



Plan for Today's Lecture(s)

- Introduction to Classification
- Classification is Purposeful, Principled, and Biased
- Classification and Standardization



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

22 October 2013
Lecture 16.1 – Introduction to Classification



What is Classification?

- "Classification is a higher order thinking skill requiring the fusion of the naturalist's eye for relationships...
 - ...with the logician's desire for structured order...
 - ...the mathematician's compulsion to achieve consistent, predictable results...
 - ...and the linguist's interest in explicit and tacit expressions of meaning"
- Louise Gruenberg, "Faceted Classification, Facet Analysis and the Web"



Distinguishing Categorization and Classification (1)

- Categories are EQUIVALENCE CLASSES - sets of resources, processes, and events that we treat the same
- A Classification (noun) is a SYSTEM OF CATEGORIES, ordered according to a PRE-DETERMINED SET OF PRINCIPLES and used to organize a collection of resources
- Classification (verb) is the process of systematically assigning resources to intentional (often institutional) categories in a classification system



Distinguishing Categorization and Classification (2)

- The contrasts among cultural, individual, and institutional categories” yield a precise definition of classification:

The systematic assignment of resources to a system of intentional categories, often institutional ones

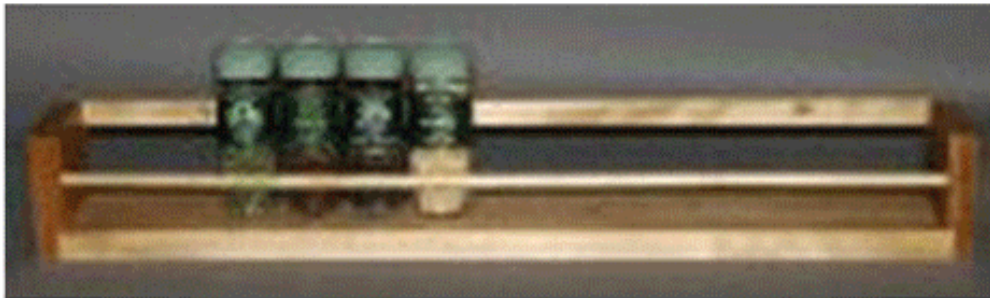
- This definition highlights the intentionality behind the system of categories, the systematic processes for using them, and implies the greater requirements for governance and maintenance that are absent for cultural categories and most individual ones.



Classification vs. Resource Arrangement

- With physical resources the end result of classification is their arrangement in some physical location
- In libraries these two activities are strongly associated
- We often describe the physical location using the category name
- But the physical location is often arbitrary and it is better to think of classification in purely logical terms, unless the categories are explicitly designed in terms of co-location (as they are in taskonomies...)

Separating the Organizing Principle From Physical Arrangement





Classification Schemes (1)

- A CLASSIFICATION SCHEME is a realization of one or more organizing principles
- An ENUMERATIVE scheme itemizes all possible categories to which resources can be assigned (into one and only one of the categories)
- Alphabetic or chronological ordering creates an implicit or latent scheme that arranges resources in a predictable place without the need to create the categories in advance



Classification Schemes

- A HIERARCHICAL or TAXONOMIC scheme emerges when multiple resource properties are used by organizing principles; each property creates another level
- A scheme can be both HIERARCHICAL and ENUMERATIVE at the lowest level where resources are categorized
- A FACETED classification scheme uses multiple resource properties, but does not require every resource to have a value for every property and allows the properties to be considered in any order



