SPIIRAS

# The Ontological Approach for SIEM Data Repository

**Igor Kotenko,** Olga Polubelova, and Igor Saenko

Laboratory of Computer Science Problems,
Saint-Petersburg Institute for Information and Automation of Russian Academy of Sciences

Saint-Petersburg, Russia

**3SL 2012 Workshop: 2012:Besançon,France,November 20,2012**

# Content (1)

- **Introduction**
- Related work
- Existing SIEM systems and standards
- Choice of the ontological approach
- Ontological data model
- Ontological repository implementation
- Conclusion

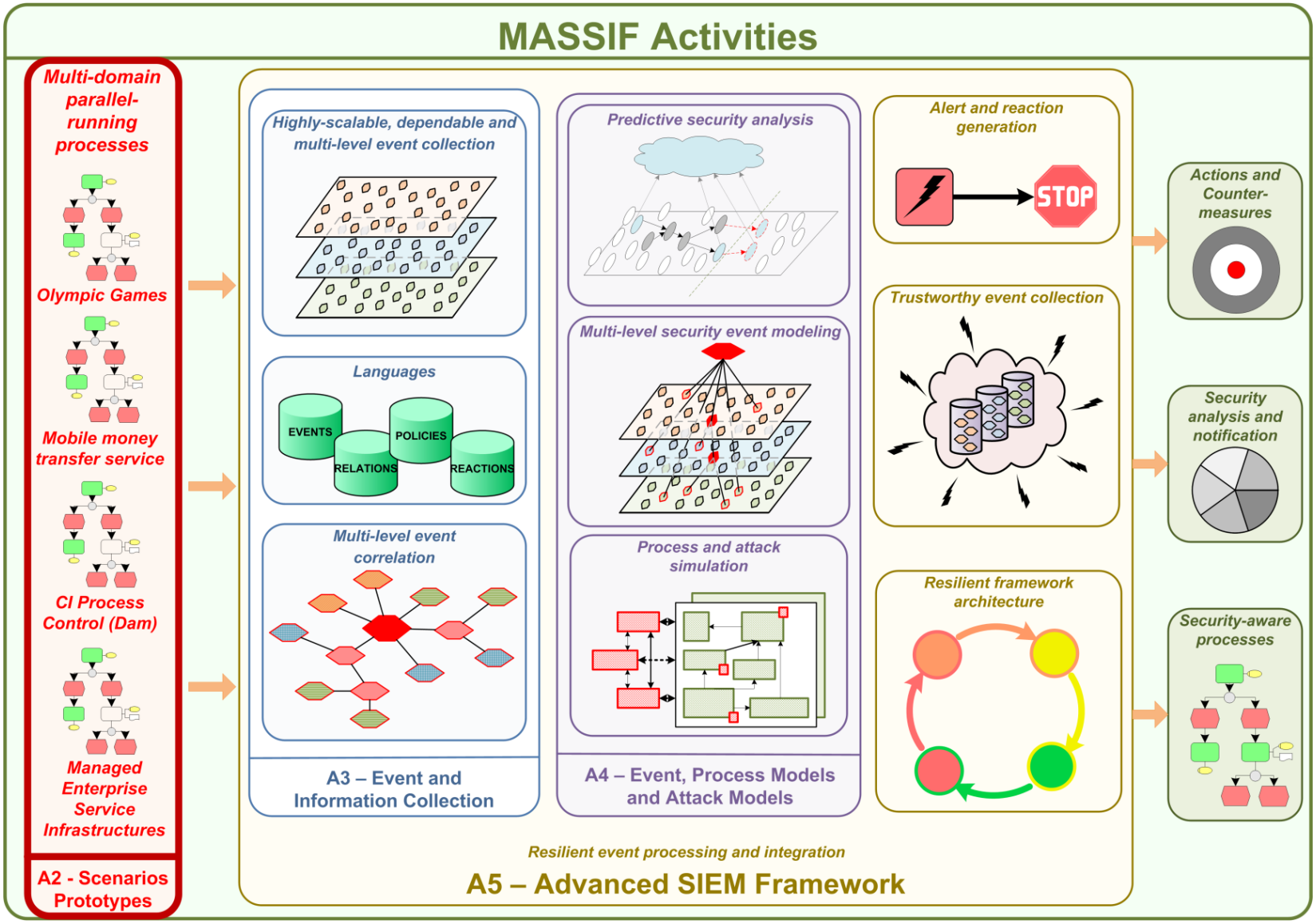
# What is SIEM?

**Security Information and Event Management (SIEM)=**

= technology which is intended

- to provide a coherent collection of security logs from various sources,
- their long- or short-term storage in a system repository,
- to use them for
  - security analysis,
  - attack and countermeasure modeling and
  - forecasting
- and to generate efficient security solutions based on event correlation, data mining, logical reasoning and data visualization.

# MASSIF Project Activities



# MASSIF Scenarios

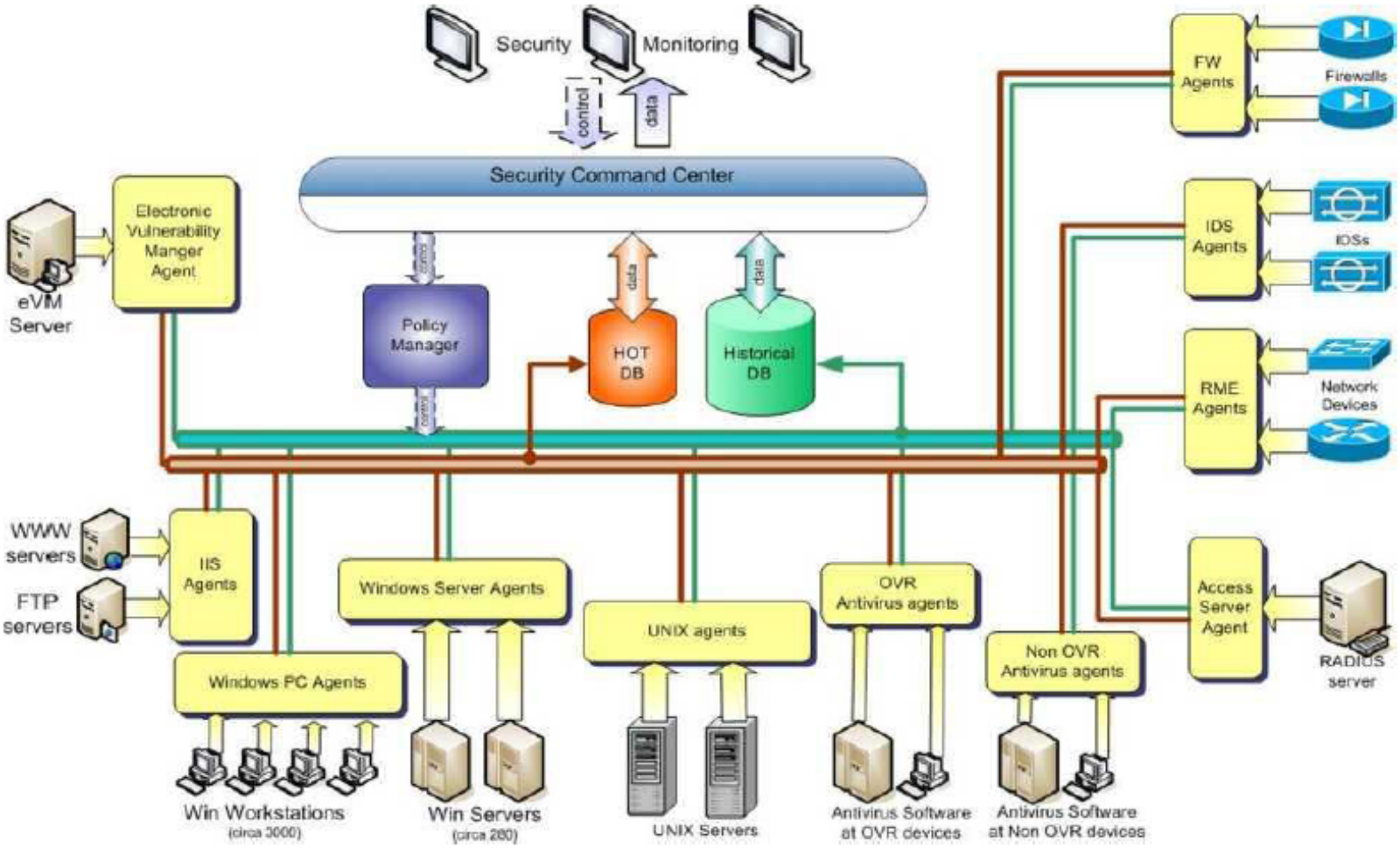
- MASSIF results will be demonstrated in:
  - Four field scenarios:

| Olympic Games IT infrastructure | Mobile phone based money transfer service | Managed Enterprise Service Infrastructures | Critical infrastructure process control |
|---|---|---|---|

  - Existing SIEM solutions:


# Service Infrastructures for SIEM Systems (1)

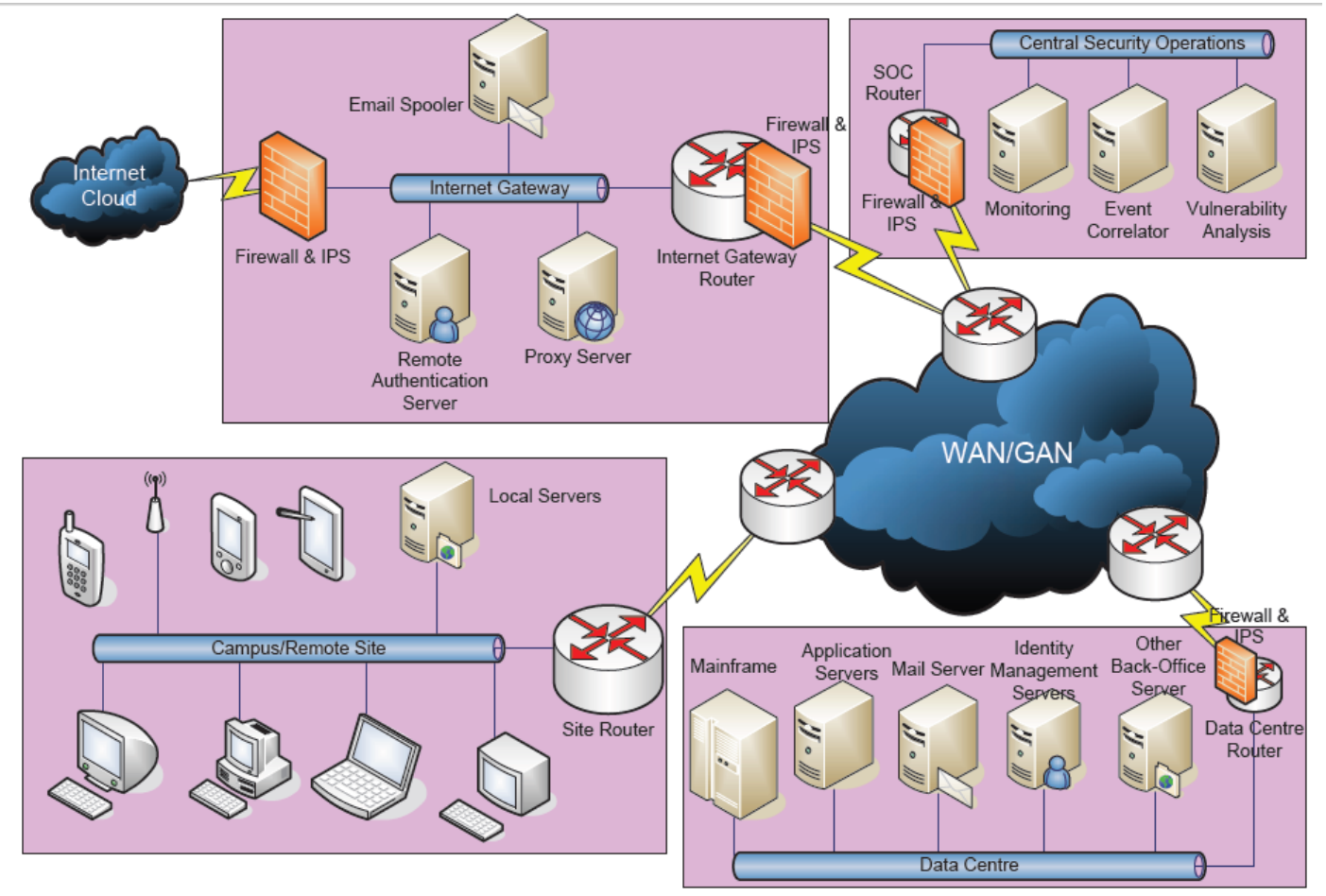
Case Study 1: **Computer network for Olympic Games**




20,000 types of security events; 10,000,000 alarms / day; 40,000 elements; 35,000 users.

# Service Infrastructures for SIEM Systems (2)

## Case Study 2: **Distributed Computer network for Transnational Company**




**Challenges:**

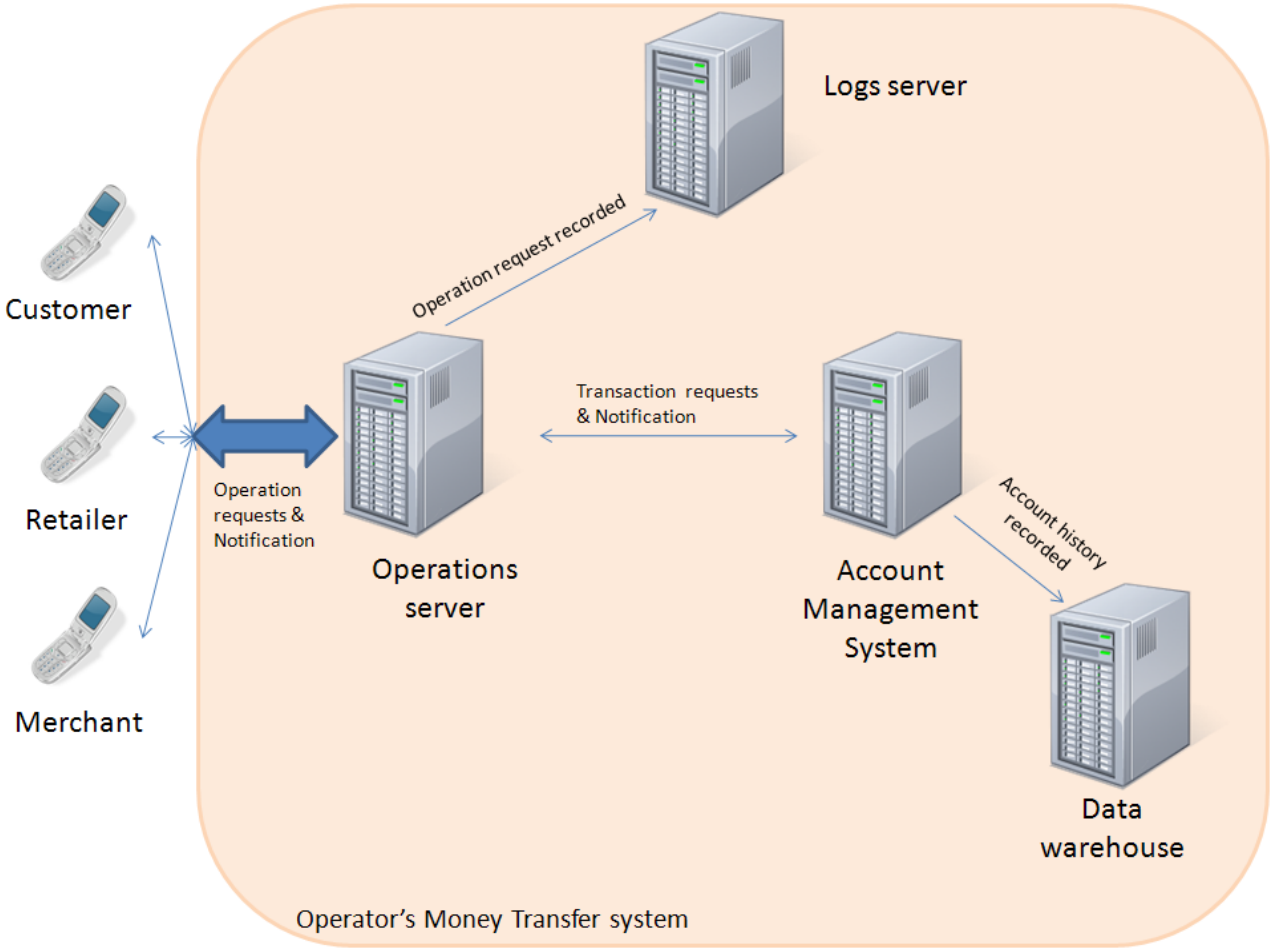
1) Massive attacks
2) Low data reliability
3) Inadequate capability of disaster recovery

