

## The Semantic Web – An Overview

Extracting Information from Data

### Problem

Your organization has information from many related and unrelated sources. Your organization may be able to add some additional data to obtain the data you need for your investigation, however, you may have that information elsewhere in your organization in those dozens of stovepipe systems. It might even be available outside of your organization, and you may want to access the data -- if you know how to relate it to other data that you can access.

Now you have other problems. The structured data (spreadsheets, databases, etc.) among your various sources are varied and the data fields have similar names but are defined differently. You also have unstructured data (position and research papers, etc.) that lack any data tags that would enable you to say what is in the document, let alone be able to relate it to other data you can access.

Finally, you do not want to simply do a search and get a list of the unrelated data sources; you want those data sources to be related, so that when identifying an item of interest, the system will actually be able to expand on that item and investigate more thoroughly.

How do you access that data in a way that is productive? How do you manage the constant change that occurs in the applications that support these data sources? How do you manage dynamic data in a chaotic environment where data may be structured or unstructured?

### Semantic Web

Semantic Web provides the next generation of solutions for managing this environment. Semantic Web is based on ontologies, basically the relationships among the data. Based on the ontology, data is collected from the various sources, "semantisized" and stored into a Semantic Store for accessing. Ontologies are developed within the context of the environment using the data. The ontology is structured in a way that is compatible with its use. As a result, different fields of pursuit could require different ontologies. Also, since ontologies are dependent on the use of the data, the users define how the ontology is structured.

One of the advantages of this technology is that a full ontology need not be defined completely. It can be done incrementally so that its capability can be identified as experience is gained. Incremental development is highly recommended.

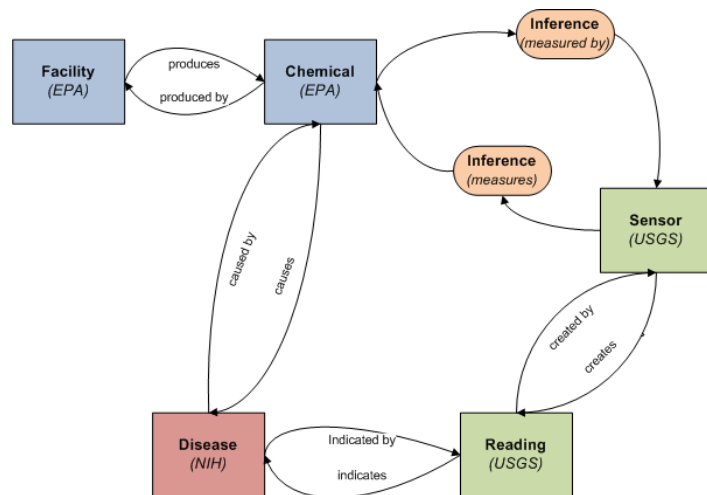


Figure 1: Semantic Ontology Example

## Semantic Web Operation & Implementation

While ontologies define the desired structure of the data, how does your data get added to the process? You still have all of these varied sources, in varied formats, with systems that seem to be in a constant state of change. You still have the structured and unstructured data.

### Roles & Resources

#### Data Sources & Access

##### Client

- Domain Experts
- Database Administrators
- Data Access

##### Intervise

- Data Alignment

### Semantic Data Organization

##### Client

- Domain Experts
- Existing Ontologies
- Existing Taxonomies

##### Intervise

- Develop Architecture
- Develop Ontology
- Implement Ontology
- Develop Interfaces to Data Sources (Input)
- Develop Data Interfaces to Data Requests (Output)
- Develop Technical Documentation

### Visualization (User Interface)

##### Client

- Define Visualizations of Data

##### Intervise

- Implement Visualizations
- Develop User Documentation

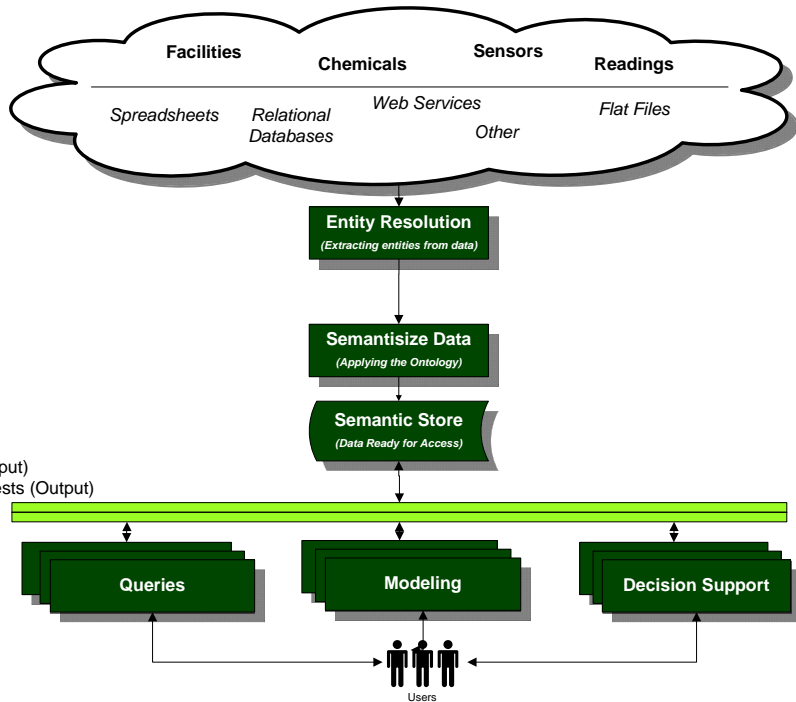


Figure 2: Semantic Web Operation & Implementation

Data mappings are done for the structured data. Data alignment is used for structured and unstructured data. It is possible to tag your text documents down to the paragraph or sentence so that the actual location of the reference can be identified. Where data cannot be directly related, inferences can be defined that will connect the data. Probabilities can be defined to provide the researcher or general user the likely relationship level. The semantizing of the data is loading the data and relating it according to the defined ontology. The source of the data is maintained with the data so that the researcher can obtain the entire context of the original data.

## Benefits

This technology is user centric. From how the information is organized to how the user sees it, the user is the primary center for design and implementation. The organization of this data can be developed and changed in much less time because the technology is not database centric. As a result, the technology is responsive to changes in the environment. The technology assists the user in accommodating masses of data from multiple sources, relating the data even when the relationships are not well defined.

## Contact Us

For an overview of the our Semantic Web services and more information on how Intervise can address your unique challenges and benefit your specific areas of interest, call (240) 599-9323 or contact Intervise at [jpriddy@intervise.com](mailto:jpriddy@intervise.com).