Semantic Wikis
for Collaboration, Information Sharing, & Knowledge Management

Giving Communities of Interest (COIs) tooling and best practices to implement the Federal Enterprise Architecture Data Reference Model (DRM) and build trusted reference knowledge

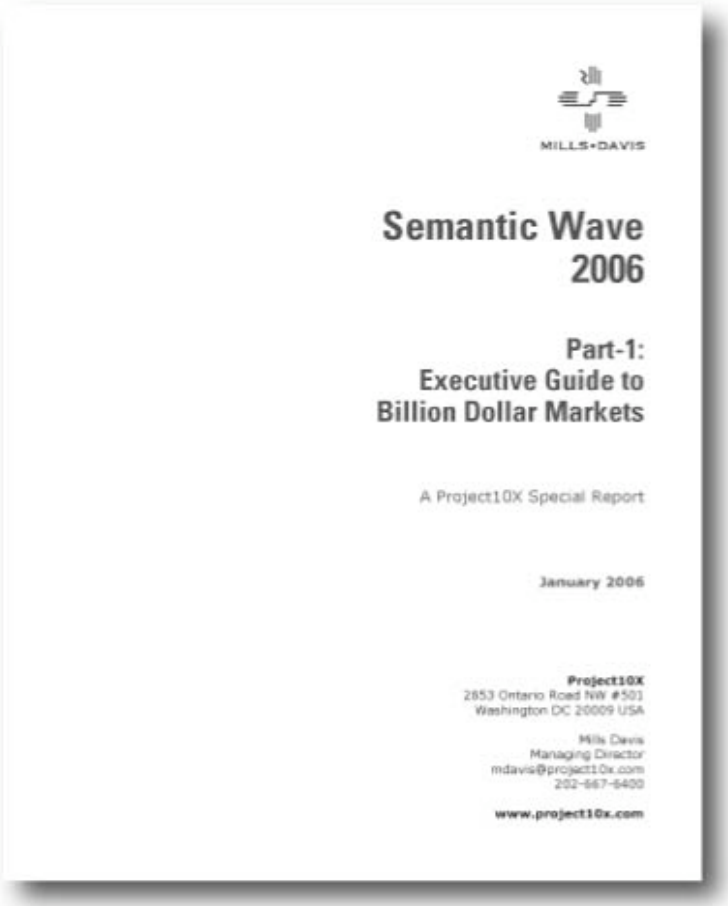
Mills Davis
Managing Director
Project10X

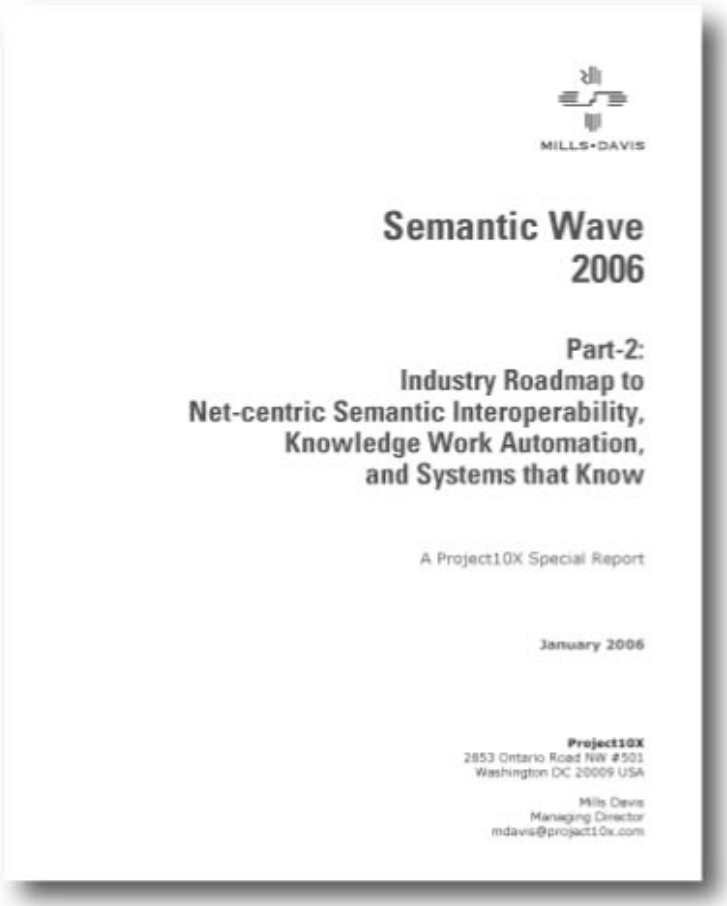
April 12, 2006

SO, WHAT HAPPENS WHEN YOU HAVE LOTS OF WEB, LOTS OF SEMANTICS, & LOTS OF SOCIAL INTERACTION WITH IT?

Semantic Wave 2006:

First comprehensive study of business applications and markets for semantic technologies





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

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

- Federal information sharing
  - Current situation
  - SICOP/SWIM approach
- Semantic wiki
  - Semantic technology
  - Wiki concepts
  - Semantic wiki concepts
- Federal sector piloting of semantic wikis
  - Ten scenarios where semantic wikis add value
  - DoD Net-centric data strategy as a case in point
- Conclusions

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

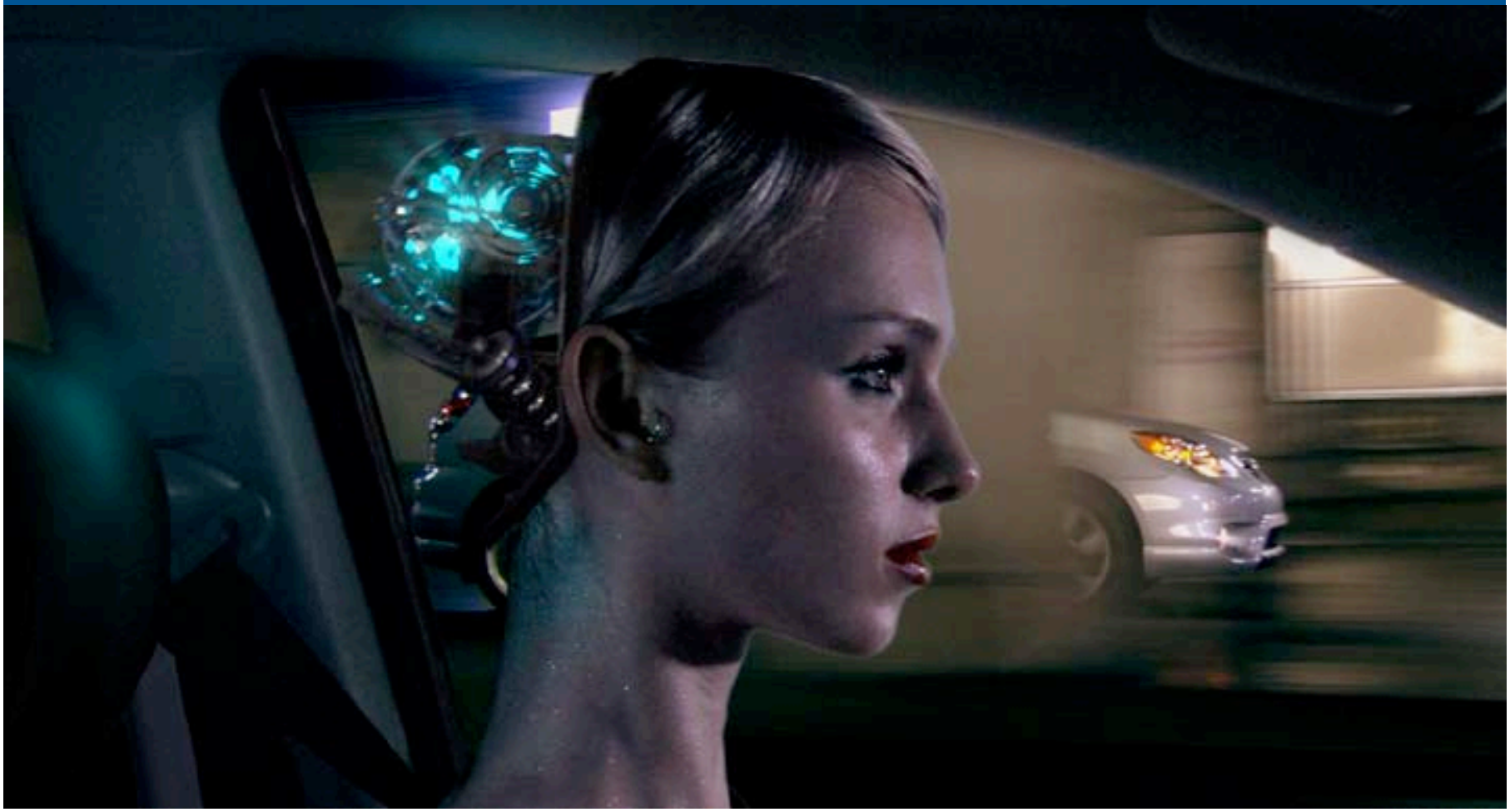
Strategy:

