Ontology and Taxonomy: Strange Bedfellows

by Michael Uschold
Semantic Arts
www.semanticarts.com

Keynote Talk
International Conference on Semantic Computing
June 16, 2014
Newport Beach, CA

The Situation

- Knowledge assets in large enterprises are very complex
- It got that way for many reasons

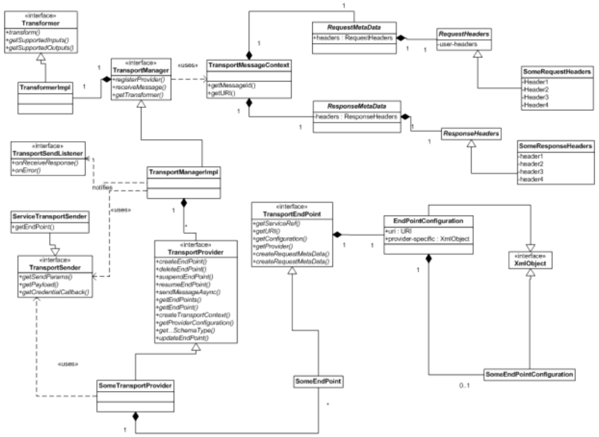
Mega Corp

Roots of Complexity

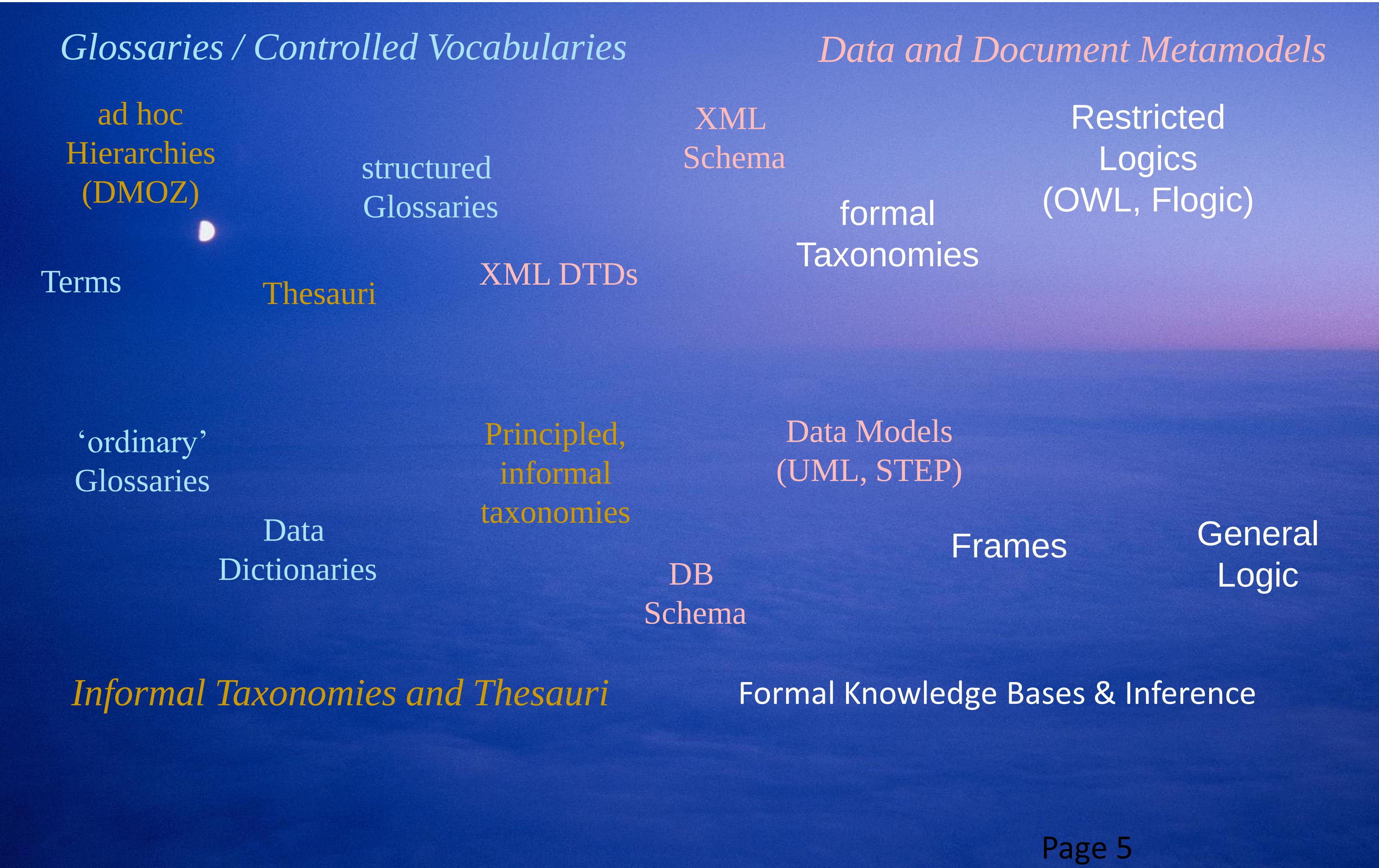
- Ambiguity is pervasive
- Systems are developed independently
- One database for each application
- Lots of metadata but:
  - No reuse of data models
  - Heterogeneity reigns supreme
- Lets look at Mega Corp

Themes at Mega Corp

- People would lament the growing complexity of their information systems
- But their focus was on short term results
- They realized they needed some good models, ideally one model to rule them all…
- But they kept acquiring more companies
- “Let’s not re-invent the wheel” led to more models (and more wheels)



Many Modeling Structures



Glossaries / Controlled Vocabularies

Data and Document Metamodels

ad hoc
Hierarchies
(DMOZ)

structured
Glossaries

XML
Schema

Restricted
Logics
(OWL, Flogic)

formal
Taxonomies

Terms

Thesauri

XML DTDs

'ordinary'
Glossaries

Principled,
informal
taxonomies

Data Models
(UML, STEP)

Data
Dictionaries

Frames

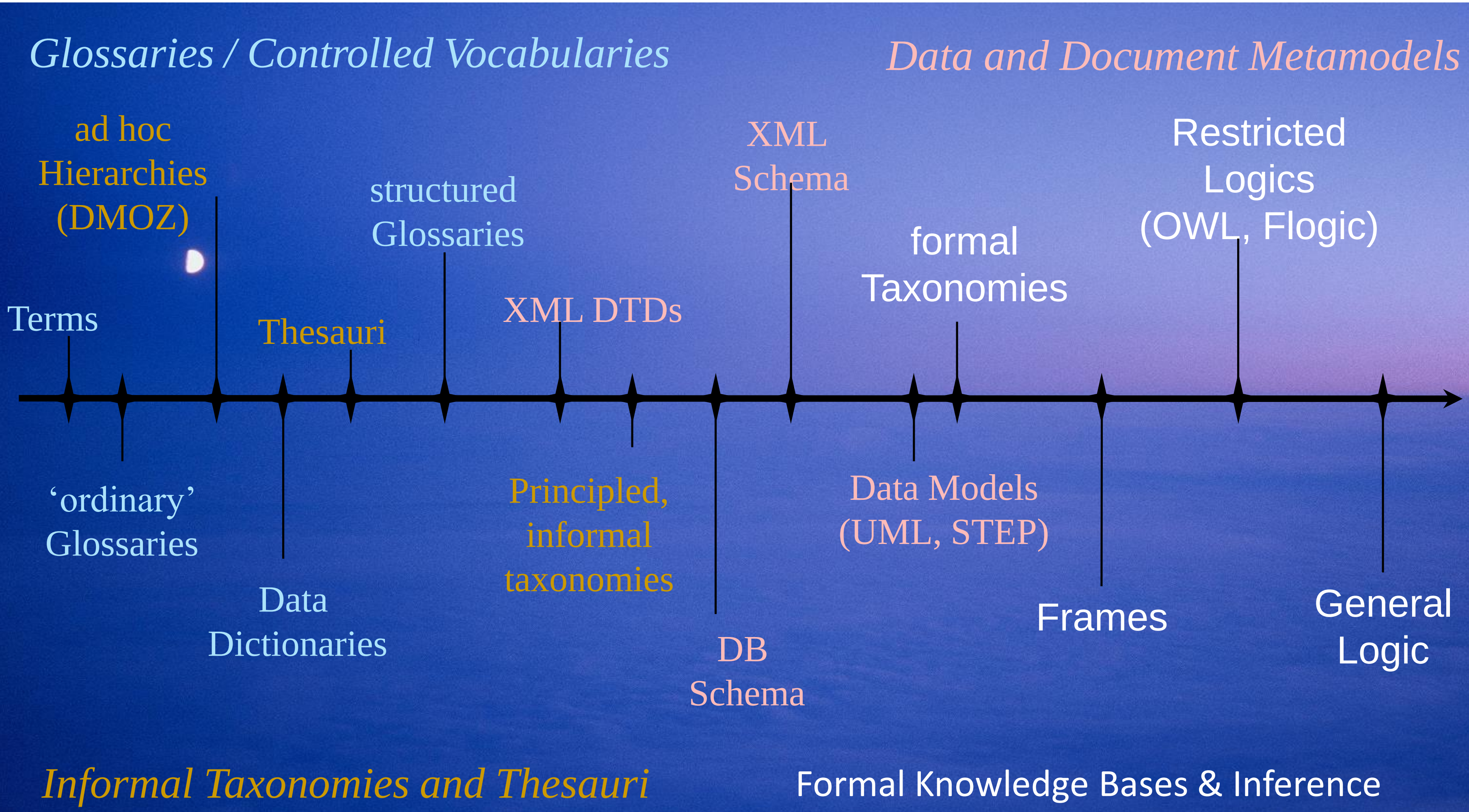
General
Logic

DB
Schema

Informal Taxonomies and Thesauri

Formal Knowledge Bases & Inference

Many Modeling Structures



Glossaries / Controlled Vocabularies

Data and Document Metamodels

Terms

'ordinary' Glossaries

ad hoc Hierarchies (DMOZ)

Data Dictionaries

Thesauri

structured Glossaries

XML DTDs

Principled, informal taxonomies

DB Schema

XML Schema

Data Models (UML, STEP)

formal Taxonomies

Frames

Restricted Logics (OWL, Flogic)

General Logic

Informal Taxonomies and Thesauri

Formal Knowledge Bases & Inference

Many Tools and Approaches: Informal to Formal

Tools and Approaches

- There are many tools
  - Spreadsheets, Spreadsheets & more Spreadsheets
  - Vocabulary managers
  - Indexing and search
  - XML editors
  - ER modeling
  - Taxonomy and Ontology tools
- Reuse and sharing is next to impossible
- These tools and approaches usually mix like oil & water

Oil & Water

- Different reasons for organizing knowledge
- Different cultures: both technological & social
- Different levels of formality (neats/scruffies)
- Conceptual vs. Design vs. Implementation
- Governance: who gets to control what?
- The menu vs. the meal

The Menu vs. the Meal

Taxonomy and Thesauri:

- focus is on words not concepts (the menu)
- relationships are between terms: synonym, hyponym, broader/narrower term
- each term should refer to just one concept

Don’t eat the menu…

Ontology:

- focus is on concepts (the meal)
- relationships are between concepts
- formal definitions
- automated inference

Eat the meal

Holy Grail: Bring It All Together

- Understand where each approach adds the most value
- Find the touch points and link them all up
- Can everyone and every tool can live in harmony?
- An impossible dream?
- We are pushing hard on this
  It's getting a lot less impossible
- Lets look at a Case Study at Mega Corp

