

The Discipline of Organizing

Robert J. Glushko

Information Architecture

EDITOR'S SUMMARY

It is normal to organize our world, but doing so systematically is key and the subject of the book *The Discipline of Organizing (TDO)*. The driving concept is that, while organization of resources is fundamental to library and information science, it is a central issue for many professional fields employing different organizational strategies and descriptive vernacular. To bring the diverse perspectives together, a broadly applicable, abstract framework can be used, based on an assessment of what is being organized, why, how much, when and by what means. These points of analysis of the resources to be organized inform organizational design decisions, considerations of stakeholders and costs and strategic planning for tools and methods. Principles underlying an organization system's design may draw on frequency of resource use or coordination of items, alphabetic or chronological ordering or unique approaches to manage hybrid and novel resources. The TDO philosophy reflects an information management approach that spans disciplinary silos and avoids field-limited terminology, while building the critical skills of resource organization and management.

KEYWORDS

organization of information	information resources management
knowledge organization systems	terminology
analytic models	interdisciplinarity

Robert Glushko is an adjunct full professor in the School of Information, University of California, Berkeley. His recent book, *The Discipline of Organizing*, has received much attention and is being widely adopted as a text in iSchool and other programs. He has over 30 years of R&D, consulting and entrepreneurial experience in information systems and service design, content management, electronic publishing, Internet commerce and human factors in computing systems. He can be reached at glushko@berkeley.edu.

What do books in libraries, spices in kitchen pantries, boats in marinas, weather observations in a data repository, songs on a smartphone or music player, paintings in a museum, animals in a zoo and a professor's lecture notes on his personal computer have in common?

At first glance, the answer seems to be “absolutely nothing.” This list contains highly diverse things that are selected and organized according to different principles, kept in different physical or digital environments and used for different purposes for different types of users.

But if we stand back a bit and take a more abstract look, we can see that all the things can be seen as the same thing. Books, spices, boats, weather data, music files, paintings, animals and lecture notes are all “resources” – “things that have value that can support goal-oriented activity” – that have been intentionally selected and organized. Similarly, despite their obvious differences, libraries, pantries, marinas, data repositories, recorded music collections, museums, zoos and computer folder and file hierarchies can each be described as an “organizing system” – “an intentionally arranged collection of resources and the interactions they support.”

A set of resources is transformed by an organizing system when the resources are described or arranged to enable interactions with them. Explicitly or by default, this transformation requires many interdependent decisions about the identities of resources; their names, descriptions and other properties; the classes, relations, structures and collections in which they participate; and the people or technologies that interact with them. These decisions and the analysis needed to make them have been systematized in *The Discipline of Organizing*, recently published by MIT Press in both print and eBook formats [1].

The Discipline of Organizing (TDO) compares and contrasts how

organizing takes place in different contexts and domains, presents common principles and design patterns and proposes that organizing systems typically follow a common life cycle of resource selection, organizing, interaction design, and maintenance. The book's terminology is trans-disciplinary and generic in order to demonstrate its applicability to many different domains, with a great many specific examples that illustrate how organizing takes on many different forms even when the underlying principles remain the same.

I initiated the project to write *TDO*, assembled nearly 20 co-authors and guided them as the principal author and editor to create a book that defines a new field while respecting the essential contributions of the “feeder” disciplines – notably library and information science, computer science, informatics, cognitive science, law, economics and business. In this brief essay I will present the foundation ideas of *TDO* and describe the experiences we've had using the book in a diverse set of university courses in library science, informatics and “Information School” programs.

Organizing Is Ubiquitous, but We Rarely Think About It

Organizing is such a common activity that we often do it without thinking much about it. We organize the clothes in our closet, the books on our bookshelves, the tools in our garage and the folders into which we file printed and digital records for tax and other purposes. Quite a few of us have jobs that involve specific types of organizing tasks. We might even have been explicitly trained to perform them by following specialized disciplinary practices. We might learn to do these tasks very well, but even then we often do not reflect on the similarity of the organizing tasks we do and those done by others or on the similarity of those we do at work and those we do at home. We take for granted and as givens the concepts and methods used in the organizing system we work with most often.

