

2008

## Collaborative R01 with NCBO Semantics and Services Enabled Problem Solving Environment For Trypanosoma Cruzi

Amit P. Sheth

Wright State University - Main Campus, amit@sc.edu

Rick Tarleton

Prashant Doshi

Mark Musen

Natasha Noy

*See next page for additional authors*

Follow this and additional works at: <https://corescholar.libraries.wright.edu/knoesis>



Part of the [Bioinformatics Commons](#), [Communication Technology and New Media Commons](#), [Databases and Information Systems Commons](#), [OS and Networks Commons](#), and the [Science and Technology Studies Commons](#)

---

### Repository Citation

Sheth, A. P., Tarleton, R., Doshi, P., Musen, M., Noy, N., Sahoo, S. S., & Weatherly, D. B. (2008). Collaborative R01 with NCBO Semantics and Services Enabled Problem Solving Environment For Trypanosoma Cruzi. . <https://corescholar.libraries.wright.edu/knoesis/260>

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact [library-corescholar@wright.edu](mailto:library-corescholar@wright.edu).

---

**Authors**

Amit P. Sheth, Rick Tarleton, Prashant Doshi, Mark Musen, Natasha Noy, Satya S. Sahoo, and Daniel B. Weatherly

# Semantics and Services Enabled Problem Solving Environment For Trypanosoma Cruzi

Amit Sheth (PI), Rick Tarleton, Prashant Doshi, Mark Musen, Natasha Noy, Satya Sahoo, D. Brent Weatherly, Pablo Mendes

## Driving Biological Problem

### Trypanosoma Cruzi (T. cruzi):

T. cruzi is a protozoan parasite and a key causative agent of Chagas disease. Chagas afflicts 18 million people in Latin America leading to heart disease and sudden death.

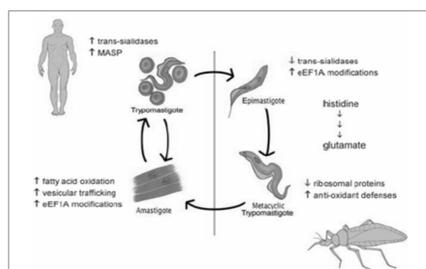


Figure: Whole organism – all life cycle stages of T. cruzi

Driving factors in T. cruzi research:

- Identification of vaccine candidates in T. cruzi
- Diagnostic techniques for identification of best antigens
- Identify genes for knockout in T. cruzi

Questions related to these factors involve tedious sifting of vast amounts of information of different types (genomic, proteomic, publications). This problem is prevalent for other organisms as well.

## Specific Objectives of Project

### Development and Use of a Suite of Ontologies Using a Phased Approach

Ontologies enable a semantic approach to information management and analysis. Phased approach promotes reuse of existing ontologies for developing new ones

### Semantic Annotation of Experimental Data Using Multiple Ontologies

Promotes integration of disparate data sources at multiple levels. Diverse ontologies will be aligned

### Efficient Query Processing, Visualization and Extensibility

Expressive querying on semantically annotated data using query builders and tools for intuitive visualization of results

### Semantics-driven PSE for T. cruzi hosted in BioPortal at NCBO

Generic framework that can be specialized for other pathogens. T. cruzi data available from the University of Georgia

## Preliminary Work

### T. cruzi Research:

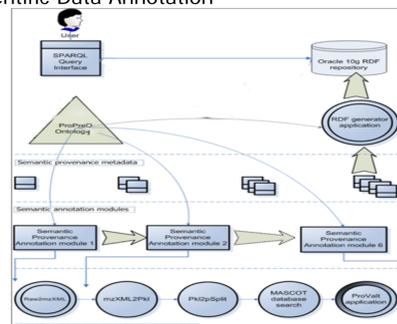
Large amounts of experimental genomic and proteomic data on T. cruzi has been produced. For example, proteomic data includes:

- Whole proteome analysis of 4 life cycle stages (600 files)
- Proteomics of glycoproteins from 3 life cycle stages (27 files)
- Quantitative proteome analysis of plasma membrane proteins from 3 life cycle stages (48 files)
- Proteomics of RNA binding proteins from 2 life cycle stages (24 files)

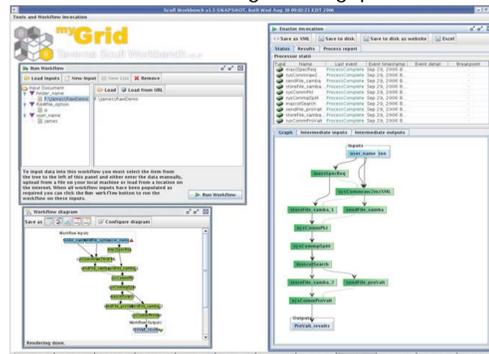
These data are from over 700 individual experiments in eight separate proteome projects. These analyses resulted in the accumulation of greater than 1 million tandem mass spectra and 12,000 individual peptide identifications.