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

22 October 2013
Lecture 16.2 – Classification is
Purposeful, Principled, and Biased



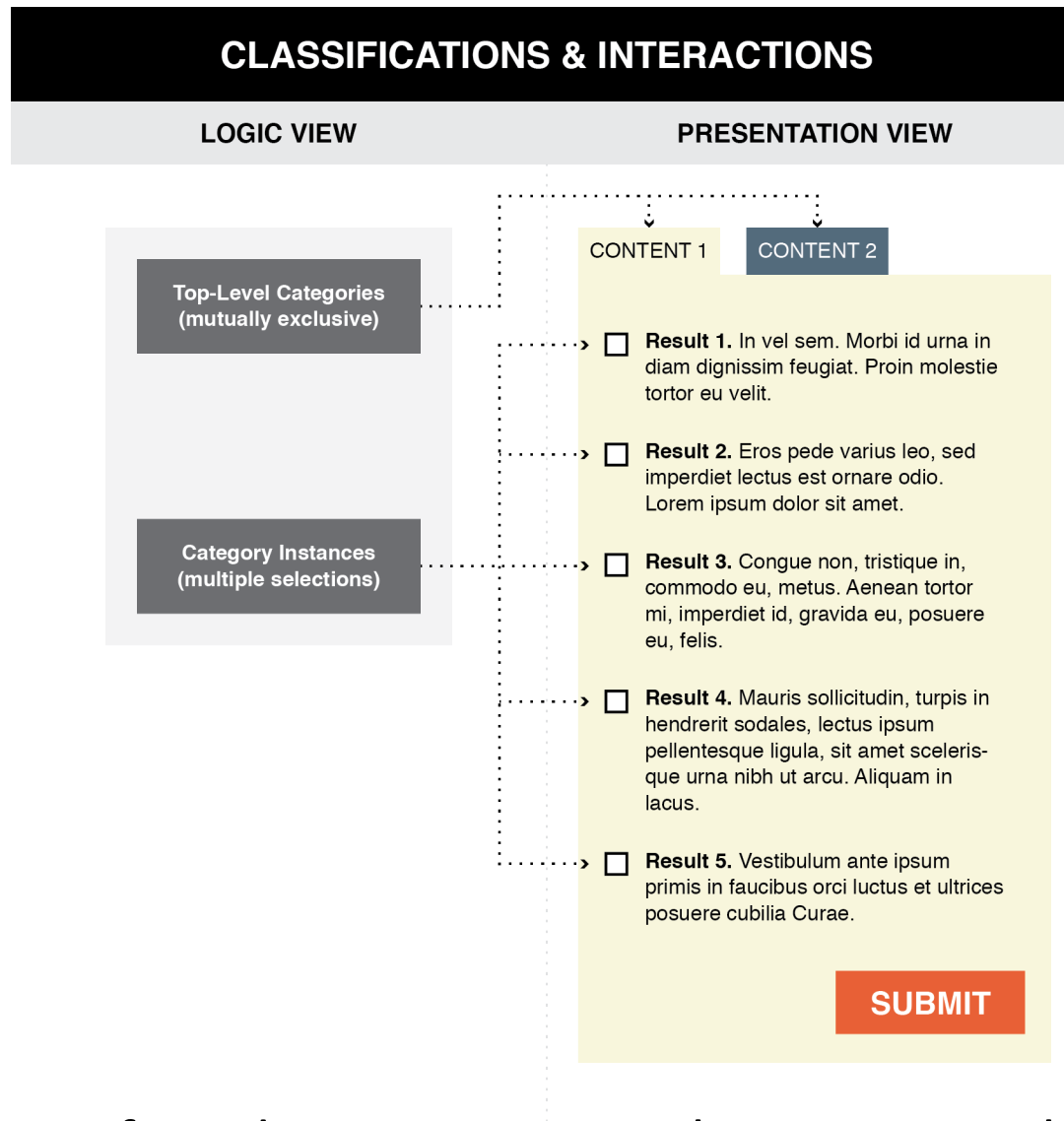
Classification Is Purposeful

- Classifications arrange resources to provide structure for and increase the precision of interactions in organizing systems
- Classifications are essential in useful work
- They serve as a reference model (a semantic road map) to individual domains and the relationships among them to help people understand the concepts and relationships
- They improve learning and communication



Classification and Interactions

- With physical resources, activity-based classification makes co-location more useful
- Effective user interfaces to physical or digital systems present and reinforce the logical classifications embodied in the system
 - ... while poor user interfaces don't
- Standard classifications for products and business services make it easier to discover suitable resources and integrate them into your own processes



Good user interface design creates a clear mapping between a classification scheme and the arrangements and interactions that users see (TDO, Figure 7.1)

