



Knowledge Assertions and Knowledge Organization Systems

Presented to the 16 July 2009 Joint DoD-IC Data Meeting

Office of the Director of National
Intelligence
Chief Information Officer



Goals for Knowledge Management

- **Currently, the content of intel resources is buried in the form of the resource (text, image, video), and is not able to be processed directly**
- **Identifying and describing the content of intel resources, in a structured (unambiguously interpretable) way, with the meaning (semantics) explicitly stated, will improve intelligence analysis, correlation, fusion, discovery, and overall usability of those resources**
- **Developing standards for describing content will help the IC move towards interoperability so these goals are achieved**



Knowledge Assertion Defined

- **Knowledge Assertions (KAs) are the identification of entities in intel resources, such as:**
 - People
 - Organizations
 - Places
 - Weapons
- **KAs also are the identification of attributes of entities and relationships between entities**
- **“Intel resources” include documents, web pages, database records, images, audios, videos, and other digital forms of content**



Resource Segment

- **A specific, possibly non-contiguous, portion of an intel resource about which the *Knowledge Assertion* is being made**

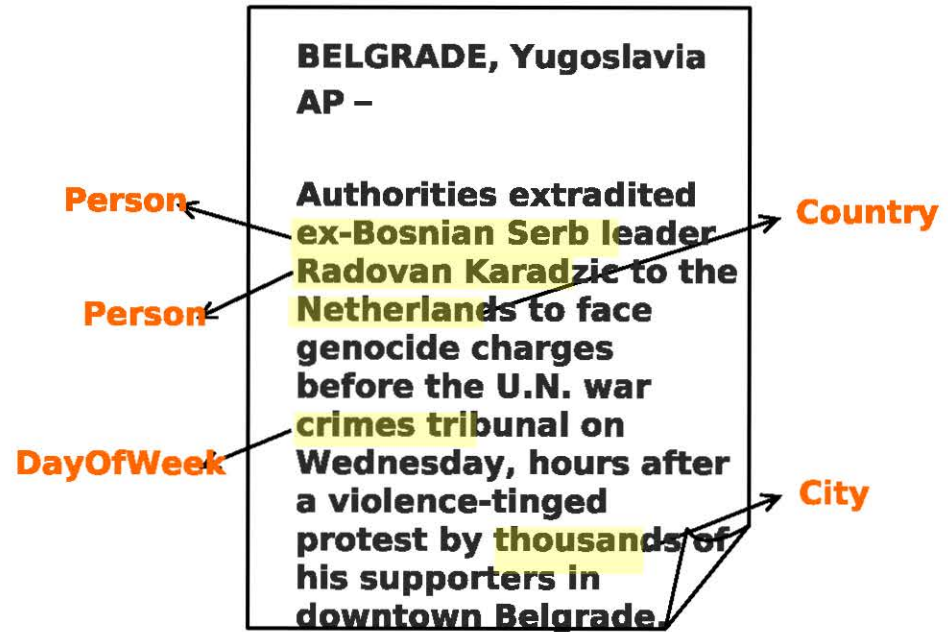
**BELGRADE, Yugoslavia
AP –**

Authorities extradited ex-Bosnian Serb leader Radovan Karadzic to the Netherlands to face genocide charges before the U.N. war crimes tribunal on Wednesday, hours after a violence-tinged protest by thousands of his supporters in downtown Belgrade.



Entity Class

- **Entity** = something that has a distinct, separate existence
- **Entity class** = the asserted type of the entity that is referred to by the **Resource Segment**





Entity Segment Attribute

- **Entity Segment Attribute** = an asserted property of an entity whose value is the **Resource Segment**





Entity Attribute

- **Entity attribute = any additional asserted property of the entity whose value is a quantitative or qualitative characteristic*, that is not a Resource Segment**

