The Discipline of Organizing

Robert J. Glushko

glushko@berkeley.edu

CMU BrainHub Lecture

16 June 2015

Today’s Talk

- Project Motivation
- The “Organizing System”
- Design Dimensions and Frameworks for Organizing Systems
- The Book, eBooks, Customization, and Collaboration
Motivation (Personal)

- In our daily lives organizing is a common personal and group activity that we often do without thinking much about it.
- It is also an important part of most business and professional activities.
- Organizing in any context can be more effective and satisfying if we are more self-aware and systematic about how we organize.

Bob’s Garage
Motivation

Making sense of the world is to organize it...

- Categorization (Creating “equivalence classes” of resources that we treat the same)
- Classification (Creating models for assigning resources to existing categories)
- Integration (Combining categories)
- Segmentation (Discovering categories computationally, assigning resources to them)
- Recommendation (Identifying “matching” resources in different categories)
- …

Motivation (Professional)

- Teaching the UC Berkeley School of Information “gateway” course since 2005
- The ~60 ISchools are highly diverse in disciplinary focus
  - As well as in student populations, typical employers, etc.
  - *What should an ISchool core course teach?*
The Undefined Intersection of “Information, People, and Technology”

- No textbook existed that explained this intersection {confluence, synthesis, synergy, blahblah…}

Organizing

- Organizing: *Creating capabilities by intentionally imposing order and structure*

- We organize:
  - Things
  - Information
  - Information about Things
  - Information about Information about {Things, Information}
  - …
We Organize…

- Libraries, museums, business information systems, scientific data… and other institutional resource collections
- Different types of documents – from narrative to transactional – which have characteristic content, structures, and presentations
- Personal information and artifacts of all kinds in our kitchens, closets, personal computers, smartphones…
- People
Motivating the Concept of “Organizing System”

- We can emphasize how all of these domains and types of collections differ… or we can emphasize what they have in common.
- They are all “Organizing Systems”

A collection of resources
intentionally arranged
to enable some set of interactions

The Organizing System [1]

- RESOURCES are “anything of value that can support goal-oriented activity”
- A COLLECTION is a group of resources that have been selected for some purpose
The Organizing System [2]

- **INTENTIONAL ARRANGEMENT** captures the idea that the system requires explicit or implicit acts of organization by **AGENTS** – human or computational ones
- These arrangements follow or embody one or more **ORGANIZING PRINCIPLES**, ideally expressed in an *implementation-neutral* manner
- The agents might also organize themselves in a bottom-up or “collective intelligence” fashion, especially human agents

Organizing Principles [1]

- Almost any property of a resource might be used as a basis for its arrangement, and multiple properties are often used simultaneously
- For physical resources the properties are often perceptual, material ones, or task-oriented ones
- For information resources the properties are often semantic ones
Organizing Books By Content

(LOC Classification)

Organizing Spices by Cuisine
Organizing Principles [2]

- Other typical arrangements are based on ownership, origin, taxonomic, or “taskonomic” properties (usage frequency, correlated usage)
- Any resource with a orderable name or identifier can have alphabetic or numeric ordering
- Any resource with an associated date (creation, acquisition) can have chronological ordering

Organizing People by Work Role
Organizing People by Family, Religion, Class, Race, Year of Death…

The Best Principles don’t Specify Implementation: “Organize Spices Alphabetically”
The Three-Tier Architecture

Organizing Principles should be logically separated from Implementation and Presentation considerations

The Activities in Organizing Systems

- We can identify four activities in the lifecycle of every organizing system:
  - Selecting resources
  - Organizing resources
  - Supporting resource-based interactions and services
  - Maintaining resources
Activities in a Closet Organizing System...