However, there is a cost to taking organizing for granted. The properties of resources you choose as the basis for organizing them make some interactions easy but might make other interactions difficult or even impossible. Arranging things by color and size makes sense in your clothes closet but not in your refrigerator. If you live alone, “frequency of use” is an effective organizing principle for your spices, cooking utensils and other resources in

your kitchen, but if you have a roommate, any principle based on individual behavior rather than static resource properties is bound to cause conflicts.

A discipline of organizing provides concepts and guidance to make these day-to-day organizing activities more effective and even enjoyable. You might appreciate the aesthetic appearance of color-sorted books on your living room bookshelves more when you imagine the impossibility of using this organizing principle in a research library.

Organizing from a “Siloed” Perspective

No one reading an article in the *Bulletin of the Association for Information Science and Technology* would dispute that organizing is a fundamental issue in many professional fields. However, these fields have only limited agreement in how they approach problems of organizing and in what they seek as their solutions. For example, library and information science has traditionally studied organizing from a public sector bibliographic perspective, paying careful attention to user requirements for access and preservation and offering prescriptive methods and solutions. In contrast, computer science and informatics tend to study organizing in the context of information-intensive applications with a focus on process efficiency, system architecture and implementation.

In addition, we most often make distinctions about organizing on the basis of the type of resources being organized. We contrast law libraries from tool libraries, knowledge management systems from data warehouses and personal stamp collections from coin collections primarily because they contain different kinds of resources. Similarly, we distinguish document collections by resource type, giving them different names, as when we contrast narrative document types like novels and biographies with transactional ones like catalogs and invoices, with hybrid forms like textbooks and encyclopedias in between.

Finally, even though the activity of adding a resource to a collection occurs in all organizing endeavors, each of the “organizing professions” uses specialized vocabulary for describing it. For example, adding a resource to a library collection is called acquisition, adding to a museum collection is called accessioning and adding to an archive is ingesting. In

business information systems, adding resources could involve loading, integrating or inserting. Similar diversity in vocabulary occurs with the activity of maintaining the collection and interactions with it over time. Libraries and museums engage in preservation and curation, but the analogous activities with business information systems are often described as data cleaning, data cleansing, governance or compliance.

Introducing the Discipline of Organizing

Library and information science, informatics, computer science and other fields focus on the characteristic types of resources and collections that define those disciplines. This focus spawns disciplinary and domain-specific vocabulary that makes it challenging to apply concepts, methods and insights across disciplines.

In contrast, the fundamental premise of the discipline of organizing is that the diverse perspectives and concepts about organizing from its feeder disciplines can be subsumed and synthesized under a more abstract framework. *TDO* complements the focus on specific resource and collection types with a framework that views organizing systems as existing in a multi-dimensional design space in which we can consider many types of resources at the same time and see the relationships among them. *TDO* introduces five groups of design decisions, phrased in generic language to emphasize their broad applicability:

- What is being organized?
- Why is it being organized?
- How much is it being organized?
- When is it being organized?
- By what means is it being organized?

In the following sections I will briefly describe each of these decision dimensions.

What's being organized?

What is the scope and scale of the resource domain? What is the mixture of physical things, digital things and information about things to be organized? Is the organizing system being designed to enable a resource

collection to be created for an existing and closed resource collection, or for a collection in which resources are continually added or deleted? Are the resources unique or are they interchangeable members of a class?