**Contribution**:

1) Advanced event correlation
2) Management of complex security events

# Service Infrastructures for SIEM Systems (3)

## Case Study 3: **Mobile Money Transfer**




**Challenges:**

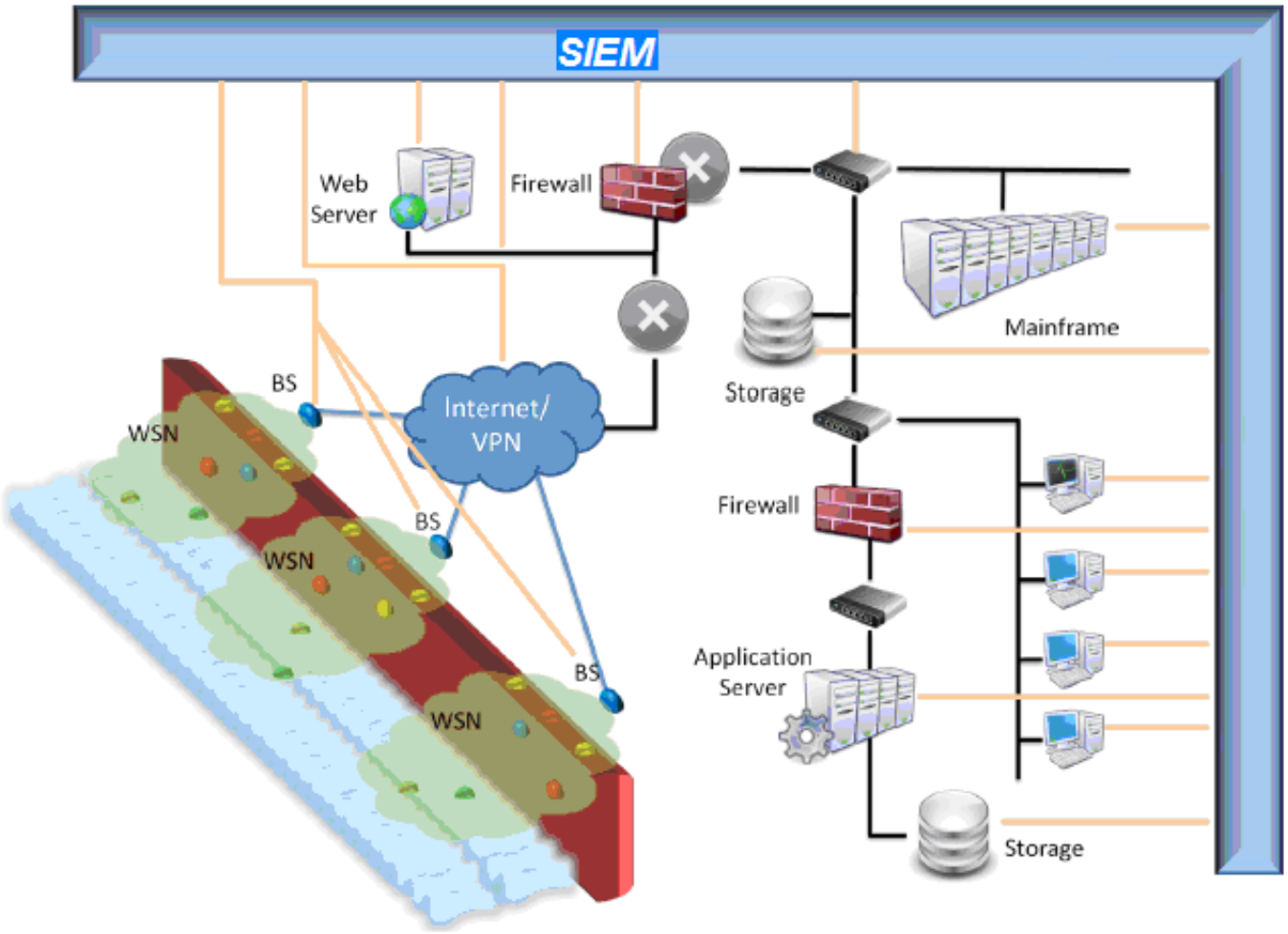
To counter money-laundering and other types of fraud

**Contribution**:

Event correlation at various levels and between levels

# Service Infrastructures for SIEM Systems (4)

## Case Study 4: **Critical Infrastructure (Dam)**




**Challenges:**

1) distinguish real threats on multiple warnings
2) Law reliability of data sources

**Contribution**:

based on the correlation of events, to have a full view of all the potentially critical security events

# What is *Repository Role* in SIEM systems?

1. ***Repository*** is one of the important components of SIEM systems.

2. ***Repository*** is a **data warehouse** that enables to store security information and event data in an internal format and extracts it at the request of other components for identifying security threats and attacks and generating countermeasures.

3. In ***SIEM systems*** it is possible to use the advanced modeling and simulation modules, which also use the **data stored in the repository** to build the attack and countermeasure graphs.

# Goals for Repository Development

- **Develop an approach and techniques for common repository support**
  - Develop unified repository, languages and tools for effective management of security information, events and security policies, logical inference about security, flexible visualization
  - Implement software prototypes for storing, manipulating, visualization and validation of security information, events and policies based on the unified repository

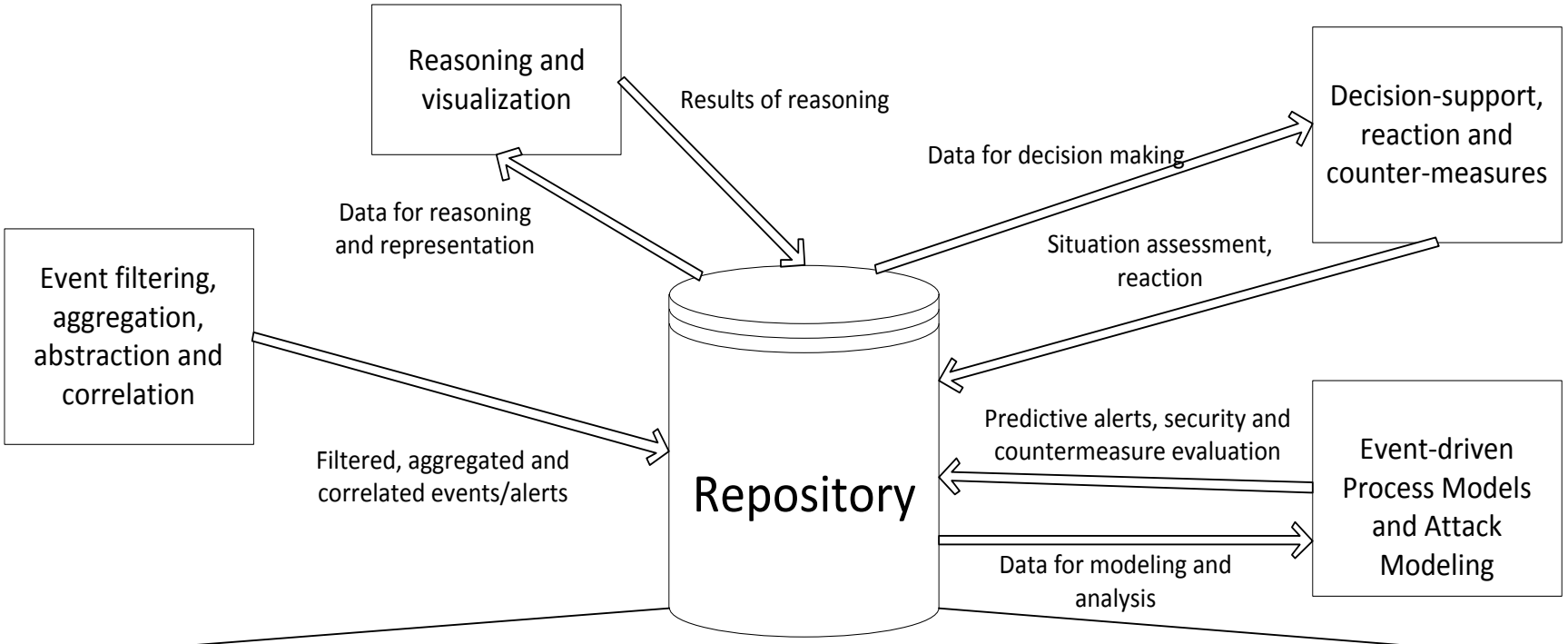
# Is it a critical problem?

We examine the issues of data repository development and implementation for new generation SIEM systems.

