



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Creating the Multivalent Book

Robert J. Glushko
glushko@berkeley.edu

@rjglushko

Books in Browsers

24 October 2014

Today's Talk

- The title of talk was made up by Peter Brantley as a placeholder and I never got around to changing it
- So that's the title even though I have never used the word "multivalent" and am not sure what it means
- But my talk this year is a follow-on to my talk at last year's B in B so I'll start there

An Avant-Garde Ruminati on Negative Space and Skimming/Groking/Mastering the Multivalent Book

Robert J. Glushko
glushko@berkeley.edu

@rjglushko

Books in Browsers

24 October 2014

My talk last year (I meant “transdisciplinary”)



UNIVERSITY OF CALIFORNIA, BERKELEY
SCHOOL OF INFORMATION

Collaborative Authoring, Use and Maintenance of a Multidisciplinary Textbook

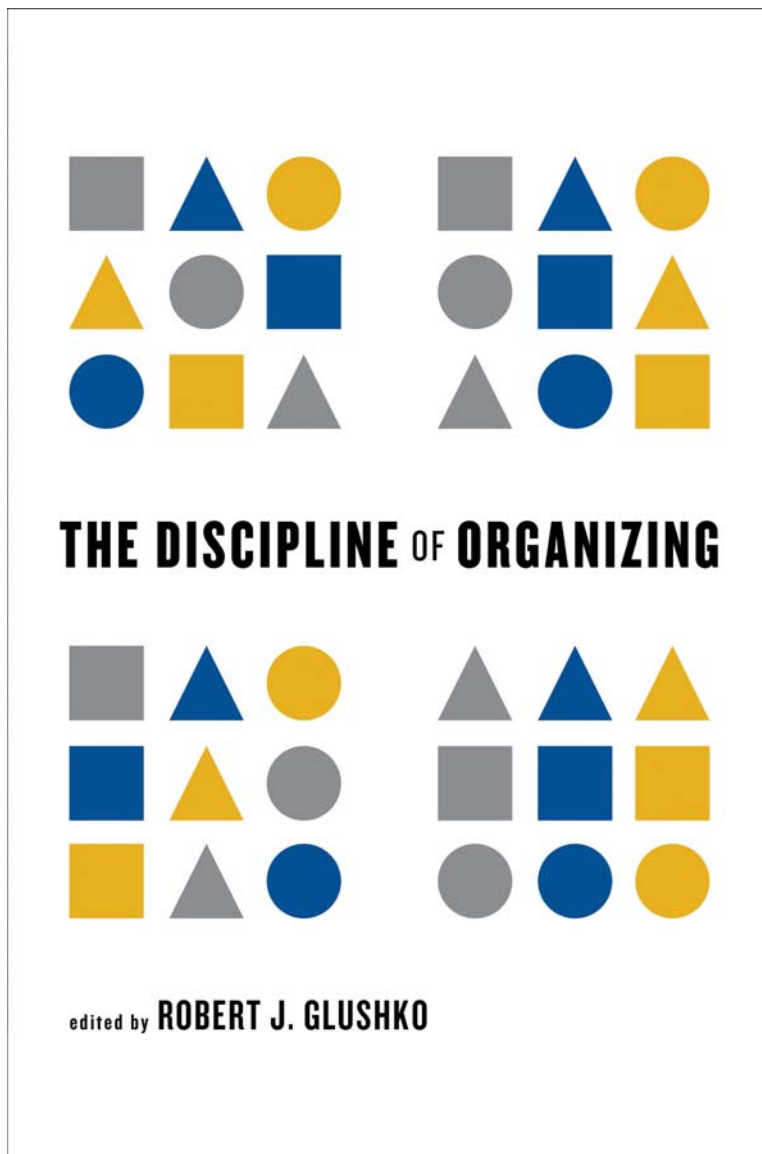
Robert J. Glushko
glushko@berkeley.edu

@rjglushko

Books in Browsers

25 October 2013

The Discipline of Organizing: The Crucible / Testbed



- Published by MIT Press in May 2013 simultaneously in print and ebook formats
- Used O'Reilly's Atlas single-source publishing platform to do that