Case Study

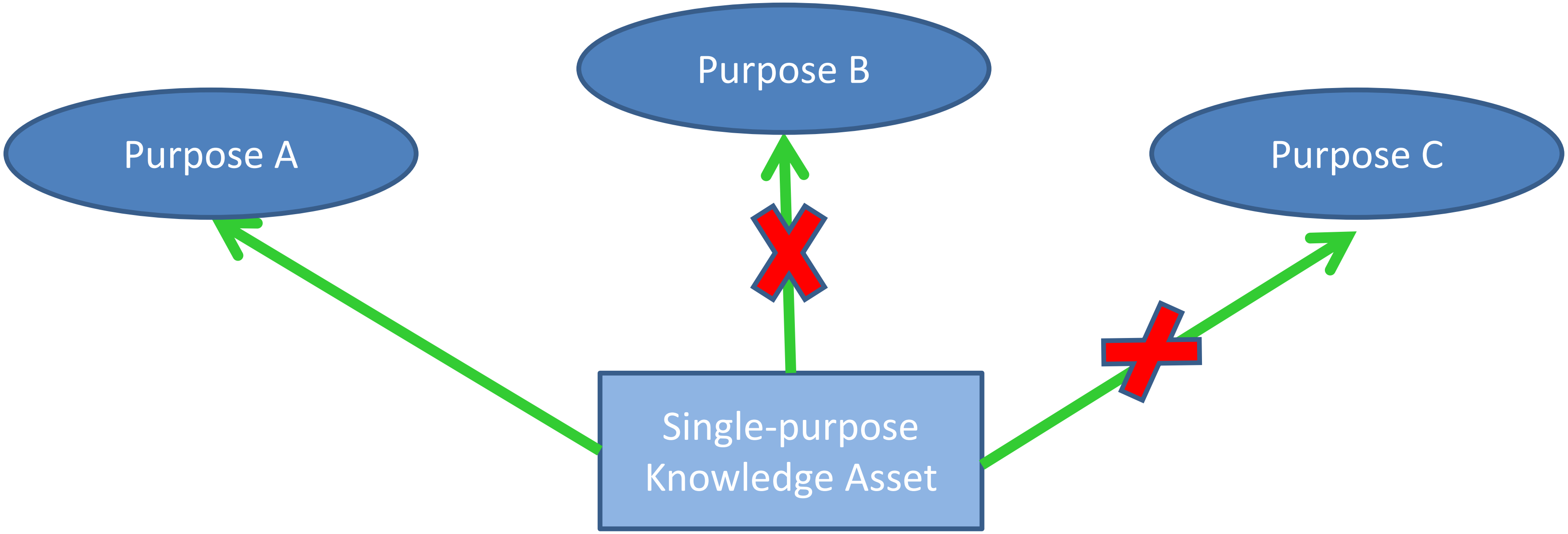
- A certain kind of thing needs to be managed
- There are millions of them, and 1000s of new ones arrive every day
- They wanted to track these items and see which ones were having what impact where (purpose A).
- So they created many hundreds of "buckets" which they would use to classify the items.
- It must be possible to classify every item into exactly one bucket.

Case Study

- The set of buckets were defined and enshrined in a spreadsheet where each row represented a bucket and there were two main columns:
  - Name of the bucket
  - Text description of the bucket
- And they saw that it was good (for their purpose)

Others Noticed

- There were some other groups that managed similar items.
- They went ahead and tried to use those same buckets for their different purposes.



Reuse not so Easy: Why?

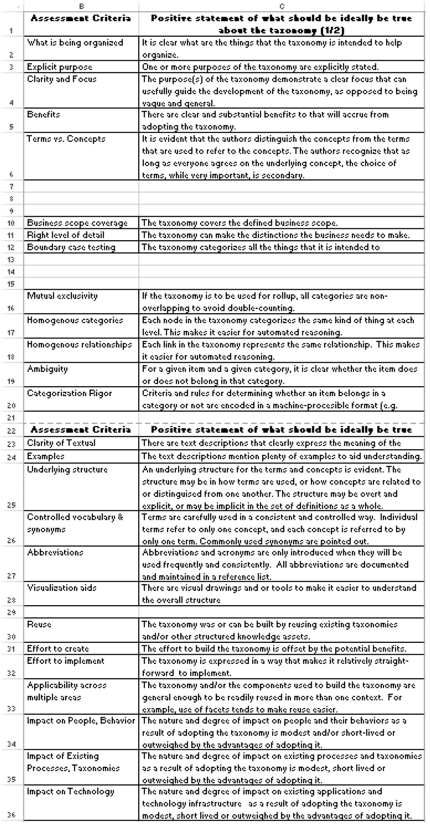
- Despite strong similarity in the underlying items for all groups, there were large differences in how they managed the items
- This was maddening:
  Similarity so near, yet Reuse so far away
- Much head-scratching ensued
- How to get to the bottom of this?

Lets Walk before we Run

- While it is important that the asset become reusable across different groups.
- We wanted to first look carefully at that first asset purely in the context of its original purpose.
- Then we will get back to reuse.

Gather Background Information

- Subject and Scope of what is being organized
  - – what is known to be out of scope?
  - – criteria for deciding in/out of scope?
- Intended audience, purpose & current uses
- Notation: syntax and semantics
- Provenance

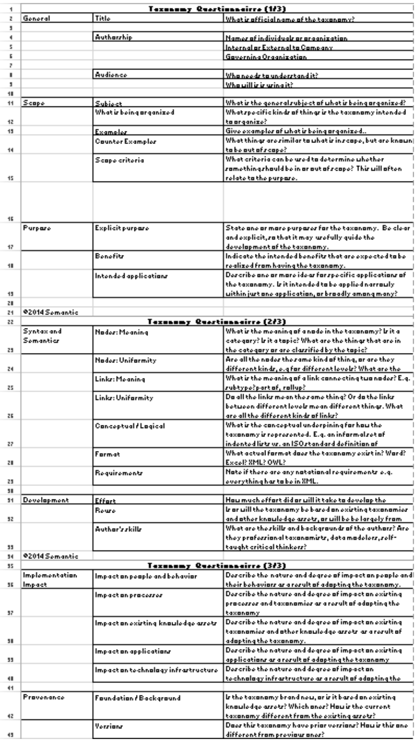


	B	C
1	**Assessment Criteria**	**Positive statement of what should be ideally be true about the taxonomy (1/2)**
2	What is being organized	It is clear what are the things that the taxonomy is intended to help organize.
3	Explicit purpose	One or more purposes of the taxonomy are explicitly stated.
4	Clarity and Focus	The purpose(s) of the taxonomy demonstrate a clear focus that can usefully guide the development of the taxonomy, as opposed to being vague and general.
5	Benefits	There are clear and substantial benefits to that will accrue from adopting the taxonomy.
6	Terms vs. Concepts	It is evident that the authors distinguish the concepts from the terms that are used to refer to the concepts. The authors recognize that as long as everyone agrees on the underlying concept, the choice of terms, while very important, is secondary.
7		
8		
9		
10	Business scope coverage	The taxonomy covers the defined business scope.
11	Right level of detail	The taxonomy can make the distinctions the business needs to make.
12	Boundary case testing	The taxonomy categorizes all the things that it is intended to
13		
14		
15		
16	Mutual exclusivity	If the taxonomy is to be used for rollup, all categories are non-overlapping to avoid double-counting.
17	Homogenous categories	Each node in the taxonomy categorizes the same kind of thing at each level. This makes it easier for automated reasoning.
18	Homogenous relationships	Each link in the taxonomy represents the same relationship. This makes it easier for automated reasoning.
19	Ambiguity	For a given item and a given category, it is clear whether the item does or does not belong in that category.
20	Categorization Rigor	Criteria and rules for determining whether an item belongs in a category or not are encoded in a machine-procesible format (e.g.
21		
22	**Assessment Criteria**	**Positive statement of what should be ideally be true**
23	Clarity of Textual	There are text descriptions that clearly express the meaning of the
24	Examples	The text descriptions mention plenty of examples to aid understanding.
25	Underlying structure	An underlying structure for the terms and concepts is evident. The structure may be in how terms are used, or how concepts are related to or distinguised from one another. The structure may be overt and explicit, or may be implicit in the set of definitions as a whole.
26	Controlled vocabulary & synonyms	Terms are carefully used in a consistent and controlled way. Individual terms refer to only one concept, and each concept is referred to by only one term. Commonly used synonyms are pointed out.
27	Abbreviations	Abbreviations and acronyms are only introduced when they will be used frequently and consistently. All abbreviations are documented and maintained in a reference list.
28	Visualization aids	There are visual drawings and or tools to make it easier to understand the overall structure
29		
30	Reuse	The taxonomy was or can be built by reusing existing taxonomies and/or other structured knowledge assets.
31	Effort to create	The effort to build the taxonomy is offset by the potential benefits.
32	Effort to implement	The taxonomy is expressed in a way that makes it relatively straight-forward to implement.
33	Applicability across multiple areas	The taxonomy and/or the components used to build the taxonomy are general enough to be readily reused in more than one context. For example, use of facets tends to make reuse easier.
34	Impact on People, Behavior	The nature and degree of impact on people and their behaviors as a result of adopting the taxonomy is modest and/or short-lived or outweighed by the advantages of adopting it.
35	Impact of Existing Processes, Taxonomies	The nature and degree of impact on existing processes and taxonomies as a result of adopting the taxonomy is modest, short lived or outweighed by the advantages of adopting it.
36	Impact on Technology	The nature and degree of impact on existing applications and technology infrastructure as a result of adopting the taxonomy is modest, short lived or outweighed by the advantages of adopting it.

Evaluate along Specific Criteria

Critera:

- Clarity & Focus
- Scope coverage
- Right level of detail
- Categorization rigor
- Consistency and Uniformity
- Rate on a scale: 1-5
- Review with client



Taxonomy Questionnaire (1/3)		
General	Title	What is official name of the taxonomy?
	Authorship	Names of individuals or organization
		Internal or External to Company
		Governing Organization
	Audience	Who needs to understand it?
		Who will is is using it?
Scope	Subject	What is the general subject of what is being organized?
	What is being organized	What specific kinds of things is the taxonomy intended to organize?
	Examples	Give examples of what is being organized..
	Counter Examples	What things are similar to what is in scope, but are known to be out of scope?
	Scope criteria	What criteria can be used to determine whether something should be in or out of scope? This will often relate to the purpose.
Purpose	Explicit purpose	State one or more purposes for the taxonomy. Be clear and explicit, so that it may usefully guide the development of the taxonomy.
	Benefits	Indicate the intended benefits that are expected to be realized from having the taxonomy.
	Intended applications	Describe one or more ideas for specific applications of the taxonomy. Is it intended to be applied narrowly within just one application, or broadly among many?
©2014 Semantic		
Taxonomy Questionnaire (2/3)		
Syntax and Semantics	Nodes: Meaning	What is the meaning of a node in the taxonomy? Is it a category? Is it a topic? What are the things that are in the category or are classified by the topic?
	Nodes: Uniformity	Are all the nodes the same kind of thing, or are they different kinds, e.g for different levels? What are the
	Links: Meaning	What is the meaning of a link connecting two nodes? E.g. subtype?part of, rollup?
	Links: Uniformity	Do all the links mean the same thing? Or do the links between different levels mean different things. What are all the different kinds of links?
	Conceptual / Logical	What is the conceptual underpining for how the taxonomy is represented. E.g. an informal set of indented lists vs. an ISO standard definition of
	Format	What actual format does the taxonomy exist in? Word? Excel? XML? OWL?
	Requirements	Note if there are any notational requirements e.g. everything has to be in XML.
Development	Effort	How much effort did or will it take to develop the
	Reuse	Is or will the taxonomy be based on existing taxonomies and other knowledge assets, or will be be largely from
	Author's skills	What are the skills and backgrounds of the authors? Are they professional taxonomists, data modelers, self-taught critical thinkers?
©2014 Semantic		
Taxonomy Questionnaire (3/3)		
Implementation Impact	Impact on people and behavior	Describe the nature and degree of impact on people and their behaviors as a result of adapting the taxonomy.
	Impact on processes	Describe the nature and degree of impact on existing processes and taxonomies as a result of adapting the taxonomy
	Impact on existing knowledge assets	Describe the nature and degree of impact on existing taxonomies and other knowledge assets as a result of adapting the taxonomy.
	Impact on applications	Describe the nature and degree of impact on existing applications as a result of adapting the taxonomy
	Impact on technology infrastructure	Describe the nature and degree of impact on technology infrastructure as a result of adapting the
Provenance	Foundation / Background	Is the taxonomy brand new, or is it based on existing knowledge assets? Which ones? How is the current taxonomy different from the existing assets?
	Versions	Does this taxonomy have prior versions? How is this one different from previous ones?

Some Problems with Initial Asset

- It turned out that the original set of buckets was hard to use and manage…
- .. even for the original intended purpose
- If a change was needed, it was hard to see which of the hundreds of buckets would be affected.
- Impact analysis was next to impossible
- But the core asset was very important

What's Going On?

