Plan for Today's Lecture(s)

- Resources and Resource Identity over time
- Intro to Resource Description
- Resource Properties and Description
- Creating Resource Descriptions

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

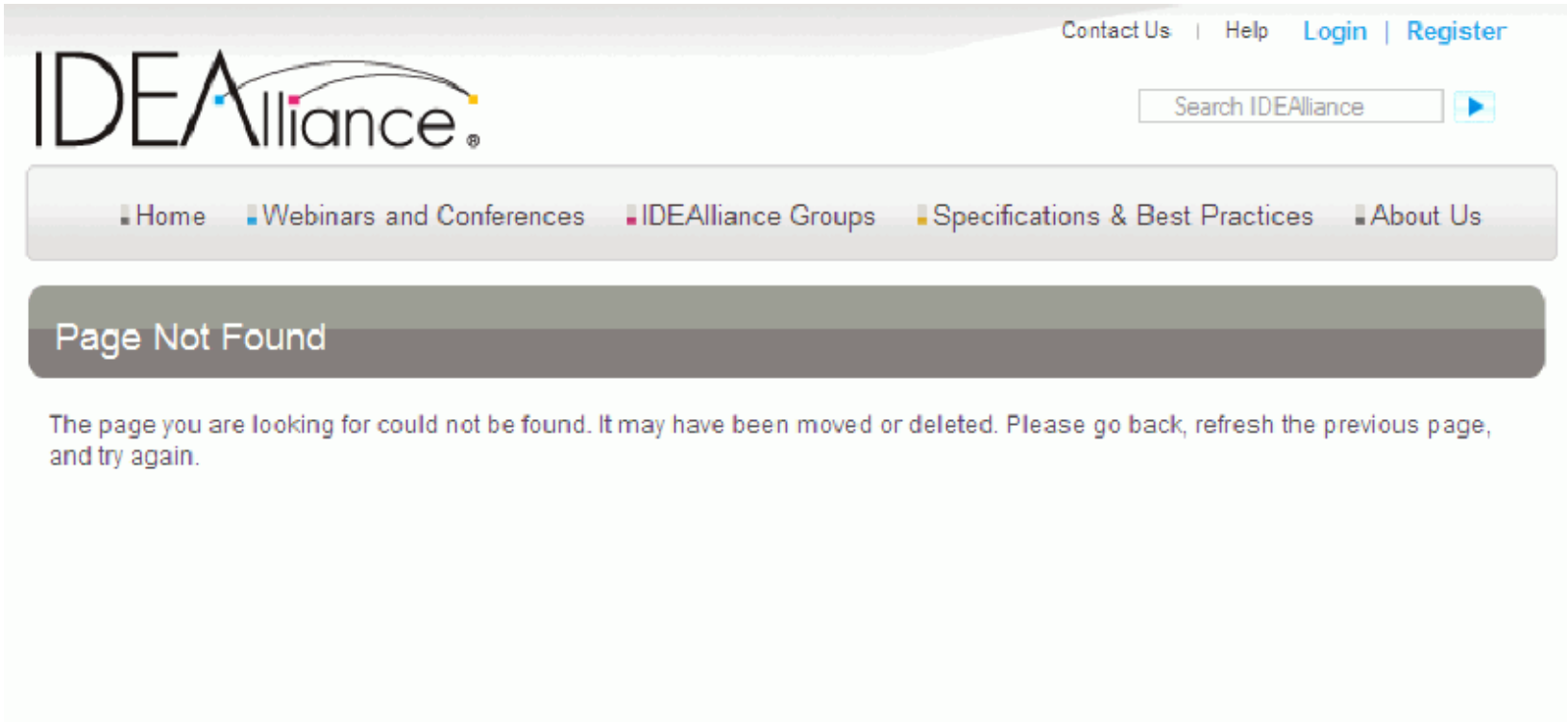
Robert J. Glushko
glushko@berkeley.edu
@rjglushko

19 September 2013
Lecture 6.4 – Resources and Resource Identity Over Time

Resources and Identity Over Time

- Organizing systems continually need to adapt and evolve in response to changes in content and context
- Resources are created or added, and other ones are destroyed or culled
- Resources might change in controlled ways or in uncontrolled ways
- The organizing system might need automated or formal mechanisms for ensuring persistence and authenticity

Persistence



When you don't want persistence:

"Reputation Management" "Mugshot Cleaner"

Authenticity

Buy the photo at Ustar.net , but consider http://www.fineartstitle.com

Authenticity in a Digital Environment – 1

- We often use judgments about the physical integrity of recorded information to stand in for a judgment about the integrity of the text
- Digital resources have no independent physical manifestations or indications of usage that can provide evidence "about their fate in the world”

