UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

12 September 2013
Lecture 5.1 – Resources and Resource Identity

Plan for Today's Lecture(s)

- Resources and resource identity
- Four distinctions about resources
- An introduction to conceptual modeling

What is a Resource?

- When you look at, refer to, or pick up a natural physical object, it usually seems that it is "one thing"
- Many things have highly visible parts or internal structures, and we must consider whether to treat them as assemblies or aggregations of these separate things
- For digital and computational resources, it is much harder to know how many parts we can or should identify
- Once you've decided what to treat as a resource, you can create a name or identifier so you can refer to it

The Unit of Analysis

- Identifying the unit of analysis is a central problem in every intellectual or scientific discipline - and in every organizing system
- But if the basic unit of an organizing system isn't usually very clear or naturally given to us, we need systematic methods for determining which resources will have separate identities and how they are related to each other

"Carving Nature at its Joints"

- Over 2000 years ago Plato (in Phaedrus) has Socrates arguing that species are distinguished by "carving nature at its joints" - where the natural differences between things are the largest or most salient or where there are few connections
- "Joints" and the boundaries they create between things exist to some extent with physical resources and are manifested in their surface or perceptible properties

Carving Information at its Joints?

- When we deal with resources that are made of information, any "joints" are man-made and arbitrary
- There might be indications "on the surface" that suggest the "joints" between information resources, but these represent design decisions, might be more structure or presentation oriented than content oriented distinctions, and are subject to cognitive interpretation

"Thing" vs "Type of Thing"

- Oops... we have been blurring the distinction between individual things or instances of things and classes of things
- We often say that that two objects are the "same thing" when we mean they are the same "type of thing"
- Identifying a resource as an instance is not the same as identifying the category or "equivalence class" to which it belongs

Two Aspects of “Thingness”

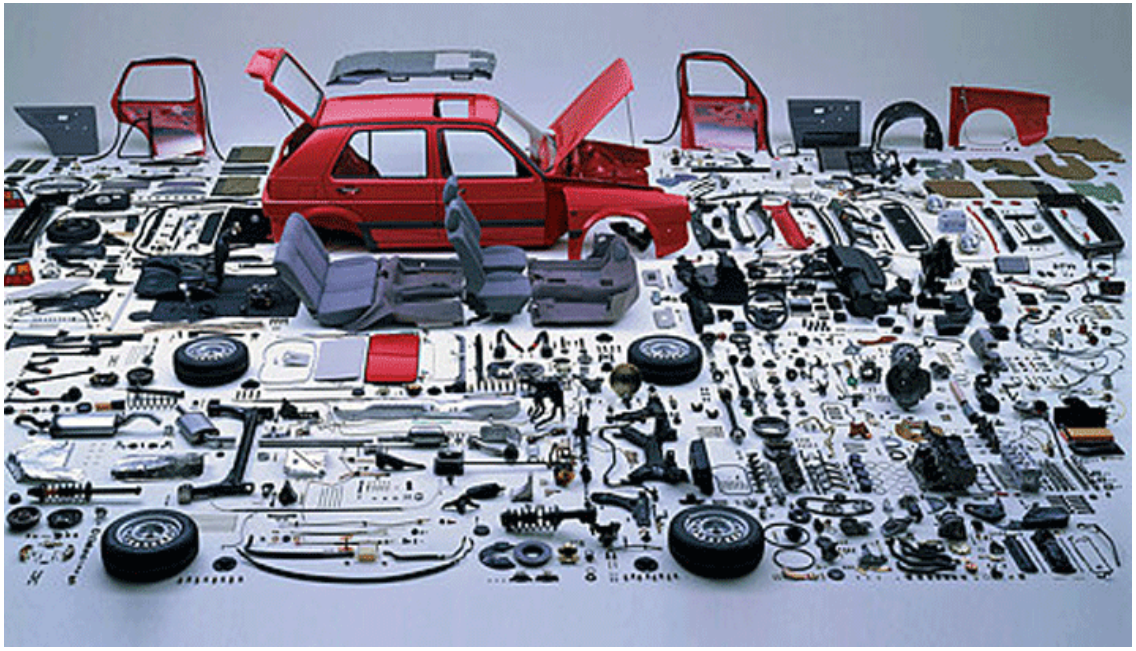
- Two separate aspects cut across the "thingness" distinctions:
  - Granularity
  - Abstraction

Granularity: Whole vs Parts

- Resources that are aggregates or composites of other resources, or that have internal structure, pose questions about the granularity of their "thingness"
- We might need to organize and manage the granular resources, the composite resources, and the relationships between them - all at the same time
- The "part things" can be identified at various levels of contexts/containers/collections in which they occur
- The granularity of physical resources is more easily determined that for information resources

Red Car as 1 Resource

Red Car : Not 1 Resource



Abstraction - 1

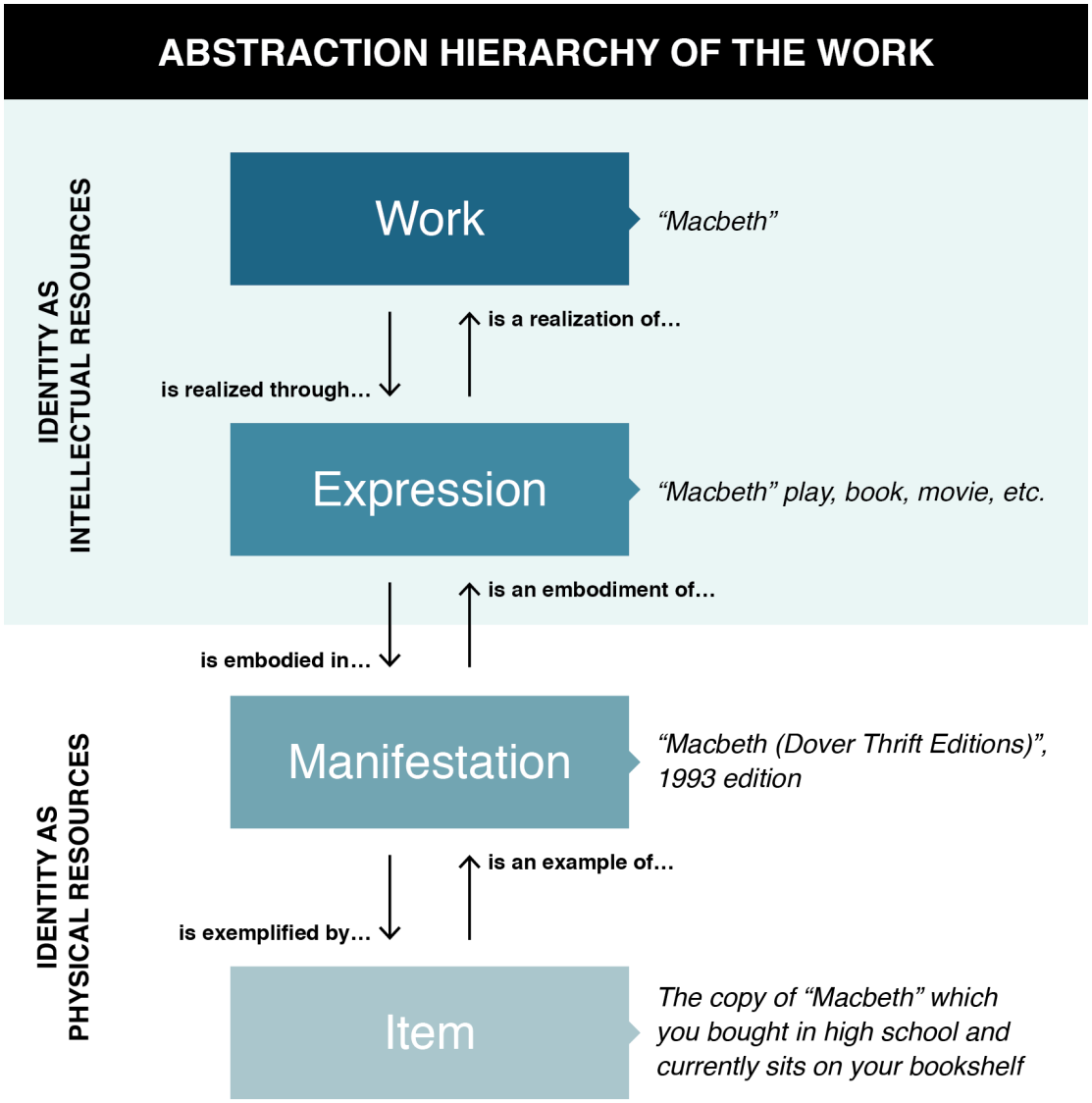
- We can identify a resource as a unique instance or as a member of a class of resources
- The size of this class - the number of resources that are treated as equivalent - is determined by the properties or characteristics we consider when we examine any instance
- The choice of these properties depends on context and intent, so the same resource can be identified abstractly in some situations and very concretely in others

Abstraction - 2

- For tangible resources, there may be "natural" levels of abstraction (See Plato…)
- For intangible resources, there aren't