- Selecting: Should I hang up my sweaters in the closet or put them in a drawer?
- Organizing: Should I sort my shirts by color, sleeve type, or season?
- Supporting Interactions: Do I need separate places for laundry or dry cleaning?
- Maintaining: Should I toss out my clothes only when they wear out, based on how long I’ve owned them, or based on whether I’m tired of them?
Activities in a Data Warehouse Organizing System…

- **Selecting**: *which data sources should be included?* How is their quality assessed?
- **Organizing**: *which data formats and schemas will enable effective processing?* Are needed transformations made at load time or query time?
- **Supporting Interactions**: *what are the most important and frequent queries that need to be pre-configured?*
- **Maintaining**: *data governance… retention, compliance, privacy issues*

Organizing Organizing Systems

- We can classify organizing systems by:
  - resource type
  - dominant purpose
  - creator
  - size of intended user community
  - or many other ways

- But these classifications overlap without clear boundaries or necessary and sufficient features
Categorizing by Resource Type

- Organizing Systems
  - Collections of Books (Libraries)
  - Collections of Art (Museum)
  - Collections of Documents (Archive)
  - Collections of Data (Repository)
  - Collections of Spices (Pantry) ...

Categorizing by Purpose: Resource Preservation as Means vs. End

- Organizing Systems
  - Memory Institutions
    - Libraries
    - Museums
    - Archives
  - Business Info Systems
    - Content Management
    - CRM
    - ERP
What is a Library?

• A collection of resources
  • Organized to enable “access” and “reuse”
  • Curated for “public good” and “community creation”
  • Conventional interaction is “circulation” – borrowing and return of resources…

A Library
A Library?

Tool Lending Library

Branch Address
5205 Telegraph Avenue
Oakland, CA 94609
(510) 597-5089

Branch Hours
Monday: Closed
Tuesday: 12:30pm - 8:00pm
Wednesday: 10:00am - 5:30pm
Thursday: 10:00am - 5:30pm
Friday: 12:00pm - 5:30pm
Saturday: 10:00am - 5:30pm
Sunday: Closed
A Library?

**Consequences of Category Thinking**

- Many types of resource collections have conventional characteristics that are deeply embedded in cultural and linguistic categories.
- Using an established category to describe an organizing system reinforces these characteristics, even if we add qualifiers.
- … and marginalizes any atypical characteristics of the organizing system being categorized.
A “Design Space” or “Dimensional” Perspective

- In addition to using categories like Library or Museum or Business Information System, consider a specific organizing system as a point in a multidimensional design space and these categories as regions in that space...
- This treats the familiar categories as “design patterns” that embody typical configurations of design choices

Consequences of Dimensional Thinking

- Overcomes the bias and conservatism inherent in familiar categories
- Design patterns support multi-disciplinary work that cuts across familiar categories and applies knowledge about them to new domains
- Creates a design vocabulary for translating concepts and concerns from category and discipline-specific vocabularies
The 5 Dimensions of an Organizing System

1. What Is Being Organized?
2. Why Is It Being Organized?
3. How Much Is It Being Organized?
4. When Is It Being Organized?
5. Who (or What) is Organizing It?

1. What Is Being Organized?

• Identifying the unit of analysis is a central problem in every intellectual or scientific discipline - and in every organizing system

• Resources that are aggregates or composites of other resources, or that have internal structure, or that can have many attributes, pose questions about the granularity of their "thingness"
How Many Things is a Car?

When you build it? When you sell it?
When you repair it?

“Thing” vs. “Type of Thing”

• It is easy to blur the distinction between individual things or instances of things and classes of things
• We often say 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
What is Macbeth?

ABSTRACTION HIERARCHY OF THE WORK

Work → "Macbeth"

Expression → "Macbeth" (play, drama, novel, etc.)

Manifestation → "Macbeth (Dover Thrift Editions)", 1993 edition

Item

The copy of "Macbeth" which you bought in high school and currently sits on your bookshelf.

"Shamu" -- Instance or Type?
Resource Focus

• We often designate some resource as primary because it is the focus of our attention

• We often create other resources that are descriptions of or otherwise associated with the primary resource

• We call these “Description resources” (a more general term than “metadata”)
Fantasy Football: One Person’s Description is another Person’s Resource
But Description is Challenging!

- People use different words for the same things, and the same words for different things - what would a "good" description be like, and how can it be created or discovered?
- Describing and organizing always (explicitly or implicitly) takes place in some context
- The context shapes which resource properties are important and the organizing principles that use those properties, introducing bias

Organizing Books by Color

Trading aesthetics for scalability in organization and retrieval

Photo by See-ming Lee (http://www.flickr.com/photos/seeminglee/4556156477) Creative Commons CC BY-SA 2.0 license
A DJ Organizes His Records – Beats per Minute

"This is the Hip-Hop and Dancehall (Jamaican) section of my record collection."