- The buckets were mainly text descriptions.
- No structure means no automation, every bucket had to be examined manually.
- Yet, there *was* evidence of much structure lurking behind the text.
- It was plain to see, if you were paying attention and knew what to look for

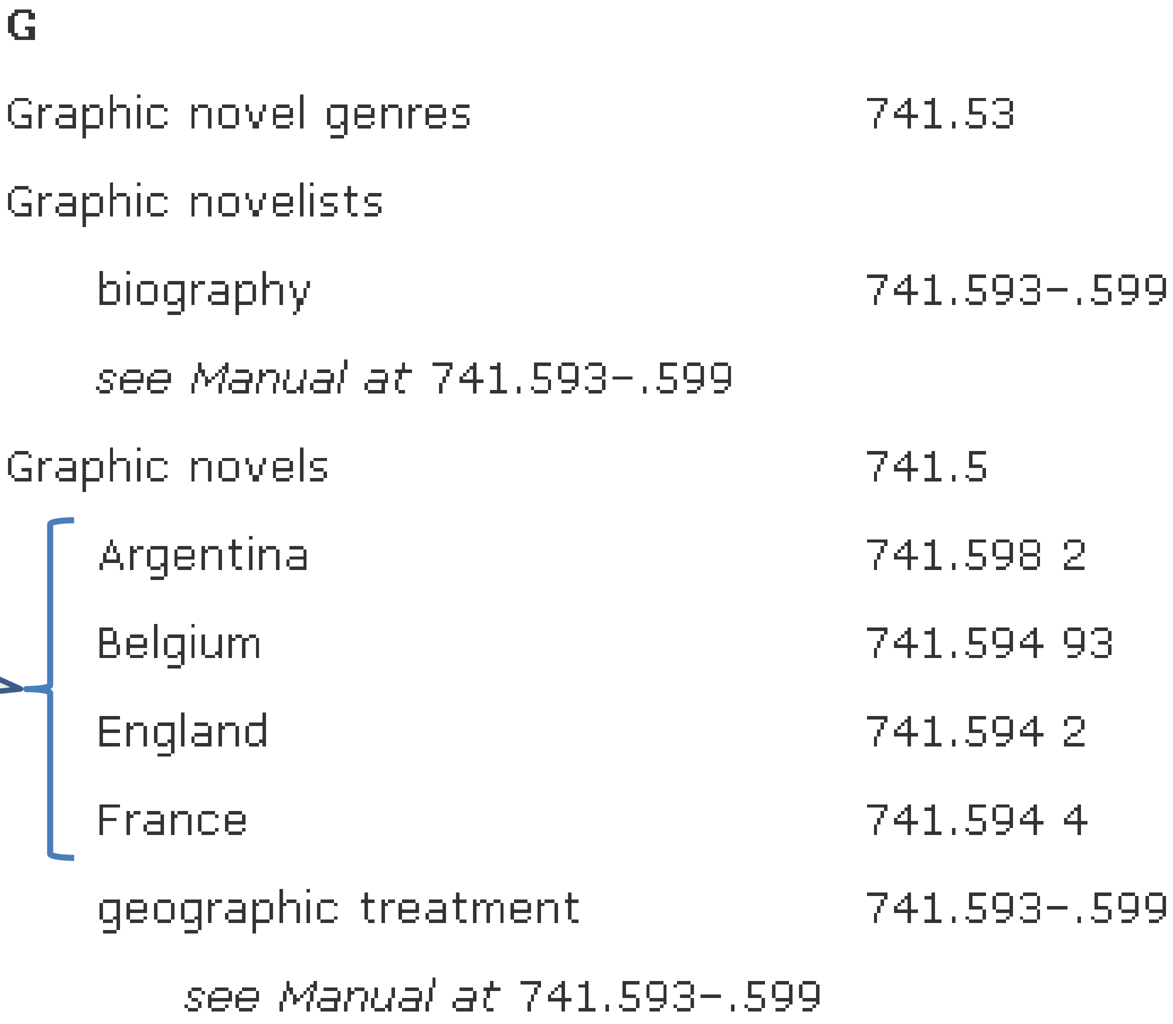
Finding Structure

- Look closely at the text descriptions and look for patterns.
- There were some recurring themes.
- Frequent mention of Goal, Region and Product, but these ideas are not captured or used in a uniform manner.

Making Implicit Structure Explicit

- Try to reword descriptions for the buckets in a uniform way. For example:
  - Work on iPhones in Africa to reduce service call wait times.
  - Work on product P in region R to achieve goal G
- The manual rewording was not always so easy.
- Not all ways to capture structure are equal
- Let's consider an old and familiar structure…

Dewey Decimal: Geography



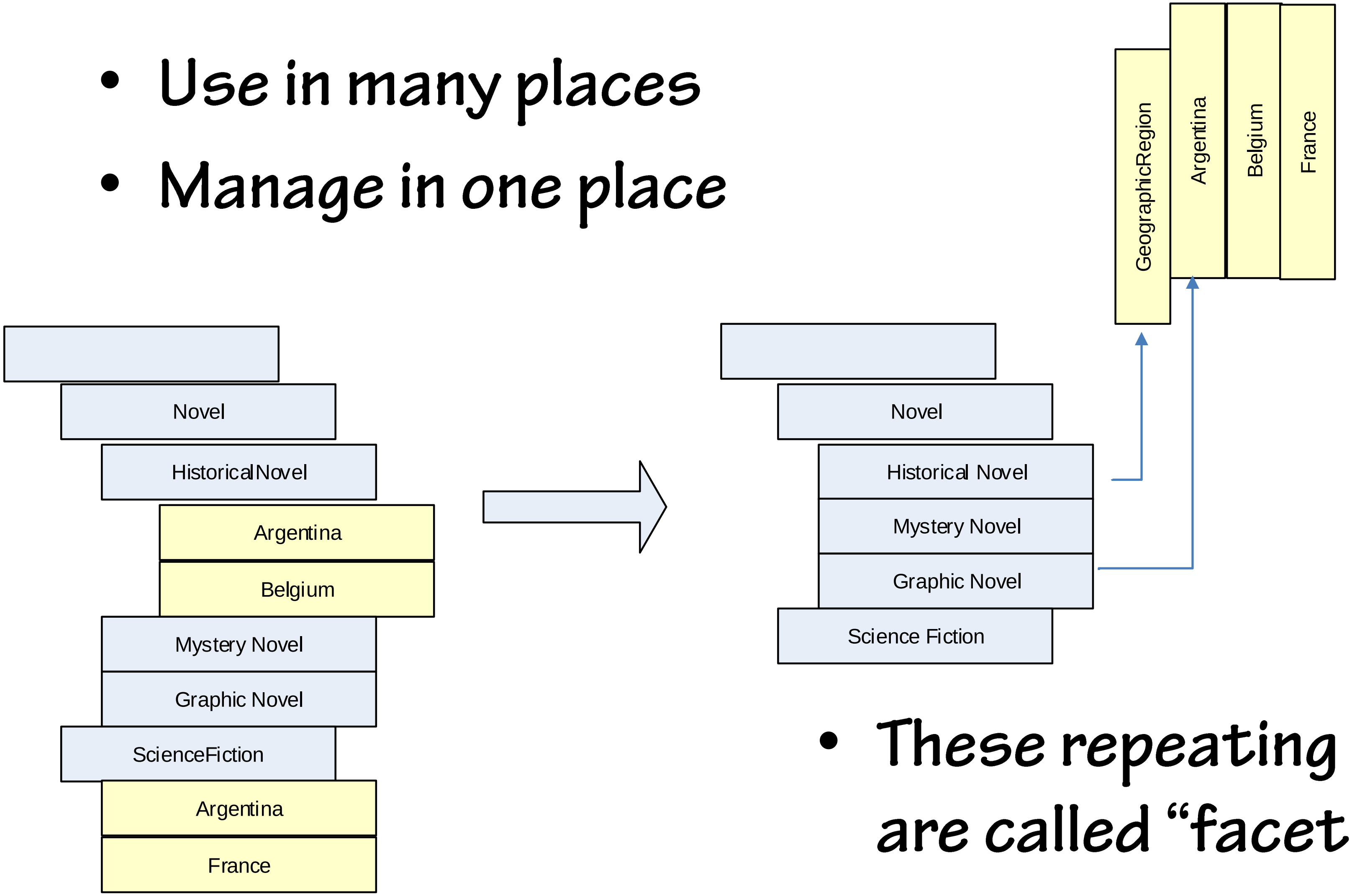
G

Graphic novel genres	741.53
Graphic novelists	
biography	741.593-.599
see Manual at 741.593-.599	
Graphic novels	741.5
Argentina	741.598 2
Belgium	741.594 93
England	741.594 2
France	741.594 4
geographic treatment	741.593-.599
see Manual at 741.593-.599	

Geography will turn up in many different places.

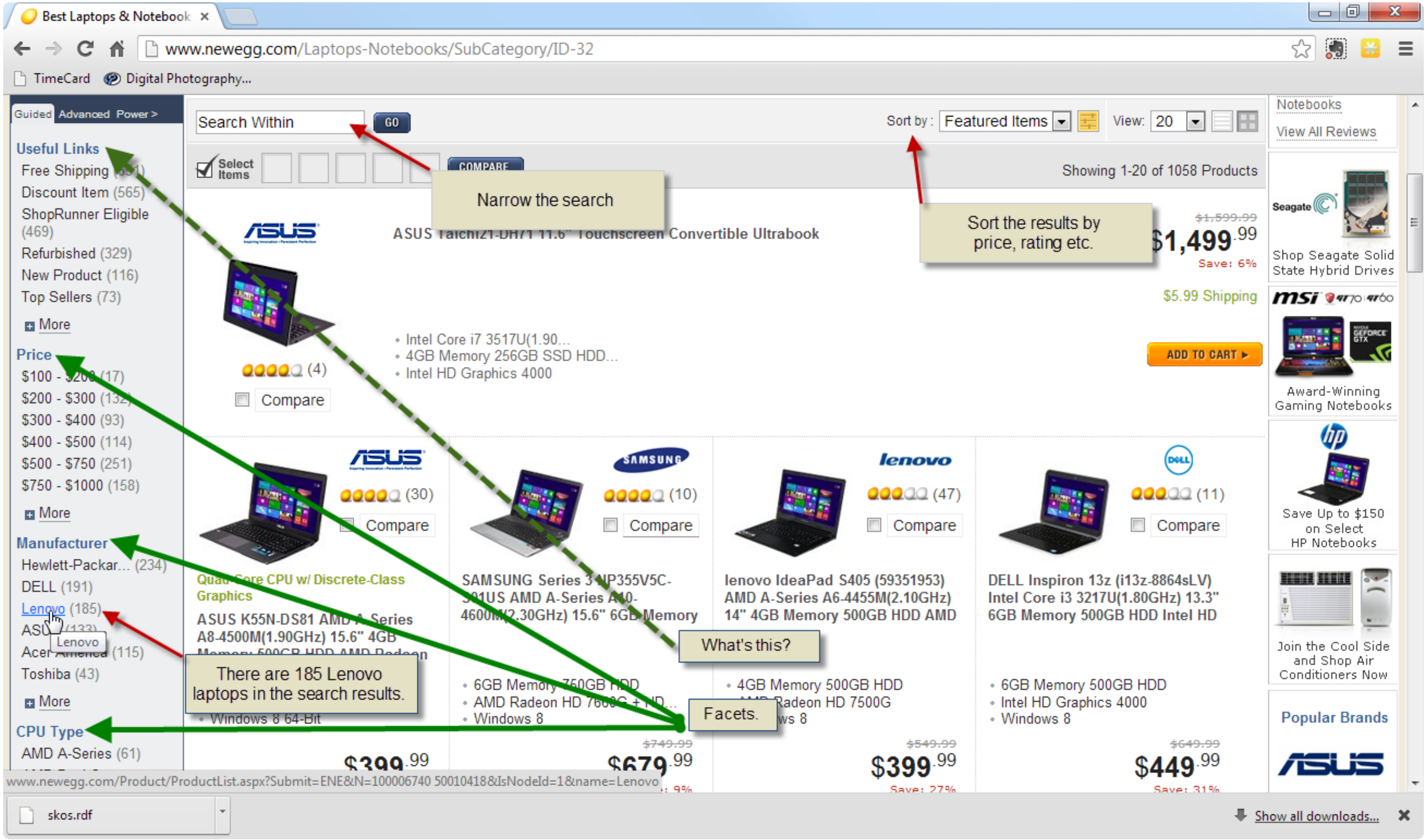
Dewey Decimal: Geography

- Use in many places
- Manage in one place



- These repeating ideas are called "facets"
- E.g. faceted search

Faceted Search: www.newegg.com



Best Laptops & Notebook

www.newegg.com/Laptops-Notebooks/SubCategory/ID-32

TimeCard · Digital Photography...

