



POWERING *the process of* INVENTION

# INTEGRATION IN THE 21<sup>ST</sup>-CENTURY ENTERPRISE

Thomas Blackadar  
American Chemical Society Meeting  
New York, September 10, 2003



# THE INTEGRATION BILL OF RIGHTS

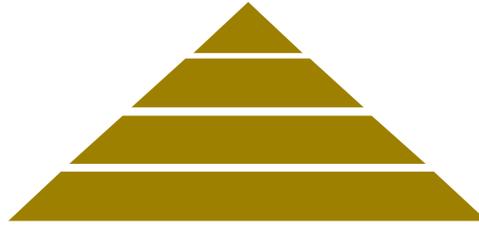
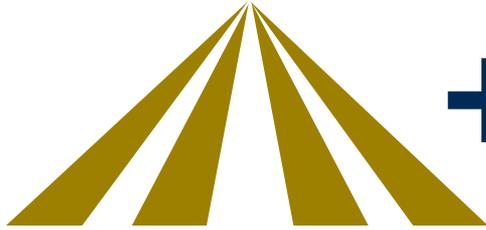
Integrate = to form, coordinate, or blend into a functioning or unified whole...

The right information in the right format with the right context to the right people with the right tools at the right time



# DELIVERING THE BILL OF RIGHTS

- ❑ Right information – comprehensive data capture, archiving, search, delivery, and retrieval systems – for public and proprietary data
- ❑ Right format – standards, common and enforced vocabularies, ontologies
- ❑ Right context – agreed thesauri, annotation tools, metadata for experiments and protocols
- ❑ Right people – validated and secure access to all appropriate team members, managers
- ❑ Right tools – easy- and ready-to-use tools for analysis, visualization, reporting, what-if scenarios
- ❑ Right time – up-to-date, good response times



