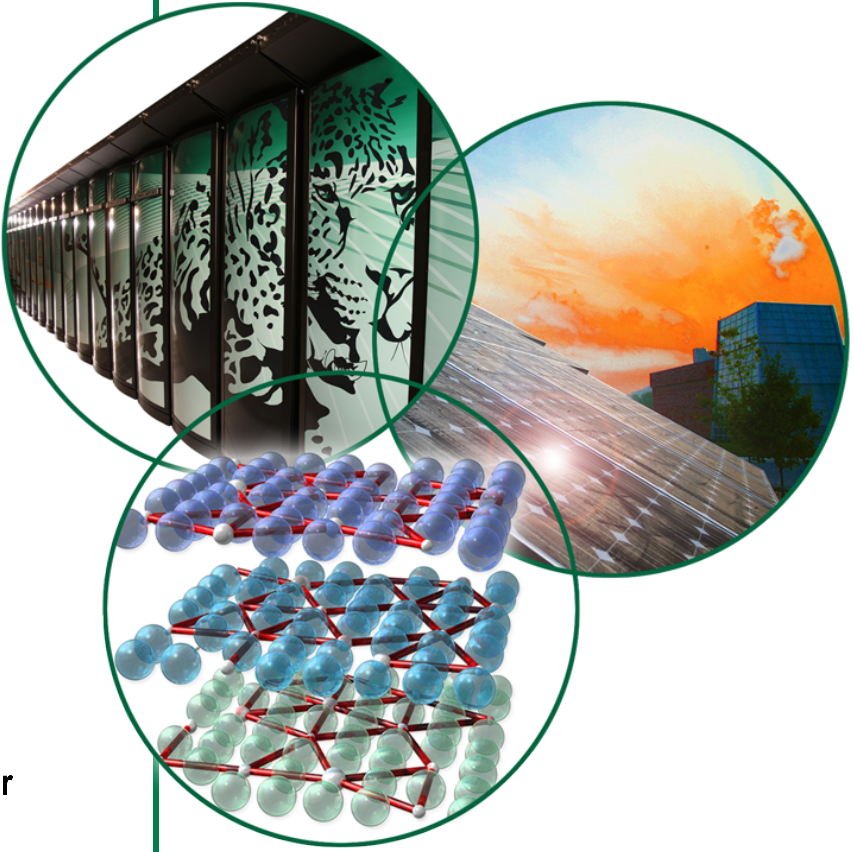
Using Semantic Web Technologies to Develop Intrinsically Resilient Energy Control Systems

Frederick Sheldon, Jingshan Huang, Jiangbo Dang, Daniel Fetzer, David Manz, Thomas Morris, Dong Wei, Jonathan Kirsch, and Stuart Goose



Presented by: Daniel Fetzer, Ph. D.

The 7th International Conference on Semantic Technologies for Intelligence, Defense, and Security (STIDS 2012)

George Mason University, Fairfax, Virginia Campus ♦ Oct. 24, 2012

U.S. DEPARTMENT OF ENERGY

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

Introduction

- To preserve critical energy control functions while under attack, it is necessary to perform comprehensive analysis on root causes and impacts of cyber intrusions, *without sacrificing the availability of energy delivery*
- We propose to design an *intrinsically resilient energy control system* where we extensively utilize Semantic Web technologies, which play critical roles in knowledge representation and acquisition
- Our ultimate goal is to ensure availability/resiliency of energy delivery functions and the capability to assess root causes and impacts of cyber intrusions

Research Motivation

- An Energy Delivery System (EDS) consists of complex and geographically dispersed network architectures with vast numbers of interconnected components
- These systems must maintain high availability and reliability even when under attack
- The incident response team needs the ability to investigate and determine the root cause, attack methods, consequences, affected assets, impacted stakeholders, and other information to inform an effective response
- Such analysis and response must be done without interrupting the availability of the energy delivery systems

