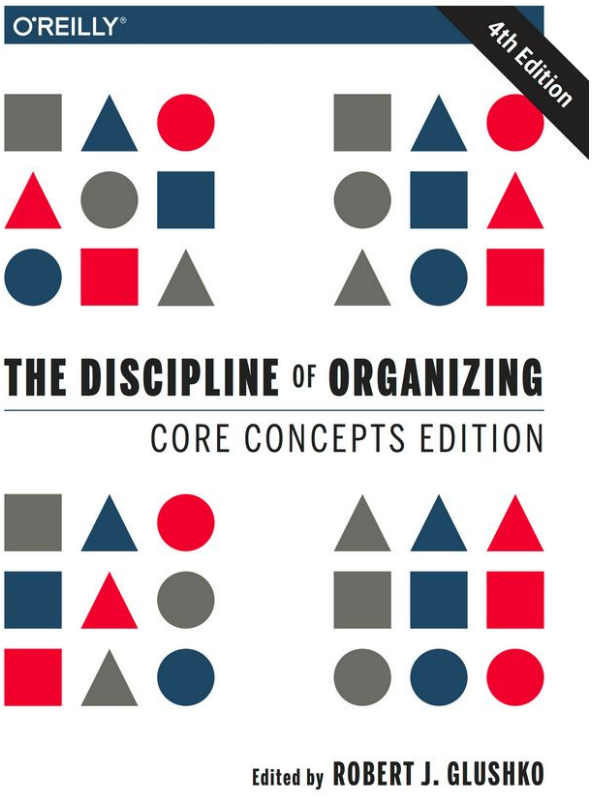


ABRIDGED EDITION

Draft – January 2018

The 4th editions of The Discipline of Organizing were published in August 2016.

The "Core Concepts" edition does not include any of the discipline-specific content contained in the "Professional "or "Informatics" editions. Even so, for many instructors and readers the Core Concepts edition is too encyclopedic in its breadth and depth. Furthermore, TDO has been adopted by schools all over the world, and some of the examples in TDO do not always "travel well."

This Abridged Edition shortens and simplifies the Core Concepts edition to make it more readable and to make it easier to translate.

bob glushko

Chapter 1

Foundations for Organizing Systems

Robert J. Glushko

1.1 The Discipline of Organizing

Organizing is intentionally imposing order and structure on a collection of things or information.

We frequently organize, but we don't always seriously or consciously think about it. We organize shoes in our closet, books on our shelves, spices in our kitchen, receipts and records in tax preparation folders, and the people on business projects or sports teams. If we use the concept of "resources" to include both tangible things and information, we can see that "organizing resources" is an extremely common and general activity.

Additionally, organizing is a fundamental issue in many fields of study, notably: library and information science, computer science, systems analysis, informatics, law, economics, and business.

Furthermore, quite a few of us have jobs that require specific organizing tasks, and we might have been explicitly trained to perform these tasks. Even though we might learn to excel at these tasks, we often do not reflect on the similarity of the organizing tasks we do and those done by others, or the organizing tasks we do at work and those we do at home. We take for granted the concepts and methods underlying the Organizing Systems we work with most often.

Taking these common concepts and methods for granted comes at a cost. The properties of resources chosen 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 a clothes closet but not in a refrigerator. If you live alone, "frequency of use" is an effective organizing principle for your spices, cooking utensils and other resources in your kitchen. However, if you have a roommate, any principle based on individual behavior rather than static resource properties may cause conflicts.

Fields of study differ a lot in how they approach problems of organizing and what they seek as their solutions. For example, library and information science has traditionally analyzed 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 view organizing in the context of information-intensive business applications with a focus on process efficiency, system architecture, and

implementation. Management and industrial organization deal with the organization of human, material, and information resources in contexts shaped by commercial, competitive, and regulatory forces.

