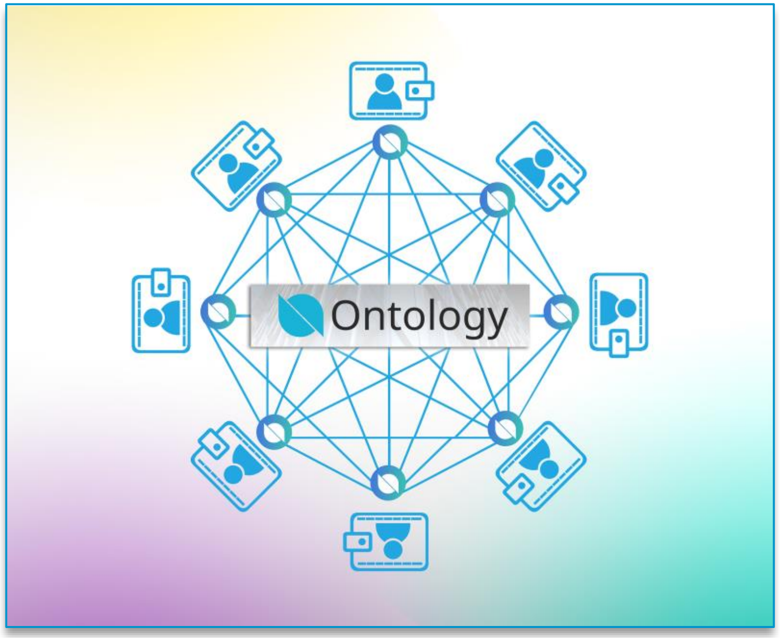


Building Ontologies for Knowledge Discovery

Heather Hedden | Taxonomy Consultant and Trainer – Hedden Information Management

James Morris | Sr. Information Scientist – Smartlogic

Today's Speakers

Smartlogic™

Heather Hedden is a Taxonomy Consultant and Trainer for Hedden Information Management. She is an Instructor of taxonomy-creation online courses, and previously served as Senior Vocabulary editor at Gale, A Cengage Company

Heather is the Author of *The Accidental Taxonomist* (2010, 1st ed.; 2016, 2nd ed.), published by Information Today Inc.

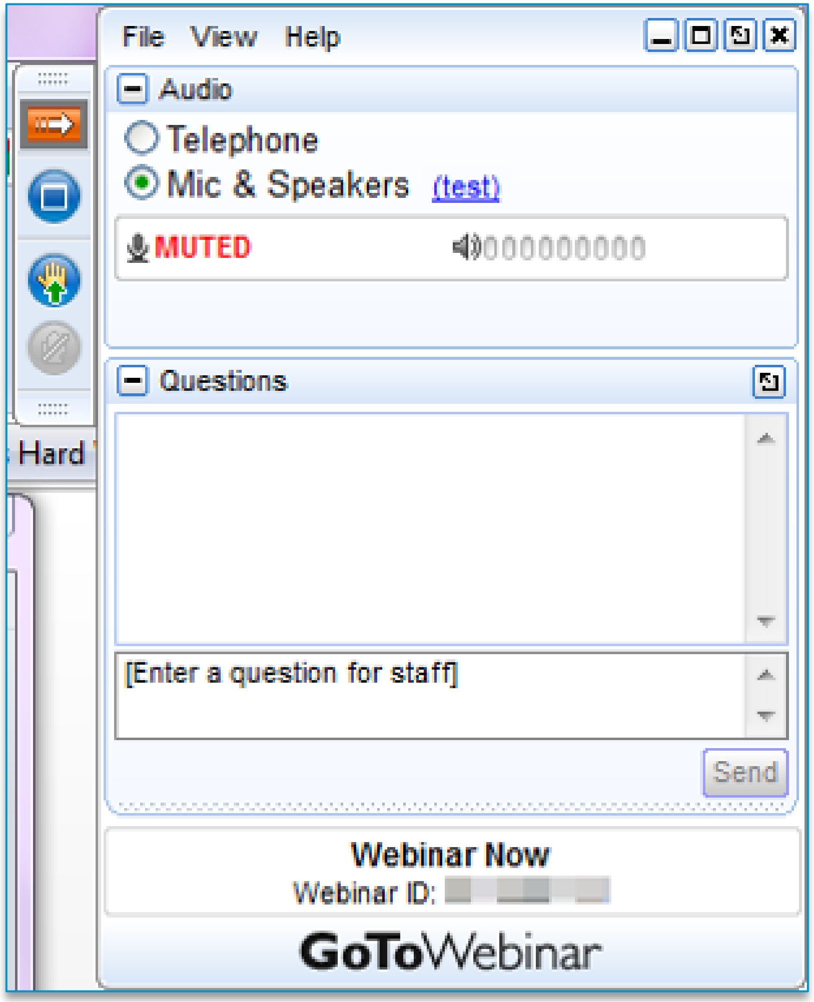
Jim has over 20 years of information systems expertise and has developed proficiency in many library, content, document, taxonomy, and semantic technologies. He thrives on building relationships, influencing across boundaries, and synthesizing different perspectives.

Prior to joining Smartlogic, he worked in pharmaceutical research, academia, and software services. He is a frequent presenter at library and knowledge management conferences, and his paper "Vocabulary Interoperability in the Semantic Web" was published in the Bulletin of the SLA Taxonomy Division in 2013. Jim holds a Master's degree in Information Studies from Drexel University in Philadelphia, PA.

