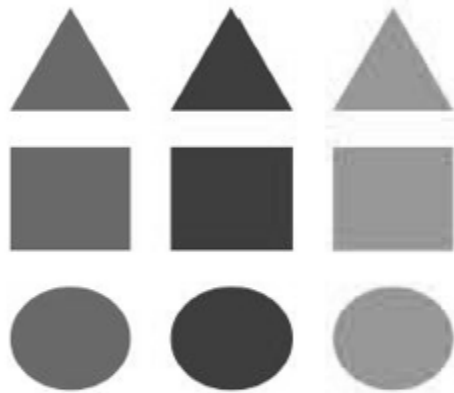


THE DISCIPLINE OF **ORGANIZING**

PROFESSIONAL EDITION



Edited by **ROBERT J. GLUSHKO**

This version of the 4th Edition (2016) is available under Creative Commons license CC BY-NC -
<https://creativecommons.org/licenses/by-nc/4.0/>

Chapter 1
**Foundations for Organizing
Systems**

Robert J. Glushko

1.1. The Discipline of Organizing 25
1.2. The “Organizing System” Concept 34
1.3. The Concept of “Resource” 36
1.4. The Concept of “Collection” 38
1.5. The Concept of “Intentional Arrangement” 41
1.6. The Concept of “Organizing Principle” 44
1.7. The Concept of “Agent” 50
1.8. The Concept of “Interactions” 51
1.9. The Concept of “Interaction Resource” 52
1.10. Organizing This Book 53

1.1 The Discipline of Organizing

To *organize* is to create capabilities by intentionally imposing order and structure.

Organizing is such a common *activity* that we often do it without thinking much about it. We organize shoes in our closet, books on our book shelves, spices in our kitchen, receipts and records in tax preparation folders, and people on business projects and sports teams. 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.

The goal of this book is to help readers become more self-conscious about what it means to organize resources of any type and about the principles by which the resources are organized. In particular, this book introduces the concept of an *Organizing System*: an intentionally arranged collection of resources and the interactions they support. The book analyzes the design decisions that go into any systematic organization of resources and the design patterns for the interactions that make use of the resources, as follows:

We organize physical things. Each of us organizes many kinds of things in our lives—our books on bookshelves; printed financial records in folders and filing cabinets; clothes in dressers and closets; cooking and eating utensils in kitchen drawers and cabinets. Public libraries organize printed books, periodicals, maps, CDs, DVDs, and maybe some old record albums. Research libraries also organize rare manuscripts, pamphlets, musical scores, and many other kinds of printed information. Museums organize paintings, sculptures, and other artifacts of cultural, historical, or scientific value. Stores and suppliers organize their goods for sale to consumers and to each other. Sports leagues organize players into teams, and the teams organize players by position or role.

We organize information about physical things. Each of us organizes information about things: when we inventory the contents of our house for insurance purposes, when we sell our unwanted stuff on eBay, or when we rate a restaurant on Yelp. Library card catalogs, and their online replacements, tell us what books a library's collection contains and where to find them. Sensors and RFID tags track the movement of goods—even library books—through supply chains, and the movement (or lack of movement) of cars on highways.

We organize digital things. Each of us organizes personal digital information—email, documents, ebooks, MP3 and video files, appointments, contacts—on our computers, smartphone, ebook readers, or in “the cloud,” —through information services that use Internet protocols. Large research libraries organize digital journals and books, computer programs, government and scientific datasets, databases, and many other kinds of digital information. Companies organize their digital business records and customer information in enterprise applications, content repositories, and databases. Hospitals and medical clinics maintain and exchange electronic health records and digital X-rays and scans.

We organize information about digital things. Digital library catalogs, web portals, and aggregation websites organize links to other digital resources. Web search engines use content and link analysis along with relevance ratings, to organize the billions of web pages competing for our attention. Web-based services, data feeds and other information resources can be interconnected and choreographed to carry out information-intensive business processes, or aggregated and analyzed to enable prediction and personalization of information services.

