

Net-Centric Enterprise Services  
(NCES)  
2<sup>nd</sup> Working Group  
Mediation and Discovery

March 5, 2004  
Status Report

Chuck Heazel  
cmheazel@opengis.org  
(703) 380-7433

A.J. Maren, Ph.D.  
ajmaren@theeagleforce.com  
(703) 506-4610 x536

# Overview of Presentation

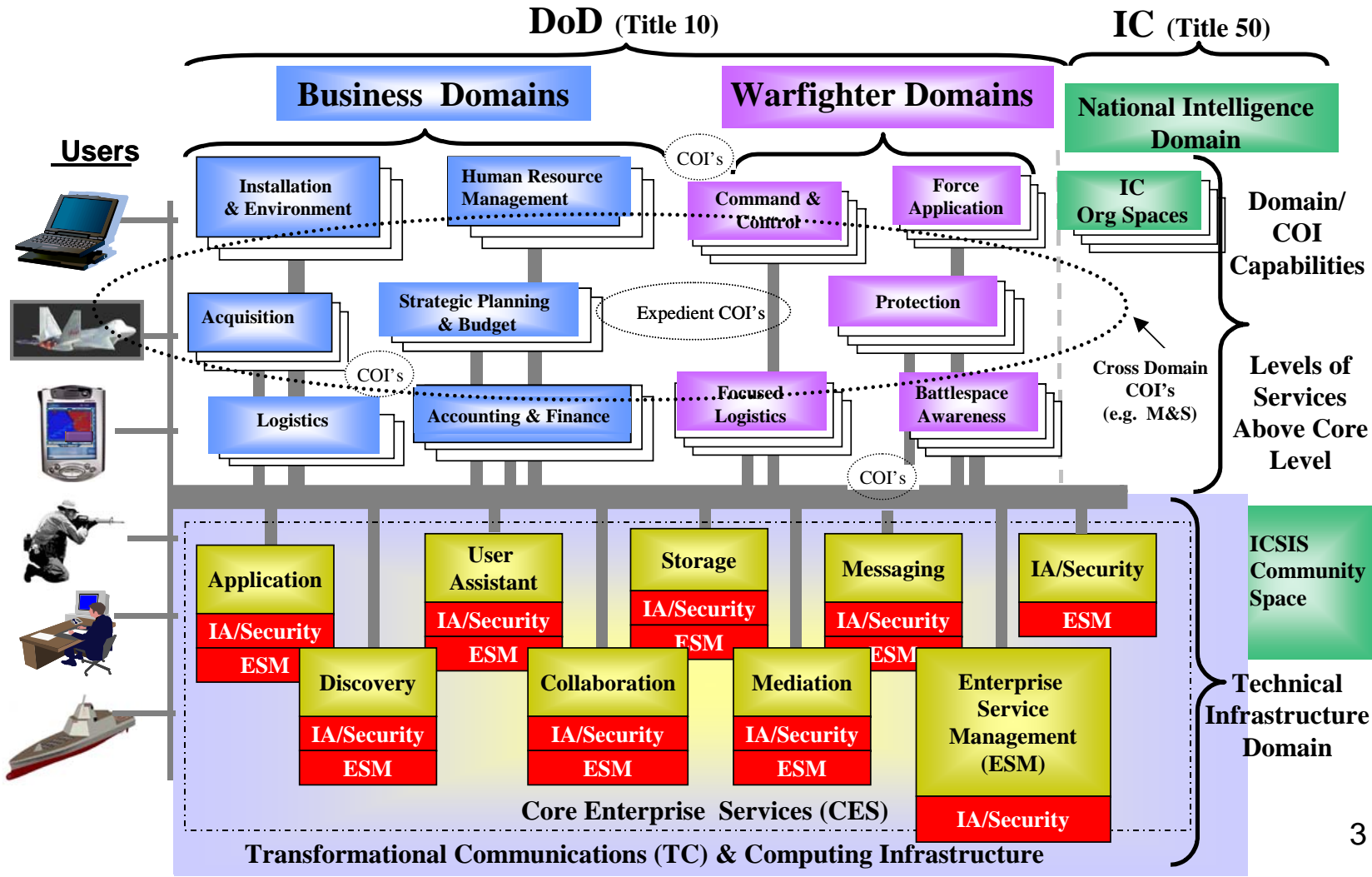
## Mediation and Discovery

- **Introduction to Issues / Role of Mediation and Discovery** (5 min) **C.M Heazel**
- **Mediation** (15 min) **C.M Heazel**
- **Discovery**
  - **Key Concepts**
    - **Focused Discovery** (10 min) **C.M. Heazel**
    - **General Discovery** (20 min) **A.J. Maren**
  - **Recommendations** (10 min) **A.J. Maren**
  - **Examples of General Discovery** (10 min) **A.J. Maren**
- **Q&A** (20 min) **CMH & AJM**

# Introduction to Issues / Role of Mediation and Discovery

## GIG Enterprise Services

Support real-time & near-real-time warrior needs, and business users



# Mediation

Based on the Mediation Whitepaper of the OGC  
Enterprise Architecture SIG

# Mediation Key Concepts

## Purpose

- In a large enterprise of autonomous systems, such as the GIG, the definition of a single set of standards that are suitable for everyone is nearly impossible.
- Each system has its' own unique requirements that the developers must address.
- Individual system environments will be built to the set of standards, data models, and technologies that best address their requirements.
- The GIG will be heterogeneous. Systems will be deployed on different schedules using different vendors and versions of software.
- To participate in the greater enterprise, there must be a way to bridge the incompatibilities between these individual IT environments.

# Mediation Key Concepts

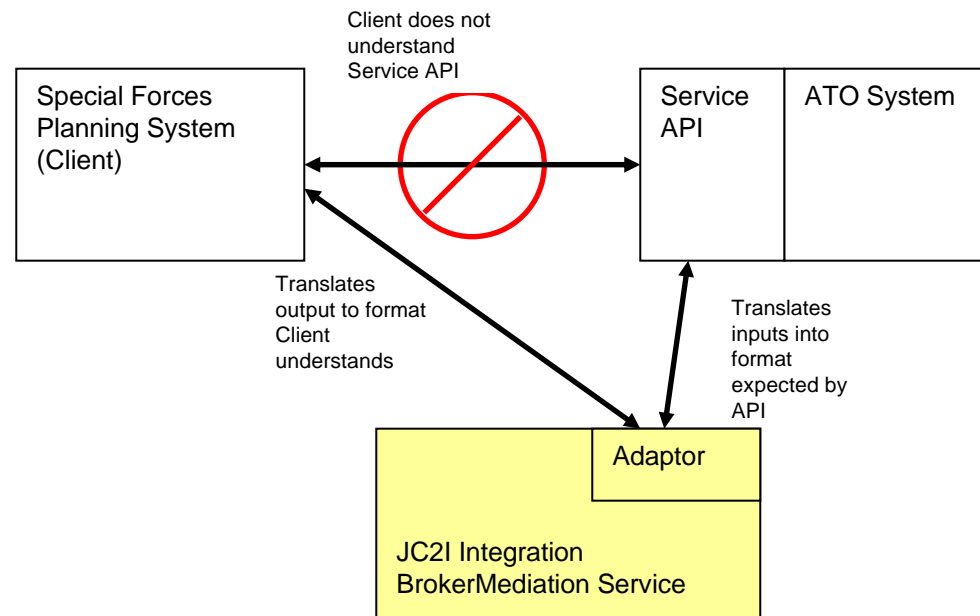
## Axes of Mediation

- Data Mediation - integrating dissimilar information
- Service Mediation - integrating dissimilar services
- Across Providers - mediation involving many sources/actors
- Single Provider - mediation involving a single provider/consumer pair