A Few Housekeeping Items

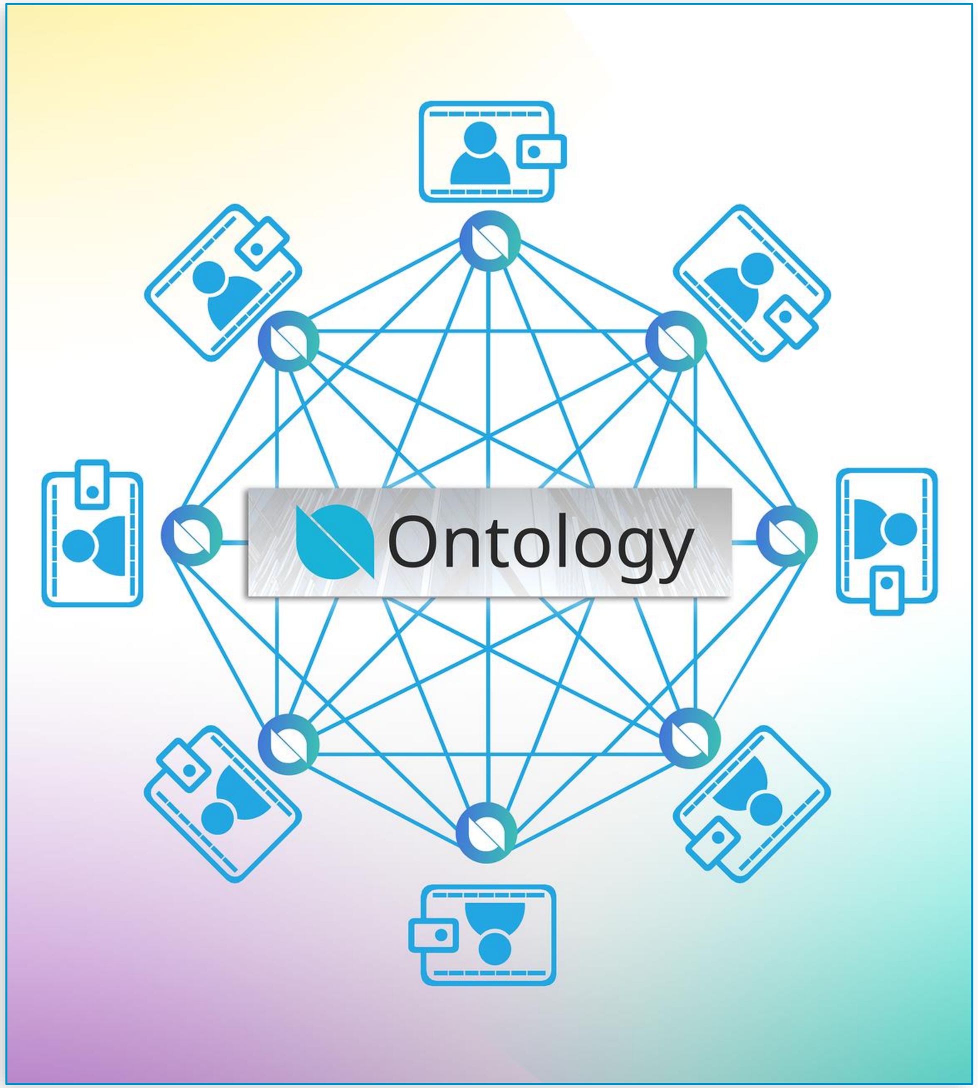
Smartlogic™

- → This webinar is in broadcast mode all participants are muted.
- → Please put your questions in the GoToWebinar panel and we'll answer as many as we can in the Q & A session.
- → This broadcast is being recorded – replay information will be sent to all registrants following the broadcast.



Agenda: Building Ontologies for Knowledge Discovery

- → What are Ontologies
- → Building Ontologies
- → Sources for Ontologies
- → Using Ontologies



What are Ontologies?

Ontologies as Knowledge Models

What is a Knowledge Model?

- → A knowledge model describes the entities and organization of information in a domain for a project.
- → A knowledge model comprises the concepts, their labels, their metadata/attributes, their relationships, and the rules for usage.
- → It could be a thesaurus, a taxonomy (as a set of hierarchies, not just one), or an ontology.

Ontologies as Knowledge Organization Systems

Controlled Vocabularies / Knowledge Organization Systems

Less **Support for Complexity** *More*

Term List	Synonym Ring	Name Authority	Taxonomy	Thesaurus
Ambiguity control		Ambiguity control	Ambiguity control	Ambiguity control
	Synonym control	Synonym control	(Synonym control)	Synonym control
			Hierarchical relationships	Hierarchical relationship
				Associative relationships

Ontologies as Knowledge Organization Systems

Smartlogic™

Term List	Synonym Ring	Name Authority	Taxonomy	Thesaurus	Ontology
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships	Ambiguity control (Synonym control) Semantic relationships Classes

Ontologies as Knowledge Organization Systems

Smartlogic™

Ontology				
Term List	**Synonym Ring**	**Name Authority**	**Taxonomy**	**Thesaurus**
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships

Ontologies as Knowledge Organization Systems

Smartlogic™

Controlled Vocabularies / Knowledge Organization Systems

Less **Support for Complexity** *More*

		Ontology		
Term List	**Synonym Ring**	**Name Authority**	**Taxonomy**	**Thesaurus**
Ambiguity control	Synonym control	Ambiguity control Synonym control	Ambiguity control (Synonym control) Hierarchical relationships	Ambiguity control Synonym control Hierarchical relationship Associative relationships

Ontologies Defined

Smartlogic™

What is an ontology?

- → A complex type of knowledge organization system,
  or a more abstract layer in describing a knowledge organization system.
- → A form of "knowledge representation."
- → A kind of knowledge model to comprehensively describe a domain of knowledge.
- → A formal naming and definition of the types, properties and interrelationships of entities in a particular domain.

Ontologies Defined

Smartlogic™

Ontology Characteristics

(from the perspective of controlled vocabularies/knowledge organization systems)

- → Concepts have attributes
- → Concepts belong to classes, based on having shared attribute types
- → Classes have semantic relationships between them, which are inherited by the concepts in the classes.
  - o Optionally, the ability to support Linked Data

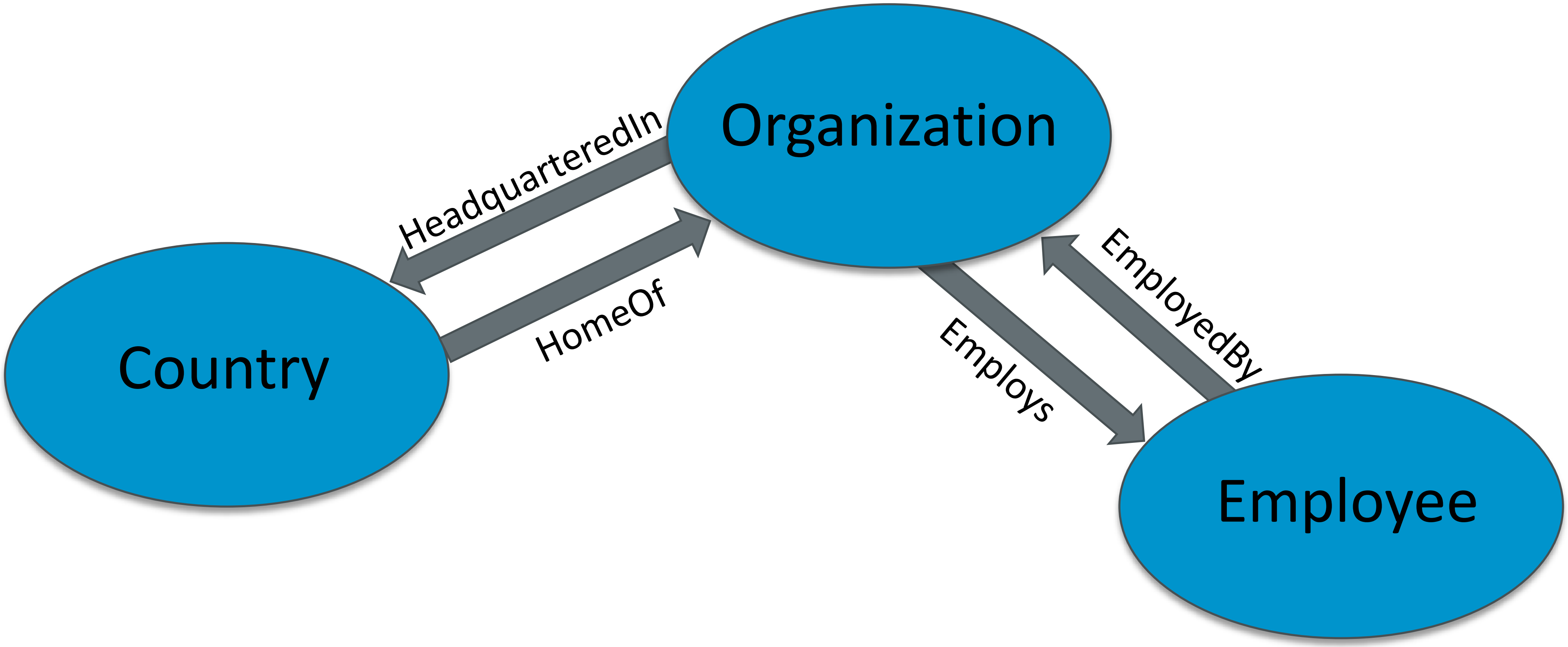
Ontology Types

Smartlogic™

- → **Upper or core ontologies (top-level ontology, upper model, foundation ontology)**
  - A generic, standard framework to serve as a model for a domain ontology, taxonomy, or other KOS
  - Examples: Basic Formal Ontology (BFO), gist, SUMO (Suggested Upper Merged Ontology), SKOS, BIBFRAME, FOAF
- → **Domain or custom ontologies**
  - Concepts belong to a specific subject domain
  - Examples: Systems Biology Ontology, Gene Ontology, BBC Ontology, Financial Industry Business Ontology (FIBO)
- → **"Ontology" may also refer to a combination of a taxonomy with a custom ontology layer.**

Ontology Types

Domain Ontology Example



Relevant Standards

Smartlogic™

OWL (Web Ontology Language)

- → A World Wide Web (W3C) specification https://www.w3.org/OWL
- → First published in 2004; OWL 2 (with extended features), published in 2009 https://www.w3.org/TR/owl2-overview
- → “A Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things”
- → To provide a common way to process the content of web Information. Enables knowledge linking on the web/Semantic Web.
- → A computer-readable language, a declarative language (not a programming or schema language)
- → Enables knowledge linking on the web/Semantic Web
- → Based on RDF and RDFS. OWL is W3Cs attempt to extend RDFS.