Authenticity in a Digital Environment – 2

- The "permeability" of digital environments makes it harder to ensure that resources have not been altered
- Scholars trust technological solutions like time stamps. encryption, watermarks, and digital signatures; technologists are skeptical of them
- In the digital world selecting for preservation has become a process of constant re-selection

Provenance

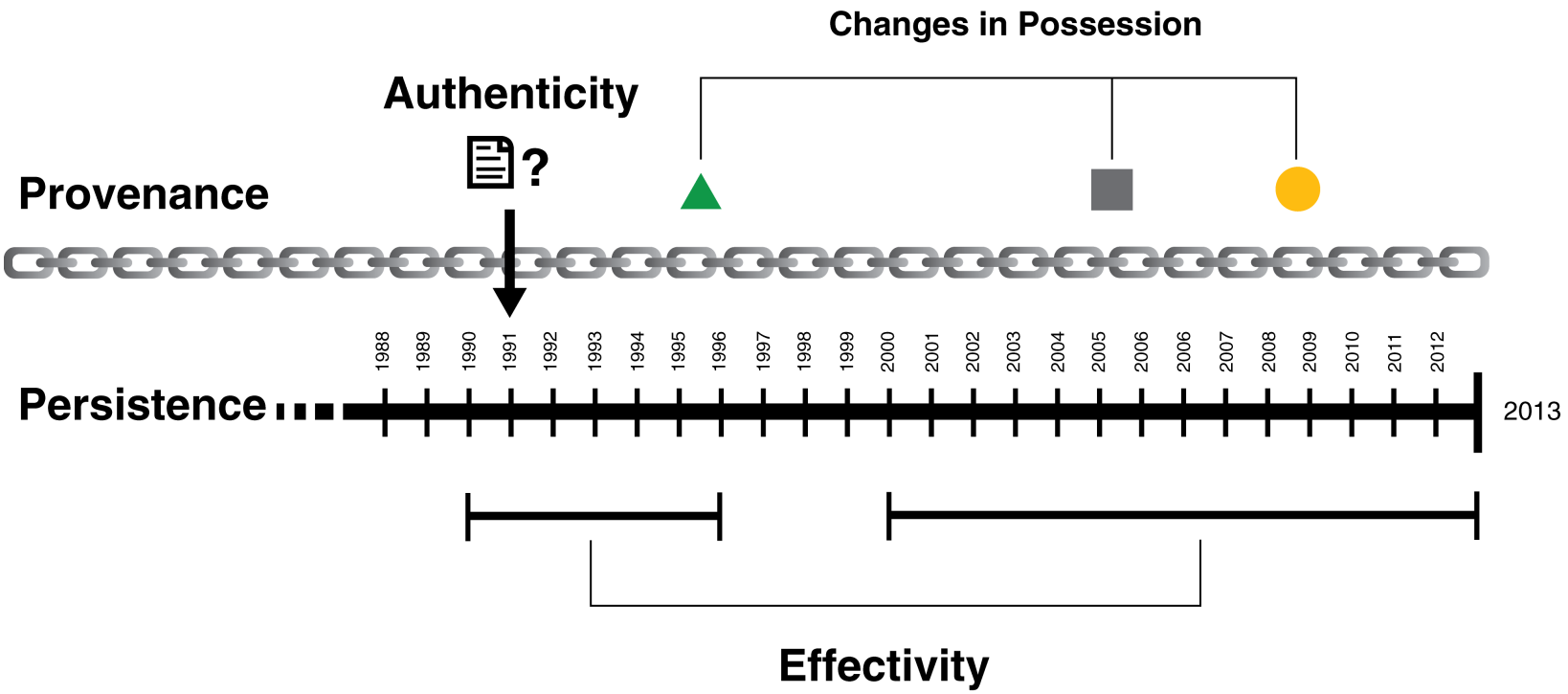

巧狀層峯數尋蒼色如煙
合一片盤根似蘚封院宇
接連常籍竹池亭掩映
却應松分明裝出依叢寺
只欠清實幾韻鍾
風霜正臘晨早見幾枝新
預荷東皇化偷回北苑春
旗槍雖不類茶學似堪倫
已有清榮諭終難混棘蓁

Effectivity

Sec. 108. This title shall take effect on the first day of the sixth month which begins after the date of the enactment of this title, except that the amendments regarding sections 19(b)(7) and 19(b)(8)(D) of the Federal Reserve Act shall take effect on the date of enactment of this title.

http://www.fdic.gov/regulations/laws/rules/8000-2200.html

Putting it all Together (TDO Figure 3.6)



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

19 September 2013
Lecture 7.1 – Introduction to Resource Description

An Overview of Resource Description

- What is a resource description?
- Why do we describe resources?
- What resource properties should be described?
- How are resource descriptions created?
- What makes a good resource description?

