



IO to Include New Semantic Tools in Next Version of Sentient, Extends Working Group

[January 11, 2008]

By [Laurie Wiegler](#)

Life science data integration firm IO Informatics said this week that it is extending its Working Group on Semantic Applications for Hypothesis Generation in Translational Research — a collective was formed last July to develop semantic applications that may be used to generate hypotheses from translational research data.

The company said that the initiative has already resulted in a number of new semantic features — including the use of Unified Resource Identifiers and data sources that use the Research Description Framework, Web Ontology Language, and Notation 3 standards — that will appear in the next version of its Sentient software platform, which it plans to launch Feb. 14 of this year.

The expanded effort will run through March and will focus on developing a “thesaurus framework for terms and relationships” that will be able to integrate entities, ontologies, and relationships to provide “meaningful representation across disciplines,” the company said.

The working group is similar in concept to another best practices-sharing model, the Bio-IT Alliance, of which IO Informatics is a member [[BioInform 10-20-06](#)].

The IO Informatics group is smaller than the Microsoft-led Bio-IT Alliance, and is also distinguished by its focus on semantic technology.

Erich Gombocz, VP and CSO of IO Informatics and chair of the working group, told *BioInform* that the group includes six organizations represented by a total of eight individuals. He said the members have been meeting about every three weeks via web conferencing.

“Particularly, we focused on biomarker qualification and biomarker validation in the context of bringing together experts from industry and academic centers of excellence to understand the ... needs [of the bioinformatics community] in respect to reasoning with data from different ‘omics,” Gombocz said.

Members include Pat Hurban, executive director for scientific affairs at Cogenics; Jonas Almeida, professor of bioinformatics at the division of quantitative sciences at MD Anderson Cancer Center; Ted Slater, associate director of knowledge management informatics at the Indications and Pathways Center of Emphasis at Pfizer; Mark Wilkinson, assistant professor of medical genetics and bioinformatics at the James Hogg iCAPTURE Center; Bruce McManus,

director of the James Hogg iCAPTURE Center; and Alan Higgins, VP of preclinical development at Viamet, a pharmaceuticals firm.

Higgins told *BioInform* that Viamet was formed about two months ago and has been “focused on finding agents that interfere with ways in which cells use metals.”

“For the first time, data from different ‘omics’ disciplines like proteomics, genomics, metabolomics and the like [will have a] clinical outcome – [and] can be understood in the context of their biological functions.”

He said that the company decided to participate in the working group because it was already working with IO Informatics as part of an Advanced Technology Program grant to develop systems to analyze and integrate data and also look at gene expression through microarrays.

He said that Viamet was interested in expanding its use of IO’s Sentient Suite because it “allows you to combine different data sets ... without asking too many questions.”

Higgins said that the working group has helped Viamet “pull in other studies and say what’s common [among the members].”

Gombocz said that IO’s data-integration system provides a “secure, compliant, and auditable framework” that “really helps to accumulate and leverage the knowledge of the organization.”

The addition of semantic web technology to the suite will allow users to create “formal ontologies, formal taxonomies,” said Gombocz.

“The semantics aspect of the working group is that for the first time, data from different ‘omics’ disciplines like proteomics, genomics, metabolomics and the like [will have a] clinical outcome – [and] can be understood in the context of their biological functions.”

He added that the semantic technology allows users to create a “network view of the data based on their relationships to understand, for instance, what pharmacodynamics changes mean in respect to biological change.”

Gombocz said that the group decided to expand the scope of the project in order to develop more user-friendly features.

The first part of the initiative “well exceeded our expectations,” he said. Deliverables for that phase of the effort included “the direct integration of local data with web based URIs,” as well as “direct-to-knowledgebase function” and semantic data formats and a taxonomy of medical disease codes for hypothesis generation, he said.

He said the group discovered that these tools needed to be more user-friendly for people not focused on network analysis or biological networks, “so they could integrate data independent of the standard terminology they come from.”

Copyright © 2008 GenomeWeb LLC. All rights reserved.

