Do More With Less: Semantic Technologies

Colin Bell

Director, Enterprise Architecture

Overview

- Data Management Challenges
  - Classic Master Data Management (MDM)
  - Data Warehousing w/ ETL
- Semantic Technologies Primer
- Big Data with Semantics
- Customer Relationship Management (CRM)
  - Student Retention
- Value Proposition

Data Management at Waterloo: Challenges

- Data exists today in a number of silos.
- Every system has its own domain specific models:
  - HRMS knows about ‘Employees’
  - SIS knows about ‘Students’
  - WatIAM knows about ‘Identities’
- But they all actually ‘know’ (meaning) about the same things- **Individuals**.
- Systems don’t understand that, humans need to explicitly form those links.

Data Management at Waterloo: Challenges

- Traditional Approach is to work towards **'Master Data Management' (MDM)**
  - Steadily work through all the silos with the business to ensure that there is one master definition for all the data in an organization.
  - Summarize it all down into one 'Master File' that the organization can refer to.
  - It helps, but it is still fraught with risk. Poor-quality data is still is a major risk. Garbage in, garbage out.
  - "Big Data" (although we don't really have it), risks data quality issues at scale.

Data Management at Waterloo: Today

- ETL and build Data Marts + Data Warehouses
  - Extract
  - Transform
  - Load
- If something does not make sense and a consistency check misses it, bad data is silently introduced to your warehousing.
- Quality of your decision support service drops.

Semantic Technologies

Semantics

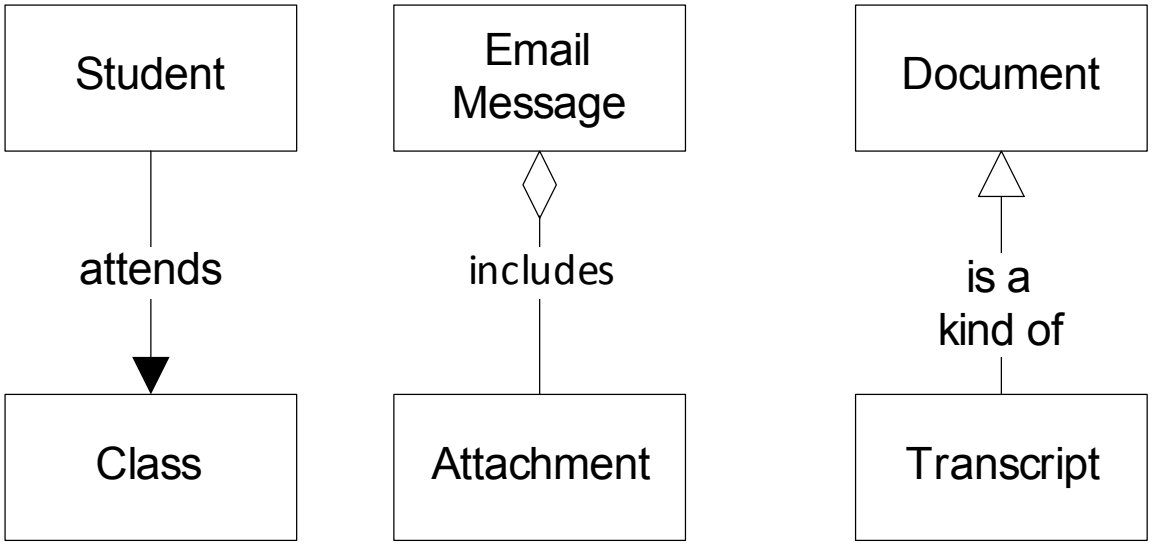
The branch of linguistics or philosophy concerned with meaning in language; the study of analysis of meaning in words, sentences, etc.

http://bit.ly/1Qeh7qA

Semantic Technologies

- While the Web (and Web 2.0) focused on structure of information and links it. Humans are required to understand the connections.
  - Example: Wikipedia page for “Dog” might link to “Wolf” but only as linked pages, not as related Species.
- Enter Web 3.0 -> The “Semantic Web”
- **Goal:** encode more than links between data. Capture meaning in a way that systems can understand.