# Mediation Key Concepts

## Adaptation

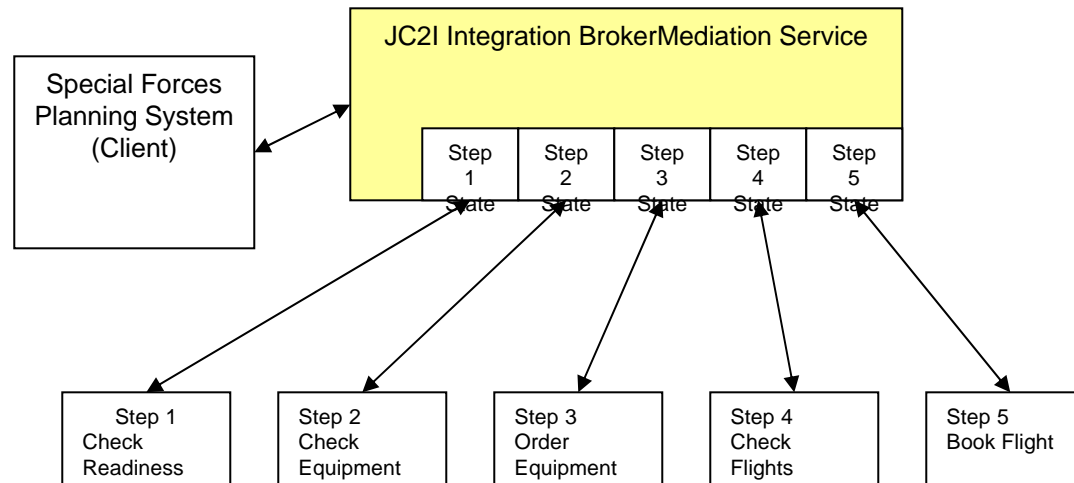
- Description
  - Used when an invoking application can not communicate directly with an outside service. Adaptors provide service mediation when systems need to communicate point to point.
- Technologies
  - Enterprise Application Integration (EAI)



# Mediation Key Concepts

## Orchestration

- Description
  - When a service request triggers a whole chain of events, orchestration services assemble and manage the integrated services (workflow)
- Technologies
  - Business Process Management (BPM)

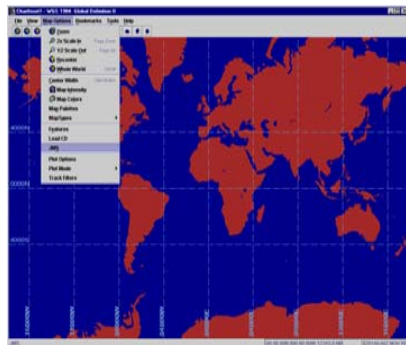




# Mediation Key Concepts

## Transformation

- Description
  - When an application requests information that is not available in the fashion that the requestor desires, transformation services convert the information into the desired format
- Technologies
  - Enterprise Application Integration (EAI)
  - Extract/Transform/Load (ETL)
  - Enterprise Information Integration (EII)



81.71143,  
26.266085,  
81.89853,  
26.409826,  
81.78501,  
26.374152

Transformation  
Service

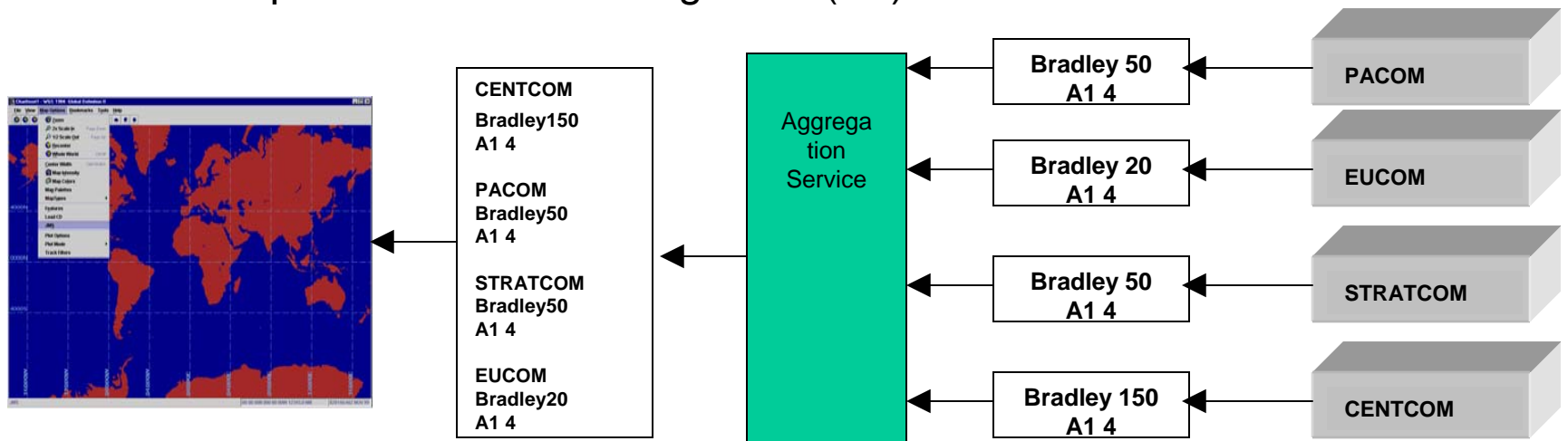
5 Airfield Dr,  
256 Base Ave,  
11055 Andrews  
AF



# Mediation Key Concepts

## Aggregation

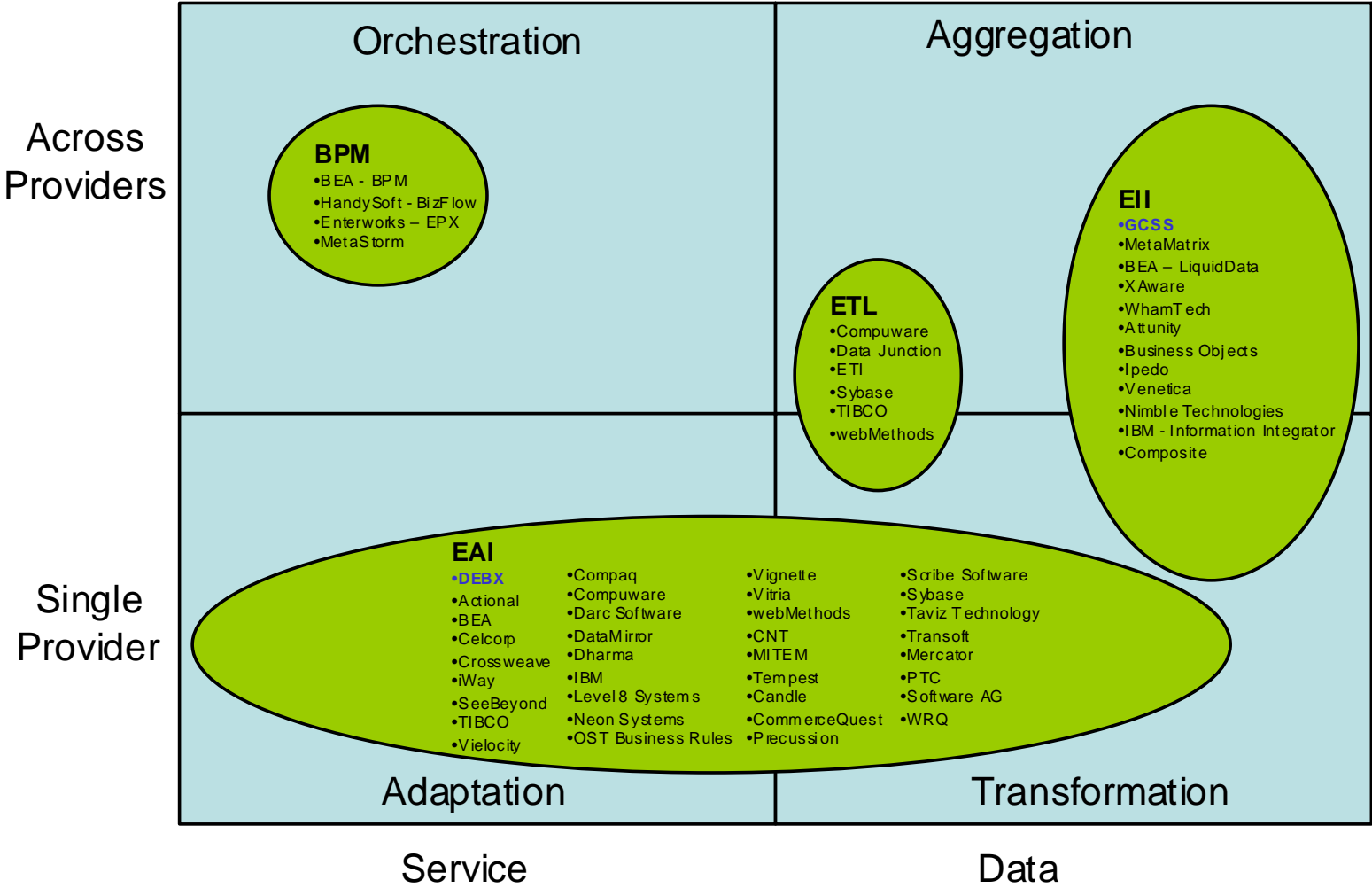
- Description
  - Provides a central point of interaction when requesting information. There are usually multiple information sources points being integrated into the single point of interaction
- Technologies
  - Extract/Transform/Load (ETL)
  - Enterprise Information Integration (EII)