Guided · Advanced · Power >

Useful Links
- Free Shipping (551)
- Discount Item (565)
- ShopRunner Eligible (469)
- Refurbished (329)
- New Product (116)
- Top Sellers (73)
- More

Price
- $100 - $200 (17)
- $200 - $300 (132)
- $300 - $400 (93)
- $400 - $500 (114)
- $500 - $750 (251)
- $750 - $1000 (158)
- More

Manufacturer
- Hewlett-Packar... (234)
- DELL (191)
- Lenovo (185)
- ASUS (133)
- Acer America (115)
- Toshiba (43)
- More

CPU Type
- AMD A-Series (61)

Search Within · GO

Sort by : Featured Items · View: 20

Select Items · COMPARE

Showing 1-20 of 1058 Products

ASUS Taichi21-DH71 11.6" Touchscreen Convertible Ultrabook
- Intel Core i7 3517U(1.90...
- 4GB Memory 256GB SSD HDD...
- Intel HD Graphics 4000

(4) · □ Compare

~~$1,599.99~~ $1,499.99 · Save: 6% · $5.99 Shipping · ADD TO CART ▸

(30) · □ Compare

Quad-Core CPU w/ Discrete-Class Graphics

ASUS K55N-DS81 AMD A-Series A8-4500M(1.90GHz) 15.6" 4GB Memory 500GB HDD AMD Radeon
- Windows 8 64-Bit

$399.99

(10) · □ Compare

SAMSUNG Series 3 NP355V5C-S01US AMD A-Series A10-4600M(2.30GHz) 15.6" 6GB Memory
- 6GB Memory 750GB HDD
- AMD Radeon HD 7660G + HD...
- Windows 8

~~$749.99~~ $679.99 · Save: 9%

(47) · □ Compare

lenovo IdeaPad S405 (59351953) AMD A-Series A6-4455M(2.10GHz) 14" 4GB Memory 500GB HDD AMD
- 4GB Memory 500GB HDD
- AMD Radeon HD 7500G
- Windows 8

~~$549.99~~ $399.99 · Save: 27%

(11) · □ Compare

DELL Inspiron 13z (i13z-8864sLV) Intel Core i3 3217U(1.80GHz) 13.3" 6GB Memory 500GB HDD Intel HD
- 6GB Memory 500GB HDD
- Intel HD Graphics 4000
- Windows 8

~~$649.99~~ $449.99 · Save: 31%

Notebooks · View All Reviews

Shop Seagate Solid State Hybrid Drives

Award-Winning Gaming Notebooks

Save Up to $150 on Select HP Notebooks

Join the Cool Side and Shop Air Conditioners Now

Popular Brands

www.newegg.com/Product/ProductList.aspx?Submit=ENE&N=100006740 50010418&IsNodeId=1&name=Lenovo

skos.rdf · Show all downloads...

Narrow the search

Sort the results by price, rating etc.

What's this?

Facets.

There are 185 Lenovo laptops in the search results.

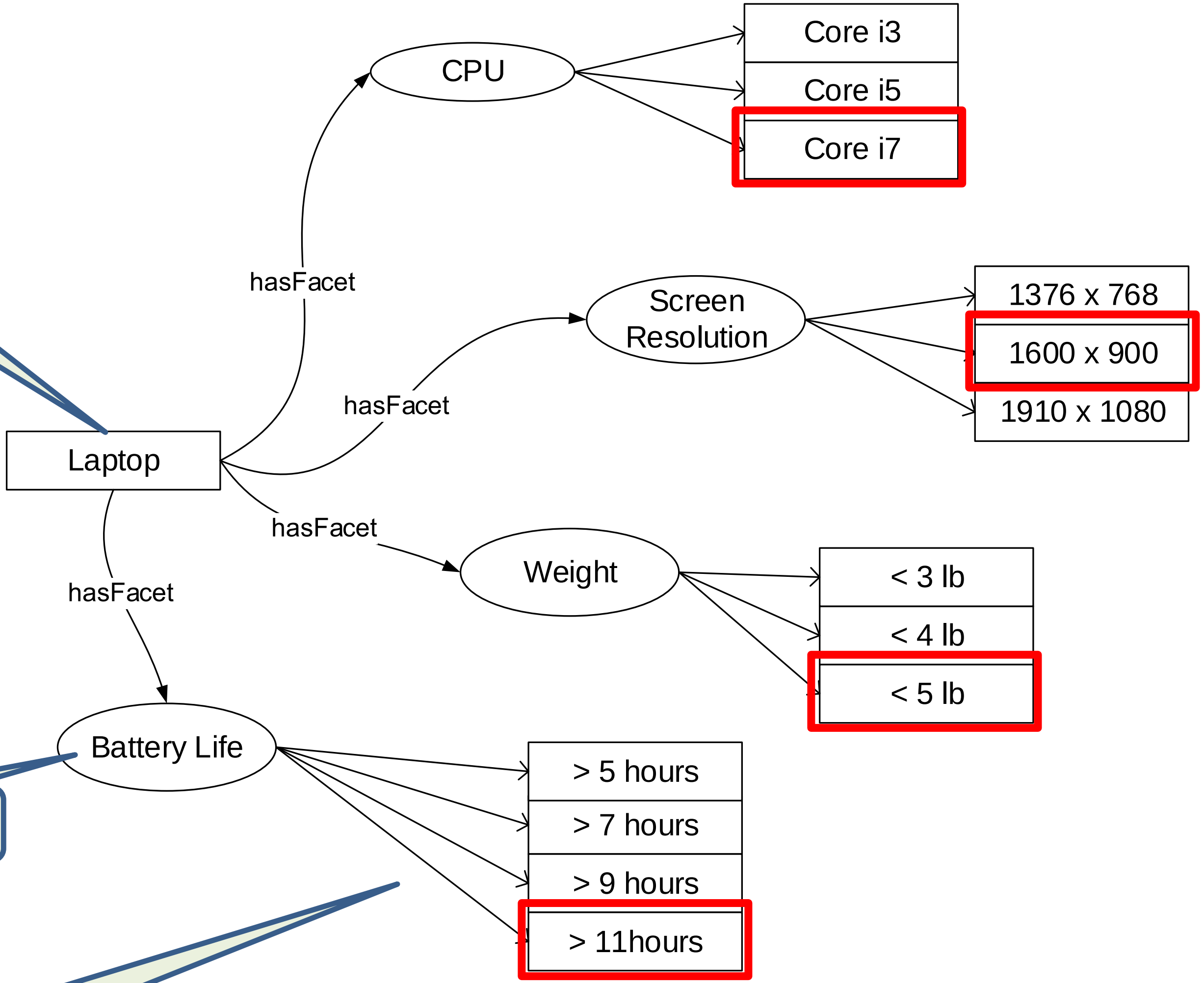
A Faceted Taxonomy for Laptops

The item being classified

Sample category or "bucket":

Under 5 pounds, Battery life more than 11 hours, 1600 x 900 resolution, i7 CPU

Each bucket is highly structured



Facet

Possible values

Back to our Structured English

- We have Goal, Region and Product
- They are candidate "facets" for characterizing the items in question.
- But there were about a dozen other potential facets that we saw in the text descriptions

- Which ones really mattered?
- Which ones are just incidental?
- Can facets really help?

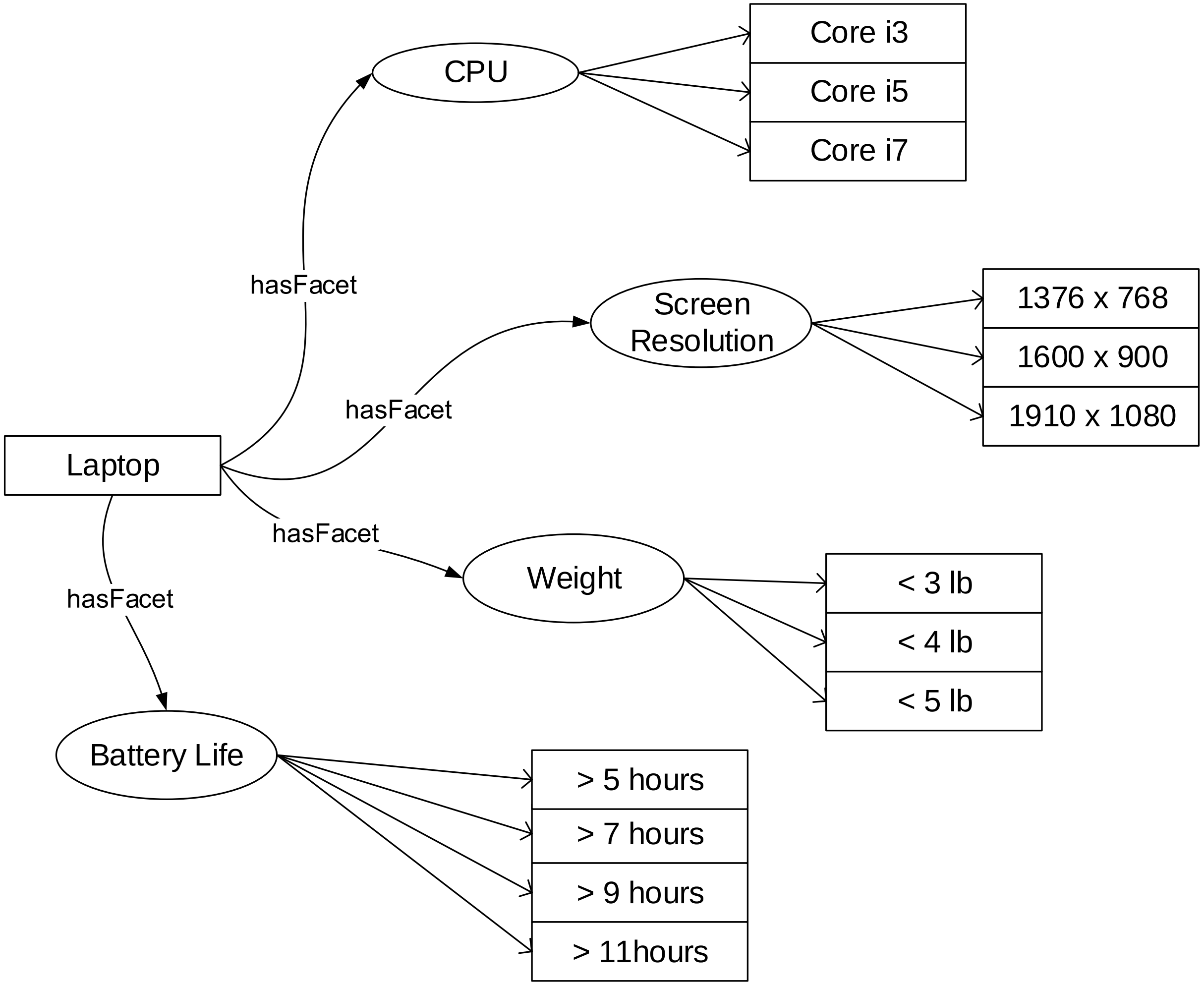
Facet Math

Without facets: there are exponentially many buckets: 3 x 3 x 3 x 4 = 108

108 things to learn and remember is a lot.

The faceted approach means there are:

- 4 facets + 13 values
- = 17 things to learn and remember



This is a Big Deal

Exponentially reducing the number of things to learn to classify things has numerous benefits

- Faster to train people
- More accurate classification
- Easier to evolve and maintain moving forward.
- The more facets & values, the greater the savings

- But how do we know we have the right facets?

Many Meetings with Stakeholders

- Get experts about the items in question.
- Ask them to identify the ways that items are different from one another
- Brainstorm to identify candidate facets
- Then evaluate them

- Example Criteria: Ideally each facet value should be unique for a given facet.