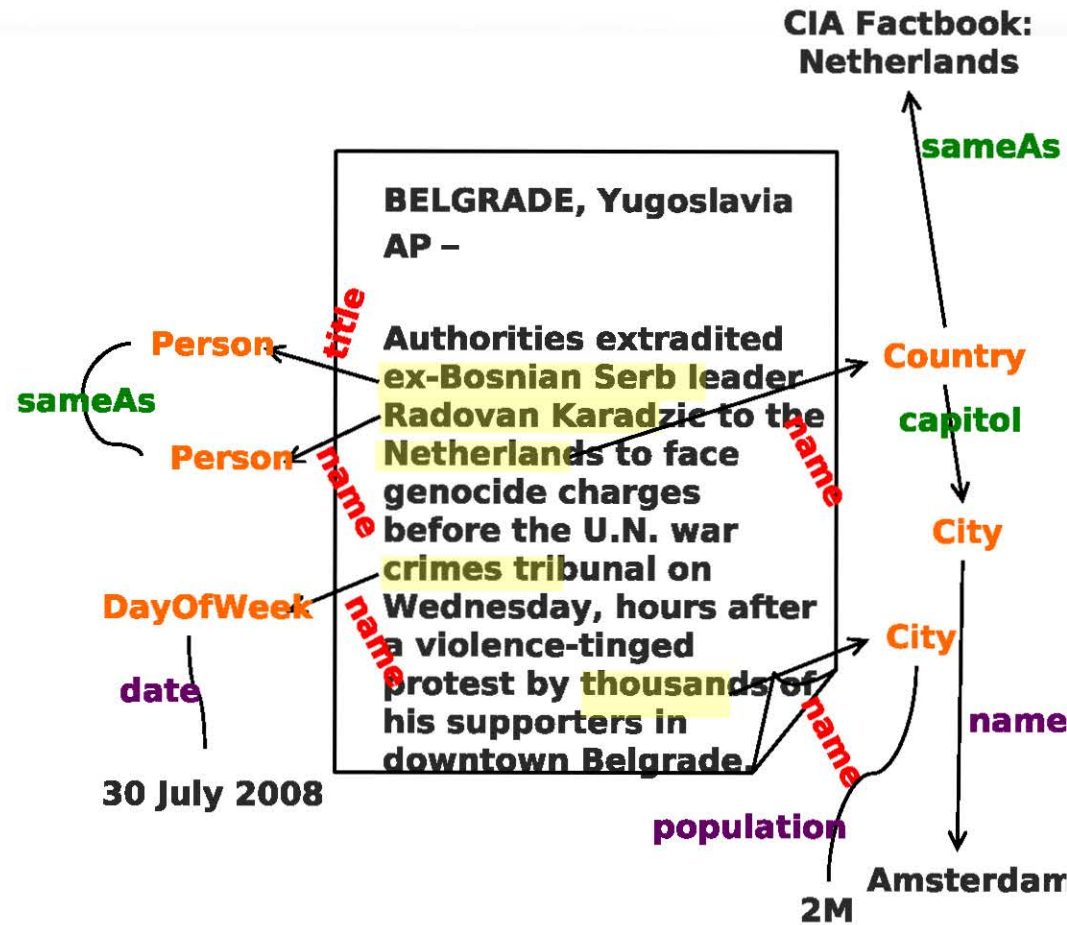
**A number, string, Boolean, quantity with units, etc.*





Entity Relationship

- **Entity Relationship**
= an asserted property of an entity whose value is another entity
- **Variants**
 - Identified by *Resource Segment*
 - Not identified by a *Resource Segment*
 - Identified external to the intelligence resource





Assertion Metadata

- Information about a KA, but not what is being asserted
- Includes:
 - Source
 - Creator
 - Contributor
 - Date
 - Confidence
 - Knowledge Organization System Reference
 - Encoding
 - Publisher
 - Knowledge Assertion Security Mark

