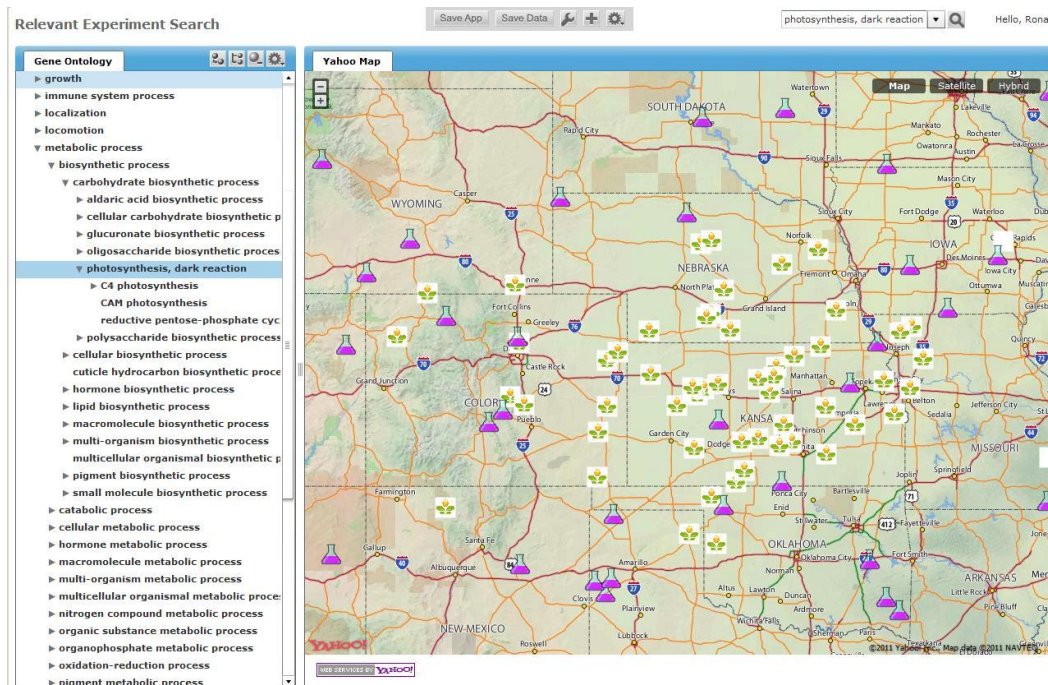


**Intent:** Cross-reference between multiple data sources, using public information to enhance understanding of proprietary information.

**Solution Story Example:**

## Investigation of Experimental Results

A company that manufactures pesticides and plant hybrids must innovate continuously in order to keep a competitive advantage. They regularly perform experiments on crops in the field to understand environmental impact on the plant. They receive several reports of a damaging blight to one of their crops; the blight impacts the "dark phase" photosynthesis function of the plant. This effect is not observed in the laboratory, only in the field. In order to understand the blight, they must perform experiments on plants in the affected areas. Setting up experiments of this sort is time-consuming and expensive. If they can find a relevant experiment that has already been performed, they can save time and money by avoiding unnecessary repetition.

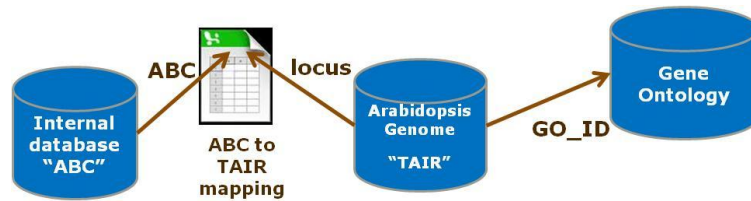


In this example, getting actionable information requires integration of several public data sources with data internal to the company. Integration was accomplished using TopBraid Suite.

The display above shows a hierarchical breakdown of metabolic processes (source: Gene Ontology). The processes in turn are associated with plant anatomy and genetics (source: public Arabidopsis gene data). The company has performed many experiments on genes associated with dark reaction photosynthesis (source: proprietary company data).

When a user clicks on an item in the Gene hierarchy, system runs a query to find all the field experiments performed on genes that are associated with functions that are related to dark-phase photosynthesis, where the location of the experiment was within five miles of one of the known sites of the studied blight.

The location of these experiments is then shown on the map (beaker icon), alongside the known locations of incidents of the blight (plant icon). Beakers near plant icons indicate a possible match - an experiment on an associated gene, already performed near a site of the blight.



The diagram above shows an overview of the public/private integration implemented in this solution. The internal private data about experiments is in a relational database denoted as “ABC”. Genome data (“TAIR”) is publicly available for download as a spreadsheet. It was automatically converted into RDF. Gene Ontology is a publicly available RDF/OWL ontology. Private ABC data and public Genome data use different types of values as their primary keys. A cross-reference between these keys was compiled as an Excel file. It was also automatically transformed into an RDF dataset. This resulted in three RDF datasets and one relational database, all accessed through a federated query.

---

**Business Benefits/Results:**

The value of internal information is increased when combined with pre-competitive data. Re-use of investigations (experimental results) is made possible by geographical cross-reference to data about known incidents (blights). The business saves time and money by locating important investigation information that has already been collected.

---

**Applicable Technologies and Products:**

TopBraid Suite™ provides a federated representation of information in RDF, including company data in a relational database, allowing the data to be queried using the SPARQL query language. The results are displayed using TopBraid Ensemble™ allowing to coordinate selected metabolic processes and the map display and to customize icons for ease of visualization.