Uniqueness: An Example

Suppose we are classifying quality control activities. One facet is the goal. Values might be:

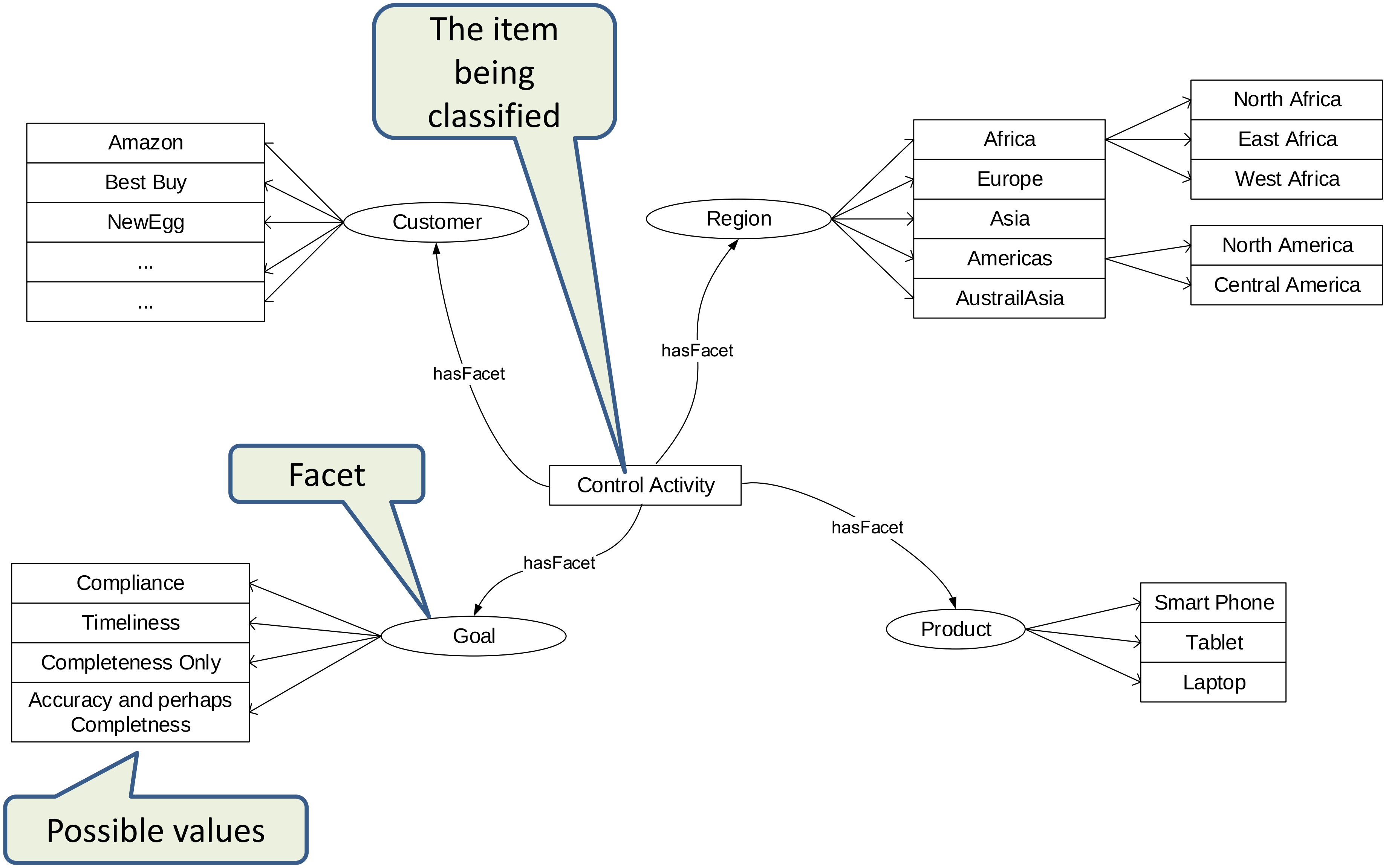
- timeliness
- completeness
- accuracy

- timeliness
- completeness only
- accuracy and possibly completeness

What happens if some control actions are for both completeness and accuracy?

Then it is hard to uniquely classify the item.

A Faceted Taxonomy for Control Activity



The item being classified

Control Activity

hasFacet

Customer

Amazon

Best Buy

NewEgg

...

...

hasFacet

Region

Africa

Europe

Asia

Americas

AustrailAsia

North Africa

East Africa

West Africa

North America

Central America

hasFacet

Product

Smart Phone

Tablet

Laptop

hasFacet

Goal

Facet

Compliance

Timeliness

Completeness Only

Accuracy and perhaps Completness

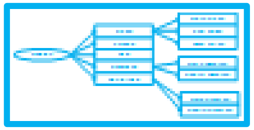
Possible values

Space of Possible Values for a Facet

The values for the facets/properties may be:

- Flat: a flat list of a handful of possible values (e.g. Amazon, Best Buy, New Egg)

- Hierarchical: a simple taxonomy (e.g. geographic regions)



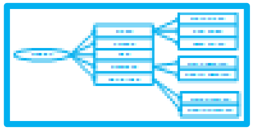
- Can anyone think of another possibility?
- What about Laptops?

Space of Possible Values for a Facet

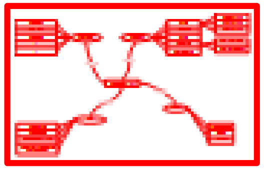
The values for the facets/properties may be:

- Flat: a flat list of a handful of possible values (e.g. Amazon, Best Buy, New Egg)

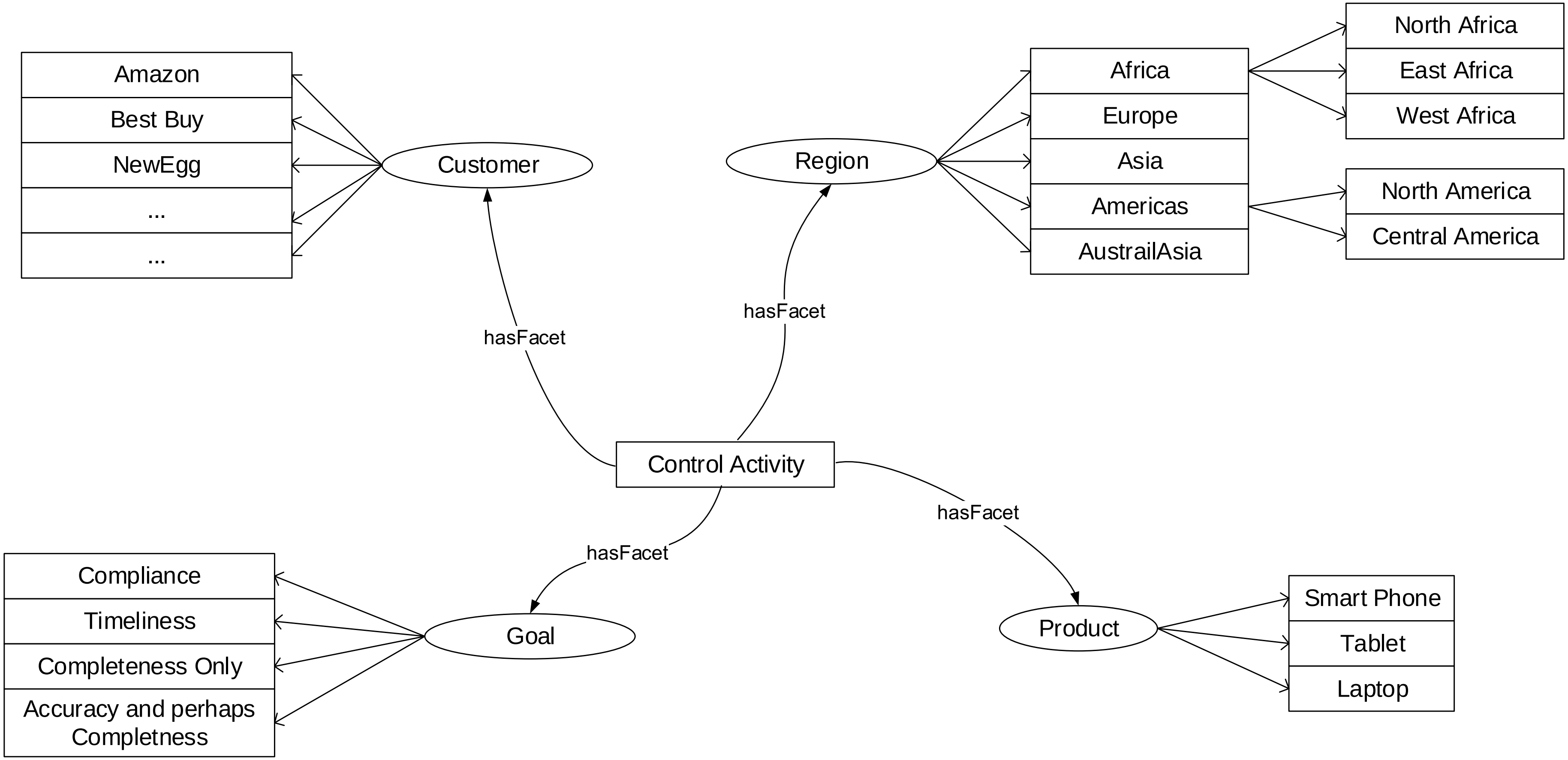
- Hierarchical: a simple taxonomy (e.g. geographic regions)



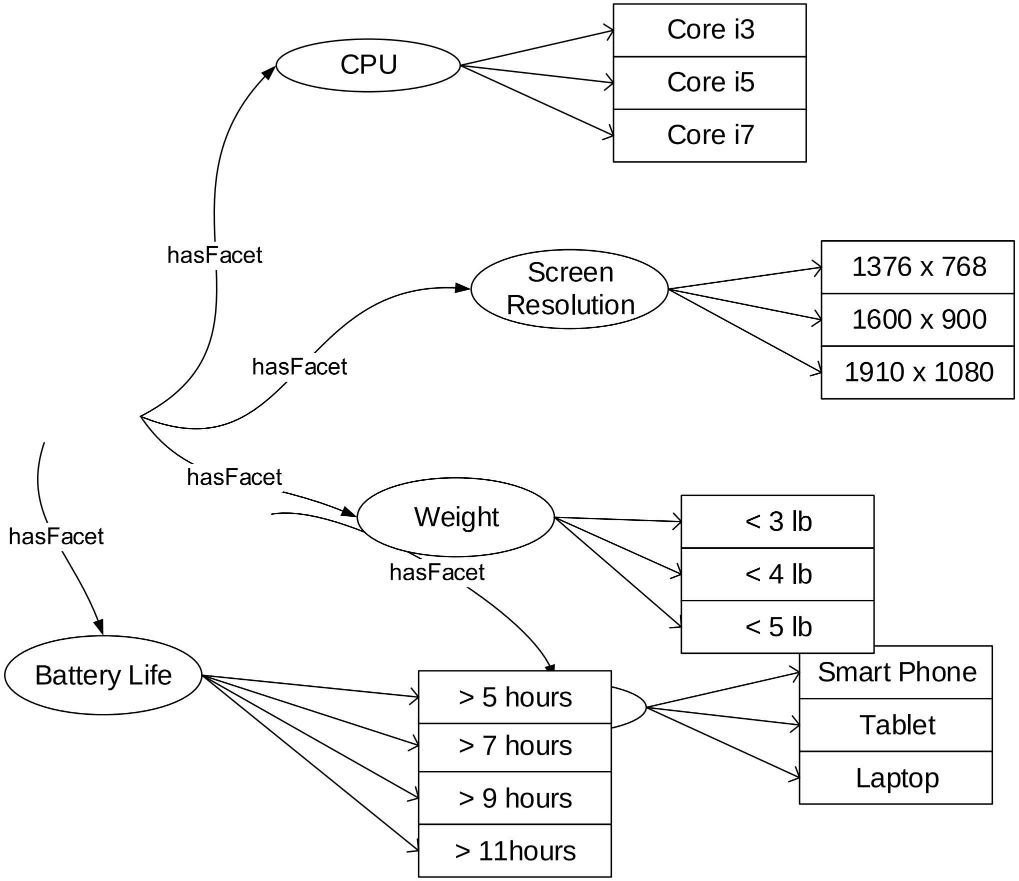
- Faceted: another faceted taxonomy embedded in the prior faceted taxonomy (e.g. products)



