Tools and Approaches for Enhancing Semantic Interoperability
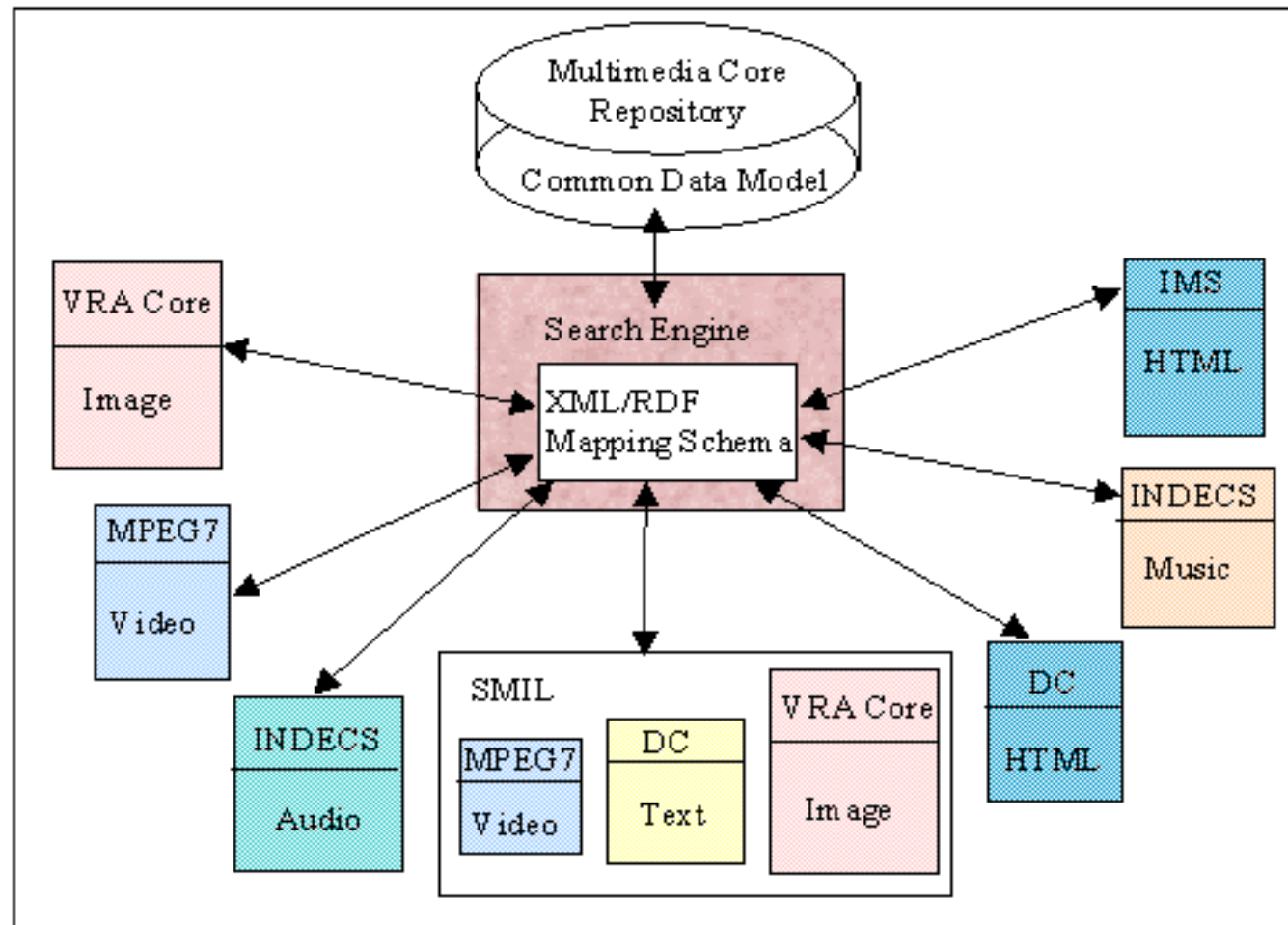
Jane Hunter
jane@dstc.edu.au