- 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.

SEMANTIC TECHNOLOGY

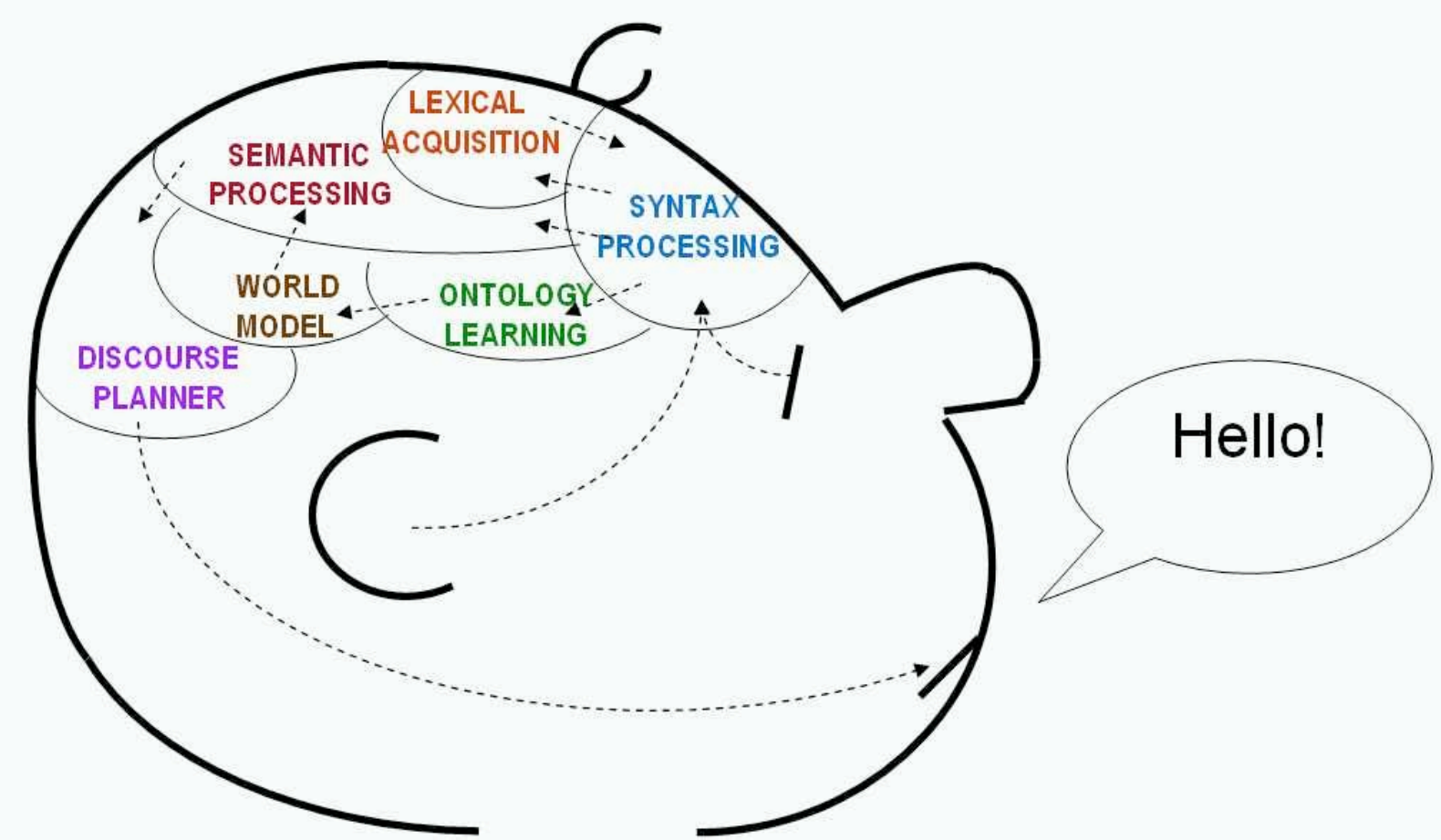
Semantic Technologies:

A shift in paradigm, technology & economics

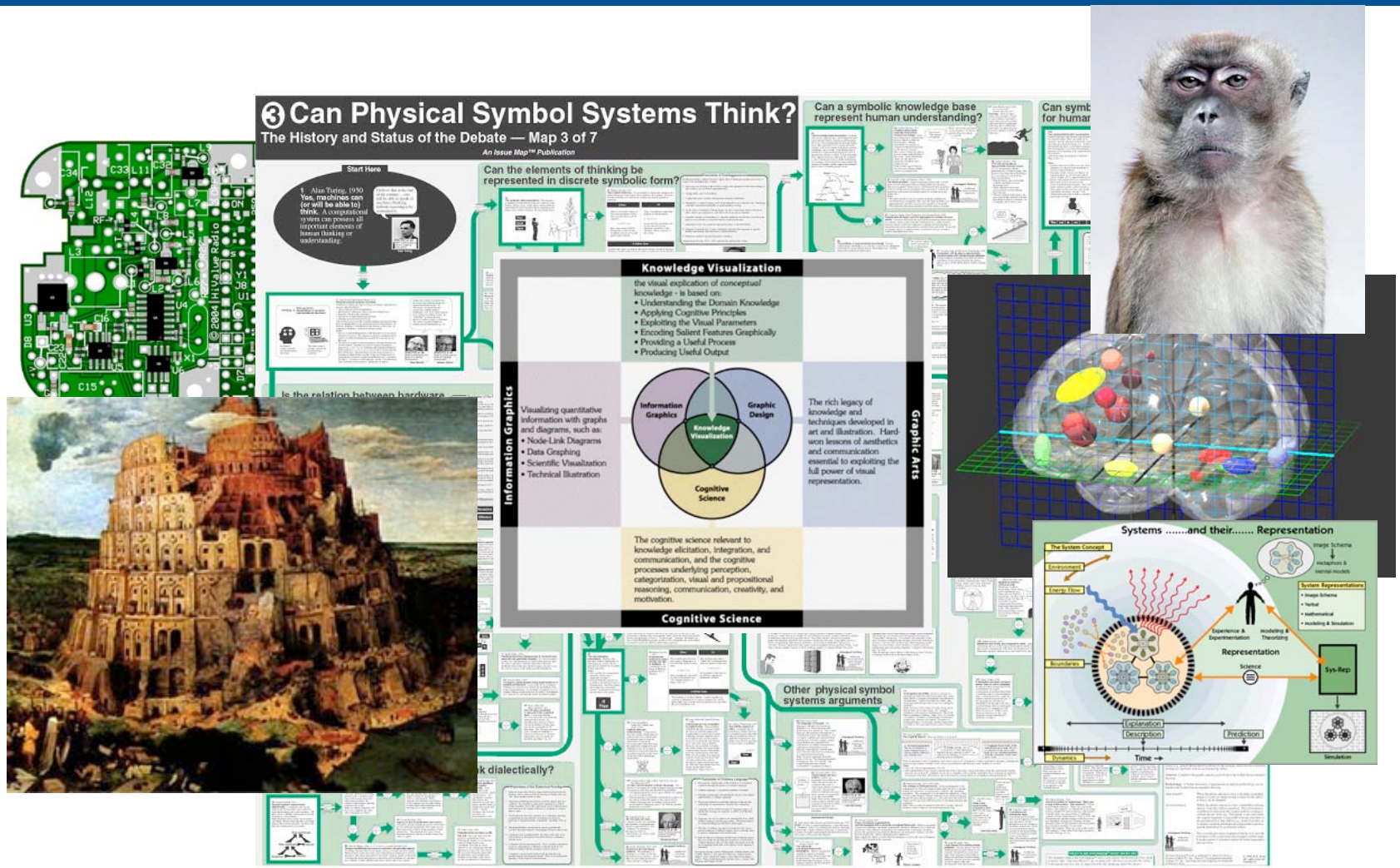


Semantic Technologies:

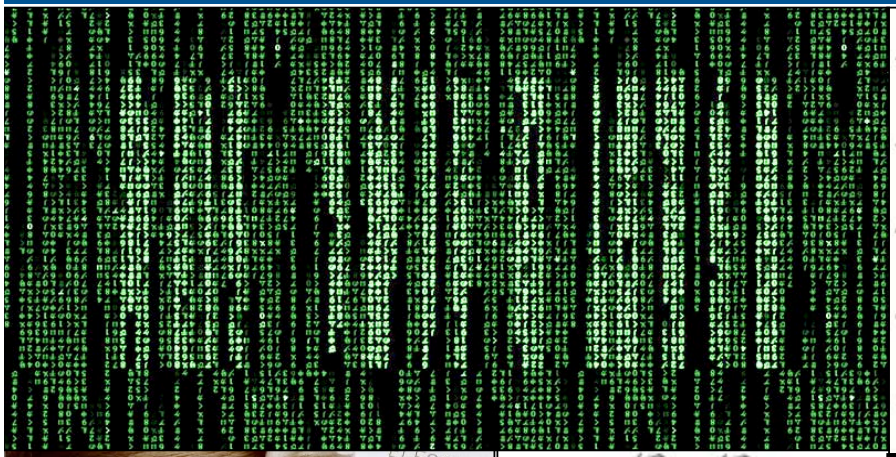
Representing meanings & knowledge about things so both computers and people can work with it

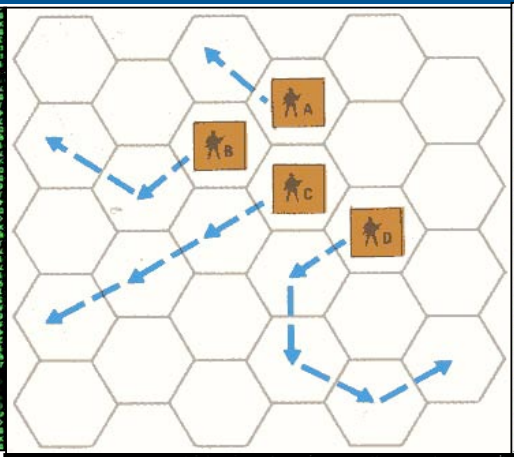


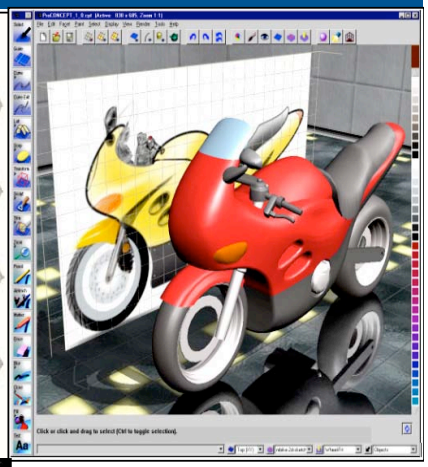
So, what do semantic technologies do?

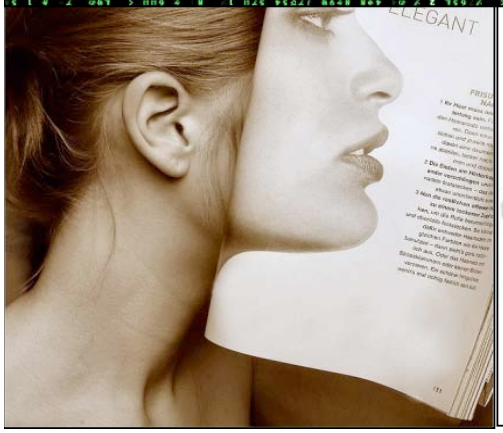


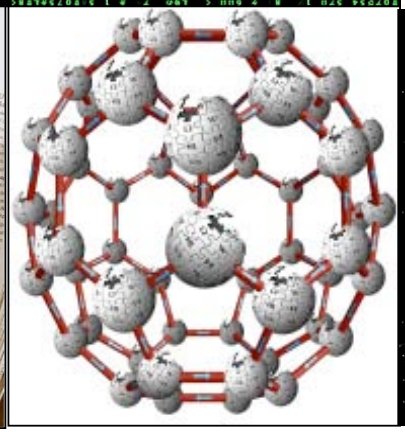
Semantic technologies model knowledge about infrastructure, information, behavior, & domain expertise separately from programs and data…