## ***Barriers along***

- Functions
- Geography
- Organizations

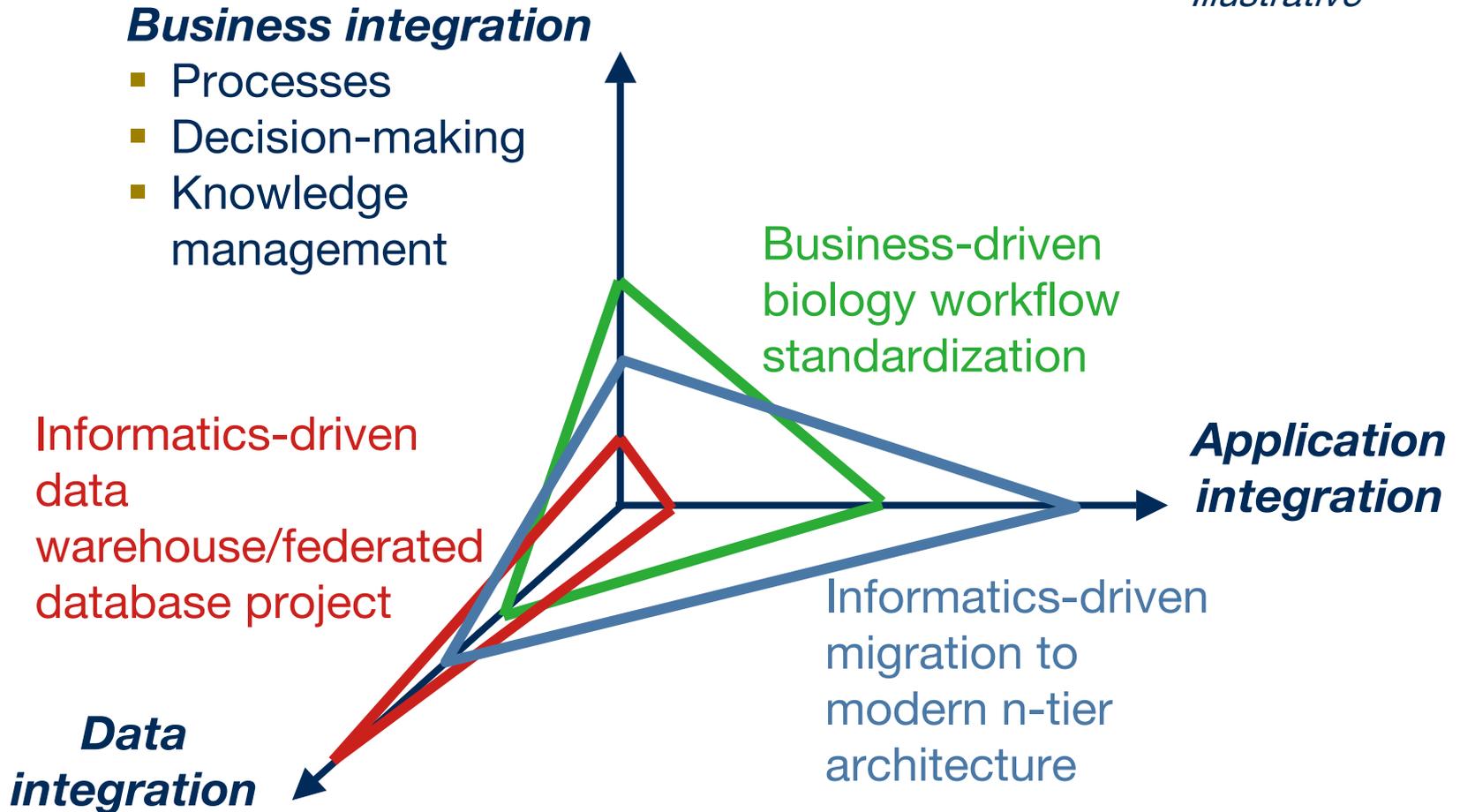
## ***Variations along***

- Processes
- Technology
- Taxonomy

***Information islands  
limiting effective  
and efficient  
information sharing***

- 
- ***Complex workflows integrating multiple project teams***
  - ***Information sharing across traditional disciplinary boundaries***
  - ***Globally dispersed organizations***
  - ***Increased volume and complexity of data***

*Illustrative*



- ❑ Warehousing – Data is processed into an integrated format for searching
  - Data model may be application-specific
  - Performance and concurrency issues
- ❑ Federation – Data is kept separate but searching is coordinated
  - Complex system architecture
  - General case is almost unsolvable

*Both approaches have their places and neither can deliver everything*



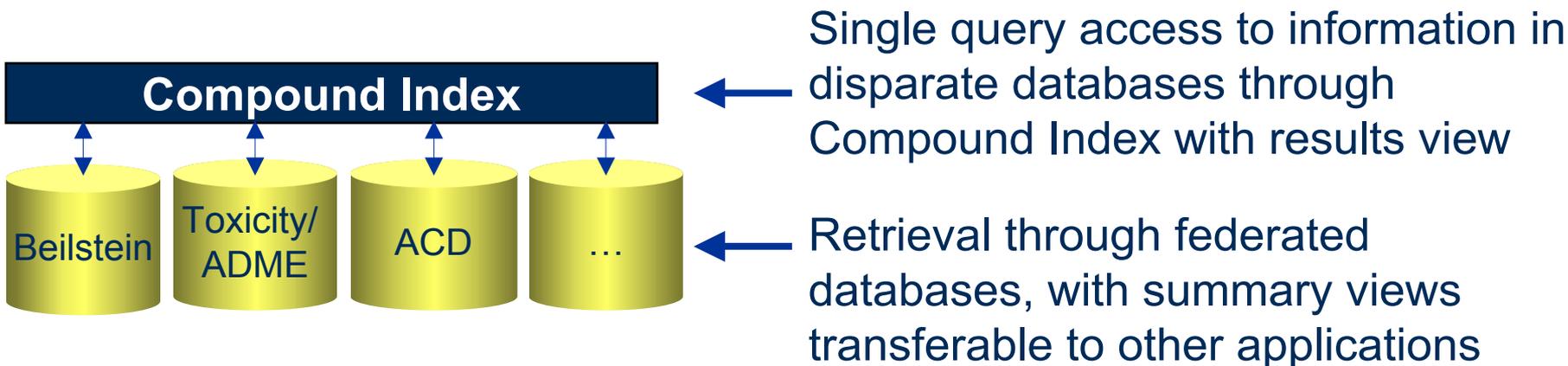
## DiscoveryGate

- Single query access to primary, secondary and tertiary literature
- Accessible through Internet
- MDL Compound Locator with access to
  - ✓ Over 12 million compounds
  - ✓ Over 250 million facts