The Breadth vs. Depth Challenge

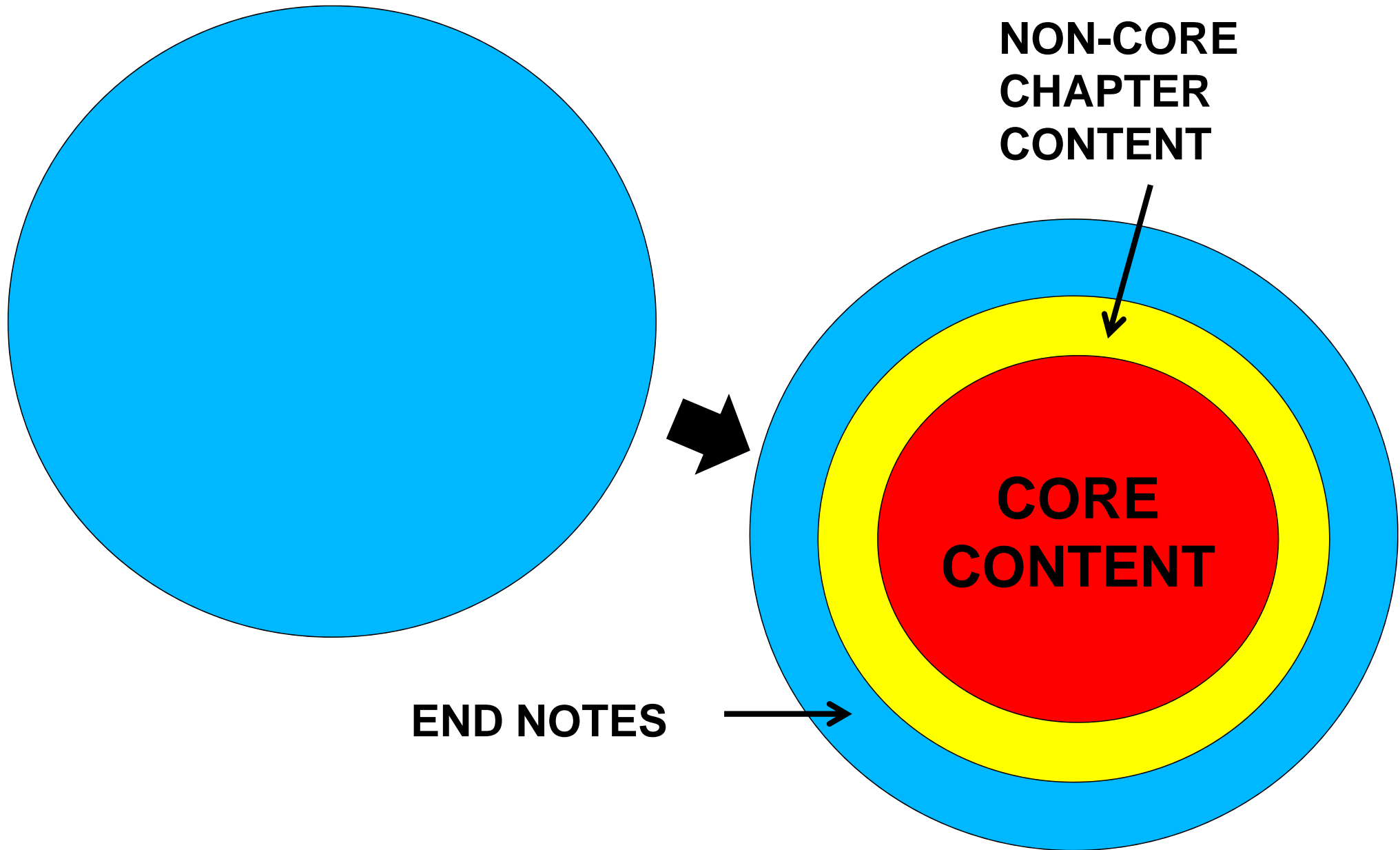
- This new discipline emerges from the intersection of computer science, library science, informatics, business, law, and other disciplines that organize collections of resources and design interactions with them
- It is a BROAD textbook to represent all the disciplines that contribute to it
- It seeks to be DEEP to treat all the disciplines with appropriate rigor and nuance

***How can it be deep and broad
at the same time?***

Using “Tagged Content” to Address the “Breadth” vs. “Depth” Challenge

- About 24% of the content in TDO was converted to endnotes tagged by discipline
- This separates discipline-specific from core content and makes depth into a choice rather than a distraction or confusion
- Reader can use these tags to decide whether or not to read the note