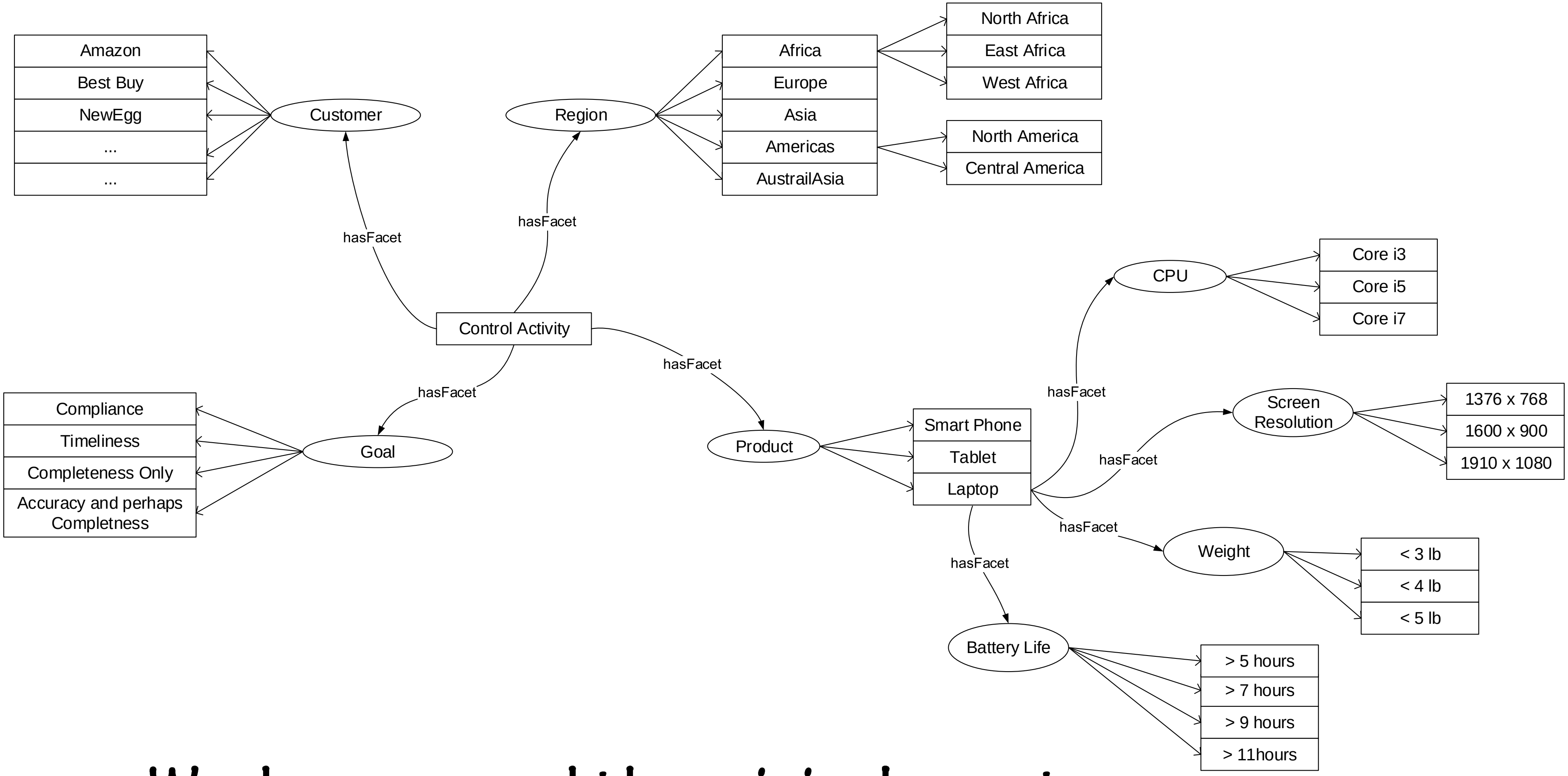
A Faceted Taxonomy for Control Activity



A Faceted Taxonomy for Controls



A Faceted Taxonomy for Controls

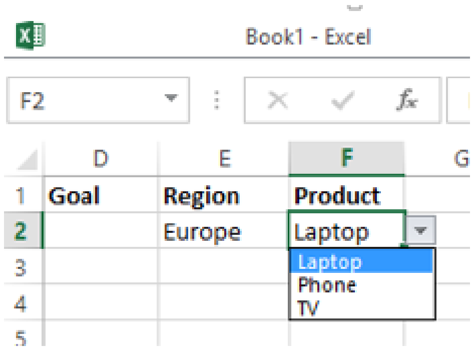


- We decomposed the original asset
- Next: re-compose it from the pieces

Re-Characterizing the Items

- For each of hundreds or thousands of item descriptions, re-characterize them using the facets.
- Many ways to do this:
  - Manually reword them one by one

  - Use a spreadsheet to create a form
    - One field in the form for each facet
    - Values may be selected from a dropdown or entered into a text field



Book1 - Excel

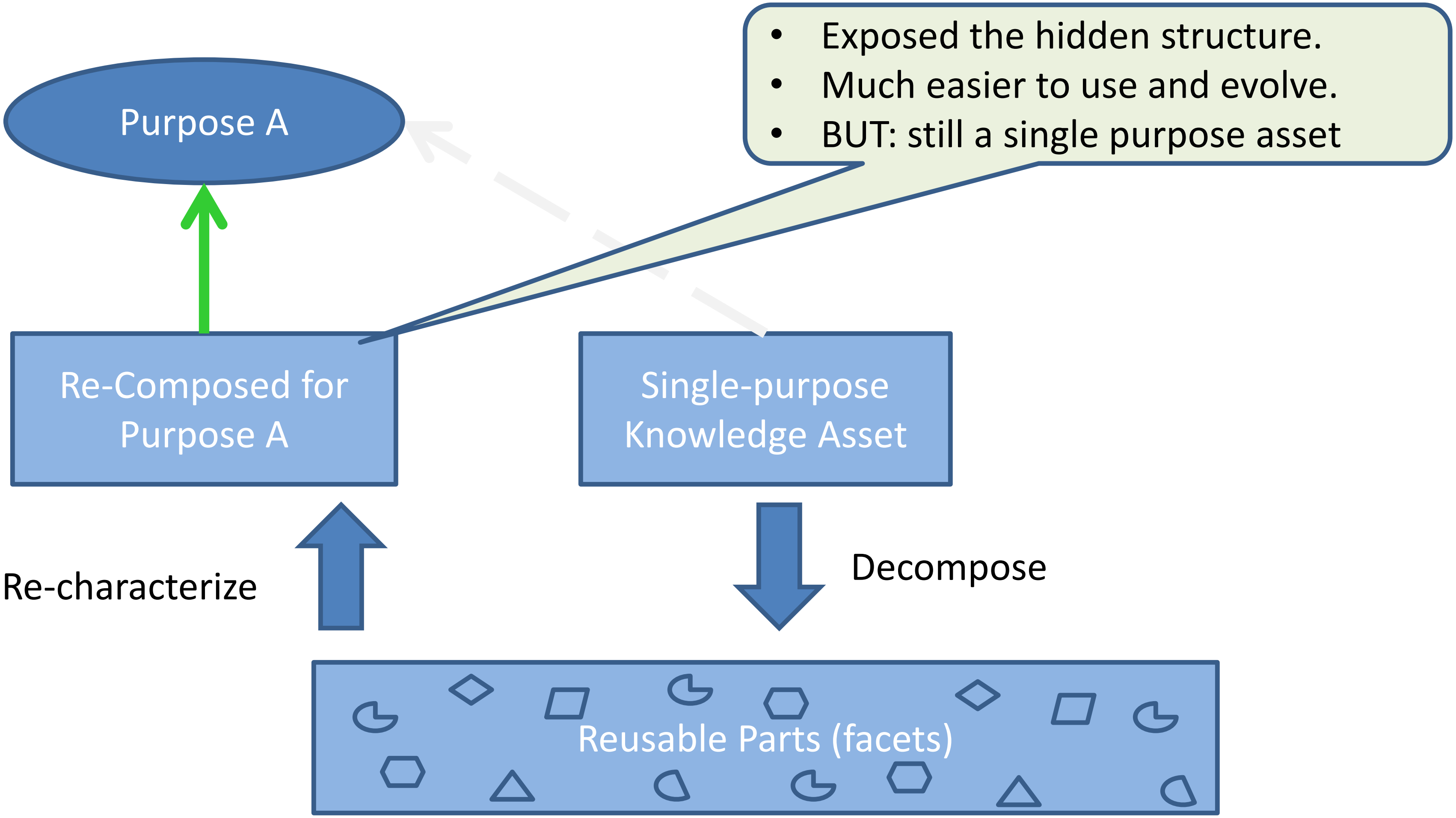
F2

	D	E	F	G
1	Goal	Region	Product	
2		Europe	Laptop	
3				
4				
5				

Laptop
Phone
TV

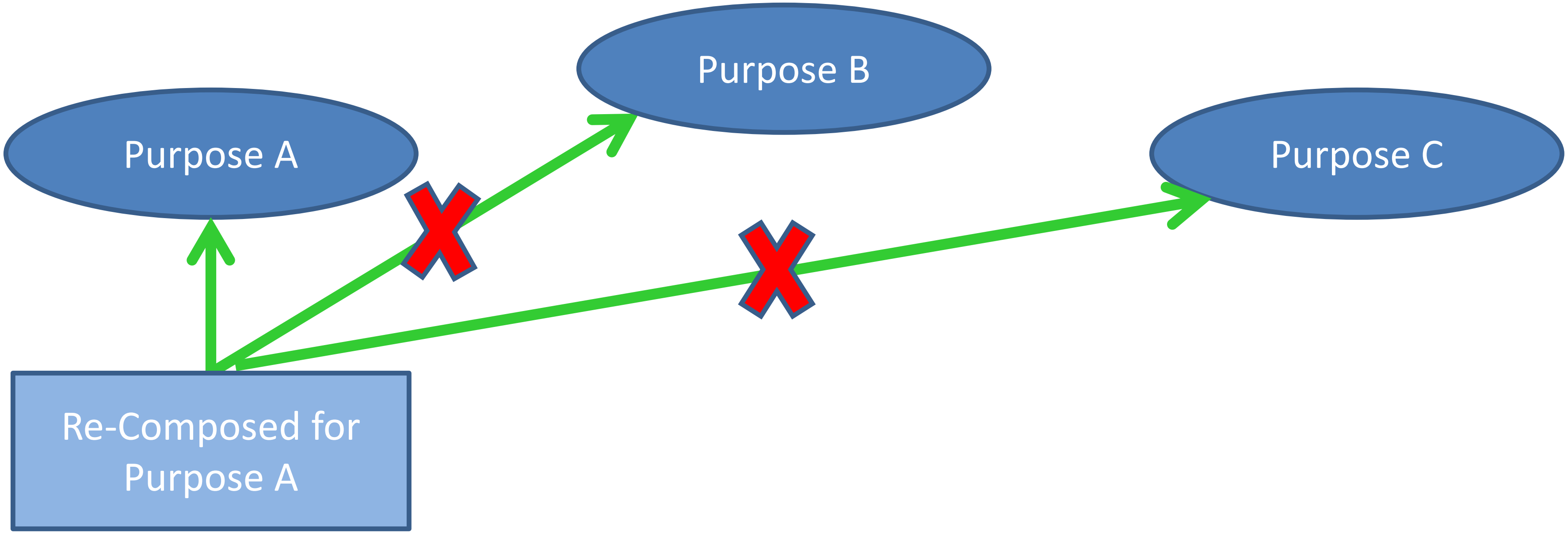
  - Build a simple app that automates the form
    - Create a simple ontology
    - Use it to drive the form
    - The taxonomy becomes a set of triples that can be queried