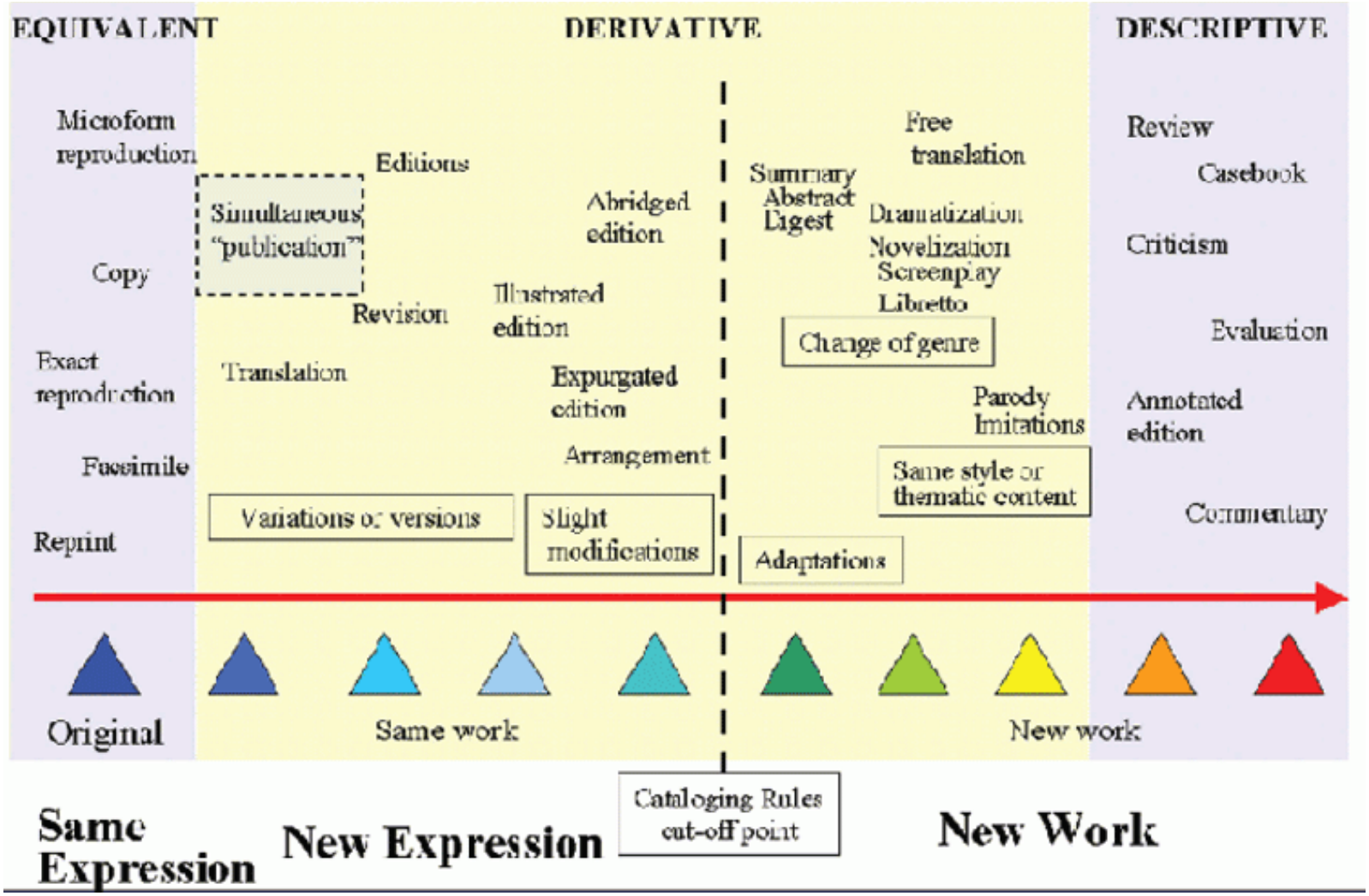
The "Abstraction Hierarchy" of the "Work"

- WORK - an abstract entity; the distinct intellectual or artistic creation; it has no single material manifestation
- EXPRESSION - the multiple realizations of a work in some particular medium or notation, where it can actually be perceived
- MANIFESTATION - each of the formats of an expression that have the same appearance; but not necessarily the same implementation
- ITEM - a single exemplar of a manifestation; if we distinguish this level it is because otherwise identical manifestations have some differentiation



What is Macbeth?

The Equivalence Continuum (Tillett)



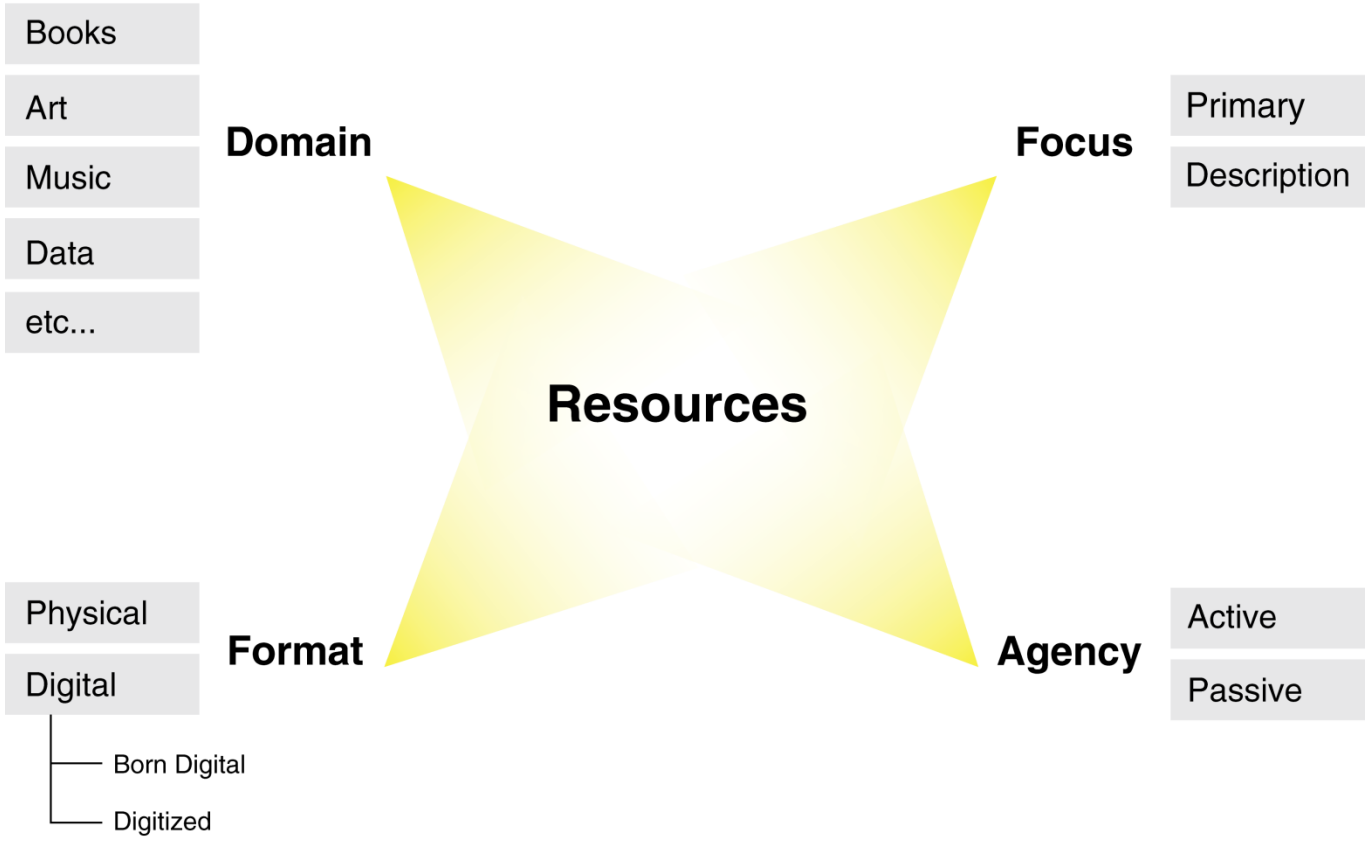
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