The goal of this book is to help readers become more self-conscious about the process and principles of organizing. In particular, it introduces the concept of an Organizing System: an abstract framework for analyzing the issues and problems of organizing that emphasizes design decisions to enable the systematic organization of resources. Every system of organization involves a choice of properties and principles used to describe and arrange resources, and ways of supporting interactions with those resources. By comparing and contrasting how these activities take place in different contexts and domains, we can identify patterns of organizing and see that Organizing Systems often follow a common life cycle.

We can begin by considering four different types or contexts of organizing:

- We organize tangible things.
- We organize information about tangible things.
- We organize digital things.
- We organize information about digital things.

The first contrast is between "organizing things" and "organizing information." We often arrange tangible things according to intrinsic and physical properties such as shape, size, or material. For example, we might sort shirts in a closet by color and style. The same is true of information represented in tangible form. For most of human history, information has been recorded on tangible media such as clay tablets, printed books, vinyl albums, or music CDs. These items have physical and measurable properties such as size, weight, and shape that influence how we organize them.

However, we more often organize our "information things" according to what they are about, rather than by their physical properties. At home, we may sort our CDs by artist or genre; we may keep cookbooks separate from travel books, and fiction books apart from reference books. Likewise, we pay little attention to the visible properties of tangible things when we arrange them according to functional or task properties; tools such as hammers, drills, and screwdrivers may not look alike, but we store them together because they are often used together.

The second contrast is between "organizing things" and "organizing information about things." This difference is clear when considering the traditional library card catalog, whose printed cards describe the books on library shelves. When the things and the information about them are both in a tangible format, it is easy to see that the former is a primary resource and the latter is a surrogate or associated resource that describes or relates to it. When it comes to "organizing information about digital things," the contrast is much less clear. When we find a web page through a hyperlink in a page of search results or access an electronic book through an online

catalog record, it can be harder to distinguish the "information" from the "thing."

What do we mean by “information?” Many definitions use equally hard-to-define words like "data," "knowledge," and "communication." While we use the word in casual conversation (e.g. “Can you give me some information about the train schedule?”), we also talk about information in scholarly or legal debates about the nature of knowledge and evidence. Additionally, some people distinguish between “data,” “information,” “knowledge,” and “wisdom” based on the level of structure, processing, or value of a set of resources.

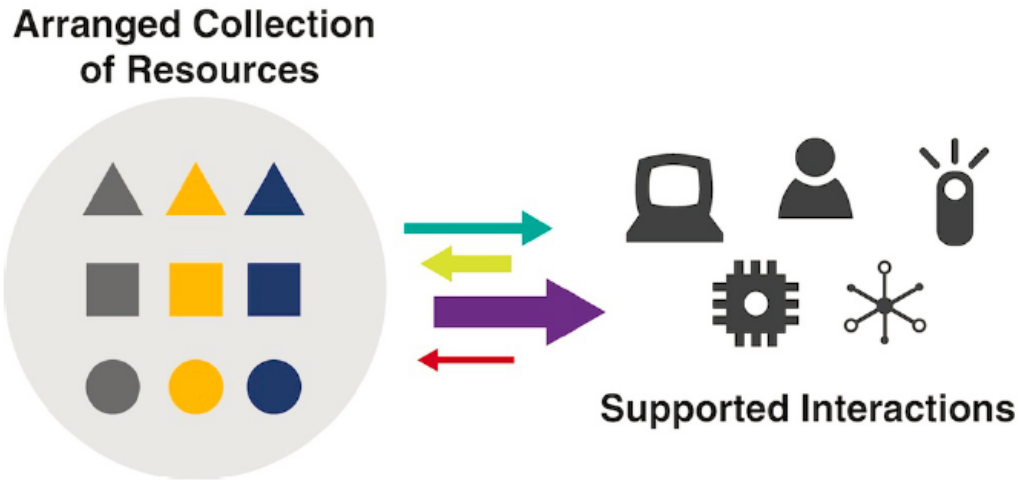
However, for professionals in information-intensive industries, precise distinctions between these concepts have little practical value. Rather than making granular distinctions, we prefer the abstract and encompassing concept of “resources” over “information” and other terms. In our context, “resource” can include everything and anything from books in a library to animals in a zoo or sales figures in a database. Calling everything that is organized a resource helps to bridge the intellectual gulf that separates the many disciplines that share the goal of organizing but differ in what they organize.

