Federal Big Data and DoD Ontologies 2014

- Goodier
Agenda

Federal big data and ontology in DoD – The current state of innovation

- How does DoD categorize big data vs. ontology research?
- How much research in big data vs. ontology is there in DoD?
- What is the status of DoD research in big data vs. ontology – open or private?
- Who are the major publishers of big data vs. ontology research in DoD?
- Augmented governance of federal big data is a semantic big data use case.
- Demo of automated compliance toolkit
## Recent DoD Research on Ontology versus Big Data

<table>
<thead>
<tr>
<th>Ontology Subject Taxonomy</th>
<th>Big Data Subject Taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>medicine and medical research (25)</td>
<td>medicine and medical research (407)</td>
</tr>
<tr>
<td>ontology (16)</td>
<td>theses (289)</td>
</tr>
<tr>
<td>computer programming and software (25)</td>
<td>computer programming and software (272)</td>
</tr>
<tr>
<td>algorithms (11)</td>
<td>algorithms (270)</td>
</tr>
<tr>
<td>cybernetics (12)</td>
<td>cybernetics (257)</td>
</tr>
<tr>
<td>foreign reports (12)</td>
<td>government and political science (206)</td>
</tr>
<tr>
<td>decision making (12)</td>
<td>military forces and organizations (201)</td>
</tr>
<tr>
<td>psychology (12)</td>
<td>anatomy and physiology (194)</td>
</tr>
<tr>
<td>information science (22)</td>
<td>training (190)</td>
</tr>
<tr>
<td>software engineering (11)</td>
<td>logistics, military facilities and supplies (183)</td>
</tr>
</tbody>
</table>
Recent DoD Research on Ontology versus Big Data

<table>
<thead>
<tr>
<th>Ontology Subject Taxonomy</th>
<th>Big Data Subject Taxonomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>medicine and medical research (61)</td>
<td>medicine and medical research (789)</td>
</tr>
<tr>
<td>ontology (27)</td>
<td>theses (325)</td>
</tr>
<tr>
<td>computer programming and software(34)</td>
<td>computer programming and software(393)</td>
</tr>
<tr>
<td>situational awareness (30)</td>
<td>algorithms (453)</td>
</tr>
<tr>
<td>cybernetics (60)</td>
<td>cybernetics (488)</td>
</tr>
<tr>
<td>sbir(small business innovation research) (62)</td>
<td>sbir(small business innovation research) (777)</td>
</tr>
<tr>
<td>sbir reports (61)</td>
<td>sbir reports (771)</td>
</tr>
<tr>
<td>military intelligence (35)</td>
<td>anatomy and physiology (355)</td>
</tr>
<tr>
<td>information science (73)</td>
<td>electrical and electronic equipment (311)</td>
</tr>
<tr>
<td>computer systems (24)</td>
<td>pe65502f (302)</td>
</tr>
</tbody>
</table>

Ontology research is private - over half -127 of 250 reports. Big data is a bit more public - with over 60% choosing public publication. SBIR is private release.
Publishers of DoD Ontology research include:

- charles river analytics inc cambridge ma (11)
- surviac (10)
- csiac-bco (8)
- aptima inc woburn ma (7)
- intelligent automation inc rockville md (6)
- army center for environmental health research fort detrick md (5)
- intelligent software solutions colorado springs co (4)
- modus operandi inc melbourne fl (4)
- stottler henke associates inc san mateo ca (4)
- army command and general staff college fort leavenworth ks (3)
Publishers of DoD Big Data research include:

- surviac (403)
- naval postgraduate school monterey ca (185)
- library of congress washington dc congressional research service (112)
- riac (77)
- government accountability office washington dc (71)
- army research lab aberdeen proving ground md weapons and materials research directorate (45)
- army command and general staff college fort leavenworth ks (44)
- army war college carlisle barracks pa strategic studies institute (41)
- carnegie-mellon univ pittsburgh pa software engineering inst (41)
- air univ maxwell afb al air force research inst (24)
DoD Big Data requires semantically AUGMENTED GOVERNANCE for dynamic event management.

to retain relevance
- in loosely coupled
- multi-tenant environments
Why? It is needed to enable the BIG DATA scale of IT.

- 25 Point Plan - Nov. 19 2010
  - focused on SPEED OF CHANGE by removing the barriers that get in the way of consistent execution

  – www.cio.gov
Cloud First Challenge – automated audit assertions for FedRAMP clouds using open, structured, industry accepted formats

- The Standards Acceleration to Jumpstart Adoption of Cloud Computing (SAJACC) project at the National Institute of Standards and Technology (NIST) generated concrete data about how different kinds of cloud system interfaces support portability, interoperability, and security.

- The SAJACC project facilitates Standards Development Organizations in their efforts to develop high-quality standards that address these important needs.
NIST–based open, structured, industry accepted formats

- All use cases, test codes, and test results are on the openly-accessible NIST Cloud Portal:
  - for use by any interested parties.
    - CloudAudit/A6 URI Ontology - The Automated Audit, Assertion, Assessment, and Assurance API
      - [http://cloudaudit.org/](http://cloudaudit.org/)
    - SCAP
    - Excluded – private formats
      - CloudTrust
      - ISACA's Cloud Computing Management Audit Assurance Program
Big Data is a semantic problem

- **DoD's big data semantics problem**
  - Sorting through millions of daily logs and records and communication exchanges to pinpoint key individuals or groups that may be crucial to a given investigation is ultimately driven by semantics.
  - By including sophisticated semantic analytics, we vastly reduce the time and budget that might otherwise be needed for a substantive analysis of the regulatory compliance for any set of records.
AUTOMATED COMPLIANCE TOOL (ACT) demo
Version 2
Automated Compliance Tool

ACT is an enhanced parser/data extraction design tool as shown in this demonstration.