What is a Resource Description?

- Information that is created intentionally and associated with a resource to enable it to be organized and interacted with
- A resource description is also a resource from the perspective of any other resource that uses it, ad infinitum (which is why Resource Focus is an important issue)
- A resource description is often a functional substitute for the resource it describes when the latter can't be accessed or used

What is a Resource Description?

- Coyle cautions us:

  Resource descriptions developed on theoretical, religious, or philosophical principles may be intellectually pleasing, but are unlikely to get the job done

Why We Describe Resources

- We describe resources so we can refer to them, organize them, and interact with them
- Each purpose might require different descriptions and different methods of using them
- Different resource domains can have characteristic or standard resource descriptions (or description categories)

Why We Describe Resources

- But different types of resources must also have differentiating properties, otherwise there would be no reason to distinguish them as different types
- We often combine descriptions... and we often compare them
- But over time as a collection of resources grows and as requirements for interactions change, the reasons for describing resources will also change

Purposes and Description

- Selection: Descriptions enable discovery, assessment of fit to requirements, authentication, appraisal
- Organizing: Tangible or perceivable properties for physical resources, "aboutness" for digital ones
- Interaction Design: Organizing is the prerequisite for this purpose, which has been highlighted in LIS since 19th century
- Maintenance: "Preservation Metadata" often emphasizes technological requirements to use resources

Objectives of Bibliographic Description

- Finding a resource that you know exists
- Identifying a resource to make sure you have the one you were looking for
- Selecting a resource from a set of candidates in a collection
- Obtaining the resource if what you have at this point is just a resource description

Objectives of Bibliographic Description

- These four purposes and the information components needed in resource descriptions for satisfying them are systematized in LIS as the Functional Requirements for Bibliographic Records (FRBR)
- …but the ideas apply more broadly to organizing systems:
  Tesco smartphone subway shopping in Korea

Complications

- The properties of resources that are easiest to describe are not always the most useful ones, especially for information resources
- For non-text information resources this problem is magnified because the content is often in a semantically opaque format that cannot usefully be analyzed by people.
- Business strategy and economics strongly influence the extent of resource description
- But as bibliographic collections grow larger, we need more descriptions to satify the “frbr”

Process of Describing Resources

- Identify / scope the resources to be described
- Determine the purposes or uses of the descriptions
- Study the resource(s) to identify descriptive properties
- Design the description vocabulary
- Design the description form and implementation
- Create the descriptions (either "by hand" or by some automated / computational process
- Evaluate the descriptions

(TDO FIGURE 4-3)

Scoping and Description

- How explicit and thorough these 7 steps need to be depends on what we are calling "scope“
- Do we know who the users are? Are the describers the users?
- Do we have all the resources, a representative set, or just the first set ... do we know what else is coming?

Scoping and Description

- Sometimes we describe resource instances; sometimes we just describe the collection they are in
- As a collection grows, the language for describing resources must become more rigorous, and descriptions created when the collection was small often require revision because they are no longer adequate for their intended purposes

Kimra' Kitchen Chair

How you describe "the folding kitchen chair" depends on…

- Whether the description is for your own purposes, for other people, or a computer
- Whether those people have seen the chair before
- How many other chairs are nearby and what they look like
- What your audience knows about chairs
- Whether you can try again if your first description doesn't satisfy its purpose