Our Solution

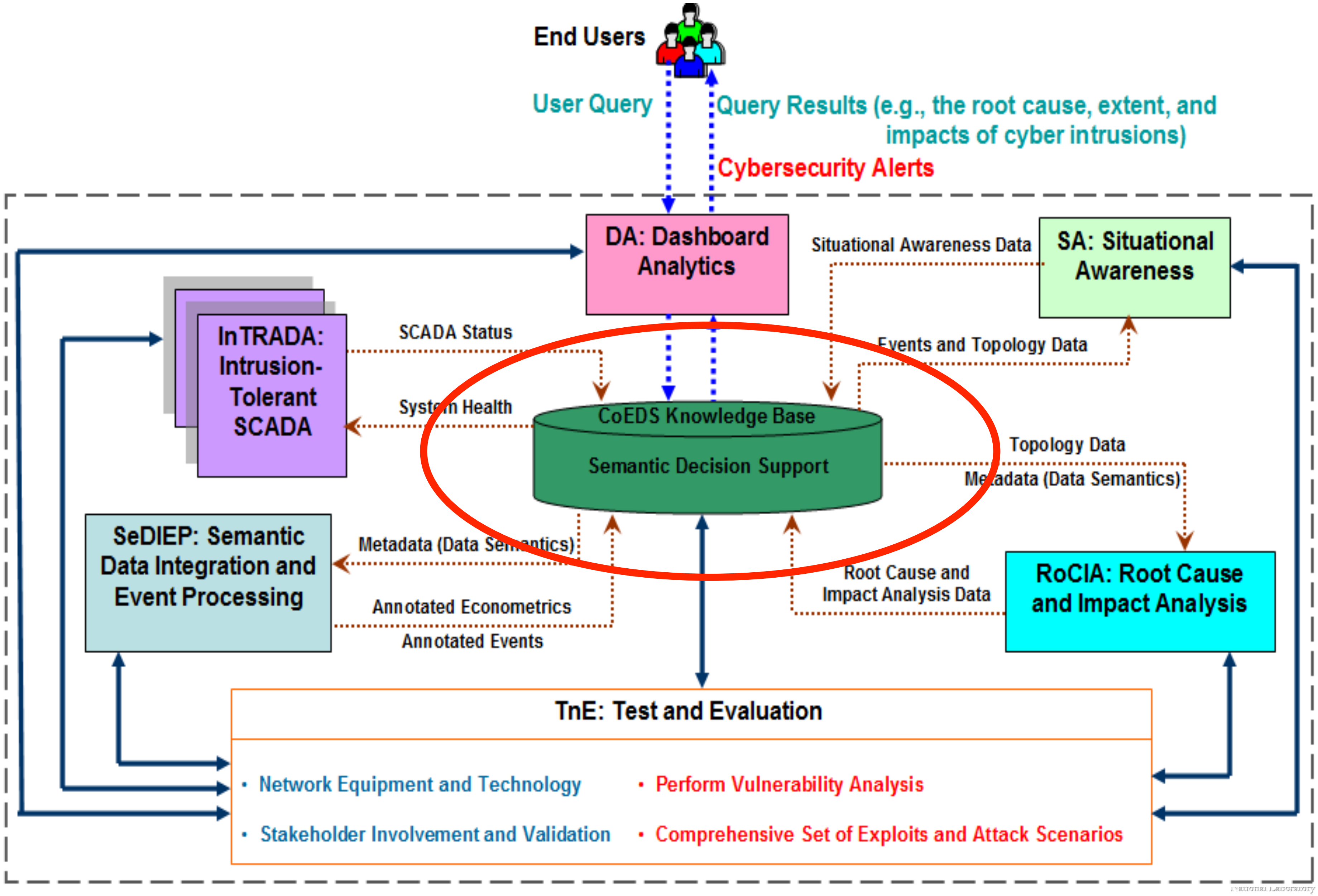
An InTrinsically Resilient Energy Control System

- To provide tools and technologies to ensure the availability/resiliency of energy delivery functions, along with the capability to assess root causes and impacts of cyber intrusions
- Extensively applies Semantic Web technologies (SWT), including cybersecurity domain ontologies, a comprehensive knowledge base, and semantic data annotation & integration techniques

SWT vs. Relational Databases

- While relational databases focus on *syntactic* representation of data and lack the ability to explicitly encode semantics, Semantic Web technologies support rich *semantic* encoding, which is critical in *automated knowledge acquisition*
- *Powerful tools* exist for capturing and managing ontological knowledge, including an abundance of reasoning tools readily supplied for ontological models, making it much more convenient to query, manipulate, and reason over available data sets. As a result, *semantics-based queries, instead of SQL queries, are made possible*
- Advances in an EDS require *changes to be made regularly* regarding underlying data models; it is preferable to represent data at different levels and/or with different abstractions. There are *no straightforward methods for performing such updates if relational models are adopted*
- Semantic Web technologies better enable EDS researchers to append additional data into repositories *in a more flexible and efficient manner*. The formal semantics encoded in ontologies makes it possible *to reuse data in unplanned and unforeseen ways*, especially when data users are not data producers, which is now very common