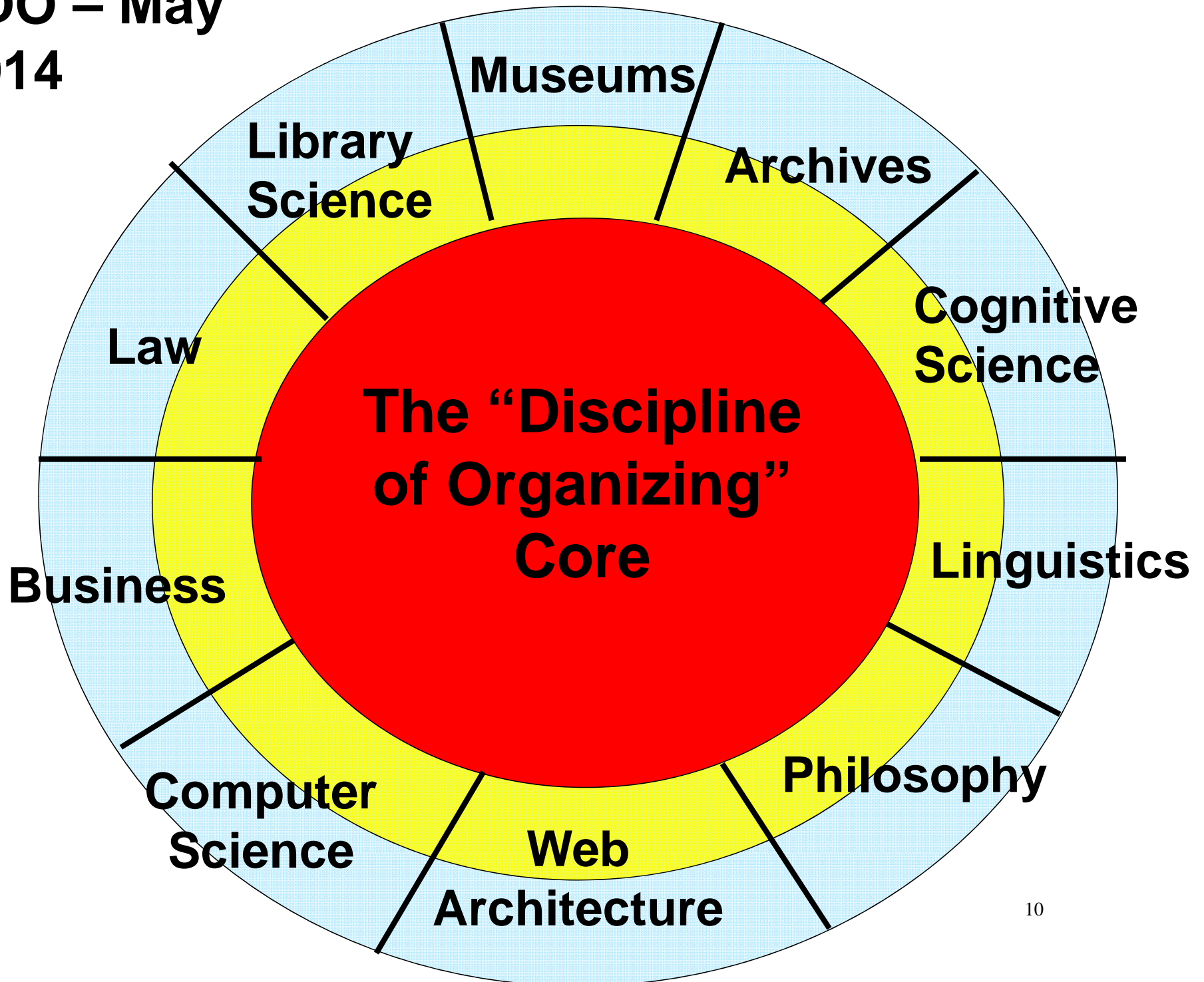
“Factoring” the Book’s Content



It Really Wasn't Factoring

- The book was collaboratively authored by 17 people – mostly former students and colleagues at Berkeley
- The refactoring into core + disciplinary supplemental content was my effort as the editor to save the book from itself
- It would be nice to have tools to support authoring this way from the outset
- We know a lot about what those tools should do... but we haven't started to build them

**TDO – May
2014**



Tagged Endnotes in Print Book

(at end of each chapter)

44. [Computing] Web resources are typically discovered by computerized “web crawlers” that find them by following links in a methodical automated manner. Web crawlers can be used to create topic-based or domain-specific collections of web resources by changing the “breadth-first” policy of generic crawlers to a “best-first” approach. Such “focused crawlers” only visit pages that have a high probability of being relevant to the topic or domain, which can be estimated by analyzing the similarity of the text of the linking and linked pages, terms in the linked page’s URI, or locating explicit semantic annotation that describes their content or their interfaces if they are invocable services (Bergmark et al. 2002), (Ding et al. 2004).

45. [CogSci] In this book we use “property” in a generic and ordinary sense as a synonym for “feature” or “characteristic.” Many cognitive and computer scientists are more precise in defining these terms and reserve “property” for binary predicates (e.g., something is red or not, round or not, and so on). If multiple values are possible, the “property” is called an “attribute,” “dimension,” or “variable.” See (Barsalou and Hale 1983) for a rigorous contrast between feature lists and other representational formalisms in models of human categories.

46. [LIS] Libraries and bookstores use different classification systems. The kitchen in a restaurant is not organized like a home kitchen because professional cooks think of cooking differently than ordinary people do. Scientists use the Latin or binomial (genus + species) scheme for identifying and classifying living things to avoid the ambiguities and inconsistencies of common names, which differ across languages and often within different regions in a single language community.