12 September 2013
Lecture 5.2 – Four Distinctions About Resources

Four Distinctions About Resources



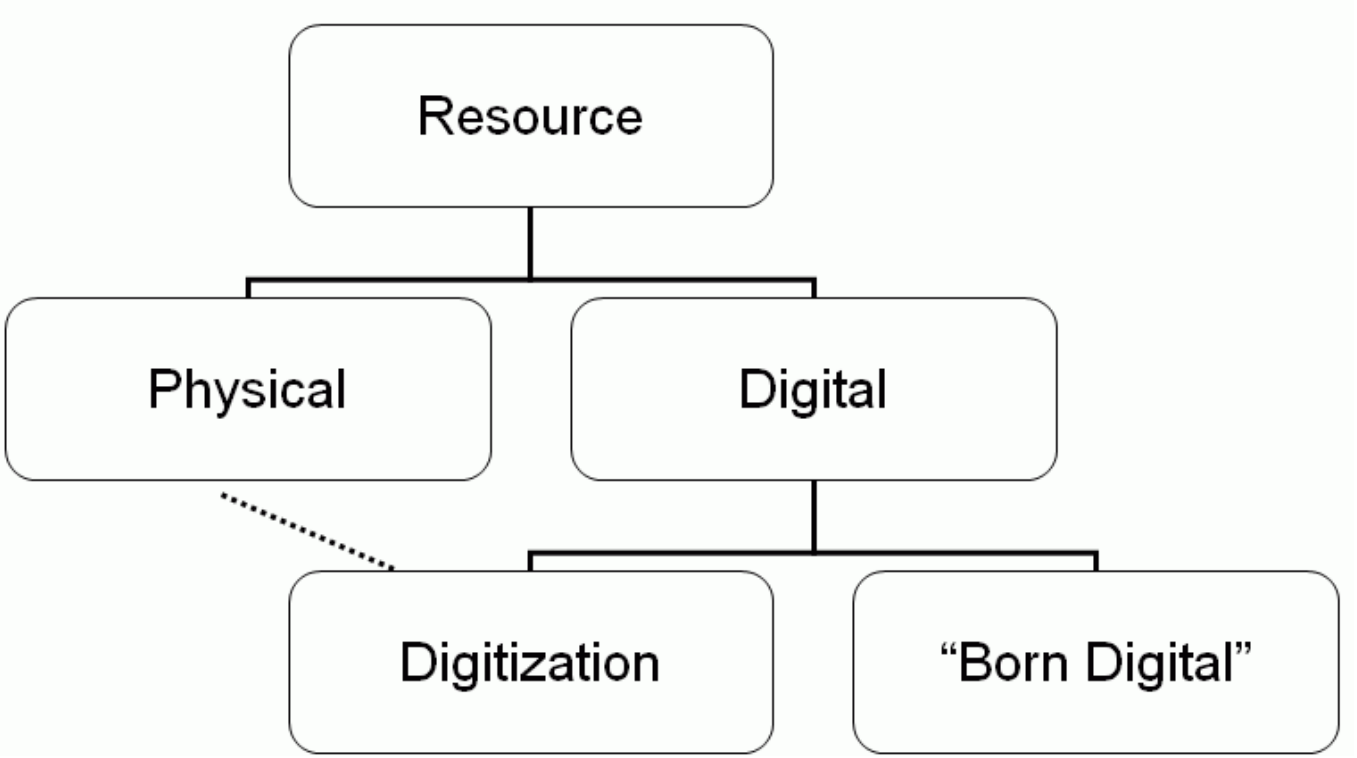
Resource Domain

- Physical resources - often distinguished by material of composition and other visible properties
- Bibliographic resources - traditional focus of library and information science; most commonly used cataloguing rules have 11 categories that don't cleanly distinguish semantics and formats
- Information resources - the content of information-intensive artifacts and applications - "Document types"

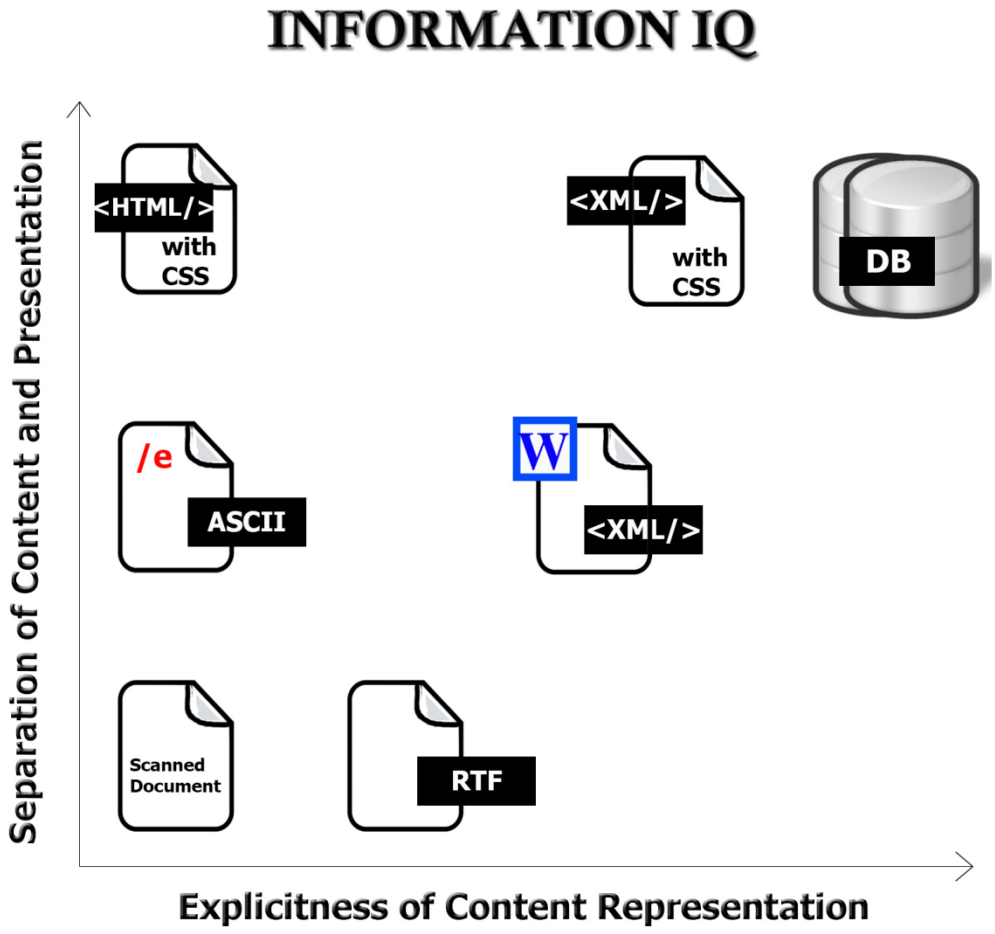
Resource Format

- In Chapter 3 of TDO
- Bits vs Atoms (Negroponte)
- Three orders of order (Weinberger)

Resource Format in TDO



Format Matters!



Bits vs Atoms (Negroponte)

- A contrast between information as "bits" and information as "atoms" was first expressed by Nicholas Negroponte of the MIT Media Lab (see Being Digital (1995)
- Information encoded as bits can move several orders of magnitude faster than atoms can
- It can be in many places at once (broadcasting, networking) unlike atoms that have to be in one place
- ...and there are many other ways in which bits and atoms differ

Movement of Bits != Movement of Atoms

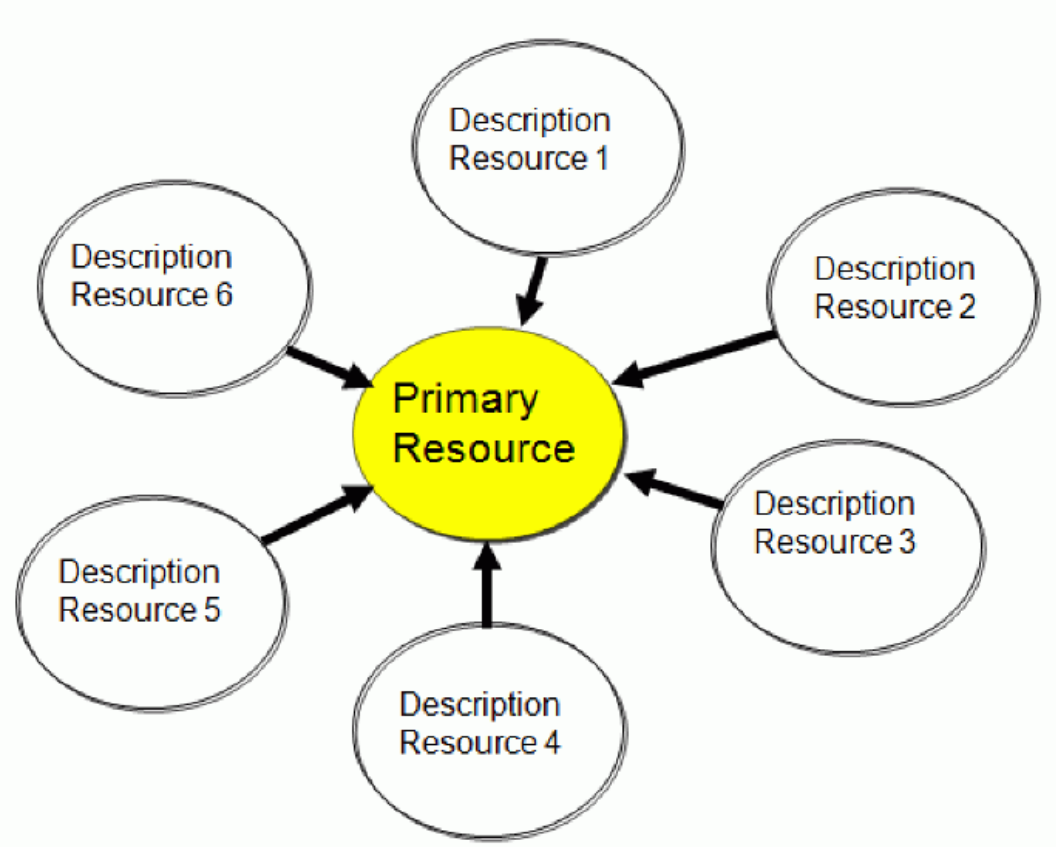
Three Orders of Order (Weinberger)

- An incremental extension of Negroponte's thinking is in Weinberger's Everything is Miscellaneous (2007)
- First order: Organize physical things themselves
- Second order: Organize physical surrogates of things
- Third order: Organize digital surrogates of things, or digital things themselves

