



GOVERNMENT COMPUTER NEWS

The online authority for government IT professionals

Oracle tackles the context conundrum

- By [Joab Jackson](#)
- Feb 14, 2008

The trouble with keeping data in a standard relational database is that such information does not typically come with machine-readable descriptions of what the data is. Searching through an Air Force database will not provide any indication that what you are looking at is an Air Force database.

As long as the database is used by its intended audience, this lack of database self-awareness is not a problem. But when another system needs to access the data, how will it make sense of the columns and rows of data? It is this semantic cluelessness that slows the process of gathering intelligence from the data.

At least one commercial database vendor is addressing this problem. Oracle is using two Semantic Web tools for the job: the Resource Definition Framework (RDF), support for which was added in Oracle 10g, and the Web Ontology Language (OWL), added in Oracle 11g.

RDF is the starting point.

'By storing RDF and applying rules to it, you can infer new information' and render explicit context about the data, said Xavier Lopez, director of spatial and semantic technologies at Oracle.

RDF offers the ability to link two data elements along with a term that describes the relationship between the two. The resulting three terms are called a triple. For instance, the database can ingest this statement: Lopez works at Oracle. 'Lopez' would be the subject, 'Oracle' the object and 'works at' the predicate tying together the two, according to the description of RDF from the World Wide Web Consortium (W3C), which oversees the framework ([GCN.com/960](http://gcn.com/960)).

But that's just the start. Having these relationships in machine-readable language allows further reasoning about the data 'making it machine-readable by other systems.

'Once you have triples in the database, you can start doing things you couldn't do before,' Lopez said. 'It is essentially designed to find patterns. Previously, it was available everywhere, but you couldn't find patterns. Now you can find patterns across it.'

Unlike the traditional data cubes used in data mining applications, the schema of the data does not need to be established beforehand, making ad hoc queries a lot easier to execute.

Lopez has seen customers compile billions of triples. After enough data has been rendered into this format, additional inferences can be made.

For instance, if you have 'Xavier works at Oracle' and 'Oracle is a software company' then an inference could be that Xavier works for a software company.

This is a simple example, but a logical step-by-step process can generate new information.

And this is where OWL comes in.

OWL extends the range of inferencing that can be done on a dataset, Lopez said. Another W3C standard ([GCN.com GCN.com/961](http://gcn.com/GCN.com/961)), OWL is a 'richer rule base,' he said. It offers sets of hierarchies that allow data to be described in terms of property characteristics, equalities and inequalities, data types, and restrictions on how the data can be defined.

Oracle has tools that can render standard relational data into the RDF format. And many software tools exist that parse unstructured data from Web sites, e-mail messages, blog sites and any other text-based documents into the RDF format.

About the Author

Joab Jackson is the senior technology editor for Government Computer News.



© 1996-2010 1105 Media, Inc. All Rights Reserved.