Let us take a closer look at these four different types or contexts of organizing. We contrasted “organizing things” with “organizing information.” At first glance it might seem that organizing physical things like books, compact discs, machine parts, or cooking utensils has an entirely different character than organizing intangible digital things. We often arrange physical things according to their shapes, sizes, material of manufacture, or other intrinsic and visible properties: for example, we might arrange our shirts in the clothes closet by style and color, and we might organize our music collection by separating the old vinyl albums from the CDs. We might arrange books on bookshelves by their sizes, putting all the big, heavy picture books on the bottom shelf. Organization for clothes and information artifacts in tangible formats that is based on visible properties does not seem much like how you store and organize digital books on your Kindle or arrange digital music on your music player. Arranging, storing, and accessing X-rays printed on film might appear to have little in common with these activities when the X-rays are in digital form.

It is hardly surprising that organizing things and organizing information sometimes do not differ much when information is represented in a tangible way. The era of ubiquitous digital information of the last decade or two is just a blip in time compared with the more than ten thousand years of human experience with information carved in stone, etched in clay, or printed with ink on papyrus, parchment, or paper. These tangible information artifacts have deeply embedded the notion of information as a physical thing in culture, language, and methods of information design and organization. This perspective toward tangible information artifacts is especially prominent in rare book collections where books are revered as physical objects with a focus on their distinctive binding, calligraphy, and typesetting.

Nevertheless, at other times there are substantial differences in how we organize things and how we organize information, even when the latter is in physical form. We more often organize our “information things” according to what they are about rather than on the basis of their visible properties. At home we sort our CDs by artist or genre; we keep cookbooks separate from travel books, and fiction books apart from reference books. Libraries employ subject-based classification schemes that have a few hundred thousand distinct categories.

Likewise, there are times when we pay little attention to the visible properties of tangible things when we organize them and instead arrange them according to functional or task properties. We keep screwdrivers, pliers, a hammer, a saw, a drill, and a level in a toolbox or together on a workbench, even though they have few visual properties in common. We are not organizing them because of what we see about them, but because of what we know about to use them. The task-based organization of the tools has some similarity to the subject-based organization of the library.

We also contrasted “organizing things” with “organizing information about things.” This difference seems clear if we consider 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 physical format, it is easy to see that the former is a primary resource and the latter a surrogate or associated resource that describes or relates to it.

What Is Information?

Most of the hundreds of definitions of information treat it as an idea that swirls around equally hard-to-define terms like “data,” “knowledge,” and “communication.” Moreover, these intellectual and ideological perspectives on information coexist with more mundane uses of the term, as when we ask a station agent: “Can you give me some information about the train schedule?”

An abstract view of information as an intangible thing is the intellectual foundation for both modern information science and the information economy and society. Nevertheless, the abstract view of information often conflicts with the much older idea that information is a tangible thing that naturally arose when information was inextricably encoded in material formats. We often blur the sense of “information as content” with the sense of “information as container,” and we too easily treat the number of stored bits on a computer or in “the cloud” as a measure of information content or value.

Geoff Nunberg has eloquently explained in *Farewell to the Information Age* that information is “a collection of notions, rather than a single coherent concept.” Michael Buckland’s oft-cited essay *Information as Thing* argues against the notion that information is inherently intangible and instead defines it more broadly and provocatively based on function. A resource that can be learned from or serve as evidence is “information-as-thing,” a definition that treats the tangible objects in museum or personal collections as information.¹[\[LIS\]](#)

When it comes to “organizing information about digital things” the contrast is much less clear. When you search for a book using a search engine, first you get the catalog description of the book, and often the book itself is just a click away. When the things and the information about them are both digital, the contrast we posed is not as sharp as when one or both of them is in a physical format. And while we used X-rays—on film or in digital format—as examples of things we might organize, when a physician studies an X-ray, is it not being used as information about the subject of the X-ray, namely, the patient? And when businesspeople make marketing and pricing decisions by analyzing digital information about what and when people buy, we can think of this as organizing customers into categories, or as organizing customer information.

These differences and relationships between “physical things” and “digital things” have long been discussed and debated by philosophers, linguists, psychologists, and others. (See the sidebars, [What Is Information?](#) (page 28) and [The Distinction between Data and Information](#) (page 30).)