This transdisciplinary view lets us emphasize what the different fields that organize resources have in common and how they fit together rather than what distinguishes them. Resource selection, organizing, interaction design, and maintenance are taught in every discipline, but these concepts go by different names. A vocabulary for discussing common organizing challenges and issues that might otherwise be obscured by narrow disciplinary perspectives helps us understand existing systems of organizing better while also suggesting how to invent new ones by making different design choices.

1.2 The “Organizing System” Concept

We propose to unify many perspectives about organizing and information with the concept of an Organizing System: an intentionally arranged collection of resources and the interactions they support. This definition brings together several essential ideas that we will briefly introduce in this chapter and then develop in detail in subsequent chapters.

Figure 1.1 depicts a conceptual model of an Organizing System that shows intentionally arranged resources, interactions (distinguished by different types of arrows), and the human and computational agents interacting with the resources in different contexts.



An Organizing System is a collection of resources arranged in ways that enable people or computational agents to interact with them.

An Organizing System is an abstract view of how the resources in a collection are described and arranged to enable human or computational agents to interact with them. The Organizing System is an architectural and conceptual perspective that is distinct from the physical arrangement of resources that might embody it and from the person, enterprise, or institution that implements and operates it. These distinctions are sometimes hard to maintain in ordinary language; for example, we might describe some set of resource descriptions, organizing principles, and supported interactions as a "library" Organizing System. However, we also need at times to refer to a "library" as the institution in which this Organizing System operates, and of course, the idea of a “library” as a physical building is deeply engrained in language and culture.

The concept of an Organizing System reflects the inherent connections between the activities of organizing resources and making use of their organization which have traditionally been treated as separate by the fields of information organization and information retrieval. A systems view of information organization and information retrieval provides structure for the tradeoffs inherent between these activities. The more effort put into organizing information “on the way in,” when it is created or added to a collection, the more efficiently it can be retrieved. The more effort willing to be put into retrieving information “on the way out,” the less it needs to be organized first.

Sometimes a collection of resources is highly organized, but because it was organized by someone else for different purposes than we have in mind, we need to reorganize it “on the way out.” This is especially common with digital text or datasets, where previously organized resources or their descriptions can be sorted, translated in format or language, combined, summarized, or otherwise transformed to fit into a new Organizing System.

Finally, a systems view can enable more nuanced consideration of the biases inherent in any Organizing System. Whether accidental or intentional, these biases implicitly or explicitly create winners or losers, treat some interactions as preferred while deprecating others, or otherwise impose or overlay a set of values on the stakeholders of the system. By consciously considering the causes and effects of these biases, it is possible to mitigate unintended outcomes in the Organizing System.

1.3 The concept of "Resource"

A resource is anything of value that can support goal-oriented activity. This definition means that a resource can be anything you want to organize, whether a tangible thing, an intangible thing, information about tangible things, or information about intangible things. Other words that aim for this broad scope are entity, object, item, and instance. That a resource can be anything is an important generalization of the concept because it enables web pages, web-based services, data feeds, objects with RFID tags, sensors, other smart devices, or computational agents to be part of Organizing Systems.

A resource can have any number of description resources associated with it to facilitate finding, interacting with, or interpreting it. In this association, the resource being described is considered to be the "primary resource." Using description resources makes it easier to organize tangible resources, which can only be in once place at a time. For example, when books or documents are described by author, subject, title, or creation date, sets of these descriptions serve as indexes or catalogs that point to the primary resource. In the digital age, the results in web search engines are familiar examples of description resources that link to the primary resources.
Resources that describe, or are associated with other resources are sometimes called *metadata*, which means data about data. However, we prefer the more general term of resource description for several reasons, the most obvious one being that the resource being described can be a tangible thing or person, not data. Also, when descriptions are embedded in resources, as in the title page of a book, the masthead of a newspaper, or the source of web pages, separating the resource content into primary and descriptive parts is arbitrary.