Agenda

- Why is Semantic Interoperability important?
- Ontologies, Ontology Languages
- Ontology Tools, Harmonization, Extensions
- Semantic Inferencing/Logic/Querying
- Visualization/Presentation
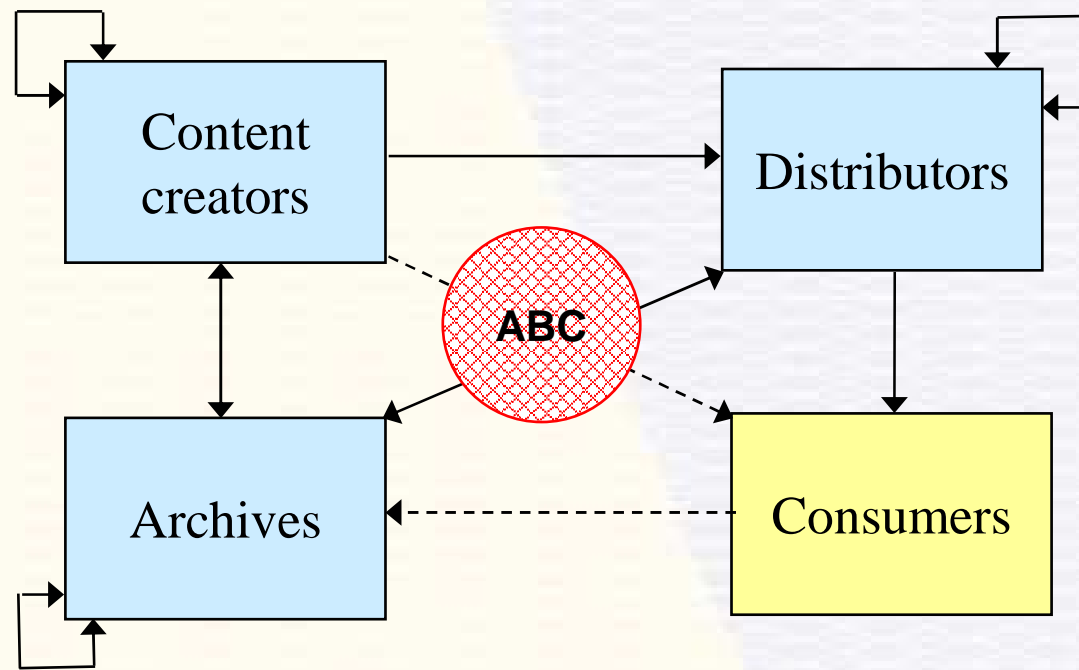- Annotation Tools
- DSTC Projects
- The Future?

Semantic Interoperability

- Provide well-defined meanings for information on the web
- Enable people, machines, services, agents to find, consume, process and produce data
- Determine relationships between web resources – extend hyperlinks to semantic relationships
- Seamless, automated query mediation, content integration and knowledge extraction
- Discover, match and combine web services and resources
- Across communities, organizations, disciplines

DSTC Pty Ltd

Problem 1 – Query Mediation

Problem 2 – Harmonization

Content creators → Distributors

ABC

Archives | Consumers

Sharing, Exchanging and Tracking Information

Problem 3 – Different Views

Cultural Metadata

Scientific Metadata

Historical Metadata

Educational Metadata

Indigenous Elder

Scientist

Historian

Teacher

Problem 4. Integration

NMAI

MNH

LoC

Qld Museum

metadata

metadata

metadata

metadata

Search, retrieve, integrate,
aggregate, value-add

File   Play   View   Content   Help

Examples of chiaroscuro in the works of
Rembrandt van Rijn

Clair-obscur (French) and chiaroscuro (Italian) both
mean "light-dark". Both terms are used to describe
strong contrast of light and dark shading in paintings,
drawings and prints. Although the effect had already
been used for many years, the term only came into
fashion in the late sixteenth century. Originally, the
word came from Italy. The painter Caravaggio
(1573–1610) made chiaroscuro his trademark.

DSTC Pty Ltd

Problem 5. Knowledge Extraction

Examples:

- Shopping Assistant
  - Aggregate price information, ratings, evaluations, locations
- Virtual Travel Agent
  - Find me a hotel close to the Telstra building and water sports, which is 4 star and under $120
- E-Science
  - Bio-informatics
  - Environmental modeling
    - Seismic, hydrological, ecological, geodetic, physical data

DSTC Pty Ltd

MAENAD Objectives

Semantic Interoperability Tools to enable:

- indexing, querying, processing, assimilation, presentation, preservation, sharing, exchange, reuse, integration

of **Mixed media** across:

- Domains
- Enterprises
- Networks/Grids
- Media types
- Granularity

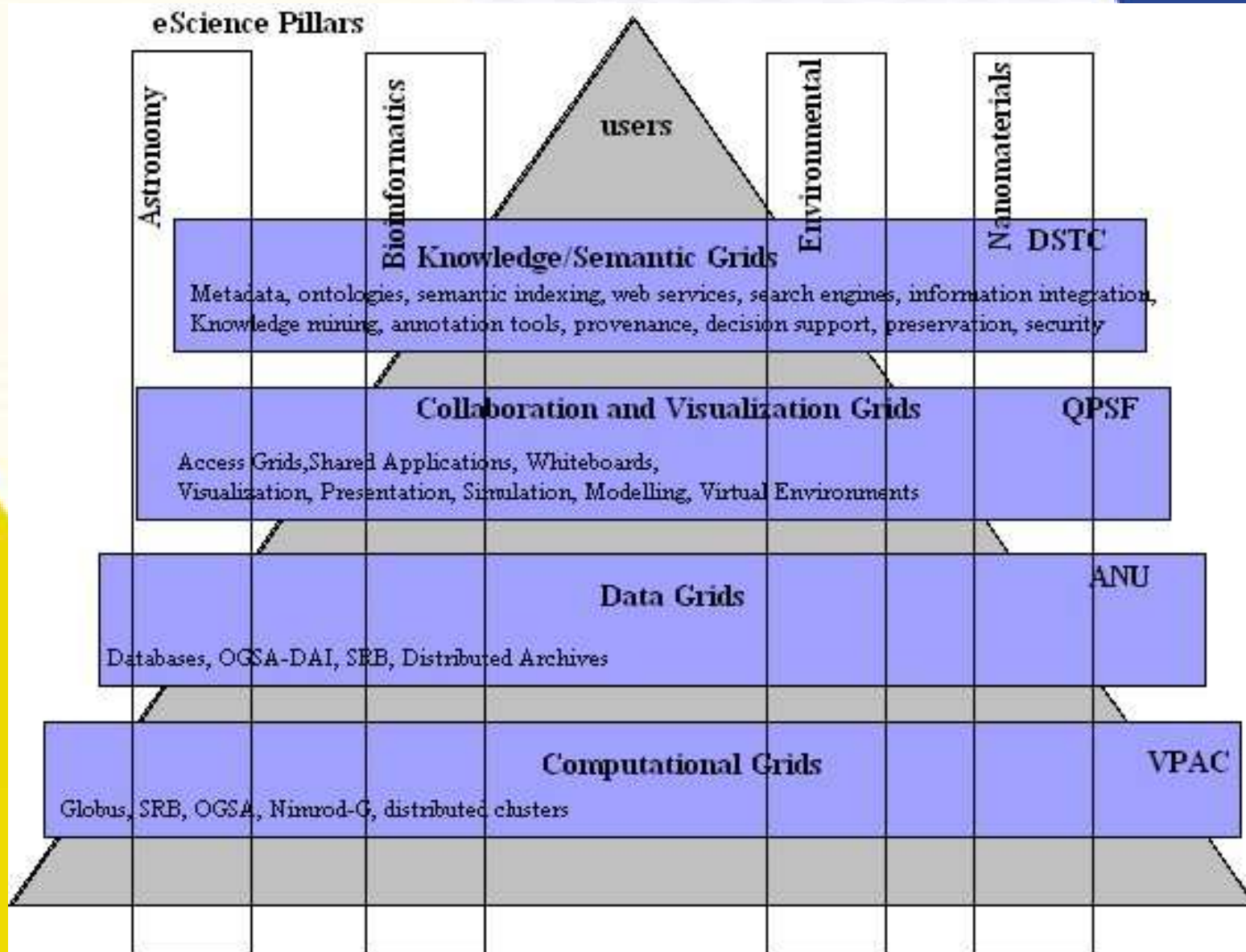Knowledge capture, extraction, layering

Key Components

- Data models, Ontologies
- Metadata standards/schemas
- Metadata capture/generation tools
- Workflows for capture/analysis/reuse
- Semantic Indexing Tools
- Semantic Inferencing Rules
- Search, browse and query interfaces
- Integration, Assimilation, Correlation
- Presentation/Visualization Tools
- Collaborative, Annotation Tools
- Knowledge Mining and Capture
- Preservation Tools

For:

- Resources &
- Services
- Computers
- Instruments
- People
- Software Agents

DSTC Pty Ltd

Middleware Categories

eScience Pillars

Astronomy | Bioinformatics | users | Environmental | Nanomaterials

Knowledge/Semantic Grids   DSTC

Metadata, ontologies, semantic indexing, web services, search engines, information integration, Knowledge mining, annotation tools, provenance, decision support, preservation, security

Collaboration and Visualization Grids   QPSF

Access Grids, Shared Applications, Whiteboards, Visualization, Presentation, Simulation, Modelling, Virtual Environments

Data Grids   ANU

Databases, OGSA-DAI, SRB, Distributed Archives

Computational Grids   VPAC

Globus, SRB, OGSA, Nimrod-G, distributed clusters

Pty Ltd

Why Ontologies?

- Approximate description of a domain by defining *a set of concepts, axioms and relationships* using a formal language
- Definitions of classes, properties and relationships between them – used to describe and represent an area of knowledge
- Semantic interoperability
  - Integration of and mapping between different metadata schemes/vocabularies
  - Richer, more accurate search and browse capabilities
  - Relate information in a page/resource to associated knowledge structures

DSTC Pty Ltd

Ontologies

- UMLS – Unified Medical Language Ontology
- CIDOC/CRM – Conceptual Reference Model
- GONG, GALEN, SNOMED - Biomedical
- MPEG-7 – Multimedia Content Description
- Harmony ABC Core Ontology
- DOLCE – Lab. For Applied Ontology
- IEEE SUO – Standard Upper Ontology
- FUSION – Fuel Cell Ontology

DSTC Pty Ltd

MPEG-7

- Multimedia Content Description Interface
- Developed by ISO/IEC SC29 WG11
- Standardized representation of multimedia metadata in XML (defined using XML Schema Language)
- Describes audiovisual content at multiple levels (*features*, *structure, semantics, models, collections*)
- Published as ISO Standard in March 2002