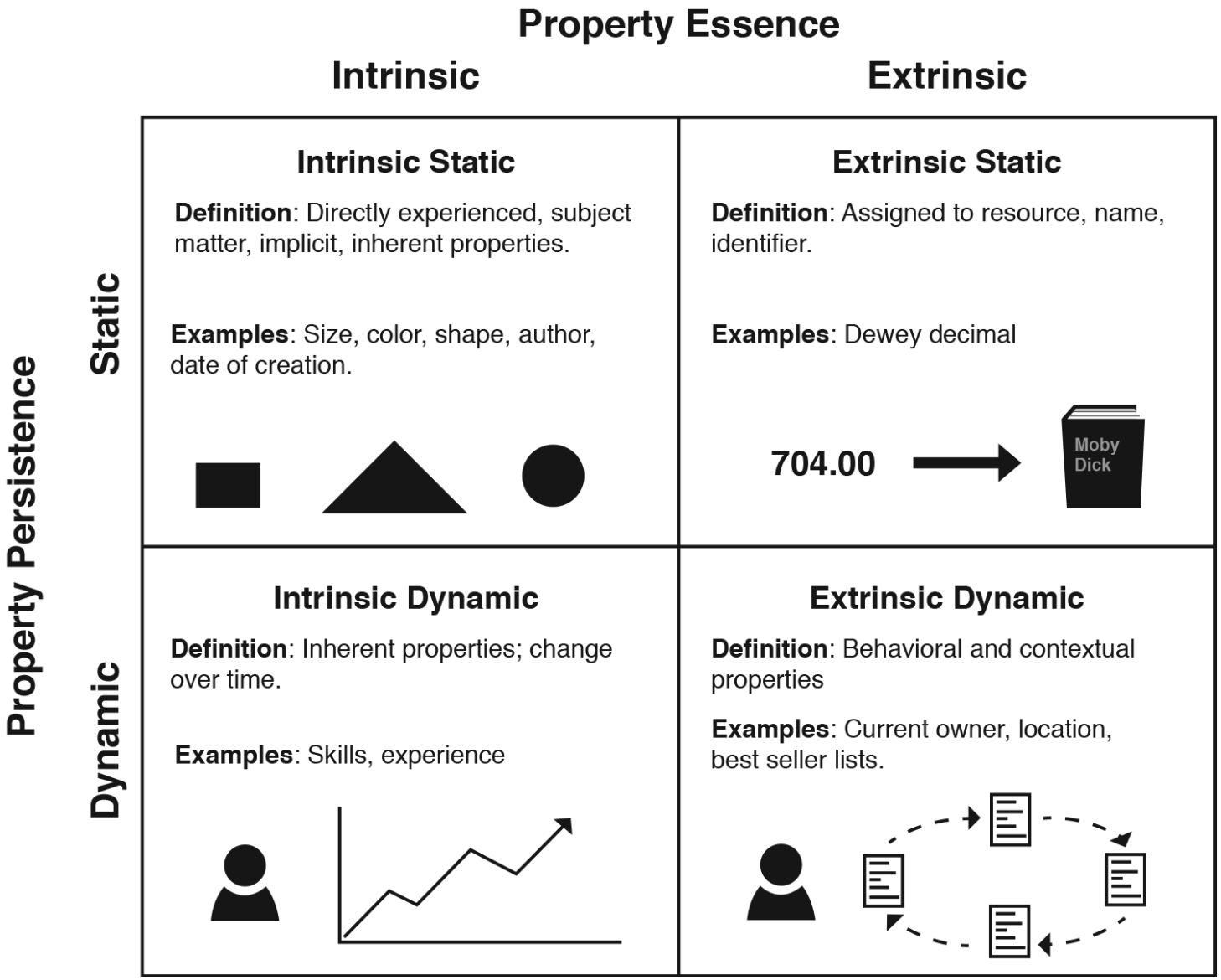
UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

19 September 2013
Lecture 7.2 – Resource Properties and Description

Classifying Resource Properties



Intrinsic Properties

- Intrinsic properties are inherent in a resource
- Some are static, never changing their values
- Others are dynamic, but they change “from the inside of the resource” not “from the outside” by actions or efforts of outside agents (like “developmental” properties – age, skill, experience)
- Intrinsic properties can sometimes be used as an identifier - like a computed "signature" for a document or media object

Physical Properties

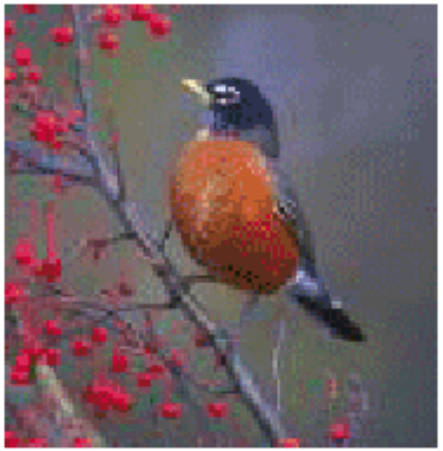
- Physical or perceptible properties are those "at the surface" that are immediately apparent
- For "natural" things, these physical properties often make excellent descriptors because they intrinsic or inherent rather than assigned
- They occur in consistent, predictable, and correlated combinations, which makes them reliable aids to identifying instances or types of things (the “joints of nature”

Distinctive Physical Properties

Physical Properties

- For "manmade" things surface properties are less predictable ("innovation" enables the separation of form and function)
- For "information" things in physical form the correlation between appearance and content varies across the Document Type Spectrum
- For "born digital" resources the appearance is often extrinsic because it is assigned or associated (a style sheet, for example) instead of being an inherent part of the resource when it was created

Which Properties Best Identify a Thing? A Type of Thing?