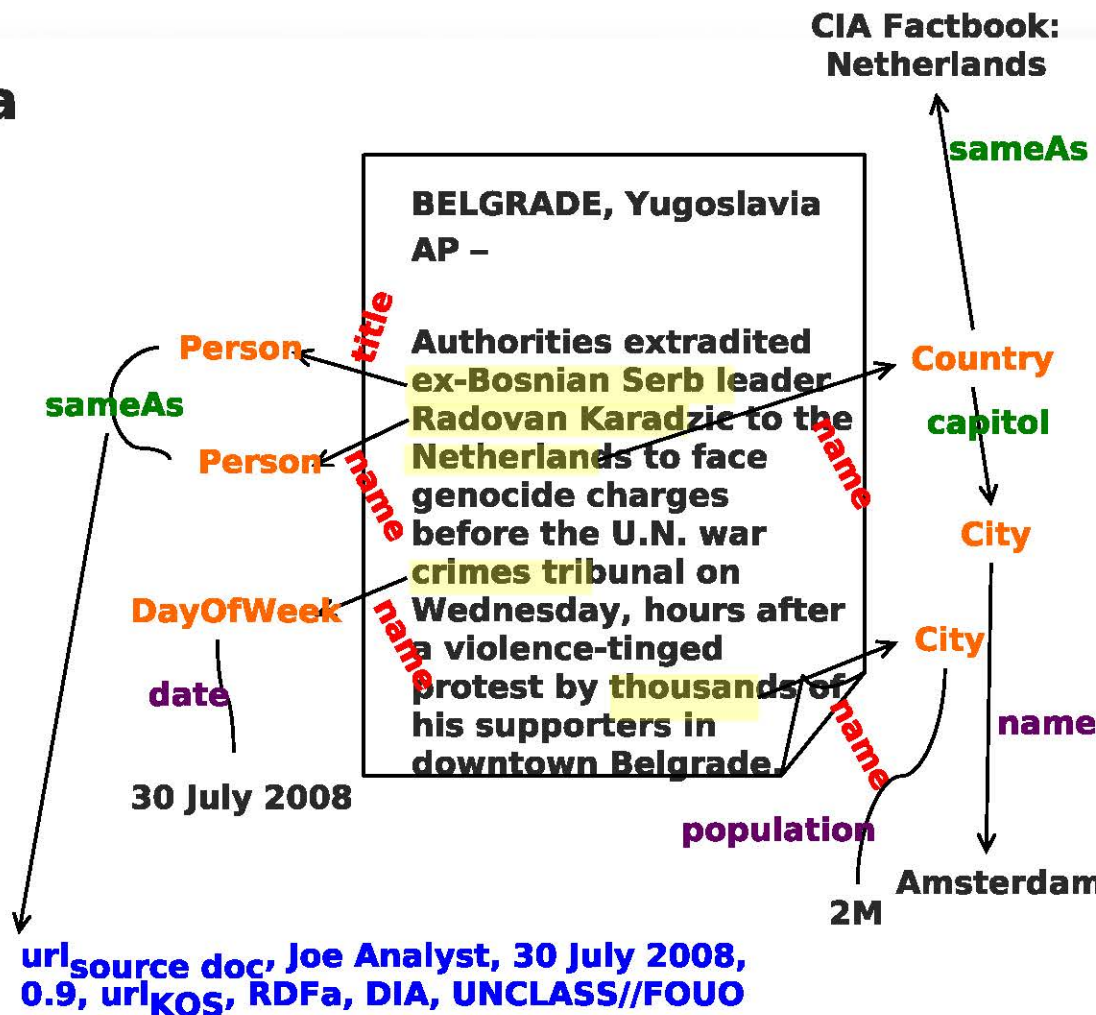




Image Example



Jane's "All the World's Aircraft":
McDonnell Douglas DC-9

E/OImage

Fixed-wingAircraft → MD DC-9
airframe

sameAs

maxTake-off
Weight

164,000 lbs

locatedAt

E/OImage

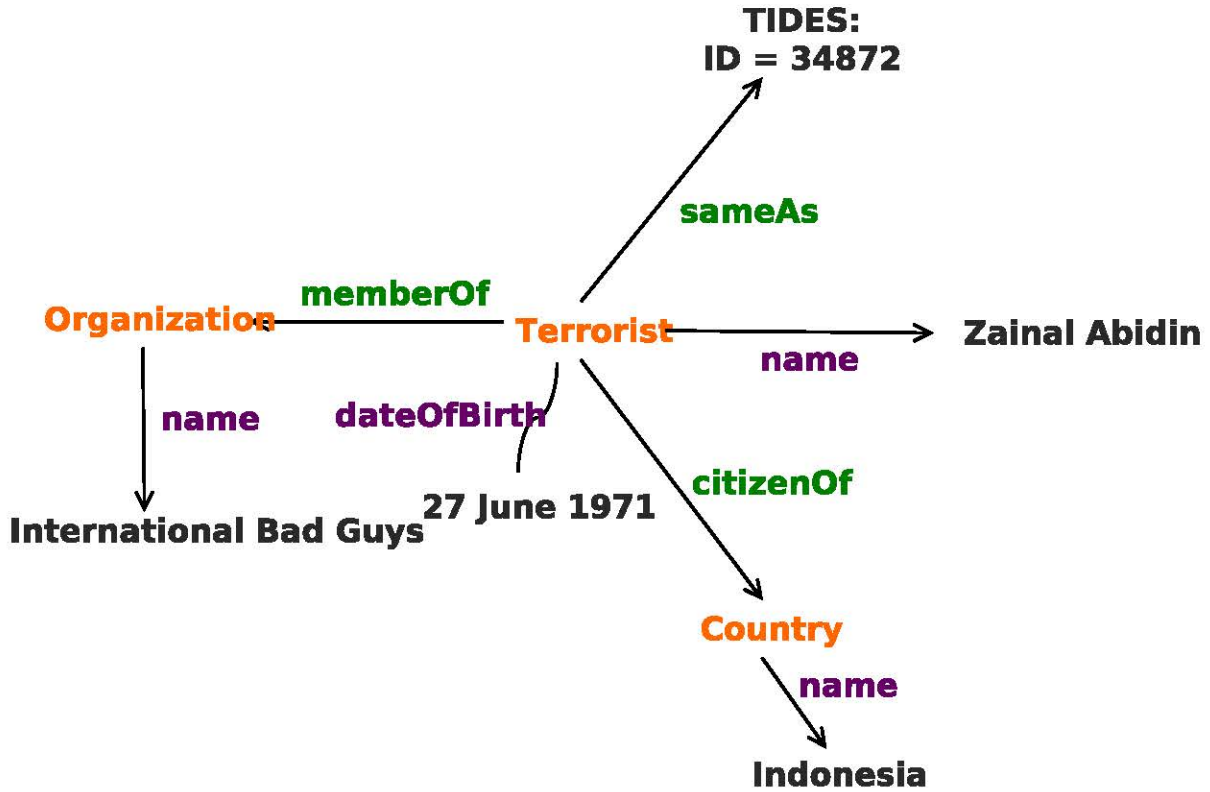
AirportTerminal

name

Ronald Reagan
National Airport,
Terminal B



Example with No Explicit Resource





Knowledge Organization System Defined

- **In previous examples, how did I know to use:**
 - Classes: **Person**, **Country**, **City**, **Fixed-wingAircraft**, **Terrorist**, **Organization**?
 - Attributes: **name**, **title**, **population**, **E/OImage**, **airframe**, **dateOfBirth**?
 - Relationships: **capitol**, **locatedAt**, **citizenOf**, **memberOf**, **sameAs**?
 - Controlled vocabulary: **DayOfWeek name** = Wednesday, **Country name** = Indonesia?
- **Maybe I should have asserted a different attribute?**
 - “Bosnian Serb leader” is **role** of Person, instead of **title** of Person?
- **Aren’t classes related?**
 - Isn’t a **Terrorist** a kind of **Person**?
- **Answer: a *Knowledge Organization System (KOS)* gives meaning (semantics) to the classes, properties, and controlled vocabulary values of the Knowledge Assertions**



Types of KOSs

- **One that deals with *concepts* in a domain**
 - Often called an *ontology*
 - An explicit and formal collection of concepts of interest in a specific intelligence domain
- **One that deals with *terms* in a domain**
 - Often called a *controlled vocabulary (CV)*
 - A collection of natural language (lexical) terms explicitly allowed to be used to refer to aspects of a specific intelligence domain