DSTC Pty Ltd

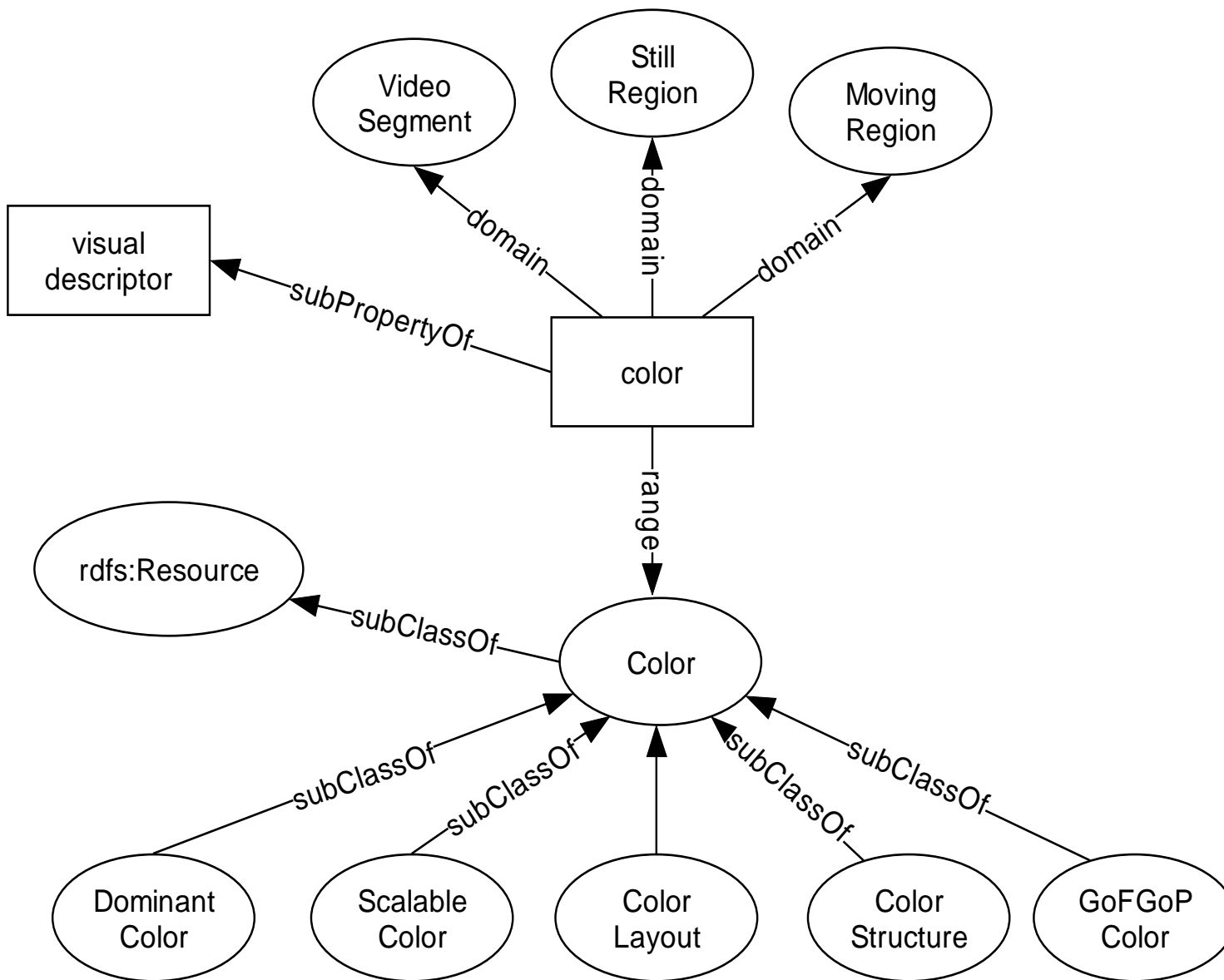# Historical Perspective

- 1999 – RDF Schema rejected as DDL
- 1999-2000 – UML modeling diagrams
- 2000 – Need for Semantic Interoperability/Conceptual Model
- 2001 – RDF/DAML+OIL Ontology
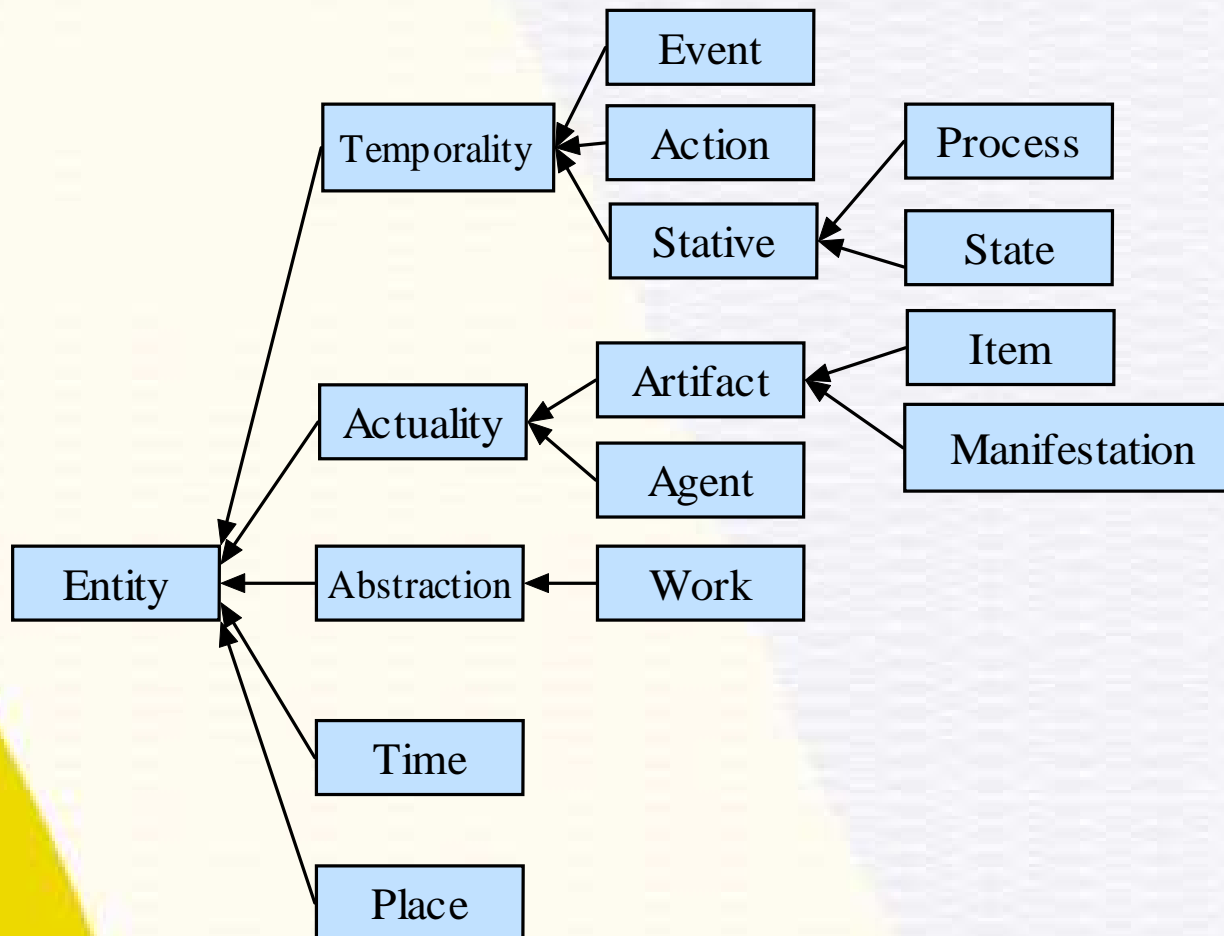- 2002 – OWL Ontology