More generally, what serves as resource description for one person or process can function as the data or primary resource for another one. Rather than being an inherent distinction, the difference between primary and associated resources is often just a decision about which resource we are focusing on in a particular situation. An animal specimen in a natural history museum might be a primary resource for museum visitors and scientists interested in anatomy, but information about where the specimen was collected is the primary resource for scientists interested in ecology or migration.

Organizing Systems can also refer to people as resources, and we often use that term to avoid specifying the specific role of an employee or worker, as in the management concept of the "human resources" department. Human resources in Organizing Systems can be understood much the same way as inanimate physical or digital resources: they are selected, organized, and managed, and can create value individually or through their interactions with others inside and outside of the system. However, human beings are uniquely complicated resources, and any Organizing System that uses them must take into account their rights, motivations, and relationships.

1.4 The concept of "Collection"

A collection is a group of resources that have been selected for some purpose.

Any set of resources, whether personal or institutional, tangible or intangible, can be called a collection. A collection can contain identifiers for resources along with, or instead of, the resources themselves. This distinction between a resource and its identifier enables a resource to be part of more than one collection, like songs in playlists.

The concept of a collection has deep roots in libraries, museums, and other institutions that select, assemble, arrange, and maintain resources. Organizing Systems in these domains can often be described as collections of collections that are usually organized according to resource type, author, creator, or collector of the resources in the collection. Businesses can also be thought of as collections of employees, products, customers, and the tangible and intangible assets used to create the products and run the business.

Collection is preferred over similar terms such as dataset (science and business) and corpus (linguistics and literary analysis) because it has fewer specialized meanings which makes it better fit with the abstract and architectural approach taken in this book. Using a familiar category name like library, museum, or data repository would reinforce the typical instances and characteristics of that category and marginalize those that are atypical. By avoiding domain-specific vocabulary, it becomes easier to avoid these pitfalls and apply concepts, methods, and insights across disciplines.

1.5 The Concept of "Intentional Arrangement"

Intentional arrangement emphasizes explicit or implicit acts of organization by people, or by computational processes acting as implementations of human intentionality.

Intentional arrangement is easiest to see in Organizing Systems created by individual people who can make all the necessary decisions about organizing their own resources. It is also easy to see in Organizing Systems created by institutions

like libraries, museums, businesses, and governments where the responsibility and authority to organize is centralized and explicit in policies, laws, or regulations.

However, top-down intentionality is not necessary to create an Organizing System, and organization can emerge from the bottom-up through collective behavior without central control. Decisions made by actors intentionally interacting with resources and each other create traces, records, or other information that accumulates over time. The Organizing Systems that emerge from this accumulated information are sometimes called self-organizing and often change their internal structure or their function in response to feedback or changed circumstances.

A good example of emergent, or bottom-up, organization involves path systems where people, as well as ants and other animals, follow and thereby reinforce the paths taken by their predecessors. The resulting organization comes from the indirect communication that occurs when agents modify their environment and others respond to those modifications such as when paths are created. Likewise, even though there is no top-down organization, the web as a whole, with its more than a trillion unique pages, is a self-organizing system that, at its core, follows clear organizing principles.

While many naturally occurring patterns and arrangements, such as rock strata or the movements of stars and planets, contain a great deal of information, these are not Organizing Systems since they are not intentionally arranged by intelligent agents according to Organizing Principles. However, when people create descriptions or models using this information—for instance, defining geological periods based on strata or creating constellations based on the stars—these are Organizing Systems for their observations and measurements.

Taken together, the intentional arrangements of resources in an Organizing System are the result of decisions about what is organized, why it is organized, how much it is organized, when it is organized, and how or by whom it is organized (each of these will be discussed in greater detail in Chapter 2). An Organizing System is defined by the composite impact of the choices made on these design dimensions. Because these questions are interrelated their answers come together in an integrated way to define an Organizing System.

1.6 The Concept of "Organizing Principle"

Organizing Principles are directives for the design or arrangement of a collection of resources that are expressed in a way that does not assume any particular implementation. In any particular Organizing System, abstract organizing principles are used and made concrete just as algorithms are selected and then coded in a specific programming language like Java or Python.