## MDL Compound Locator

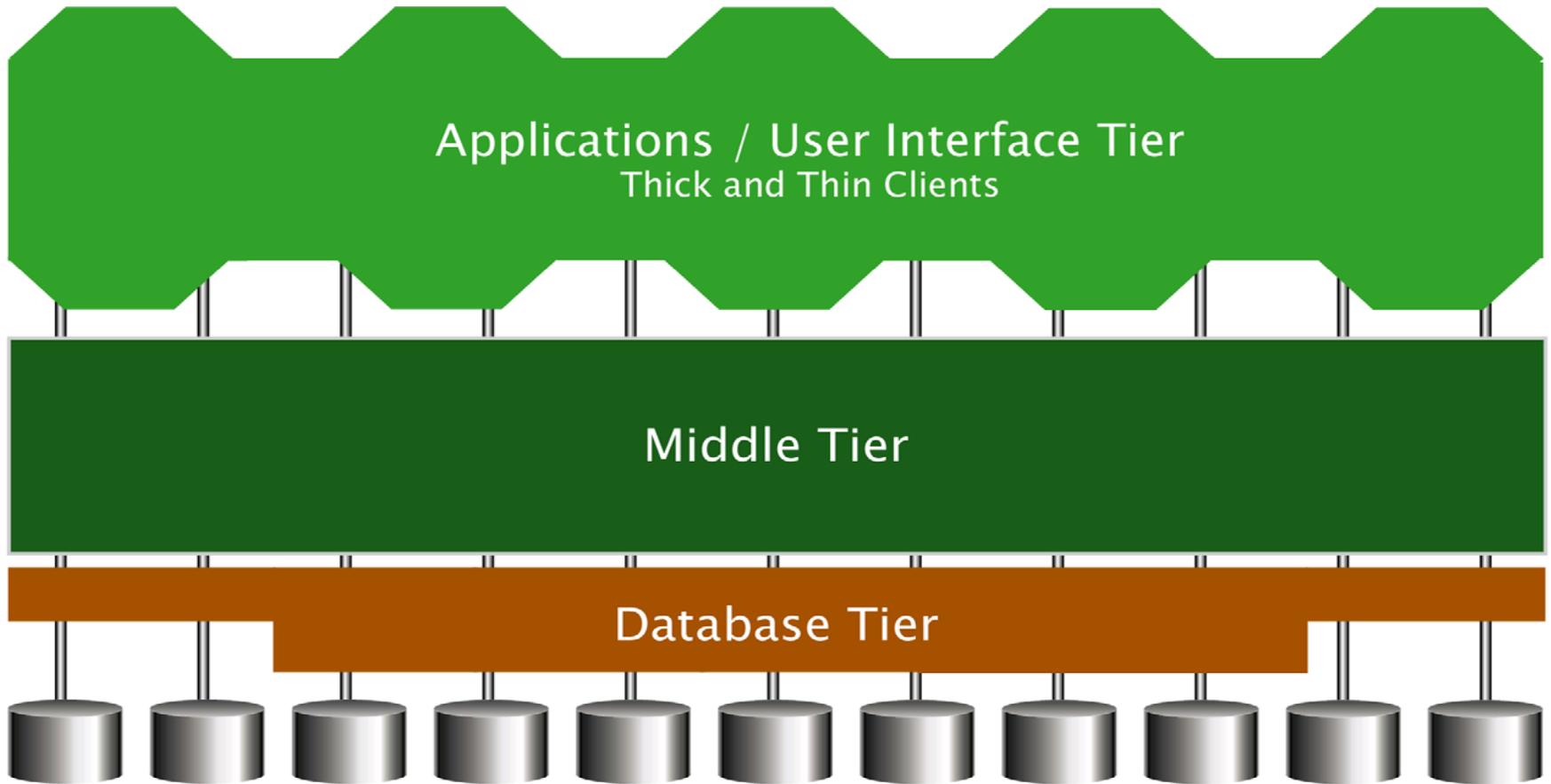
- Concordance table of normalized chemical structures
- Information about original sources
- Set of relational tables to store metadata describing type of data available from original data source (Compound Index)