The distinctions among organizing physical things, organizing digital things, or organizing information about physical or digital things are challenging to describe because many of the words we might use are as overloaded with multiple meanings as “information” itself. For example, the library science perspective often uses presentation or implementation properties in definitions of “document,” using the term to refer only to traditional physical forms. In contrast, the informatics or computer science perspective takes an abstract view of “document” to refer to any self-contained unit of information, separating a document's content from its presentation or container.^{2[Com]}

The most abstract definition of “document,” presented in [What is a Document?](#) follows from Buckland's “information as thing” idea. Because it can be studied to provide evidence, an antelope is both “information as thing” and also a “document” when it is in a zoo, even though it is just an animal when it is running wild on the plains of Africa. However, in 2015 the United States Supreme Court rejected this expansive definition in a case that hinged on whether a fish could be viewed as a document.^{3[Law]}

Similar definitional variation occurs with “author” or “creator.” When we say that “Herman Melville is the author of *Moby Dick*” (Melville 1851) the meaning of “author” does not depend on whether we have a printed copy or an ebook in mind, but what counts as authorship varies a great deal across academic disciplines. Furthermore, different standards for describing resources disagree in the precision with which they identify the person(s) or organization(s) primarily responsible for creating the intellectual content of the resource. People who are serious about music description rightly criticize streaming services and online stores that have only a single “artist” field because this fails to distinguish the composer, conductor, orchestra, and other people with distinct roles in creating the music.

If we allow the concept of information to be anything we can study—to be “anything that informs”—the concept becomes unbounded. Our goal in this book is to bridge the intellectual gulf that separates the many disciplines that share the goal of organizing but differ in what they organize. This requires us to focus on situations where information exists because of intentional acts to create or organize. (See the sidebar, [The Discipline of Organizing](#) (page 31))

Many of the foundational topics for a discipline of organizing have traditionally been presented from the perspective of the library sector and taught as “library and information science.” These include bibliographic description, classification, naming, authority control, curation, and information standards. In recent

The Distinction between Data and Information

Astute readers might have noticed that we included sensor data as “information about physical things” and data feeds as “information about digital things.” Many textbooks in the information science and knowledge management fields distinguish data and information in a more precise way. To them, data sits at the bottom of an Information Hierarchy, Knowledge Pyramid, or DIKW Hierarchy in which Data is transformed into Information, which is transformed into Knowledge, which is then transformed into Wisdom.

In this framework, data are raw or elementary observations about properties of objects, events, and their environment. Data becomes information when it is aggregated, processed, analyzed, formatted, and organized to add meaning and context so it can be used to answer questions. This processing can include calculation, inference, or refinement operations on the data. For example, measurements of temperature, precipitation, and wind speed are data. When combined and summarized, a set of data becomes statistical information about the weather on a particular day. When collected over a period of months or years, these datasets become information about the climate of the location where they were collected.

The Discipline of Organizing does not make this sharp contrast between data and information in the Hierarchy/Pyramid. People who read this book are likely to be aspiring or practicing professionals in information-intensive industries where information and data are often treated as synonyms to mean the content of a database or data-managing application. A distinction between data and information might be useful in theory, but not in these applied settings.

The distinction between data and information is also being blurred by the expansion in the scope of the definition of data in the emerging career field of *data science*. Indeed, a popular introductory text eliminates information entirely from the Hierarchy/Pyramid with its title, *Discovering knowledge in data: an introduction to data mining*.^{4[DS]}

decades these foundations have been built on and extended by computer science, cognitive science, informatics, and other new fields to include more private sector and non-bibliographic contexts, multimedia and social media, and new information-intensive applications and service systems enabled by mobile, pervasive, and scientific computing. The latest additions to the discipline of organizing are coming from *data science* and *machine learning*, introducing considerations of speed and scale that arise when massive computational power and new statistical techniques are harnessed to organize and act on information.