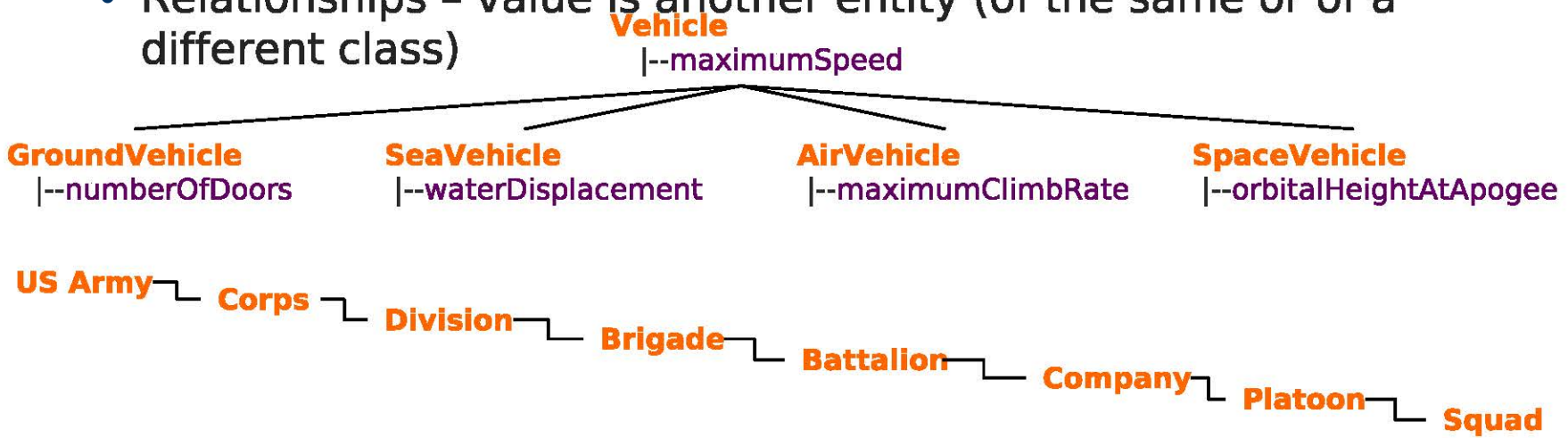
Ontology Elements

● Class

- A grouping of a number *entities* regarded as forming this group by reason of specific properties, characteristics, qualities, or traits

● Properties

- Attributes - value is qualitative or quantitative; attributes can be restricted to have values in a CV
- Relationships - value is another entity (of the same or of a different class)





Controlled Vocabulary

- **Controlled Vocabulary: a defined set of terms**
 - Usage: Preferred term(s), alternate term(s), deprecated term(s)
 - Can include a natural language description of values
- **Common controlled vocabularies in IC**
 - Language Codes
 - Currency Codes
 - Units of Measure
 - Subject Area Taxonomy
 - Orgs and Alliances
 - Classification Markings
 - Dissemination Control Markings
 - Special Access Required Markings
 - Intelligence Disciplines
 - MIME Types
- **Enumeration**
 - A subset of the values of a *single* attribute of an entity class
 - Example: RELTO countries



Connection between Ontology and CV

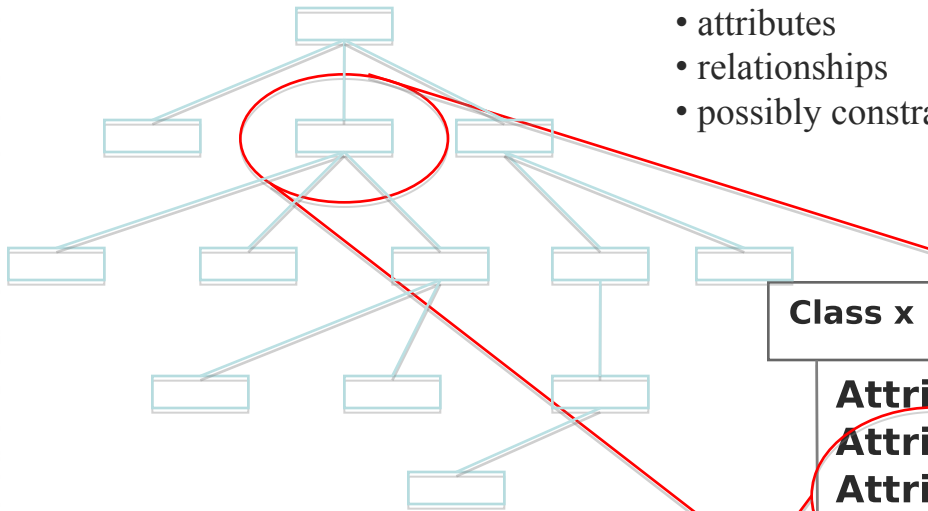
- **CV: a list, where each entry in the list is a set of terms that are allowable values of some attributes of an entity class**
- **Example: a CV of country names and ISO trigraphs, with one entry:**
 - Afghanistan | Islamic Republic of Afghanistan, AFG
- **Although this definition of a CV is in the context of an ontology, the connection is often implicit**
 - But it wouldn't make sense to have the CV without stating what the terms mean, which can be done informally (e.g., in English) or formally (e.g., in an ontology)



Unified View of KOSs

Ontology

- class hierarchy
- attributes
- relationships
- possibly constraints, rules, etc.