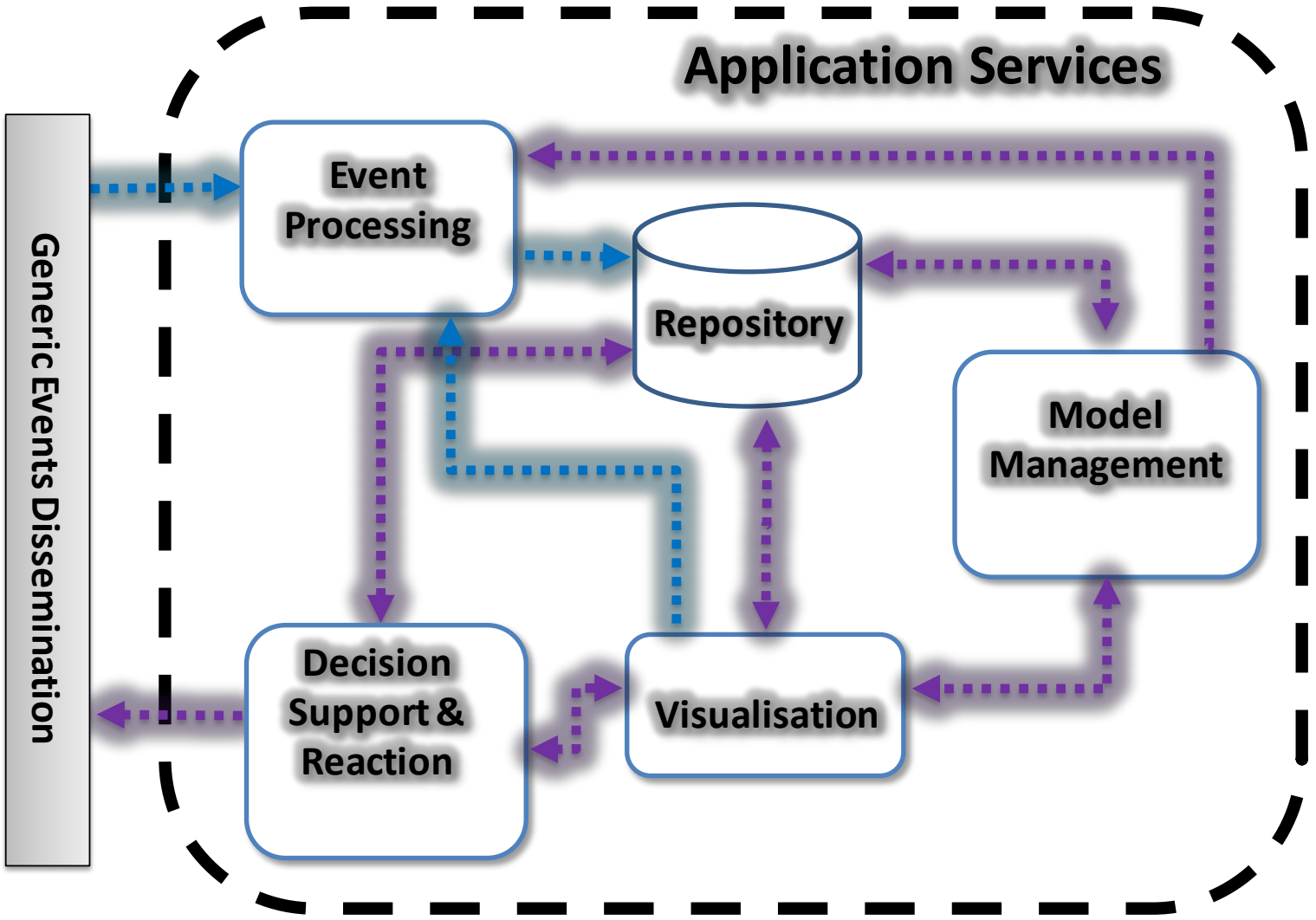
This is a critical problem, because

- logs are saved to the repository from the multitude of security event sources,
- the data from the repository are used in a multitude of data processing, modeling and decision making modules of SIEM systems, and
- it is necessary to provide near real time data processing

# Main Data Flows with Repository in MASSIF



# Repository and Application SIEM Services



# Objectives of the Repository Implementation

1. Design (selection) of a unified repository, languages and tools for effective management of security information, events and policies, and logical inference about security.

2. Implementation of software applications for storing, manipulating, visualizing and validating the security information, events and policies based on the unified repository.

# Research and Development Goals

1. To examine the main issues of data model design and repository development for new generation SIEM systems.

2. To investigate:
   - *Ontological approach* for data model development
   - *Hybrid approach* to implement the repository
   - Advanced *repository architecture*

# Content (2)

- Introduction
- **Related work**
- Existing SIEM systems and standards
- Choice of the ontological approach
- Ontological data model
- Ontological repository implementation
- Conclusion

# Basic Directions in SIEM (1)

## Verification of security policies:

[**Cruz et al., 2008** ]: an ontology forming the formal basis to model the dynamic aspects of role-based access control (for Olympic Games).

[**Da Silva et al., 2007**]: a security ontology that is used to extract knowledge from natural texts.

[**Kolovski et al., 2007**] and [**Rochaeli et al., 2005**]: an ontological approach to engineering the security policies, using the ontological “services-actors-resources” model and the paradigm of ontological templates respectively.

[**Fitzgerald et al., 2007**]: an ontological approach for configuring the firewall management policy for Linux Netfilter.

# Basic Directions in SIEM (2)

## Vulnerability analysis:

[**Rochaeli et al., 2007**]: an ontological approach to construct the knowledge representation system for their vulnerabilities modelling.

## Security monitoring:

[**Kenaza et al., 2006**]: the use of an ontology to provide contextual security event monitoring and intrusion detection (by converting a set of warnings into a set of formatted data).

## Forensics:

[**Schatz et al., 2004**]: an ontological approach for domain-specific event-based knowledge. In the case of unification with the language rules it is sufficient to apply the standard methods of correlation in the automated forensics.

# Content (3)

- Introduction
- Related work
- **Existing SIEM systems and standards**
- Choice of the ontological approach
- Ontological data model
- Ontological repository implementation
- Conclusion

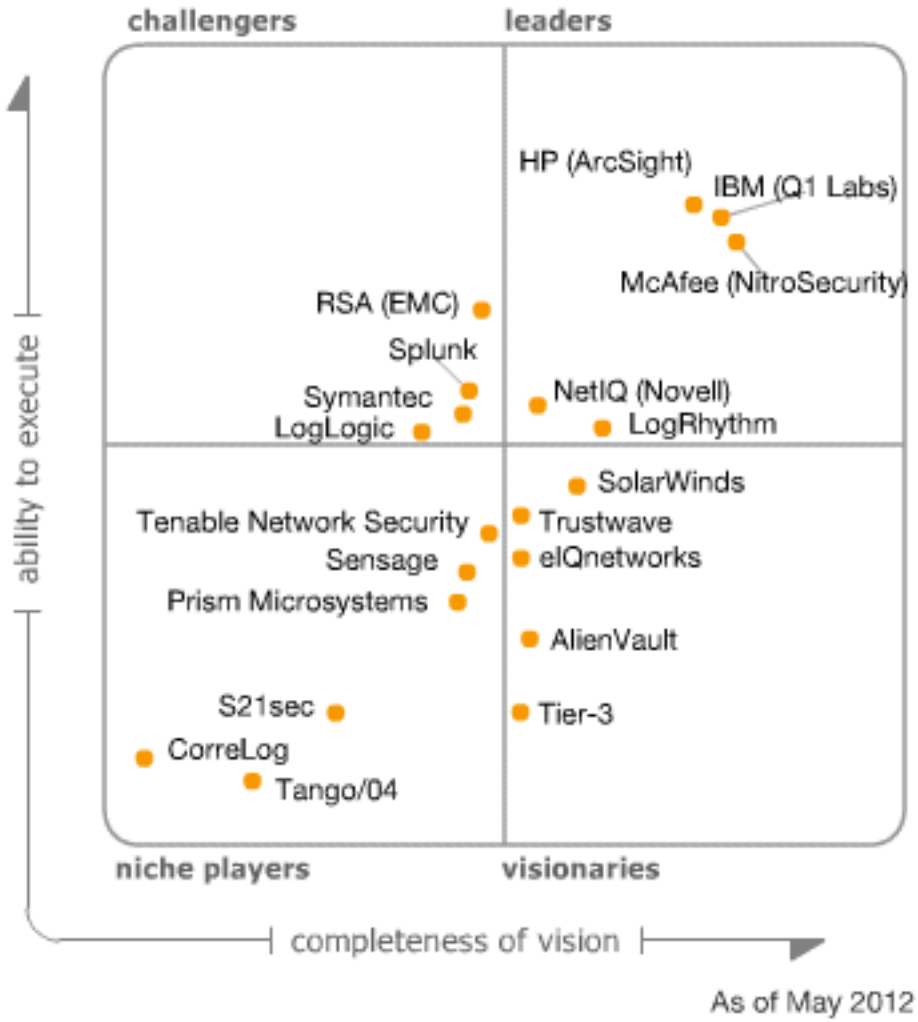
# Information and Event Management Standards

- The **Security Content Automation Protocol (SCAP)** is the synthesis of interoperable specifications derived from community ideas:
  - Common Platform Enumeration – CPE
  - Common Configuration Enumeration – CCE
  - Common Vulnerabilities and Exposures – CVE
  - Common Vulnerabilities Scoring System – CVSS
  - Common Event Expression – CEE

**Other standards**:

- Common Base Event (CBE, IBM)
- Common Event Format (CEF)
- The Common Intrusion Specification Language (CISL)
- The Intrusion Detection Message Exchange Format (IDMEF)
- The Incident Object Description Exchange Format (IODEF)
- Common Information model (CIM)

# Existing SIEM Systems (*Gartner, 2012*)




- ArcSight
- RSA (EMC)
- Symantec Security Information Manager (SIM)
- LogLogic
- IBM Tivoli Security Information and Event Manager (TSIEM)
- CA
- Novell
- LogRhythm
- netForensics Open Security Platform
- Tenable
- LogMatrix

All most popular existing SIEM systems use **SQL databases**!!!

