facilitates gene discovery projects, including genomics, proteomics and drug targeting and delivery. The system enables researchers to quickly and efficiently retrieve all known biological data on a given topic, independent of the size of the database or the complexity of the query.

Researchers can quickly search multiple biological databases simultaneously using convenient Web or command-line interfaces. Queries and query results can be stored in a personal archive, which offers greater flexibility for searching. For example, long queries can be run in the background or stopped manually by the user. Results can be saved to a clipboard for future processing, and data on the clipboard can be downloaded in a variety of data and compression formats.

Integration of different formats

System administrators can easily integrate flat-file and relational databases using convenient Web or command-line interfaces and standardized data formats using a system based on Extensible Markup Language (XML). Both public and proprietary databases are easily integrated, independent of the source format.

The BioRS system supports the use of relational databases (Oracle® and MySQL™) with complex schemes. Using a Web-based graphical interface, table columns are easily integrated and Structured Query Language (SQL) conditions allow precise databank schemes to be built for easy searching. Relational database management systems can be accessed directly for searching. Alternatively, indices of relational data can be accessed to speed up searching.

Integration of remote data

To facilitate data sharing between members of a large working group, databanks integrated at any location can be accessed by users at other locations. Once a databank and corresponding indices have been integrated from a remote BioRS instance, cross-references can be created locally for quick access to information in other integrated databanks. Without making copies, data is shared across different locations while maintaining both security and user-specific preferences.

User and group administration

System administrators can assign users to groups. Access to data is granted by assigning privileges to individual users and user groups.

Integration of other applications

It is possible to easily integrate other applications (e.g., export search results for further processing by external tools or for visualizing database entries). These integrated applications can be accessed from the Search Tool graphical user interface (GUI).

Software version These technical specifications apply to version 5.6 of the BioRS Integration and Retrieval System.

Architecture The BioRS Integration and Retrieval System is based on client–server architecture. Individual modules for searching, indexing and parsing are provided as services within a distributed computing environment. Communication between modules is based on Common Object Request Broker Architecture (CORBA®). This distributed programming approach allows operations to be performed on multiprocessor systems and on any number of workstations.

Each module is open for seamless integration of external applications. Application programming interfaces (API) are available for Perl, C++, Java™ and others.



Supported database formats Format-specification files are available for commonly used databases listed below. Other formats are supported: contact Biomax about databases, including proprietary data, not included in this list.

ACE	International Protein Index	Place
Androgen Receptor Gene Mutations	InterPro	PRINTS
BIND	Kabat	ProDom
Blocks	KEGG	PROSITE, PROSITEDOC
CATH	LocusLink	PubChem
COG	MEDLINE	RefSeq
dbSNP	NCBI EntrezGene	S/MARt DB
DIP	NCBI HomoloGene	SCOP
EMBL	NCBI NR_nucleotide	SWISS-PROT
ENSEMBL	NCBI NR_protein	TRANSCompel
ENZYME	NCBI Taxonomy	TRANSFAC
GenBank	NCBI UniGene	TRANSPATH
GENESEQ	OMIM	TREMBL
GenPept	PathoDB	UniRef
GOA	PDB	UniProt
HSSP	Pfam-A, Pfam-B	
HUGO	PIR-PSD	

Supported relational databases The following relational database management systems (RDBMS) are supported:

- Oracle (versions 8/9/10)
- MySQL (versions 4/5)

Database updates Updates of data are processed by the dedicated Biomax Database Update Tool. This tool facilitates the management of downloads as well as processing and indexing of data. The tool is closely integrated with the BioRS system.

Requirements for parsing Parsers are written in the proprietary Biomax Format-Specification Language, a high-level, compiled language, which uses a set of predefined terms. The syntax of the parsing rules follows the Yet Another Compiler-Compiler (YACC) grammar. The language uses an XML-like structure and is easy to use (see the manual for more information). If new parsers are to be compiled, the following are required:

- C++ compiler
- GNU make version 3.80
- Bison (GNU Project parser generator), version 1.75 or higher
- FLEX (Fast Lexical Analyzer Generator), version 2.5.4 or higher

Platforms and operating systems The following platforms and operating systems are supported:

- Personal computer (PC) running SuSE® Linux® version 8.2
- PC running SuSE Linux Enterprise Server version 9
- IBM® running AIX® version 5.2L
- Sun™ running Solaris™ version 10



Supported browsers The following browsers are supported:

- Internet Explorer, version 6
- Mozilla®, Firefox® version 1.5

Both Java and JavaScript must be enabled.

Hard disk size The size of the hard disk required depends on the databases that will be included. The following information is given to allow you to estimate the hard disk size you will need.

Flat-files (approximate sizes)

- EMBL, 300 gigabyte (GB, September 2006 version 88)
- GenBank, 300 GB (October 2006 version 156)
- Other databases, usually less than 300 GB

BioRS indices:

- Depending on the characteristics of the indexed database and the indices created, the index files usually require 100% of the flat-file size. (For example, for a flat-file requiring 150 GB, the corresponding indices would also require 150 GB. The *AllText* index would require an additional 140 GB.)

Note: when updating databanks, additional disk space will be required temporarily. The required space is the same as that needed for the corresponding indices.

Note: additional space may be required for incoming databases or temporary backups. Therefore, it is important to ensure there is always extra space available.

Processors and memory For all platforms, 1.5 GB random access memory (RAM) are required for each indexer. The following processors are the minimum required for the indicated computers:

- x86/AMD64 or x86/EM64T processor
- UltraSparc™ IV for Sun Solaris
- Power5 IBM Power PC

Note: when processing and indexing large databanks that consist of several flat-files, it is strongly recommended to use several processors or several clustered computers in parallel. For example, at least six processors should be used when indexing the GENBANK or EMBL databanks. When large databanks are to be searched or there are many users, it is possible to increase the number of parallel threads or to start multiple searches on several clustered computers.

Bandwidth When using several computers in a cluster for indexing databases, the bandwidth of the network could be a bottleneck. Therefore, using a gigabit network for the databanks server is strongly recommended.

Biomax and BioRS are registered trademarks of Biomax Informatics AG in Germany and other countries. The BioRS software was developed in collaboration with RingRows GmbH, Martinsried. Oracle is a registered trademark of Oracle Corporation. MySQL is a trademark of MySQL AB. AIX, and IBM are registered trademarks of IBM. CORBA is a registered trademark of the Object Management Group, Inc. Java, Solaris, Sun, and UltraSparc are trademarks of Sun Microsystems, Inc. Linux is a registered trademark of Linus Torvalds. Mozilla and Firefox are registered trademarks of The Mozilla Organization. SuSE is a registered trademark of SuSE AG. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law. BIORSTPR0703