"7" is the format Jamaican singles are still released, and these are two of my 5 baskets. In an ideal world, the baskets are loosely arranged by BPM (Beats Per Minute), roughly how fast the music is. This tells the DJ what records generally mix with what other records."

"When I return hip-hop records to the shelf, they just go in the front of that particular pile, with the net result that stuff that gets played less drifts to the back.."

~87-97 BPMs
~98-110 BPMs

Photo by Matt Earp aka Kid kameleon

Classifying Resource Properties

Property Essence

<table>
<thead>
<tr>
<th>Intrinsic Static</th>
<th>Extrinsic Static</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition:</strong> Directly experienced, subject matter, implicit, inherent properties.</td>
<td><strong>Definition:</strong> Assigned to resource, name, identifier.</td>
</tr>
<tr>
<td><strong>Examples:</strong> Size, color, shape, author, date of creation.</td>
<td><strong>Examples:</strong> Dewey decimal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Intrinsic Dynamic</th>
<th>Extrinsic Dynamic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition:</strong> Inherent properties; change over time.</td>
<td><strong>Definition:</strong> Behavioral and contextual properties</td>
</tr>
<tr>
<td><strong>Examples:</strong> Skills, experience</td>
<td><strong>Examples:</strong> Current owner, location, best seller lists.</td>
</tr>
</tbody>
</table>

Property Persistence

Static

Dynamic
Resource Properties for Kitchen Organization

• **Intrinsic static properties**: Store pots, pans, and dishes in different cabinets and nest by size

• **Extrinsic static properties**: A spice rack with the spices arranged in alphabetical order

• **Intrinsic dynamic properties**: Arrange perishable goods by expiration date, a “useful life remaining” property that decreases to zero over time

• **Extrinsic dynamic properties**: Put the most frequently used condiments or spices in the front
Resource Properties for 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

Summary of TDO as a Discipline

- The concept of Organizing System unifies a vast multidisciplinary body of design and analysis practice

- Design dimensions overcome the limitations and inertia of the traditional categories

- Abstractions about resource properties and organizing principles enable conversations between people who lacked common language

- It is a generative, forward-looking approach that encourages and accommodates innovation while preserving conventional theory and practice as design patterns
Case Studies in the 2014 Edition

11.1. A Multi-generational Photo Collection
11.2. Knowledge Management for a Small Consulting Firm
11.3. Smarter Farming in Japan
11.4. Single-Source Textbook Publishing
11.5. Organizing a Kitchen
11.6. Netflix
11.7. Luxury Brand Store
11.8. Weekly Newspaper
11.9. The CODIS DNA Database
11.10. IKEA
11.11. The Antikythera Mechanism
11.12. My Vegetable Garden
11.13. IP Addressing in the Global Internet
11.14. Knitting Supplies
11.15. Making a Documentary Film
11.16. Open Knowledge Management for Chemical Hazard Assessment and Alternatives Analysis
11.17. Managing Information About Data Center Resources
11.18. Neuroscience Lab

The Discipline of Organizing as a Book

Published by MIT Press in May 2013 as a printed book and as epub3 and Kindle ebooks

Two “enhanced ebook editions” were published by O’Reilly Media in Aug 2014

Third “polyvalent” ebook edition aiming for July 2015

In use in nearly 60 courses in 20 countries as of June 2015

DisciplineOfOrganizing.org
The First Editions

• The first print and ebook digital editions (epub and Kindle mobi) were produced from the same source files using the O'Reilly Atlas publishing system

• They are essentially identical except for the interactions like search and hyperlinking that are intrinsic to the digital formats

Investing in Markup

• We waited for two months while the printed books were published and worked their way into distribution channels

• We decided to invest heavily in markup that was not used in the print version

• We produce enhanced ebooks now

• And this markup will be even more valuable in “open data” or “semantic web” contexts in the future
Key Points in Chapter Two

Which activities are common to all organizing systems?

