“Connecting the Dots with Dynamically Linked Data from Diverse Sources”
Webcast Agenda

• **Problem**: Cost and complexity of distributed data integration (multiple types, multiple sources)

• **Solution**: *TopBraid Insight* -- data virtualization and federated query

• **Explanation**: Semantic Technologies are uniquely suited to solving the Problem

• Discussion and Questions
Healthcare/Clinical Research/Life Sciences face:

• A Biomedical Information Tsunami
  – Overwhelming volume of data
  – Multitude of sources, types, etc.

AND

• An Informatics Tower of Babel
  – Each community speaks its own “dialect”
  – Integration critical to achieve promise of “personalized” medicine
The TMC
(example domain)

21st-Century Data Integration across the Translational Medicine Continuum (multiple stakeholders multiple systems, multiple contexts*)

* “Re-purposing” of data requires separation of context from data/meta-data, i.e. explicit representation of context-specific meta-data.

Used by permission: National Cancer Institute Cancer Bioinformatics Grid (caBIG)
Connect the dots for new insights. Ease Big Data Variety

- Right Insights
- Right People
- Right Time

- Dynamic Interactive Exploration
- Supports the Logical Data Warehouse approach
  - Flexible, Adaptive Information Structuring
So. You’ve got the data. You’ve got good people. You’ve got systems.

What’s the problem?

Big Data and Legacy Information Structures – Bad match

- Adds costly overhead
- Stresses “brittle” operational and IT systems
- Can’t know questions that enable effective data structuring
- Ties up key personal and organizational “bandwidth
- Obscures key patterns and critical information

Doubles Annually.

(infrastructure, financial)

(infrastructure)

(human, infrastructure)

(human)

(human, opportunity)
Real world problems are solved using insights built from the components of meaning – questions, concepts, perspectives.

Topbraid Insight offers a layer of connection and meaning to the user and protects them from the distracting mechanics of computerized data access.
Find all the 3-way valves across all vehicles that correspond to the valves in this vehicle that are showing intermittent malfunctions at this point in checkout since we changed to this new supplier and the associated change orders and work authorizations.

Ontologies allow the meaning of data to be expressed so that data can be related across databases with different schemas.

Data is in different places with no simple way to achieve integration.
Big Data Variety - The IT Problem

The Typical Physical Data Warehouse solution:

- Bio activity of small molecules
- Phys/Chem properties
- Derwent Patent Index
- Sample Test Results

ETL

Chembl
Chebi
DrugBank
Data virtualization with a unified query interface

Real time, on-demand datamarts

Documents/IP

Physical Data Warehouse

Sample Test Results

Any source

Documents/IP

Physical Data Warehouse

Sample Test Results

Any source

Documents/IP

Physical Data Warehouse

Sample Test Results

Any source
A Virtual Data Warehouse with a unified query interface
**Key Concepts**

**ConnectSet** - integration of several data sources

- Identifies data to be included in a Logical Warehouse and how it connects (maps) together.
- TBI server can provide multiple ConnectSets

**ExploreSpace** – an exploration space.

- Dynamically created by a user based on a ConnectSet for a specific data exploration and investigation.
- Unlimited number of ExploreSpaces can be created based on a ConnectSet.
Results: “Things” and “Relationships”

For example:

- Shakespeare wrote Play
- Play hasName Hamlet
Results: Our Demo – Drug Related Results

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

SideEffect

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug

Drug
Results: Random Info vs. Key Patterns
TopBraid Insight™ (TBI): Connects the Dots

• Tames Big Data to empower businesses
• Offers on-demand integrated access to diverse data, making it possible to discover information just in time
• Delivers new levels of creativity and infrastructure flexibility
What’s Different about this Approach to Data Integration

• **Build incrementally. Change easily.** Handle Volume, Velocity and Variety. Adapt to changes and unanticipated needs

• **Better Value** - Accomplishes what statistics do with lots of time and expense on limited or stale data. Users interact directly with rich data. Enables what they do now but faster, better, easier and less expense.

• **New perspectives** - We provide a true 360 degree view. Less Limitation and Bias – Discover patterns, questions rather than picking in advance

• **No PhD in informatics needed** Enable a broader group to explore and discover and collaborate. Reduce “bottleneck” dependence on expensive skills and info-middlemen.

• **Develops valuable semantic assets** (Workspaces) that can be seen, used and shared for collaboration
The Problem results from...

- Multiple data sources
- Multiple stakeholders
- Multiple data-collection contexts
- Multiple data-usage scenarios
“Connecting the dots” means “machine understanding the semantics” of the data/meta-data being “connected”

Data integration (by machines) means \textit{sharing semantics}

\[ \text{Two (or more) systems given same data/meta-data will produce the same results when performing the same function} \]

There are no silver bullets or secret sauces...
What’s different about “semantic technologies”?

• Semantics as a “first-class citizen”

• Remove the “non-semantic” barriers to integration
  – RDBMS table structures
  – XML document hierarchies
  – Vendor-specific implementations
Connecting the dots

• The “secret sauce” in TopBraid Insight isn’t really so secret.

• TopBraid Insight uses W3C semantic standards

• Aligning semantics isn’t simple...

• TBI’s application of semantic technologies make it as easy...

• ...and when it’s done, it opens the door to insights that were previously difficult, time consuming, and expensive to achieve.
In Conclusion ...

The whole is greater than the sum of its parts

Aristotle

...Questions?
“A wonderful harmony is created when we join together the seemingly unconnected.”

- Heraclitus