Clear Communication of a Classification Scheme



Photo taken by Bob Glushko at Union Square Greenmarket, NYC

Poor Communication of a Classification Scheme

[BART Station List](#)

12th St. Oakland City Center	Lafayette
16th St. Mission (SF)	Lake Merritt (Oakland)
19th St. Oakland	MacArthur (Oakland)
24th St. Mission (SF)	Millbrae
Ashby (Berkeley)	Montgomery St. (SF)
Balboa Park (SF)	North Berkeley
Bay Fair (San Leandro)	North Concord/Martinez
Castro Valley	Orinda
Civic Center/UN Plaza (SF)	Pittsburg/Bay Point
Coliseum/Oakland Airport	Pleasant Hill/Contra Costa Centre
Colma	Powell St. (SF)
Concord	Richmond
Daly City	Rockridge (Oakland)
Downtown Berkeley	San Bruno
Dublin/Pleasanton	San Francisco Int'l Airport
El Cerrito del Norte	San Leandro
El Cerrito Plaza	South Hayward
Embarcadero (SF)	South San Francisco
Fremont	Union City
Fruitvale (Oakland)	Walnut Creek
Glen Park (SF)	West Dublin/Pleasanton
Hayward	West Oakland

Better Might Be Something Like...

San Francisco

Embarcadero

Montgomery

Powell

Civic Center

...

Oakland

West Oakland

12 Street

19 Street

MacArthur

Rockridge

...

Berkeley

Ashby

Berkeley Downtown

Berkeley North



Classification is “Principled”

- The principles for defining categories (enumeration, properties, similarity, family resemblance, etc.) are embodied in the classifications that use these principles
- But the emergent classification system follows other principles about its purposes, scope, scale, intended lifetime, extensibility, and other considerations
- Being "principled" means that these design choices are made they should be consistently followed



Principles Embodied in the Classification Scheme

- Warrant: What is the justification for the choice of categories and their names?
 - Literary Warrant: Classify only the resources we have?
 - Scientific Warrant: Use expert categories and names
 - Use Warrant: Use categories and names from “ordinary” people
- Breadth and depth of classification hierarchy
- Degree of enumerativeness



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Principles for Assigning Resources to Categories

- Hierarchy
- Uniqueness



Hierarchy and Uniqueness Principles

- Deeply engrained in our intellectual and physical worlds
- Each category is successively divided into smaller subdivisions
- Every level of the hierarchy is divided according to a "feature" or "character of division"
- Every resource is classified in only one subdivision



The "One and Only One Place" Rule

- If each resource goes in one and only one category it won't always be easy to decide what metadata or property determines the categorization
- For example, a book's title is often a clue to the subject, but should never be the only thing analyzed
 - *Who Moved My Cheese?* is a work on coping with change, not a work related to the culinary arts
 - Should *Leaves of Grass* be assigned to the "Gardening" category?
- Needless to say, the rules for categorizing books take a long time to learn and are not always easy to apply



Principles for Maintaining the Classification

- Lifetime management
- Category integrity: Classify once and for all, or change categories and assignments as we learn?
- Flexibility / extensibility / hospitality



"Principled" Maybe ... But Every Classification is "Biased"

- Every classification is arbitrary to some extent because the criteria used to establish the categories reflect a point of view or perspective on the domain that intentionally excludes all other perspectives
- There is no "one true way" about how things go together
- Every classification system implicitly or explicitly distinguishes between "good" or "standard" and "bad" or "nonstandard" ways of understanding things
- [Borges](#): “there is no classification of the Universe that is not arbitrary and full of conjectures”



Classification Biases

- Natural vs. man-made disaster?
- Welfare vs. entitlement?
- Taxes vs. government investments?
- Illegal vs. undocumented immigrants?

