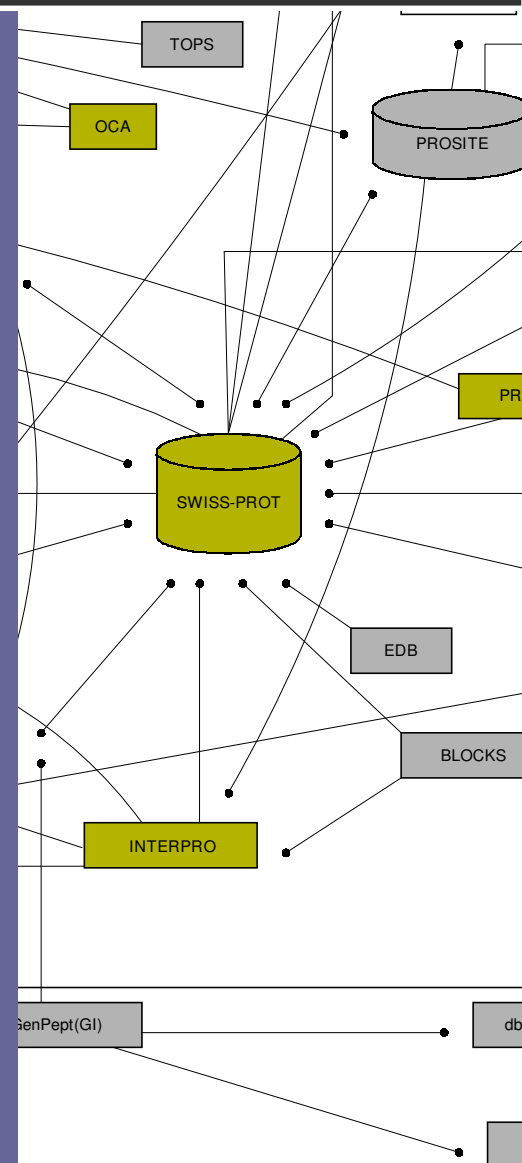


XMELD

XML-Based Data Integration Platform for the Life Sciences

- Comprehensive XML-based data integration platform
- The power and openness of XML: the data-format of choice in pharmaceutical and clinical research
- Sophisticated graphical query and search
- Standards-based data representation and schemas
- Ideal for enterprise research, as well as academic research and education
- Suitable for installations of all sizes



Xmeld: Data Integration for the Life Sciences

Datasheet

One of the most significant requirements underlying modern life sciences research is systematic integration, organization, and clear visual representation of complex research and enterprise data: Data that is dispersed throughout the enterprise, scattered across disparate platforms, compiled within specialized applications and databases.

The Xmeld platform for Data Integration provides a compelling data-integration technology devised for data-driven drug discovery, clinical trials analyses and knowledge management for life sciences and pharmaceutical research.

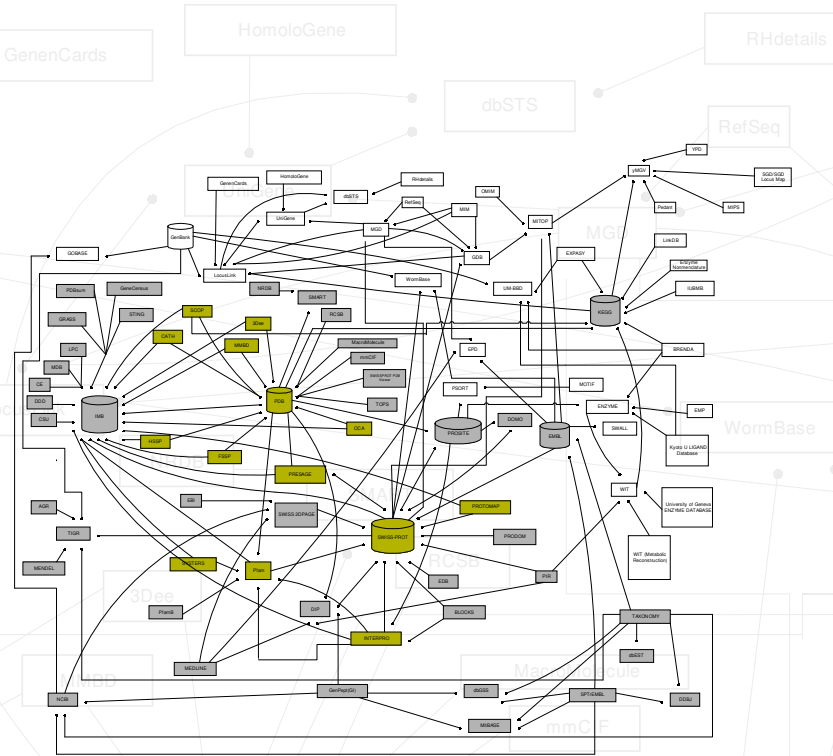
Xmeld allows customers to integrate disparate and heterogeneous forms of data to provide the basis for a holistic approach to research and a systemic foundation for business intelligence and data mining. Xmeld relies on its ability to put together physically disparate sources of data and then allowing end-users to conduct single-point queries across this vast pool of data, as though it were a single resource.