### Semantic Bioinformatics for Glycoproteomics:

- Ontology Development and Alignment  
In previous projects we have developed 3 deep domain ontologies – Glyco, ProPreO and EnzyO. Participated in designing and developing tools for ontology engineering and alignment such as Protégé, PROMPT and OPTIMA
- Scientific Data Annotation



### Scientific Workflows for High Throughput Proteomics



## Proposed Research

### Problem Solving Environment for T. cruzi

- Intuitive querying of multiple sets of heterogeneous databases
- Formulate scientific workflows to test hypotheses
- Cognizance of usability issues to promote ease of use among biologists

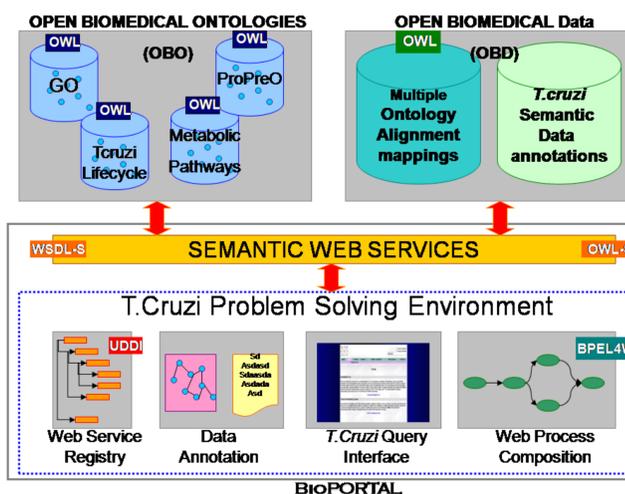


Figure: System architecture of the PSE

### Querying multiple sets of heterogeneous data

Involves:

- Developing new ontologies for T. cruzi lifecycle stages and using existing ontologies  
Extend ProPreO and Glyco
- Automatically aligning parts of different ontologies that target similar domains  
Extend PROMPT and integrate OPTIMA
- Semantic annotation of experimental data  
Automatic entity identification and disambiguation
- Extending query languages such as SPARQL

### Formulate Workflows For Hypotheses Testing

Involves:

- Development of semantic Web services  
Expose common operations such as BLAST as Web services  
Utilize SA-WSDL for describing Web services  
Utilize Web services within queries
- Techniques for automatically composing semantic Web services  
Extend Haley and investigate new approaches for composition

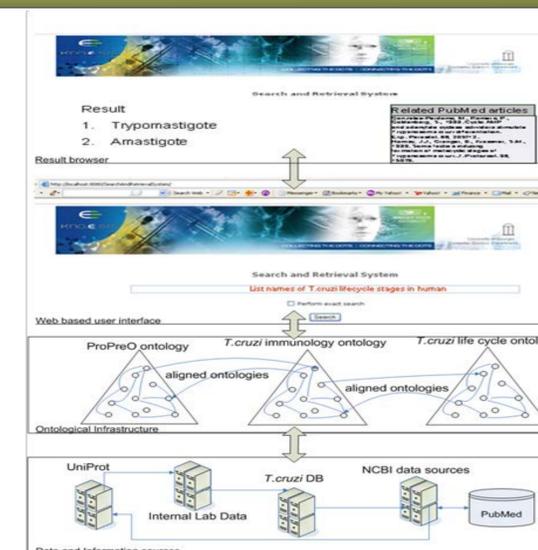


Figure: Schematic representation of the semantic T. cruzi PSE illustrating the simple Web based user interface, the underlying suite of ontologies and data/information sources

## Evaluation

### Validation

- Several example queries will be selected for validation
- Correctness of the answers provided by the PSE to the queries will be rigorously tested
- Usability and computation time will be qualitatively and quantitatively measured and compared with previous manually intense approaches

### Exploration

- New metabolic pathways will be investigated and auxiliary questions will be answered
- PSE will facilitate the discovery of new potential therapeutic and immunological targets within the T. cruzi biology

## Acknowledgment

This project is supported by grant R01HL087795 from NHLBI. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.