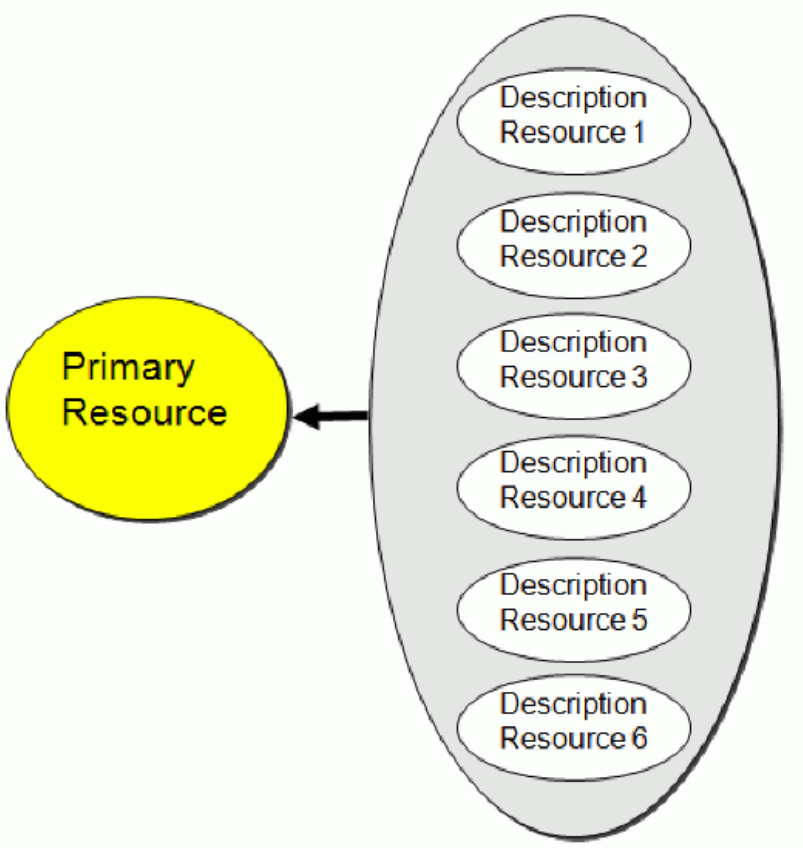
Resource Focus

- Any resource can have other resources associated with it, usually to describe it in some way to facilitate finding it, interacting with it, or interpreting it
- So we can designate a resource as a primary one or as a description resource
- Description resources are often called METADATA

