Managers and decision makers in modern enterprises, both private and public, need timely and actionable information (business intelligence) to accomplish their goals and tasks.

Often this actionable information must be gathered and merged from disparate data sources. Combining the needs of business intelligence and information interoperability, all in “real-time” (i.e., without incurring IT labor), presents a huge challenge to most large organizations today.

The need for merging disparate information in real-time is especially acute in government agencies. For example, the Environmental Protection Agency is charged with monitoring and governing actions by private enterprise that could possibly harm the common water supply or the air quality. Being able to collect and merge information from hundreds of databases, across the federal, state and tribal level, is of vital importance to the EPA in discharging its governance role effectively.

To cope with this need to exchange and merge information across disparate databases, the EPA and other federal agencies have adopted an ambitious plan whereby various “database owners” (e.g., state agencies in particular governance areas such as Water Quality) are strongly encouraged to “map” the particular tables and columns in their databases to a set of Data Standards, implemented through standard XML tags.

This approach is problematic for several reasons:

1) Many “legacy” systems were put into use before the Data Standards were published (January 2006), and were not designed with XML compliance mapping in mind
2) Mapping individual tables and columns to a given Data Standard set of XML files is difficult, do to the wide variance in database design (a “standard” data element may consist of the values in a single column in one database, but could be an aggregation of multiple values in another database)
3) There are many overlapping Data Standards
4) Many state and tribal agencies don’t have IT personnel on staff who are trained in the level of data modeling and XML, or don’t have a budget for such compliance mapping

The real problem is the need for mapping entities and attributes sharing the same meaning, even if the table structures are not the same. In essence, mapping information through a common meaning is a “semantic problem” more than a database problem. According to Dave Hollander, CTO of Contivo:

“The real impediment to interoperability is semantics – analysts estimate semantics may account for up to 95% of the effort. ”

The rest of this paper outlines a new “semantic technology” approach that can result in huge productivity gains in both Business Intelligence and information integration areas.

1. Business Intelligence (BI).

Today’s systems and databases are designed with ‘built-in IT labor components’: they require the intervention and assistance of high-priced IT professionals to construct and deliver even simple reports and answers to their business users. Thus these labor-
intensive architectures are ill-designed for getting actionable information to executives and analysts – the vast majority of the BI community. The IT labor bottleneck in BI illustrated below comes from the findings of a report by Giga Research:

End User Market: 86%  
Technical User Market: 14%

The present constructs of data and metadata do not allow information to be accessed directly by end users through natural language requests. This is because databases do not 'understand' the meaning of natural language requests: their metadata lacks semantic or conceptual information of the database content¹.

Instead, database metadata is designed only to understand database query languages, such as Structured Query Language (SQL), and their scripting language derivatives (stored procedures).

In effect, Semantic Technology (ST) brings exactly what is needed to BI: the ability to “understand” the meaning of users’ common language requests, and automatically generate the database query to deliver an immediate answer to the user.

The ideal ST solution combines two new technologies to solve BI problems: ontology and natural language database query generation. These two technologies are discussed below.

Ontology.

Ontology, in a nutshell, is a set of conceptual objects (entities, attributes, entity relationships and business rules) that describe a single area of user interest. Ontologies are composed of Conceptual Objects' (COs), organized in a hierarchy of subject areas and conceptual graphs. Some (but not all) of these COs are mapped to tables and columns in enterprise databases. Enterprise ontologies are business-related domains and subject areas. Even more specifically, corporate ontologies represent the set of subject areas comprising a specific corporation's enterprise model.

¹ This is often called the “semantic gap” between users and databases.
One implementation of enterprise and corporate ontologies is OntoloNet™. In OntoloNet, a set of base ontologies are constructed for abstract domains such as general business, government, sports, etc. These base ontologies capture general knowledge of an abstract domain in the form of entities and their relationships, business rules, and common terminology (terms that the “man on the street” can understand).

Stacked on top of these base ontologies are more specific concept models, describing a particular industry or subject area domain. On top of this “concept stack”, specific concept models are constructed which closely mirror targeted organizational databases. The metadata of an organization’s subject area concept models become its enterprise ontology stack. Then a short process of 'semantifying' the subject area ontologies is conducted by which an organization's subject area experts (SAEs) add organization-specific sentence phrases (internal lingo and acronyms) to OntoloNet’s lexicon.

A basic linguistic design of semantic technology is to supply semantic phrases at each ontology level. At the most basic 'Common IO' level, semantic phrases describe the common word patterns and phrases by which one can ask, for example, about a person's residence, such as "where does ((person)) live?" and "((person)) resides at ((residence address))". At the general business ontology level, semantic phrases refer to 'business information objects' such as sales revenue, gross profit, etc.

Since an employee "is-a" person, the listener subconsciously maps these lexical objects between the layers of conceptual compartments stored away in the brain, and quickly makes sense of the question.

Corporate ontologies can be viewed as a “Corporate Intelligence Stack”: a knowledge repository of the entities, business rules and semantics organized by the organization’s subject area Lines of Business. The primary benefit for utilizing corporate ontologies in Business Intelligence is to provide end users information from corporate databases without IT people having to intercede to write reports and queries for them. Below is a graphic representing a corporate ontology for a health care company:
By ‘semantifying’ each subject area ontology at its level in the hierarchy, new client databases can be quickly semantified without having to train the semantic lexicon every time from scratch.