It enables rapid decision making that supports legal, regulatory, and policy compliance using cognitive metadata.
ACT represents violations and the configurations that are causing them.

**Policy violations**

**State of the system**

**ACT Module**

**ACT database**

**Cognitive Metadata**

**Dynamic policies related to violations**

**Applicable events**

**Augmented Governance**
Start/Stop the xml database

```
08 Feb 2011 03:12:49.221 [main] INFO  (JettyStart.java [run]:98) - Configuring a
  Context from C:\exist\conf.xml
08 Feb 2011 03:12:49.252 [main] INFO  (JettyStart.java [run]:91) -
08 Feb 2011 03:12:49.252 [main] INFO  (JettyStart.java [run]:92) - Running with
  Java 1.6.0_21 [Sun Microsystems Inc. (Java HotSpot(TM) Client VM) in C:\Program
  Files (x86)\Java\jre1.6.0_21]\"
08 Feb 2011 03:12:49.252 [main] INFO  (JettyStart.java [run]:97) -
08 Feb 2011 03:12:49.252 [main] INFO  (JettyStart.java [run]:101) - [Exist Versi
  on : 1.4.0]
08 Feb 2011 03:12:49.252 [main] INFO  (JettyStart.java [run]:103) - [Exist Build:
  20091111]
08 Feb 2011 03:12:49.268 [main] INFO  (JettyStart.java [run]:105) - [Exist Home:
  C:\exist]
08 Feb 2011 03:12:49.268 [main] INFO  (JettyStart.java [run]:107) - [SUN Revisio
  n : 154480]
08 Feb 2011 03:12:49.268 [main] INFO  (JettyStart.java [run]:115) - [Operating S
  ystem : Windows 7 6.1 x86]
08 Feb 2011 03:12:49.268 [main] INFO  (JettyStart.java [run]:118) - [jetty.home:
  C:\exist\tools\jetty]
08 Feb 2011 03:12:49.268 [main] INFO  (JettyStart.java [run]:120) - [log4j.confi
  guration : file:/C:/exist/log4j.xml]
08 Feb 2011 03:13:09.453 [main] INFO  (FileResource.java [client]:68) - Checkin
  g Resource aliases
08 Feb 2011 03:13:09.545 [main] INFO  (HttpServer.java [setStatsOn]:1130) - Stat
  istics on = false for org.nutzbay.jetty.Server@112da480
08 Feb 2011 03:13:09.561 [main] INFO  (HttpServer.java [doStart]:684) - Version
  Jetty/5.1.12
08 Feb 2011 03:13:09.298 [main] INFO  (Container.java [start]:74) - Started org.
  nutzbay.jetty.servlet.WebApplicationHandler@13d14e82
  Logging already initialized. Skipping
08 Feb 2011 03:13:10.195 [main] INFO  (Container.java [start]:74) - Started WebA
  pplicationContext[/exist, exist.xml Database]
08 Feb 2011 03:13:10.031 [main] INFO  (SocketListener.java [start]:205) - Starte
  d SocketListener on 0.0.0.0:8080
  nutzbay.jetty.Server@112da480
08 Feb 2011 03:13:10.632 [main] INFO  (JettyStart.java [run]:174) - ----------
08 Feb 2011 03:13:10.633 [main] INFO  (JettyStart.java [run]:175) - exist-db has
  started on port 8080. Configured contexts.
  host:8080/exist
08 Feb 2011 03:13:10.638 [main] INFO  (JettyStart.java [run]:179) - ----------
```
Start/Stop JBOSS middleware
run Binding Factory
ACT Module

ACT database

applicable events
Administer ACT event database
Establish character sets for multiple languages
Administer ACT event database
Establish global variables for cognitive metadata
Administer ACT event database
Establish schema privileges for cognitive metadata
Administer ACT event database
Establish session variables for cognitive metadata
Administer ACT event database
Establish constraints for cognitive metadata
Administer ACT event database
Review cognitive metadata table catalog
ACT database

applicable events

Cognitive Metadata

ACT Module
Administer ACT event database
Code smart query - cognitive metadata
Administer ACT event database
Code smart query – set cognitive metadata timer
Administer ACT event database
Establish cognitive metadata collection sites
Administer ACT event database Establish collection sites - backup
ACT represents violations and the configurations that are causing them.

Policy violations

State of the system

ACT Module

ACT database

Applicable events

ACT Use cases

Cognitive Metadata
Administer ACT event database
Establish cognitive metadata collection sites

- **Currently these are “test” event collection sites.**
  - For each ACT Use Case the event collection site varies
  - For example,
    - Web Data warning Use Case ACT will collect from “known” problem sites like CRYPTOME, etc.
    - Public Cloud Data Spillage Prevention Use Case ACT collects from the Public Cloud Storage site
NIST Public Clouds
ACT Use Cases are defined

Use Cases expand as Dynamic Policies and Violations change
ACT Use Cases – Uses effective reporting standards set by GAO

- Focus on risks
- Involve stakeholders
- Assign accountability
- Link to business goals

Source: GAO.
Establish LINK TO POLICIES

ACT database

applicable events

Cognitive Metadata

Dynamic policies related to violations

ACT Module

ACT Use cases

state of the system

policy violations
ACT represents violations and the configurations that are causing them.

Policy violations

State of the system

ACT Module

ACT database

Applicable events

Cognitive Metadata

Dynamic policies

Augmented Governance