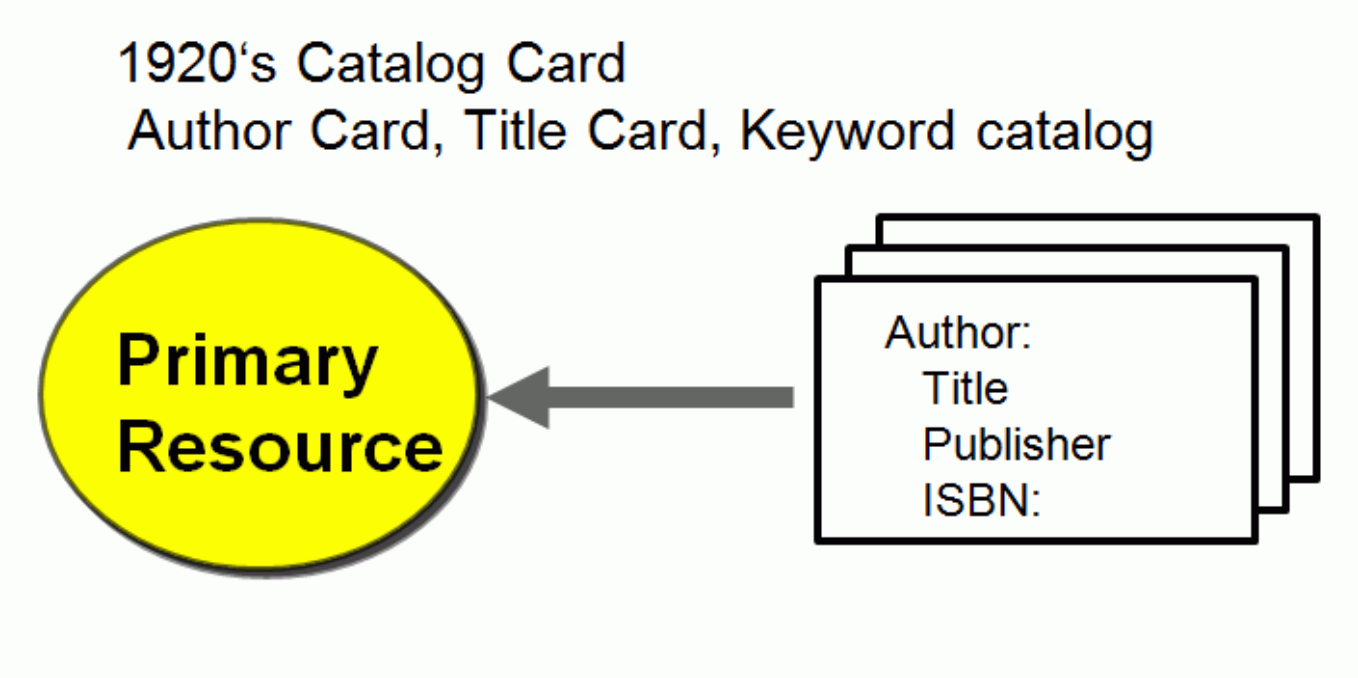
Description Resources



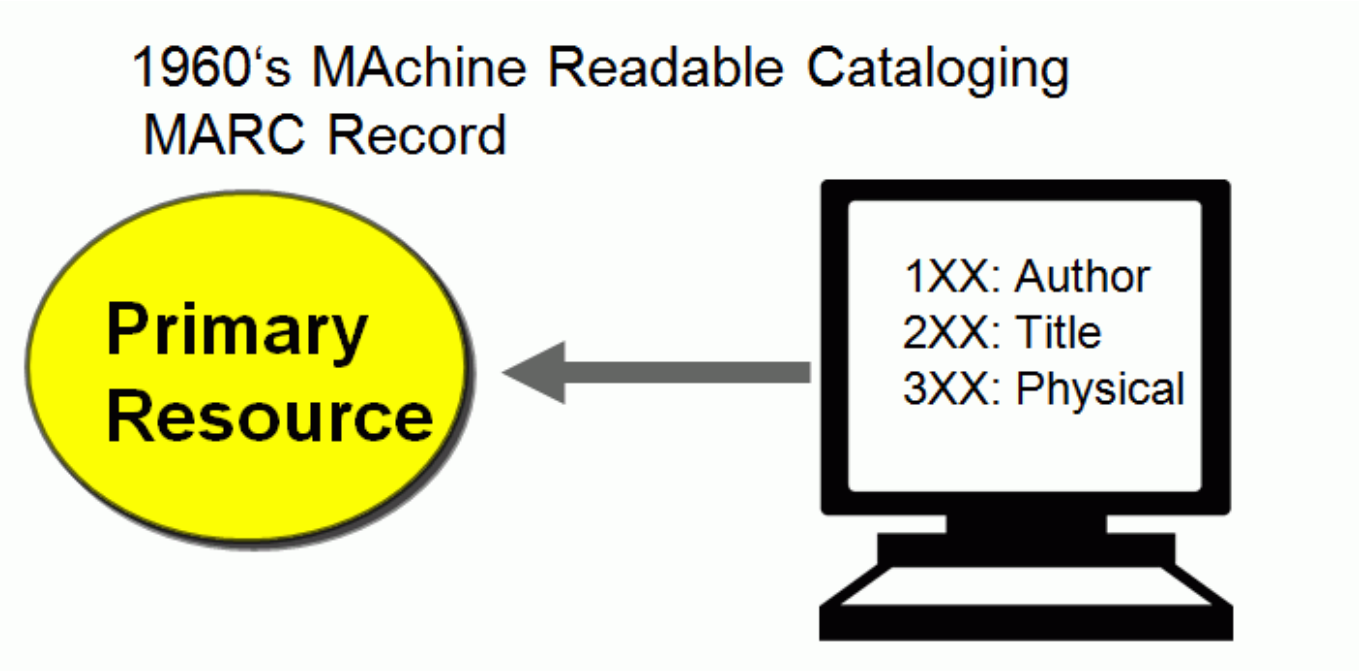
Aggregated Description Resources



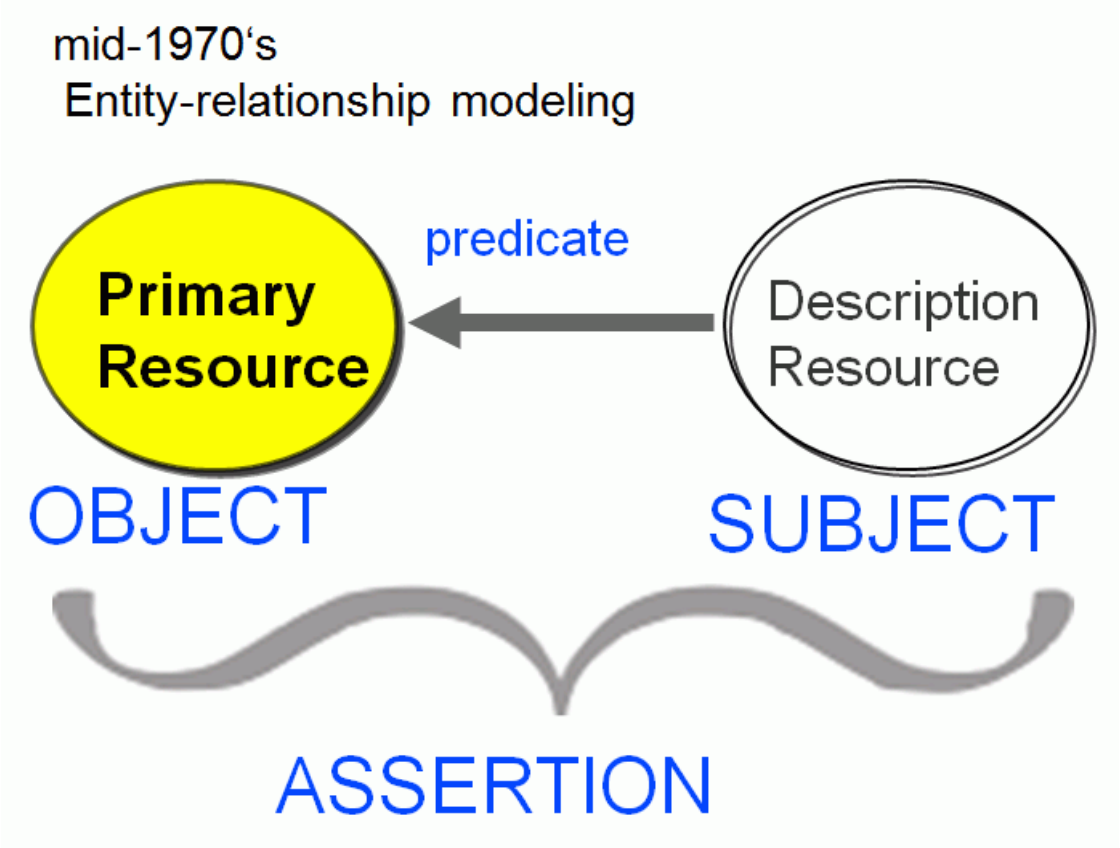
Bibliographic Description Resources - 1920



Bibliographic Description Resources - 1960



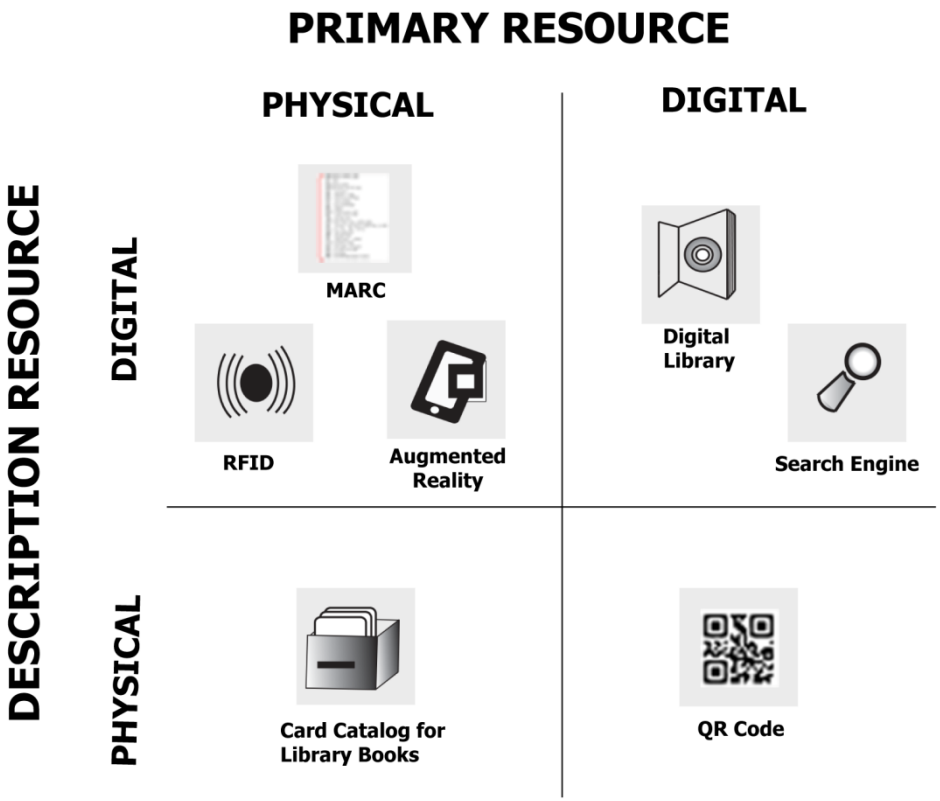
Resource Descriptions as "Triples"



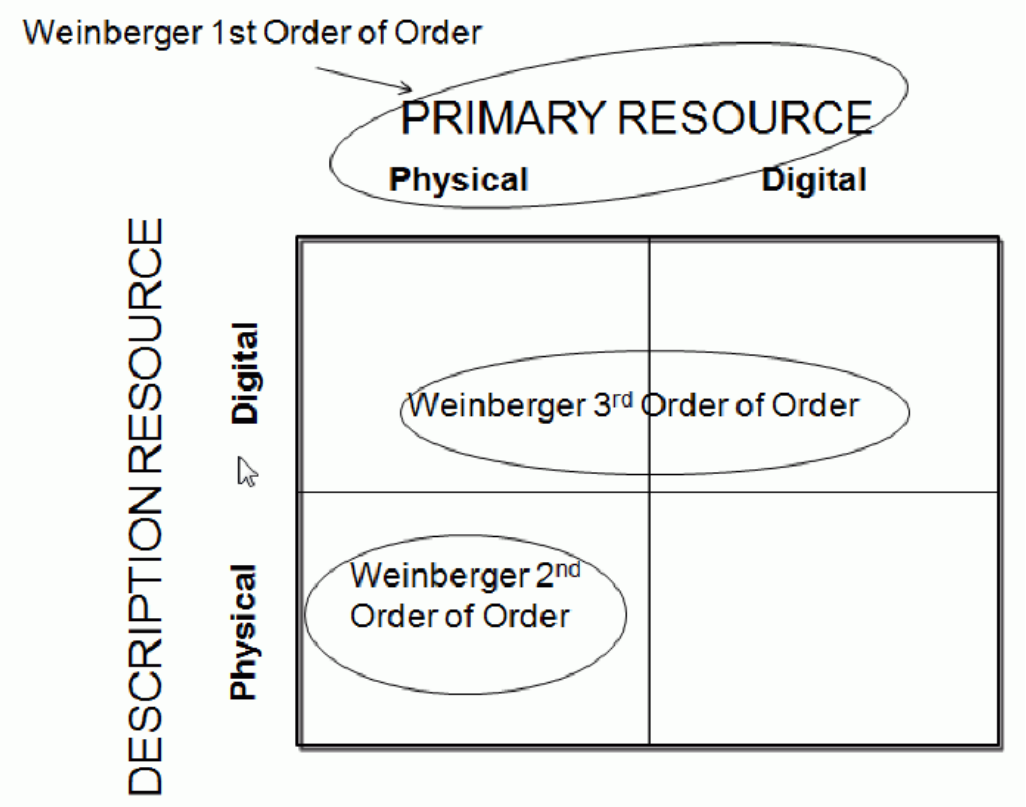
Resource Focus

- In the bibliographic domain the resource vs. description distinction is deeply embedded
- It other domains it can be quite arbitrary and "focus" is an important decision
- One person's data is another person's metadata -- and now substitute "process" for "person"

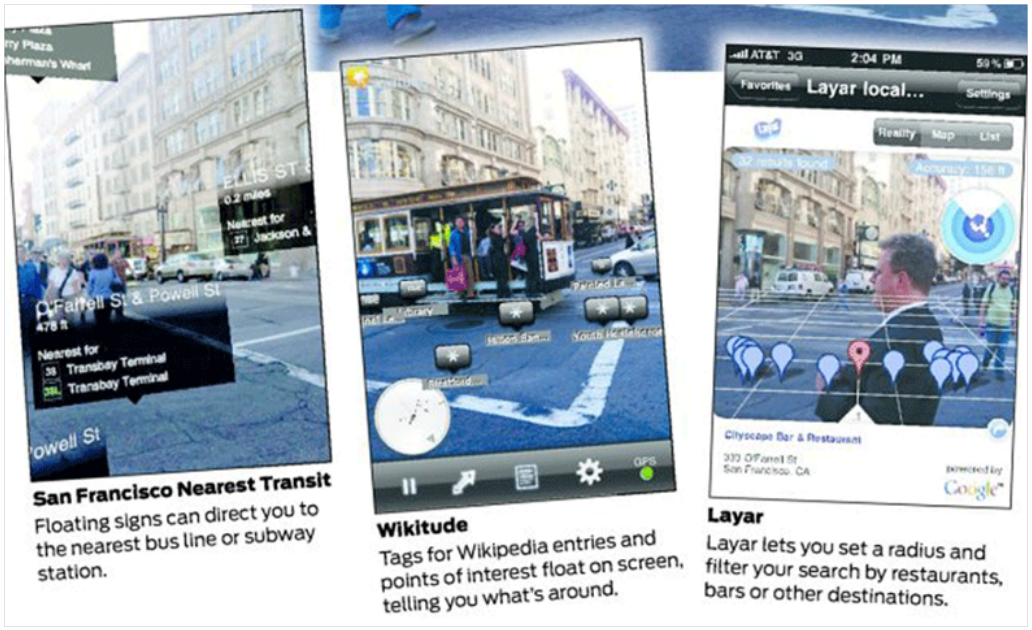
Format x Focus



Format x Focus & Weinberger



Augmented Reality – Overlay Digital on Physical



San Francisco Nearest Transit
Floating signs can direct you to the nearest bus line or subway station.