- The properties best for identifying an instance might be those that aren't typical of its class

Extrinsic Properties

- Extrinsic properties are assigned to a resource rather than being inherent in it
- Some extrinsic properties are static: assigned names or identifiers don't usually change
- Other extrinsic properties change often: the current location, popularity, price, etc. of a resource)

Cultural Properties

- Cultural properties are those that derive from conventional language use or culture and often involve analogy
- Because they derive from knowledge of culture or language, they might be unintelligible to people who don't have the same perspective and experience
- .. and they might lose their cultural salience

“Holbein” Carpets

“Darmstadt Madonna”

“The Ambassadors”

[More about Holbein and Henry VIII]

Contextual Properties

- Contextual properties relate to the situation or context in which the thing being described exists
- Dey et al (2001) define context as “any information that characterizes a situation related to the interactions between users, applications, and the surrounding environment.”
- This open-ended definition implies many contextual properties that might be used in a description
- Since context changes, context-based descriptors might be appropriate when assigned but not later; see "persistence" and "effectivity"

Contextual Description == Private Language?

We found **334,015 results** matching guest.

View: Most relevant • Most recent • Most interesting Show: Details • Thumbnails

My Regular Guest by Araleya

72 comments 36 faves

Tagged with macro, home, nature, beautiful ...

Taken on March 22, 2007, uploaded April 13, 2007

Taken in Bang Kruai , Nonthaburi (map)

See more of Araleya's photos, or visit her profile.

A "guest" tag makes sense to the photographer but no one else would use it to describe this photo if they don't share the context

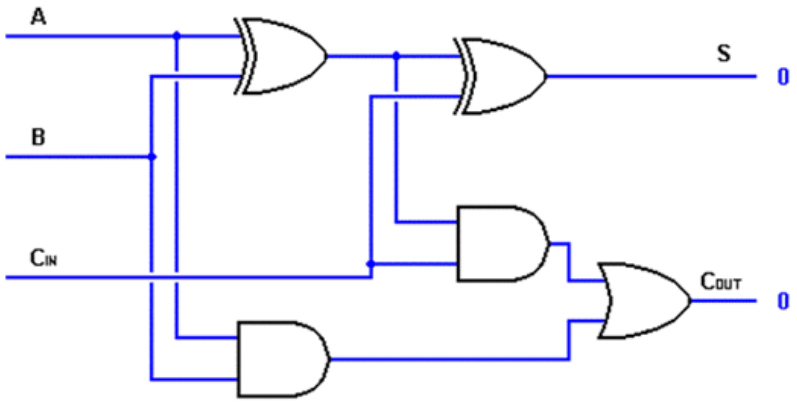
Structural Properties

- The internal and external structure of a thing can be a useful part of its description
  - The number and arrangement of component parts (e.g, the number of chapters or pages in a book)
  - The number and type of connections with other resources (e.g., the number of Facebook friends or Twitter followers)