Before we can begin to organize any resource we often need to identify it. It might seem straightforward to devise an organizing system around tangible resources, but we must be careful not to make assumptions about resources. In different situations, the same thing can be treated as a unique item, as one of many equivalent members of a broad category or as component of an item rather than as an item on its own. For example, in a museum collection, a hand-carved chess piece might be identified as an individual entity, as part of a set of chess pieces or be treated as one of the 33 unidentified components of an item identified as a chess set. When the resources being organized consist of information content, deciding on the unit of organization is challenging because it might be necessary to look beyond physical properties and consider conceptual or intellectual equivalence. A high school student told to study Shakespeare's play *Macbeth* might treat any printed copy or web version as equivalent and might even try to outwit the teacher by watching a film adaptation of the play. To the student, all versions of *Macbeth* seem to be the same resource, but librarians and scholars make much finer distinctions.

Why is it being organized?

What interactions or services will be supported and for whom? Are the uses and users known or unknown? Are the users primarily people or computational processes? Does the organizing system need to satisfy personal, social or institutional goals? Does the organizing system have a limited timeframe, or must it and the resources it contains be maintained indefinitely?

Almost by definition, the essential purpose of any organizing system is to describe or arrange resources so they can be located and accessed later. The organizing principles needed to achieve this goal depend on the types of resources or domains being organized, and in the personal, social or institutional setting in which organization takes place. The fine distinctions between organizing systems that have many characteristics in common

reflect subtle differences in the priority of their shared goals. For example, many organizing systems create collections and enable interactions with the goals of supporting scientific research, public education and entertainment. We can contrast zoos, animal theme parks and wild animal preserves in terms of the absolute and relative importance of these three goals, for example.

When the scale of the collection or the number of intended users increases, not everyone is likely to share the same goals and design preferences for the organizing system. In formal or institutional organizing systems conflicts between stakeholders can be more severe, and the organizing principles might even be specified in commercial contracts or governed by law. For example, physicians view the creation of patient records as central to diagnosis and treatment, while insurance companies think of them as evidence needed for payment and reimbursement and researchers think of them as primary data.

How much is it being organized?

What is the extent, granularity or explicitness of description, classification or relational structure being imposed? What organizing principles guide the organization? Are all resources organized to the same degree, or is the organization sparse and non-uniform?

The organizing system for a small collection can sometimes use only the minimal or default organizing principle of co-location – putting all the resources in the same container, on the same shelf or in the same email inbox. But as a collection grows in size, the time to arrange, locate and retrieve a particular resource becomes more important and the collection must be explicitly organized to make these interactions efficient. As a result, most organizing systems employ organizing principles that make use of properties of the resources being organized (for example, name, color, shape, date of creation, semantic or biological category), and multiple properties are often used simultaneously.

Unlike those for physical resources, the most useful organizing properties for information resources are those based on their content and meaning, and these properties are not directly apparent when you look at a book or document. Significant intellectual effort or computation is necessary to reveal

these properties when assigning subject terms or creating an index. The most effective organizing systems for information resources often are based on properties that emerge from analyzing the collection as a whole. For example, the relevance of documents to a search query is higher when they contain a greater than average frequency of the query terms compared to other documents in the collection, or when they are linked to relevant documents.

Different preferences and disagreements between stakeholders in an organizing system about how much organization is necessary often result because of the implications for who does the work and who gets the benefits. Physicians prefer narrative descriptions and broad classification systems because those practices make it easier to create patient notes. In contrast, insurance companies and researchers want fine-grained, “form-filling” descriptions and detailed classifications that would make the physician’s work more onerous.

The cost-effectiveness of creating systematic and comprehensive descriptions of the resources in an information collection has been debated for nearly two centuries, and in the last half century the scope of the debate has grown to consider the role of computer-generated resource descriptions.

When is it being organized?

Is the organization imposed on resources when they are created, when they become part of the collection, when interactions occur with them, just in case, just in time, all the time? Is any of this organizing mandated by law or shaped by industry practices?

The organizing system framework recasts the traditional tradeoffs between information organization and information retrieval as the decision about when the organization is imposed. We can contrast organization imposed on resources “on the way in” when they are created or made part of a collection with “on the way out” organization imposed when an interaction with resources takes place.