Wikitude
Tags for Wikipedia entries and points of interest float on screen, telling you what's around.

Layar
Layar lets you set a radius and filter your search by restaurants, bars or other destinations.

QR Code – Connect Physical to Digital

http://qrcode.kaywa.com/

Resource Agency

- Passive or operand resources ("nouns") must be acted upon or interacted with to produce an effect
- Traditional organizing systems contain tangible and static resources that are "natural operands"
- Active or operant resources ("verbs") create effects or value on their own, sometimes when then initiate interactions with operand resources.
- Sensors, web-based services, information feeds are operant resources that often are combined to implement business processes or business models

Smart Things

- Sensors, RFID tags, GPS, other computation and communication built into physical objects (and phones) to identify them or their context
- Technical challenges in sorting and identifying the objects in the "information torrent" especially when the objects exhibit some intentional or discretionary behavior
- In an "Internet of Things" any object can have an IP address that uniquely identifies it

Tracking the Food Supply

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

12 September 2013
Lecture 5.3 – Introduction to Modeling

An Introduction to Modeling - 1

- Models are *simplified descriptions* of a subject that abstract from its complexity to emphasize some features or characteristics while intentionally de-emphasizing others
- Models enable us to describe and communicate systems regardless of the specific domain or discipline that they represent

An Introduction to Modeling - 2

- A model can represent a human activity, a natural system, or a designed system
- We can model structures – objects, their characteristics, their static relationships with each other like hierarchy, and reference
- We can model functions, processes, behaviors – dynamic activities that create and affect structures

Recipes as Everyday Models

- A recipe describes both objects and structures (ingredients) and the processes (instructions) for creating a food dish
- You can FOLLOW the recipe to create the dish
- You can COMMUNICATE the recipe to someone else who can then create the same dish
- You can use the recipe as a GUIDE FOR EXPERIMENTATION with the objects or the processes in the recipe to create alternative dishes

Not a Model of Wonton Soup

A Model of Wonton Soup

Wontons

½ lb. ground chicken,
veal or lean pork
2 green onions, finely chopped
2 tsp. grated fresh ginger root
1 tsp. cornstarch (corn flour)
1 tsp. lite soy sauce
6 drops hot pepper sauce
18 wonton wrappers

Soup

4 cups chicken broth
2 tsp. lite soy sauce
1 tsp. sesame oil
¼ tsp. hot chili paste
Salt and freshly ground black pepper
1 cup shredded fresh spinach
or watercress
1 green onion, thinly sliced

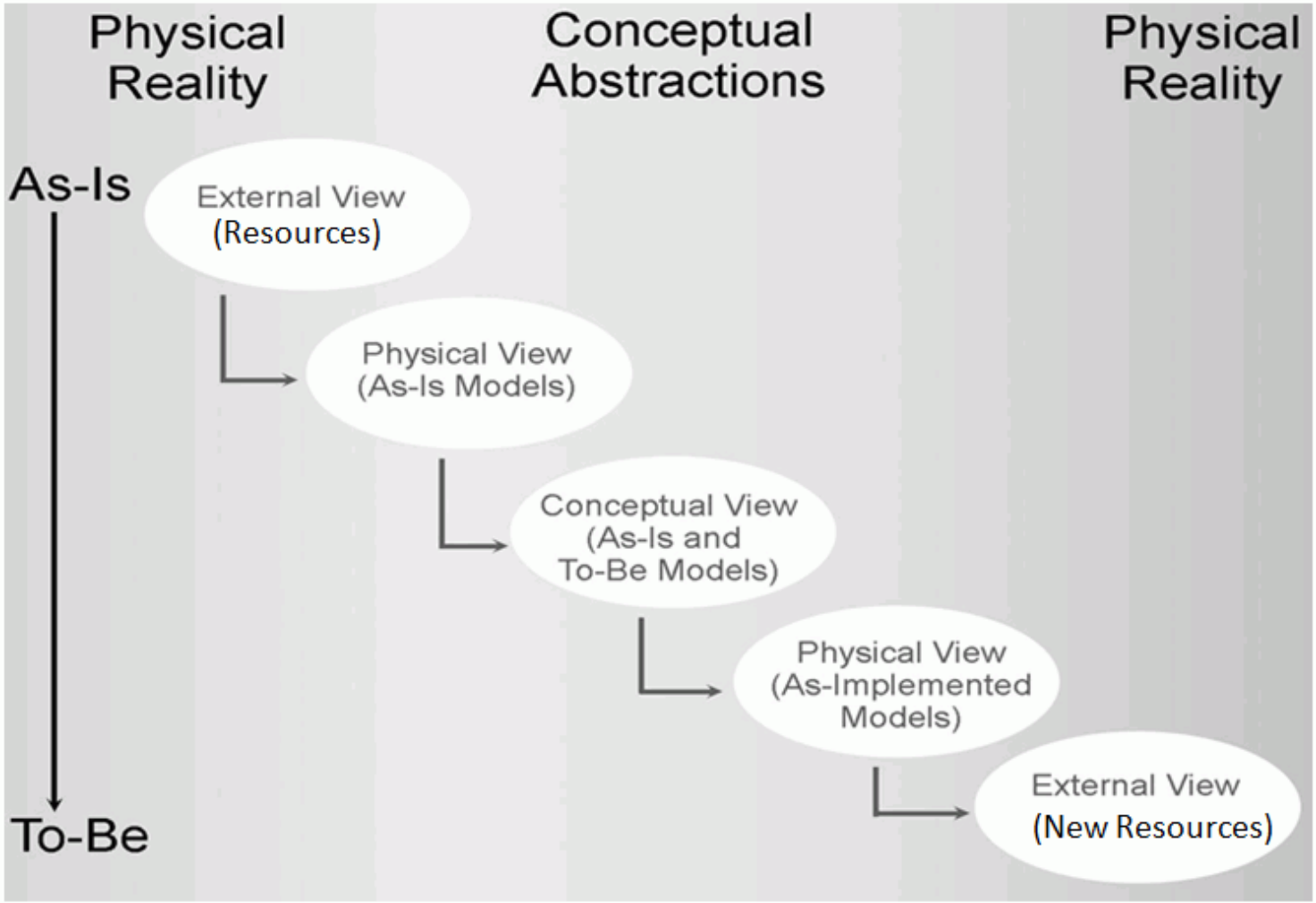
In bowl, combine ground chicken, green onions, ginger root, cornstarch, soy sauce and hot pepper sauce; mix until well blended. Lay wonton wrappers on work surface. Place spoonful of ground chicken mixture in center of each one.

Moisten edges with water. Fold each wrapper in half to form triangle; press edges to seal. Arrange in steamer basket. Cover and steam over boiling water for 3 minutes.

Meanwhile, in large saucepan, combine broth, soy sauce, sesame oil, and hot chili paste. Bring to boil. Season to taste with salt and pepper.

Just before serving, stir in spinach until wilted. Place 3 wontons in each bowl. Ladle soup over top. Garnish with green onion rings.

