SEMANTIC TECHNOLOGY IN THE FEDERAL SECTOR

Status and Prospects

April 28, 2006

Presentation to:
Semantic Days 2006
Stavanger, Norway

Mills Davis
Managing Director
Project10X

MILLS DAVIS

202-667-6400 • mdavis@project10x.com

- Mills Davis is Project10X's founder and managing director for industry research and strategic programs. He consults with technology manufacturers, global 2000 corporations, and government agencies on next-wave semantic technologies and solutions.

- Mills serves as lead for the Federal CIO council's Semantic Interoperability Community of Practice (SICoP) research into the business value of semantic technologies. Also, he is a founding member of the AIIM interoperable enterprise content management (iECM) working group, and a founding member of the National Center for Ontology Research (NCOR).

- A noted researcher and industry analyst, Mills has authored more than 100 reports, whitepapers, articles, and industry studies.

Topics

- Semantic technology in the Federal sector
  - Quick overview of current status and prospects

- Semantic Wikis — for collaboration, information sharing, knowledge management, and knowledge computing
  - Drill down into one area of current focus

- Pilot scenario: Federal information sharing
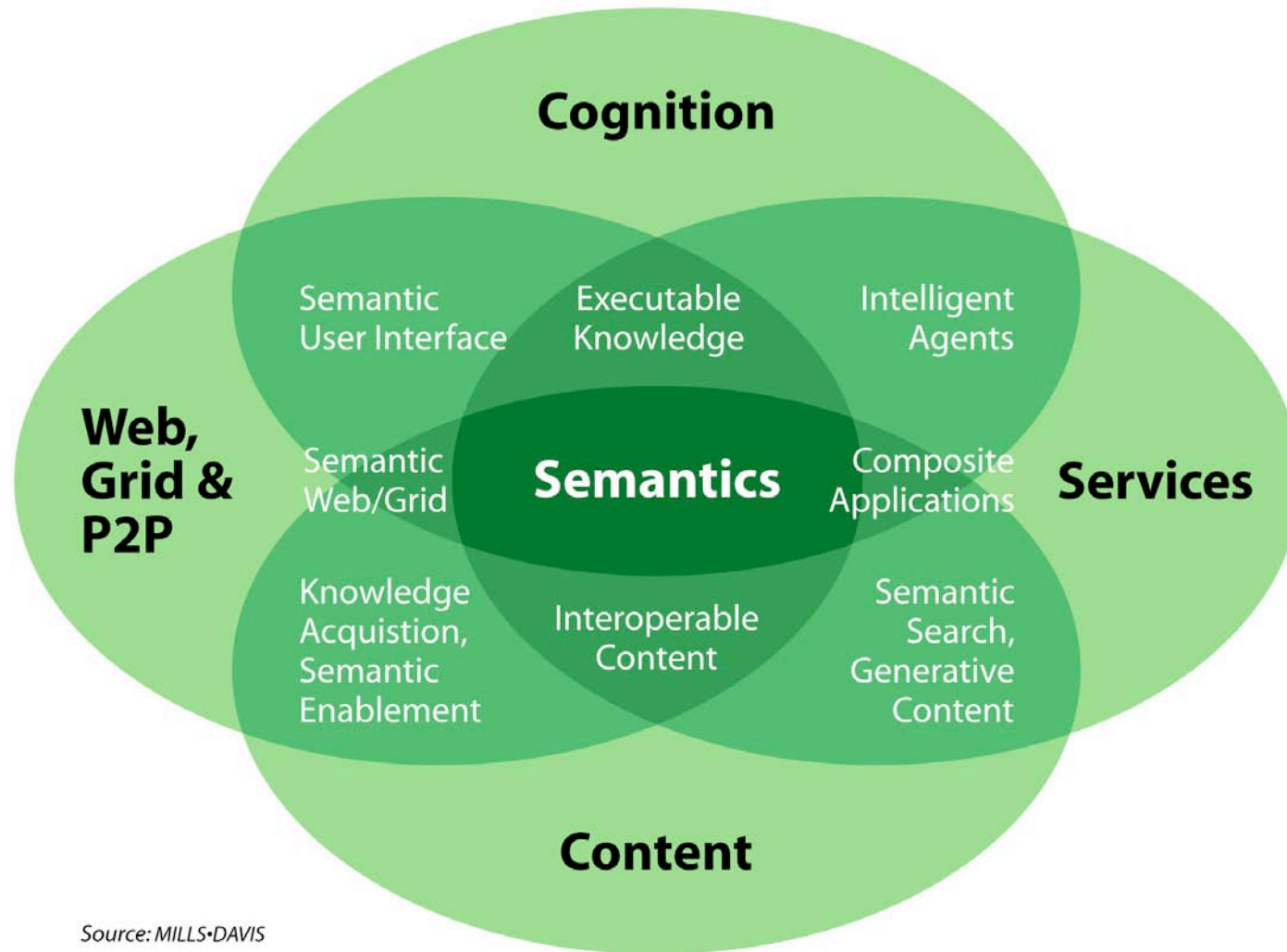 - One scenario where semantic technologies will add value

SEMANTIC TECHNOLOGY
IN THE
FEDERAL SECTOR

Status of Semantic Technologies in the Federal Sector

- Semantic technology is a central R&D theme across federal agencies
- More than 50 Federal agencies and departments are actively engaged.
- Expenditures for semantic technology accelerating in DoD, Intelligence, Security, and Life Sciences, and civilian agencies, including increasing SBIR activity
- Ongoing pilot programs in enterprise architecture, net-centric infrastructure, service oriented architecture, situation awareness, intelligence, composite applications, emergency management, logistics, compliance, content interoperability, knowledge management, semantic search, semantic wikis, e-learning, in silico research, simulation-based acquisition, robotics, decision support, and other areas.
- Production deployments planned, but also gated by issues of scalability, complex reasoning performance, reference knowledge life cycle management, version control and change management, trust, and simplification of user interface.

Semantic Wave R&D:
Semantic technologies are a central theme across a broad array of Federal research and development initiatives.