## Technology behind MDL Compound Locator/DiscoveryGate



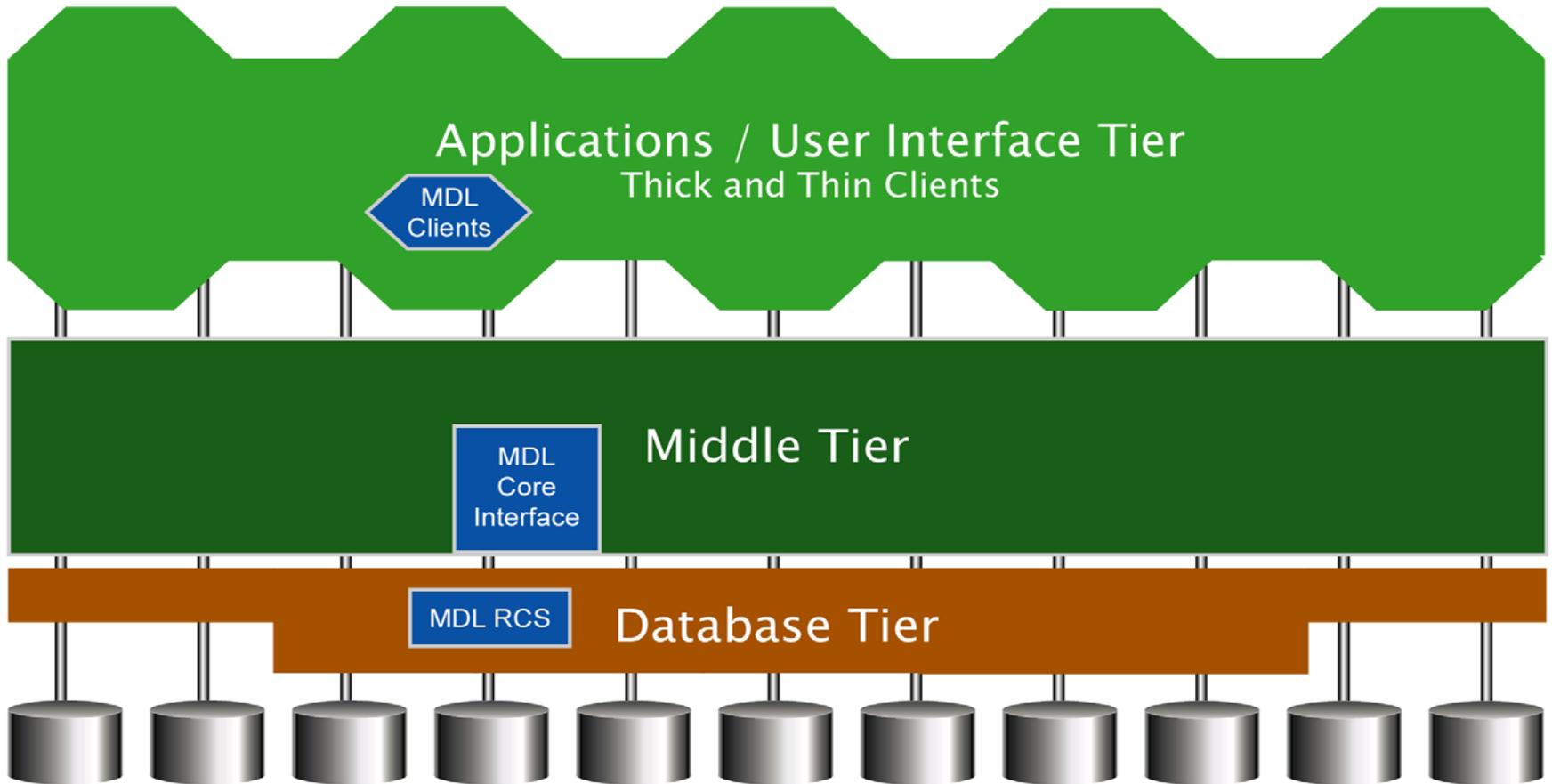
- ❑ General move to three-tier models
  - UI tier: presentation and user interaction
  - Middle tier: standard and specialized services, data integration, business rules and logic
  - Database tier: standard and specialized data stores
  
- ❑ Multiple flavors available from horizontal and vertical vendors: mix and match best of breed
  - Horizontal vendors provide enterprise-wide integration tools
  - Vertical vendors provide scientific added value

# MULTI-TIER MODEL IN GENERAL





# A VENDOR'S PLACE IN THE MULTI-TIER MODEL



## ❑ 1 – Compatibility

- Products work together but are relatively independent

## ❑ 2 – Linkage

- Products need access to the same data or objects
- Products must work together as one

## ❑ 3 – Adoption

- One product or standard is so broad or powerful that everyone uses it
- One product is always the basis for the other