The Classical Modeling Approach



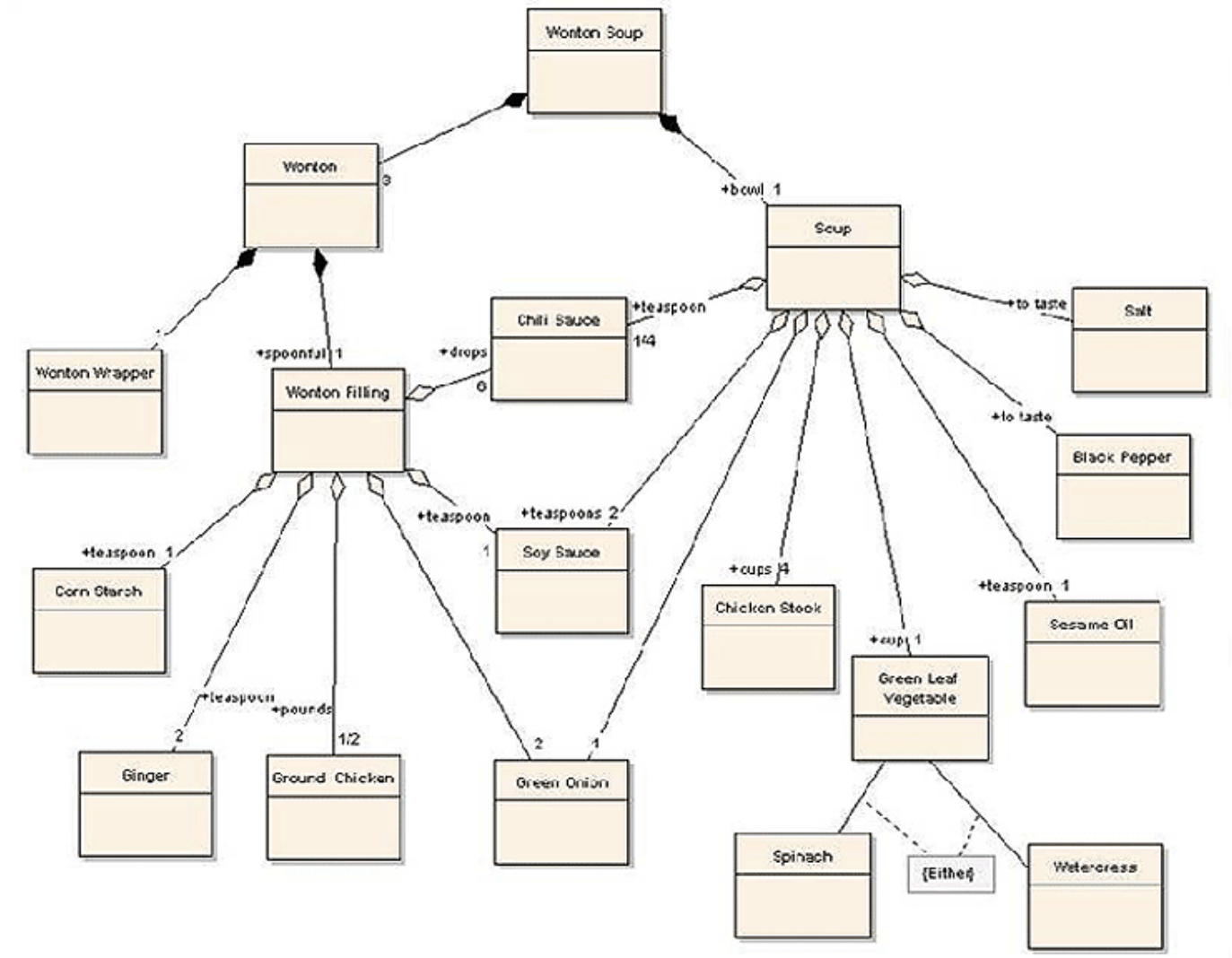
Physical Modeling for Analysis - 1

- The primary purpose of modeling is to better understand some existing system or environment and its resources and to describe this understanding so it can be communicated
- Models of things as they currently exist are Physical or As-Is models
- This modeling activity is usually called "systems analysis" or simply "analysis"
- A basic task of modeling for analysis is capturing the languages and practices of the people who work with the resources in the "real world"

Physical Modeling for Analysis - 2

- Any system (especially business and other institutionalized ones) has groups of stakeholders who do not fully understand the "big picture" – an analysis model can be used to prevent or repair misunderstandings
- A physical model synthesizes different views or observations into a more complete or generic perspective that accurately accounts for all of them

Physical View of Wonton Soup



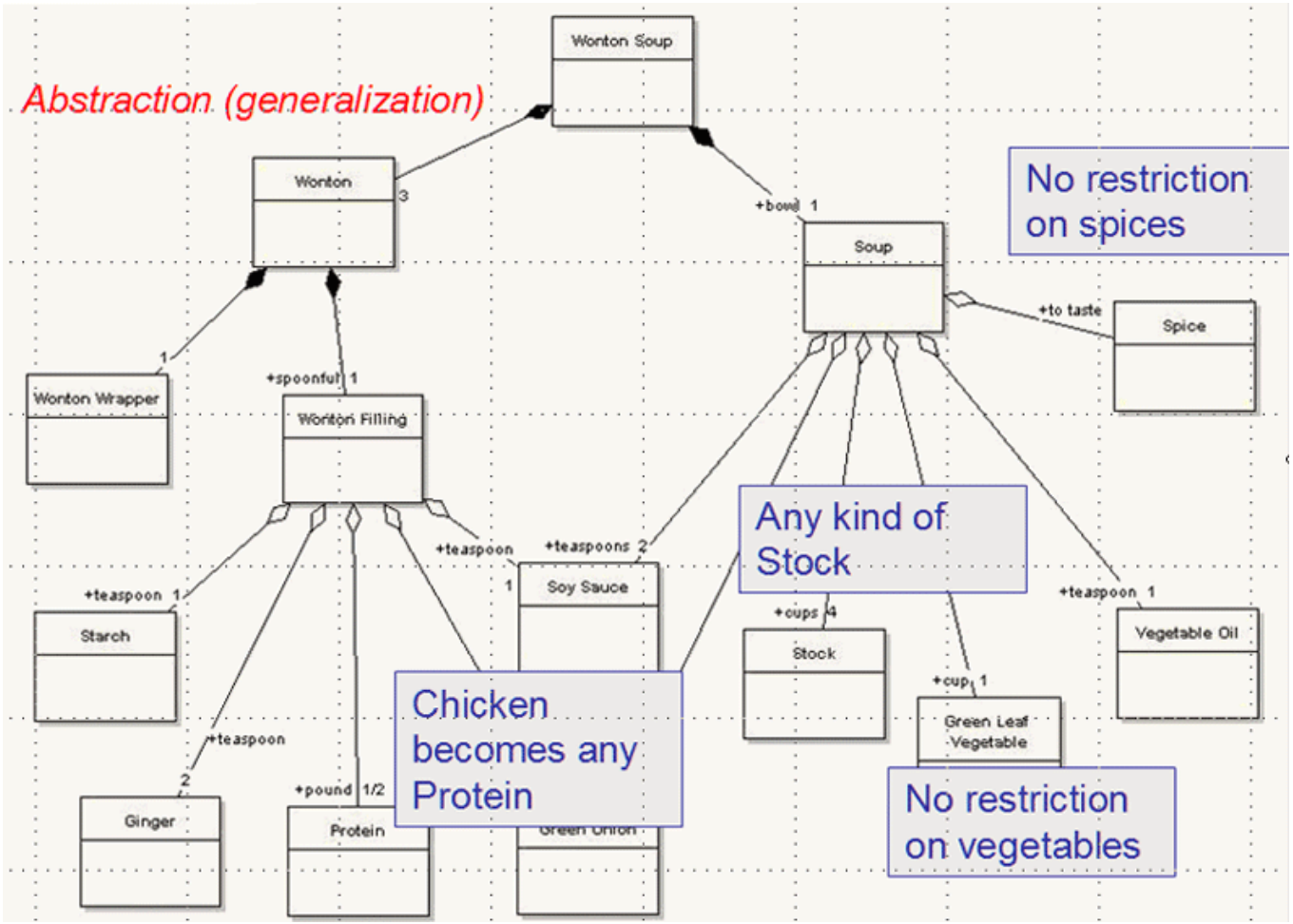
Conceptual Modeling for (Re-)Design - 1

- The next purpose of modeling is to assist in the design or re-design of a system or set of resources
- This modeling activity is usually called "systems design" or simply "design"
- Models of resources as they could be are Conceptual or To-Be models
- Design abstracts away or generalizes from the technology and implementation details in the physical model to create a conceptual model

Conceptual Modeling for (Re-)Design - 2

- A conceptual model that is implementation-independent is often easier to talk about than one that is encoded in a specific technology context
- When technologies change, the implementation model may change but the conceptual model won't.
- When implementations in different technologies are based on the same conceptual models, they can more readily understood because of their common conceptual components.

Conceptual View of Wonton Soup



Using Conceptual Models

- Resources can be manipulated as conceptual components without impacting the real world that they describe, or in ways that are impossible in the real world
- This encourages the re-use of common components via standardization, patterns and libraries
- It facilitates the rationalization of components and the removal of redundancies and inefficiencies