Cognition

Web, Grid & P2P

Services

Content

Semantic User Interface

Executable Knowledge

Intelligent Agents

Semantic Web/Grid

Semantics

Composite Applications

Knowledge Acquistion, Semantic Enablement

Interoperable Content

Semantic Search, Generative Content

Source: MILLS·DAVIS

Federal Agencies involved with Semantic Technologies:
Defense, Intelligence, Security, Law Enforcement, Life Sciences, Medicine, Administration, Revenue, Health, Energy, Environment, Interior, Archives, Budget, Acquisition, Human Capital, Electronic Government, Executive Offices,
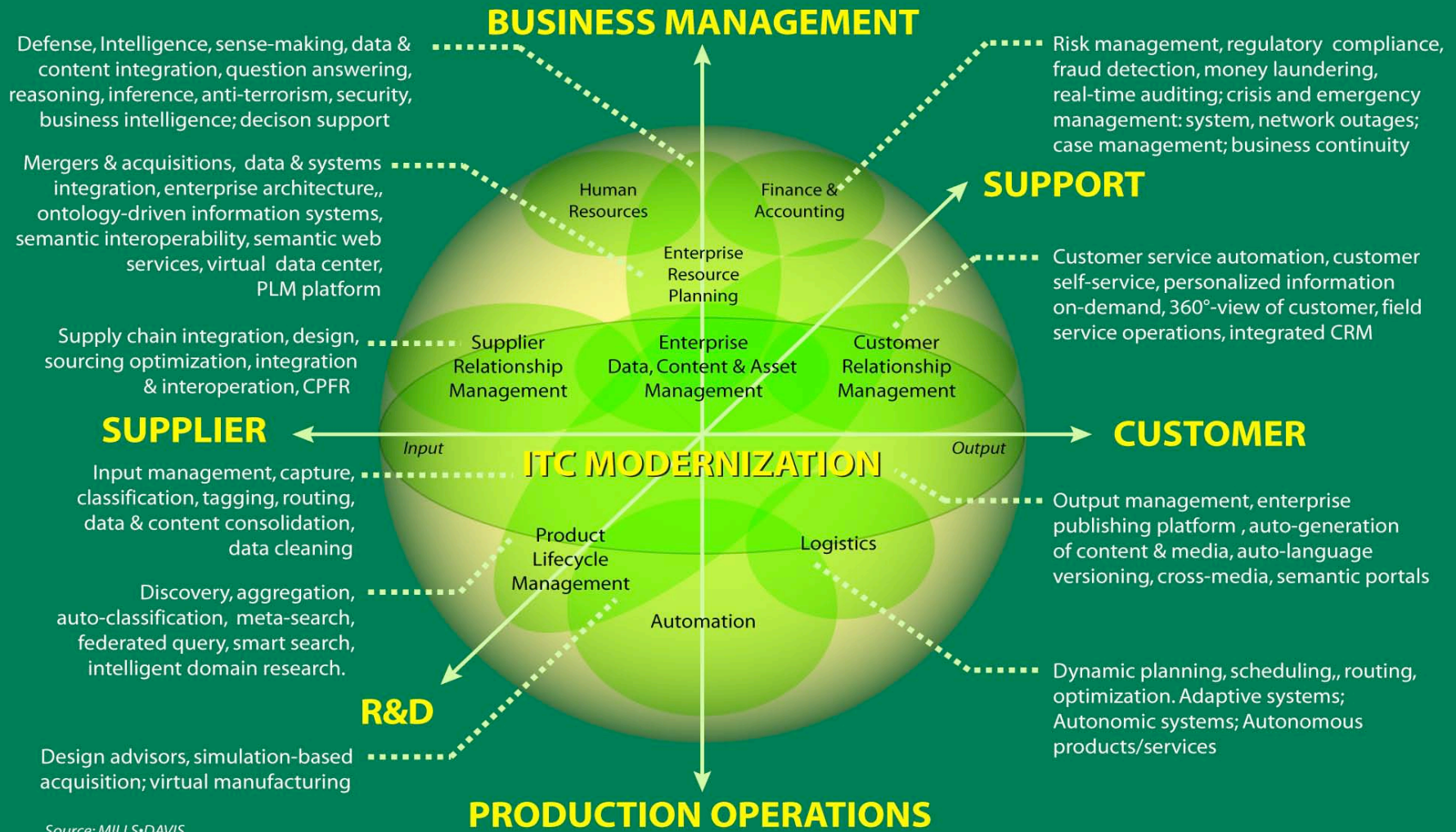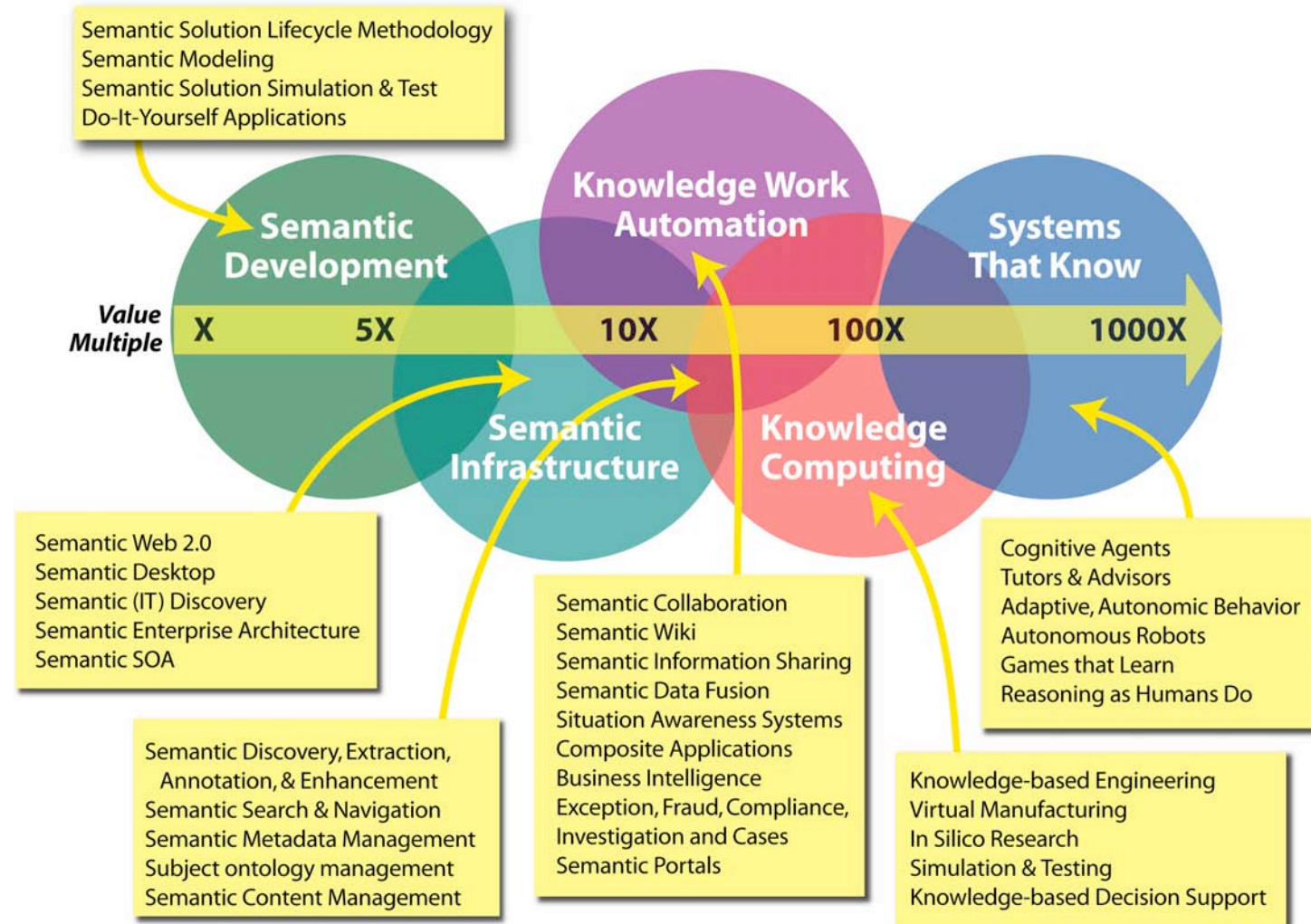
Semantic Applications Everywhere:

Customer-facing services, "smart" products & services, design & manufacture, research, supplier-facing processes, infrastructure & integration, intelligence, security, risk, & decision support

BUSINESS MANAGEMENT

Defense, Intelligence, sense-making, data & content integration, question answering, reasoning, inference, anti-terrorism, security, business intelligence; decison support

Mergers & acquisitions, data & systems integration, enterprise architecture,, ontology-driven information systems, semantic interoperability, semantic web services, virtual data center, PLM platform

Supply chain integration, design, sourcing optimization, integration & interoperation, CPFR

SUPPLIER

Input management, capture, classification, tagging, routing, data & content consolidation, data cleaning

Discovery, aggregation, auto-classification, meta-search, federated query, smart search, intelligent domain research.

R&D

Design advisors, simulation-based acquisition; virtual manufacturing

Risk management, regulatory compliance, fraud detection, money laundering, real-time auditing; crisis and emergency management: system, network outages; case management; business continuity

SUPPORT

Customer service automation, customer self-service, personalized information on-demand, 360°-view of customer, field service operations, integrated CRM

CUSTOMER

Output management, enterprise publishing platform , auto-generation of content & media, auto-language versioning, cross-media, semantic portals

Dynamic planning, scheduling,, routing, optimization. Adaptive systems; Autonomic systems; Autonomous products/services

PRODUCTION OPERATIONS

Human Resources

Finance & Accounting

Enterprise Resource Planning

Supplier Relationship Management

Enterprise Data, Content & Asset Management

Customer Relationship Management

Input

ITC MODERNIZATION

Output

Product Lifecycle Management

Logistics

Automation

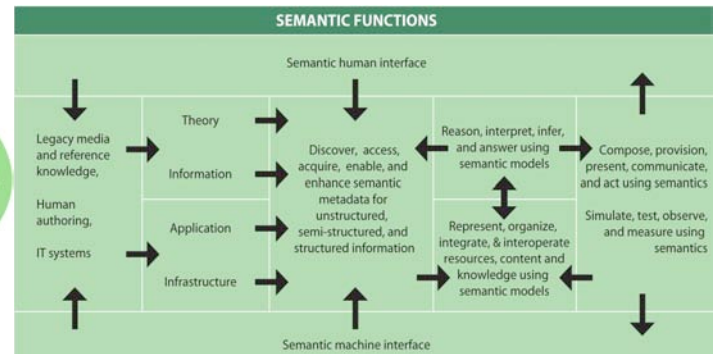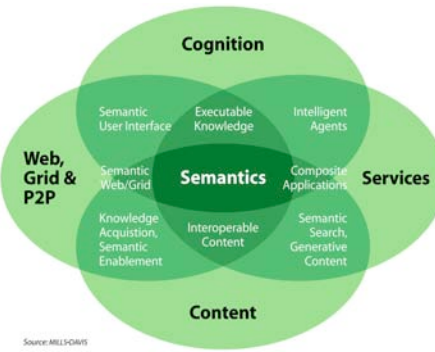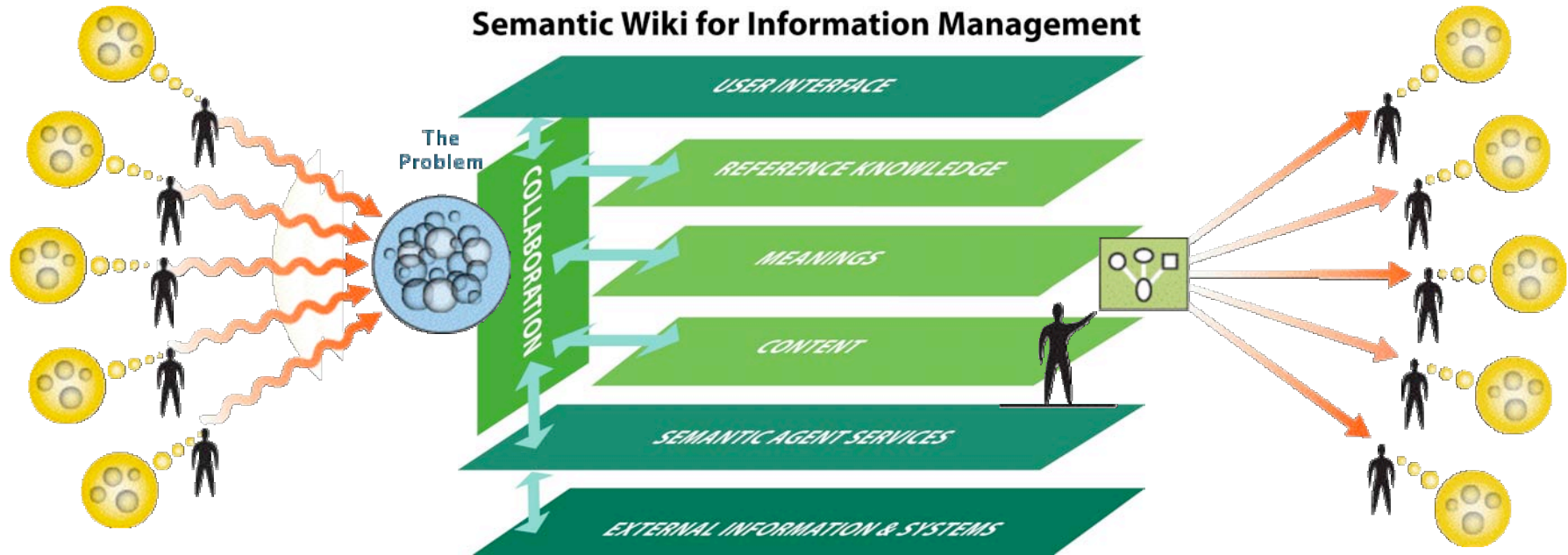Source: MILLS·DAVIS