Organizing Systems for small collections can function with very simple organizing

principles and a small amount of intentional arrangement. In a small collection, the minimal or default organizing principle of *colocation* – putting everything in the same container, on a single shelf, or in the same e-mail inbox – may be sufficient. In a kitchen with few spices, you do not need to alphabetize them because it is easy to find the one you want. Over time, resources may be implicitly organized by the *frequency of use* principle: the most commonly used spices in a drawer often migrate to the front because that is the easiest place to return them after use. However, as a collection grows in size, it becomes increasingly inefficient to arrange, locate, and retrieve specific resources this way, and more complex organizing principles and a greater degree of intentional arrangement become necessary.

For this reason, most Organizing Systems employ organizing principles that make use of properties of the resources being organized, such as their name, color, shape, date of creation, or semantic category. These properties are often used simultaneously. For example, in your kitchen, you might arrange your cooking pots and pans by size and shape so you can nest them and store them compactly, but you might also arrange things by cuisine and separate grilling equipment from the wok and other items for making Chinese food.

For organizing information resources, the most useful properties are those that reflect their content and meaning. These types of properties are not directly apparent when you look at a book, document, or collection of data, and significant intellectual effort or statistical computation is necessary to reveal them. Additionally, they are often based on statistical properties that emerge from analyzing the collection as a whole rather than individual resources. For example, the relevance of documents to a search query is higher when they contain a higher than average frequency of the query terms compared to other documents in the collection, or when they are linked to relevant documents. Likewise, algorithms for classifying email messages continuously recalculate the probability that words like "beneficiary" or "congratulations" indicate whether a message is "spam" or "not spam" in the collection of messages processed.

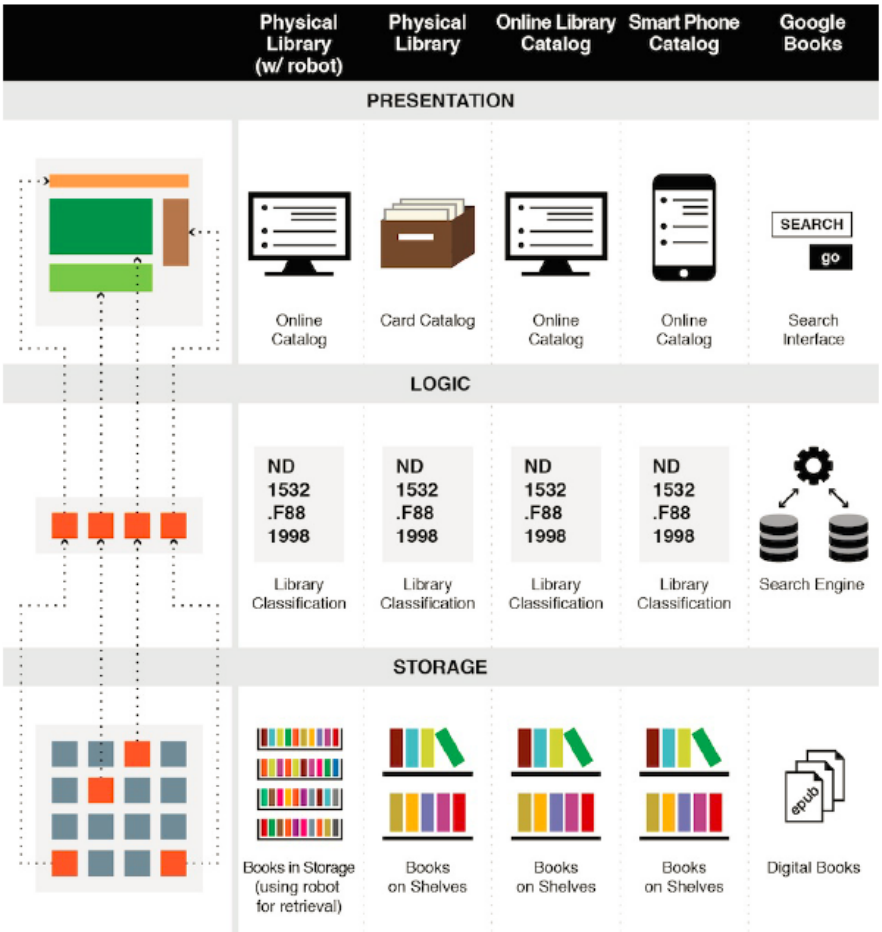
1.6.1. Organizing Principles and Architectural Thinking