Selection, organizing, interaction design, and maintenance activities occur in every organizing system.

See section-2.1.

Selection, organizing, interaction design, and maintenance activities occur in every organizing system. These activities are not identical in every domain, but the general terms enable communication and learning about domain-specific methods and vocabularies. The most fundamental decision for an organizing system is determining its resource domain, the group or type of resources that are being organized. Even when the selection principles behind a collection are clear and consistent, they can be unconventional, idiosyncratic, or otherwise biased.
Key Points in Chapter Two

Which activities are common to all organizing systems?

Selection, organizing, interaction design, and maintenance activities occur in every organizing system. See §2.1, “Introduction.”

These activities are not identical in every domain, but the general terms enable communication and learning about domain-specific methods and vocabularies. See §2.4, “Introduction.”

Quiz Mode” Presents the Questions in the eBook

Self-Review

Question

Which activities are common to all organizing systems?

Show Answer  Next Question Previous Question Back to Key points
The Mandate and Challenge with Multiple Disciplines

- The concept of “organizing system” as the transdisciplinary synthesis of the disciplines that deal with “organizing” mandates a book with many authors
- It must balance the breadth of representing all the contributing disciplines with enough depth for each to be credible
- It must introduce vocabulary that is discipline-neutral to enable interdisciplinary communication
- It must incorporate discipline-specific concepts and examples in the context of the transdisciplinary content and vocabulary

Battling the Breadth vs. Depth Challenge

- Using the draft book at different schools increased its breadth, but every adopter and reviewer would suggest topics and examples from their own disciplines to “flesh out the book”
- The book grew bigger and bigger and bigger...making it more credible to experts but less accessible for students
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
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 invokable 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 (Bursalou 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.
Collaborative Authoring
Implications for Book Architecture (1)

• A “bottom up or “emergent consensus” approach
  • Assembles disciplinary experts to develop a shared plan
  • Imposes relatively weak constraints on possible changes to the book; any part is potentially revisable because all authors have comparable authority
Collaborative Authoring
Implications for Book Architecture (2)

• In TDO’s “top down” approach
  • One person proposes a vision and outline for the book and recruits experts with complementary expertise to become co-authors” with specific topical and disciplinary writing tasks
  • Evolution takes place via annotation “at the leaves” with new examples, sidebars, and endnotes, but the overall structure and sequence of the book’s table of contents is mostly preserved

“Polyvalence” – A Family of Books

• Our disciplinary tagging turned our book source content into a “polyvalent” 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 and conditional compilation enables the creation of many similar software systems from a shared set of software assets
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

No endnotes, but inline student quizzes

All the endnotes, no inline student quizzes

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 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 Personalization

• We modified the book-building process to convert 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

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 2. Activities in Organizing Systems
Robert J. Glushko
Erik Wilde
Jess Hemeny

2.1. Introduction

Chapter 8. The Forms of Resource Descriptions
Ryan Shaw
Murray Maloney

8.1. Introduction
Summary of Ebooks and Personalization

- 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 information visualizations and user interface controls for specifying content to be included and excluded, readers can personalize their reading experience with any configuration of disciplines.
- We expect to have this capability in the 3rd edition TDO for use in fall 2015 courses.

But We’re Not Done:
The Vision of an Extensible “Network” Book

- Any instructor or institution should be able to create supplemental content.
- An XML-savvy instructor should be able to publish a local edition with this supplemental content by configuring the “build instructions” in an XML editor.
- This content can be submitted for incorporation in new standard editions published on a regular cycle.
- Supplemental content should be discoverable from anywhere in the “network textbook” at any time.
The Extensible “Network” Textbook

A DEPLOYED EBOOK

Content discovery and inclusion from non-shared resources (e.g., each university or course)

Pre-packaged channels

Online reading... content discovery and inclusion from shared repository (e.g., used by all the Information Schools)

Offline reading... no content discovery possible

With the help of:

> 20 principal and contributing authors since 2011
> Berkeley TAs, GSRs, and students
> 50 instructors, > 1000 students

markup guru and production editor Murray Maloney

programmers Bob Stayton, Alex Milowski, Pablo Arvizu, Jirka Kosek

info viz Katey Basye, Isabelle Sperano