This solution provides the basis for integrating any form of complex, unstructured research data. Xmeld assumes that such data is largely heterogeneous, physically disparate and available off different platforms (flat files, ftp sites, legacy database systems etc.).

Besides integrating existing enterprise data, this solution also allows the end users to create and maintain their own databases on an ongoing basis.

A single point, expressive query language ensures that a variety of queries can be implemented across these data sources to pinpoint to data that matters, seamlessly.

By representing the data consistently and logically using today's most legible data-format, XML, this solution makes is conducive for productive and collaborative R&D.



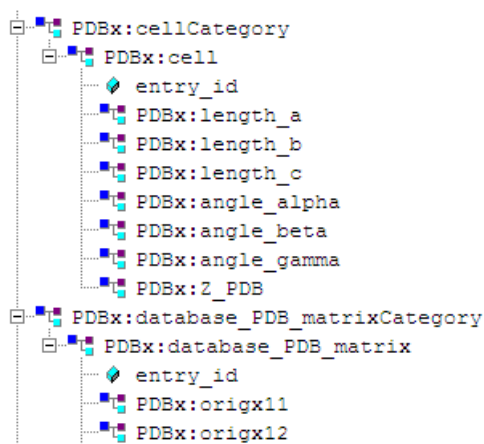
Why use XML to represent data?

XML opens the doors to clear and appropriate data representation. This functionality arises from the ability to use XML to define the structure of elements and data and the inter-relationship of data and data elements. Such a structure allows developers to create 2D representations of chemical and biological structures using data from a number of sources or to represent complex scientific data in a graphical format (see Fig.1).

The utilization of XML creates powerful opportunities for developers in sectors such as pharmaceutical, biotechnology, genomics etc. where complex data is increasingly interpreted visually. This data representation functionality can be fully captured and rendered by problem-specific applications.

What makes XML Data integration attractive?

Most database interactions are defined in a particular programming language, so that results are returned in that language's native data types. While Standard Query Language (SQL) is a widely accepted method for specifying the database question, there is not yet a language-independent and database-neutral specification for the response. XML affords an opportunity to provide this. XML provides an open source solution for data migration between programming languages. This proves highly attractive in a sector like drug discovery and clinical studies characterized by a large number of heterogeneous applications using a number of programming languages to derive complex data in a number of formats.



Why now?

Virtually all-large and medium sized biopharmaceutical companies are using XML, primarily driven by the need to rapidly and cost-effectively integrate data from a number of sources. Eighty three percent of a survey population is using XML¹. Ninety seven percent expect to be using it by 2004, in effect making XML an integral part of all applications and databases used by life science organizations.

Users across clinical trials, drug discovery and regulatory submissions are becoming more familiar and comfortable with the technology and by the efforts of various data-representation initiatives such as CDISC (Clinical Data Interchange Standards Consortium) and to raise awareness of XML and to create an industry-wide standard over the web.

The FDA in the United States has already made known its desire to use XML as the underlying technology for data related to evaluating drug approval application^{2,3}, thereby matching the Agency's strategic direction for data submissions and aligning with other industry initiatives such as the Operational Data Model (ODM) defined by CDISC.