Investing Across the Semantic Value Spectrum:
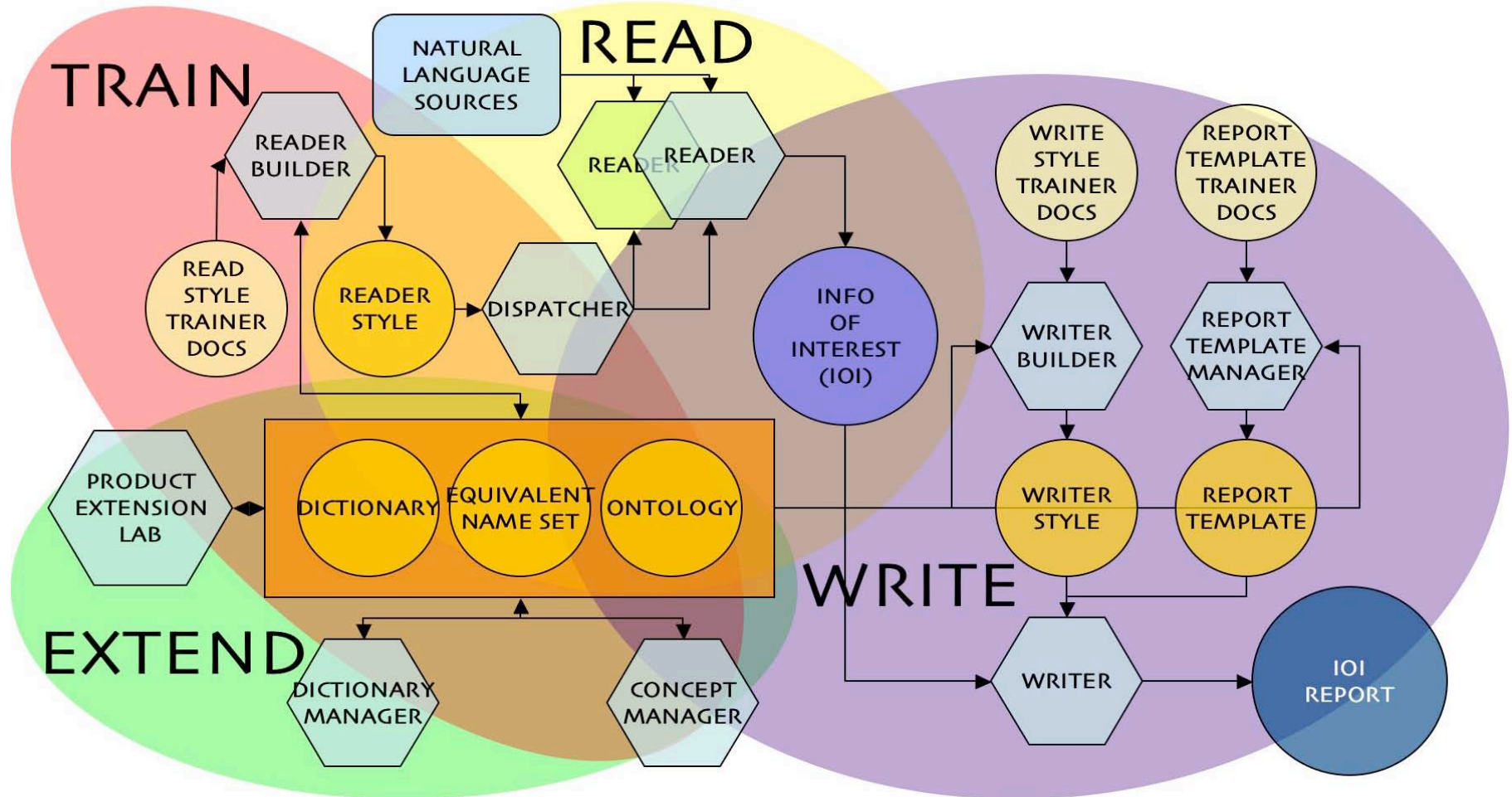## *Different capabilities power different levels of return*

Semantic Solution Lifecycle Methodology
Semantic Modeling
Semantic Solution Simulation & Test
Do-It-Yourself Applications