- Mapping ontologies to relational databases

There is a close commonality of ontology and conceptual models (entities, entity relationships, entity attributes). And logical data models (also called entity relationship diagrams or ERDs) are closely related to both conceptual models and to relational database metadata (assuming a normalized database). Therefore it is possible to map a subject area’s conceptual / logical model elements to its 'target database' physical database elements (tables, columns, relationships, primary and foreign keys).
Natural language database query generation.

The second major technology employed by ST is natural language database query generation. This is really a “round trip automation” process that integrates several components: a Semantic NLP (natural language processor), a knowledgebase of ontologies, a database query generator and an “answer formatter”. The end user's natural language information request is parsed ('understood') in the context of enterprise ontologies, and an SQL query is generated automatically. The answer-set from the SQL command is automatically formatted and returned to the user.

Natural language query parsing requires both upper-level and lower-level ontology. An upper-level terminological ontology, such as WordNet, recognizes phrases and word parts other than just nouns. Lower-level subject area ontologies discriminate meaning according to the context and purpose of the natural language inquiry.

Semantic technology architects saw that if a natural language user request could be 'understood' by a parser in the context of a subject area ontology's Concept Objects and its Business Rules, then the 'last mile' of automating the query would be to automatically generate an SQL query against the database tables and columns which map to the ontological COs.

Below is a graphic that illustrates the ‘round trip automation’ of a natural language database query generator patented by Marvin Elder. The graphic illustrates a real-world example where a user types in a request in natural language: “Total seafood product orders”. This sentence dynamically generates an SQL command (shown at the right), which then is executed against a corporate database. The result set is captured and automatically formatted as a “default report”.

A solution that converts natural language to database queries through ontologies illustrates how semantic technology can dramatically boost IT productivity.
2. Information Integration.

Today's web-centric, on-demand business model imposes a need for a new real-time database architecture that facilitates systems interoperability across disparate systems and databases. Due to the independent design of these disparate information systems, a new set of “semantic disconnect” problems has emerged.

To take full advantage of real-time information architecture, these semantic problems must be overcome. How large a problem is this need for semantic methodologies in systems interoperability? According to one semantics expert,

“Leading analysts have estimated that 35-65% of System Integration costs are due to Semantic issues. And in every sector of the market, ... our biggest software challenges come down to creating and resolving meaning. In other words: semantics.” [Mills-Davis, 2006]

Semantic technology allows disparate data sources to be interrelated through a novel approach that the author calls “conceptual joins™”. This semantic solution is far superior to the conventional data warehouse approach: it is much easier and faster to map individual databases to subject area concepts and to “concept views”. These conceptual objects are already inter-related in a hierarchy of enterprise ontologies.

One implementation of conceptual joins is Marvin Elder's Cohesive Intelligence™ solution, which employs a distributed natural language processor, coupled with its networked OntoloNet repository, deployed on multiple disparate databases.

This Cohesive Intelligence solution allows a common language request to be multicast to targeted data sources, where logic and inferences in individual Corporate Intelligence Stacks can determine whether the disparate data source can contribute any facts being requested. The individual facts are visible to an analyst (or a system-wide query) across the corporate enterprise, then collated and merged into a single answer.

Below is an example of a single natural language request that automatically generates individual SQL query results on thousands of disparate databases to be merged together.

“Find Persons who attended an engineering school and who lives in the Southwest and who ...”

One plain language query is distributed to hundreds of ‘hub’ systems, each of which distribute to hundreds of ‘spoke’ systems. Net distribution: thousands of systems simultaneously executing dynamically generated SQL.
Summary.

IT professionals are interested in approaches and products that best serve their corporate constituents. Sometimes a new approach may reach "outside the box" of computer science historically used as the foundation for Information Technology. This is especially true when new business models introduce problems that are insoluble with conventional IT methodologies. Such is the case with today's global business model: Semantic Technology is better equipped to solve the new types of semantic problems that come with the new business model.

For BI use, Semantic Technology (ST) is already being used to generate precise answers for end users requesting information from corporate databases, through the use of automated round-trip query and report automation tools.

For interoperability of disparate systems and data sources, ST is even more compelling because of the high content of semantic disconnects between disparate information sources.

From a technical design standpoint, semantic technology methods and solutions, coupled with existing IT technology, illustrate how a difficult problem (full-cycle query automation) can be solved by incorporating and combining technologies from many science disciplines.

Author's bio:

Marvin Elder is the founder and Real Time Answers™, a company utilizing Semantic Technology solutions for the new real-time, on-demand global enterprise model. He and his company have patents pending for several claims in natural language database queries and in a novel method of conceptual joins over disparate data sources.

Mr. Elder has over thirty years of experience in data architecture, IT management, and IT consulting. Mr. Elder's research in relational theory and natural language spans over twenty-five years.

Mr. Elder is also a successful software inventor. Among his inventions were:

- one of the first real-time wireless business applications (ready-mix concrete dispatching and order entry),
- the first PC-based fourth generation language (4GL) with built-in natural language (Salvo) -- a feat for which he was voted Future Computing's Software Entrepreneur of the year,
- a CASE tool which included a built-in IDEF1X data modeling tool and which automatically generated ready-to-run relational database applications (Inroads),
- a business rules engine used for real-time pricing of yellow page ads, and
- a business rules engine used in Medicare / Medicaid billing in long-term health care.
- A Natural Language Business Intelligence and Analytics product, SemTrue™.

Mr. Elder holds an undergraduate degree in engineering management from the University of Texas (Austin) and an MBA degree from Southern Methodist University.