The organizing systems perspective makes it easier to identify and apply the inherent tradeoffs between information organization and information retrieval that are obscured by the silos of traditional disciplinary and category perspectives. It is clear that the more effort put into organizing

information or other resources, the more effectively they can be retrieved, and the more effort put into retrieving resources, the less they need to be organized first.

How or by whom, or by what processes, is it being organized?

Is the organization being performed by individuals, by informal groups, by formal groups, by professionals, by automated methods? Are the organizers also the users? Are there rules or roles that govern the organizing activities of different individuals or groups? Is this organization imposed in a centralized, top-down manner or in a distributed, bottom-up manner?

Professional indexers and cataloguers undergo extensive training to learn the concepts, controlled descriptive vocabularies and standard classifications in the particular domains in which they work. Many of today's content creators are unlikely to be professional organizers, but presumably the creator best understands why something was created and the purposes for which it can be used. Non-creator users in the populace at-large are most often creating organization for their own benefit. Not only are these ordinary users unlikely to use standard descriptors and classifications, the organization they impose sometimes so closely reflects their own perspective and goals that it isn't useful or accurate for others.

The organizing systems view no longer contrasts information organization as a human activity and information retrieval as a machine activity, or information organization as a topic for library and information science and information retrieval as one for computer science. Instead, we readily see that computers now assist people in organizing and that people contribute much of the information used by computers to enable retrieval.

In particular, machine learning is a subfield of computer science that develops and applies algorithms that accomplish tasks that are not explicitly programmed; creating categories and assigning items to them is an important subset of machine learning. Two subfields of machine learning that are particularly relevant to organizing systems are supervised and unsupervised learning. In supervised learning, a machine-learning program is trained by giving it sample items or documents that are labeled by category, and the program learns to assign new items to the correct categories. In unsupervised

learning, the program gets the samples but has to come up with the categories on its own by discovering the underlying correlations between the items. Both approaches are fundamental as "big data" continues to spread like an epidemic from its roots in computer science into all the organizing disciplines.

Designing an Organizing System

The arrangements of resources in an organizing system follow one or more organizing principles, even if those principles are not explicit in the mind of the organizer. Organizing principles are directives for the design or arrangement of a collection of resources that are ideally expressed in a way that does not assume any particular implementation or realization.

When we organize a bookshelf, home office, kitchen or the MP3 files on our music player the resources themselves might be new and modern but many of the principles that govern their organization are thousands of years old. For example, we organize resources using easily perceived properties to make them easy to locate, we group resources that we often use together and we make resources that we use often more accessible than those we use infrequently. Commonly used organizing principles include alphabetical ordering (arranging resources according to their names) and chronological ordering (arranging resources according to the date of their creation or other important event in the lifetime of the resource). Some organizing principles sort resources into pre-defined categories and other organizing principles rely on novel combinations of resource properties to create new categories.

Expressing organizing principles in a way that separates design and implementation aligns well with the three-tier architecture familiar to software architects and designers: user interface (implementation of interactions), business logic (intentional arrangement) and data (resources).

The logical separation between organizing principles and their implementation is easy to see with digital resources. In a digital library it does not matter to a user if the resources are stored locally or retrieved over a network. How the resources and interactions with them are implemented are typically of little concern.

The separation of organizing principles and their implementation is harder to recognize in an organizing system that only contains physical

resources, such as your kitchen or clothes closet, where you appear to have unmediated interactions with resources rather than accessing them through some kind of user interface or “presentation tier.” Nevertheless, you can see these different tiers in the organization of spices in a kitchen. Different kitchens might all embody an alphabetic-order organizing principle for arranging a collection of spices, but the exact locations and arrangement of the spices in any particular kitchen depends on the configuration of shelves and drawers, whether a spice rack or rotating tray is used and other storage-tier considerations.