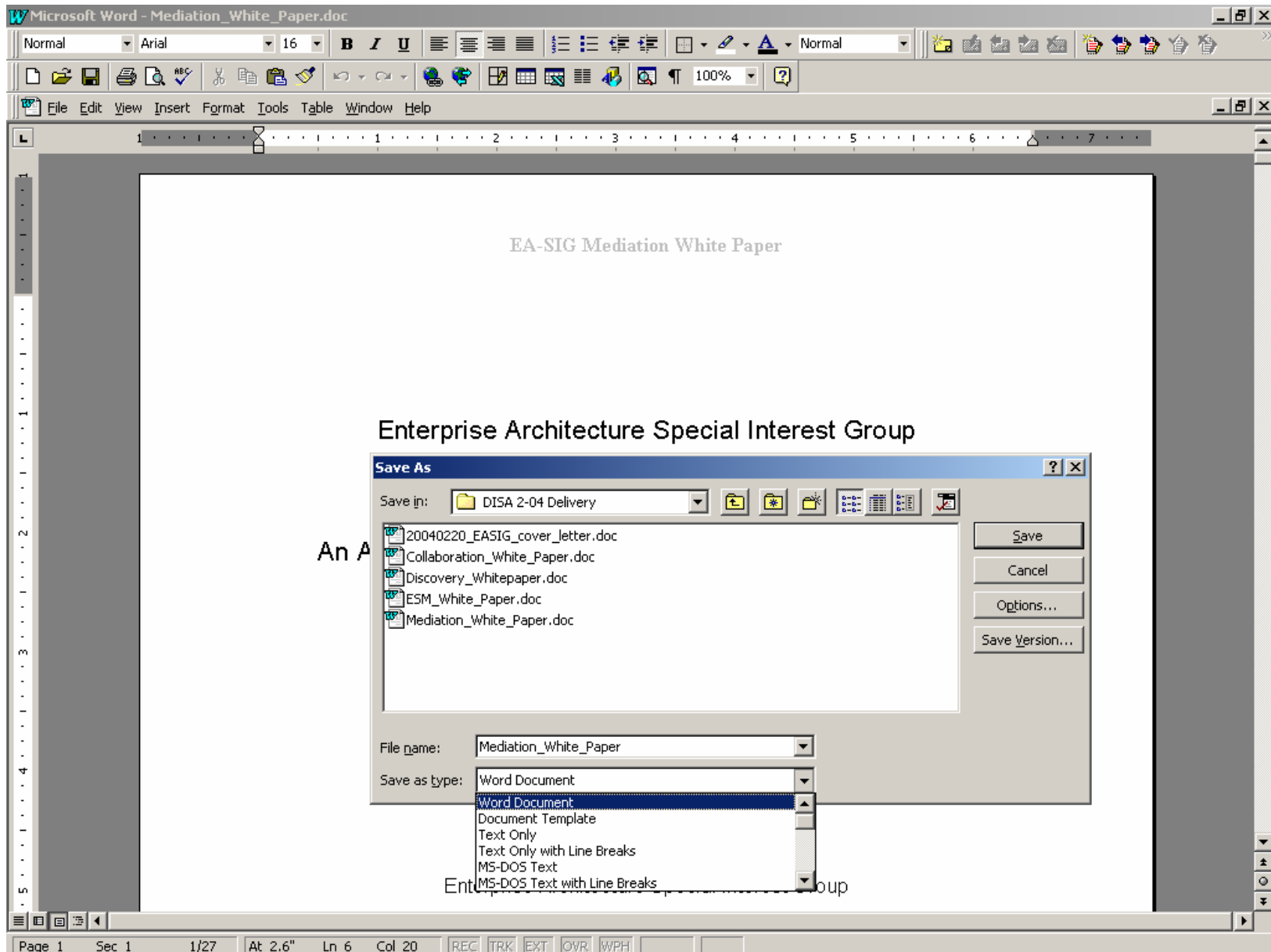
# Industry Concepts

- Enterprise Application Integration (EAI)
  - Adaptation and Transformation
  - Supply chain enablement
- Extract/Transform/Load (ETL)
  - Transformation and Aggregation
  - Data Warehouses, business application focus
- Enterprise Information Integration (EII)
  - Transformation and Aggregation
  - Federated information integration, information focus
- Business Process Management (BPM)
  - Orchestration
  - Workflow management
- Embedded
  - Ubiquitous

# Mediation Key Concepts



# Embedded Mediation



# Recommendations

## (Transformation)

- Develop an understanding of the current operational information communities through the DoD XML Registry
- Identify opportunities for normalization and transformation
  - Use DoD XML Registry Namespace Managers Forum to identify information constructs that are common across most DoD information communities.
  - Use Namespace Managers to identify information constructs that cross information communities in the course of operations. This is the subset of information that is a candidate for transformation.
  - Work with the Namespace Managers to identify which members of that subset cannot be used by the receiving organization in its' native format. Identify the transformations needed to make the conversion.
- Continue to use the forum to manage the definition of DoD information transformations and the nature of those transformations.

# Recommendations

## (Adaptation)

- Enable the Core Enterprise Services across multiple distributed computing platforms
  - Establish adapters allowing J2EE based clients to invoke the CES Discovery Services using J2EE discovery interfaces.
  - Establish adapters allowing J2EE based clients to invoke other CES services using J2EE interfaces on an as needed basis.
  - Establish adapters integrating legacy information security infrastructures (such as Active Directory) with the CES security services.
  - Establish a taxonomy of the services types needed to support DoD operations and identify or define standard interfaces for those service types.

# Recommendations

## (Adaptation) Continued

- Develop a process for managing the portfolio of services and interfaces. Define the services and interfaces using UML models so that semantically equivalent implementations can be generated as new distributed computing technology becomes available.
- Develop adapters for key legacy systems to integrate them into the greater GES community.



# Recommendations

## (Aggregation)

- Aggregation does not appear to be a near term need
- Some aggregation capabilities exist in the C4I community already
- Additional aggregation capabilities can be expected to appear as an integral part of emerging services
- The NCES program office should focus their aggregation efforts on identifying the requirements as they develop and shepherding development efforts toward common implementations

# Recommendations

## (Orchestration)

- Orchestration is of little use without a rich portfolio of deployed services to orchestrate
- Near term efforts should focus on orchestration pilots and prototypes to discover the capabilities of this technology and to identify possible opportunities to use it.

# Recommendations

## (General)

- There is a lot of work in the research and standards communities in the area of mediation and semantic interoperability
- It is doubtful that any of this is ready for robust operational use
- The NCES program should monitor and, where possible, participate in these activities
- As technology matures, be prepared to insert it into the program

# Cross Service Issues

- Discovery
  - Mediation is a critical element in all but the most basic discovery
- ESM
  - It is not likely that there will be a single set of ESM standards. Mediation will be required to integrate the ESM communities
- Collaboration
  - As collaboration becomes more information centric mediation will play a critical role
- Architecture Patterns
  - Construction of a federation of systems of the size of GES will require mediation services to integrate the inevitable incompatibilities