Class x

Attribute x₁
Attribute x₂
Attribute x₃
Attribute x₄
Attribute x₅
⋮
Relationship x₁
⋮

Controlled Vocabulary

- allowed value sets for multiple attributes
- possibly relationships

(Attribute x₂, Attribute x₃, Attribute x₄) =
{ (a₂, a₃, a₄) | (b₂, b₃, b₄) | ... | (z₂, z₃, z₄) }

Enumerated list

Attribute x₄ = • allowed value subset for
{ a₄ | b₄ | ... | z₄ } single attribute



KOS Metadata

- **Each class/property/term has metadata**

- Contributor
- Creator
- Date
- Description
- Class/property/terms Security Mark
- Source
- Title

- **Entire KOS has metadata**

- Contributor
- Creator
- Date
- Description
- Identifier
- Encoding
- Publisher
- KOS Security Mark
- Source
- Title



Knowledge Object Defined

- **A set of Knowledge Assertions all of which are “about” the same entity is a *Knowledge Object***
- **Formally, a KO is the set of assertions that defines the class of the entity plus all attributes and relationships of that entity, according to a unified ontology**
- **An important objective of intel collection and analysis is the integration of “everything” the IC knows about an entity, which can be represented as a KO**

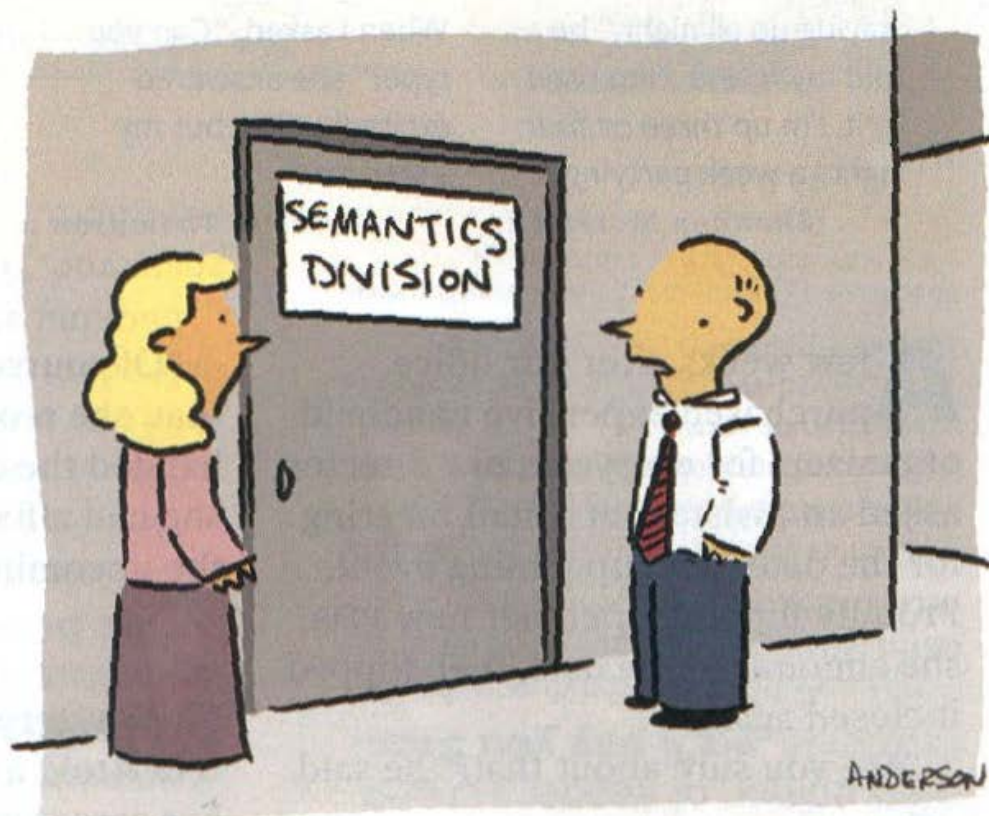


Summary

- **A *Knowledge Assertion* is a means to express specific content of an intelligence product using a formal mechanism**
- **A *Knowledge Organization System* provides contextual meaning (semantics) to KAs**
- **A *Knowledge Object* is the set of KAs about a single entity, which taken together is what is known about that entity**



Questions?



"We're really more of a department."