Semantic Development

Knowledge Work Automation

Systems That Know

Value Multiple

X   5X   10X   100X   1000X

Semantic Infrastructure

Knowledge Computing

Semantic Web 2.0
Semantic Desktop
Semantic (IT) Discovery
Semantic Enterprise Architecture
Semantic SOA

Cognitive Agents
Tutors & Advisors
Adaptive, Autonomic Behavior
Autonomous Robots
Games that Learn
Reasoning as Humans Do

Semantic Discovery, Extraction,
   Annotation, & Enhancement
Semantic Search & Navigation
Semantic Metadata Management
Subject ontology management
Semantic Content Management

Semantic Collaboration
Semantic Wiki
Semantic Information Sharing
Semantic Data Fusion
Situation Awareness Systems
Composite Applications
Business Intelligence
Exception, Fraud, Compliance,
   Investigation and Cases
Semantic Portals

Knowledge-based Engineering
Virtual Manufacturing
In Silico Research
Simulation & Testing
Knowledge-based Decision Support

Source: MILLS•DAVIS

SEMANTIC WIKIS

Semantic Wiki:
Emerging platform for collaborative authoring, information sharing, knowledge management, and knowledge computing
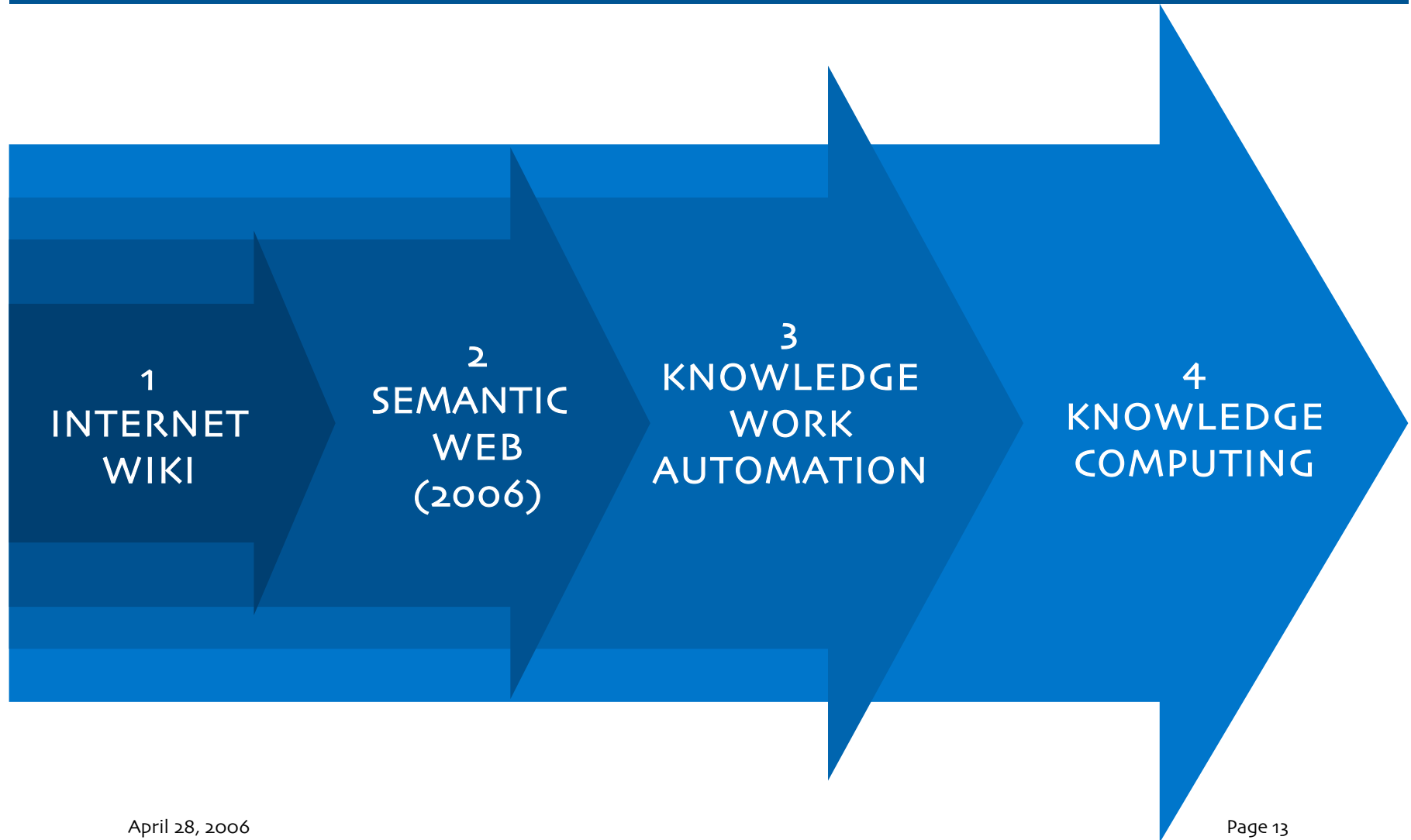

Semantic Wiki for Information Management

Semantic Wiki Automation: Train, Read, Write

Source: Semantic Insights

Semantic Wiki:
Capability Roadmap

1
INTERNET WIKI

2
SEMANTIC WEB
(2006)

3
KNOWLEDGE WORK AUTOMATION

4
KNOWLEDGE COMPUTING

Stage-1.   Internet Wiki

CAPABILITY	DESCRIPTION
Development	• Web-based environment for multi-user reading and writing / editing of content • Light-weight, easy to use • Little or no training required
Infrastructure	• Web site • Internet, intranet, desktop deployments • Manual roles and security • Multi-user access
Collaborative Authoring	• Manual authoring, editing, reading • Text and wiki syntax • Roll-back • Evolving to include all types of content – table, chart, diagram, image, audio, video (370 formats…) • Cleaner presentation • Version histories
Search & Navigation	• Link navigation • Text search

Stage-2. Semantic Web (circa 2006)

CAPABILITY	DESCRIPTION
Development	• Essentially manual collaboration environments for content and reference knowledge development, using semantics to improve access to wiki content through semantic search and navigation.
Infrastructure	• 3-layer web site (content– semantics– UI) using semantic web standards (RDF & OWL) • Semantics / policy-based roles and security • Network appliances
Collaborative Authoring	• Manual authoring, tagging, annotation of semantics and reference knowledge • Several semantic wiki projects for group development and management of ontologies for web-based information, including editing to achieve semantic interoperability of ontologies, including upper level ontologies
Search & Navigation	• 15-20 semantic wiki projects adding tags & annotations to content to create & navigate a semantic layer • Search and navigate via topic maps, search via clusters, categories, facets • Context-sensitive UI • Some semantic wikis provide limited inferences about content by reasoning over RDF and OWL models.