# Discovery

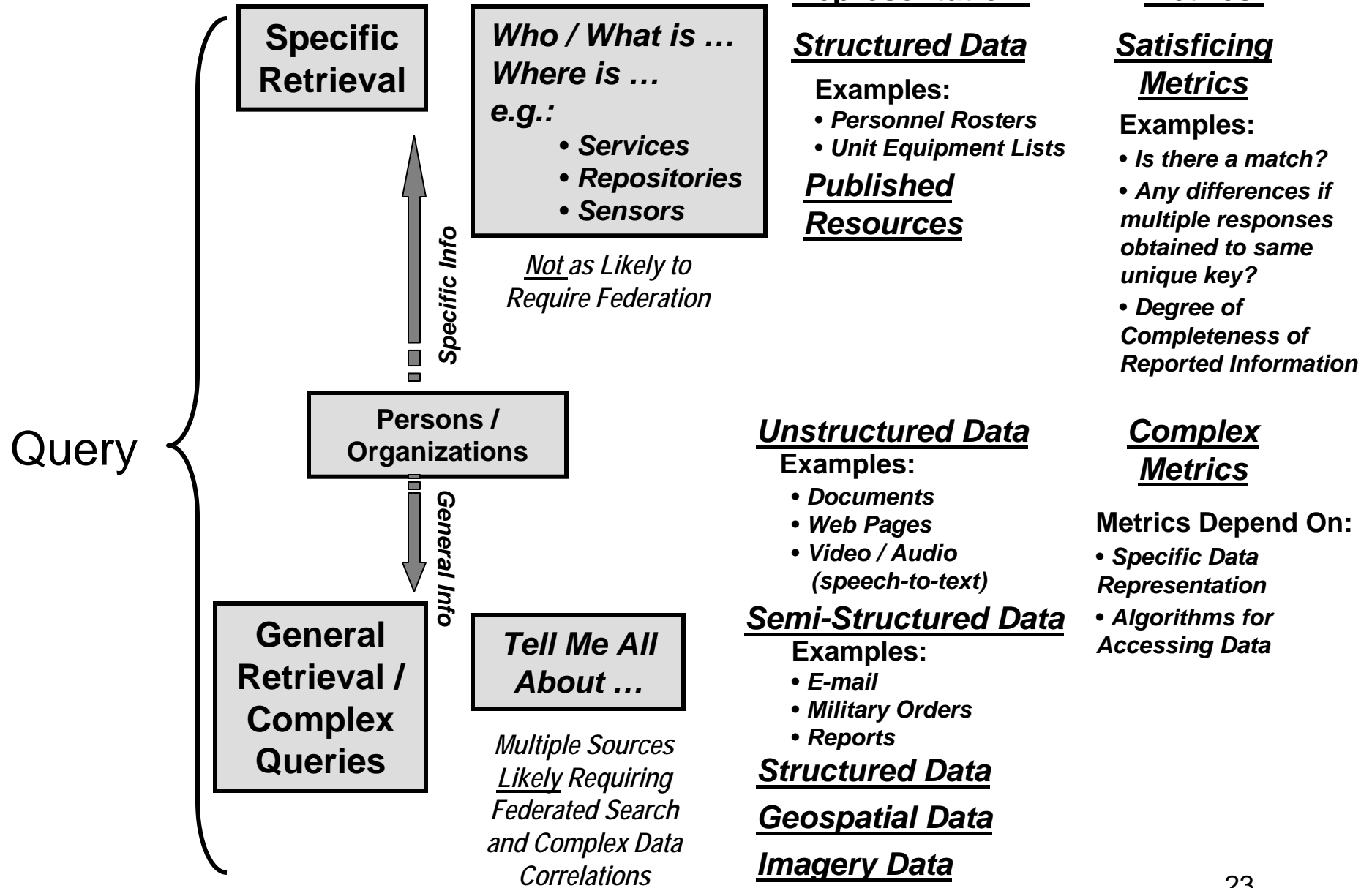
Based on the Discovery Whitepaper of the OGC  
Enterprise Architecture SIG

# Discovery Key Concepts

## Categories of Query

- **Specific**
  - Query for entities with known attributes
  - Typically targeted toward structured data
  - Use for discovering ***specifics*** about people, organizations, places, services, repositories, and registries.
- **General**
  - Query for broad information about a given topic
  - Targets structured, unstructured, and semi-structured data
  - Involves some degree of knowledge management processing

# The Query Continuum



# Discovery Key Concepts

## Taxonomy of Discovery Capabilities

- ***Focused (Specific) Retrieval:*** Metadata Search
- ***General Retrieval / Complex Query:*** Multiple Tools Plus “Control” Mechanism for Dynamic Orchestration

***Tools Operate at Different Levels of “Knowledge Discovery / Representation”***

***Five “Levels” for General Retrieval of Knowledge:***

- Concept Extraction
- Concept Correlation
- Syntactic Discovery
- Context-based Discovery
- Semantic Discovery

***“Control” Methods allow selection of federated tools and services, dynamic orchestration***

***Two Basic Control Levels for General Retrieval:***

- Feedback Control with Utility
- Reasoning-based Metacontrol



# Focused (Specific) Discovery: Key Concepts

## ***Metadata Search***

- *Function* - Match query terms against a defined metadata model.
- *Methodology* - Relational database key word search
- *Technology:*
  - LDAP
  - UDDI
  - EbXML (EbRIM)
  - Z39.50

# General (Broad) Discovery Key Concepts

## Level 1: Concept Extraction

- Function - Identify and extract concepts; apply concept-based descriptive metatags to linguistic corpora elements and their segments, as well as appropriately indexed images.
- Methodology - Statistically-based methods, including Bayesian Logic, enhanced with Shannon's Information Theory (and alternatively) Semantic Nets  
Technology:
  - Multiple COTS tools implementing Bayesian / Shannon algorithm
  - Multiple COTS tools implementing Semantic Nets

# General (Broad) Discovery Key Concepts

## Level 2: Concept Correlation

- Function - Identify those concepts that are statistically close within corpora elements. Serves to identify concepts that are associated with each other.
  - “Related concepts” can generate additional focused searches
  - Works best with small number of inputs ( $O(10^2)$  –  $O(10^4)$ )
  - Previous concept-extraction stage needed to “scale down” starting corpora to provide appropriate inputs to this processing
- Methodology - Co-occurrence matrices ( $N^2$  process); Latent Semantic Indexing (LSI)
- Technology:
  - Multiple COTS tools implementing Co-occurrence matrices and/or (LSI)
  - Variations on type of input to co-occurrence matrix (nouns and noun-phrases, etc.)

# General (Broad) Discovery Key Concepts

## Level 3: Syntactic Discovery

- Function - Identify “relationships” (verbs) linking “concepts” (nouns) => yields an “intelligence primitive”
- Methodology - Syntactic analysis (complexity  $>N^2$ ; computationally expensive)
- Technology:
  - Multiple COTS tools performing various forms of:
    - Verb extraction, leading to verb bundling and overall “relationship” identification
    - Syntactic labeling of *all* words in small, selected corpora elements
    - Extraction of “people, places, and things” from selected elements
    - Enables further queries based on **concept plus verb (or relationship)** identification as opposed to simpler “noun-” based queries (Level 1), or even “noun” plus “noun” queries (concept-plus-concept queries; Level 2)

# General (Broad) Discovery Key Concepts

## Level 4: Context-Based Discovery

- Function :
  - 1) Identify “context” associated with any “intelligence primitive” (concept-relationship-concept)
  - 2) Enable “handover” of primitive to structured data processing and analytics
  - 3) Enable “handover” of an event to geospatial / temporal representation and reasoning
- Methodology - Multiple methods, many computationally expensive
- Technology:
  - Context extraction: TBD
  - Handover for structured data analytics: TBD
  - Handover for geospatial / temporal representation and analytics: Metacarta, others TBD

# General (Broad) Discovery Key Concepts

## **Level 5: Semantic Discovery**

- Function - Ontologies and their taxonomies, provide inputs to feedback loops governing classification / concept categorization.
- Methodology - Very computationally expensive; also typically long-term investment of representing organizational or knowledge infrastructure.
  - Ontologies and taxonomies must be specified first
  - Topic Maps can be used to find relations between similar taxonomic sub-graphs
- Technology:
  - TBD

# General (Broad) Discovery Key Concepts

## **Control Level: Feedback Control with Utility**

- Function - Integrate multiple categories of discovery. Provide controlled feedback between the segments to enhance the capabilities of each segment.
- Methodology - Feedback loops input values to control system, modulated by utility functions
  - Treat each available Level 1- 5 COTS tool as a “service”
  - Provide control parameters as inputs into each selected tool
- Technology:
  - Orchestration