MPEG-7 Class Hierarchy
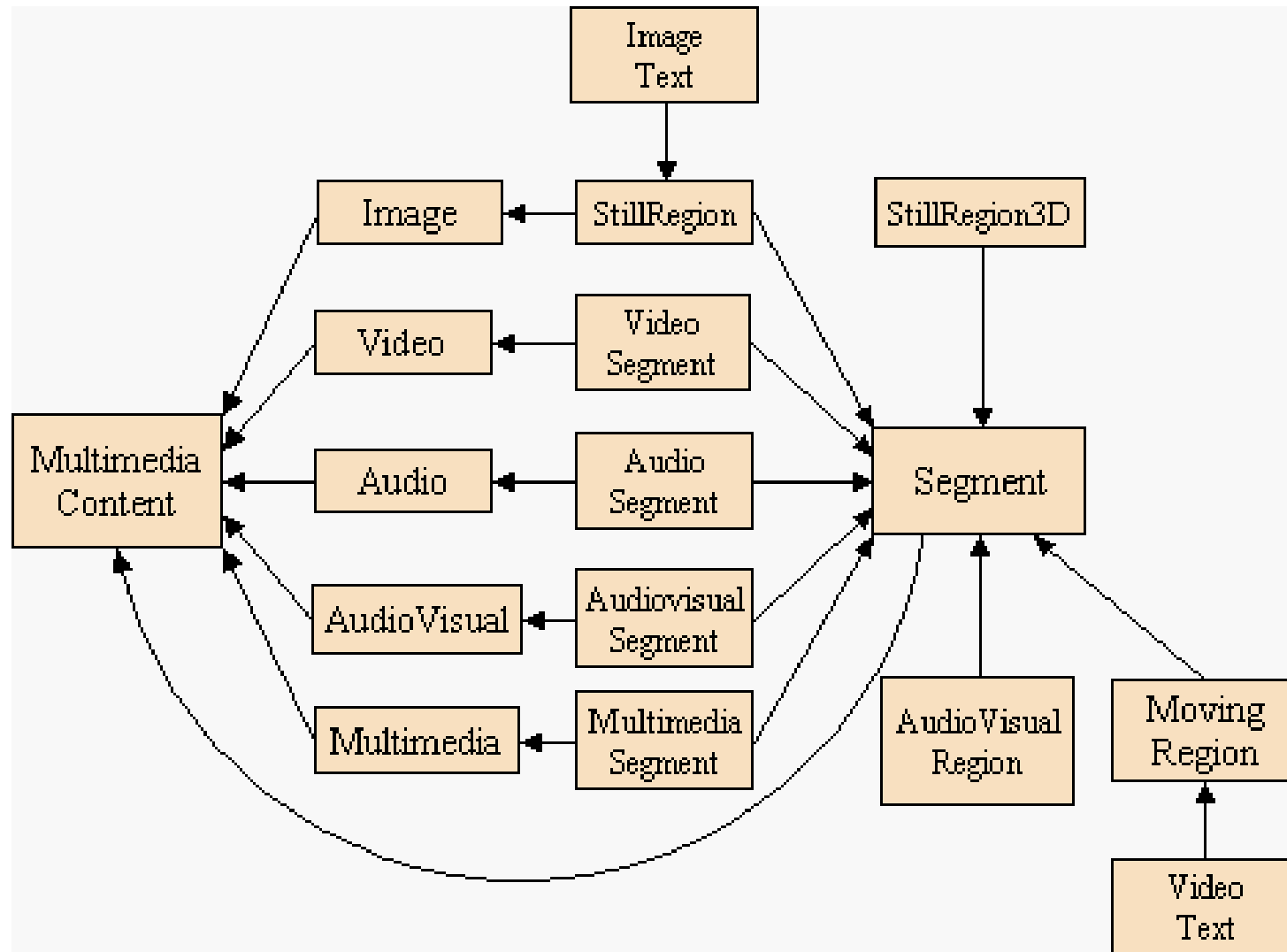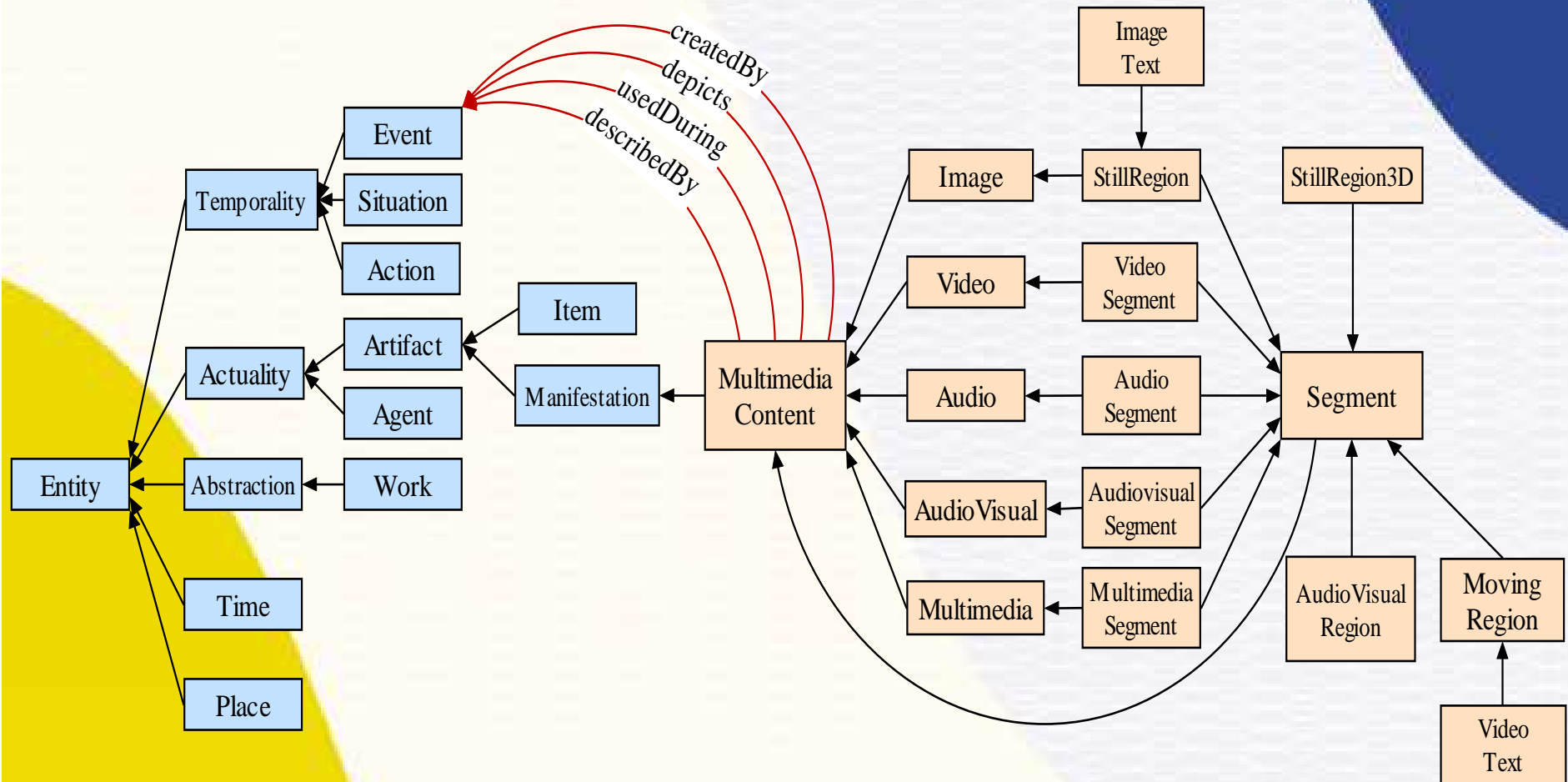
Descriptor Relationships