Pop-up “Web” Note in eBook

matters if a digital document or video resides on a computer in Berkeley or Bangalore if it can be located and accessed efficiently.⁴⁹[Web]

Moreover, because the function manifested as physical properties does not matter to digital content in

An organizing system for digital resources that are associated with them. storage capacity to digital organizing digital resources and descriptive

systems can support collections and interactions at a scale that is impossible in organizing systems that are entirely physical, and they can implement services and functions that exploit the exponentially growing processing, storage and communication capabilities available today.⁵²[Web]

2.3.2. Organizing Digital Resources

[49] [Web] Web-based or “cloud” services are invoked through URIs, and good design practice makes them permanent even if the implementation or location of the resource they identify changes (Berners-Lee 1998). Digital resources are often replicated in content delivery networks to improve performance, reliability, scalability, and security (Pathan et al. 2008); the web pages served by a busy site might actually be delivered from different parts of the world, depending on where the accessing user is located.

ts do
rces
and
ary
izing

Multivalence

- Multi == more than one; Valent == having combining power; Multivalent == capable of multiple combinations
- Our disciplinary tagging turned our book source content into a family of related books, with different versions produced based on design-time configurations
- Not really a new idea:
 - A single automobile production line can support the assembly of customized variations of a car model
 - Software product line engineering enables the creation of many similar software systems from a shared set of software assets

Software Product Line Engineering (For Reference)

Kästner, C., Apel, S., & Kuhlemann, M. (2008, May). Granularity in software product lines. In *Proceedings of the 30th international conference on Software engineering* (pp. 311-320). ACM.

Krueger, C. (2002). Easing the transition to software mass customization. In *Software Product-Family Engineering* (pp. 282-293). Springer Berlin Heidelberg.

Couto, Marcus Vinicius, Marco Tulio Valente, and Eduardo Figueiredo. "Extracting software product lines: A case study using conditional compilation." *Software Maintenance and Reengineering (CSMR), 2011 15th European Conference on*. IEEE, 2011.

Not even a New Problem for Book Publishing

- Elementary school teachers have long recognized the need for different versions of the same book because the readings skills and vocabularies of their students can differ a lot
- They want their students to read the same stories... from different versions of the same book that are skills-appropriate

Reading A-Z



Criteria for Leveling Reading A-Z Books

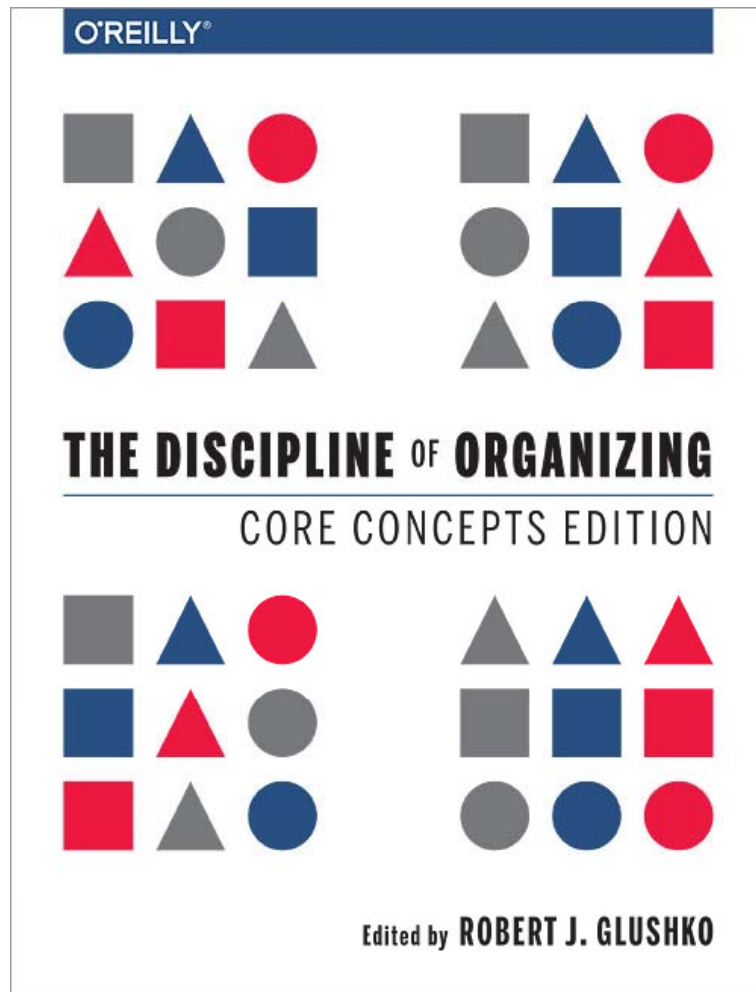
- Word count
- Number of different words
- Ratio of different words to total words
- Number of high-frequency words
- Ratio of high-frequency words to total words
- Number of low-frequency words
- Ratio of low-frequency words to total words
- Sentence length
- Sentence complexity
- Predictability
- Language pattern and repetition
- Print size, spacing, and number of words per page
- Illustration support
- Concept load
- Topic familiarity