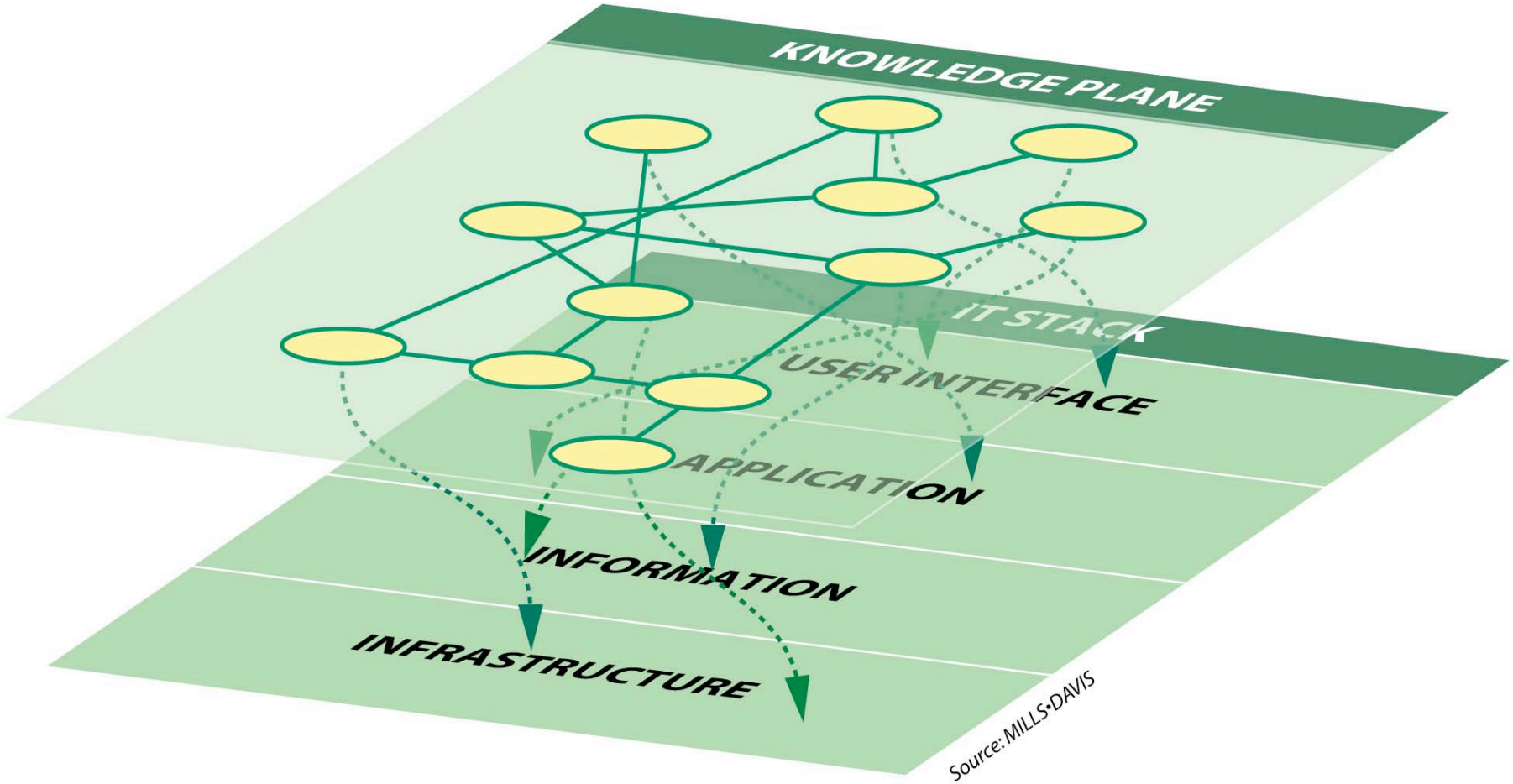






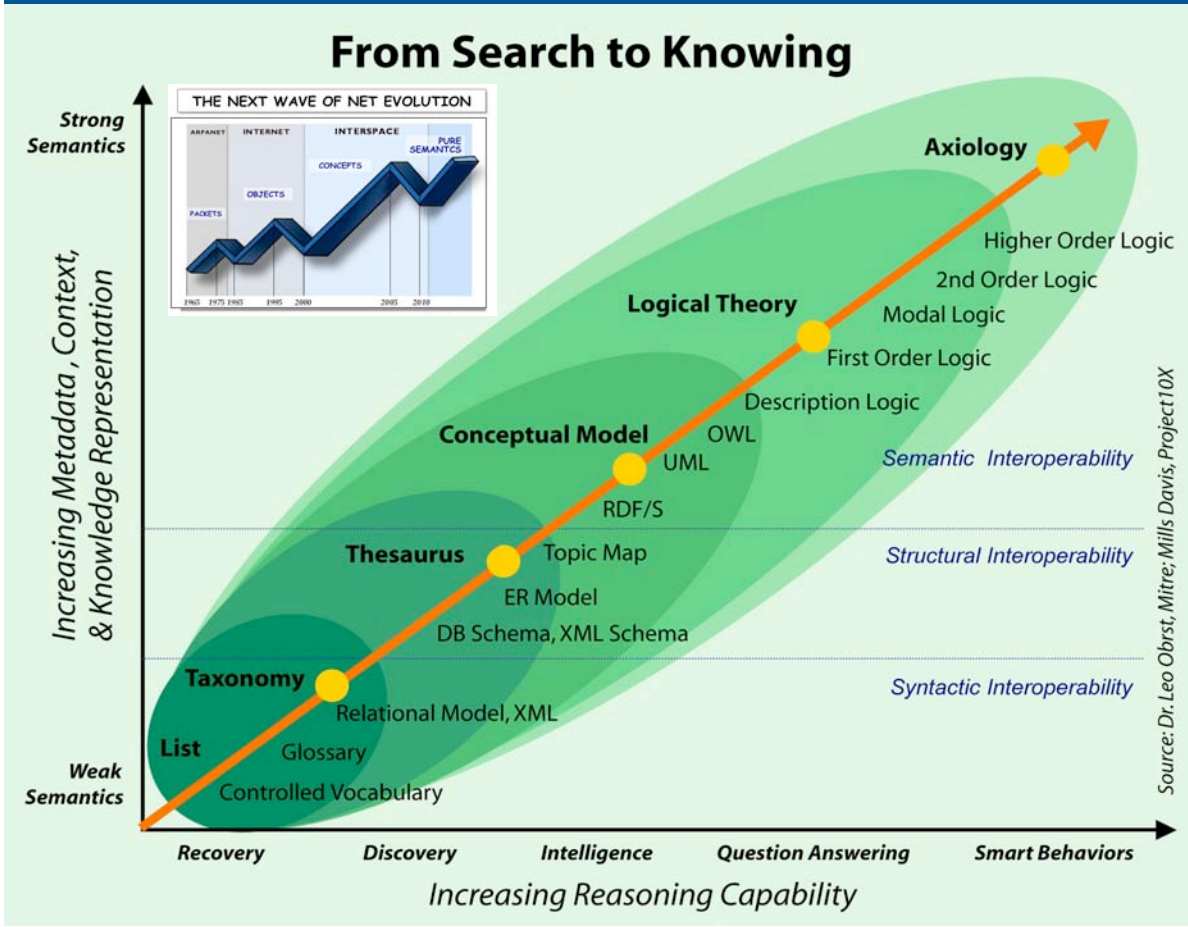
Knowledge Plane:

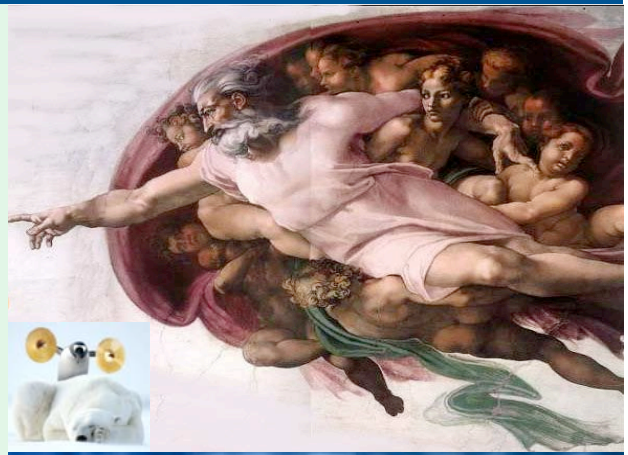
Semantic technologies affect all layers of the IT stack



Semantic Bandwidth:

More metadata, semantic modeling & knowledge representation, more reasoning capability



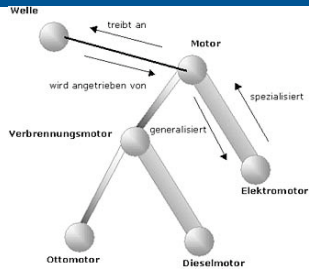


Semantics?
It's what we do every second of the day.