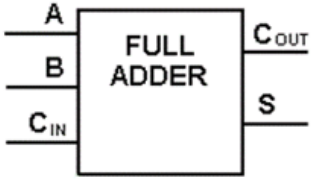
Meaning is Structure

Circuits for Binary Addition

Full Adder



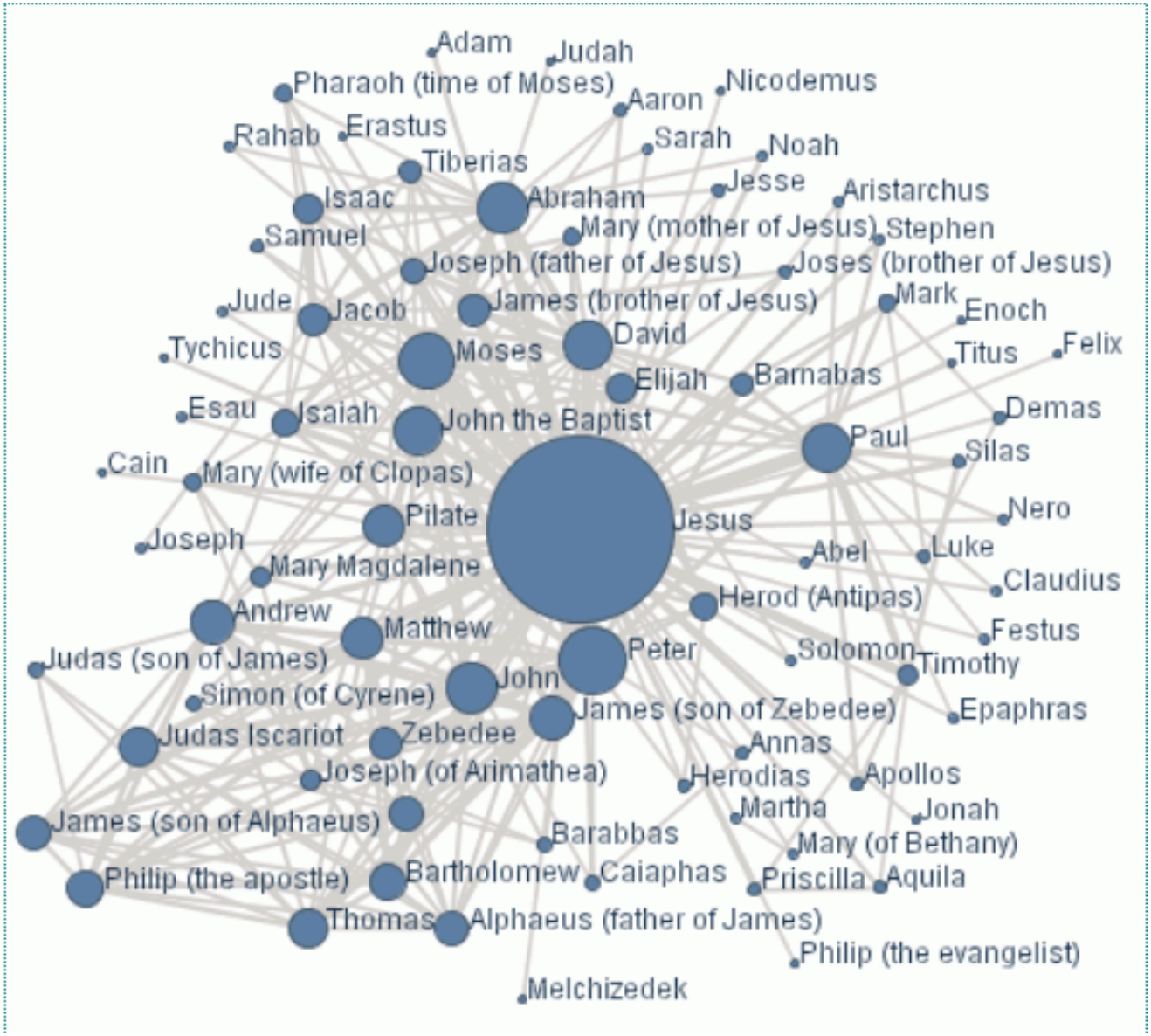
INPUTS			OUTPUTS	
A	B	C_{IN}	C_{OUT}	S
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1



Images from http://www.play-hookey.com/digital/adder.html

Could you describe this without using structural concepts?