More than 1000 leveled books in multiple genres and formats meet the unique needs of every student

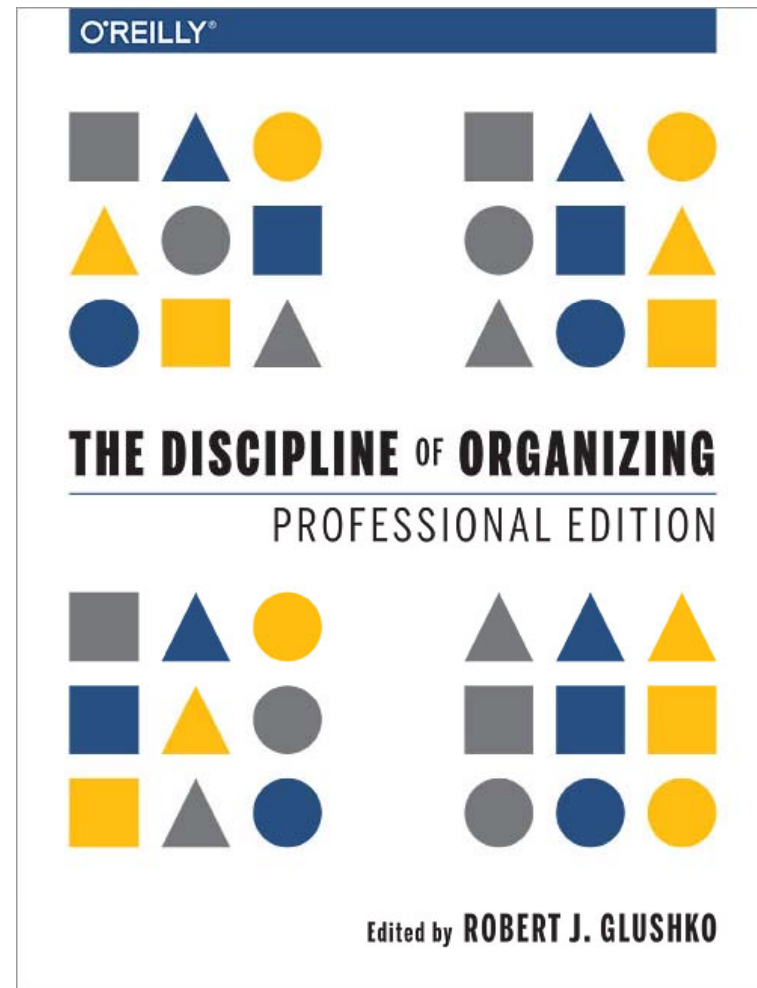
Multivalence using the TDO Discipline Tags

- Computer Science
- Web Architecture
- Cognitive Science
- Linguistics
- Philosophy
- Information Architecture (added 7/14)
- Library and Information Science
- Museums
- Archives
- Business
- Law

The O'Reilly TDO ebooks (8/14)



