
BioRS™ Integration and Retrieval System

Technical Profile

1. Release information

This *Technical Profile* is current as of December 2010. It covers the major release version 5.8 of the BioRS™ Integration and Retrieval System. BioRS v5.8 is a full public release. For more release information pertaining to this version, see the BioRS Integration and Retrieval System *Release Notes*.

2. Software features

2.1. 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.

2.2. 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 Oracle® and MySQL® relational databases 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 (RDBMS) can be accessed directly for searching. Alternatively, indices of relational data can be accessed to speed up searching.

2.3. 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.

2.4. 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.

2.5. 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).

2.6. Supported database formats

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

Table 1. Supported database formats

| | | |
|----------------------------------|-----------------------------|---------------------|
| 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 | |

3. 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.

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

4. System requirements

4.1. Supported relational database management systems

This BioRS version supports the following relational database management systems (RDBMS):

- Oracle database version 8 or higher
- MySQL version 4 or 5

4.2. 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.

4.3. 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 follow the Yet Another Compiler-Compiler (YACC) grammar. The language uses an XML-like structure and is easy to use (see the *YACC 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

4.4. Supported client platforms and operating systems

This BioRS version can run on almost any platform on modern personal computers (PCs). In particular, the following platforms and operating systems (OS) are supported by this version:

- PC running Microsoft® Windows®
- Apple® Macintosh®
- PC running Linux® systems

The Admin Tool requires Java version 5 or 6. Java version 1.4 is deprecated.

4.5. Supported client browsers

This BioRS version requires the following browsers:

- Internet Explorer, version 6 or higher
- Mozilla® Firefox® version 2.0 or higher; JavaScript™ must be enabled

Several available viewers for displaying databank entries require that the Adobe® Flash® Player plug-in (version 7 or higher) be installed.

The Search Tool requires the following browsers:

- Internet Explorer version 6 or higher
- Mozilla Firefox version 3 or higher
- Safari® browser

4.6. Supported server platforms and operating systems

This BioRS version supports the following server platforms:

- SUSE® Linux Enterprise Server version 10 or 11 (x86 and x86_64 architectures)
- Red Hat® Enterprise Server version 5 or 6 (x86 and x86_64 architectures)

4.7. 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.

The hard-disk size required for flat-files can be estimated using the following as approximate sizes:

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

The hard-disk size required for BioRS indices can be estimated as follows:

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.

4.8. 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® PowerPC®

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.

4.9. 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.

5. Contact Biomax

For more information about the BioRS system, contact Biomax:

Biomax Informatics AG
Robert-Koch-Str. 2
D-82152 Planegg
Germany
Tel: +49 89 895574-0
Fax: +49 89 895574-825
Email: info@biomax.com
Website: www.biomax.com/biors



Copyright 2011 Biomax Informatics AG

Biomax and BioRS are registered trademarks of Biomax Informatics AG in Germany and other countries. Oracle, Java, JavaScript, MySQL, Solaris, Sun and Ultrasparc are trademarks of Oracle Corporation. Perl is a registered trademark of the Perl Foundation. Microsoft and Windows are registered trademarks of Microsoft Corporation. Apple, Macintosh and Safari are registered trademarks of Apple Inc. Linux is a registered trademark of Linus Torvalds. Firefox and Mozilla are registered trademarks of The Mozilla Foundation. Adobe and Flash are registered trademarks of Adobe Systems Incorporated. SUSE is a registered trademark of Novell, Inc. Red Hat is a registered trademark of Red Hat, Inc. IBM and PowerPC are registered trademarks of IBM Corporation. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law. The BioRS software was developed in collaboration with RingRows GmbH, Germany.

Biomax Informatics AG; Sitz: Martinsried bei München, Amtsgericht München, HRB 134442; Ust-IdNr. DE 191 603 142; Steuer-Nr.: 143 / 100 / 10945; Vorstand: Dr. Klaus Heumann; Vorsitzender des Aufsichtsrates: Prof. Dr. Hans-Werner Mewes BIORSTPR1107