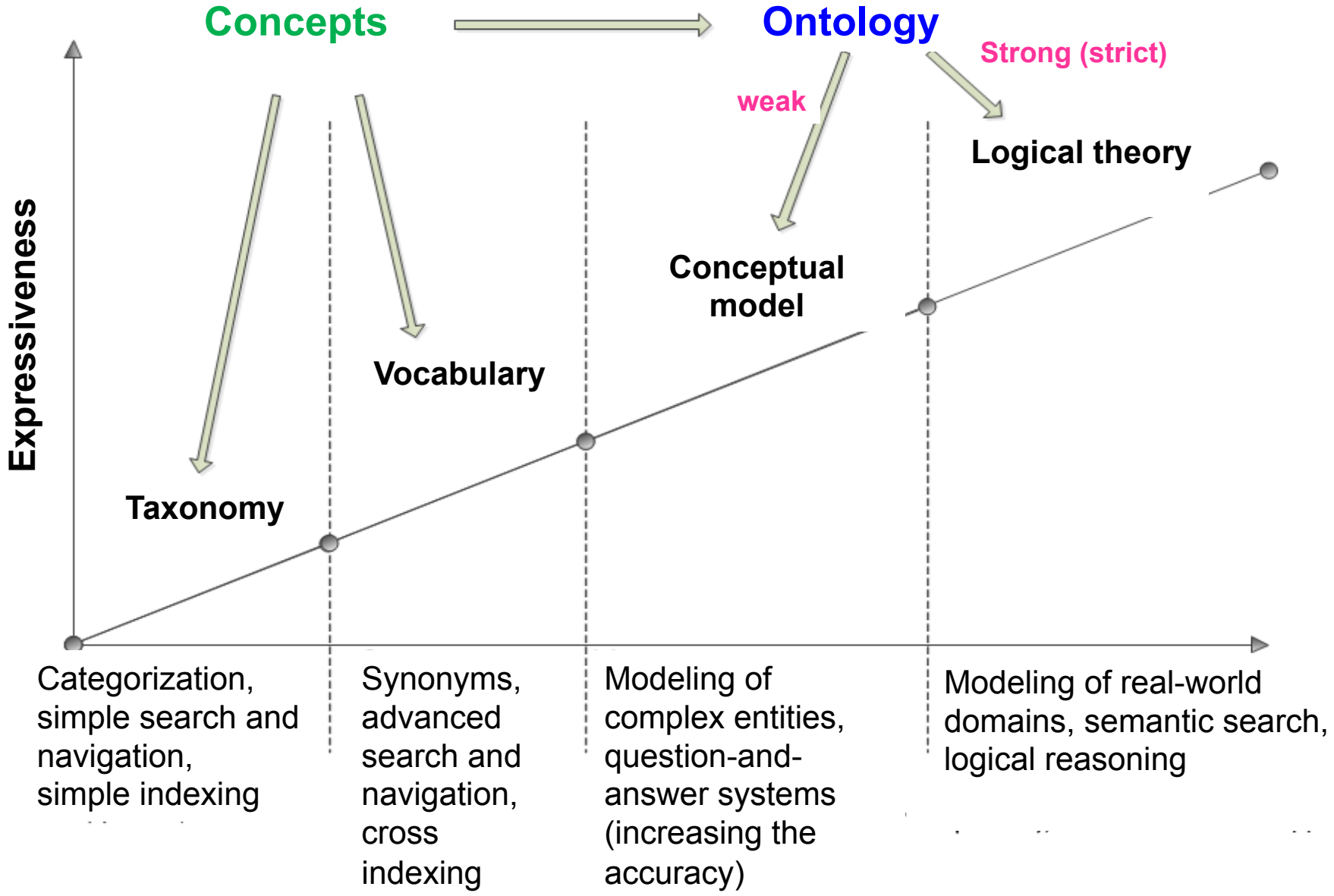
# Languages for Data Representation

- **RDFS (RDF Schema).** RDF data model is a directed graph, which is based on elementary statements (triples).
- **OWL (Web Ontology Language)** is a language of Semantic Web, created to represent ontologies.
- ***SWRL* (Semantic Web Rule Language)** is a proposal for a Semantic Web rule language, based on a combination of OWL sublanguages with RuleML sublanguages.
- **SPARQL Protocol and RDF Query Language (SPARQL)** is a query language to the data presented on the RDF model as well as the protocol for these requests and responses.

# Content (4)

- Introduction
- Related work
- Existing SIEM systems and standards
- **Selection of the ontological approach**
- Ontological data model
- Ontological repository implementation
- Conclusion

# Advantages of ontologies



# Representation of the Entity "Vulnerability" (CVE)

**Example with AND-OR operations**

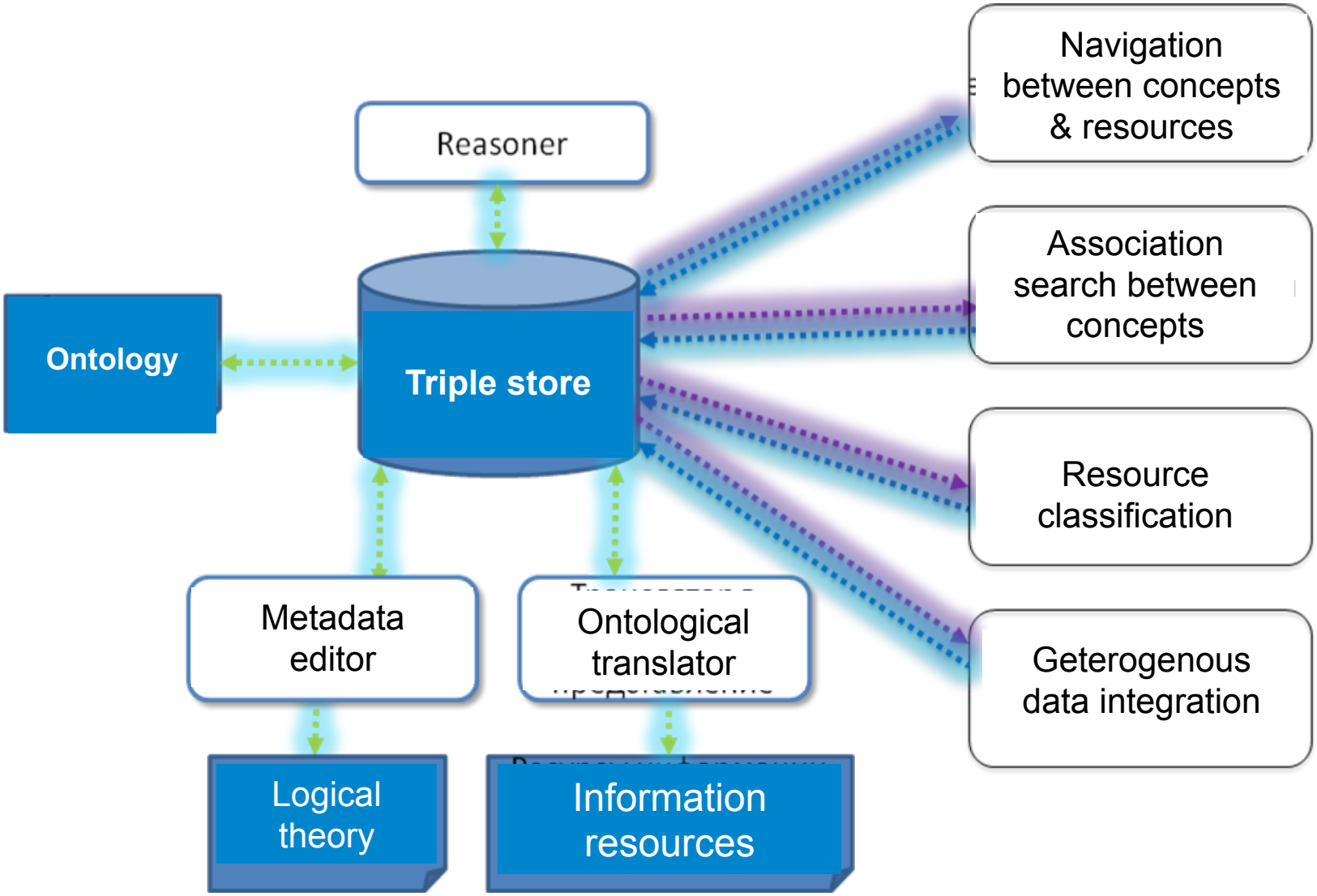
```
<vuln:vulnerable-configuration id="http://nvd.nist.gov/">
  <cpe-lang:logical-test negate="false" operator="AND">
      <cpe-lang:logical-test negate="false" operator="OR">
            <cpe-lang:fact-ref name="cpe:/a:microsoft:ie:9"/>
      </cpe-lang:logical-test>
      <cpe-lang:logical-test negate="false" operator="OR">
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_vista::sp2"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_vista::sp2:x64"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_server_2008::sp2:x86"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_server_2008::sp2:x64"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_7:::x86"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_7::sp1:x86"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_7:::x64"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_7::sp1:x64"/>
            <cpe-lang:fact-ref name="cpe:/o:microsoft:windows_server_2008:r2::x64"/>
            <cpe-lang:fact-ref
      name="cpe:/o:microsoft:windows_server_2008:r2:sp1:x64"/>
      </cpe-lang:logical-test>
  </cpe-lang:logical-test>
</vuln:vulnerable-configuration>
```

# Representation of the entity “Vulnerability” (SQL)

| | |
|---|---|
| 16 | OR(cpe:/o:hp:apollo_domain_os:sr10.2,cpe:/o:hp:apollo_domain_os:sr10.3:beta) |
| 17 | OR(cpe:/o:sun:sunos:4.1,cpe:/o:sun:sunos:4.1.1) |
| 18 | OR(cpe:/o:sun:sunos:4.0.3,cpe:/o:sun:sunos:4.1,cpe:/o:sun:sunos:4.1.1) |
| 19 | OR(cpe:/o:sun:sunos:4.0.3,cpe:/o:sun:sunos:4.0.3c) |
| 20 | OR(cpe:/o:digital:ultrix:4.0,cpe:/o:digital:ultrix:4.1) |
| 21 | OR(cpe:/a:next:next:2.1) |
| 22 | OR(cpe:/o:att:svr4:4.0) |
| 23 | OR(cpe:/o:digital:ultrix:4.2) |
| 24 | OR(cpe:/a:ncsa:telnet) |

Specification of vulnerabilities in the relational data model is **a hard task**!!!