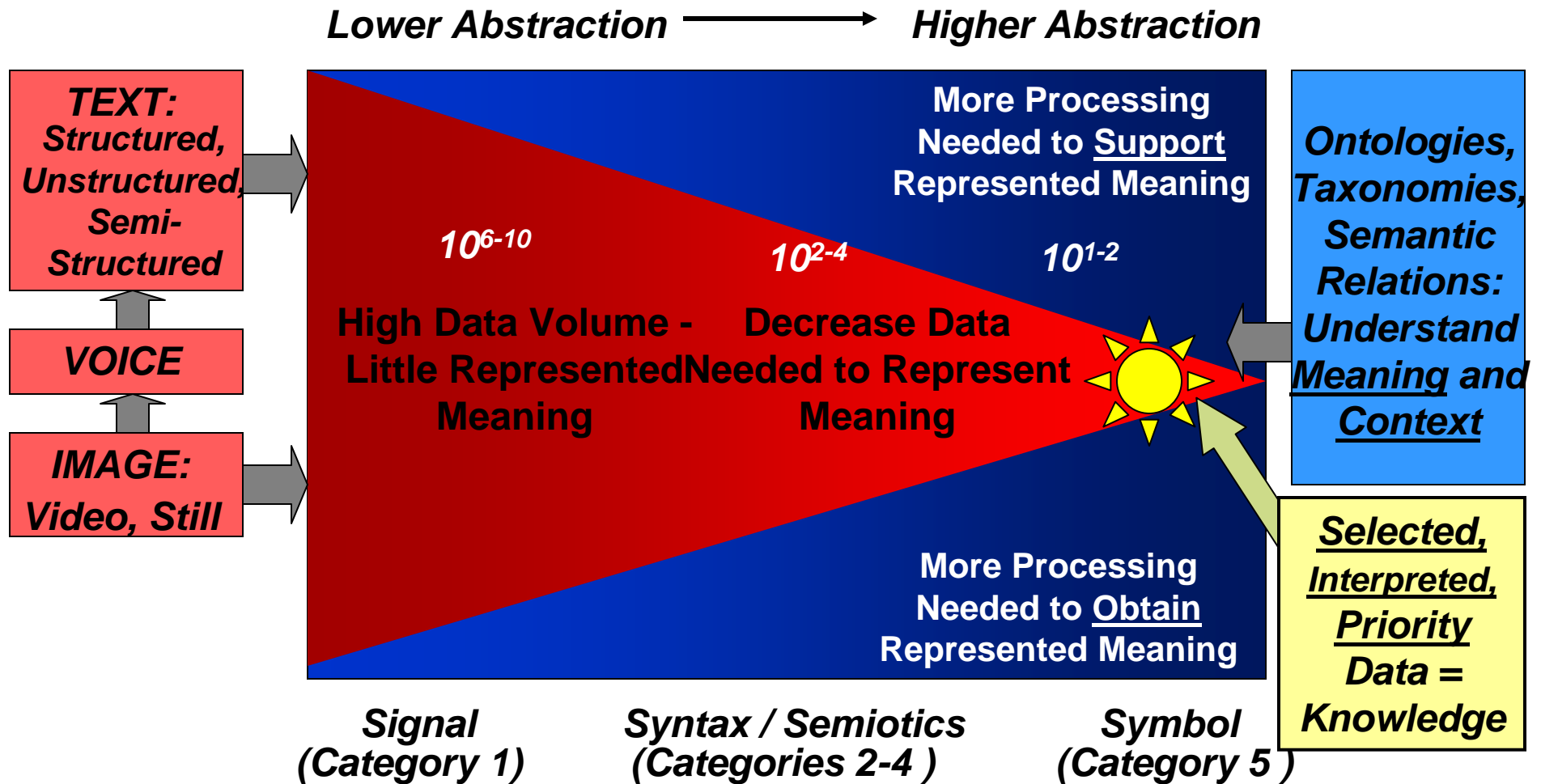
# General (Broad) Discovery Key Concepts

## Metacontrol Level: Reasoning-based Metacontrol

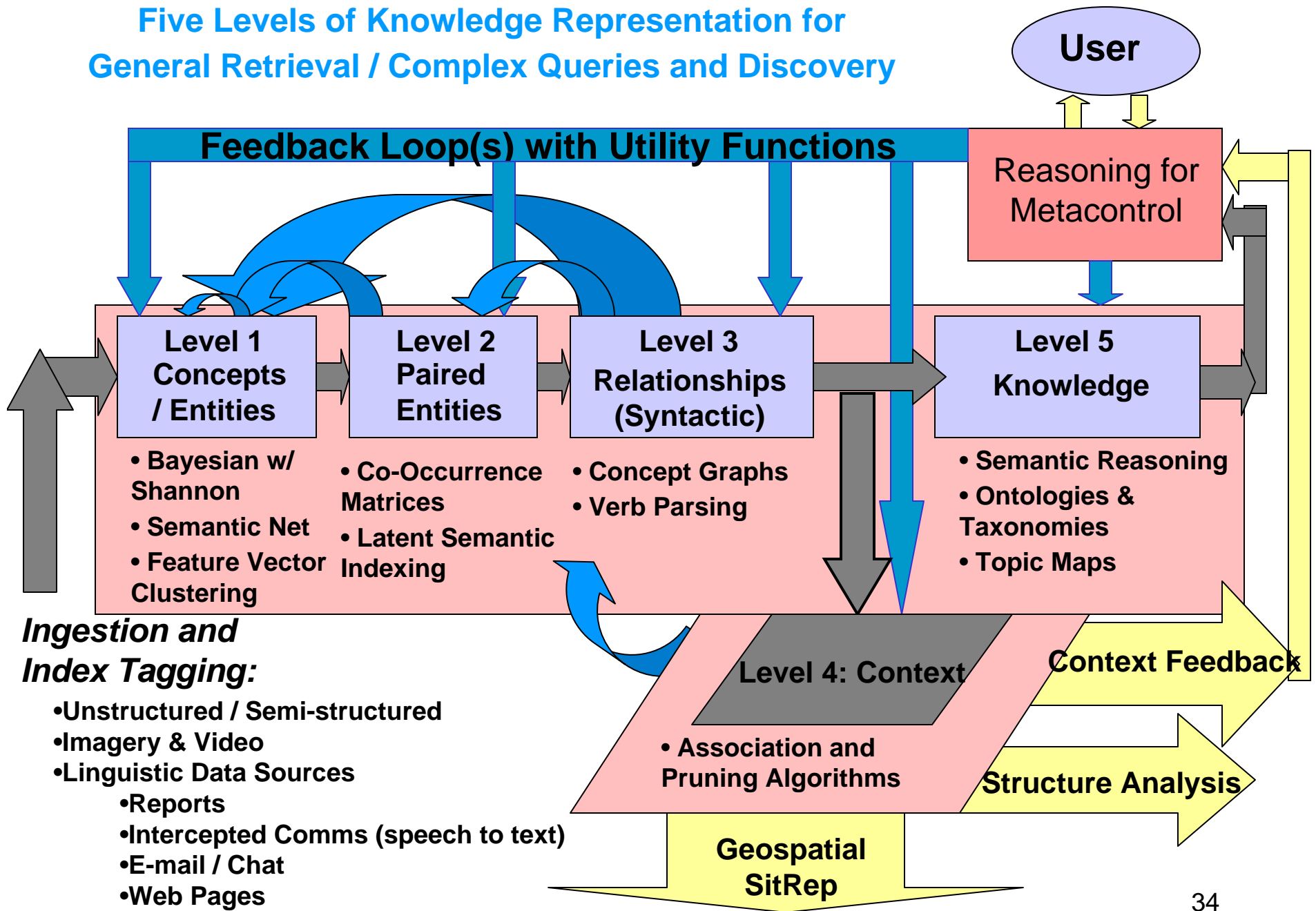
- Function:
  - Define strategy for dynamically selecting available **federated services** (Level 1 – 5 COTS Tools),
  - Define strategy for transitioning “knowledge” from one level to another,
  - Define strategy for feedback and “spinning off” related queries,
  - Define strategy for identifying when alert thresholds are reached,
  - Define orchestration of various services (multiple services simultaneously, or different services at any level invoked depending on feedback, etc.),
  - Define strategy for dynamically selecting various **federated repositories**,
  - **Define decision points** for generating alternate queries and stopping query processes.
- Methodology - Business rules, schemas, rule-based reasoning including reasoning under uncertainty and reasoning with constraints, adaptive pattern recognition
- Technology:
  - TBD



# The Knowledge Discovery Challenge



# Five Levels of Knowledge Representation for General Retrieval / Complex Queries and Discovery



# Recommendations

## Focused Discovery

Discovery of ...	Applicable Standards
Individuals ( <i>specific and profile</i> )	LDAP, UDDI, EbXML, ICML
Individuals ( <i>general</i> )	ICML, other (semantic) web-based standards – e.g., DAML, OWL
Organizations ( <i>specific and profile</i> )	LDAP, UDDI, EbXML, ICML
Organizations ( <i>general</i> )	ICML, other (semantic) web-based standards – e.g., DAML, OWL