The Social Network of Jesus

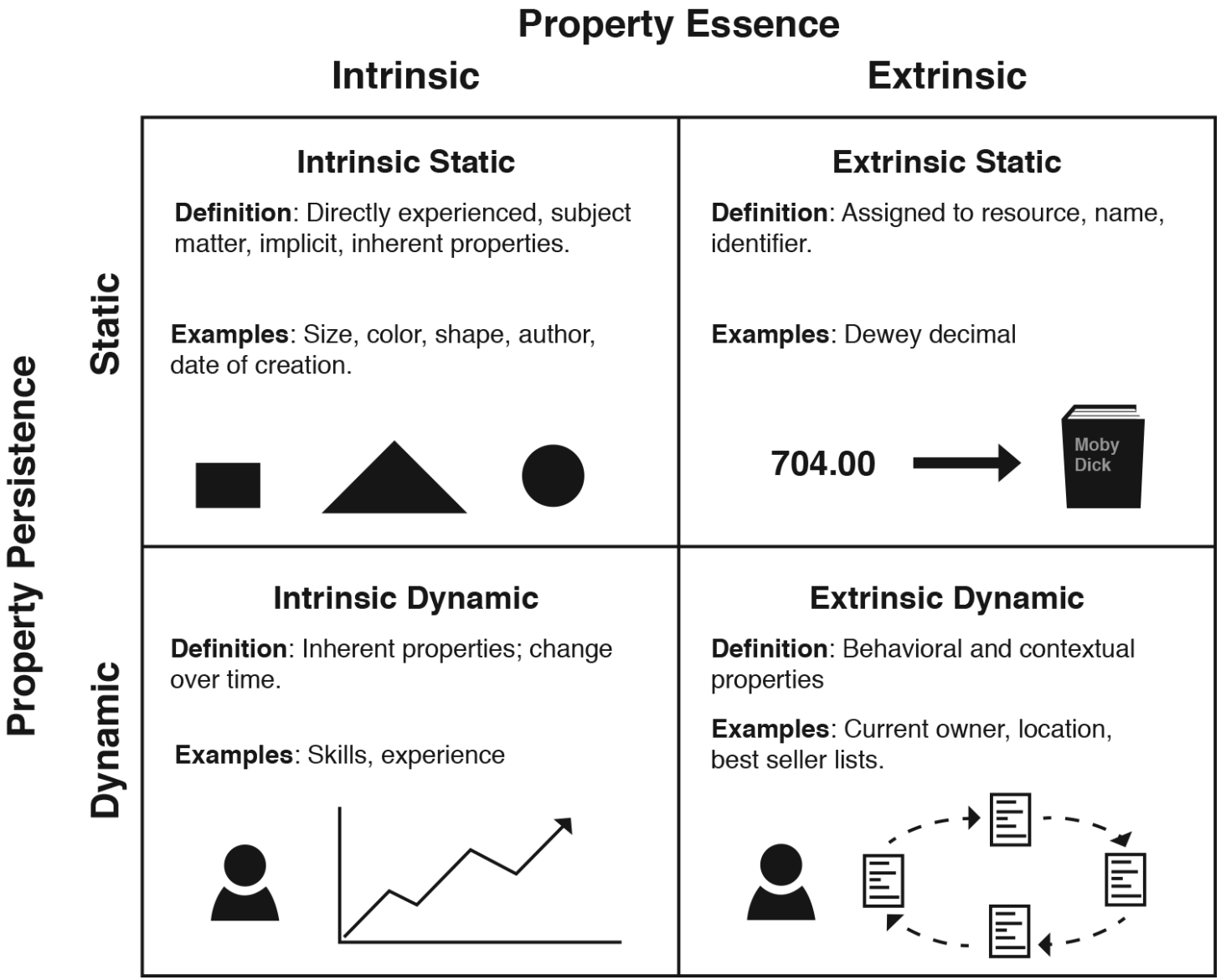


Properties vs. Relationships

- But while we can express structure within and between resources as properties, there is "better language" for expressing this as RELATIONSHIPS
  - Arrangement
  - Proximity
  - Connectivity
  - Part-whole
  - …

(See TDO Chapter 5)

Classifying Resource Properties



Principles of Kitchen Organization

- **Intrinsic static properties**: If you store your pots, frying pans, and baking pans and nest each set by size
- **Extrinsic static properties**: A spice rack with the spices arranged in alphabetical order
- **Intrinsic dynamic properties** if you arrange your milk and other perishable goods by expiration date, a "useful life remaining" property that decreases to zero as the expiration date approaches
- **Extrinsic dynamic properties** if you put the most frequently used condiments or spices in the front of a refrigerator or pantry shelf.

Principles of Document Organization

- **Intrinsic static properties:** Author, date published, words in the text
- **Extrinsic static properties**: ISBN, LOC Classifications
- **Intrinsic dynamic properties:** Effectivity (e.g., laws and regulations)
- **Extrinsic dynamic properties:** Links/citations to and from other documents

UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

INFO 202
"Information Organization & Retrieval"
Fall 2013

Robert J. Glushko
glushko@berkeley.edu
@rjglushko

19 September 2013
Lecture 7.3 – Creating Description Vocabularies and Assigning Descriptions

Designing the Description Vocabulary

- Good descriptions use terms that their intended users might use
- Good descriptions don't contain details that aren't necessary
- Good descriptions are created systematically and follow standards

The Image Game: Finding Good Descriptions

Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on "purse"

Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on "purse"

The Need for Controlled Vocabularies

- The words people use to describe things or concepts are "embodied" in their context and experiences... so they are often different or even "bad" with respect to the words used by others
- These naturally-occurring words are an "uncontrolled vocabulary"

The Need for Controlled Vocabularies

- Searches for resources using an uncontrolled vocabulary will not succeed when they fail to match resources described or indexed using a controlled vocabulary
- To agree on the words we use in descriptions will improve recall, but it means that we must use a subset of the words we would otherwise use

What is a Controlled Vocabulary?