Decomposing & Re-characterizing

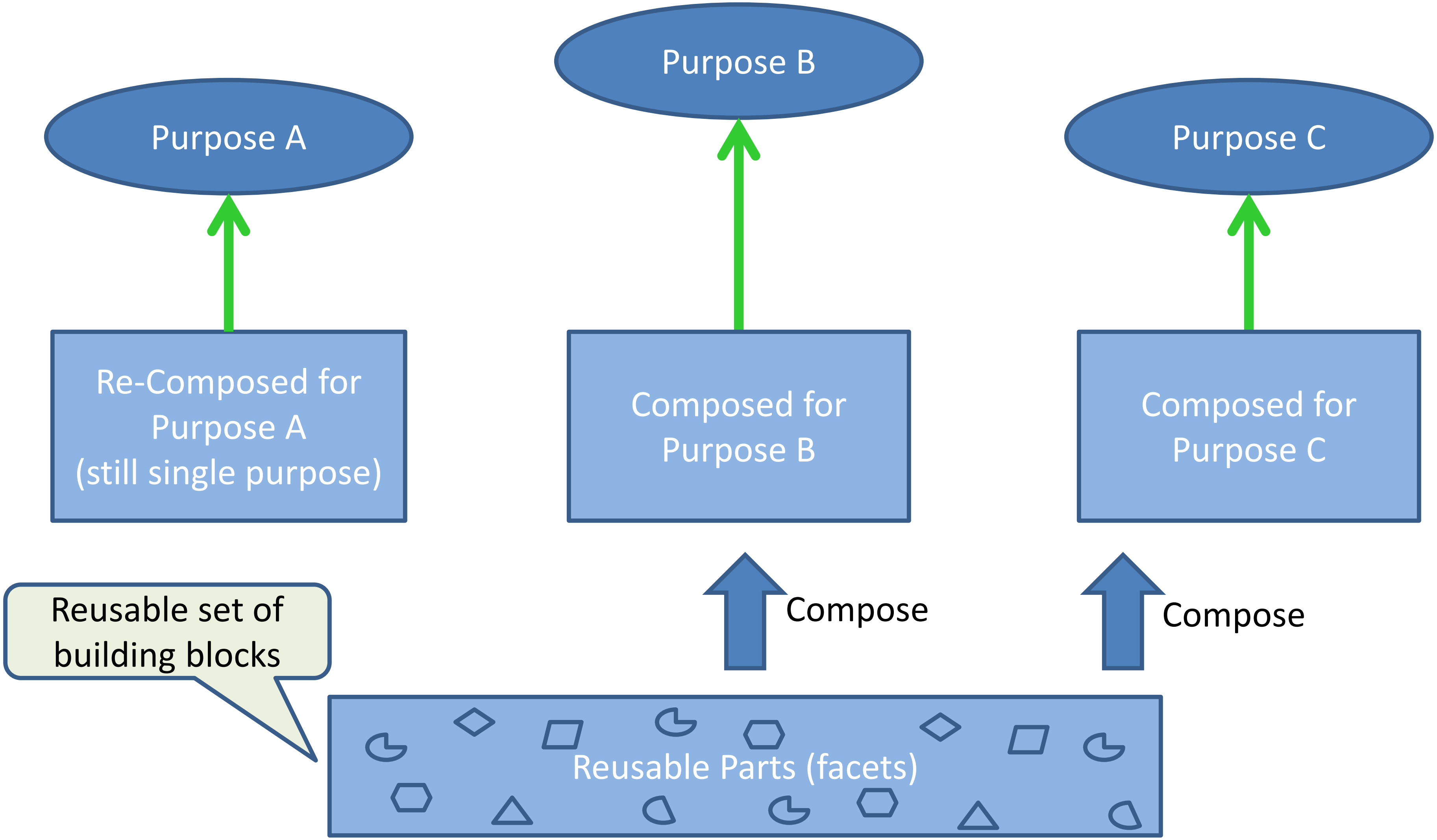


- Exposed the hidden structure.
- Much easier to use and evolve.
- BUT: still a single purpose asset

What about other uses?



What about other uses?



Purpose A

Purpose B

Purpose C

Re-Composed for Purpose A (still single purpose)

Composed for Purpose B

Composed for Purpose C

Compose

Compose

Reusable set of building blocks

Reusable Parts (facets)

More than Just a Story

- In our early work at Mega, this was just a story, a nice idea we hoped would come true.
- Several months later, we were back at Mega and asked them how things were going.
- They are doing just what the picture depicts
- Taking the facets and applying them to classify the items for their own purposes
- But wait, there's more!
- What about an ontology?

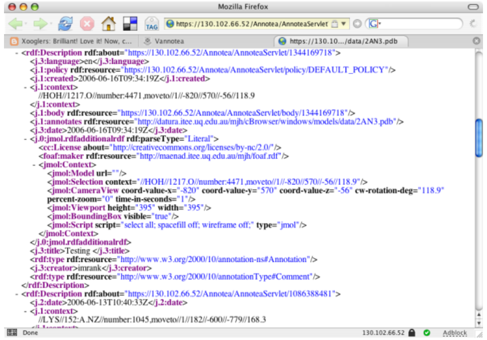
Linking Taxonomies to an Ontology

- Normally, a taxonomy of terms, or a faceted taxonomy would live independently from an ontology.
- Our vision is to have every thing connected.
  - spreadsheets with a semantic underpinning
  - multiple applications & databases
  - data models and messages
- Opens up vast possibilities for querying and analyzing data across an enterprise

Enterprise-Wide Ontology

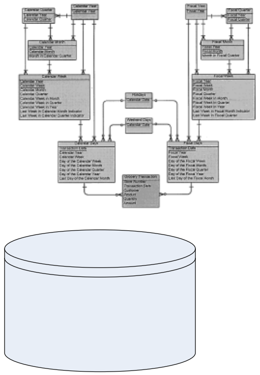
- We were also building an enterprise ontology for a major part of their business.
- They are now linking the facets to the ontology so that faceted taxonomies are living in harmony with formal ontologies.
- All the way from text definitions to informal taxonomies to faceted taxonomies to ontologies, everything linked together.

Triple as Common Denominator

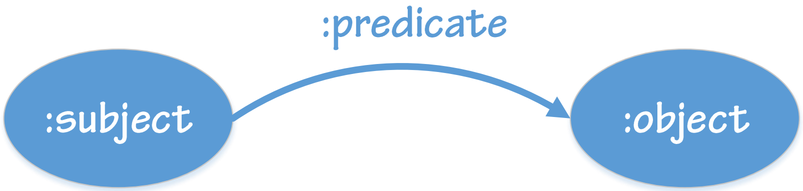

Mozilla Firefox
https://130.102.66.52/Annotea/AnnoteaServlet
Xooglers: Brilliant! Love it! Now, c...
Vannotea
https://130.10..../data/2AN3.pdb
Done
130.102.66.52
Adblock

```
- <rdf:Description rdf:about="https://130.102.66.52/Annotea/AnnoteaServlet/1344169718">
    <j.3:language>en</j.3:language>
    <j.1:policy rdf:resource="https://130.102.66.52/Annotea/AnnoteaServlet/policy/DEFAULT_POLICY"/>
    <j.1:created>2006-06-16T09:34:19Z</j.1:created>
  - <j.1:context>
      //HOH//1217.O//number:4471,moveto//1//-820//570//-56//118.9
    </j.1:context>
    <j.1:body rdf:resource="https://130.102.66.52/Annotea/AnnoteaServlet/body/1344169718"/>
    <j.1:annotates rdf:resource="http://datura.itee.uq.edu.au/mjh/cBrowser/windows/models/data/2AN3.pdb"/>
    <j.3:date>2006-06-16T09:34:19Z</j.3:date>
  - <j.0:jmol.rdfadditionalrdf rdf:parseType="Literal">
      <cc:License about="http://creativecommons.org/licenses/by-nc/2.0/"/>
      <foaf:maker rdf:resource="http://maenad.itee.uq.edu.au/mjh/foaf.rdf"/>
    - <jmol:Context>
        <jmol:Model url=""/>
        <jmol:Selection context="//HOH//1217.O//number:4471,moveto//1//-820//570//-56//118.9"/>
        <jmol:CameraView coord-value-x="-820" coord-value-y="570" coord-value-z="-56" cw-rotation-deg="118.9"
        percent-zoom="0" time-in-seconds="1"/>
        <jmol:Viewport height="395" width="395"/>
        <jmol:BoundingBox visible="true"/>
        <jmol:Script script="select all; spacefill off; wireframe off;" type="jmol"/>
      </jmol:Context>
    </j.0:jmol.rdfadditionalrdf>
    <j.3:title>Testing </j.3:title>
    <rdf:type rdf:resource="http://www.w3.org/2000/10/annotation-ns#Annotation"/>
    <j.3:creator>imrank</j.3:creator>
    <rdf:type rdf:resource="http://www.w3.org/2000/10/annotationType#Comment"/>
  </rdf:Description>
- <rdf:Description rdf:about="https://130.102.66.52/Annotea/AnnoteaServlet/1086388481">
    <j.2:date>2006-06-13T10:40:33Z</j.2:date>
  - <j.1:context>
      //LYS//152:A.NZ//number:1045,moveto//1//182//-600//-779//168.3
```

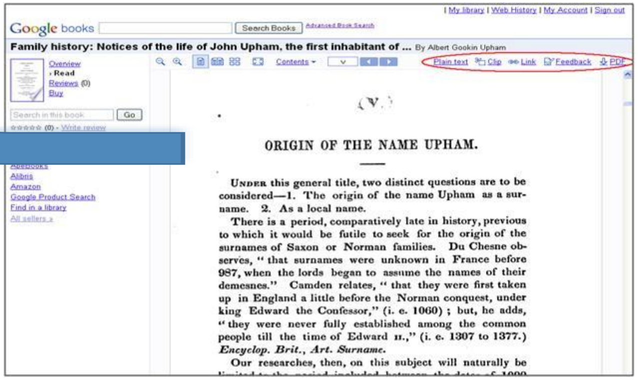
XML



Databases



Spreadsheets



Family history: Notices of the life of John Upham, the first inhabitant of ... By Albert Gookin Upham

(V.)

ORIGIN OF THE NAME UPHAM.

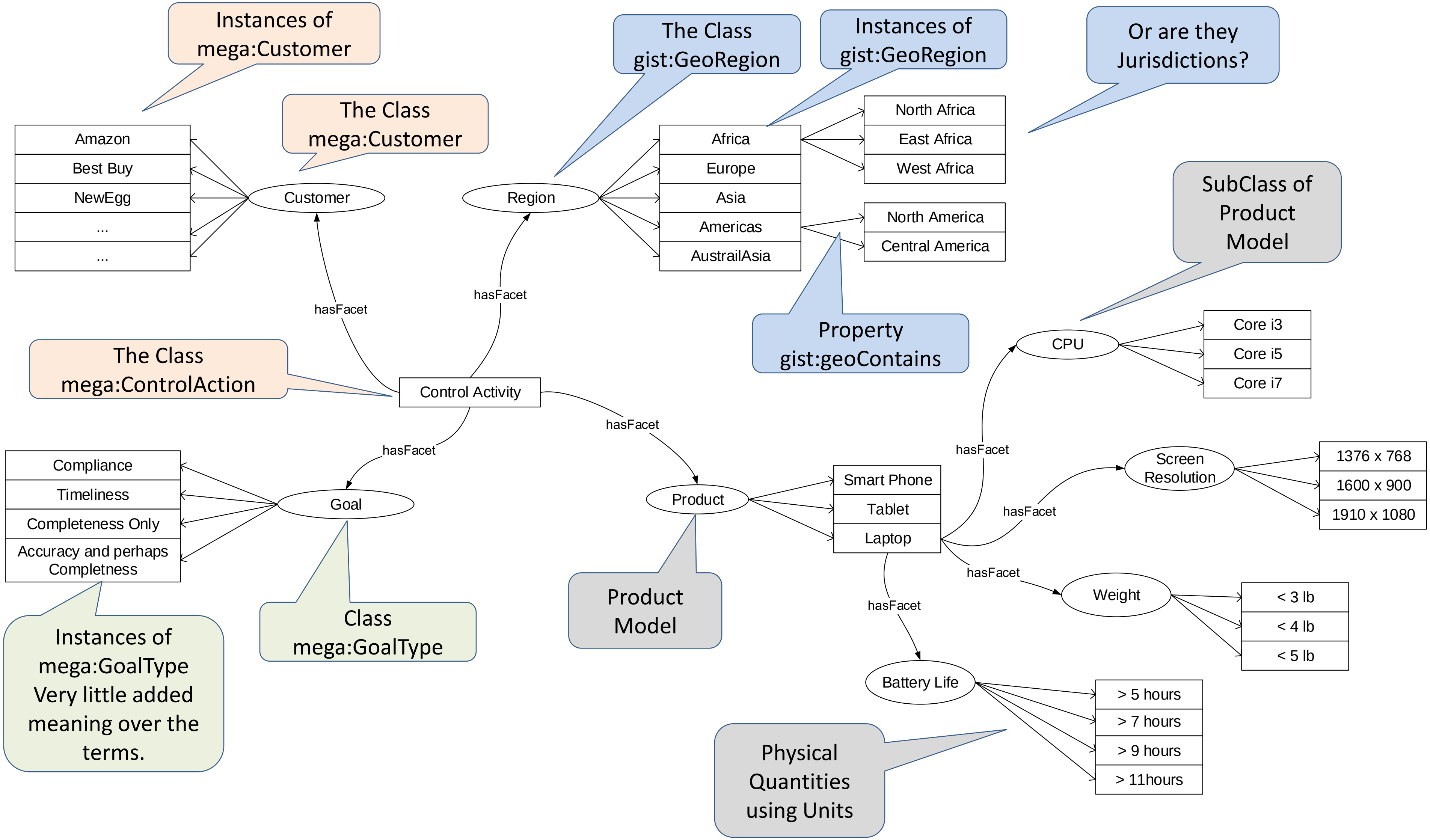
Under this general title, two distinct questions are to be considered—1. The origin of the name Upham as a surname. 2. As a local name.

