The Need for Scientific Data Annotation

Alick K. Law, Ph.D., M.B.A.
Marketing Manager
IBM Life Sciences
alaw@us.ibm.com
Cross disciplinary research approach requires organizations to address diverse needs.

**Drug Discovery Process**

**Disciplines**
- Biology and Molecular Biology
- Chemistry and Molecular Chemistry

**Key Stages**
- Research
- Target discovery and validation
- Drug Design

**Key Activities**
- Gene Sequencing/Mapping
- Gene Expression Analysis And Functional Genomics
- Structural Biology/Protein Structure Determination
- Rational Drug Design
- Lead Identification, Validation and Optimization
- Pharmacogenomics
- Clinical Trials

**Business Challenges**
- Managing Target Diversity
- Maintaining Quality Control
- Increasing Scale
- Improving Fail Fast Strategy
- Accelerating Time To Market

**Discovery Solutions**
- Data integration tools
- Data management tools
- Data analysis tools
- Knowledge management tools
- Workflow / business process management tools

Source: Vaxa Study (March '01); Pharma Univ. Pre-Meeting (Oct. 02); Discovery Community Workshop (Jan '03)
Typical data integration scenarios

- **Genomics**
  - Discover all information relevant to a nucleotide or protein sequence from a variety of public domain sources

- **Cheminformatics**
  - Join information from structured repositories via Oracle cartridges for structure-activity relationships (SAR); absorption, dispersion, metabolism and excretion (ADME) and other applications

- **Data Analysis**
  - Integrate local scientific data (e.g. Excel), library and warehouse data for high-performance computation
Integration with Chemical Informatics
Information Integration

- Decision Support
  - Analyze, Discover & Measure

- Collaborate & Share
  - Manage, share, leverage knowledge

- Adaptive Workflow
  - Domain-specific protocols

- Integrated Applications
  - Web Services Enabled Computational Components

- Integrated Data
  - SQL Enabled Semantically Harmonious Information

Scientific Data Annotation
Discovery research organizations can utilize scientific data annotations to:

- Quickly identify related activities throughout the research organization
- Avoid repeating previously failed research efforts
- Preserve knowledge and expertise across the organization
- Create an enterprise-wide system to capture the decision-making process
- Identify subsets of data for further analysis based on annotation content along with qualifying information
Facts belong with data, insights are missing knowledge.

<table>
<thead>
<tr>
<th>Unstructured</th>
<th>Structured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structure yields common input and searchability</td>
<td></td>
</tr>
</tbody>
</table>

Annotation Use Value:
- **Facts**
  - Unstructured
  - Structure yields common input and searchability
- **Insights**
  - Structured
  - Highest Value

Facts belong with data, insights are missing knowledge.
Annotations of scientific applications

- Provide a knowledge capture mechanism to track the progression of compounds in a project from “active” status to “hit” status

- Enable selection of additional targets or anti-targets to screen to assure selectivity of the compound chosen for advancement
Where in the Drug Discovery Process is annotation valuable?

IBM InsightLink™ provides cross-platform annotation capability to generate, capture and harvest new insights across the organization.
The Drug Discovery Process is highly complex and costly. Duplication of effort occurs quite frequently, especially in global organizations with scientists located in laboratories distant from one another. Annotation provides an effective way to reduce non-productive activities, shorten the time from early discovery to development, and provide an overall enhancement to the process. Enterprise-wide access to annotations, including searching and retrieval of the annotation context along with the related original data results in a very powerful tool.
For more information

- IBM Life Sciences –
  http://www.ibm.com/solutions/lifesciences

- IBM Life Sciences Drug Discovery Solutions -

- IBM DiscoveryLink® -

- IBM InsightLink™ -