CAFE Categories: Find the Truck(s)

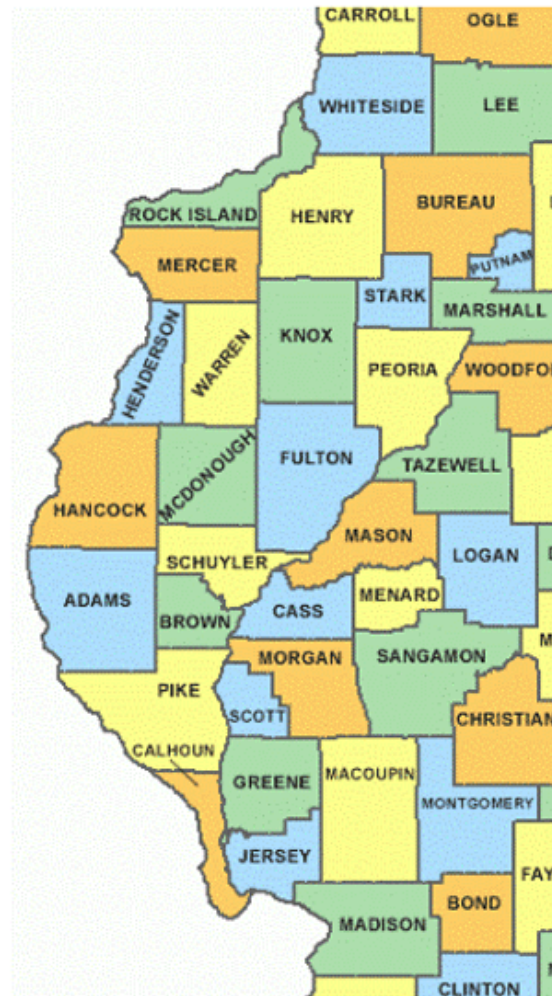
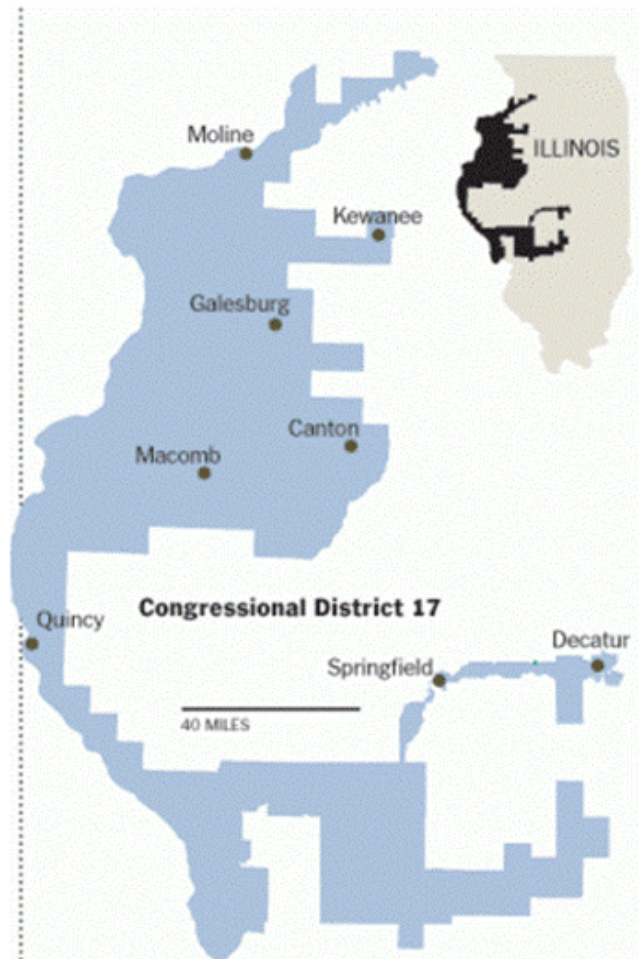