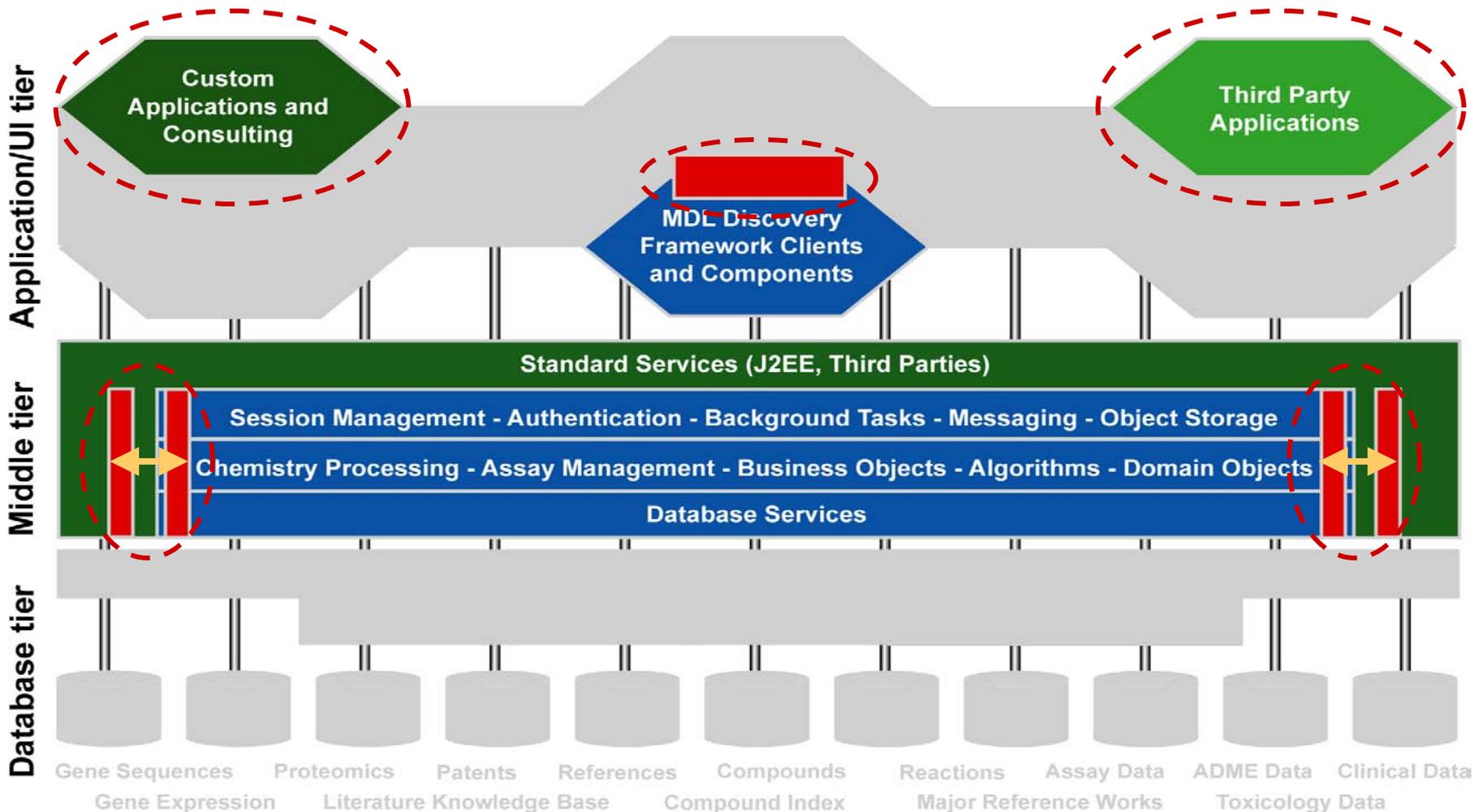
## LEVEL I - COMPATIBILITY

- ❑ Systems need to be presented together to the user, but require relatively little linkage
- ❑ Examples: Portals, application servers, developer tools, many current applications
- ❑ Approach: Use a common technology
- ❑ Advantages:
  - Easy development and maintenance
  - Parts can be updated or replaced independently
  - User feels like systems are integrated
- ❑ Disadvantages:
  - Cost to develop connections each time
  - Little cross-system information sharing
  - Overlapping technologies may need to be used

- ❑ Systems must act as one, sharing data or objects
- ❑ Examples: authentication, data source connections
- ❑ Approach: Write an adapter that allows one system to see the other's data or objects
- ❑ Advantages:
  - Done once, works for all related applications
  - Integrated for both users and system
  - Works even without globally accepted standard
- ❑ Disadvantages:
  - Cost to develop linkage once
  - Overlapping technologies

- ❑ System is expected to be built on top of a global standard
- ❑ Examples: Oracle database, Windows, J2EE
- ❑ Approach: Build on top of the standard technology
- ❑ Advantages:
  - Functionality built only once
  - Highest level of interoperability
  - Broad developer support
- ❑ Disadvantages:
  - Lack of choice – we must really agree on a universal standard
  - Constraints on functionality