- A controlled vocabulary is a standardized set of terms (such as subject headings, names, classifications, etc.) assigned by organizers / cataloguers / indexers of resources
- A CV can be a content standard for the values used in (or as) descriptive elements
- A CV can be thought of as a fixed or closed dictionary in which everything must be defined using the same set of terms

Types of Controlled Vocabularies

- Dictionaries
- Authoritative names
- Authority control for places and time periods
- Identifiers
- Code lists
- Subject heading lists (like the Library of Congress – LCSH)
- Thesauri
- Classification systems

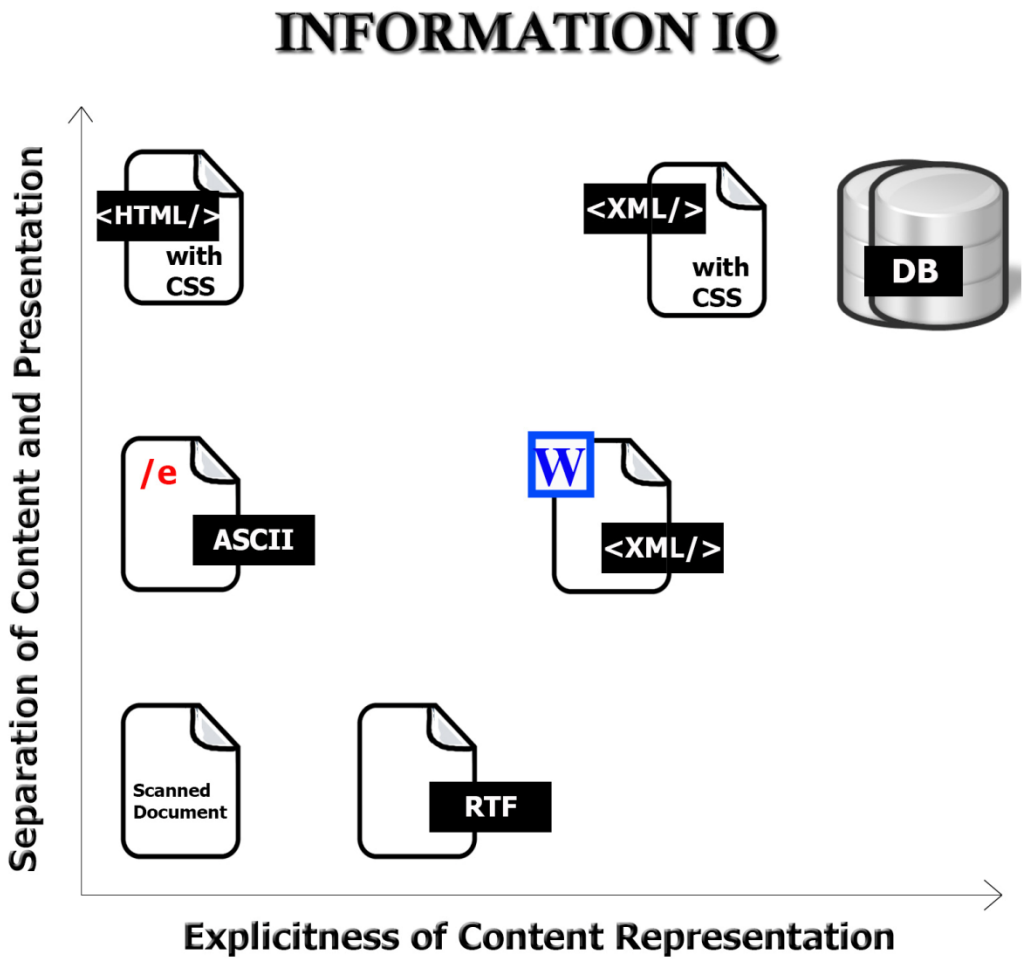
Horizontal and Vertical Vocabularies

- Horizontal
  - Concepts that are common to all (or a large number) of domains
- Vertical
  - Particular industry or vertical market
  - Specialized product or process semantics
  - Sometimes called “domain-specific languages”

"Dimensionality Reduction" in Descriptions

- Creating a controlled vocabulary reduces the number of descriptive terms that are or can be assigned to something
- When achieved by computational techniques, "dimensionality reduction" goes by names like "principle components analysis," "orthogonal decomposition," "latent semantic analysis," "factor analysis," and others
- These techniques analyze the correlations between descriptors and transform a large set into a smaller set of uncorrelated ones

Description Form and Implementation: Format Matters!



Creating Resource Descriptions

- By professionals: "Institutional" descriptions that follow standards
- By authors: Best knowledge of purpose and intended audience
- By users: Most variable, influenced by social purposes
- By automated processes: Most reliable, primarily technical / objective properties but semantic description capability is emerging

Evaluating Resource Descriptions

- Quality should be evaluated with respect to intended purposes... but what if different stakeholders have different purposes?
- Creating resource descriptions can be costly; is it worth it? How and when measure the costs?
- Computational description is vastly less expensive, but is it good enough?
- How can we incent "community" or "crowdsourced" description to be of better quality?

Readings for Next Lecture

- Next lecture on XML (September 24) will be presented by Erik Wilde

- Glushko and McGrath. Document Engineering. Chapter 2, XML Foundations
- XML In a Nutshell, Ch 2 (as a reference guide and for assignment)