W3C® OWL

Relevant Standards

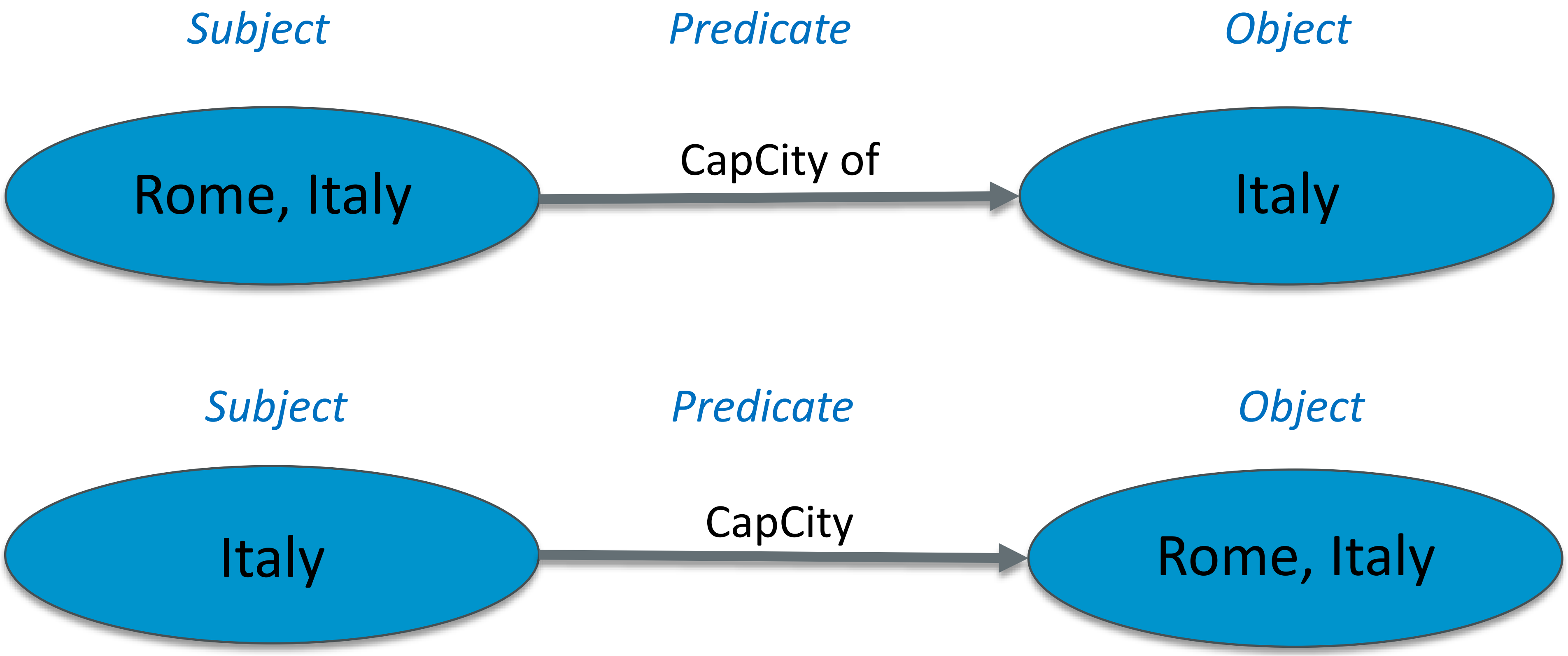
Smartlogic

RDF (Resource Description Framework)

- → A World Wide Web (W3C) recommendation https://www.w3.org/TR/rdf11-concepts
- → Adopted by the W3C as a recommendation in 1999, RDF 1.1 specification in 2014
- → More fundamental, basic, and generic than OWL.
- → "A standard model for data interchange on the Web"
- → Requires the use of URIs (Uniform Resource Identifiers) to specify things and to specify relationships.
- → Facilitates data merging even if the underlying schemas differ.
- → Models information as subject – predicate – object triples.

Relevant Standards

RDF triple: (1) Subject – (2) Predicate – (3) Object



Relevant Standards

OWL basic components

→ **Classes** – *subjects or objects (domains and ranges) of RDF triples*

- May contain individuals (instances of the class) and other subclasses
- Sets of concepts that share characteristics and relationships
- *In knowledge organization systems*: Concept schemes, top concept in a scheme, or concepts with narrower concepts

→ **Individuals – subjects or objects (domains and ranges) of RDF triples**

- Members or instances of a class.
- In knowledge organization systems: Concepts

→ **Properties – predicates of RDF triples**

- Relations between instances or classes (2-way)
- Attributes of instances or of classes (1-way)

Building Ontologies

Building Ontologies

Creating a Domain Ontology (as a custom knowledge organization system)

1. **Use a tool (such as Semaphore) that builds ontologies in SKOS, RDFS and OWL, so it's interoperable with other SKOS vocabularies and the Semantic Web.**
2. **Consider starting with a core (upper) ontology as a generic model.**
3. **Knowledge modeling is the basic task:**
   - Define the scope
   - Identify the various classes (e.g. people, places, organizations, products)
   - Identify the relationships between classes
   - Identify the attributes for classes
4. **Create specific instances within the classes and apply the relationships**
   - As combining an ontology with the specifics of a taxonomy
   - The taxonomy can already exist and be made more expressive,
     or it can be created along with the ontology as an integrated project.

Building Ontologies

Ontology knowledge modeling

Identify and define **Classes**

- Specific enough for business needs
- Generic enough to include various, multiple instances
- Not to be more complicated than necessary

Building Ontologies

Ontology knowledge modeling

Identify and define **Relations** between Classes

- OWL considers Relations as Properties.
- Relations are between pairs of Classes.
- *From* the “Domain” (Subject) class; *To* the “Range” (Object) Class.
- Relations are relevant (inherited) between all instances of each Class.
- Although not required, Relations typically are bidirectional (with an inverse).