Feature	Video Segment	Still Region	Moving Region	Audio Segment
Time	✓	-	✓	✓
Shape	-	✓	✓	-
Color	✓	✓	✓	-
Texture	-	✓	-	-
Motion	✓	-	✓	-
Audio Ds	✓	-	-	✓

DSTC Pty Ltd

ABC Ontology

Event

Temporality — Action — Process

Stative — State

Item

Artifact — Manifestation

Actuality — Agent

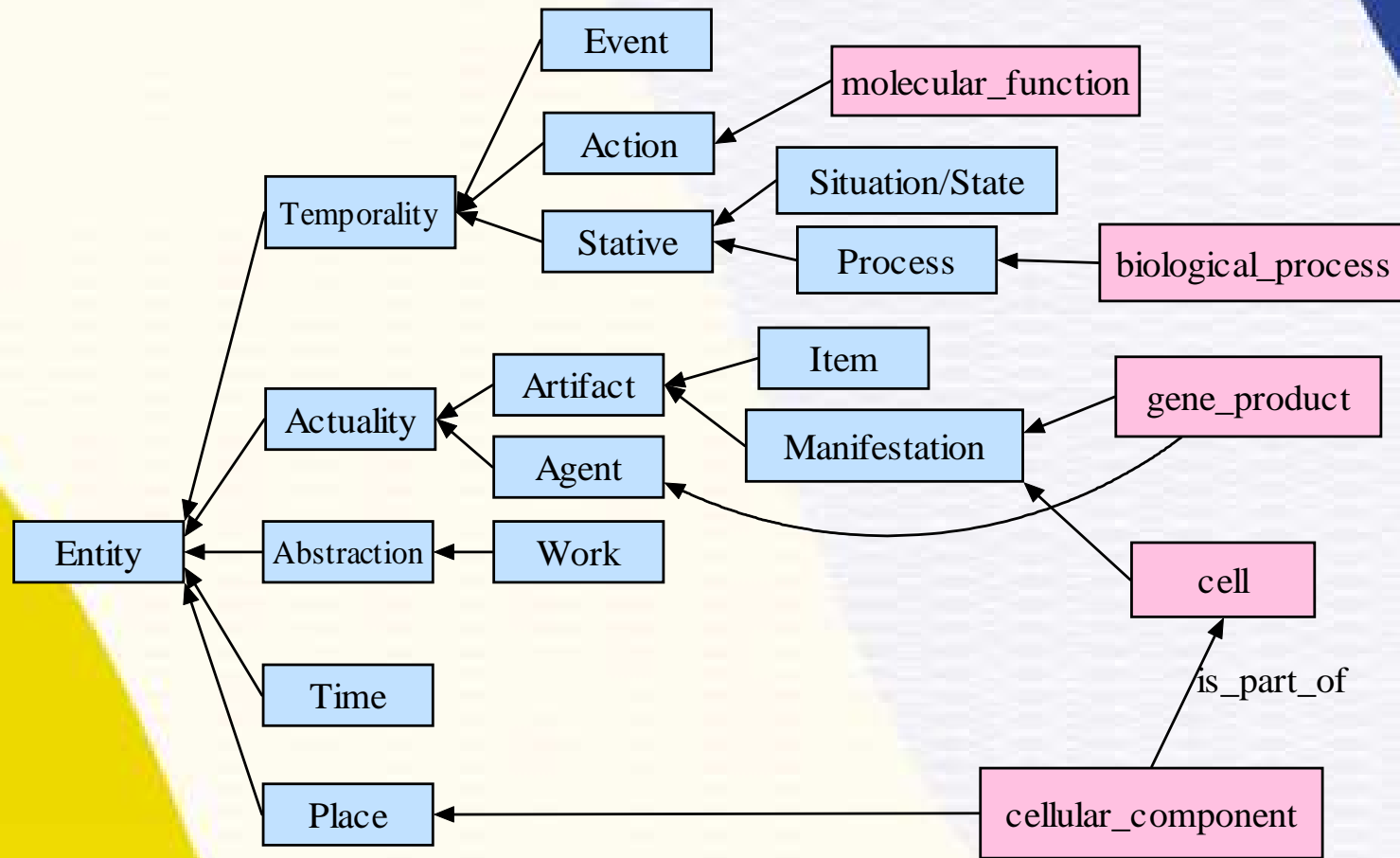Entity — Abstraction — Work
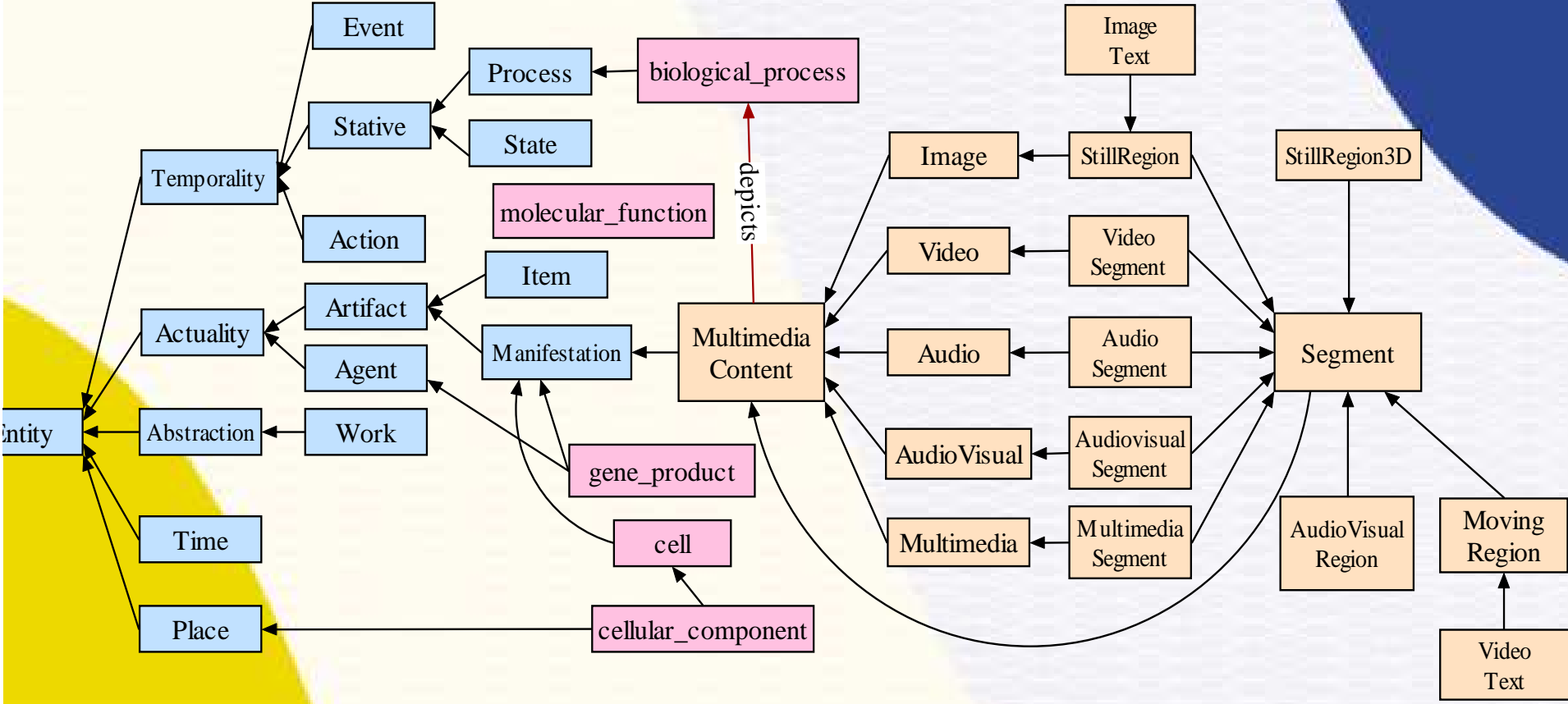
Time