¹ Survey, June 2002, "XML in pharmaceutical research and development processes", *Silico Research publication titled "eR&D Insights: Analysis of computing strategies in the life science research and development processes"*

² Stockbridge, N., *FDA Proposed ECG Interchange Standards*, November 19th, 2001, <http://www.fda.gov/cder/regulatory/ersr/ECGdata.htm>

³ Lipicky, R., Stockbridge, N., et. al., *Proposed Presentations for the Meeting on 19 November 2001*, <http://www.fda.gov/cder/regulatory/ersr/ECGpresentations.htm>

Is this an XML database?

No. Xmeld is a middleware that uses an underlying, select enterprise relational and XML databases to physically store the data that it integrates. This allows enterprise customers to have immediate access to the wealth of relational database management tools, such as backup, recovery and mirroring.

Custom Packages based on this middleware

Genvea also provides dedicated and fully integrated workbenches for data-driven drug discovery (Drug Discovery Enterprise Edition, abbreviated as DDEE) and agricultural informatics research (Agricultural Genome Informatics Workbench), *in silico* vaccine design and development (Vaccinome Informatics Workbench) that utilize this middleware to integrate relevant databases.

In the case of the DDEE, for example, this includes relevant genome sequences, including the complete annotated sequence database - Genbank, genomes of Homo Sapiens, model organisms and pathogens; protein databases, protein structure databases etc.



Life Sciences Data Integration with Xmeld

Technical Specifications

Database Types Supported

- Flat file formats
- XML file formats
- Relational database management systems (RDBMS)
- Integration of both local and remote data

Tools and Features

- Advanced XML-data integration engine – Xmeld, combined with graphical ease of use
- Standards-based query and search tools
- Powerful, graphical query interface to correlate across Genome, Protein and Structure data
- Construct complex (“impossible”) queries using graphical tools
- Simplified Form based query tools for scientists
- Consistent, XML-representation of all data

Verified Data Specifications

- Genbank specification from National Center for Biotechnology Information (NCBI), USA
- Swiss-prot, TrEMBL and InterPro XML specifications from European Bioinformatics Institute (EBI)
- PDBML/XML specification from Research Collaboratory for Structural Bioinformatics (RCSB)
- CDISC and HL7 formats for representation of data from clinical trials
- FASTA, multiple FASTA formats for sequence data
- MAGE-ML import and export for microarray experiment data
- Configurable XML formats for data export

System Features

- Genvea Biosciences Reference Architecture for optimal query performance
- Secure Client-server access

Programming and Enabling Tools

- Programmable data middleware
- W3C XQuery language for data query
- Optional integrated high-performance workflow tool, Quascale
- Graphical XQuery construction, XML visualization, and Form-Query construction tools

Hardware and System Support

- Support for Redhat Linux and Sun Solaris operating systems
- System optimization for maximum-performance
- “Data-Bank in an Appliance”, *Genomatique*, is now available with six key public repositories pre-integrated

Support and Services

- Online documentation
- System monitoring, flexible real-time notification
- Proactive system management for increased system availability

Training

- On-site one-time training of scientific staff and IT support personnel

Functional consultancy (Optional)

- On-site domain consultancy on integrating further databases and for integrating Xmeld into custom research workflows

Genvea Biosciences
53, Craig Road,
#04-01,
Singapore 089691

Chemoinformatics
20, Jalan SS2/66
47300 Petaling Jaya
Selangor Darul Ehsan

Phone: +65.62224569
Email: info@Genvea.com

Phone: +60.3.78760022
Email: info@chemoinformatics.com

©2003 Genvea Biosciences Pte Ltd. All rights reserved. Specifications subject to change without notice. Genvea Biosciences and the Genvea Biosciences logo are registered trademarks of Genvea Biosciences Pte Ltd in Singapore and/or other countries worldwide. All other trademarks mentioned herein are the property of their respective owners. Adapted image “Colored TEM of coronaviruses” courtesy of Dr Steve Patterson / Science Photo Library; Adapted image “Colored TEM of a cluster of corona viruses” courtesy of Dr Gopal Murti / Science Photo Library; Adapted image “Coronavirus” courtesy of Russell Kightley / Science Photo Library;