Building Ontologies

Ontology knowledge modeling

Specify the desired **Attributes** for each Class

- Simple properties for a specific class
- Just enough for business needs
- Also applicable to all instances within the Class

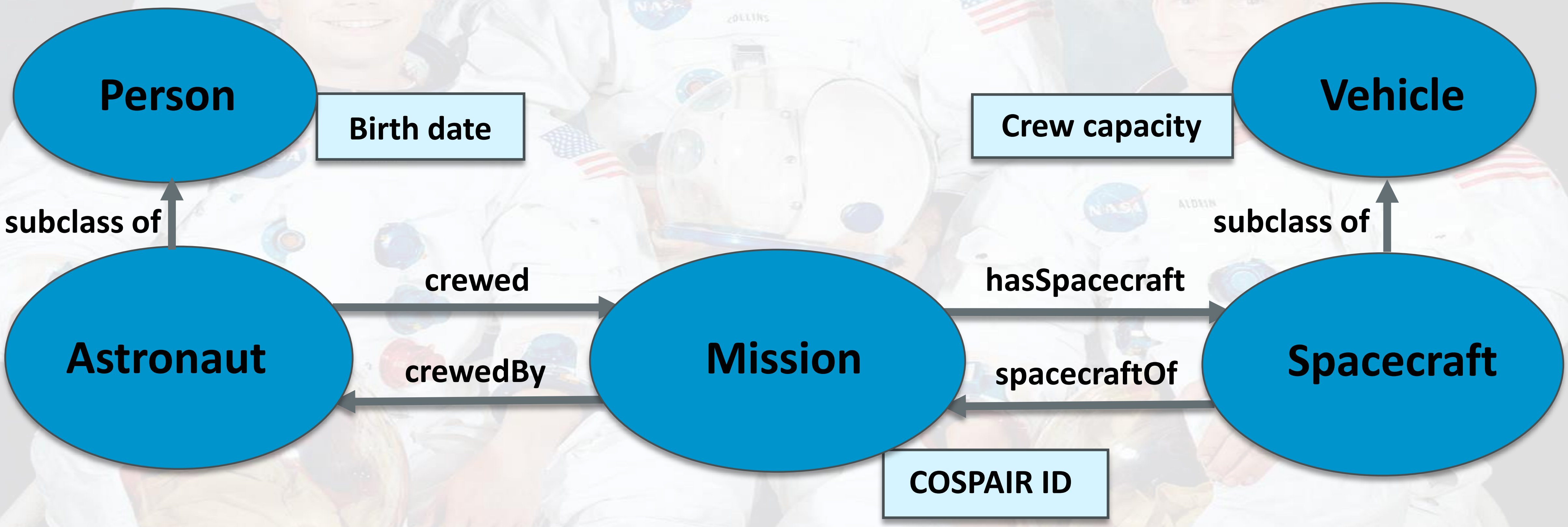
- Consider carefully if something should be an Attribute or Class (e.g. Launch vehicle).

Building Ontologies

Ontology knowledge modeling

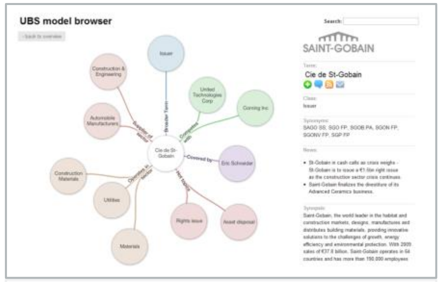
If attributes are shared, consider designating classes as sub-classes of more generic classes.

- Attributes of a class are inherited by its subclasses
- For example, birthdate applies to any “person”, not just astronauts.



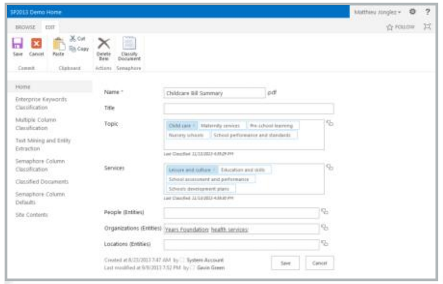
Semaphore Demonstration

Semaphore : Revealing Smarter Decisions



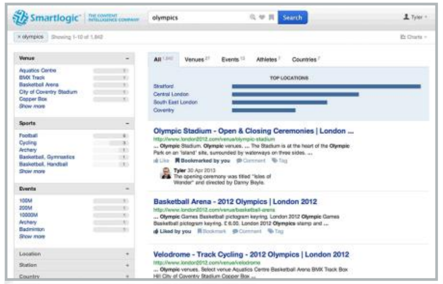
Build and manage semantic models

Simplify the ingestion, development and customization



Enrich, extract and harmonize

- Enrich information assets with complete, consistent and precise metadata
- Extract critical facts, entities and relationships for further processing
- Harmonize different information sources for unified access



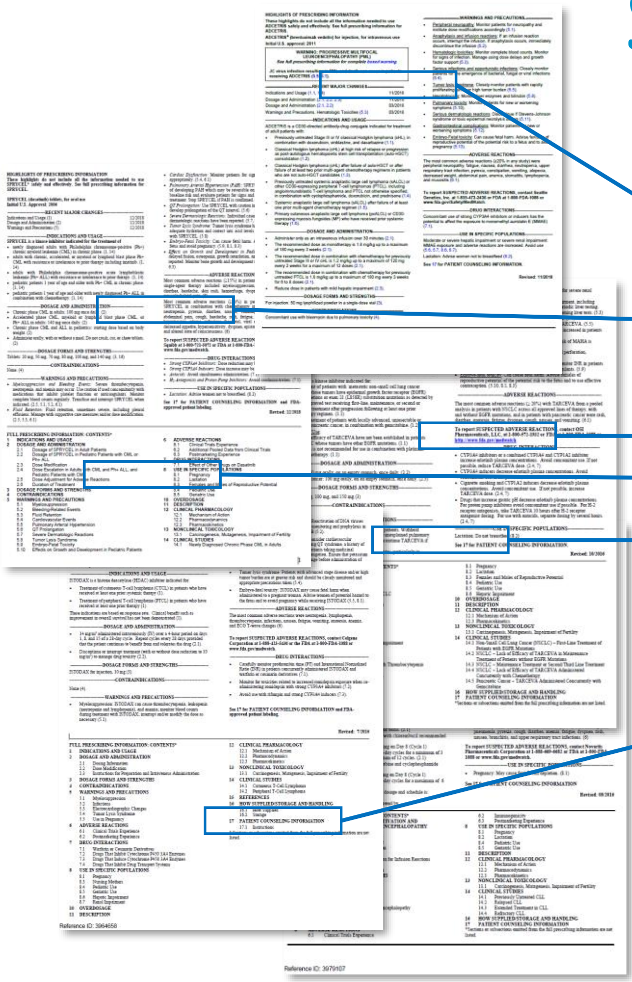
Apply semantics to your business problem

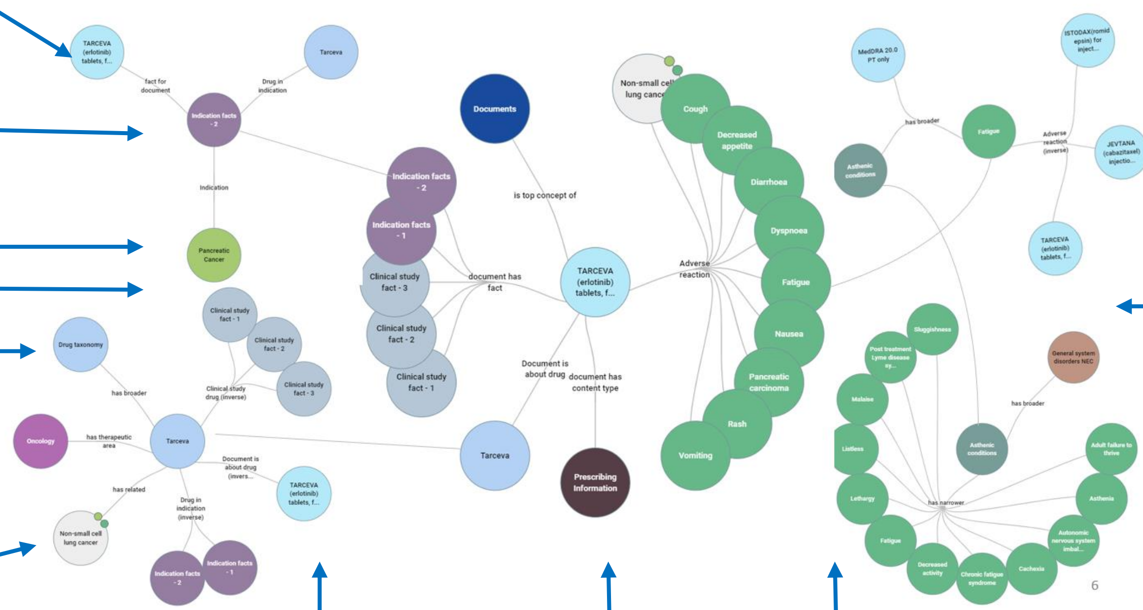
- Enable knowledge discovery
- Support investigative analytics
- Automate manual processes for higher precision