No endnotes, but inline student quizzes



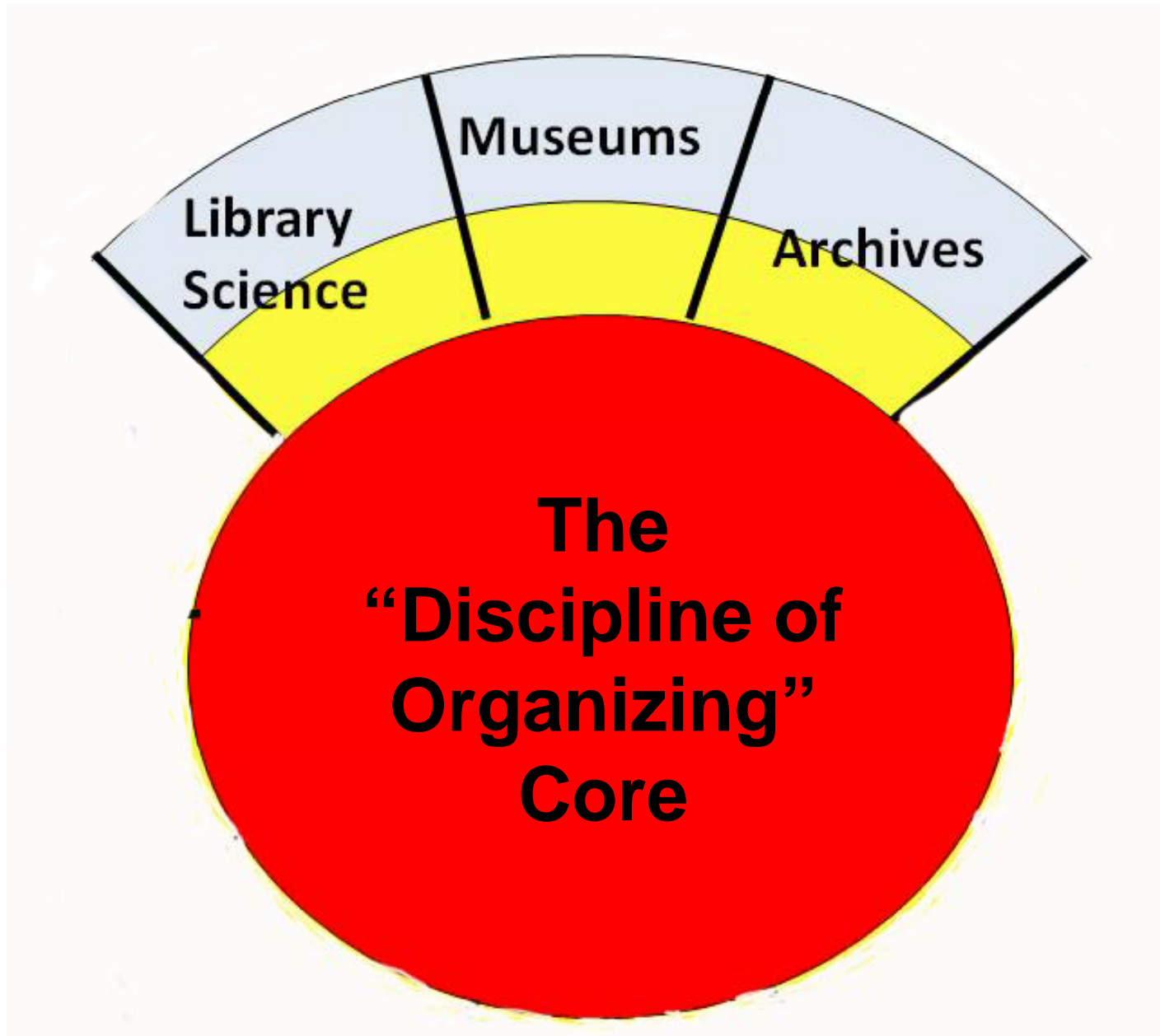
All the endnotes, no inline student quizzes

The Combinatorial Explosion of Design-Time Configuration

- With 11 disciplines the combinatorial possibilities create an extremely large “family of books” (2048)
- Even if we apply strong reasonableness or familiarity constraints it is still easy to imagine many appealing configurations of disciplines:
 - Memory Institutions (LIS, Museums, Archives)
 - Informatics (Computing, Info Architecture, Web, Business, Law)
 - Information Architecture (Info Architecture, Linguistics, Web)
 - Sensemaking (Cognitive Science, Linguistics, Philosophy)

...

“Memory Institutions” Configuration



Other Limitations of Design-Time Configuration

- Design-time configuration assumes a lot of homogeneity of the intended readers
- But TDO's target readers are mostly in "Schools of Information" like Berkeley's, which are multidisciplinary programs that attract students with different disciplinary backgrounds
- Different chapters and topics in TDO inherently differ in their multidisciplinary mix

“Reading Time” or Dynamic Multivalence

- We're now experimenting with a modified book-building process that converts the discipline tags in our XML source files to class attributes in the generated epub HTML
- This lets us use Javascript and CSS to insert controls into the ebooks to allow readers to selectively include and exclude content

Inclusion vs. Exclusion

- The conventional mechanisms for dealing with supplemental content are inclusion by traditional link following (and return) or with pop-up notes
- But inclusion mechanisms don't tell the reader anything about what might be included
- We can consider inverse mechanisms that exclude content, where we can more easily convey the consequences of the reader's act
- What's the right default? Include all the content and let the reader exclude some, or include only the core and have the reader make inclusion choices?

