Open-Source Based LIMS Systems in Biotech

Helge Weissig, Ph.D. • helgew@activx.com
Project Leader, Bioinformatics
ActivX Biosciences, Inc.

Outline

- Definition and uses of LIM Systems
- LIMS Requirements at ActivX
- Commercially available systems
- Open-source, GPL licensed systems
- The Velocity/Turbine/Torque framework
- Overview of the ActivX LIMS
- Conclusions
- Credits & Acknowledgements
Laboratory Information Management Systems

- Computerized systems to track, manage, and facilitate processing of samples and data and integrate data analysis
- Pre-LIMS
  - Notebooks, spreadsheets, files, folders…
  - Experimental data in various, incompatible formats
  - Data analysis limited to single experiments
  - Access privileges difficult to enforce

Typical LIMS Features

- Logistics management
  - Data storage and maintenance
  - Inventory tracking
  - Ownership audit trails
- Designed for laboratory use
  - Barcode tracking of samples
  - Workflow tracking (SOP)
  - Interface to instrumentation for data collection/import
- Integrated Data Analysis tools
ActivX Biosciences, Inc.

- Drug discovery & development company
- Unique protein activity analysis technology
- Founded in April of 2000
- Currently 40 employees
- Internal drug development program with focus on oncology
- Collaborations with pharmaceutical industry

LIMS Requirements at ActivX

- **Versatility**
  - Three major technology platform components
  - Several types of data
  - Need to integrate with commercial and public database contents
- **Scalability**
  - Initial storage requirements expected to grow greatly in the future
- **Maintainability**
Specific Needs - Data Storage

- **Proteomics data**
  - Protein identification by mass spectroscopy
  - Quantification by gel & capillary electrophoresis and mass spec
- **Medicinal chemistry data**
  - Commercial chemical compound database
  - Activity data for candidate drug molecules
  - Inventory of chemicals
- **Biological Sample data**
  - Tracking & categorization of samples
  - Treatment information

Specific Needs - User Access

- **Simple and intuitive user interfaces**
- **Web based access**
  - Platform independence
  - Maintainability with fewer resources
- **Project based access privileges**
- **User authentication**
- **Data & analysis result export**
Specific Needs - Analysis

• Cross-platform integration
  • Data capture from various instruments
• Cross-experiment integration
  • Data capture from commercial software packages
• Backend high-performance access for CPU/memory intensive analysis

Commercial LIMS Systems

• Many vendors
  • http://www.limsourc.com/
• No complete “off-the-shelf” product
• Very few open-source products
  • Essential for in-house adaptation & maintenance
• High license and support costs
Open-Source, GPL LIMS

- **General**
  - WIFlow - Generic workflow, pre-alpha
  - Gnosis(the LIMS) - python/C, planning stage
  - FreeLIMS - coming soon
- **Application Specific**
  - HalX - Structural Genomics, alpha
  - AMAD, ArrayDB, BASE… - Microarray

ActivX In-House Solution

- Use of mature, open-source & free components
- Tiered system architecture
- Relational database backend
- Data model driven business layer
- Template & MVC based front end
Tiered Framework Architecture

- **Velocity**: Template engine, Java-based MVC implementations
- **turbine**: Servlet framework, secure, web application layer
- **torque**: Persistence layer, peer based Object relation model
- **MySQL**: Relational database, storage system

Open-Source Throughout

- The Apache Jakarta Project
  - Turbine & Velocity
- The Apache DB Project
  - Torque

http://www.mysql.com
Data and Software Integration

- **Data import**
  - Legacy data
  - Commercial software output
  - NT Domain authentication
- **Data Analysis tools**
  - MatLab
  - gnuplot
  - R package
  - PtPlot/JFreeGraph

Open-Source!

Open-Source Advantages

- Massive influx of ideas & contributions
- Flexibility
- No “surprises”
- Reliability
- Able to modify it as needed
- “Push it forward” mentality of community
- Large support base for popular packages
- IP of open-source based system with developer
Open-Source Disadvantages

- Support not guaranteed
- Unpredictable development cycles
- (FDA validation requirements)

Information Flow Overview

- Tissues, cell lines & biological samples
- Probes
- Inhibitors
- Robotic Sample Prep
- Xhibit™
- Xceed™
- Xsite™
- Compute Cluster
- LIMS
- Robot Status & Control
Conceptual LIMS Sections

- Chemistry
- Biology
- Xhibit™ - High-throughput gel electrophoresis
- Xsite™ - MS based peptide ID & quantification
- Xceed™ - High-throughput capillary electrophoresis
- Data Analysis
- Compound Screen
- Inventory
R package
Conclusions

- Unique requirements for LIM Systems in Biotech
- Open-source framework models available
- Ease of development & integration
- Need for programming expertise in group

Acknowledgements

- Joe Zou, Ph.D.
- Baohong Zhang, Ph.D.
- Xinghai Chen
- Jeff Leathers
- ActivX Biosciences, Inc.