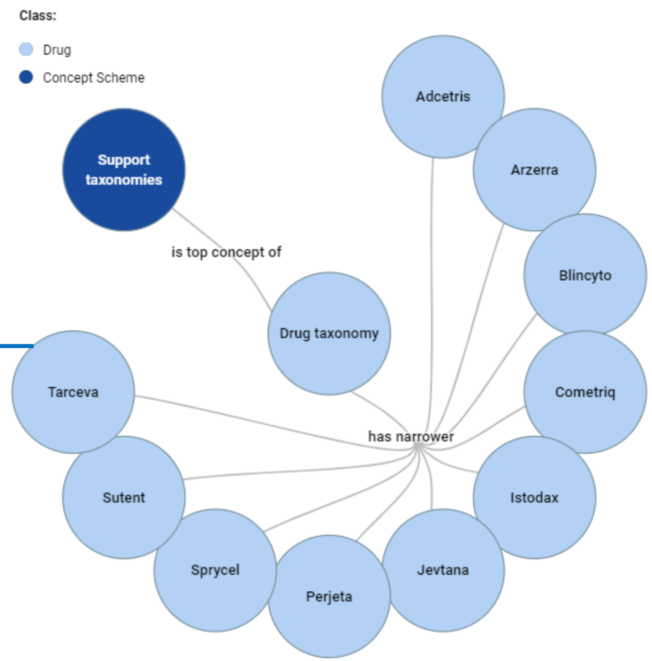
Semaphore delivers these capabilities at enterprise scale

Semaphore and Knowledge Graphs

Smartlogic™







DATA FROM DOCUMENTS

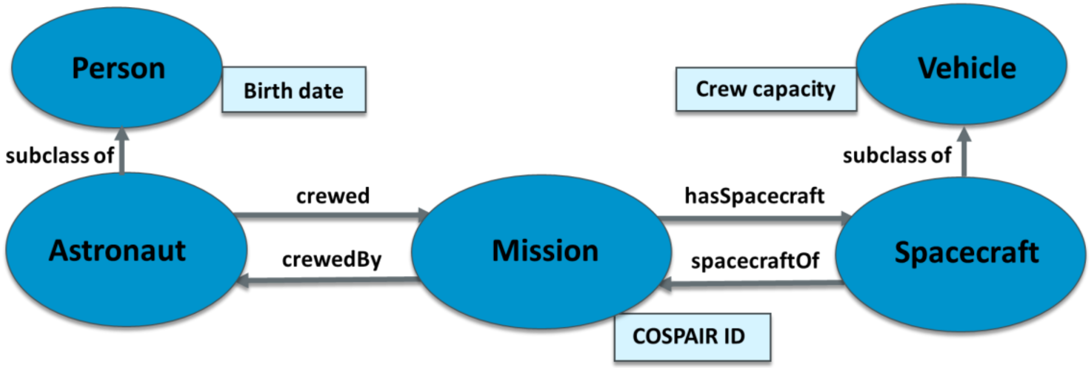
NCT Number	Study Results	Interventions	Phases	Study Type
NCT02714530	No Results Available	Drug: Erlotinib	Phase 2	Interventional
NCT02774278	Has Results	Drug: Erlotinib	Phase 2	Interventional
NCT01664533	Has Results	Drug: Erlotinib		Observational
NCT02595450	Has Results	Drug: Erlotinib		Observational
NCT01990261	Has Results	Drug: Erlotinib		Observational
NCT02013206	Has Results	Drug: erlotinib	Phase 2	Interventional
NCT01310036	Has Results	Drug: Erlotinib	Phase 2	Interventional
NCT01161173	Has Results	Drug: Erlotinib		Observational
NCT03126799	No Results Available	Drug: Erlotinib	Phase 2	Interventional
NCT02080078	No Results Available	Drug: Erlotinib	Phase 1	Interventional