Place

ABC+MPEG-7

DSTC Pty Ltd

ABC+GONG

ABC+GONG+MPEG-7

Query By Image Content

- Retrieve images based on
  - Colour, texture, shape
  - spatial relations between these
- Query by example image
- Doesn't support semantic queries

Semantic Inferencing

- If ((colour ==x)&& (texture==y)
  &&(shape==z)&&(size<a))
  then (object==b)
- Where b is a high-level semantic term in the domain-specific ontology

Domain-specific Semantic Querying

- Give me all images depicting "apoptosis"
  - A term defined in the GONG
  - Inferred from the rules + low level features
  - Low level features – from comparison with example database
  - MathML + RuleML

Application Examples

- Indigenous Knowledge Management
- FUSION
- Vannotea
- PANIC

DSTC Pty Ltd

Conclusions

- Metadata – expensive, poor quality – need for automatic extraction (real-time), workflows, controlled vocabularies, validation, authentication, metadata normalization
- Dynamic, executable, distributed, domain-specific ontologies
- Semantic descriptions/discovery/choreography of web/grid services
- Ontology harmonization – manual, difficult
- Domain-specific semantic weblets feasible
- Global Semantic Web – automated information integration & knowledge mining – still long way off

References

http://www.w3.org/2001/sw

http://metadata.net/harmony

http://metadata.net/sunago/fusion/

http://metadata.net/filmed/

http://semanticgrid.org/

DSTC Pty Ltd