# Architecture of SIEM Inference System

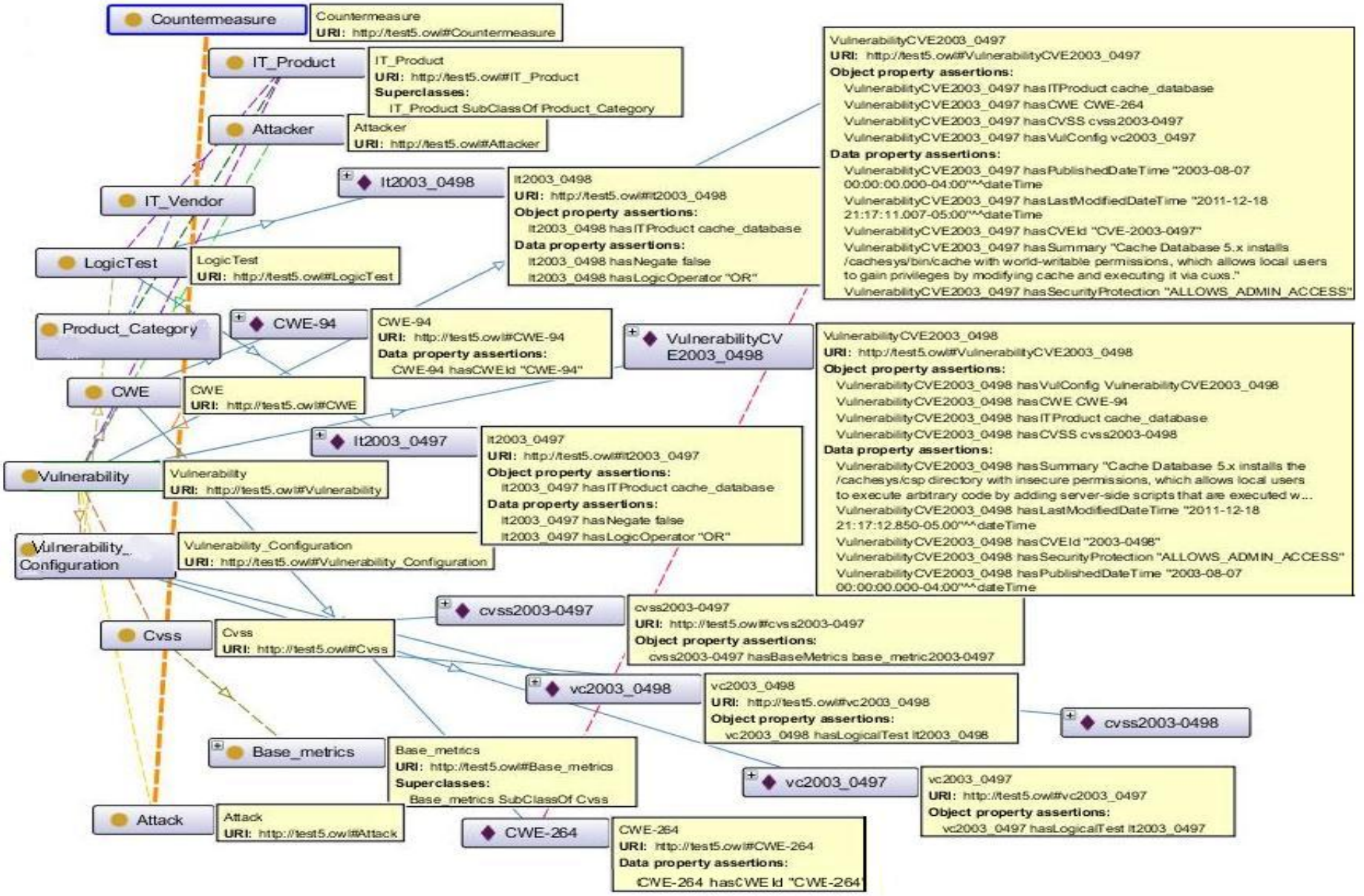


# Content (5)

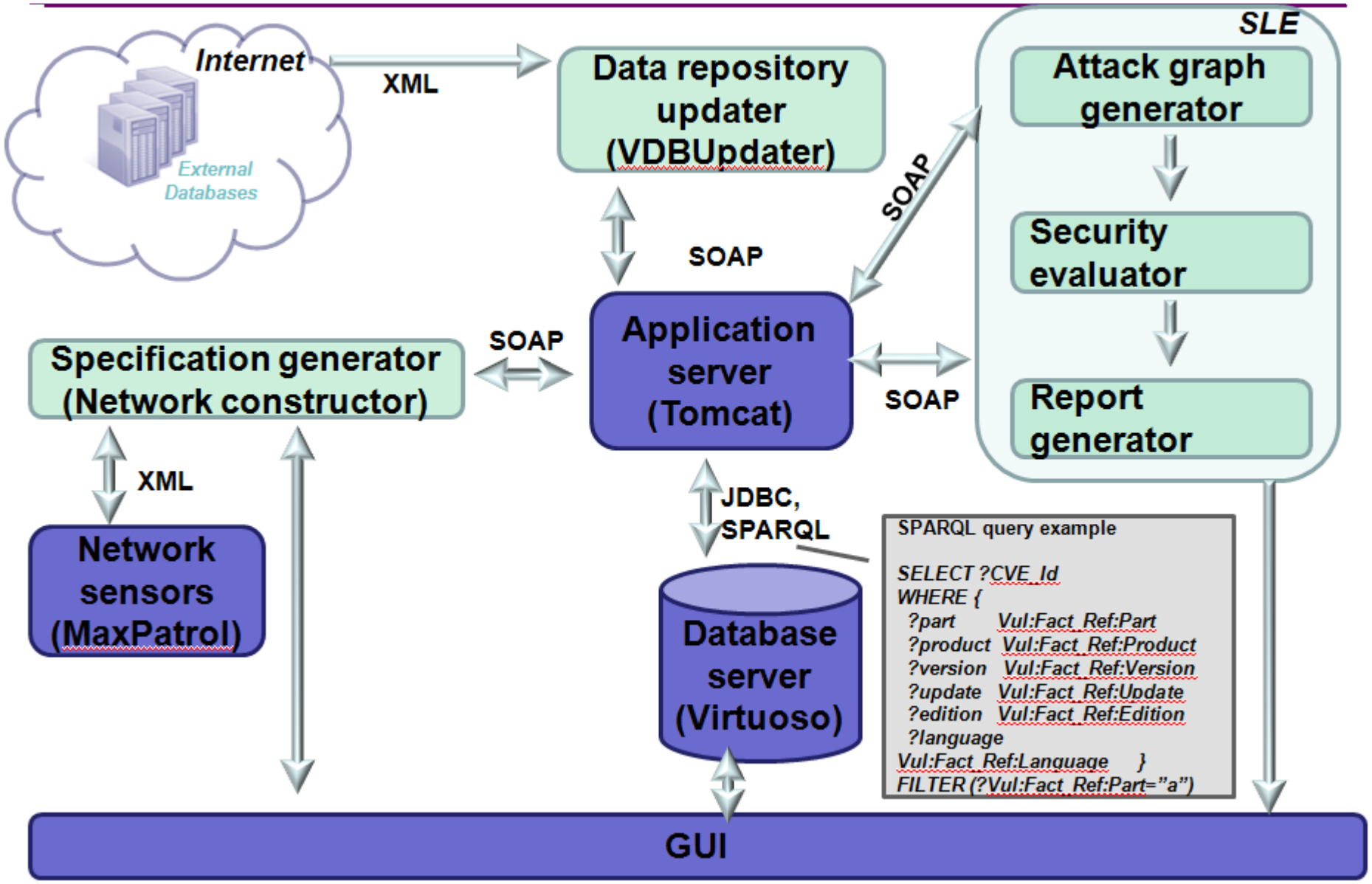
- Introduction
- Related work
- Existing SIEM systems and standards
- Choice of the ontological approach
- **Ontological data model**
- Ontological repository implementation
- Conclusion