ONTOLOGIES AND KNOWLEDGE MODELS

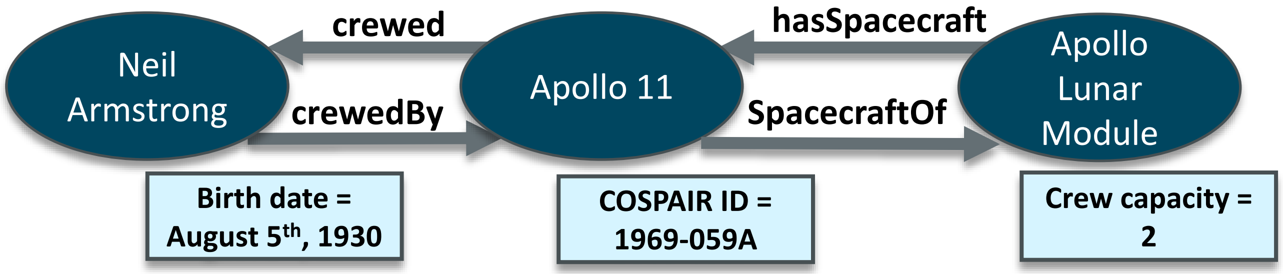
DATA FROM DATABASES

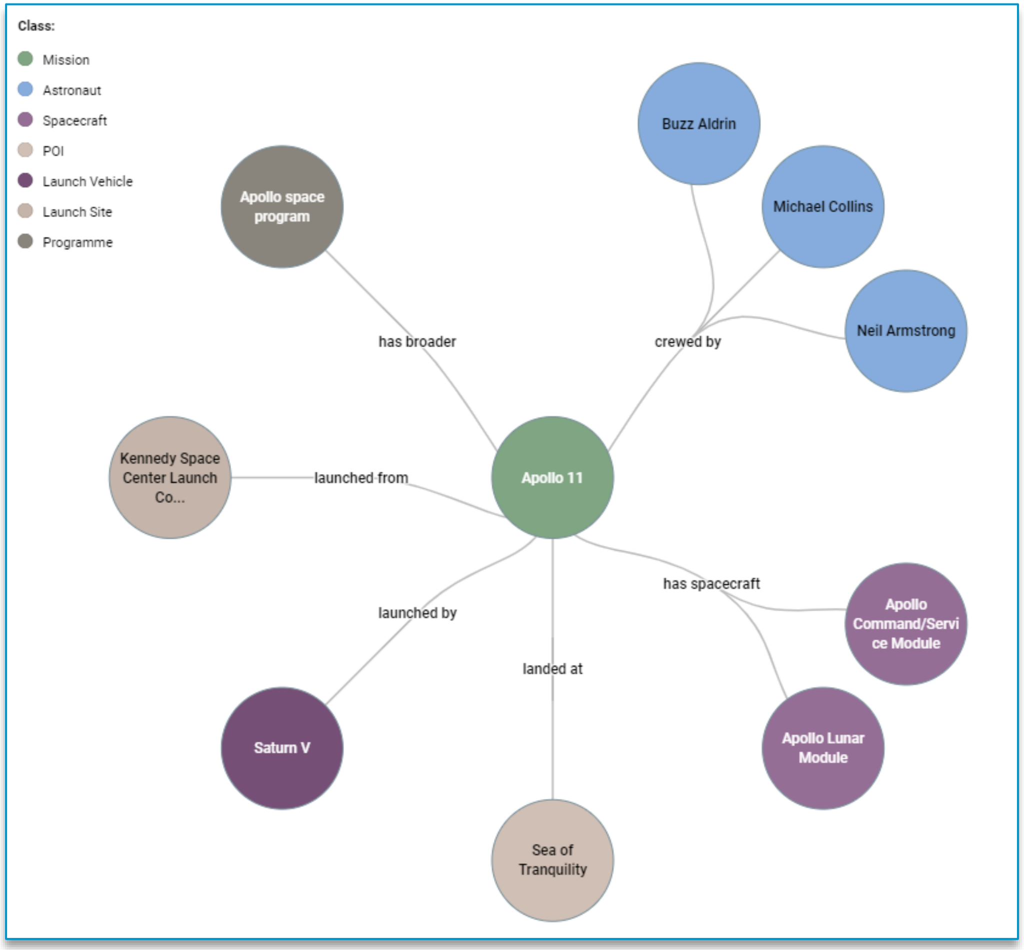
Building Ontologies in Semaphore

- Build class structure
- Add relationships and metadata to classes



- Add instance data (concepts) to classes





Ontologies Help Machines Understand Data

Smartlogic™

Q: How do we know this text is about Apollo 11?

A: Because we've modeled the evidence.

Q: What other facts are in this text?

→ This mission was the spaceflight that landed the first two humans, who were Americans, on the Moon. Mission commander Neil Armstrong and pilot Buzz Aldrin landed the lunar module Eagle on July 20, 1969, at 20:18 UTC. Armstrong became the first to step onto the lunar surface six hours later on July 21 at 02:56:15 UTC; Aldrin joined him about 20 minutes later. They spent about two and a quarter hours together outside the spacecraft, and collected 47.5 pounds (21.5 kg) of lunar material to bring back to Earth. Michael Collins piloted the command module Columbia alone in lunar orbit while they were on the Moon's surface. Armstrong and Aldrin spent just under a day on the lunar surface before rendezvousing with Columbia in lunar orbit.

→ The mission was launched by a Saturn V rocket from Kennedy Space Center in Merritt Island, Florida, on July 16, and was the fifth manned mission of NASA's Apollo program. The spacecraft had three parts: a command module (CM) with a cabin for the three astronauts, and the only part that landed back on Earth; a service module (SM), which supported the command module with propulsion, electrical power, oxygen, and water; and a lunar module (LM) that had two stages – a lower stage for landing on the Moon, and an upper stage to place the astronauts back into lunar orbit. After being sent toward the Moon by the Saturn V's upper stage, the astronauts separated the spacecraft from it and traveled for three days until they entered into lunar orbit. Armstrong and Aldrin then moved into the lunar module Eagle and landed in the Sea of Tranquility. They stayed a total of about 21.5 hours on the lunar surface. The astronauts used Eagle's upper stage to lift off from the lunar surface and rejoin Collins in the command module. They jettisoned Eagle before they performed the maneuvers that blasted them out of lunar orbit on a trajectory back to Earth. They returned to Earth and landed in the Pacific Ocean on July 24.

Sources for Ontologies

Sources for Ontologies

→ **Using existing controlled vocabularies**

- Generally, a custom-created controlled vocabulary/taxonomy is best.
  - Customized to the particular set of content
  - Customized to the particular set of users

- An existing vocabulary may be considered for a taxonomy:
  - As a starting point, to be developed and customized further
  - For indexing external content in certain subjects
  - A term list for a single metadata field or facet (filter) in a faceted taxonomy

- An existing ontology may be considered for use:
  - As a high-level knowledge model
  - As a model for a subject domain (such as biomedicine)

Sources for Ontologies

Smartlogic™

→ **Upper / foundation ontologies**

- A model for a knowledge model
- Readily available, free, no license (because they do not have specifics)
- Examples:
  - Basic Formal Ontology (BFO) http://basic-formal-ontology.org
  - gist https://www.semanticarts.com/gist/
  - SUMO (Suggested Upper Merged Ontology) http://www.adampease.org/OP/
  - BIBFRAME http://id.loc.gov/ontologies/bibframe.html
  - FOAF http://xmlns.com/foaf/spec/
  - SKOS https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html

Sources for Ontologies

Smartlogic™

SKOS

Class or Individual	*Property: Attribute*	*Property: Attribute*	*Property: Relation*	*Property: Relation*	
Concepts	Labels & Notation	Documentation	Semantic Relations	Mapping Properties	Collections
Concept	prefLabel	note	broader	broadMatch	Collection
ConceptScheme	altLabel	changeNote	narrower	narrowMatch	orderedCollection
inScheme	hiddenLabel	definition	related	relatedMatch	member
hasTopConcept	notation	editorialNote	broaderTransitive	closeMatch	memberList
topConceptOf		example	narrowerTransitive	exactMatch	
		historyNote	semanticRelation	mappingRelation	
		scopeNote			