InTRECS Overall Structure



End Users

User Query

Query Results (e.g., the root cause, extent, and impacts of cyber intrusions)

Cybersecurity Alerts

DA: Dashboard Analytics

SA: Situational Awareness

Situational Awareness Data

InTRADA: Intrusion-Tolerant SCADA

SCADA Status

System Health

Events and Topology Data

CoEDS Knowledge Base

Semantic Decision Support

Topology Data

Metadata (Data Semantics)

SeDIEP: Semantic Data Integration and Event Processing

Metadata (Data Semantics)

Annotated Econometrics

Annotated Events

Root Cause and Impact Analysis Data

RoCIA: Root Cause and Impact Analysis

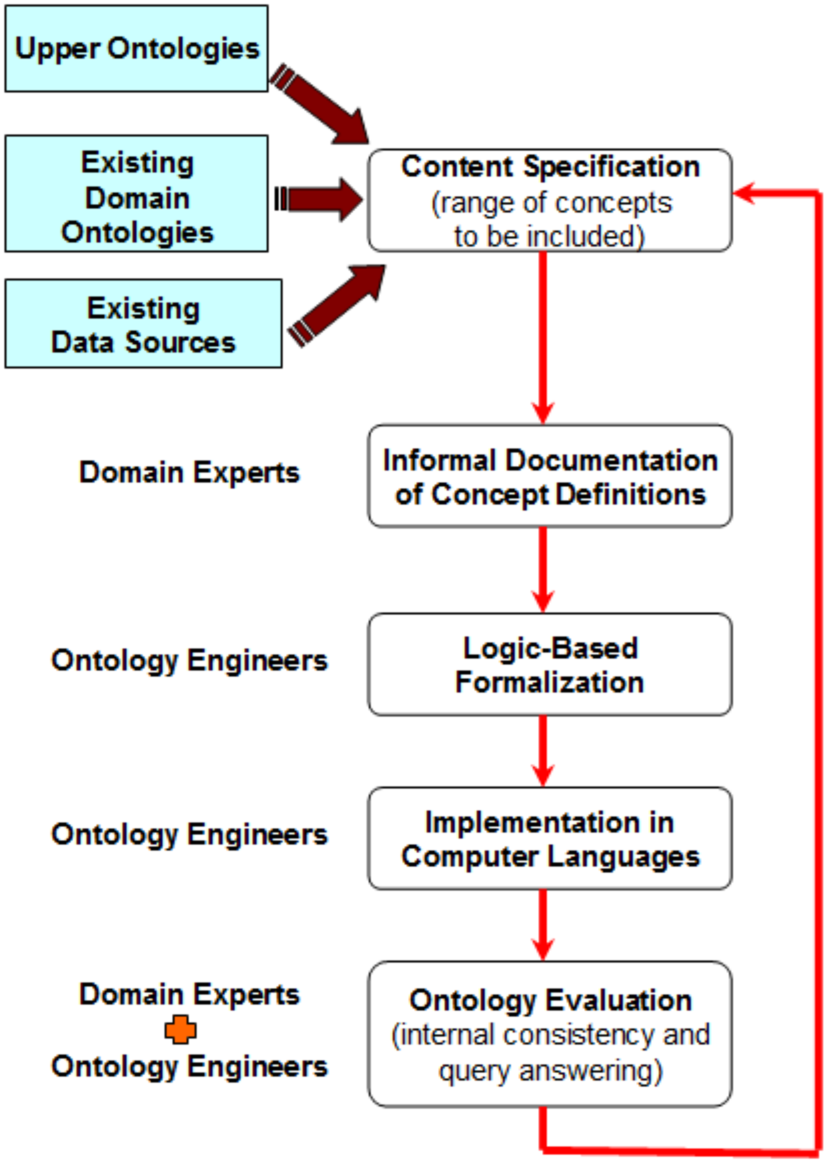
TnE: Test and Evaluation

- Network Equipment and Technology
- Stakeholder Involvement and Validation
- Perform Vulnerability Analysis
- Comprehensive Set of Exploits and Attack Scenarios