# Ontological Data Model for SIEM Repository

(vulnerabilities, software/hardware manufacturers and other concepts)



# Applying in AMSEC



# Ontological Model of “Vulnerability”


**Main features:**

- the ontology describes vulnerabilities, attacks, software/hardware manufacturers and other concepts
- the relationships between software and hardware components are specified using the description logic
- connections are represented, mainly, by the subclasses
- the logical reference is confined to the task of classification, which increases the processing speed

# Example of Vulnerability Representation as Triplets

**CVE standard:**

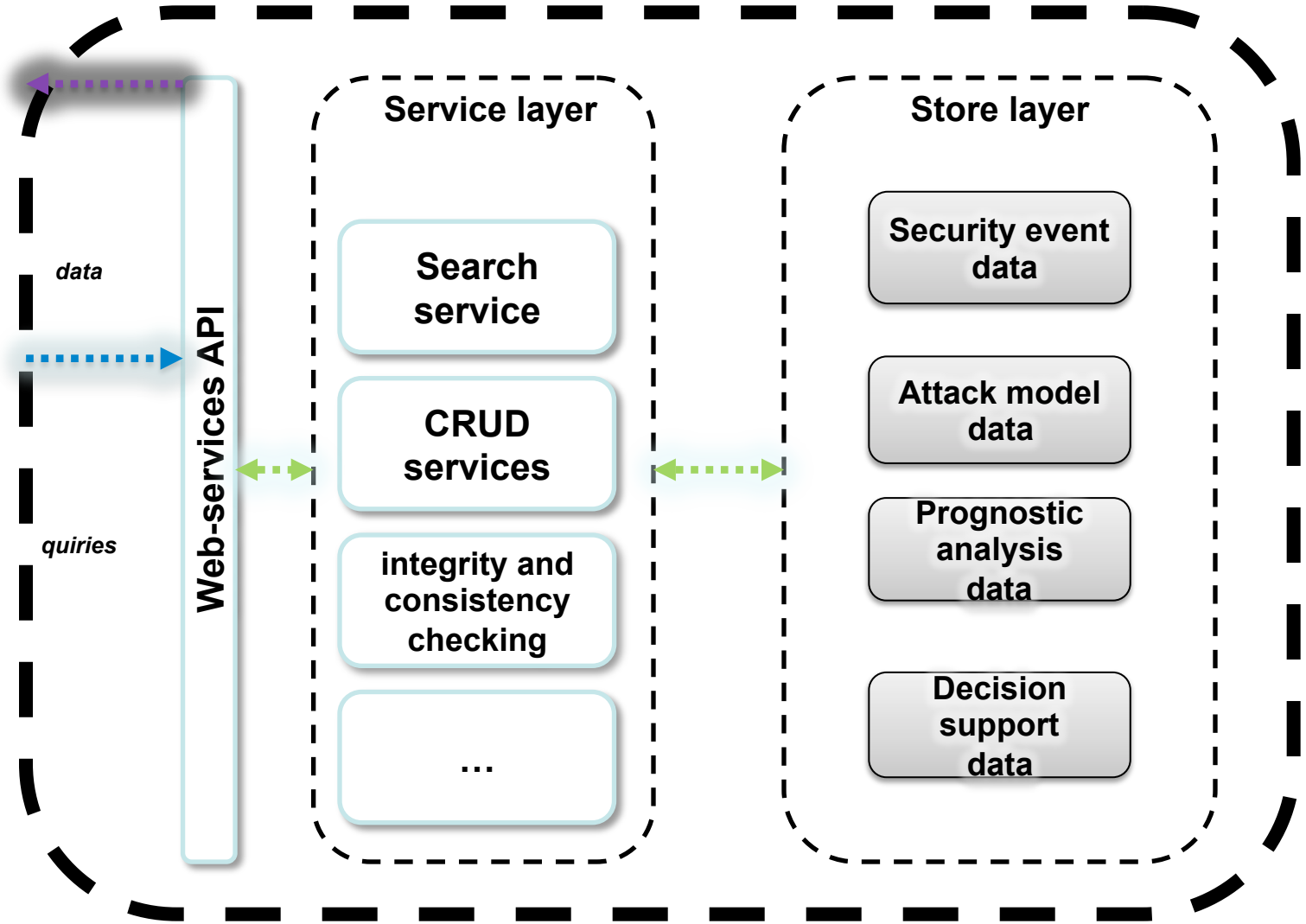
```
<entry id="CVE-2003-0497">
    <vuln:vulnerable-configuration id="http://nvd.nist.gov/">
            <cpe-lang:logical-test negate="false" operator="OR">
                <cpe-lang: fact-ref name="cpe:/a:intersystems:cache_database:5"/>
            </cpe-lang:logical-test>
    </vuln:vulnerable-configuration>
<vuln:vulnerable-software-list>
…
</entry>
```

***Ontological*** **representation in terms of *triplets:***

# Content (6)

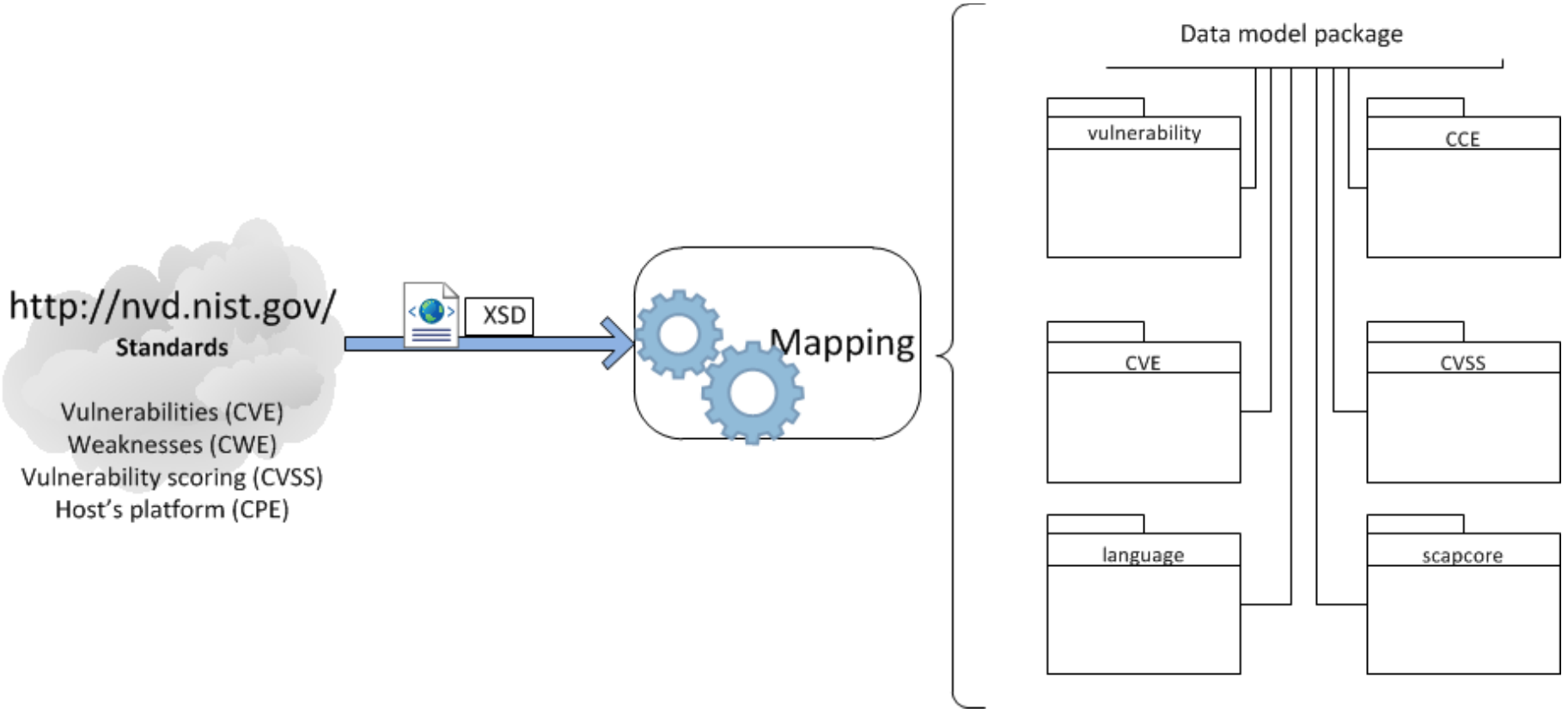
- Introduction
- Related work
- Existing SIEM systems and standards
- Choice of the ontological approach
- Ontological data model
- **Ontological repository implementation**
- Conclusion