Sources for Ontologies

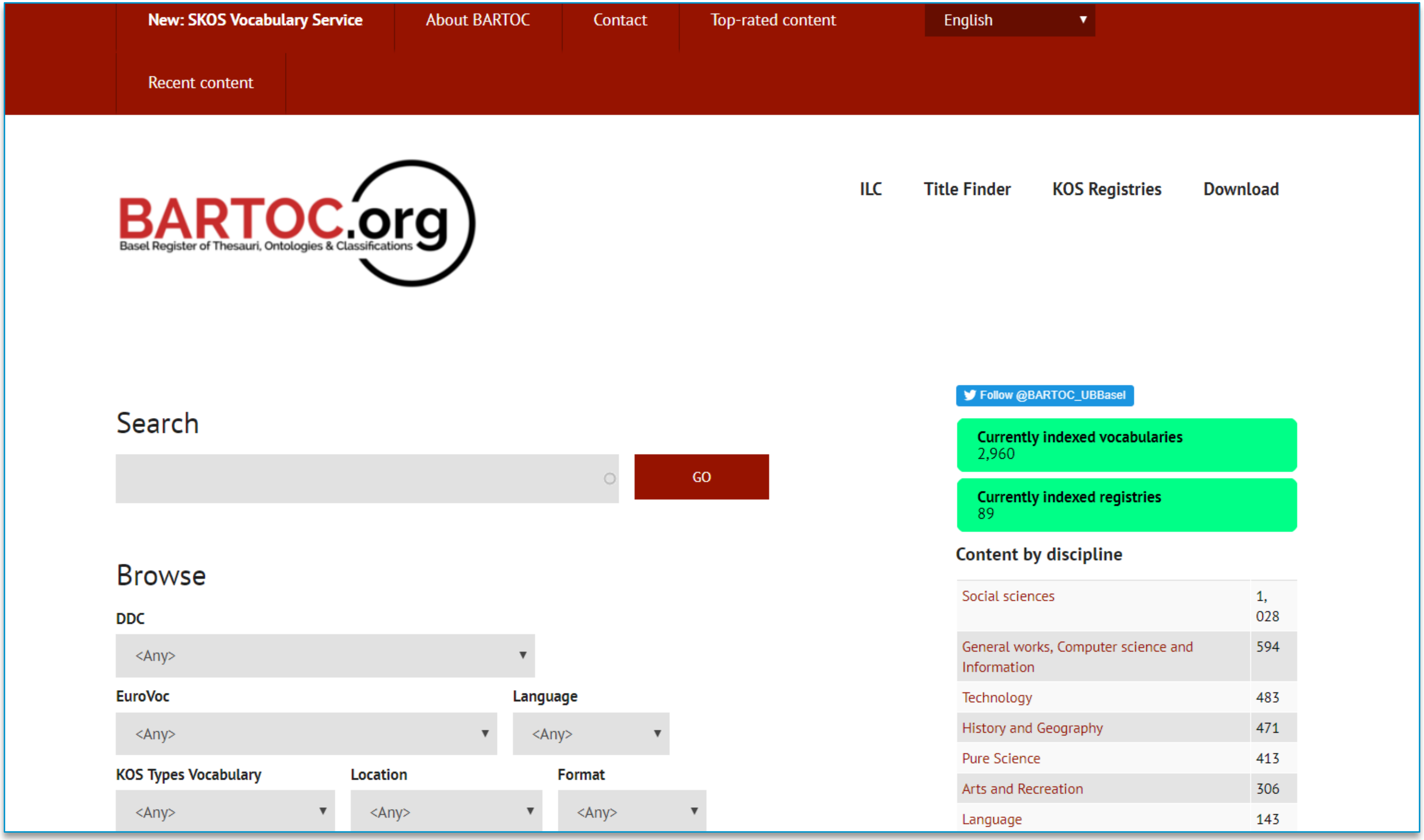
→ **Domain ontologies**

- A knowledge model customized for a subject domain
- Available for license – License types vary.
- More exist in certain domains, such as biomedical/life sciences.
- Examples:
  - Gene Ontology http://geneontology.org
  - Open Biomedical Ontologies (OBO) Foundry http://obofoundry.org
  - BBC Ontology https://www.bbc.co.uk/ontologies/bbc
  - Financial Industry Business Ontology (FIBO) https://spec.edmcouncil.org/fibo

Sources for Ontologies

→ **Directories of ontologies (and other knowledge organization systems) for license**

- Comprehensive:
  - BARTOC (Basel Register of Thesauri, Ontologies & Classification) bartoc.org
  - Research Vocabularies Australia https://vocabs.ands.org.au
  - Open Metadata Registry http://metadataregistry.org/vocabulary/list.html
  - Taxonomy Warehouse www.taxonomywarehouse.com
- Domain-specific (examples):
  - NCBO BioPortal – biomedical ontologies bioportal.bioontology.org/ontologies
  - Open Biological and Biomedical Ontology (OBO) Foundry www.obofoundry.org
  - VEST Directory of AIMS (Agricultural Information Management Standards) portal of the FAO aims.fao.org/vest-registry/vocabularies
  - Heritage Data (UK cultural heritage) www.heritagedata.org/blog/vocabularies-provided/


New: SKOS Vocabulary Service
About BARTOC
Contact
Top-rated content
English
Recent content
BARTOC.org
Basel Register of Thesauri, Ontologies & Classifications
ILC
Title Finder
KOS Registries
Download
Search
GO
Browse
DDC
<Any>
EuroVoc
<Any>
Language
<Any>
KOS Types Vocabulary
<Any>
Location
<Any>
Format
<Any>
Follow @BARTOC_UBBasel
Currently indexed vocabularies
2,960
Currently indexed registries
89
Content by discipline
Social sciences 1,028
General works, Computer science and Information 594
Technology 483
History and Geography 471
Pure Science 413
Arts and Recreation 306
Language 143

Sources for Ontologies

Ontology licensing typically does not have the issues of licensing taxonomies, thesauri, or other controlled vocabularies.

- Format is more standard (RDF)
- The ontology does not need to be as specific as a licensed taxonomy would.
- More ontologies are open source and free than are taxonomies/thesauri for license
- Ontologies are meant to be models to be modified and built out from. Free licenses for taxonomies may prohibit modification.

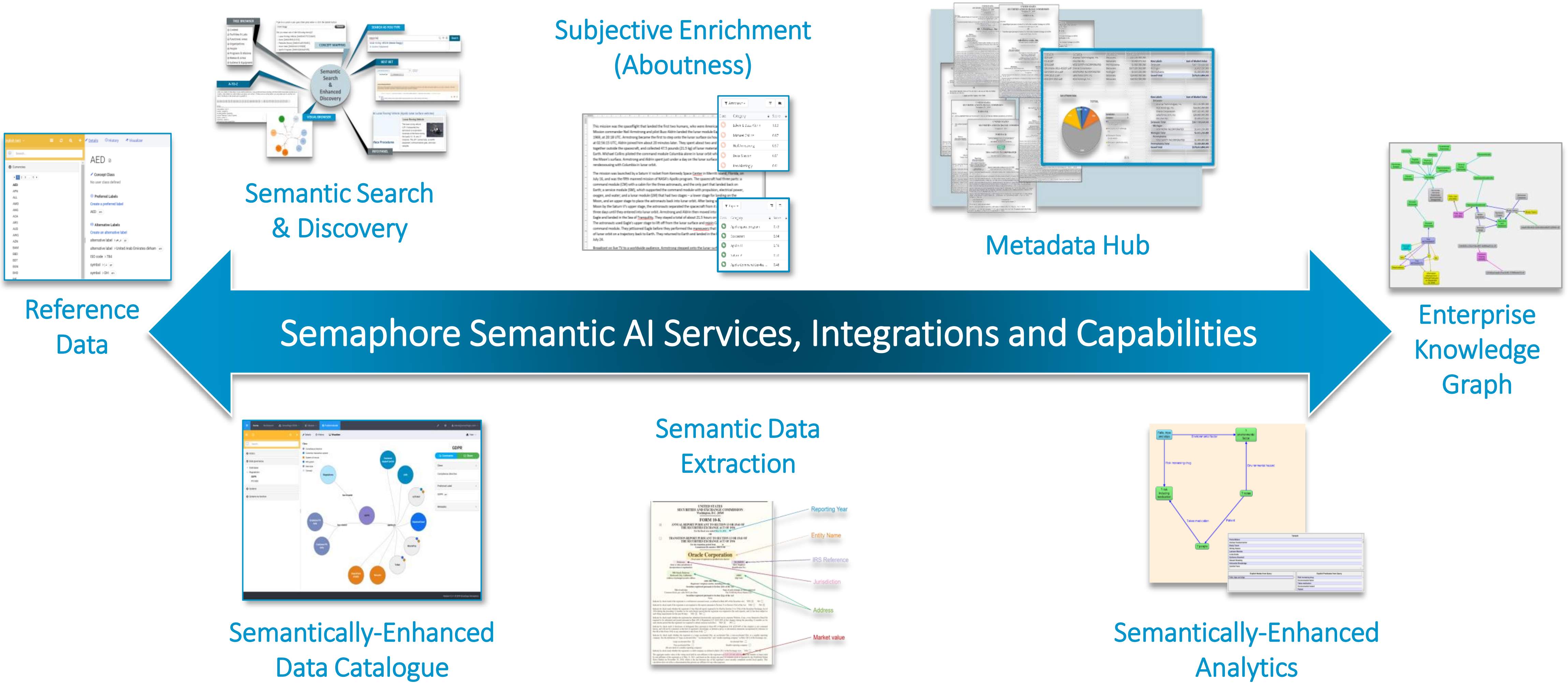
Using Ontologies

Reusing Ontologies in Semaphore

Smartlogic™

- → **All ontologies, including those using OWL, are not modeled in the same way.**
- → **While a rich source of information, an ontology's purpose may not be yours.**
- → **Planning for the reuse of an ontology:**
  - How will it be used in your organization's context?
  - With what other knowledge models will it be used?
  - What downstream systems will use the ontology?
  - How will it be governed; kept current?