CoEDS Domain Ontologies and Knowledge Base

- **C**ybersecurity **O**ntologies for **E**nergy **D**elivery **S**ystems
- Four components: (i) domain ontologies, (ii) an RDF repository, (iii) a SPARQL RDF query engine, and (iv) an inference engine
- Through automatic data integration and logic reasoning, CoEDS will:
  - Provide **a unified and consistent data layer** for analyzing data at the semantic level
  - Assist end users to **effectively obtain real-time decision support**, so that they can (i) obtain health status updates of SCADA replicas, (ii) analyze and better understand the root cause, extent, and impacts of an attack, (iii) acquire situational awareness, and (iv) recommend courses of action

CoEDS Ontology Development



Upper Ontologies

Existing Domain Ontologies

Existing Data Sources

Content Specification (range of concepts to be included)

Domain Experts — Informal Documentation of Concept Definitions

Ontology Engineers — Logic-Based Formalization

Ontology Engineers — Implementation in Computer Languages

Domain Experts + Ontology Engineers — Ontology Evaluation (internal consistency and query answering)

- Iterative, knowledge-driven approach
- Both ontology engineers and domain experts need to be involved
- The language to describe ontologies is OWL (Web Ontology Language), a format recommended by World Wide Web Consortium (W3C)
- The ontology editor is Protégé

RDF and SPARQL

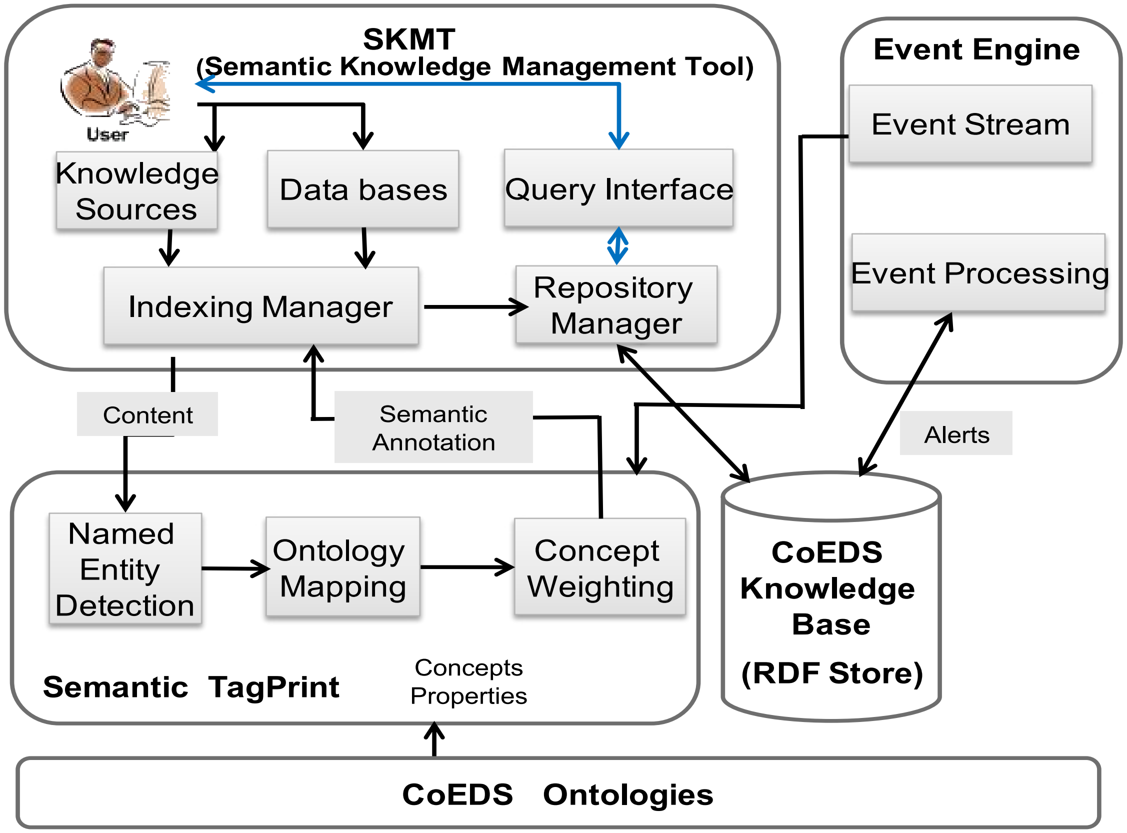
- The interoperability becomes an obstacle during knowledge discovery because heterogeneous data sources to be integrated include structured, semi-structured, or unstructured data
- RDF is a model recommended by W3C for data interchange, supporting the evolution of schemas over time
- The generic structure of RDF allows structured, semi-structured, and unstructured data to be mixed, exposed, and shared across different applications, thus helping to handle the data interoperability challenge
- Annotated RDF triples will be accumulated into a central repository
- SPARQL Protocol and RDF Query Language (SPARQL) is a query language recommended by W3C to retrieve and manipulate RDF data
- RDF queries across semantically integrated sources can be executed by a SPARQL-based query engine