What?

- Convert data into something we can comprehend
- By developing or applying concepts
- Quickly relating them to instances in the world
- Applying and revising our world models
- Sharing our models with others

How?



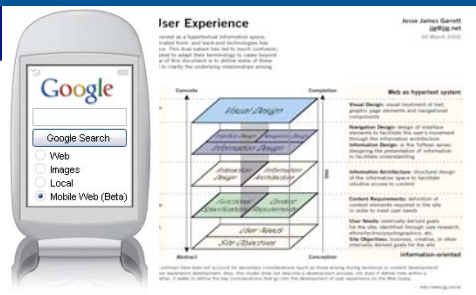
- Identification
  - Concept encoding
- Generalization
  - Organizing concepts by kind
- Aggregation
  - Aggregating complexes into simpler concepts
- Common Properties
  - Relationships (connecting properties)
  - Attributes (flat properties)
- Naming Conventions
  - Terms / Phrases
  - Language

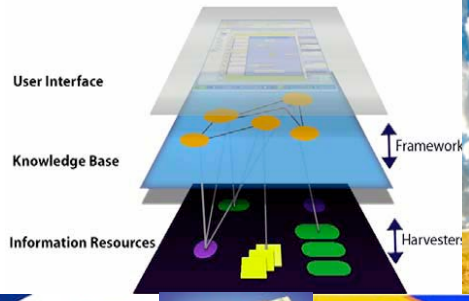
Semantics build on what we already know

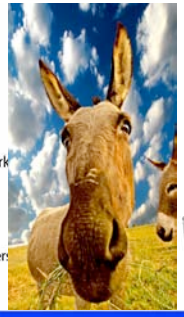
Sounds like Metadata?	Sounds like Taxonomy and Vocabularies?	What more do Semantics provide?
• Similar… • Metadata serves the implementation paradigms of the system • Is the semantics of the data structures in a system, e.g.: – Relational – ERD – Object oriented – UML – XML documents – XSD XMI	• Close… • Taxonomy provides a hierarchy of terms of concepts from a social point of view • Permits only one accepted notion of a term: – Brother? family, or religion, or union, or…	• Context • Multiple points of view • Inferred relationships • Causality • Granularity

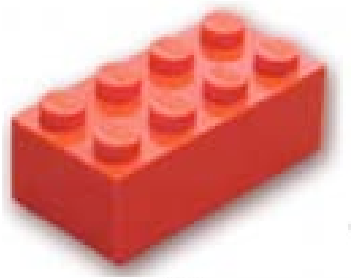
Semantic Web 2.0:

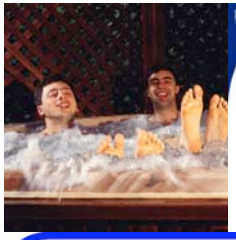
Context, social nets, & relationships are king;
Make my life more manageable & enjoyable





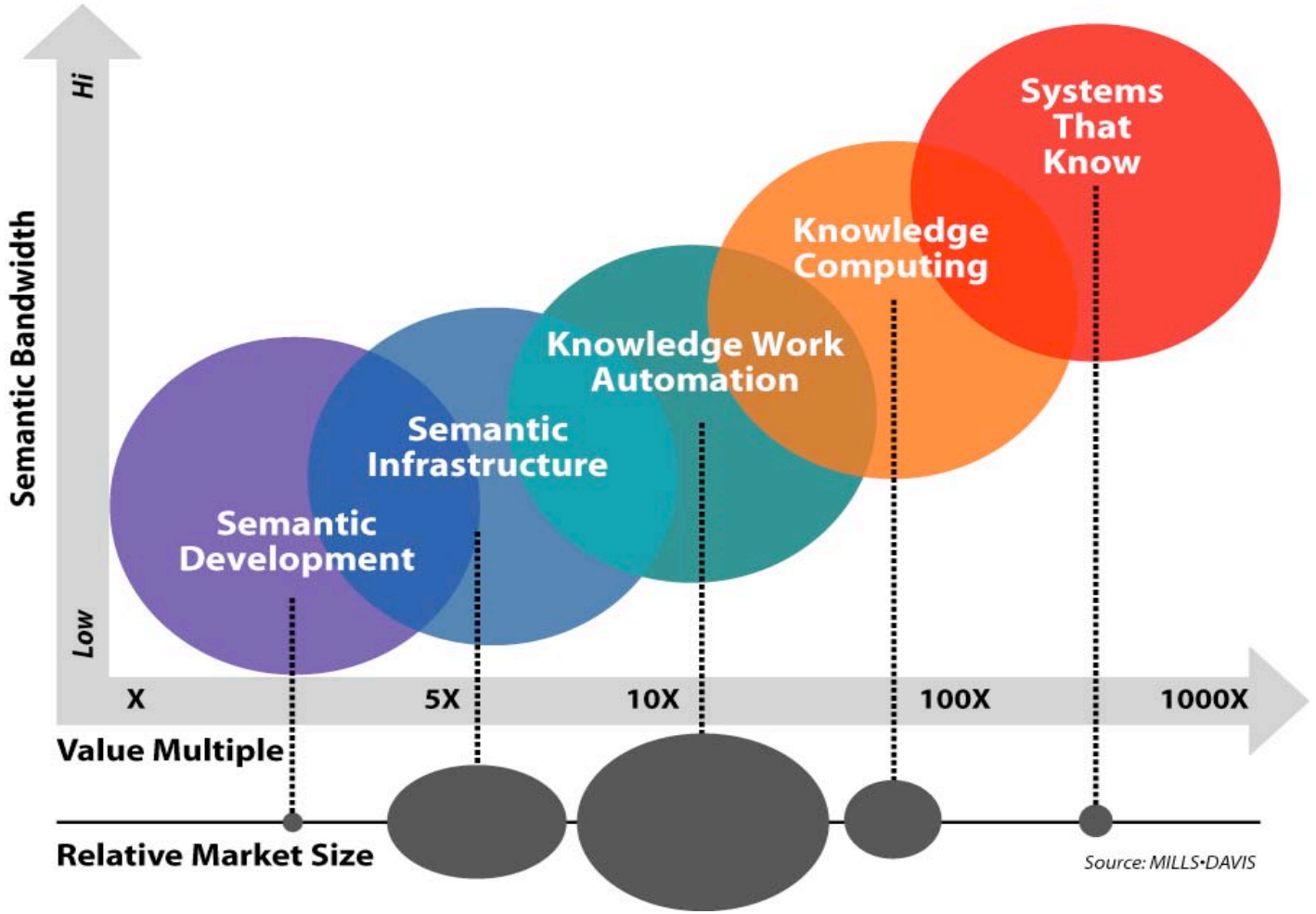






Semantic Bandwidth:

Value gains from two-fold to more than 100 times



SEMANTIC WIKI

The original WIKI* idea

- "A web site where anybody can create/edit a web page"
- Structure
  - is not pre-determined
  - invented & evolved by community
  - neither top down or bottom up
- Quick collaborative writing
- Non-linear hypertext

* **Wiki** is the short form for "wiki wiki web," from the Hawaiian expression "wiki wiki" meaning fast or quick.