Semantic Models Basics

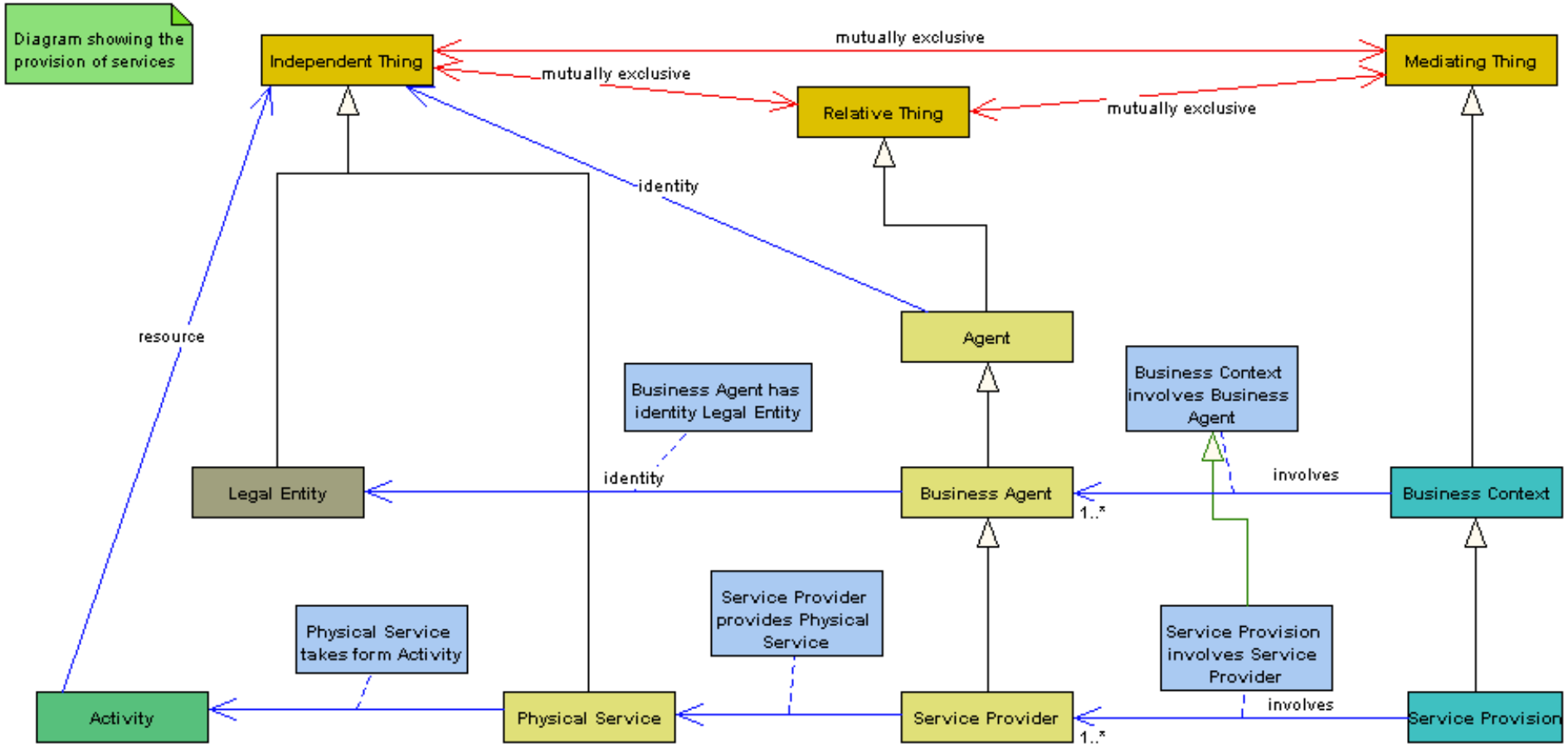


Nouns + Verb Relationship

Inclusion

Subclass / Inheritance

Example: Conceptual Semantic Model

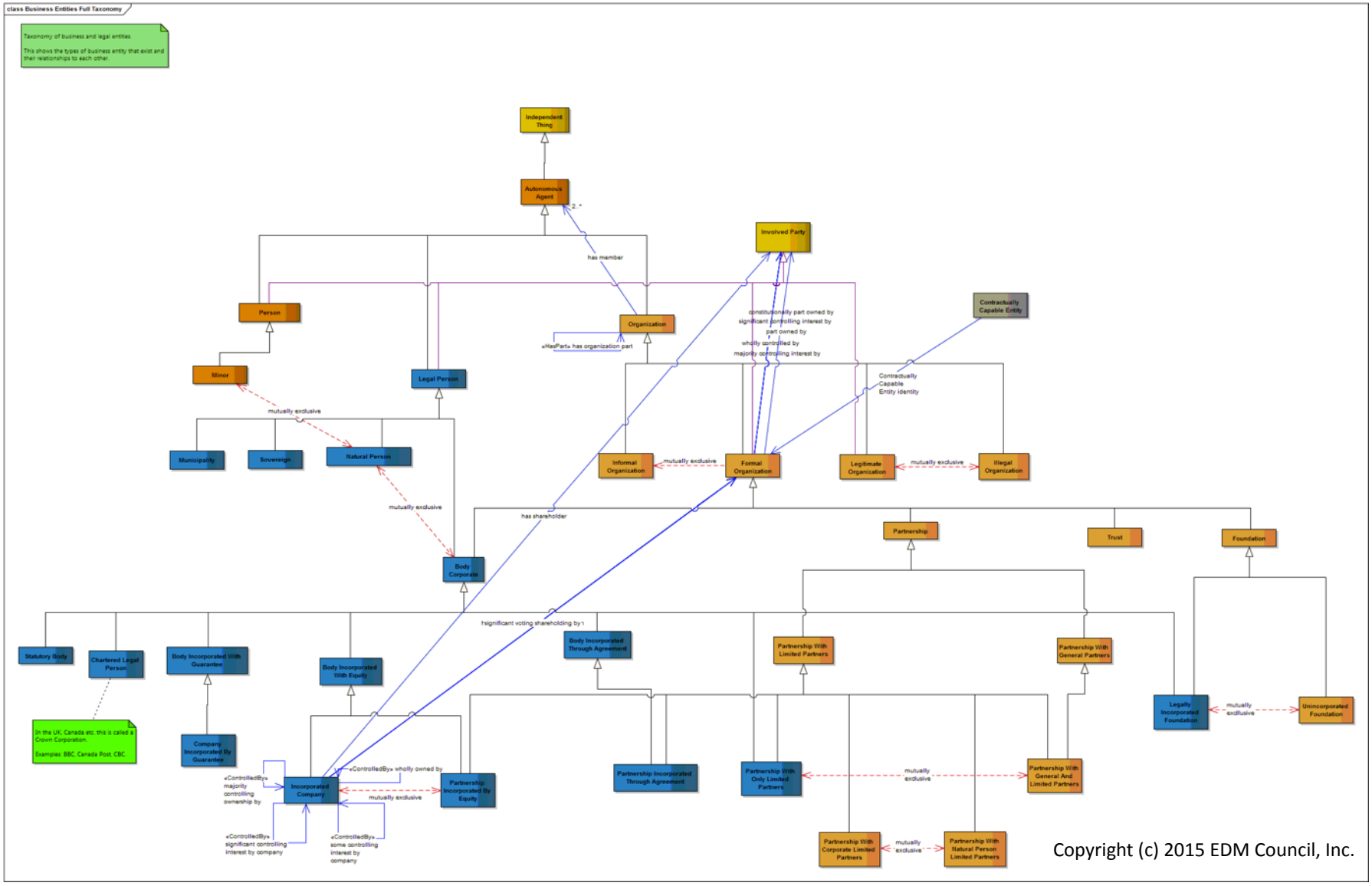


Semantic Technologies

W3C Standards. Built on top of other standards (XML, NS, UNICODE, etc.):

- **RDF (Resource Description Framework)**
  - Data modelling language for the Semantic Web.
- **OWL (Web Ontology Language)**
  - The schema / knowledge representation language of the Semantic Web.

Example: Operational Semantic Model



Example: RDF

```
<rdf:RDF
     xml:base="http://spec.edmcouncil.org/fibo/BE/AboutBE/"
     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
     xmlns:owl="http://www.w3.org/2002/07/owl#"
     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
     xmlns:dct="http://purl.org/dc/terms/"
     xmlns:skos="http://www.w3.org/2004/02/skos/core#"
     xmlns:sm="http://www.omg.org/techprocess/ab/SpecificationMetadata/"
     xmlns:fibo-fnd="“http://spec.edmcouncil.org/fibo/BE/AboutBE/">

<!-- Body of RDF goes here.  In this example, OWL goes here. --!>

</rdf:RDF>
```