Design Issues

- A reader's decision to read supplemental content is shaped by
 - the proportion of supplemental content to core content
 - the disciplinary mix of the supplemental content
 - the nature of the core content
- And also by the specific context in which the supplemental content is encountered

Chapter 4 Endnotes

Note Type	#	% of total	# chars	avg	% of total chars
Computing	16	25%	2592	162	13%
Web	7	11%	1564	223	8%
IA	2	3%	616	308	3%
Cog Sci	5	8%	3049	610	15%
Linguistics	4	6%	1241	310	6%
Philosophy	0	0%	0		0%
LIS	17	27%	5266	310	26%
Museums	3	5%	313	104	2%
Archives	0	0%	0		0%
Law	1	2%	441	441	2%
Business	8	13%	5344	668	26%
TOTAL	63	100%	20426	324	100%

Chapter 4. Resource Description and Metadata

Robert J. Glushko

Kimra McPherson

Ryan Greenberg

Matthew Mayernik

4.1. Introduction

Click. A professional photographer standing on a mountainside takes a picture with a digital camera. What information should be recorded and associated with the recorded image of the mountain scene? Modern cameras assign an identifier to the stored photograph and they also capture the technical description of the image's production: the type of camera, lens, shutter speed, light sensitivity, aperture, and other settings.^{190[Com]} Many modern cameras also record information about the geographic and temporal circumstances surrounding the image's creation: the date, time and location on Earth where the photo-



Proportion of supplemental content per chapter

Using Tile to Indicate Proportions of Includable Content

Chapter 4. Resource Description and Metadata

Robert J. Glushko

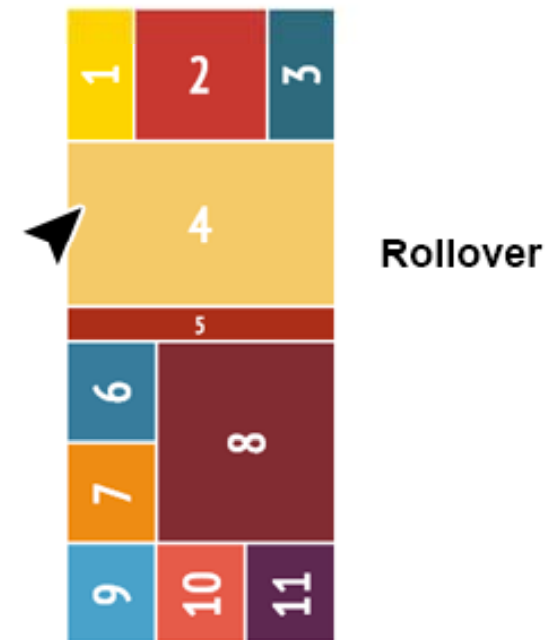
Kimra McPherson

Ryan Greenberg

Matthew Mayernik

4.1. Introduction

Click. A professional photographer standing on a mountainside takes a picture with a digital camera. What information should be recorded and associated with the recorded image of the mountain scene? Modern cameras assign an identifier to the stored photograph and they also capture the technical description of the image's production: the type of camera, lens, shutter speed, light sensitivity, aperture, and other settings.^{190[Com]} Many modern cameras also record information about the geographic and temporal circumstances surrounding the image's creation: the date, time and location on Earth where the photo-



Chapter 4. Resource Description and Metadata

Robert J. Glushko

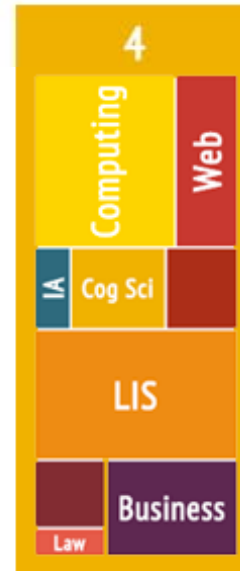
Kimra McPherson

Ryan Greenberg

Matthew Mayernik

4.1. Introduction

Click. A professional photographer standing on a mountainside takes a picture with a digital camera. What information should be recorded and associated with the recorded image of the mountain scene? Modern cameras assign an identifier to the stored photograph and they also capture the technical description of the image's production: the type of camera, lens, shutter speed, light sensitivity, aperture, and other settings.^{190[Com]} Many modern cameras also record information about the geographic and temporal circumstances surrounding the image's creation: the date, time and location on Earth where the photo-



Content breakdown appears

Chapter 4. Resource Description and Metadata

Robert J. Glushko

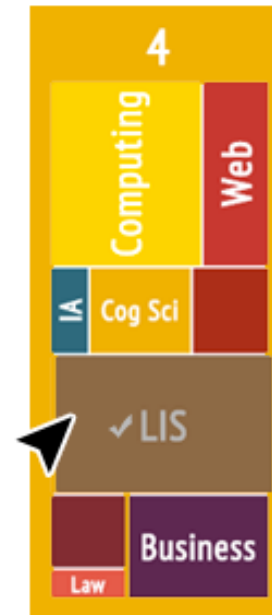
Kimra McPherson

Ryan Greenberg

Matthew Mayernik

4.1. Introduction

Click. A professional photographer standing on a mountainside takes a picture with a digital camera. What information should be recorded and associated with the recorded image of the mountain scene? Modern cameras assign an identifier to the stored photograph and they also capture the technical description of the image's production: the type of camera, lens, shutter speed, light sensitivity, aperture, and other settings.^{190[Com]} Many modern cameras also record information about the geographic and temporal circumstances surrounding the image's creation: the date, time and location on Earth where the photo-