Beverage Categories: Find the Beer



Gerrymandering: Find the Democrats

Tilting an Election Through Redistricting – NY Times, 9/25/2010



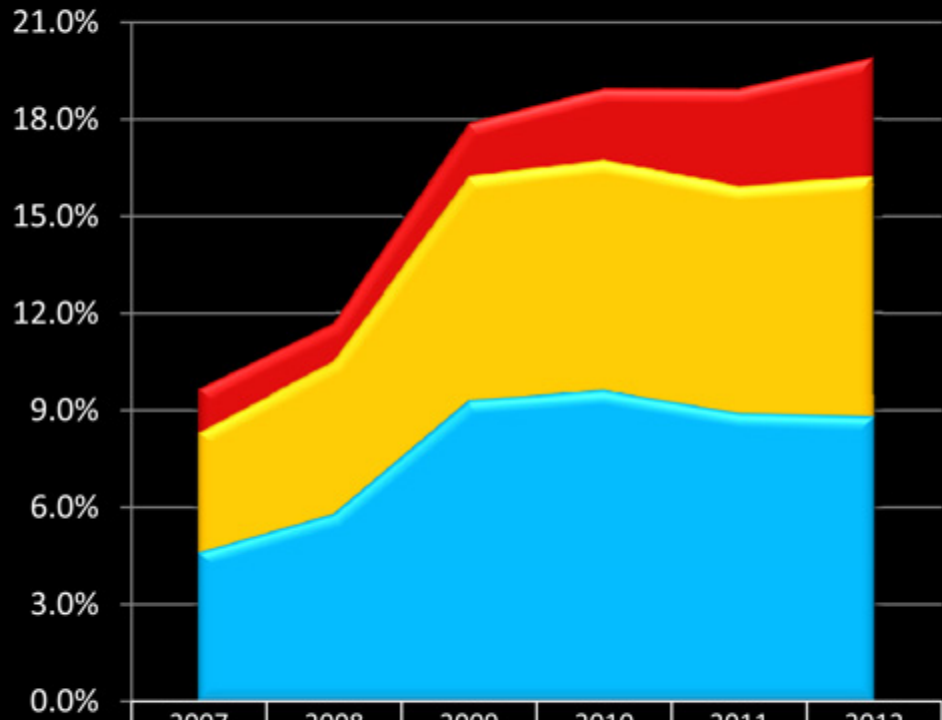
[Illinois is famous for its corrupt politicians....](#)



Classification Biases

- Minority, underrepresented minority
- Coup or regime change?
- Dietary supplement vs. beverage?
- Employee vs. contractor, full vs. part time?
- Unemployed

Graph B
Constant Workforce Participation
Unemployment Rate
 All numbers annual except 2012



■ Constant Workforce Participation	9.6%	11.7%	17.9%	18.9%	18.9%	19.9%
■ Full (U-6)	8.3%	10.5%	16.2%	16.7%	15.9%	16.2%
■ Headline (U-3)	4.6%	5.8%	9.3%	9.6%	8.9%	8.8%

**Making
 9 Million
 Jobless
 People
 Vanish**

[When you give up looking for a job, you're not unemployed](#)



Some More Examples...

- How do airlines and transit authorities define “on time” and “late” with respect to their scheduled times?
- Utilities report the number of “customers” affected by power failures... but a “customer” is a bill payer, not a person
- If a season ticketholder doesn’t show up, are they still counted in the “paid attendance” category?
- "Response time" to a fire is the time it takes for the first emergency vehicle to arrive, even if it isn't a fire truck



How Psychiatry Went Crazy

Tavris, Carol. How Psychiatry Went Crazy. Wall Street Journal, 17 May 2013

- DSM 1 in 1952 had 11 categories; new one has several hundred
- Why has the number of categories increased so much?
- How are the categories used? Are the categories “good ones” that can be used reliably?
- Why the hurry to revise?



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202

“Information Organization & Retrieval”

Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

22 October 2013
Lecture 16.3 – Classification and Standardization



Classification and Standardization (1)

- Classifications and standards both impose order on resources
- They both distinguish, explicitly or implicitly, between standard / appropriate / effective and nonstandard / inappropriate / ineffective ways of creating organizing, and using resources
- But this does not imply that a standard is a good one or that the best one will win a "standards war"



Classification and Standardization (2)

- Some classifications become standards
- Every successful standard imposes a classification system
- Standards and classification systems can have significant inertia, becoming part of the "installed base" or "infrastructure" and difficult to change