WIKI concepts

- Authoring via web browser
  - Also, uploading of arbitrary (multimedia) content
- Simplified wiki syntax
  - Very simple markup for authors
- Collaborative editing
  - Any page can be immediately contributed to, extended, revised, corrected assuming you have the right privileges
- Rollback mechanism
  - All changes are versioned, audited and transparent to the community
- Strong linking
  - "Concepts" in text can immediately become active resources (pages/links)
- Search
  - Typically, a full text search capability

WIKI application areas

962 000+ articles

- Encyclopedia systems
  - Collective knowledge in a certain area as a community effort with broad range of contributors
- Collaborative writing
  - Authors work collaboratively on a writing, which is immediately accessible to readers
- Project knowledge management
  - Project tracking, brainstorming, coordination of ideas, agenda tool to collect topics, project notes repository, knowledge base, staff directory
- Personal knowledge management
  - Sketchpad to collect ideas, addresses, dates, tasks, bookmarks, etc.
- Content management system / knowledge-base
  - Collect content, connect content, simple publishing
- Software development
  - Collaborative documentation, track bugs, e.g. most open source projects coordinate via wikis

Benefits of the wiki idea

Wiki way:

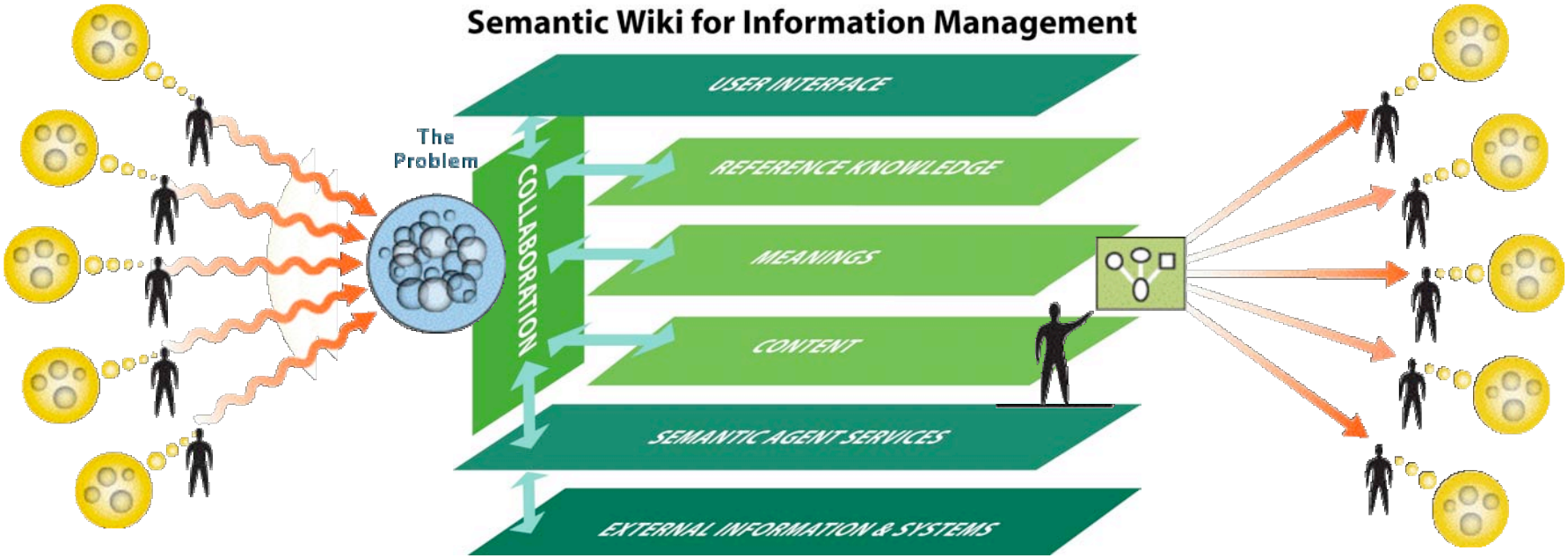
- Distinct concepts or topics are built on the fly
- Discourse forms around or in the context of a topic
- Eliminates serialized document work flow
- Team or community members can immediately see commentary in the context of a topic

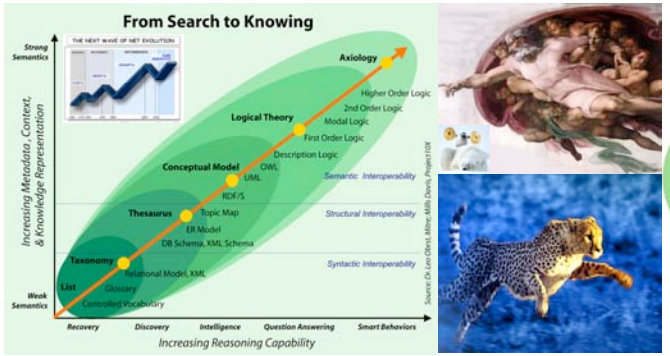
Versus:

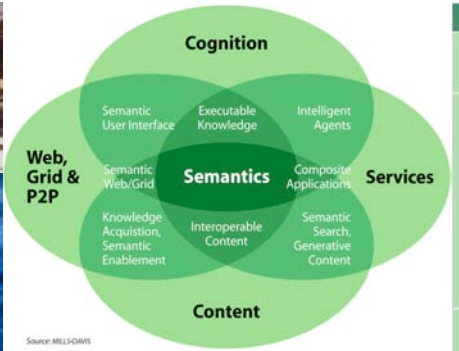
- Each person edits a copy of the document
- A poor soul merges the results
- Expensive file shares
- E-mailing bulky documents
- "Versions" of opaque documents everywhere
- "Organizing" documents in hierarchal file system

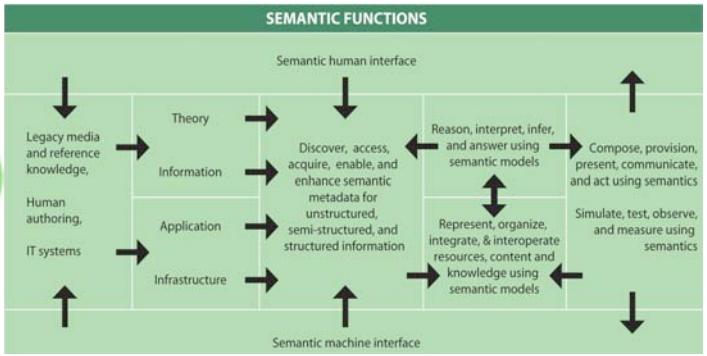
Semantic Wikis:

Collaborative authoring, editing, peer review, information sharing, and knowledge management









Semantic WIKI basic ideas

- Semantic Wikis create a "**knowledge layer**" or overlay network structure that defines concepts, attributes, and relationships of the underlying content of the Wiki. Relationships become explicit as links.
- Semantic Wikis capabilities include:
  - User identity, role authorization, security (including resolution of public vs. private content vs. secure access issues)
  - Semantic modeling of policy, process (workflow) models, schedules & calendars
  - Easy multi-user authoring and editing — read, import, author, edit, annotation, manage, and communication of both content and knowledge layers