In designing Organizing Systems and choosing appropriate Organizing Principles, we should be guided by Architectural Thinking. This design philosophy suggests that organizing principles should not assume any particular implementation, environment, or technology. Designers should think abstractly about the desired resource interactions and not constrain their design with implementation concerns like the physical environment, storage capacity, or programming language. Choosing technologies and implementations that support the interactions comes later.

Architectural Thinking creates three layers or tiers in an Organizing System: presentation, logic, and storage. Layers are easy to see with digital resources. Software developers commonly divide applications into tiers of user interface

(implementation of interactions), business logic (intentional arrangement), and data (resources). This separation allows each of the three tiers to be upgraded or re-implemented independently to satisfy changing requirements or to take advantage of new technologies. When a library shifts from a paper card catalog to an online catalog, users enjoy quicker and more convenient searches through their browser or smartphone, but the books do not have to be reclassified or moved to new shelf locations. Likewise, when an organization moves its email system from a local host to a cloud service, users might not notice or care – they can send, receive, filter, and sort messages just as before.

Figure 1.2. Presentation, Logic and Storage Tiers.



It is highly desirable when the design and implementation of an Organizing System separates the storage of the resources from the logic of their arrangement and the methods for interacting with them. This three-tier architect is familiar to designers of computerized Organizing Systems but it is also useful to think about Organizing Systems in this way even when it involves physical resources.

The separation of Organizing Principles and their implementation is harder to achieve and recognize in an Organizing System that only contains physical resources – it might seem that you can have unmediated interactions with resources rather than accessing them through some user interface or "presentation tier" that

supports the principles specified in the “middle tier” and realized in the “storage tier.” However, the three tiers can be seen clearly in a library using an old-fashioned card catalog (see figure 1.2), where users interact with a physical presentation tier (drawers full of index cards) using a complex logic (a library classification system that organizes books by their information content) to access physical resources (books stored on shelves).

Library organizing systems are highly evolved and complicated, but we can see the importance of architectural thinking even in the much simpler environment of spices in a kitchen. Different kitchens might all use an alphabetic order organizing principle for arranging a collection of spices, but the exact locations and arrangement of the spices in any particular kitchen depend on the configuration of shelves and drawers, whether a spice rack or rotating tray is used, and other storage-tier considerations. Similarly, spices could be logically organized by cuisine, with Indian spices separated from Mexican spices, but this organizing principle does not imply anything about where they can be found in the kitchen. Many common home-organizing tricks focus on the presentation tier (such as putting color-coded labels on file folders) or the storage tier (buying stackable boxes or storage containers), but these efforts will not help if there is no logical Organizing Principle to guide them.

1.7 The Concept of “Agent”

Many disciplines have specialized job titles to distinguish among the people who organize resources (for example cataloger, archivist, indexer, curator, collections manager). We use the more general word, *agent*, for any entity capable of autonomous and intentional organizing effort. This abstract view allows us to recognize that trained people working alone are not the only ones who perform organizing work. When users tag images on a photo-sharing website or use hashtags to arrange messages by topics, they are intentionally arranging resources in an Organizing System. Additionally, groups of people, such as companies or standards bodies, may work together to organize resources for efficiency, safety, and economic benefit.

Computational processes can also be agents. For example, search engines use algorithms to locate and describe web pages and rank them according to relevance for a particular user and query. Even resources can be agents, with the ability to initiate interactions and create value. Beyond the obvious case of human resources being agents, examples of non-human resources that are agents are self-driving cars and "smart" thermostats that can use sensors, physical mechanisms, and network connections to gain awareness of their environment, take useful actions, and communicate with one another.