Inference Engine

- We will incorporate an inference engine (a.k.a. logic reasoner), which is a more expressive method for querying and reasoning over available data sets
- Ontology-based (a.k.a. semantics-based) queries, instead of traditional SQL queries, are thus made possible
- Consequently, we will be able to acquire hidden knowledge and information that was originally implicit and unclear, yet critical, to end users
- With a logic reasoner, CoEDS repository will work as a comprehensive knowledge base

SeDIEP Semantic Data Integration and Event Processing

SeDIEP obtains ontological metadata from CoEDS and utilizes such metadata to automatically annotate data with semantics. Annotated data, including cybersecurity econometrics, dynamic events, are stored into CoEDS to construct and continuously update the central data repository.



SeDIEP Core Components

SeDIEP has three major components in the subsystem: (i) Semantic TagPrint, (ii) Semantic Knowledge Management Tool (SKMT), and (iii) Event Engine.

- **Semantic TagPrint** is an automatic semantic tagging engine that annotates structured data and free text using ontological entities from CoEDS ontologies.
- **SKMT** manages heterogeneous data sources for semantic annotation and integration.
- **Event engine** feeds the semantic tagging engine with dynamic events. It also perform event filtering, correlation, and aggregation or abstraction with the semantics defined in CoEDS ontologies.

SeDIEP Benefits

Meaningful data – SeDIEP will annotate terms in text with their corresponding concepts in CoEDS ontologies by finding their meanings and analyzing their context.

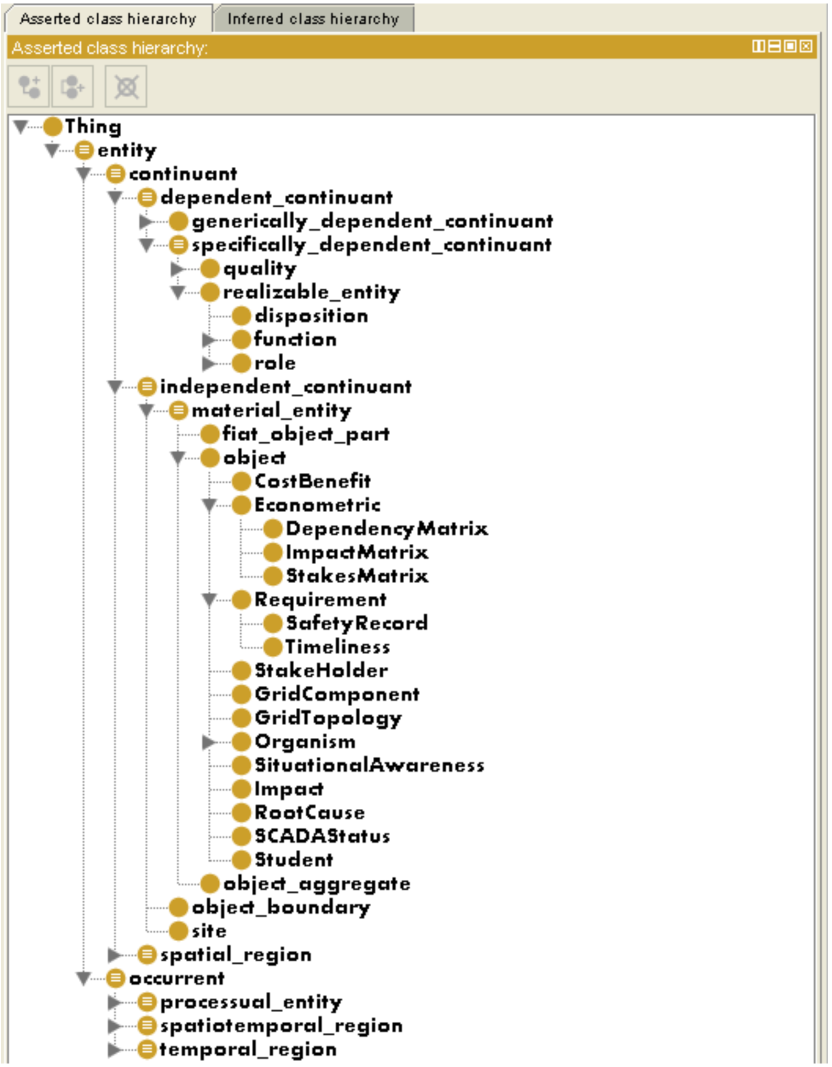
Scalability - Indexed data are stored and managed in a repository. Collected and initially processed data can be incrementally analyzed and indexed.