# Advanced Repository Architecture




- SOA architecture
- Interconnection with modules by SOAP
- Hybrid storage

# Mapping Data Schemas of NIST Standards to Data Model Layer

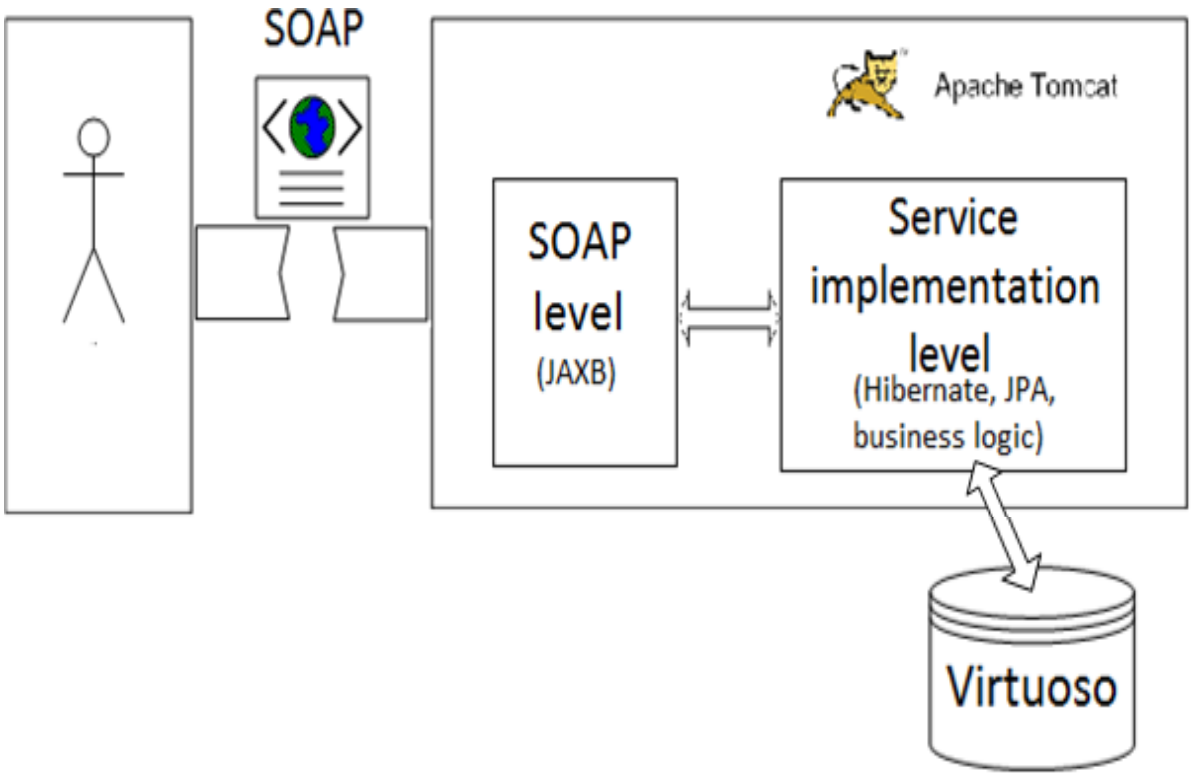


# Implementation Features

- Spring Framework
- Java Persistence API
- SOAP
- Web services
- Spring security 3.0.6

## Data access languages

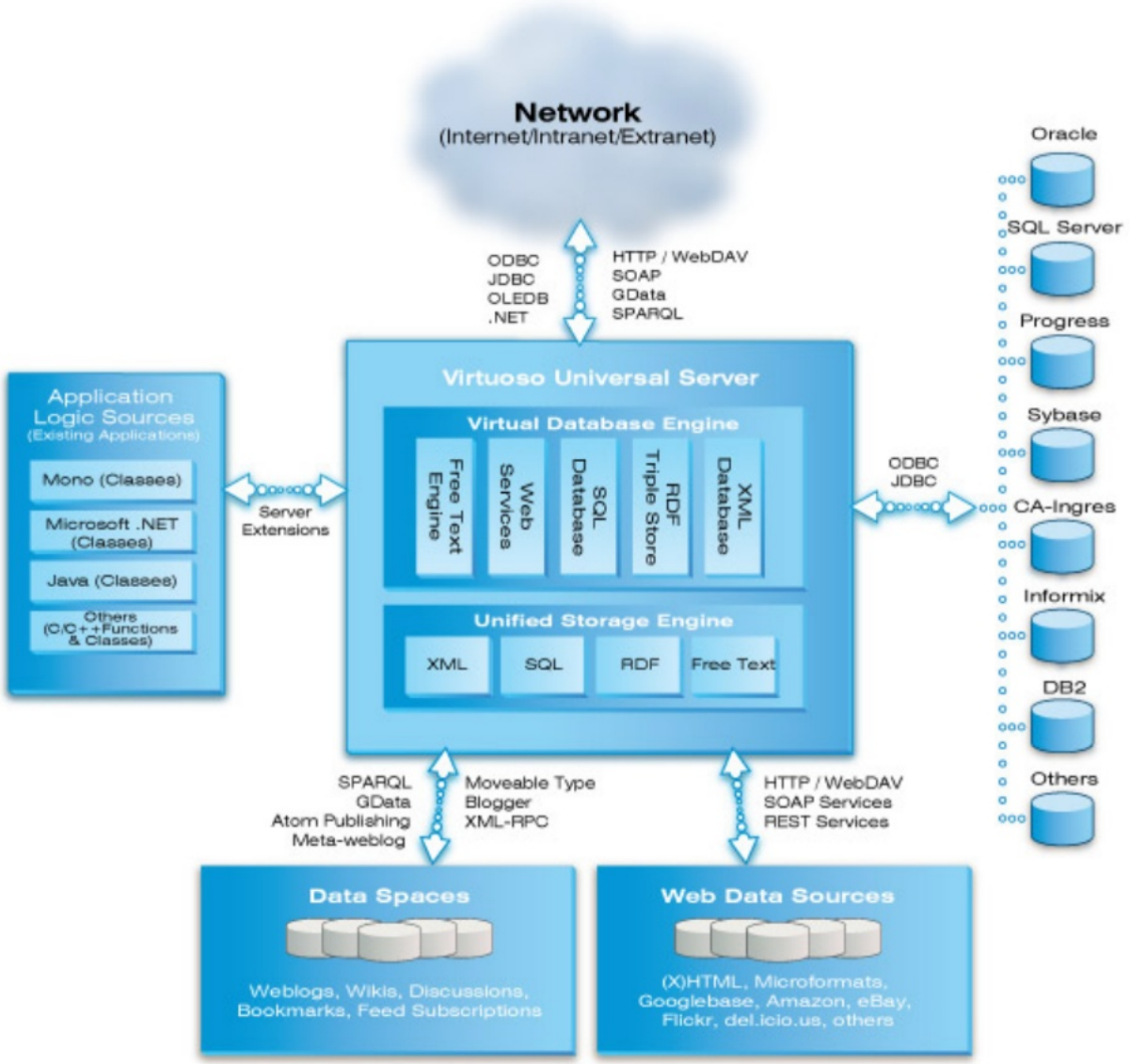
- Relational DB
  - SQL
- Triplet storage
  - SPARQL is a query language and a protocol for accessing RDF
  - SPIN is the RDF syntax for SPARQL
- XML DB
  - Xquery
  - XPath



# Architecture of *Virtuoso*

***Virtuoso*** (*OpenLink* product) - a complex semantic system repositories, which implements the ability to represent the **relational data, XML** and **triplets**.

*Triplets* is a short formal statement in the form of **"subject-predicate-object".**



# Conclusion

**Main results:**

- *Ontological approach* to provide the necessary flexibility of data representation in the repository and the possibility of more accurate and high-quality results of queering (and ontology for AMSEC)
- *Hybrid approach* to implement the repository which allows to integrate relational databases, XML databases and stores of triplets
- Advanced *repository architecture* implemented and tested with the data used for attack modeling in SIEM systems

**Future research:**

- Improving and expanding the proposed ontology
- Adding to the repository different services that provide data security, verification of security properties and policies, etc.
- Exploration of logical reasoning based on ontological repository, development of mechanisms for data visualization

# Questions? Comments?

## For more information please contact


**Igor Kotenko**

Head of Laboratory of Computer Security Problems,
St. Petersburg Institute for Informatics and Automation
of Russian Academy of Sciences (SPIIRAS)

*ivkote@comsec.spb.ru*

*http://www.comsec.spb.ru*



*This research is being supported by grant of the Russian Foundation of Basic Research (project #10-01-00826-a), Program of fundamental research of the Department for Nanotechnologies and Informational Technologies of the Russian Academy of Sciences (contract #2.2), State contract #11.519.11.4008, Science Support Foundation and partly funded by the EU as part of the SecFutur and MASSIF projects.*




# Contacts

Igor Kotenko
ivkote@comsec.spb.ru
http://www.comsec.spb.ru/kotenko/

Olga Polubelova
ovp@comsec.spb.ru
http://www.comsec.spb.ru/polubelova/

Igor Saenko
ibsaen@comsec.spb.ru
http://comsec.spb.ru/saenko/