The capabilities of the agents in an Organizing system inform its goals, capabilities, and costs. While professional human agents such as indexers, catalogers, and taxonomists can use established classifications and controlled vocabularies to

organize resources, this is very costly and time intensive. Instead, it might be better to organize resources algorithmically and sacrifice some precision and predictability to save time and money. Additionally, the terms and structures used by professionals may be too cumbersome for users, who might prefer the flexibility to organize resources based on their own experience and work practices. Understanding these dimensions of an Organizing System allows us to determine how best to combine the efforts of human and computational agents, and how it allocates effort and costs between its creators, users, maintainers, and other stakeholders.

1.8 The Concept of "Interaction"

An *interaction* is an action, function, service, or capability that makes use of the resources in a collection or the collection as a whole.

Just as with organizing principles, it is useful to think of interactions in an abstract way that does not assume an implementation. The interaction of *access* is fundamental in any collection of resources, but many Organizing Systems provide additional functions to make access more efficient and to support additional interactions with the accessed resources. For example, libraries and similar Organizing Systems implement catalogs to enable interactions for *finding* a known resource, *identifying* a resource in the collection, and discriminating or *selecting* among similar resources.

Some of the interactions with resources in an Organizing System are determined by the characteristics of the resource. Because many museum resources are unique or extremely valuable, visitors are allowed to view them but cannot touch or borrow them. In contrast, most library resources are neither unique nor extremely valuable, and a library might have multiple printed copies of a popular book and lend them all. After a printed book is checked out from the library, there are many types of interactions that might take place—reading, translating, summarizing, annotating, and so on—but these are not directly supported by the library Organizing System and are invisible to it.

Digital resources can enable a greater range of interactions than physical ones: any number of people or processes can make a request to an application program interface (API) or download a file because the resource is not used up and the marginal cost of allowing another access is nearly zero. Furthermore, with digital resources, many new kinds of interactions can be enabled through application software, web services, or APIs in the Organizing System. These can include translation, summarization, annotation, keyword suggestion, and update services – all interactions that are possible with physical resources, but at a much higher cost and not supported by the Organizing System itself.

While enabling interactions is critical, there are also valid reasons to prevent or

constrain them. For example, owners of copyrighted works that want to prohibit copying or reuse of their content can use digital rights management (DRM) technology to prevent copying, limit access to particular devices, or revoke access. However, these technologies typically do not account for legal exceptions, such as “fair use,” to the rights they enforce, and such exceptions may be blocked entirely.

1.9 The Concept of “Interaction Resource”

Interactions with physical resources sometimes leave traces or other evidence of the interaction. Many of these traces are unintentional, like fingerprints or the erosion on a shortcut path across a lawn, while others are intentional, like a student's notes in a textbook or spray-painted graffiti on a building. While clever forensic investigators like Sherlock Holmes can use interaction traces to identify or vindicate suspects, not every interaction leaves a trace, traces fade over time, and different traces associated with the same resource lack consistency. Given this, most physical interaction traces are not of much use.

However, when Organizing Systems contain digital resources or physical resources that have sensing, recording, or communication capabilities, interaction traces can be made predictable, persistent, and consistent. A user accessing, browsing, buying, highlighting, linking, or otherwise interacting with such an Organizing System creates a record which becomes an "interaction resource." These interaction resources can then be analyzed to reorganize the resource collection or otherwise influence subsequent interactions with the primary resources.

Interaction resources are often essential to the functioning Organizing Systems. For example, smart "toll tags" broadcast their identity when the car they are in passes a radio receiver at a toll location. The interaction resource this creates is then used to identify the account and credit card with which to pay the toll. Taken together, the collection of these interaction resources can then be used as the primary resources in other Organizing Systems such as those that manage traffic congestion or support road design. Similarly, interaction resources created by search engines can be used to adjust the order of search hits, select ads, or personalize the content of web pages.