There is a period, comparatively late in history, previous to which it would be futile to seek for the origin of the surnames of Saxon or Norman families. Du Chesne observes, " that surnames were unknown in France before 987, when the lords began to assume the names of their demesnes." Camden relates, " that they were first taken up in England a little before the Norman conquest, under king Edward the Confessor," (i. e. 1060) ; but, he adds, " they were never fully established among the common people till the time of Edward ɪɪ.," (i. e. 1307 to 1377.) *Encyclop. Brit., Art. Surname.*

Our researches, then, on this subject will naturally be

Free text

Linking Taxonomies to an Ontology



Instances of mega:Customer

The Class mega:Customer

The Class gist:GeoRegion

Instances of gist:GeoRegion

Or are they Jurisdictions?

SubClass of Product Model

Property gist:geoContains

The Class mega:ControlAction

Product Model

Class mega:GoalType

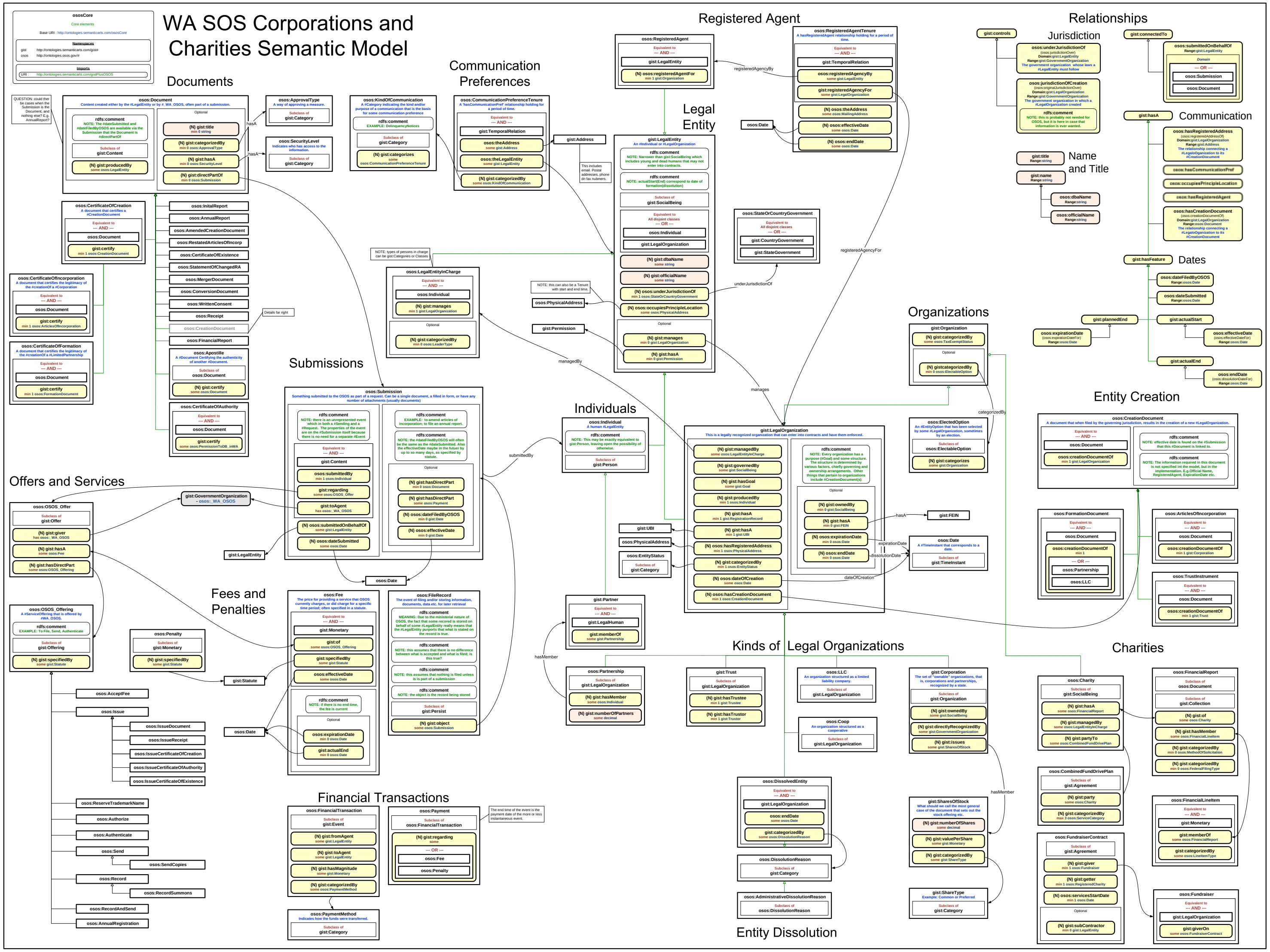
Instances of mega:GoalType Very little added meaning over the terms.

Physical Quantities using Units

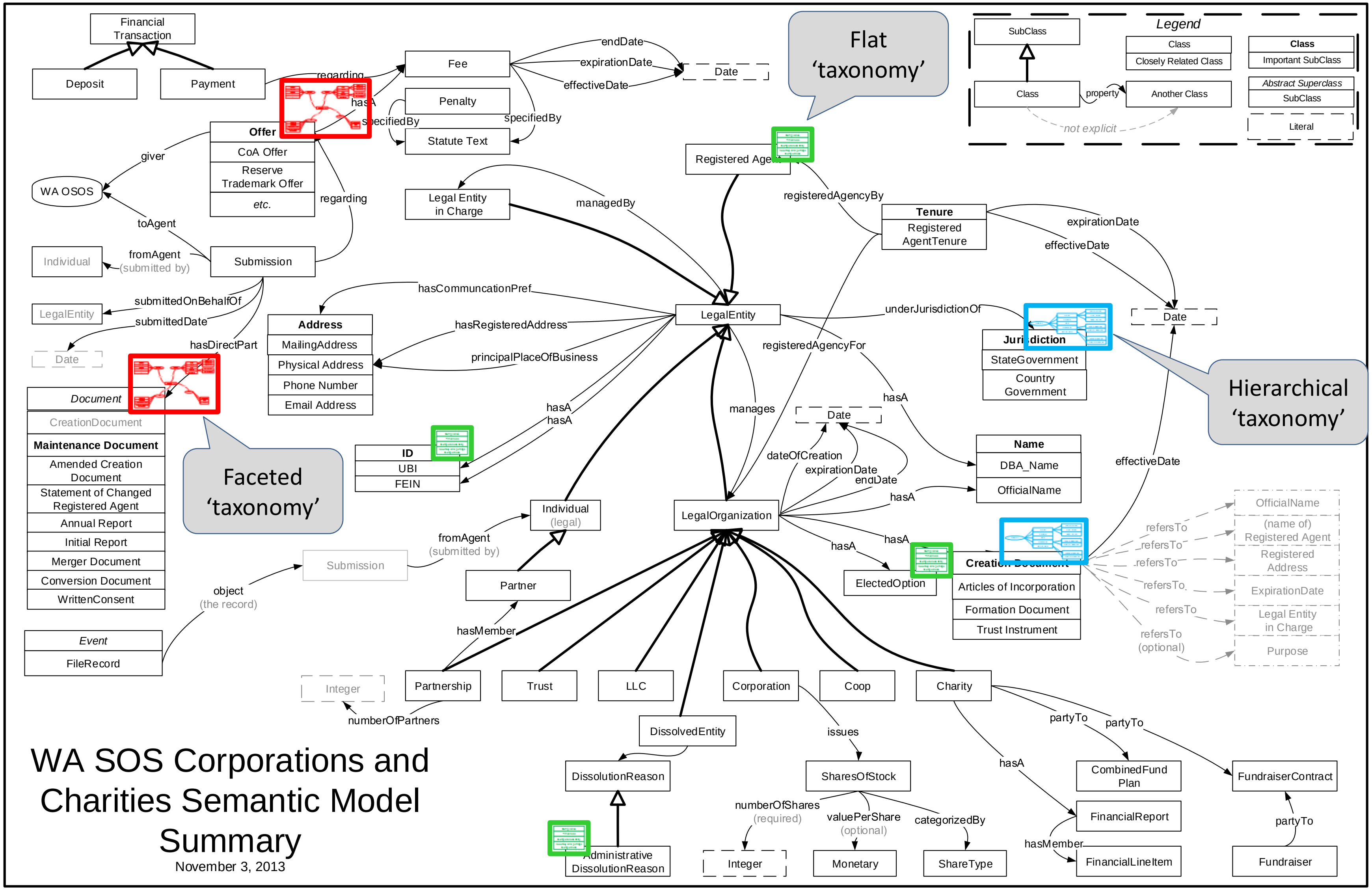
An Ontology Perspective

- We have been talking from a taxonomy perspective, and then linking to an ontology.
- The reverse is when we are building an enterprise ontology and we want to identify where there are potential taxonomies lurking.

Example: Corporations and Charities



Example: Corporations and Charities



Flat ‘taxonomy’

Hierarchical ‘taxonomy’

Faceted ‘taxonomy’

WA SOS Corporations and Charities Semantic Model Summary

November 3, 2013

Governance

- Taxonomies can be independently governed
- When changes occur, the touch points are limited so there is minimal disruption

One More Example: Codes

- A key application at Mega has over 24,000 codes grouped into over 700 code categories.
- There are only two or three people in the company who understand them
- Very time-consuming to learn,
  a risk if people are no longer around
- Impossible to do any serious analytics

One More Example: Codes

- We uploaded the codes into a triple store and explored using SPARQL queries .
- This dramatically reduces learning curve, eases risk and burden on the few experts.
- We also found some errors
- Lucky that no one used those fields
(or maybe they did, and no one noticed!)

The Learning

- A little bit of semantics can make a big difference in a surprising way
- Codes are notoriously difficult to understand (hence the name?)
- But they really do mean something, and we are starting the process of giving them meaning by linking them to the enterprise ontology and the taxonomies.
- In the long term, every one of those codes could turn into a facet value in a taxonomy.
- The vision for ontology and taxonomy to live in harmony is unfolding in another division in Mega

Summary Themes

- Knowledge assets are often a mess
- Hard to use, reuse, maintain and evolve
- Decompose into the essential components and using facet analysis
- Re-characterize the original asset so it is easier to use, maintain and evolve.
- Use the essential components as common building blocks to purpose-build other assets for other uses.

Summary Themes

- The common building blocks are linked to and/or become part of an enterprise wide ontology.
- The Enterprise Ontology has many uses:
  - reaching a common understanding
  - basis for semantic integration of heterogeneous knowledge and data assets (including countless spreadsheets)
  - supports automated inference for consistency, completeness and enhanced analytics

Summary Themes

- Most taxonomy work is about search and navigation
- We broadened it to help manage knowledge assets more generally, whatever their purpose.
- Improve understandability, use and reuse

Ontology vs. Taxonomy

- Most ontologists are not very interested in taxonomy
- Many traditional taxonomists don't understand ontology
- We are applying ontological analysis to design better taxonomies
- We find that both are critically important in the modern enterprise.
- Thus we have Strange Bedfellows…

One Happy Family of Models

- Ontology: best for core classes and properties
- Taxonomies, often faceted
  - for fine scale distinctions on the edges
  - to be governed by separate and sometimes external parties
- Data models and messages derived from the ontology, using fine grained distinctions from the taxonomies as needed.
- When Mega started doing this in their Enterprise…

Something Magic Happened

- Rather than the 1,000,000 concepts Mega had baked into all the schema of all their current systems
- Or the 100,000 elements they had captured in a metadata repository (so far)
- Or the 200,000 taxonomic distinctions they had either collected or subscribed to
- Or the 50,000 attributes they had in their fully attributed Entity Data Model
- Or the 20,000 elements they had in the sum total of all the messages in their SOA

It turned out…

- There were less that 1000 concepts that they ran their whole business on
- And of these 1000 there were 70 classes and 30 properties that shaped all other information
- Anyone who was a bit motivated could find they concepts they needed in this new simplified knowledge-scape

Semantic Computing Writ Large

Our focus today: taxonomy, ontology, the semantic web.

Early in 2014, Gartner called out the following as major trends: Cloud, Big Data and Semantics

Complementary technology:

- Machine learning
- NLP
- Big Data
- The Cloud

Big Semantic Cloud

Thank You

Our website: www.semanticarts.com

- We do consulting and training, specializing in helping large companies find their information core
- Leave me a card for a copy of this presentation
- Questions?