Discovery of ...	Applicable Standards
Services ( <i>build-time</i> )	UDDI, EbXML, Z39.50
Services ( <i>run-time</i> )	EbXML, OGC Catalog, Z39.50
Security	LDAP, ICML, SAML
Repositories	LDAP, UDDI, EbXML, ICML

Discovery of ...	Applicable Standards
Structured Data	SQL, OGC Catalog, Z39.50

# Recommendations

## General Discovery / High-Level Discovery

Discovery of ...	Applicable Standards
Semi-Structured Data	EbXML, web crawlers, OGC Catalog, ICML, Z39.50, OIL, DAML
Unstructured Data	EbXML, web crawlers, OGC Catalog, ICML, Z39.50, OIL, DAML
Real-Time Data	EbXML, ICML, OIL, DAML

Discovery of ...	Applicable Standards
Schemas	XML
Ontologies / Taxonomies	OWL
Symbols	MIL STD 2525C, NTDS, NATO

# Recommendations

## (Research)

- ***Federation capabilities*** - how do we query multiple discovery services at the same time?
- ***Registry information model*** - a common information model is needed for focused discovery
- ***Registry population*** - how do we populate and maintain registry metadata?
- ***Storage services*** - should storage services provide a discovery capability as well?
- ***Tools*** - perform a market survey for both focused and general discovery products
- ***Orchestration*** - needed to build the more advanced general discovery capabilities
- ***Architectures*** - where do the different types of discovery fit in the overall architecture?

# Cross Service Issues

- Strong connection between DISCOVERY and MEDIATION (see final slide), also some with COLLABORATION. Lesser connections with ENTERPRISE SERVICE MANAGEMENT (ESM) and SECURITY.

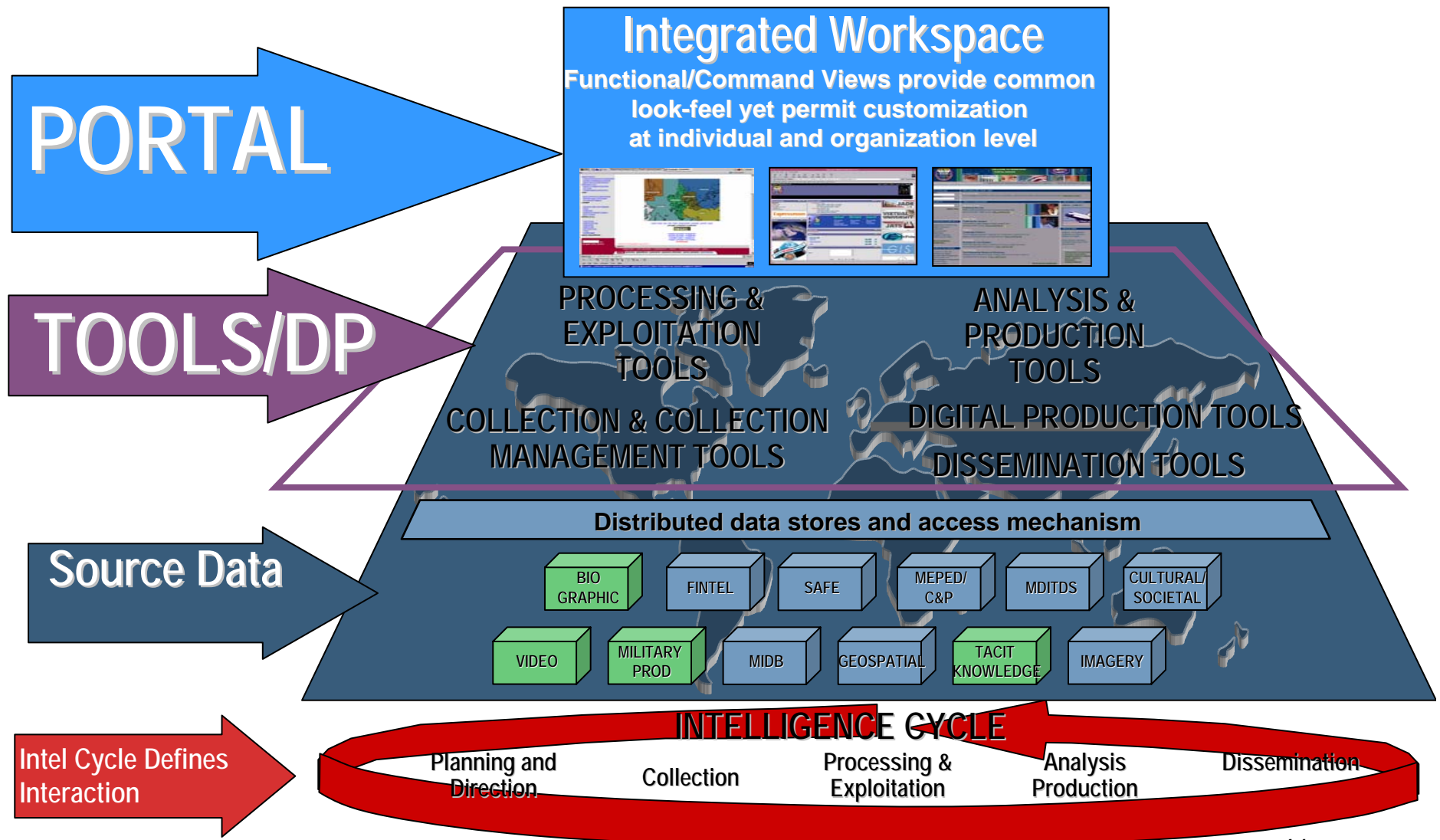
# Examples of General Discovery

# General (Broad) Discovery: Architecture Approach Evolution

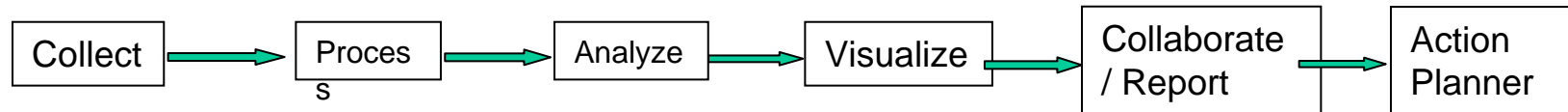
- Joint Intelligence Virtual Architecture (JIVA)
  - Collection of “best of breed” COTS products assembled as “JIVA Knowledge Discovery Toolkit”
- DIA
  - Linear architecture of services
  - Populated by “best of breed” COTS
- GCSS-AF
  - Linear with non-linear feedback control scales queries into large corpora
  - Clear definition of “processing levels” according to knowledge representation and associated algorithms
  - Level 1 “Concept Search” capability implemented through Bayesian classification modulated by Shannon Information Theory; applicable to large corpora
  - DISA-certified, fielded to 1.2 Million users by April 2004
  - Level 5 “Taxonomy”-based work now funded



# JIVA Vision



# DIA Architecture Functionality



- Non-Attributable Direct Internet Connection
- Preserved query terms & hits
- Direct Classified Connect to JWICS, SIPRNT, NIPRNET, etc..
- Automated profiling on JWICS, SIPRNT, NIPRNET
- Direct Connection to NSA for GAMMA feed
- JITF-CT SAFE Repository

- Index data across all classification levels
- Entity extraction and categorization
- PCTT Specific subject matter taxonomies for entity relationship identification
- Data Translation
- Natural Language Recognition
- Language Conversion