"Specifications"

- A specification is an explicit and detailed description about how something (a product, system, service, or practice) works that might include:
 - the definitions of the key concepts or terms used in the description
 - the materials from which it is made
 - the components (tangible or intangible) that are arranged or assembled to make it work
 - the methods or processes by which it operates
 - the inputs and outputs for each method or process



"Standards"

- A standard is a published specification that is developed and maintained by consensus of all the relevant stakeholders in some domain
 - following a defined and transparent process
 - usually under the auspices of a dedicated organization with the authority to create standards
- The rigor and transparency of the process by which a standard is created is essential to its legitimacy
- Some governments or other large procurers or products of systems require that they be based on these "de jure" standards



Why Standards Exist (1)

- Standards serve as a coordinating mechanism whenever there are alternative ways of doing or making something that might be incompatible or that otherwise wouldn't be possible
- Standards can also serve as a regulating mechanism to restrict behavior or operation that might be unsafe, unfair, or otherwise "harmful"



Why Standards Exist (2)

- Many standards define technical specifications and procedures to ensure a "common design" or "interoperability"-- TDO calls these "institutional semantics"
- Standards define institutional categories about models, behaviors, or processes - TDO calls these "institutional taxonomies" (like the UNSPSC)



Who Sets Standards? (1)

- Some technical standards are set by governments in order to protect the public interest
- Other standards are set by organizations created precisely for the purpose of setting standards
- Other standards are set by groups of companies who may or may not establish an ongoing organization and governance procedures to maintain and extend a standards "family" around the initial standard



Who Sets Standards? (2)

- Other "standards" begin as proprietary vendor specifications that will become "de facto" standards if they are widely adopted and achieve market dominance
- Often "de facto" standards are submitted to a standards organization with the goal of "rubber stamping" them as "de jure" ones



Specifications {and,or,vs} Standards

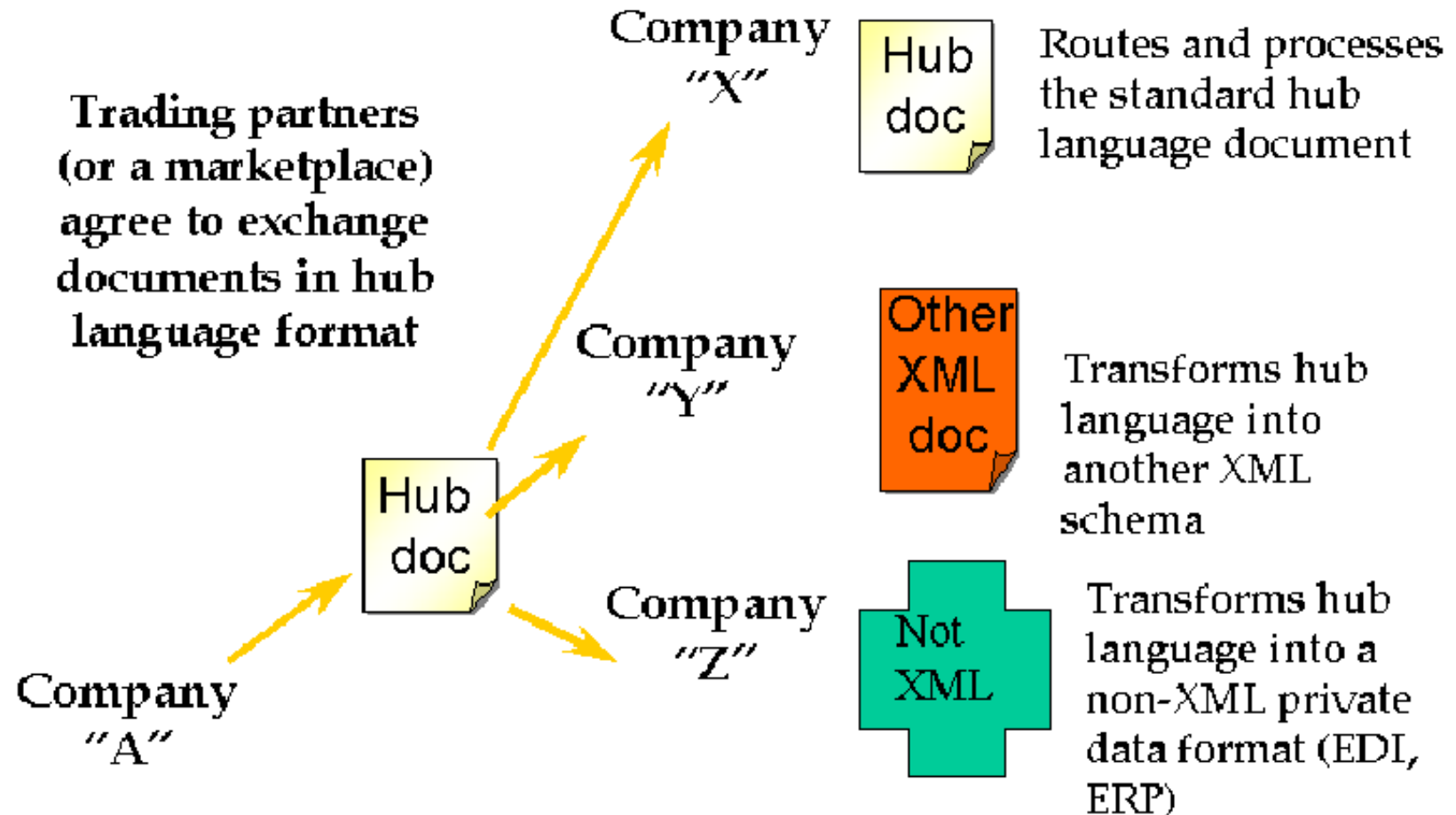
- "All standards are specifications, but not all specifications are standards"
- Any person, firm, or ad hoc group can create a specification and then use it themselves, try to get others to use it, or prevent others from using it
- In contrast, standards are sometimes created by people or groups who aren't themselves planning to use them, but standards are generally made available for anyone's use at no cost or under "reasonable and non-discriminatory" terms
- But the widespread adoption (or "traction") of a specification is often more important than whether or not it was created by a "standards" organization