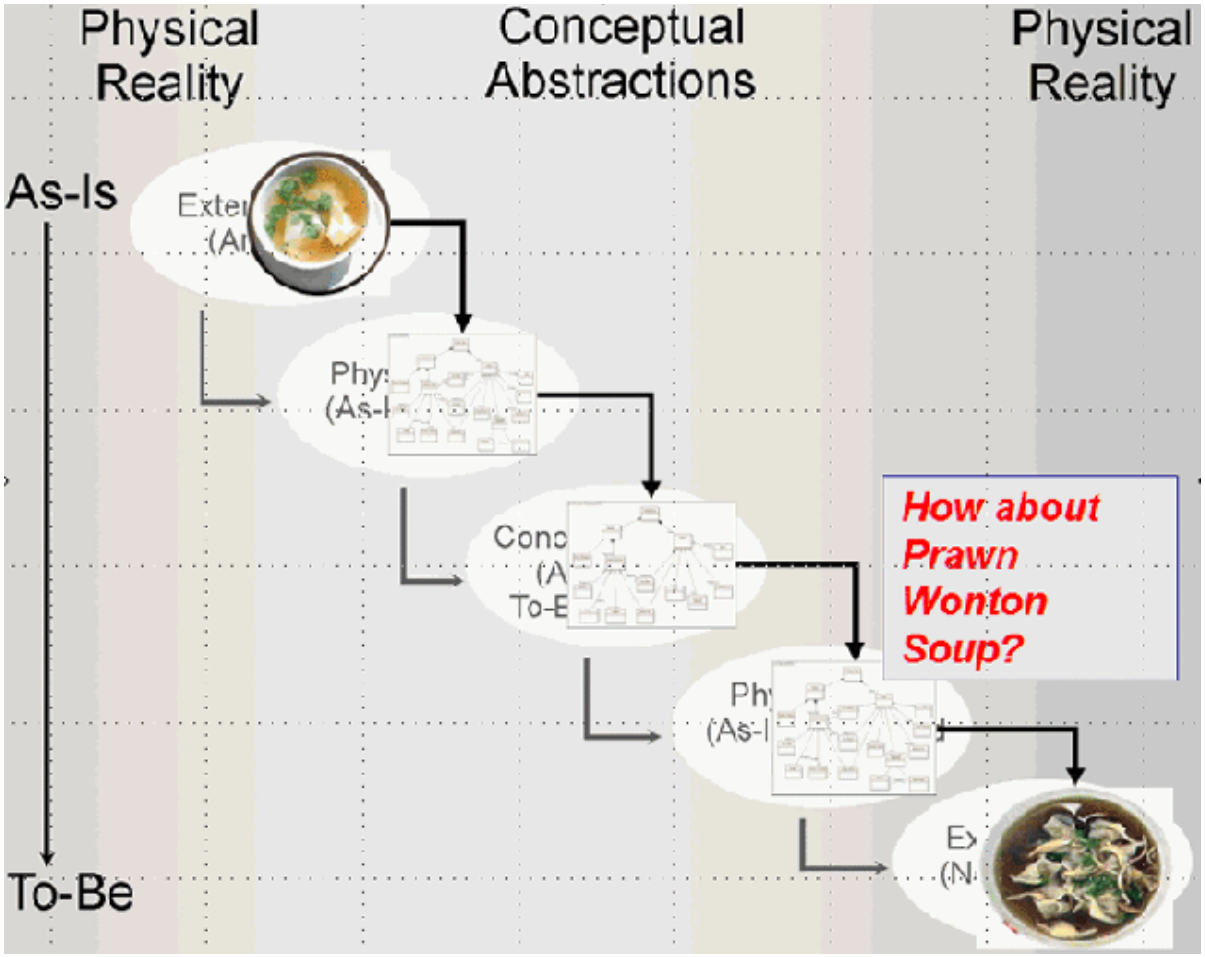
The Modeling Gaps

- There is an essential difference or gap between the real world being modeled and any models of it, or else the models would serve no purpose
- Likewise, there is always a gap between a physical model and a conceptual model, because an analysis model is often most useful when it isn't tied to specific or feasible implementations or technologies
- But this means we can sometimes see what the current world looks like and what we would like it to be without being able to see how to get from one to the other

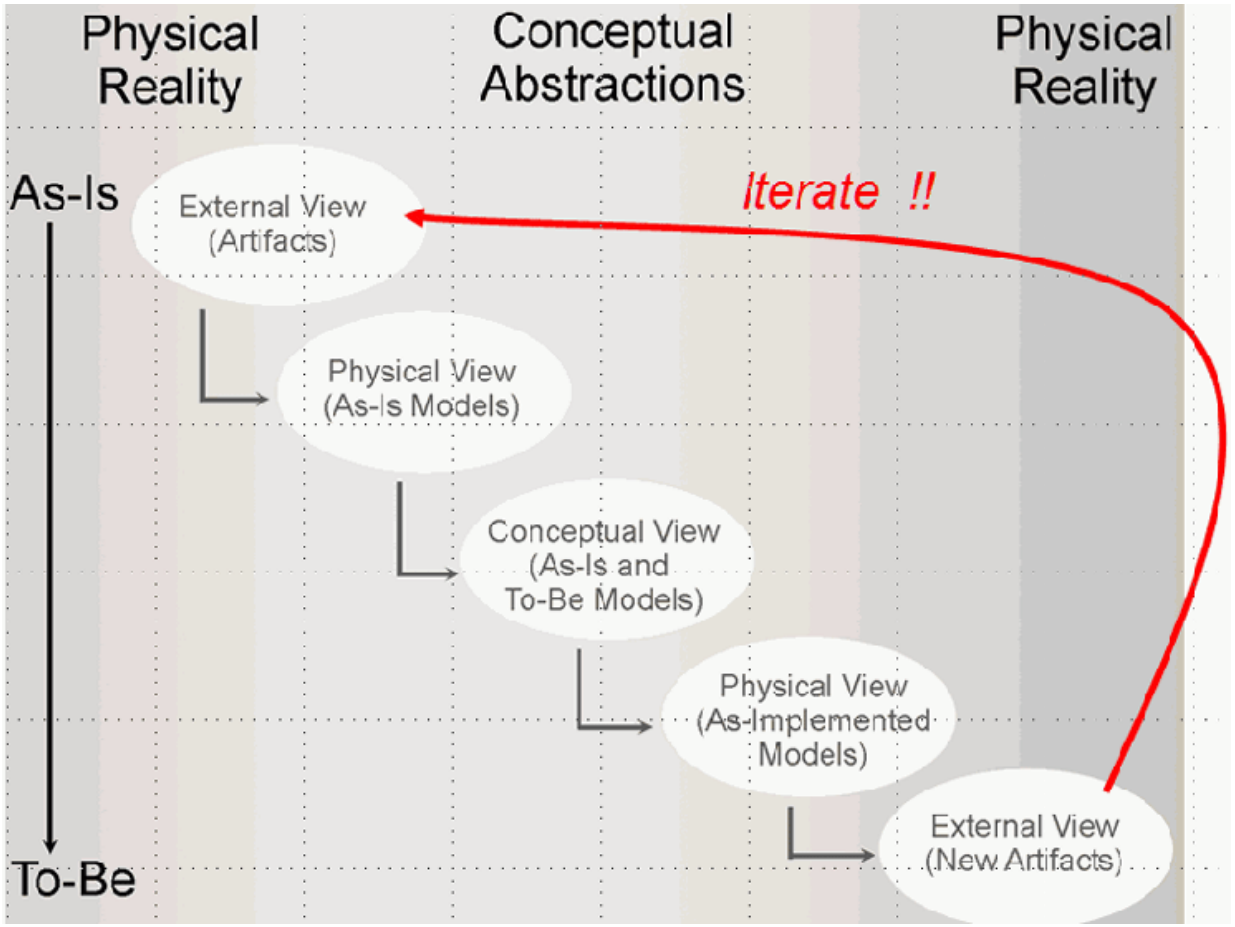
Implementing a Conceptual Model as a Physical One

- A model is purely theoretical until it is encoded in a technology that lets it operate in the real world again.
- This is often a two-stage process: encoding conceptual models as physical ones, and then applying transformations to create resource instances with desired properties
- When resources implemented in different technologies are generated or re-generated from models, they can more readily interoperate because of their common conceptual components

Applying the Conceptual Model



Iteration is Inevitable



Adapting the Classical Modeling Approach to TDO – 1

- In an Organizing System context we can analyze a domain and analyze its resources and their processes or behaviors as we normally would
- But taking an architectural perspective that distinguishes organizing principles from implementation means that we emphasize the information or intangible content of the resources
- We only model those resources and their properties that are relevant to organizing them

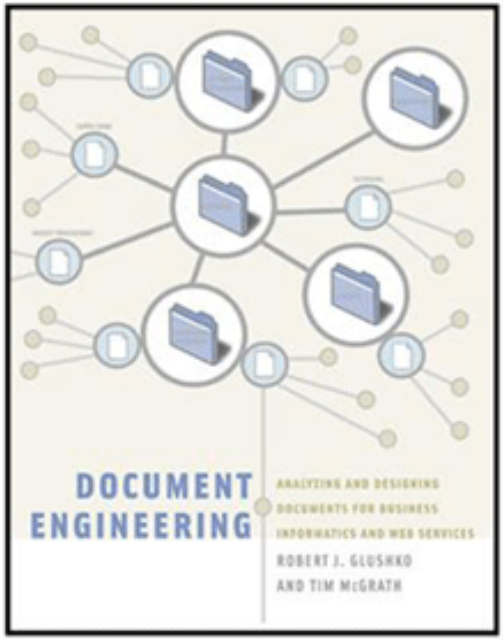
Adapting the Classical Modeling Approach to TDO – 2

- In some Organizing System contexts we start with a collection of resources to organize
- So we analyze the domain and the resources to determine appropriate properties to use in descriptions and organizing principles as we would in any other modeling activity
- But sometimes we concurrently design the Organizing System and the resources it will contain
- Systematic modeling of information components and documents to enable efficient reuse and transformation is essential

Model the Information, not the Things

```
<Book>
  <Title>Document Engineering</Title>
  <Author>Glushko</Author>
  <Author>McGrath</Author>
  <ISBN>0804900337</ISBN>
  <Publisher>MIT Press</Publisher>
</Book>
```

≠



Readings for Next Lecture

- TDO 3.5, TDO 8.3.1 and 8.3.2
- Gardiner www.wired.com/wiredscience/2013/05/sensors-listen-to-world/
- Smith, Abby. “Authenticity in perspective”
- Barbara Bazzanella, Stefano Bortoli, Paolo Bouquet, “Can Persistent Identifiers Be Cool? “