# HOW DOES THIS WORK IN PRACTICE?



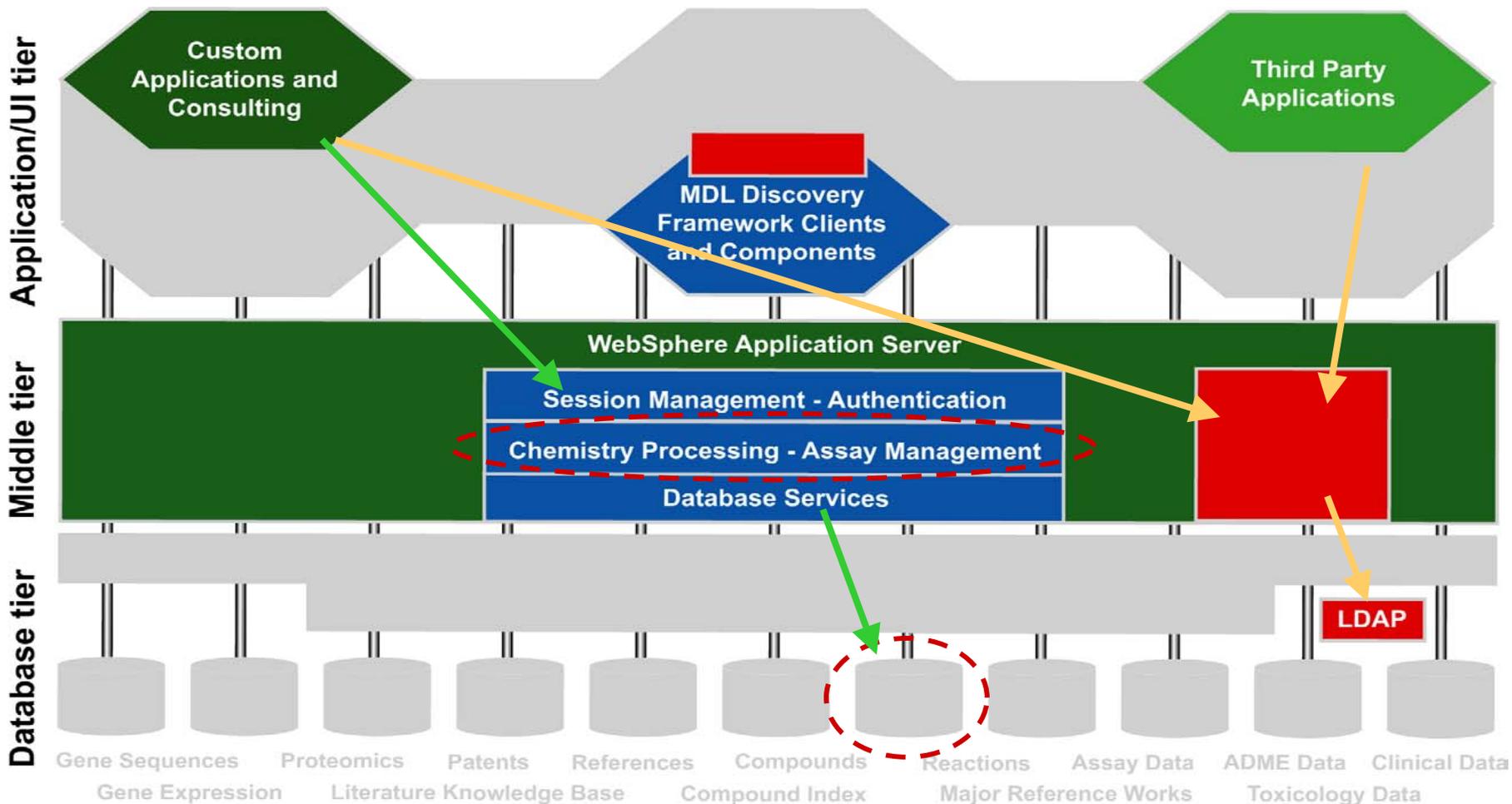


# EXAMPLE: MDL<sup>®</sup> ISENTRIS<sup>™</sup> AND LDAP AUTHENTICATION

- ❑ A large organization already has a user authentication system
  - Commercial LDAP directory server
- ❑ Organization uses IBM WebSphere application server as their application development standard
- ❑ Organization wants a reaction-based lab journal
  - Molecule and reaction tables in Oracle
  - Integrated chemistry and biology data
  - Centralized registration component
  - Custom enumeration service
  - New business rules must be available to other applications developed by vendors and in-house staff