Stage-3. Knowledge Work Automation

CAPABILITY	DESCRIPTION
Development	•Comprehensive automation of techno-social processes of knowledge work using semantic technologies •E.g., automate project life cycle research functions including auto-reading and modeling of reference knowledge (such as policy, standards, documentation); auto/semi-auto build ontology; also, auto reading issue-related information sources; semantically harmonize different points of view
Infrastructure	•Website, application server, P2P, tiered & mobile deployments •Automate infrastructure mapping functions: auto-discovery and latent semantic indexing of IT artifacts – data, systems, communications
Collaborative Authoring	•Automate content research functions: Train, Read, Understand, Write •E.g., Automatically read, interpret and report on 100 documents (or topics) in a particular writing style. •Automate semantic enablement & enhancement of content
Search & Navigation	•High quality user experience; Information in users context •Semantic search and navigation across multiple wikis and external content sources; Visual navigation via concept maps, information graphics, and other models •Automate semantically-enabled content access •E.g., Read body of knowledge/ information, to automate development of UI providing semantic access via concept/ relationship structure

Stage-4. Knowledge Computing

CAPABILITY	DESCRIPTION
Development	• Putting reference knowledge to work! E.g., Do-it-yourself applications and data mash-ups; scalable executable reference knowledge; complex reasoning and behaviors • Automate composition and provisioning of functionality, data, and services (e.g., composite applications) • Automate version control and change management
Infrastructure	• Semantic agents and services • Adaptive, autonomic, and autonomous behaviors • Composite applications
Collaborative Authoring	• Automate content authoring: • Automate knowledge modeling, ontology development, alignment, and harmonization • Knowledge-based computing — (e.g., life sciences in silico research; intelligence sense making, data fusion, and scenario analysis; emergency management and decision support)
Search & Navigation	• Question answering • Decision support • Smart behaviors

PILOT SCENARIO: FEDERAL INFORMATION SHARING

Current Situation

- Annual spending on IT by the Federal government is approximately $70 billion. Across government agencies, there exist millions of web pages, databases, document repositories, file systems, and records archives.

- For Federal agencies, discovery, understanding, and sharing of information across organizational boundaries, lines of business, and functions present a monumental challenge since information sharing activities are fragmented, across multiple agencies, information areas, and standards initiatives.

- Federal agencies sponsor and participate in a multiplicity of programs that target sharing of information internally and externally, e.g.: DOD, Intelligence Community, DHS, HHS, DOC.

- Several agencies have cross-agency missions requiring record, document, and information sharing, e.g.: OMB, GSA, GPO, NARA.

Policy Guidance

- Congressional legislation, Presidential policies, and agency directives establish requirements to categorize and share information across organizational boundaries — with citizens, business organizations, and other agencies of government at all levels, both domestic and international.

- Office of Management and Budget (OMB) guidance regarding the sharing of information requires agencies to: (a) harmonize data resources, (b) categorize them for discovery, (c) expose them as services for sharing, and (d) publish authoritative information that will enable other agencies, levels of government, industry, interested parties, and the public to find, access, and understand this information.

- Agencies are instructed and graded on how well they implement information sharing that is consistent with the policy guidance and data reference model (DRM) set forth in the Federal Enterprise Architecture.

Role of Communities of Interest

- Information sharing policy guidance for civilian and defense agencies lays emphasis *communities of interest* (COIs) as a means to harmonize related areas of information across organizational boundaries.

- A community of interest is a cross-organization, cross-business line, and cross-function group of information producers and consumers that needs to share information to accomplish some mission, business, or social purpose. An example would be a community of interest relating to geo-spatial information.

- By definition, members of communities of interest are collaborative bodies that do not report through a common management and administrative structure. Focusing on communities of interest recognizes the importance social agreements and shared meanings (semantics) for effective information sharing.

Key Challenges

- How do we empower communities of interest with cost-effective collaboration environments, governance processes, and lifecycle knowledge management to enable their work?

- How do we knit together the work of various communities to enable a larger fabric of cost-effective Federal information sharing and access to emerge?

SICOP/SWIM
Semantic Interoperability Community of Practice
Semantic Wiki for Information Management Working group

Point of departure:

- Information sharing is a techno-social process.

- Shared understanding (semantic interoperability) emerges organically and is sustained through active involvement of interested communities.

- Approaches to information sharing directed top-down have been tried repeatedly and have failed to achieve the desired result.

- Goal of SICOP/SWIM is to give communities of interest the tooling and best practices to build trusted reference knowledge and manage it across the entire lifecycle.

SICOP/SWIM
Semantic Interoperability Community of Practice
Semantic Wiki for Information Management Working group

Strategy:

- Pilot test semantic wikis as a light-weight, easily used, and extensible environment for cross-organizational and community of interest collaboration, information sharing, and knowledge management.

- Research and develop a semantic model-of-models that interrelates standards efforts being developed and used by different communities.

- Link together agency and COI based information sharing initiatives and support business line initiatives that require information sharing across multiple COIs and organizations.

DoD Net-Centric Data Strategy

DoD Net-Centric Data Strategy

What is a COI?

- COIs are described in the DoD Net-Centric Data Strategy
- A COI is …
  - a *Community*
  - *Of* people
  - who are all *Interested* in something
  - and need to share information
- What does a COI do?
  - *Work together to resolve the issues that affect their community*
  - *Establish community standards on how information will be exchanged within the COI*
- What can't a COI do?
  - *COIs do not operate systems or provide services*
  - *COIs do not submit POMs*
  - *COIs do not direct changes to ICDs, ORDs, CDDs, or CPDs*

However, *members* of COIs *do!*

A Net-Centric DoD / NIICIO

7

Pilot Scope

Scope:

(1) What programs of records or other sources will advertise data as a web-service IAW the agreed COI vocabulary?