Example: OWL

```
<owl:Class rdf:about="&fibo-fnd-aap-agt;AutonomousAgent">
        <rdfs:label>autonomous agent</rdfs:label>
        <rdfs:subClassOf>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&fibo-fnd-aap-agt;isIdentifiedBy"/>
                <owl:minCardinality rdf:datatype="&xsd;nonNegativeInteger">0</owl:minCardinality>
            </owl:Restriction>
        </rdfs:subClassOf>
        <rdfs:subClassOf>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&fibo-fnd-aap-agt;hasName"/>
                <owl:minQualifiedCardinality rdf:datatype="&xsd;nonNegativeInteger">0</owl:minQualifiedCardinality>
                <owl:onDataRange rdf:resource="&fibo-fnd-utl-bt;text"/>
            </owl:Restriction>
        </rdfs:subClassOf>
          <skos:definition rdf:datatype="&xsd;string">An agent is an autonomous individual that can adapt to and interact with its
environment.</skos:definition>
        <fibo-fnd-utl-av:explanatoryNote rdf:datatype="&xsd;string">Note that this does not necessarily imply that an agent is
free to act as it sees fit, without constraint. Rather, an autonomous thing in the sense meant here is something which may or may
not be subject to controls and constraints but is self-actualizing in its behavior in response to any such constraints. Autonomous
things may include human beings, organizations, software agents, robots and animals.</fibo-fnd-utl-av:explanatoryNote>
        <rdfs:seeAlso rdf:datatype="&xsd;anyURI">http://www.jamesodell.com/WhatIsAnAgent.pdf</rdfs:seeAlso>
        <rdfs:seeAlso rdf:datatype="&xsd;anyURI">http://www.jamesodell.com/WhyShouldWeCareAboutAgents.pdf</rdfs:seeAlso>
        <sm:relatedSpecification rdf:datatype="&xsd;anyURI">http://www.omg.org/spec/SoaML/</sm:relatedSpecification>
        <sm:directSource rdf:datatype="&xsd;anyURI">http://www.omg.org/techprocess/meetings/schedule/AMP.html</sm:directSource>
    </owl:Class>
```

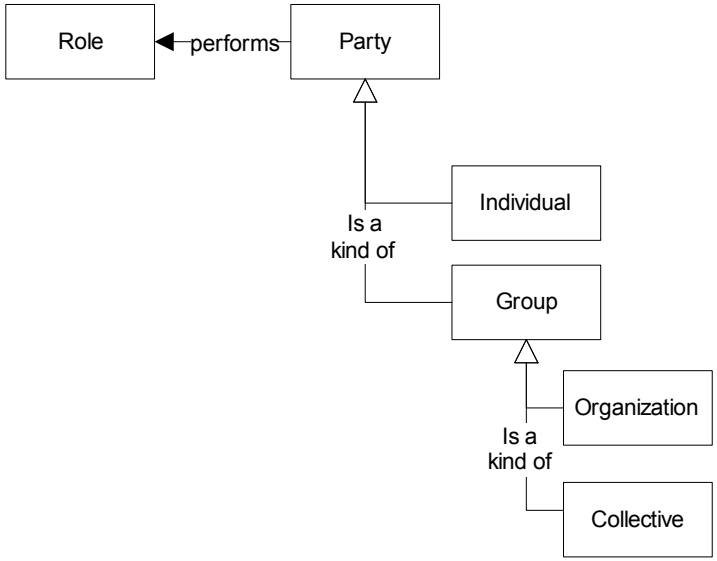
Big Data with Semantics

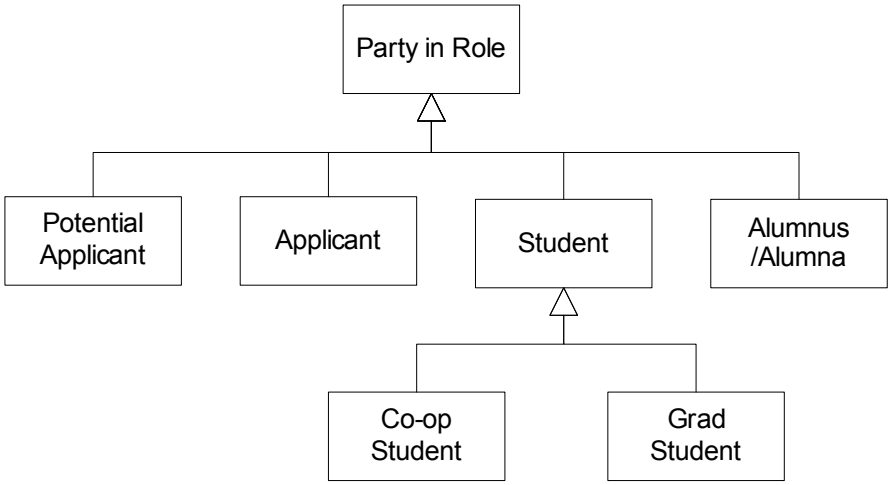
- Diagram a business readable representation of the entities, relationships, and their meanings. Use this as a **rosetta stone** when designing, implementing, and working with customers.