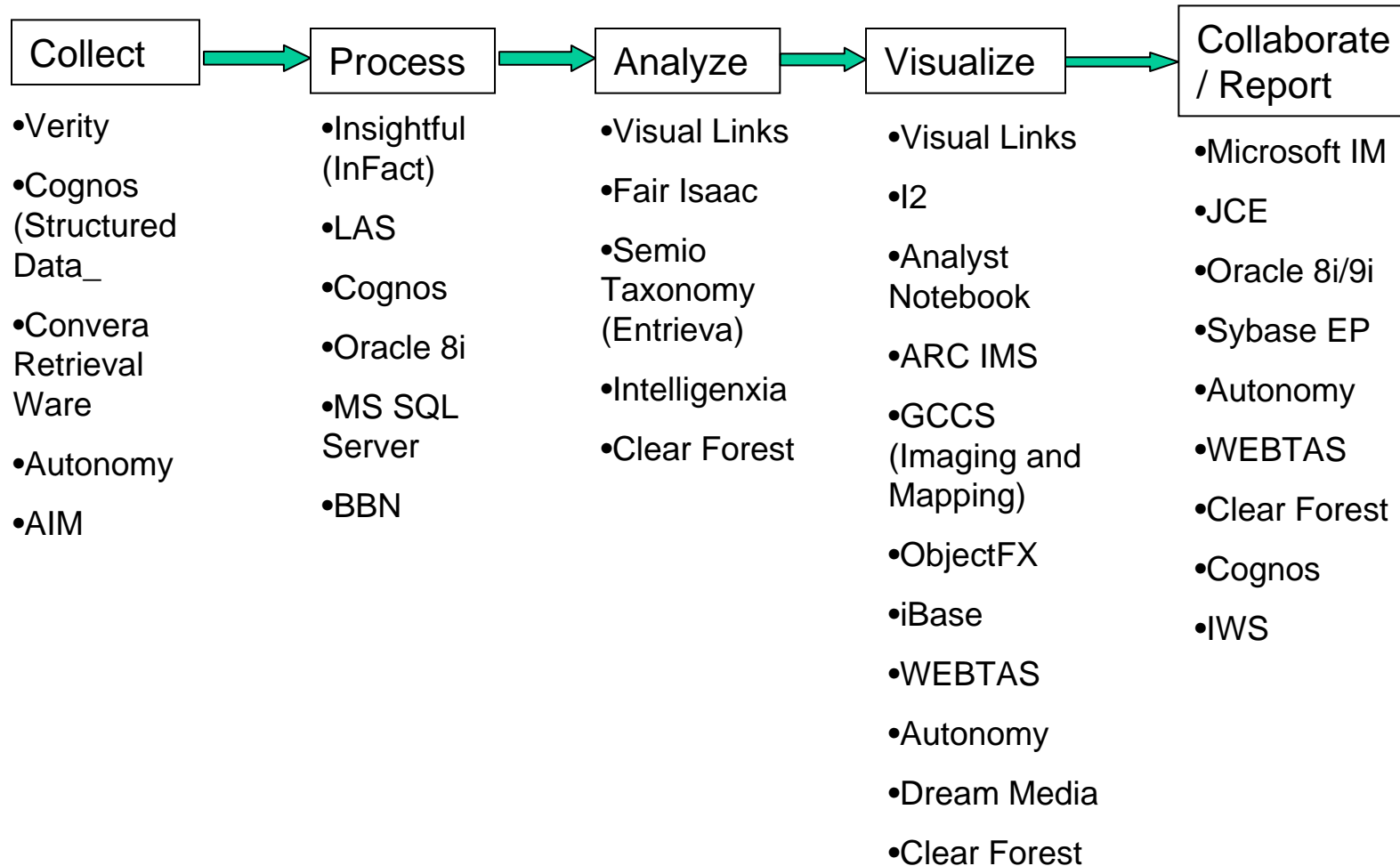
- Rapid Access to Terabytes of indexed data
- Search and Mining
- Link Analysis
- Temporal (Timeline) Analysis
- Entity Relationship Analysis
- Geospatial Analysis

- Web Sites and their relationships to other web sites
- Concept strength in large data collections
- Entity and link relationships (over time)
- Common Intelligence Picture
- Targeting

- Instant Messaging
- White Boarding
- Chat
- VTC
- Phone
- E-Mail
- Warning
- Finished Products
- Semi-Processed Intel
- Dissemination to JWICS, SIPRNT, NIPRNET

- Target ID
- Threat ID
- Sensor asset planner
- ATO planner
- OPS Re-planner

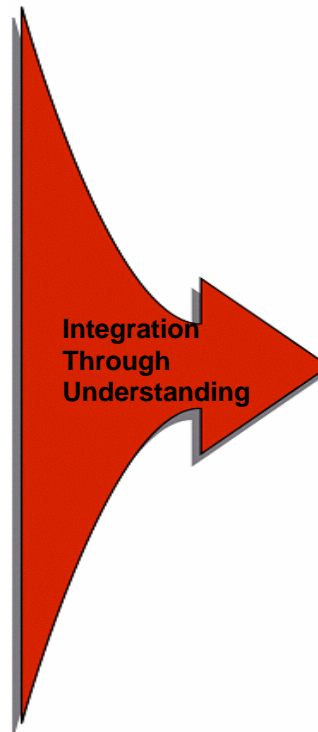
# DIA Architecture Software Components





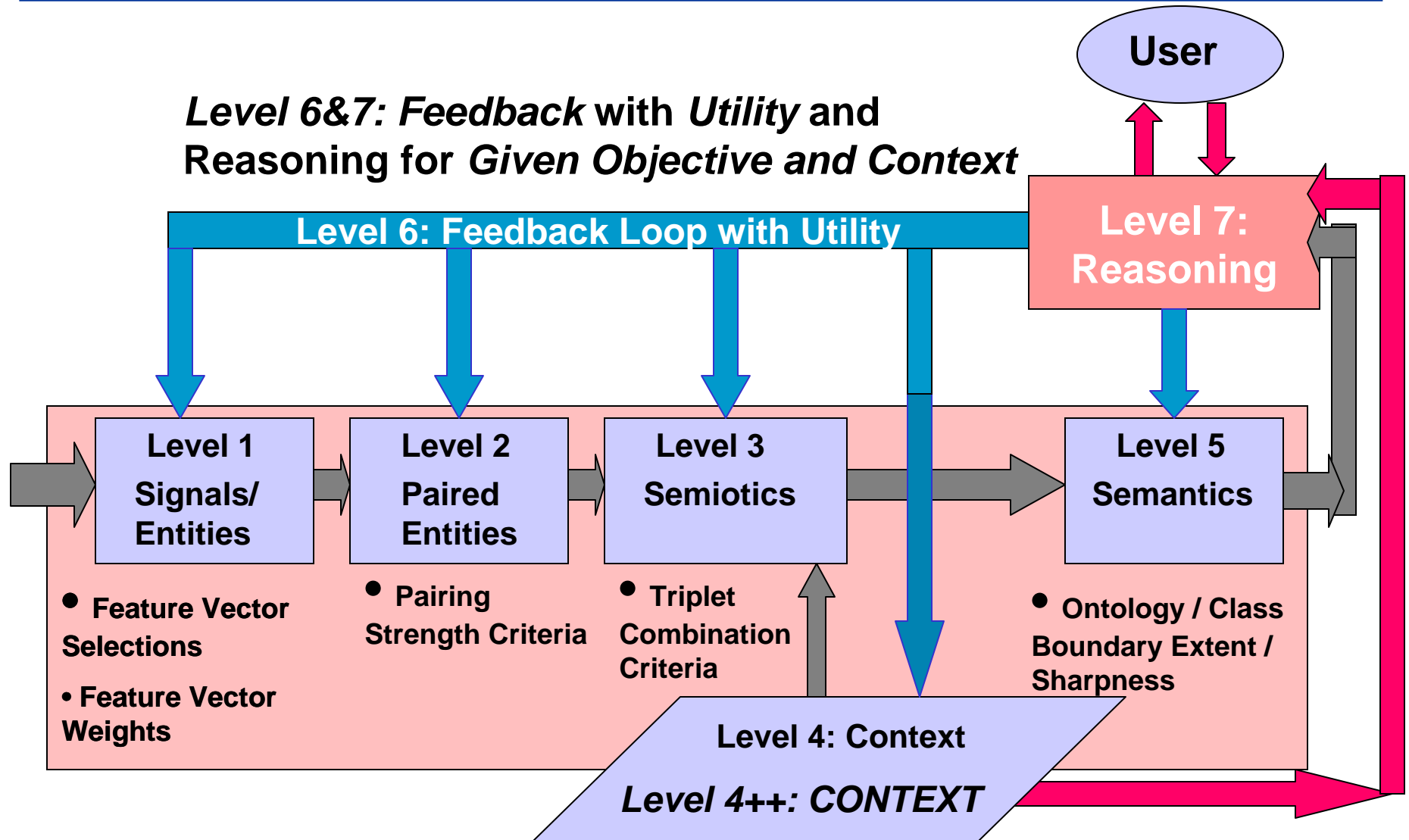
# Air Force Portal Information Processing Model

