

The background of the slide is a deep blue space scene. On the left, a large portion of the Earth is visible, showing continents and oceans. In the upper center, a bright sun or star is partially obscured by a red ring of light, with several smaller, fainter rings of light around it. The overall atmosphere is futuristic and high-tech.

NORTHROP GRUMMAN

DEFINING THE FUTURE

**Semantics
for Information Sharing and Discovery
in the Intelligence Community**

OIC 2008, Dec. 3-4, 2008

Martin Thurn
Computational Linguist

Impediments to Sharing & Discovery

- Classification Levels
 - Security levels
 - Compartment boundaries
 - Need-to-know

Impediments to Sharing & Discovery

- Classification Levels
 - Security levels
 - Compartment boundaries
 - Need-to-know
- Document-based classification
 - Entire document is treated as classified, even if most of it is unclassified
 - Analyst might only get a list of documents' titles, dates, etc.
 - Analyst might not know a document exists at all

Antidote

- Physically-separate
 - Stored away from original data (not embedded)
- Standard
 - Expressed in RDF and/or OWL
- Semantic
 - Meaning of the data, not just syntax
- Metadata
 - Data that describes other data
- Standard Semantic
 - Adhere to ontologies

A Little History Lesson

- Physically-Separate Semantic Metadata has been used for centuries
- Library card-catalog system
 - Book might be rare / delicate / protected
 - Card is never rare / delicate
 - Card can be replicated / updated / shared

Physically-Separate Metadata in the IC

- Metadata can be unclassified
 - By nature
 - By design
- Metadata can be copied / shared / stored anywhere
- Secret documents can not be shared, but
- Unclassified metadata can be shared
- Metadata can be much larger than a 3x5 card

Ontologies for Discovery

- Ontologies allow terms to be abstracted
 - with minimal loss of meaning
- Semantic Query
- Abstracted indexes

- Rich standard ontologies across the IC

Metadata for Discovery

- Physically separate metadata is typically the last resort
 - Because it is not likely to contain what the searcher wants to find
- Rich semantic metadata
 - Increases recall
 - Increases precision
- Standard semantic metadata
 - Enables automation

Scenario for Semantic Discovery

- Analyst needs to map locations of objects
- Data comes from plain-English descriptions of dates and locations
- In the past, analyst needed to read the documents to find the locations (not to mention type in all the coordinates)
 - If documents are in a restricted collection, she won't find them
- In our approach, unclassified RDF metadata is automatically created
 - Even if documents are in a restricted collection
- Analyst can easily discover that metadata
- In addition, the coordinates can be automatically mapped!

Semantic Metadata for Information Sharing



- Semantic Metadata expressed in RDF
 - Automatically generated using COTS
 - Unambiguous
 - Machine interpretable

Semantic Metadata for Information Sharing

- Semantic Metadata expressed in RDF



- Classification Guide and policies expressed in OWL
 - OWL can express logical inferences
 - Unambiguous
 - Machine interpretable

Semantic Metadata for Information Sharing

- Semantic Metadata expressed in RDF



- Classification Guide and policies expressed in OWL



- Automatic redaction by abstraction (or deletion)

Ontologies for Information Sharing

- Semantic Metadata expressed in RDF



- Classification Guide and policies expressed in OWL



- Automatic redaction by abstraction (or deletion)



- Theorem prover
 - Open Source
 - Sound and complete

Ontologies for Information Sharing

- Semantic Metadata expressed in RDF



- Classification Guide and policies expressed in OWL



- Automatic redaction by abstraction (or deletion)



- Theorem prover



- Mathematical proof that redacted RDF is unclassified

Scenario for Semantic Information Sharing



- Local Police needs names of persons of interest
- Data comes from FBI secret interrogation transcripts
- In the past, FBI agent needed to read the transcripts to find the appropriate names, then get approvals to disseminate

- In our approach, whenever FBI creates an interrogation transcript:
 - Semantic metadata is automatically created
 - Semantic metadata is automatically redacted
 - Semantic metadata is automatically sent to Police

NORTHROP GRUMMAN

DEFINING THE FUTURE