Then…

- Define a machine readable representation of the entities, relationships, and their meanings.
- Machine learning algorithms and expert systems become possible.

Customer Relationship Management (CRM): Student Retention Management

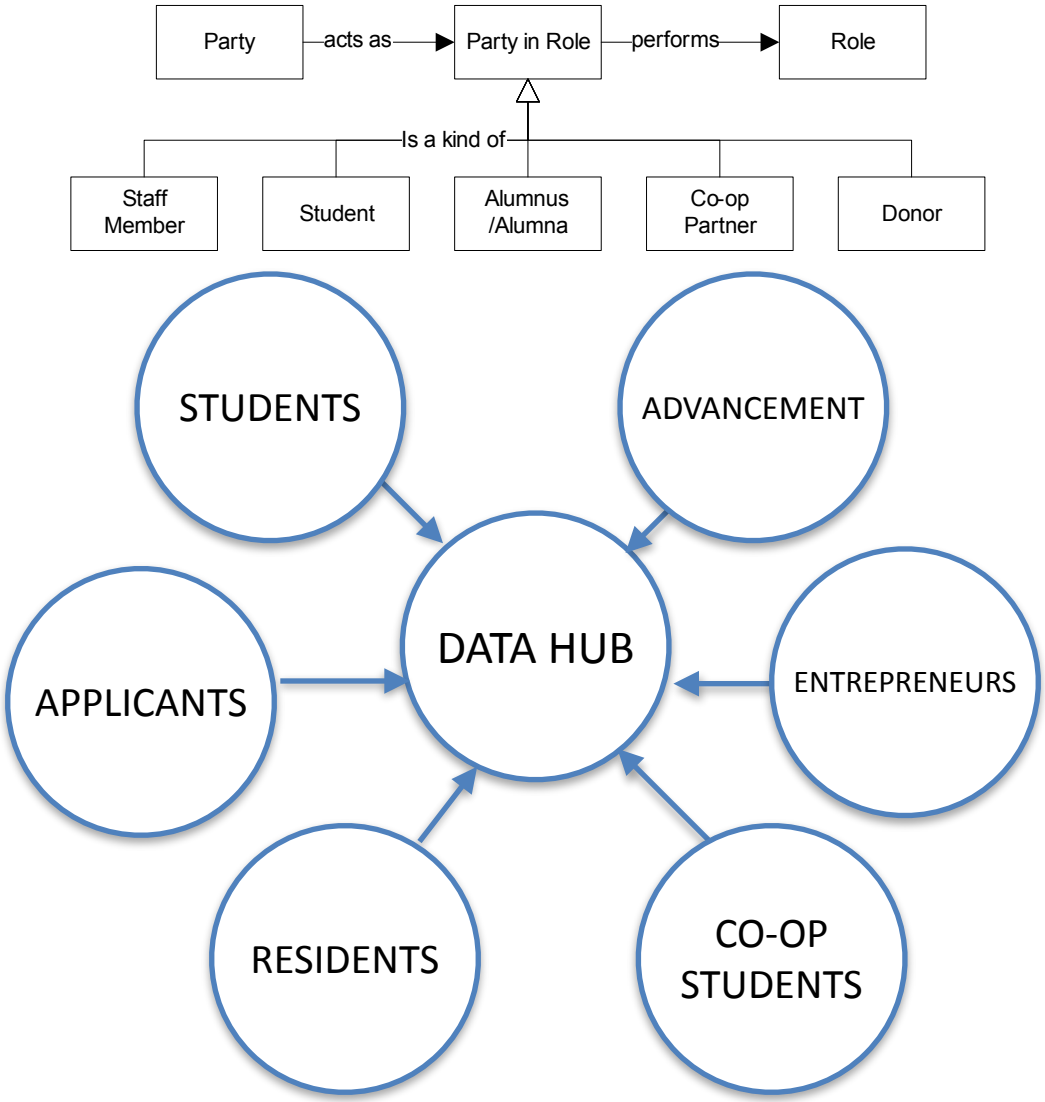




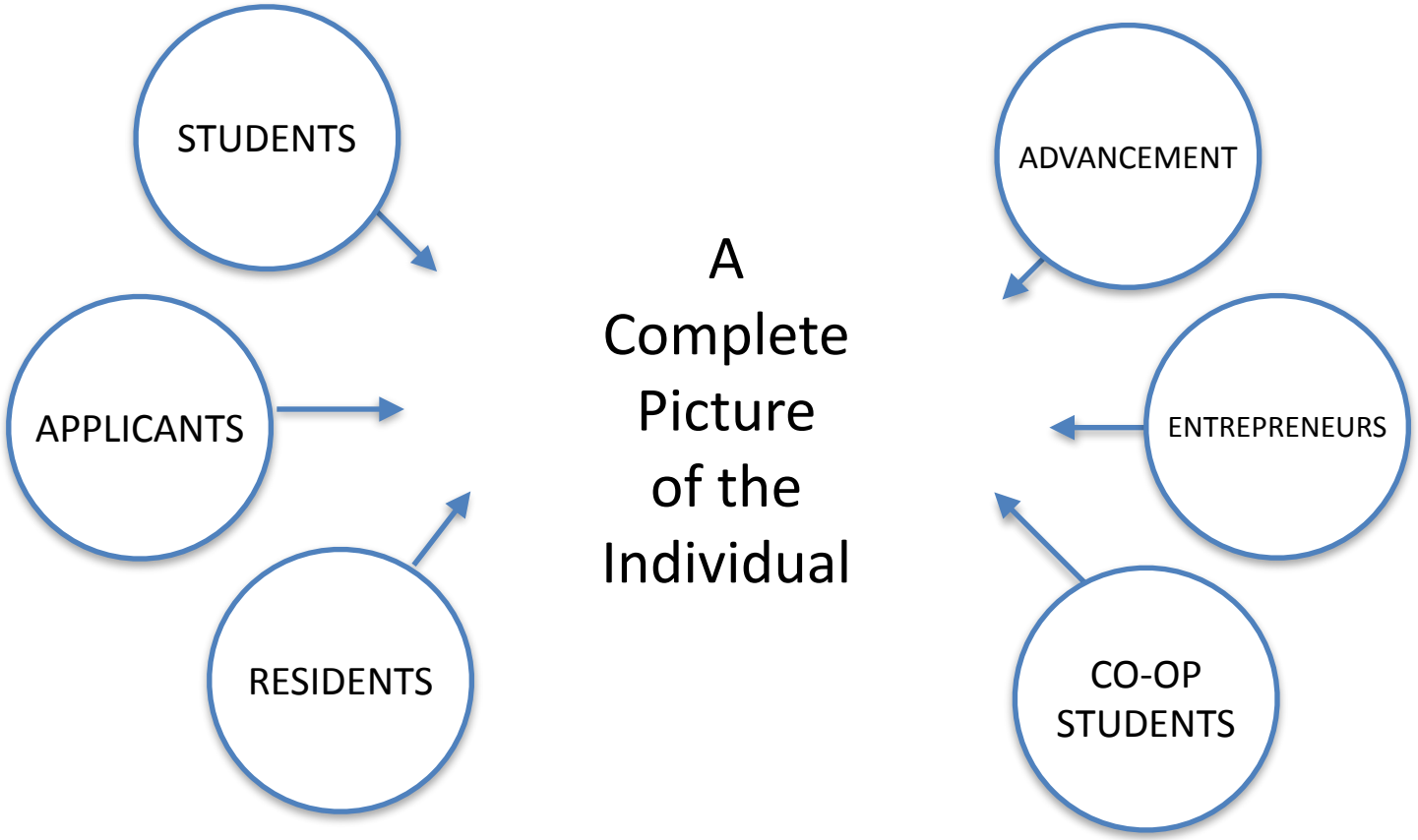
A 360 degree view of the 'Student Customer' requires that we understand that Parties can play multiple roles.

Given the federated nature of system investment at Waterloo, multiple systems are used to capture information about Parties in different Roles.

Customer Relationship Management (CRM): Student Retention Management



Customer Relationship Management (CRM): Student Retention Management



Value Proposition

- **Indirect Value-Add**
  - *Conceptual Semantic Models*
    - Powerful communication tool (rosetta stone).
    - Improve speed to deliver on ETL / Data Warehousing efforts. Inform MDM efforts.
  - *Operational Semantic Models*
    - Use machine learning technologies in conjunction with semantic technologies to improve data quality.
- **Eliminate Waste**
  - Data driven decision making quality improves with higher quality data.

Questions / Comments?