Categories {and vs. or} Design Dimensions for Organizing Systems

We can always create new categories for organizing systems by stretching the conventional definitions of *library* or other familiar ones and adding modifiers, as when Flickr is described as a web-based, photo-sharing library, or when we describe a collection of seeds for heirloom varieties as a “seed library” or “seed bank.” But whenever we define an organizing system with respect to a familiar category, the typical or mainstream instances and characteristics of that category that are deeply embedded in language and culture are reinforced, and those that are atypical are marginalized. In the Flickr case this means we suggest features that are not there (like authoritative classification) or omit the features that are distinctive (like tagging by users).

A similar categorization challenge arises with the Google Books digitization project. Google co-founder Sergei Brin characterized its ambitious project to put tens of millions of books from research libraries online as “a library to last forever.” But the Google Books project was widely criticized as not being true to library principles. We can readily identify design choices in Google Books that are more characteristic of the organizing systems in business domains, and the project might have been perceived more favorably had it been described as an online bookstore that offered many beneficial services for free.

In contrast, *TDO*’s dimensional perspective acknowledges the diversity of instances of collection types and provides a generative, forward-looking

framework for describing hybrid types that do not cleanly fit into the familiar categories. Even though it might differ from the conventional categories on some dimensions, an organizing system can be designed and understood by its “family resemblance” on the basis of its similarities on other dimensions to a familiar type of resource collection.

Thinking of organizing systems as points or regions in a design space makes it easier to invent new or more specialized types of collections and their associated interactions. If we think metaphorically of this design space as a map of organizing systems, the empty regions or “white space” between the densely populated centers of the traditional categories represent organizing systems that do not yet exist. We can consider the properties of an organizing system that could occupy that white space and analyze the technology, process or policy innovations that might be required to let us build it there.

Benefits of “Spanning the Silos” with Broader Concepts and Vocabulary

I teach in the School of Information at the University of California, Berkeley, which like most of the iSchools has a multidisciplinary curriculum and a diverse student population. ISchools often mix students with library science, computer science, social science or engineering undergraduate degrees in the same courses – even when the students have equally diverse career objectives. When we teach organizing using traditional library science or computer science texts, which narrowly focus on the concepts and terminology of a particular academic discipline and problem domain, students from other backgrounds have some difficulty. In contrast, teaching organizing from *TDO*’s broader perspective that emphasizes what different academic disciplines and domains have in common makes it easier for students with different backgrounds and working in different domains to understand and learn from each other.

Berkeley is just one of many iSchools or departments in similar fields that has adopted *TDO* as a core text or that is using it as a supplementary text to broaden and deepen the syllabus in more traditional courses in information organization or content management. These schools currently

GLUSHKO, continued

include Colorado, Humboldt, Illinois, Kentucky, Michigan, North Carolina, Rhode Island, Rutgers, St. Louis and Texas.

All of us teaching with *TDO* recognize that it can sound odd to describe the animals in a zoo as resources, to think of viewing a painting in a museum as an interaction or to say that destroying information to comply with privacy regulations is maintenance. But part of what a database administrator can learn from a museum curator follows from the rich associations the curator has accumulated around the concept of curation that are not available around the more general concept of maintenance. Without the shared concept of maintenance to bridge their disciplines, this learning could not take place.

A very practical implication of teaching organizing using more generic concepts and vocabulary is that it enables students to obtain jobs with firms

that might not otherwise hire them. A student who says she knows about *curation* can't as easily sell her skills to a business looking for someone to develop a business continuity plan as one who recognizes that "organizing resources and maintaining them over time" is the skill the company wants and the one she has. ■

Resource Mentioned in the Article

- [1] Glushko, R.J. (Ed.). (2013) *The discipline of organizing*. Cambridge, MA: MIT Press. This essay is in part derived from Chapter 1, "Foundations for Organizing Systems." For a book review in the mainstream press, see Chris Wright, "The Man Who Organized Everything," *Boston Globe*, August 25, 2013.