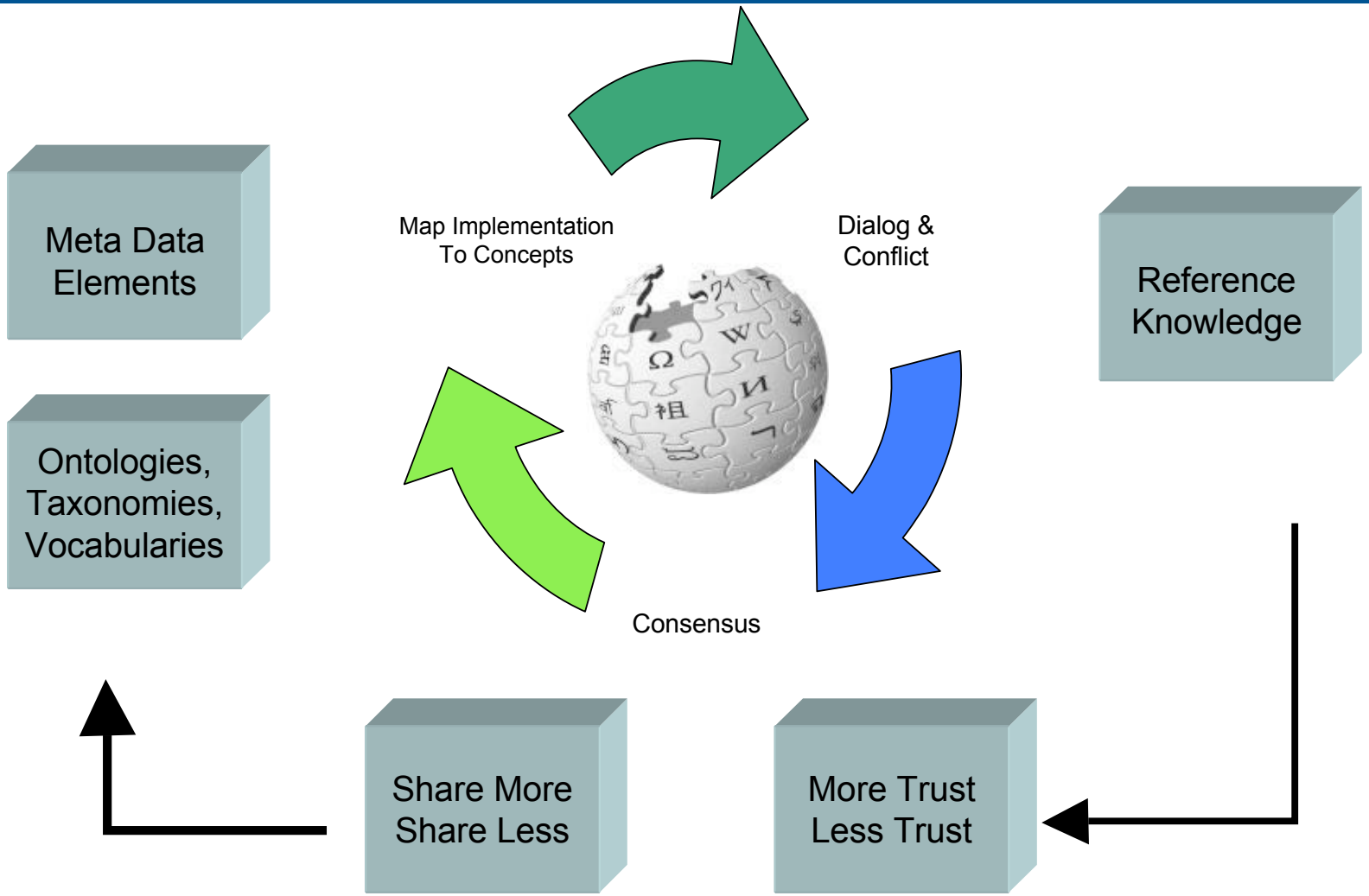
Semantic WIKI basic ideas

- Easy capture, modeling, editing, and validation of semantic metadata, through: (a) automated and semi-automated mining of concepts and relationships in content; (b) semantically enabling structured, semi-structured, and unstructured information; and (c) effortless editing of metadata to create semantic fabric between pages and topics.
- Discovery, capture and transformation of the semantics embedded in as-built data schemas, application program interfaces, comments within source libraries, and system and user documentation into a form that can be queried, harmonized, and managed by the community over the lifecycle of this knowledge.
- Effective semantic search and context-aware navigation spanning internal and external sources that is concept based rather than language-based. Queries that span vocabularies, languages, and search engines.

Semantic WIKI basic ideas

- Visualization of content and knowledge structure — such as contexts, categories, taxonomies, semantic nets.
- Richly structured navigation of content (e.g., pages, topics, services and resources) — ability to present multiple perspectives, multiple levels of abstraction, dependency and contingency relationships, etc.
- Semantic Wiki content linked to dynamic models, simulations, and visualizations. Also linked to external repositories, file systems, including desktops, servers, web-based sources, semantic-enabled feeds (e.g. RSS), etc.
- Community workflows and mechanisms to vet work in process, conduct peer review, test, and come to agreement.

Federated Trust Engines



Two key ideas

1. Problem-oriented architecture
   - Consumer vs. producer centric
   - Event-driven collaboration and composite applications
   - Semantic interoperability, not just syntax and structure
2. Community based management of information and knowledge sharing
   - Techno-social lifecycle governance
   - Fabric of trust (authoritativeness)
   - Automated change management

FEDERAL PILOT SCENARIOS

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.

The need

- Standards agnostic
- Completely pluggable
- Harmonization center for meta data and ontologies
- "Soft" wiki layer that invites user collaboration and discourse
- Configurable governance
- Federated change management

DoD Net-Centric Data Strategy

A Net-Centric DoD NII/CIO

Government Executive Panel
NDIA Net-Centric Operations Conference
March 13-16, 2006

Transforming the Way the DoD Manages Data

An Army Officer recently observed,

"The Global Information Grid (GIG) exists to connect people with information"

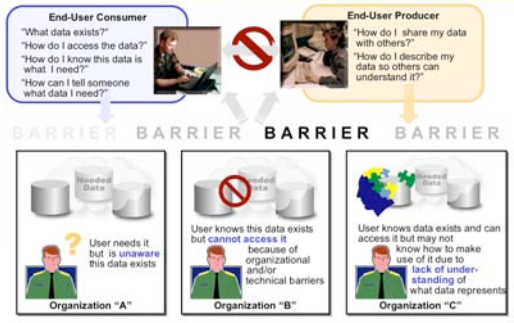
Mike Krieger
Director, Information Management
OASD(NII)/DoD CIO
michael.krieger@osd.mil
March 14, 2006