Easy integration - Various data sources can be seamlessly integrated along with their semantic indexes.

Preliminary Experimental Results

- In this ongoing research, we have developed a preliminary version of CoEDS domain ontologies and knowledge base to demonstrate a proof of concept of how Semantic Web technologies can significantly contribute to resilient energy control systems
- We also exported instances into an RDF data repository within the Sesame framework

CoEDS Ontologies



```
Thing
  entity
    continuant
      dependent_continuant
        generically_dependent_continuant
        specifically_dependent_continuant
          quality
          realizable_entity
            disposition
            function
            role
      independent_continuant
        material_entity
          fiat_object_part
          object
            CostBenefit
            Econometric
              DependencyMatrix
              ImpactMatrix
              StakesMatrix
            Requirement
              SafetyRecord
              Timeliness
            StakeHolder
            GridComponent
            GridTopology
            Organism
            SituationalAwareness
            Impact
            RootCause
            SCADAStatus
            Student
          object_aggregate
        object_boundary
        site
      spatial_region
    occurrent
      processual_entity
      spatiotemporal_region
      temporal_region
```

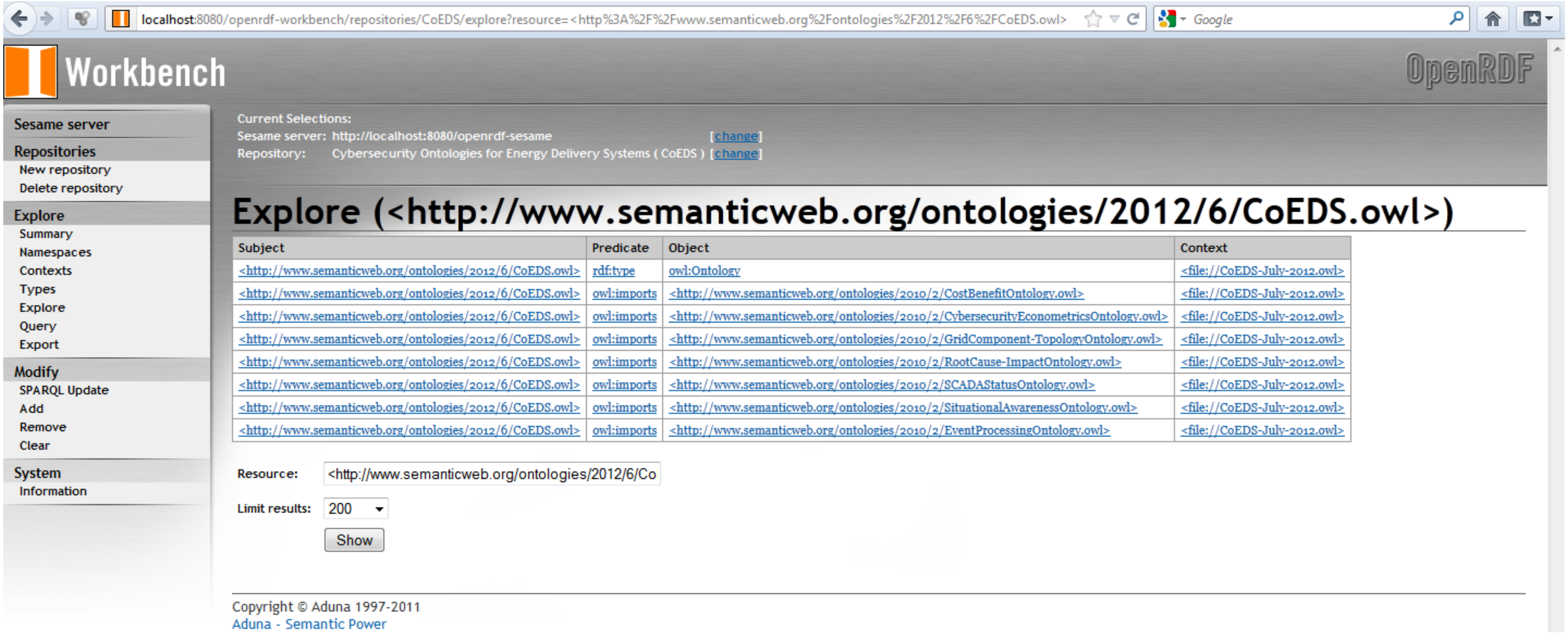
- A screen shot from Protégé GUI, which exhibits a portion of CoEDS concepts
- The well-defined, general-purpose structure from the Basic Formal Ontology (BFO), a popular upper ontology across different disciplines and research areas, was preserved in the ontology schema
- CoEDS ontologies contain 269 concepts, 232 object properties, and 110 data properties