Standards and Metcalfe's Law

- Standards are especially important in industries or domains that have significant "network effects" where the value of a product or service depends on the number of interoperable or compatible
- These contexts are sometimes called "winner take all" markets, which is why standards wars take place - because of the economic implications
- How do you encourage and enable others to conform to a standard?

Interchange / Hub Languages





Mapping in and out of Hub Language

- If all parties/applications/services rely on a hub language for their external interfaces, an exponential interoperability challenge becomes a linear one
- Mapping tools for transforming instances from an internal information model to another one are ubiquitous as standalone tools and as parts of application servers
- EXAMPLE: [Altova MapForce](#)



Using Standards in Your Information Organization Efforts

- You may be required by the government, by your customers, or by "business partners" to conform to some standards for information organization
- But to the extent that standards embody best practices, it is a good idea to follow them anyway
- Nevertheless, this is "mother and apple pie" advice because of the proliferation of standards, their sometimes competitive and sometimes complementary character



Standards for Document Description in Organizing Systems (1)

- For specific vocabularies or "domain-specific languages" (e.g., HTML, UBL, SPL, DocBook, DITA, EPUB, SVG)
- For schema languages that formally define the domain-specific languages (e.g., XSD)
- For "profiles" that suggest modeling practices like degree of abstraction and granularity, or constrain the possible values for some descriptor
- For the metalanguages used to encode the DSLs and schema languages (e.g., XML, SGML)



Standards for Document Description in Organizing Systems (2)

- For the construction of file and document names
- For identifying resources and namespaces
- For specifying versions, stages or revisions of documents or other resources



Other Standards for Organizing Systems

- Metadata standards (MARC, Dublin Core, EAD, METS, RDF, OWL)
- Standards for classification (LoC, DD, BISAC)
- Standards for thesaurus construction
- Standards for processes (WS-*)
- Standards for communication protocols



Readings for Next Lecture

- TDO rest of chapter 7