Widen Scope of Data to Reveal Better Decisions



Q&A

Thank You

SMARTLOGIC – AMERICAS
111 N MARKET ST.
SAN JOSE, CALIFORNIA, 95113
TEL: +1 408 213 9500

SMARTLOGIC – EUROPE, MIDDLE-EAST AND AFRICA
200 ALDERSGATE
LONDON, EC1A 4HD
TEL: +44 203 176 4500

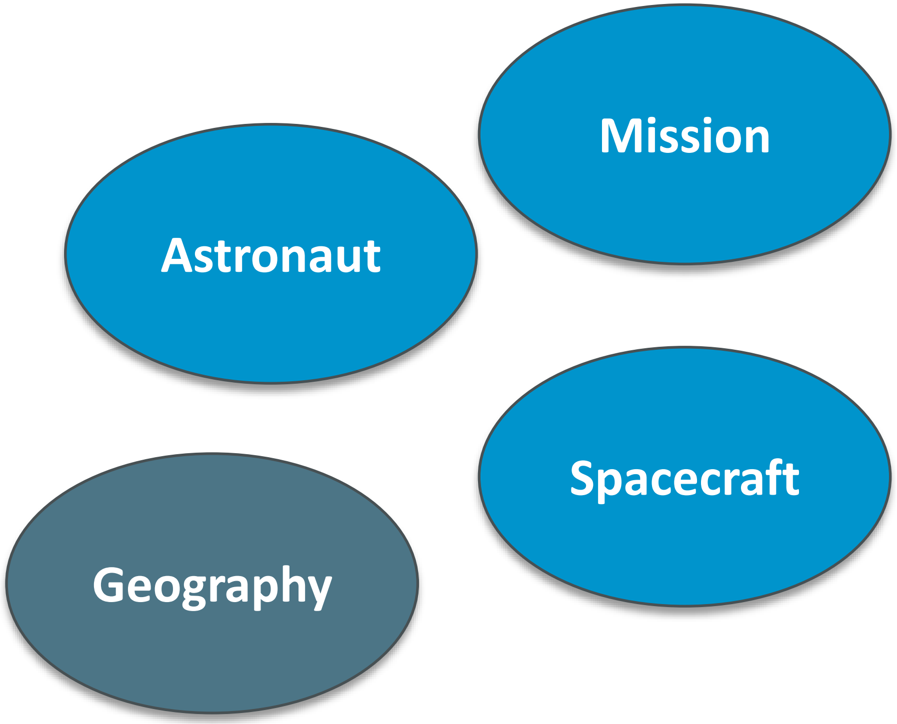
WWW.SMARTLOGIC.COM
INFO@SMARTLOGIC.COM

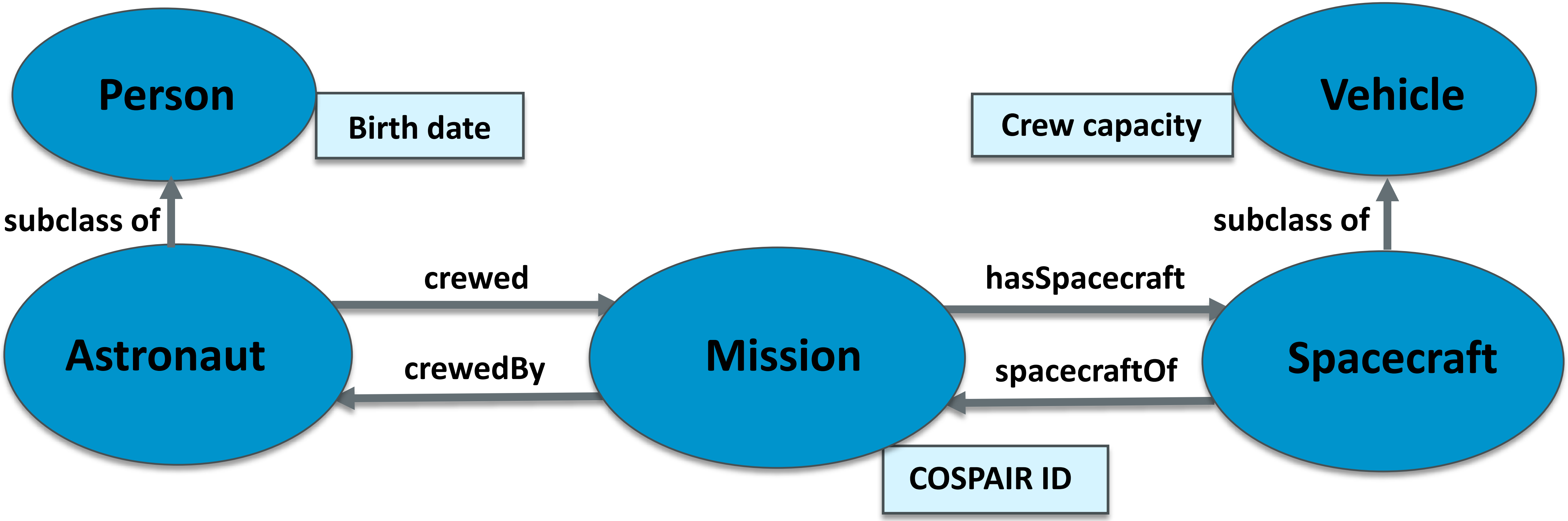
Heather Hedden – Taxonomy Consultant and Trainer | **Hedden Information Management**

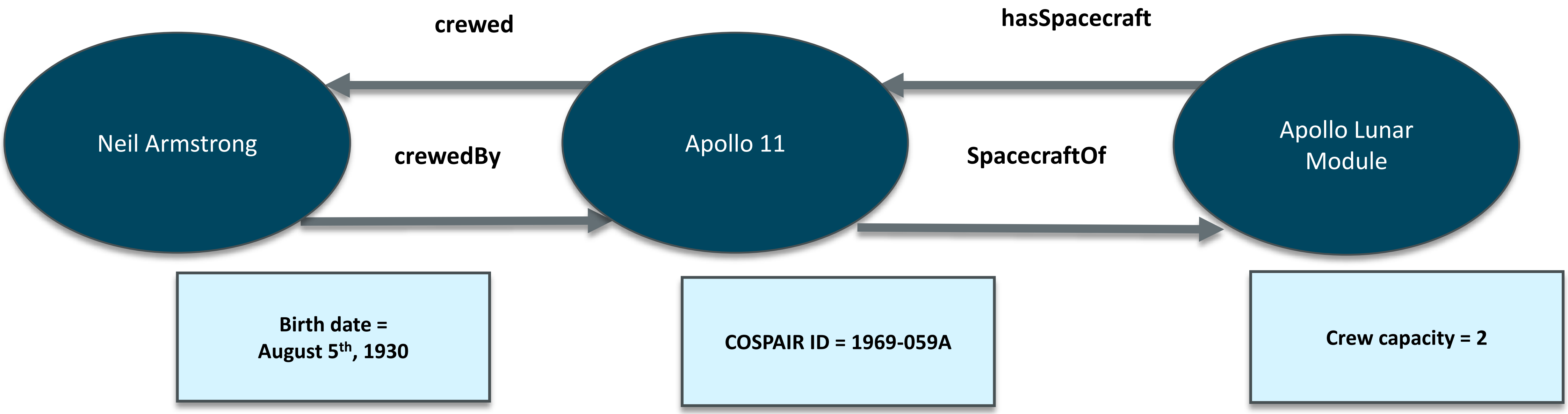
heather@hedden.net
www.hedden-information.com
www.linkedin.com/in/hedden
Twitter: @hhedden

Extra slides (take out before final version)

Ontologies as component of knowledge graphs




Person
Birth date
Vehicle
Crew capacity
subclass of
subclass of
crewed
hasSpacecraft
Astronaut
Mission
Spacecraft
crewedBy
spacecraftOf
COSPAIR ID


crewed
hasSpacecraft
Neil Armstrong
crewedBy
Apollo 11
SpacecraftOf
Apollo Lunar Module
Birth date = August 5th, 1930
COSPAIR ID = 1969-059A
Crew capacity = 2

Reusing Ontologies in Semaphore