CoEDS Knowledge Base

The current CoEDS KB contains a total of 1,223 facts (a.k.a. axioms in Protégé)

Axiom Category	Statistic Information
Class Axioms	460
Subclass Axioms	268
Equivalent Class Axioms	57
Disjoint Class Axioms	135
Object Property Axioms	217
Data Property Axioms	108
Individual Axioms	236
Annotation Axioms	202

Sesame Framework to Manage Data Repository



Current Selections:
Sesame server: http://localhost:8080/openrdf-sesame [change]
Repository: Cybersecurity Ontologies for Energy Delivery Systems (CoEDS) [change]

Explore (<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>)

Subject	Predicate	Object	Context
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	rdf:type	owl:Ontology	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/CostBenefitOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/CybersecurityEconometricsOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/GridComponent-TopologyOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/RootCause-ImpactOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/SCADAStatusOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/SituationalAwarenessOntology.owl>	<file://CoEDS-July-2012.owl>
<http://www.semanticweb.org/ontologies/2012/6/CoEDS.owl>	owl:imports	<http://www.semanticweb.org/ontologies/2010/2/EventProcessingOntology.owl>	<file://CoEDS-July-2012.owl>

Resource: <http://www.semanticweb.org/ontologies/2012/6/Co

Limit results: 200

- A screen shot from Sesame GUI, where the seven sub-ontologies and the overall CoEDS ontologies were clearly demonstrated
- Within the Sesame framework we exported all ontological instances into an RDF data repository for future storage and management

Conclusions

- We proposed to develop InTRECS, an intrinsically resilient energy control system, to address the challenges in preserving critical energy control functions while under attack
- Semantic Web technologies, which play critical roles in knowledge representation and acquisition, have been extensively adopted in our system
- We justified the research motivation, described our methodology in detail, and exhibited preliminary experimental results
- Future research directions include, but are not limited to, (i) continue CoEDS ontology development towards delivering a highly stable and more usable version; (ii) incorporate query and inference engines into the knowledge base for end users to better analyze root causes and impacts of cyber intrusions; and (iii) implement SeDIEP subsystem

Contact Information

Frederick Sheldon and Daniel Fetzer
Oak Ridge National Laboratory
Oak Ridge, TN
{sheldonft, fetzerdt}@ornl.gov

Jingshan Huang
University of South Alabama
Mobile, AL
huang@southalabama.edu

David Manz
Pacific Northwest National Laboratory
Richland, WA
david.manz@pnnl.gov

Jiangbo Dang and Dong Wei
Siemens Corporation
Princeton, NJ
{jiangbo.dang, dong.w}@siemens.com

Thomas Morris
Mississippi State University
Mississippi State, MS
morris@ece.msstate.edu

Jonathan Kirsch and Stuart Goose
Siemens Corporation
Berkeley, CA
{jonathan.kirsch, stuart.goose}
@siemens.com