1

A Net-Centric DoD NII/CIO

Barriers to Identifying, Accessing and Understanding Data

Defining The Data Problem

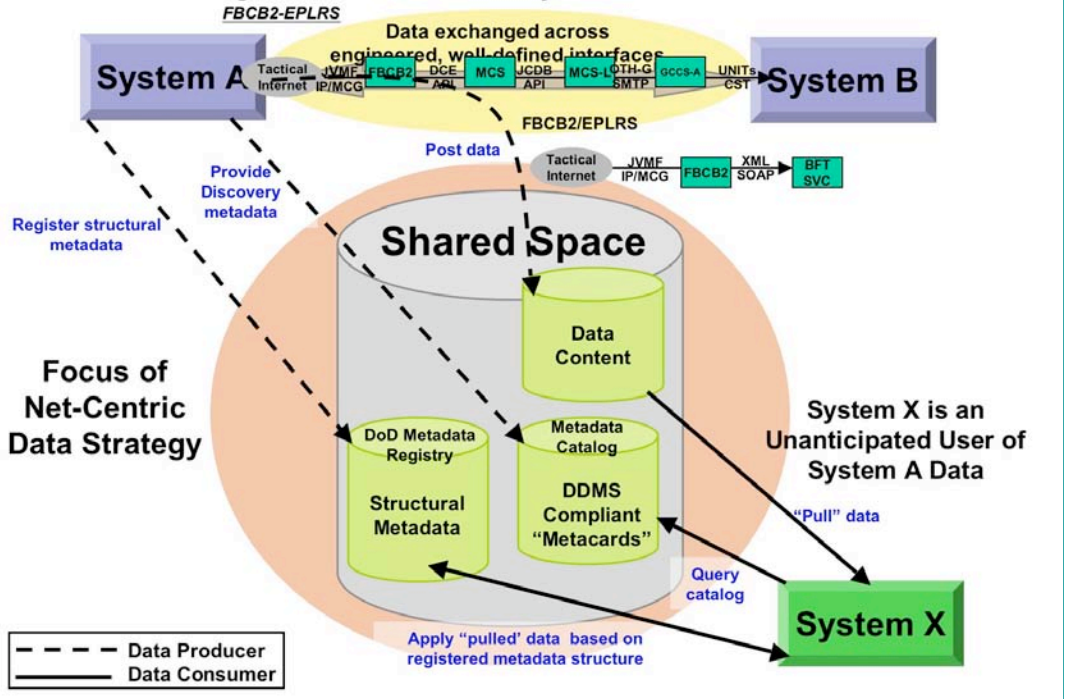


4

A Net-Centric DoD NII/CIO

Net-Centric Data Strategy Enables Unanticipated Users

Focus of Existing Data Administration Policy is Pre-defined Point-to-Point Interfaces



6

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!***

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?

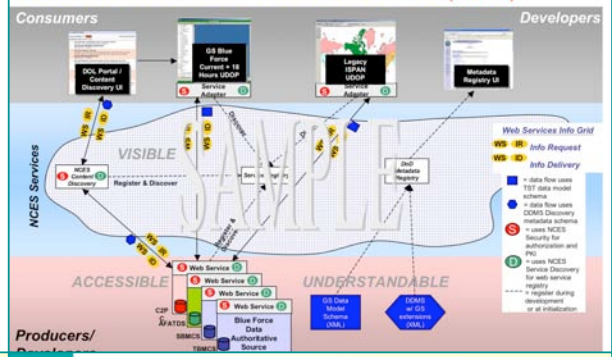
(4) What joint exercise(s) will be used to demonstrate net-centric capabilities?

(5) What organizations are participating?

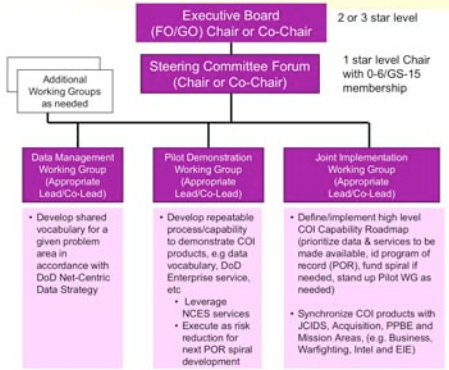
12

COI Pilot Systems Architecture

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



Sample COI Organization Chart

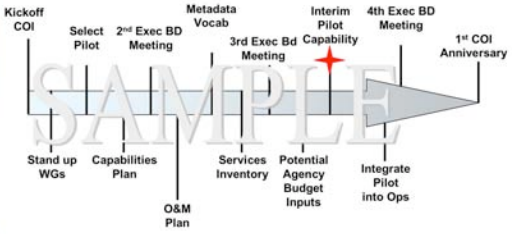


FOR DISCUSSION PURPOSES ONLY

10

COI Pilot POA&M

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



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

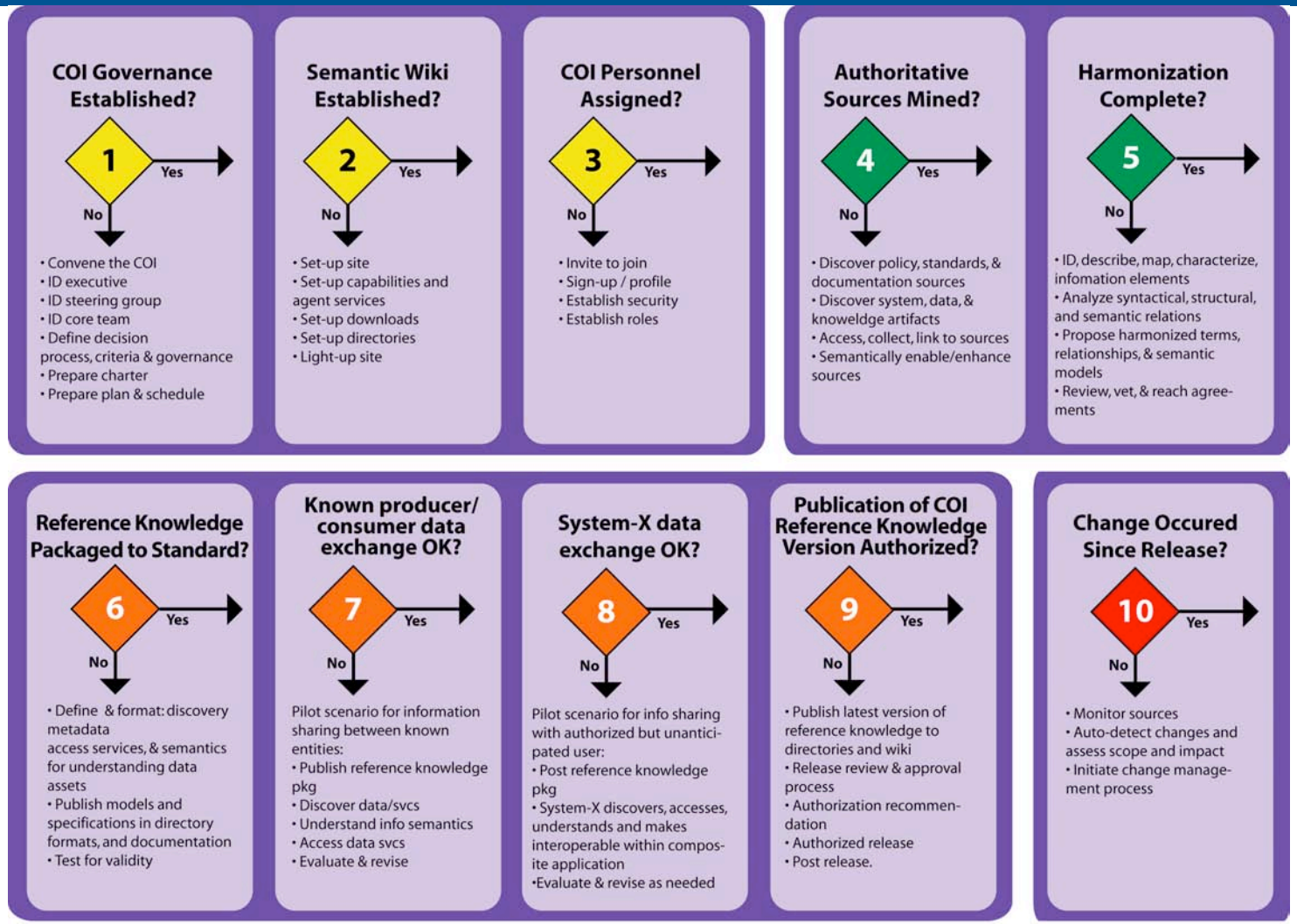
#3 - User assessment of demonstrated net-centric capabilities

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

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

16

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



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