- Notes
- Database
- Document Management
- News Feeds
- Email
- Internet
- Files
- XML
- Audio/ Media



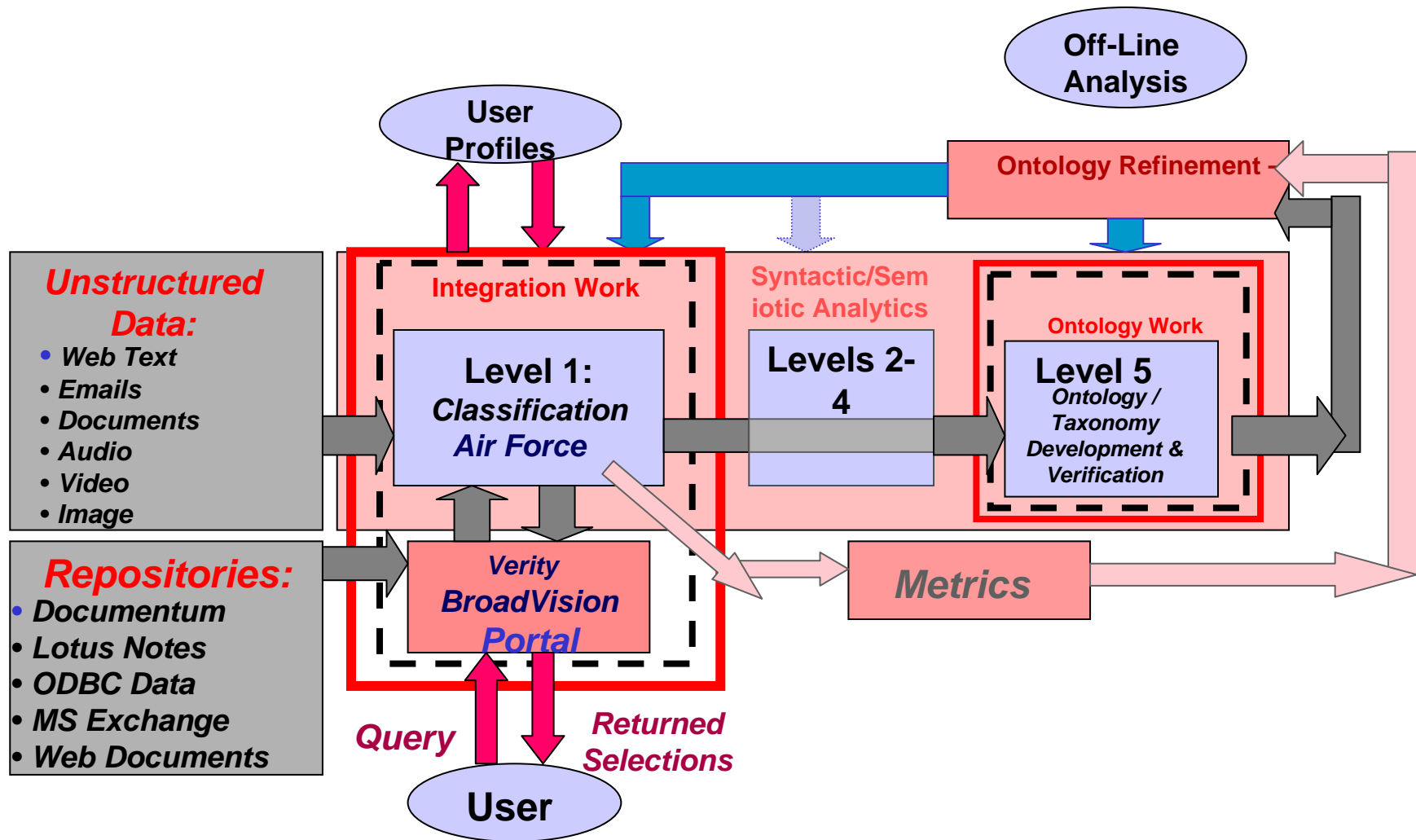


# The 7 Levels of Knowledge Discovery



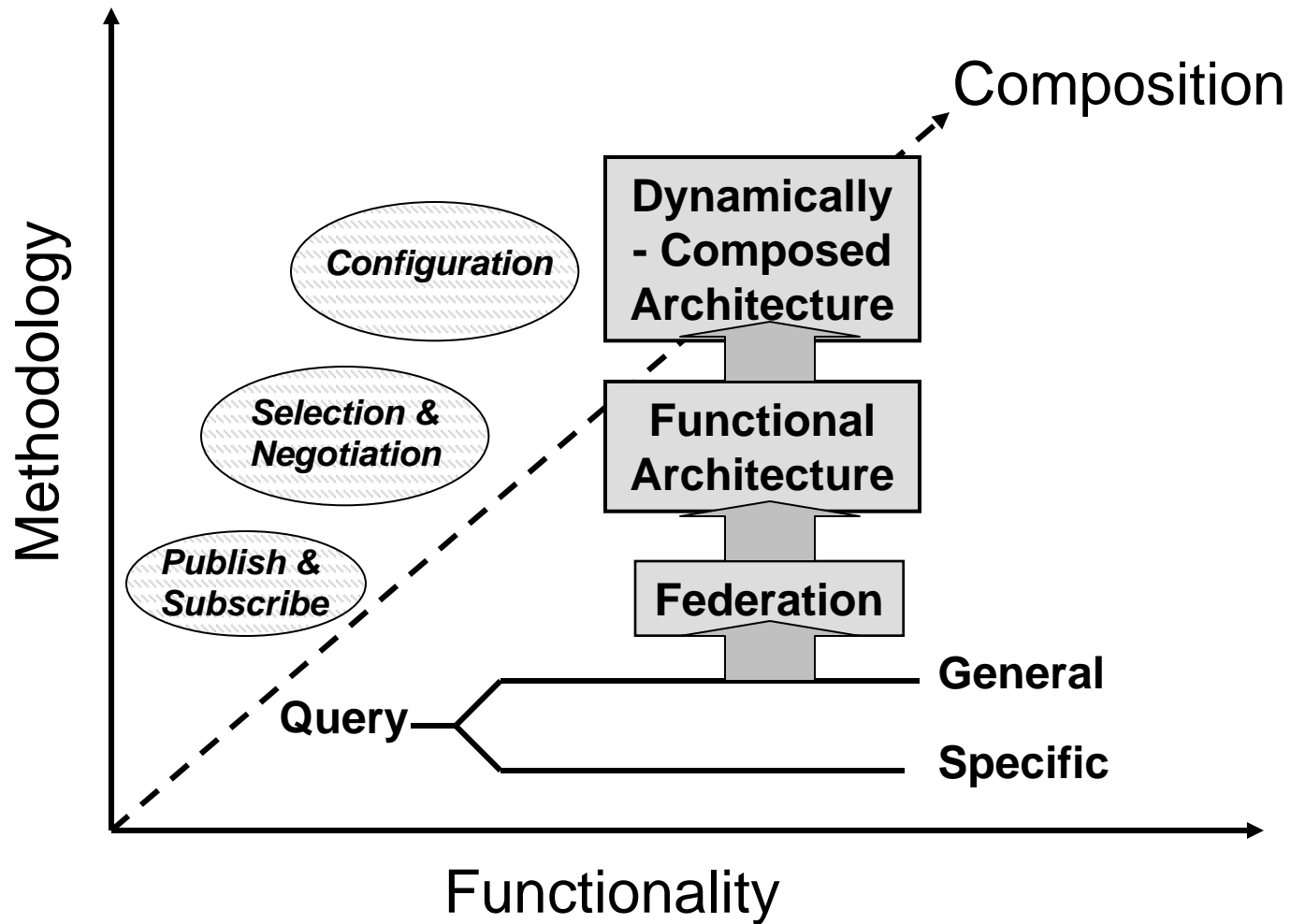


# The Air Force Knowledge Discovery Architecture



*Integrity... Service... Excellence*

# Discovery and Mediation



# Thanks To ---

- John Clink – GD
- Jeff Harrison – NGIT
- Chuck Heazel - LMCO
- Tom Huggins – Polexis
- Paul Lunceford - FGM
- A.J. Maren – EagleForce
- Steve Matney - SRA
- Brad Mediary – Booz-Allen
- Mike Meyer – SAIC
- Hans Polzer – LMCO
- Glenn Pruitt – FGM
- Mark Schiffner – FGM
- Jeff Stohlman – IBM
- Mark Young – GD