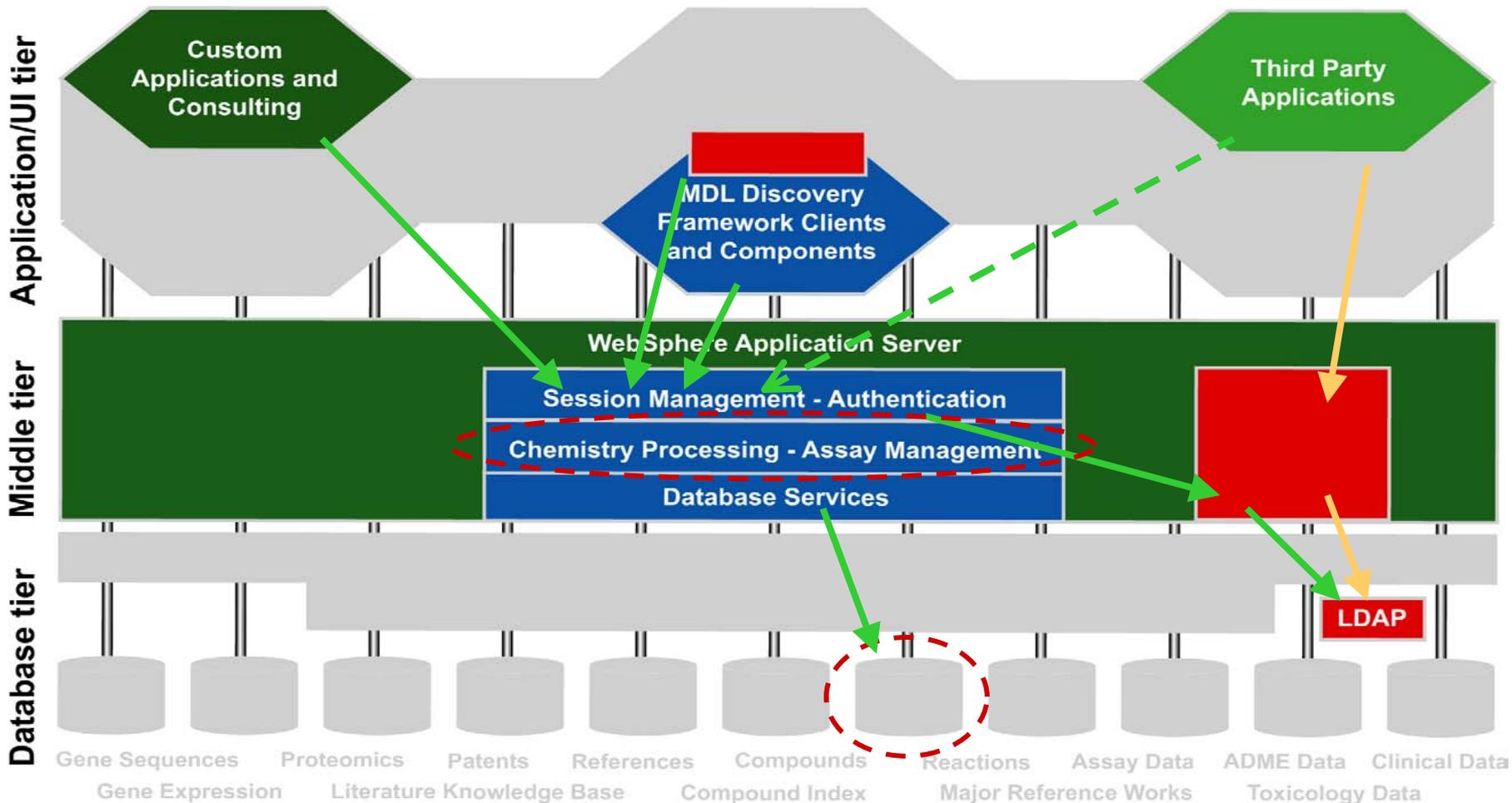


# EXAMPLE: MDL<sup>®</sup> ISENTRIS<sup>™</sup> AND LDAP AUTHENTICATION





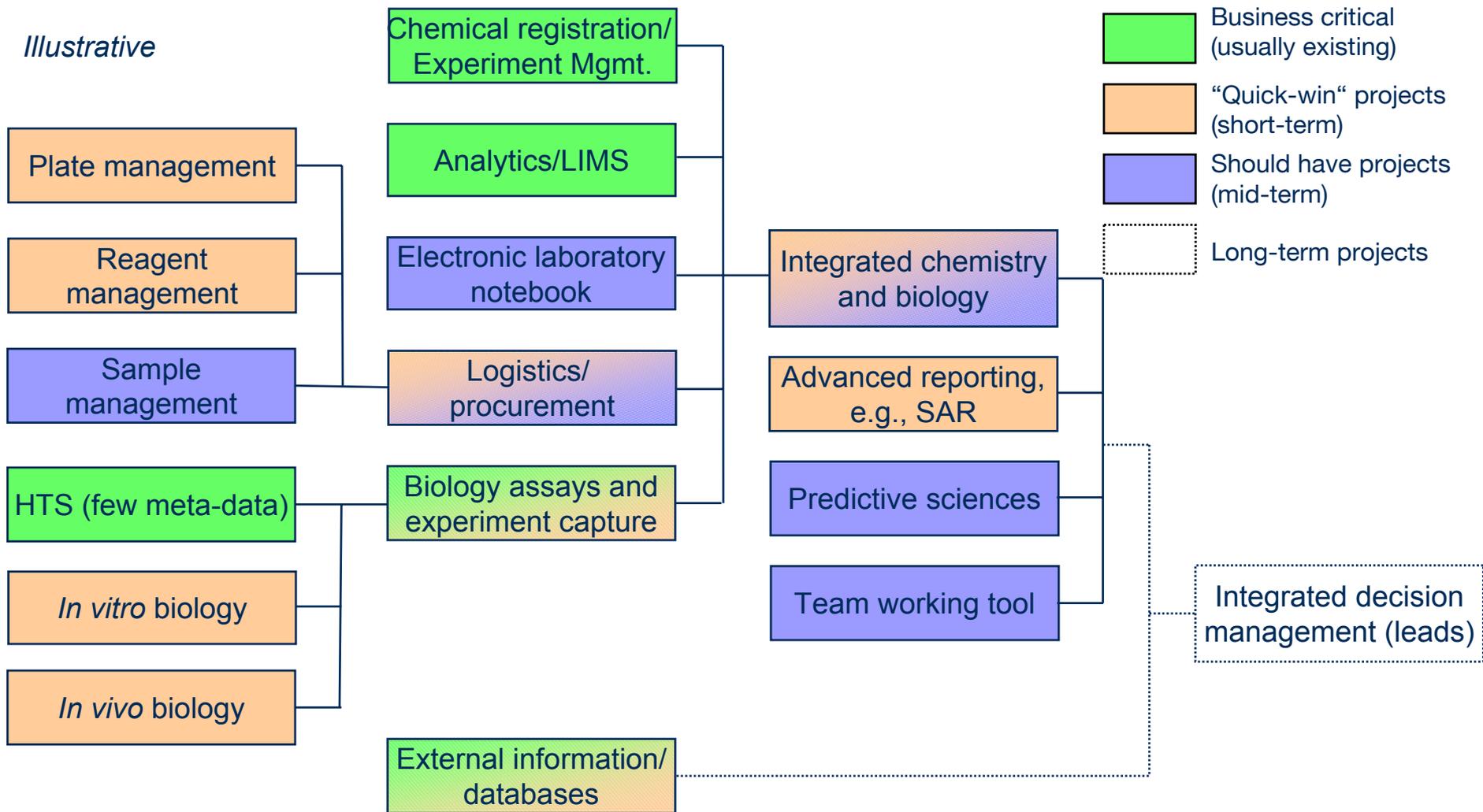
# EXAMPLE: MDL<sup>®</sup> ISENTRIS<sup>™</sup> AND LDAP AUTHENTICATION



- ❑ Vertical-market technologies will always have some overlap with other systems
  - But each system may be different
  - And some organizations may have no system
- ❑ Only use a specific middle tier where it makes sense
- ❑ Where a middle tier does make sense:
  - It should be readily linked with your own system
  - Done once, this linkage enables custom, third-party, and commercial applications to work

- ❑ Data integration alone will not deliver
  - Lack of agreed standards, formats, vocabularies, access methods can hamper integration, navigation, access, and performance
- ❑ Application integration alone will not deliver
  - Separate applications in different disciplines need a top-down mandate to integrate
- ❑ Business integration will deliver only if built solidly on the other two
  - Federation and warehousing as appropriate
  - A rational combination of standards-based systems from multiple sources

*Illustrative*





# SUMMARY: DELIVERING THE BILL OF RIGHTS

- ❑ Right information – data management strategy lets you know what you know
- ❑ Right format – standards and connections make data usable
- ❑ Right context – makes data meaningful
- ❑ Right people – systems get data to those who should have access and only those
- ❑ Right tools – applications let scientists focus on science, not bookkeeping
- ❑ Right time – fast access to current knowledge leads to crisper decision making