Click

Chapter 4. Resource Description and Metadata

Robert J. Glushko

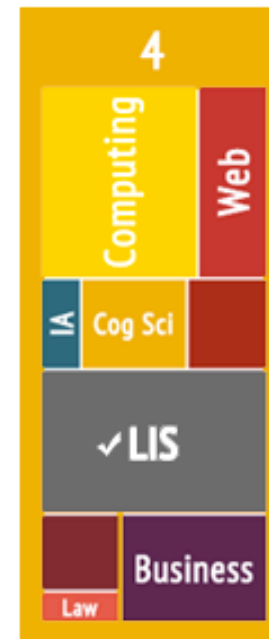
Kimra McPherson

Ryan Greenberg

Matthew Mayernik

4.1. Introduction

Click. A professional photographer standing on a mountainside takes a picture with a digital camera. What information should be recorded and associated with the recorded image of the mountain scene? Modern cameras assign an identifier to the stored photograph and they also capture the technical description of the image's production: the type of camera, lens, shutter speed, light sensitivity, aperture, and other settings.¹⁹⁰[Com] Many modern cameras also record information about the geographic and temporal circumstances surrounding the image's creation: the date, time and location on Earth where the photo-



Content added

more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

The purposes and nature of bibliographic description are the foundation of library and information science and have been debated and systematized for nearly two centuries. *Bibliographic descriptions* characterize information resources and the entities that populate the bibliographic universe, which include works, editions, authors, and subjects. Despite the “biblio-” root, bibliographic descriptions are applied to all of the resource types contained in libraries, not just books. Note also that this definition includes not just the information resources being described as distinct instances, but also as sets of related instances and the nature of those relationships.¹⁹⁵[LIS]

A bibliographic description of an information resource is typical-

Using Icons to Indicate Scope of Excludable Content

LIS

LIS

more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

The purposes and nature of bibliographic description are the foundation of library and information science and have been debated and systematized for nearly two centuries. *Bibliographic descriptions characterize information resources and the entities that populate the bibliographic universe, which include works, editions, authors, and subjects.* Despite the “biblio-” root, bibliographic descriptions are applied to all of the resource types contained in libraries, not just books. Note also that this definition includes not just the information resources being described as distinct instances, but also as sets of related instances and the nature of those relationships.¹⁹⁵[LIS]

A bibliographic description of an information resource is typical-



Rollover or touch

On roll out, the X disappears.

Touch out side to get rid of X.



more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

The purposes and nature of bibliographic description are the foundation of library and information science and have been debated and systematized for nearly two centuries. *Bibliographic descriptions characterize information resources and the entities that populate the bibliographic universe, which include works, editions, authors, and subjects.* Despite the “biblio-” root, bibliographic descriptions are applied to all of the resource types contained in libraries, not just books. Note also that this definition includes not just the information resources being described as distinct instances, but also as sets of related instances and the nature of those relationships.^{195[LIS]}

A bibliographic description of an information resource is typical-

LIS

Fades out

LIS

more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

A bibliographic description of an information resource is typical-

more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

Slides Up

A bibliographic description of an information resource is typically realized as a structured record in a standardized format that describes a specific resource. The earliest bibliographic records in the nineteenth century were those in book catalogs, which organized for each author a list of his authored books, with separate entries for each edition and physical copy. Relationships between books by different authors were described using cross-references.

LIS

The nature and extent of bibliographic descriptions were highly

LIS

more inclusive and more easily understood alternative to two terms that are well established in organizing systems for information resources: *bibliographic descriptions* and *metadata*. We will also distinguish resource description as a general concept from the narrower senses of tagging of web resources and the *Resource Description Framework (RDF)* language used to make statements about web resources and physical resources that can be identified on the Web.

4.2.2.1. Bibliographic Descriptions

The purposes and nature of bibliographic description are the foundation of library and information science and have been debated and systematized for nearly two centuries. *Bibliographic descriptions* characterize information resources and the entities that populate the bibliographic universe, which include works, editions, authors, and subjects. Despite the “biblio-” root, bibliographic descriptions are applied to all of the resource types contained in libraries, not just books. Note also that this definition includes not just the information resources being described as distinct instances, but also as sets of related instances and the nature of those relationships.^{195[LIS]}

A bibliographic description of an information resource is typical-

Using a Bar to Indicate Scope of Excludable Content

Summary

- Multidisciplinary books can balance the competing goals of breadth and depth by marking text by audience, complexity, or other attributes
- If done carefully, this creates a “mother of all books” from which many different books can be generated
- With appropriate user interface controls, readers can selectively include and exclude content to optimize their reading experience
- We will eventually figure out “appropriate”