(2) What value-added services will be demonstrated?

(3) What network(s) will be used to demonstrate net-centric capabilities?

(4) What joint exercise(s) will be used to demonstrate net-centric capabilities?

(5) What organizations are participating?

A Net-Centric DoD / NIICIO

12

COI Pilot Systems Architecture

GS COI Blue Force Current + 18 Hours Service UDOP **(DRAFT)**

Consumers — Developers

VISIBLE

ACCESSIBLE   UNDERSTANDABLE

NCES Services

Producers/

SAMPLE

Sample COI Organization Chart
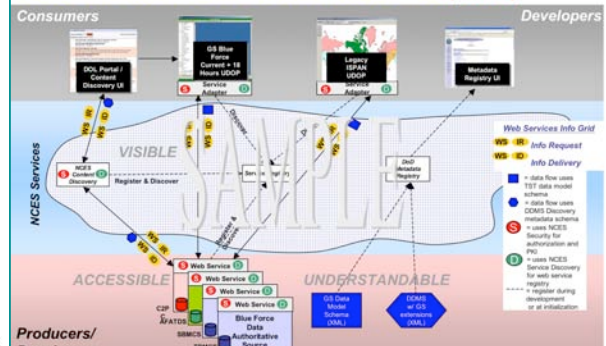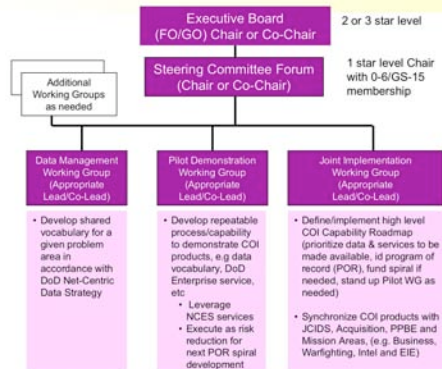
Executive Board (FO/GO) Chair or Co-Chair — 2 or 3 star level

Steering Committee Forum (Chair or Co-Chair) — 1 star level Chair with 0-6/GS-15 membership

Additional Working Groups as needed

Data Management Working Group (Appropriate Lead/Co-Lead)
- Develop shared vocabulary for a given problem area in accordance with DoD Net-Centric Data Strategy

Pilot Demonstration Working Group (Appropriate Lead/Co-Lead)
- Develop repeatable process/capability to demonstrate COI products, e.g data vocabulary, DoD Enterprise service, etc
  - Leverage NCES services
  - Execute as risk reduction for next POR spiral development

Joint Implementation Working Group (Appropriate Lead/Co-Lead)
- Define/implement high level COI Capability Roadmap (prioritize data & services to be made available, id program of record (POR), fund spiral if needed, stand up Pilot WG as needed)
- Synchronize COI products with JCIDS, Acquisition, PPBE and Mission Areas, (e.g. Business, Warfighting, Intel and EIE)

FOR DISCUSSION PURPOSES ONLY

A Net-Centric DoD / NIICIO

10

COI Pilot POA&M

High-level Graphic with dependencies, decision points, and final demonstrated illustrated.

Kickoff COI
Select Pilot
2nd Exec BD Meeting
Metadata Vocab
3rd Exec Bd Meeting
Interim Pilot Capability
4th Exec BD Meeting
1st COI Anniversary

Stand up WGs
Capabilities Plan
Services Inventory
Potential Agency Budget Inputs
Integrate Pilot into Ops
O&M Plan

SAMPLE

A Net-Centric DoD / NIICIO

14
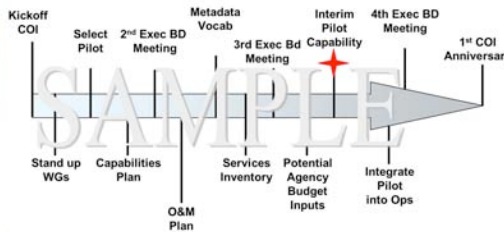
COI Pilot Metrics

Metrics to assess the return on investment (ROI) (resources as well as net-centric capabilities and agility) of the pilot.

Start-point:

#1 - Changes and impact to Programs of Record (POR) involved in the COI Pilot

#2 - Initial and incremental costs of web service interfaces to advertise Program of Record (POR) data

#3 - User assessment of demonstrated net-centric capabilities

#4 – Feedback on ease of use and adoption of CES pilot services

#5 – Level of effort to agree on initial COI vocabulary

A Net-Centric DoD / NIICIO

16

Semantic Wiki Scenario:
COI develops & publishes reference knowledge for net-centric information sharing

COI Governance Established?

1 — Yes / No

- Convene the COI
- ID executive
- ID steering group
- ID core team
- Define decision process, criteria & governance
- Prepare charter
- Prepare plan & schedule

Semantic Wiki Established?

2 — Yes / No

- Set-up site
- Set-up capabilities and agent services
- Set-up downloads
- Set-up directories
- Light-up site

COI Personnel Assigned?

3 — Yes / No

- Invite to join
- Sign-up / profile
- Establish security
- Establish roles

Authoritative Sources Mined?

4 — Yes / No

- Discover policy, standards, & documentation sources
- Discover system, data, & knowledge artifacts
- Access, collect, link to sources
- Semantically enable/enhance sources

Harmonization Complete?

5 — Yes / No

- ID, describe, map, characterize, infomation elements
- Analyze syntactical, structural, and semantic relations
- Propose harmonized terms, relationships, & semantic models
- Review, vet, & reach agreements

Reference Knowledge Packaged to Standard?

6 — Yes / No

- Define & format: discovery metadata access services, & semantics for understanding data assets
- Publish models and specifications in directory formats, and documentation
- Test for validity

Known producer/consumer data exchange OK?

7 — Yes / No

Pilot scenario for information sharing between known entities:
- Publish reference knowledge pkg
- Discover data/svcs
- Understand info semantics
- Access data svcs
- Evaluate & revise

System-X data exchange OK?

8 — Yes / No

Pilot scenario for info sharing with authorized but unanticipated user:
- Post reference knowledge pkg
- System-X discovers, accesses, understands and makes interoperable within composite application
- Evaluate & revise as needed

Publication of COI Reference Knowledge Version Authorized?

9 — Yes / No

- Publish latest version of reference knowledge to directories and wiki
- Release review & approval process
- Authorization recommendation
- Authorized release
- Post release.

Change Occured Since Release?

10 — Yes / No

- Monitor sources
- Auto-detect changes and assess scope and impact
- Initiate change management process

REFERENCES

HTTP://WWW.SEMANTIC-CONFERENCE.COM/SEMANTICWAVE.HTML

HTTP://WWW.PROJECT10X.COM/PAGES/PUBLICATIONS.HTML

HTTP://WEB-SERVICES.GOV/

HTTP://COLAB.CIM3.NET/CGI-BIN/WIKI.PL

# Ten scenarios where semantic wikis add value

Tsunami 2005	Global ad hoc emergency response involving 150+ volunteers, round the clock, use wiki to collect, organize, and share information about survivors and relief needs. This collaboration outperforms commercial news sources
Intelligence Community	Intelligence Community exploring the use of blogs and wikis as means for accelerating and improving collaboration and intelligence assessment across constituent agencies and specialty areas.
IRS Integrated Navigation System	Unified topic map access to corpus of IRS publications, FAQs authored by multiple groups and separate locations. Quarterly cross-organizational collaboration required to resolve differences.
Life Sciences Medical Health	Multi-decade efforts to standardize vocabularies, taxonomies, thesauri, and subject ontologies for life sciences, medical research, and clinical use reaching stage where responsibility for curation and ongoing development must shift to research communities
NIEM	National Information Exchange Model envisions integration of information sharing across Justice, Intelligence, Homeland Security, Transportation, Public Health, Emergency & Disaster Management involving Federal, State, Local levels of government.

Ten scenarios where semantic wikis add value

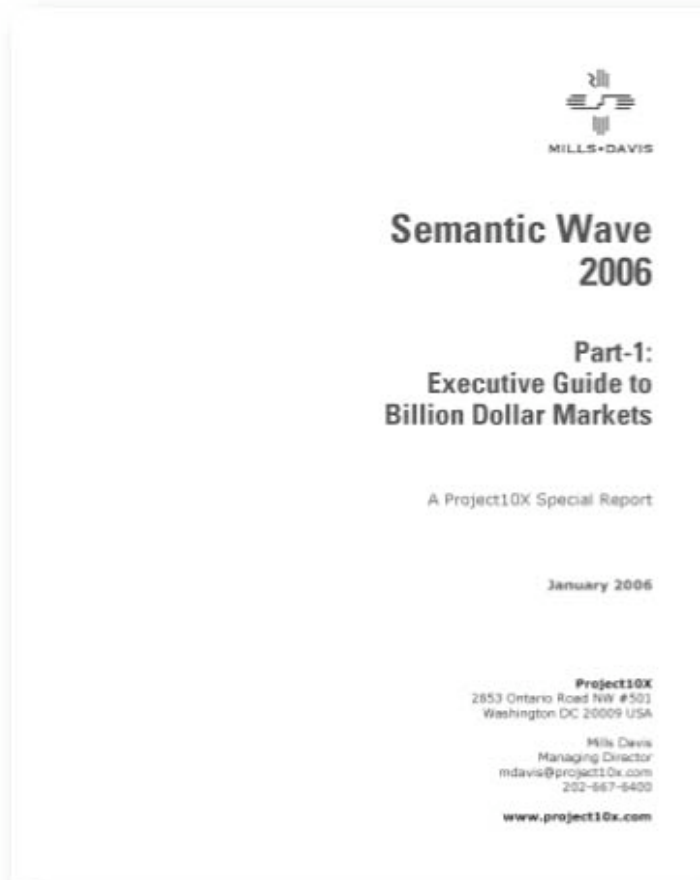| National Science Foundation	Proposals to NSF are peer-reviewed by 50,000 reviewers with documents stored in 200 different repositories. Non-invasive, collaborative process needed to handle information requests that fall outside of pre-established program areas.
Patent Office	Peer-to-Patent experiment will explore peer review process involving potentially hundreds of thousands of experts collaborating to assess prior art and evaluate patent claims before award. Requires semantic wiki and new governance approaches.
FEA DRM 2.0	9-month wiki-based process has three levels of participation: steering group, designated agency representatives, and public (over 600). Socialization of process results in well vetted recommendations, enabling OMB to issue policy guidance. Implementation will stress role of COIs in enabling sharing
DOD Net-Centric Data Strategy	DoD strategy is to move from privately owned/stored data in disparate networks and legacy systems/applications to a net-centric enterprise information environment where both known and unanticipated authorized users can publish and subscribe data and services. Implementation stresses COI pilots.
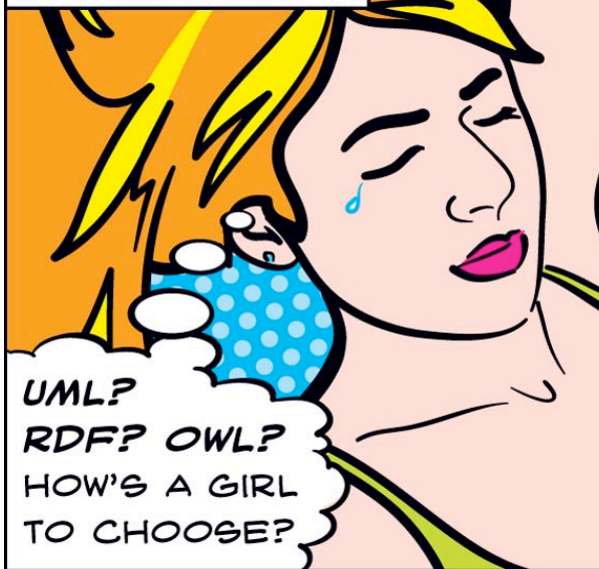
Semantic Wave 2006:
First comprehensive study of business applications and markets for semantic technologies

(http://www.semantic-conference.com/semanticwave.html)