The Discipline of Organizing

A *discipline* is an integrated field of study in which there is some level of agreement about the issues and problems that deserve study, how they are interrelated, how they should be studied, and how findings or theories about the issues and problems should be evaluated. A *framework* is a set of concepts that provide the basic structure for understanding a domain, enabling a common vocabulary for different explanatory theories.

Organizing is a fundamental issue in many disciplines, most notably library and information science, computer science, systems analysis, informatics, law, economics, and business. However, these disciplines have only limited agreement in how they approach problems of organizing and 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.^{5[LIS]} In contrast, computer science and informatics tend to study organizing in the context of information-intensive business applications with a focus on process efficiency, system architecture, and implementation. The disciplines of management and industrial organization deal with the organization of human, material, and information resources in contexts shaped by commercial, competitive, and regulatory forces.

This book presents a more abstract framework for issues and problems of organizing that emphasizes the common concepts and goals of the disciplines that study them. Our framework proposes that every system of organization involves a collection of resources, and we can treat physical things, digital things, and information about such things as resources. Every system of organization involves a choice of properties or principles used to describe and arrange the resources, and ways of supporting interactions with the 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 create a discipline of organizing in a disciplined way.

The new methods and tools of *data science* and *machine learning* let us organize more information, to do it faster, and to make predictions based on what people have clicked on, bought, or said. But this is not the first time that new ideas and technologies have challenged how people organized and interacted with resources. Fifty years ago, searchable online catalogs radically changed how people used libraries. The web, invented less than thirty years ago so that scientists could share technical reports, is now an essential part of many human activities. It is important not to view the latest new thing as changing every-

thing, because new things will continue to come, and these technology breakthroughs still depend on and complement the organizing work done by people. Data science will not replace human organizers, any more than any other science has replaced humans. (See sidebar, **Data Science and the Discipline of Organizing** (page 33)).

This is why we need to take a transdisciplinary view that lets us emphasize what the different disciplines 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 be otherwise 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.

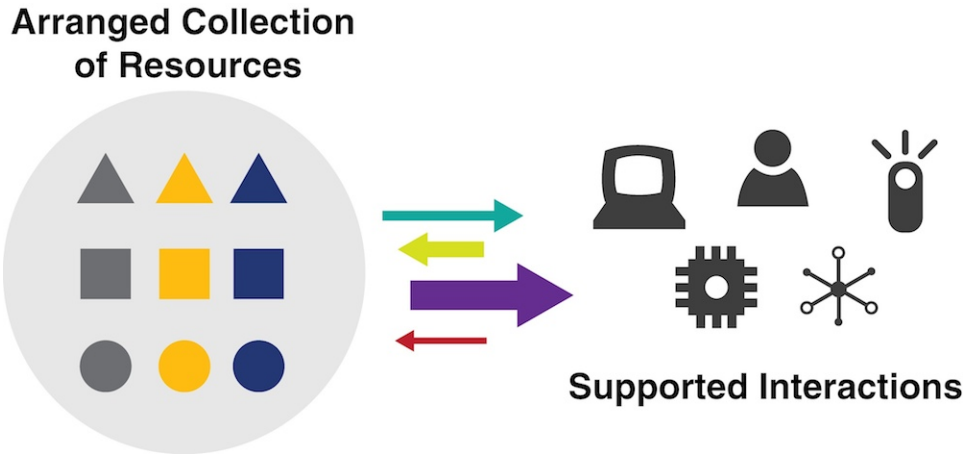
Data Science and the Discipline of Organizing

Advances in computing power and statistical techniques are making it possible to identify patterns in data and extract meaningful information at a scale never before possible. Many books and articles about *data science*, machine learning, and predictive analytics make bold predictions that these emerging fields will radically change the world. These claims are both provocative and promising, but at its core, data science is about how resources are selected, described, and organized; concepts with a long tradition in information and library science. Instead of organizing and describing the books in a library or the products in a warehouse, a data scientist might organize information about books or products into massive data tables, treating each resource as a row and its descriptive properties as the columns. After people might have organized books or products into categories, machine learning techniques might classify new books or products using those categories, or perhaps discover new categories based on access or purchasing behaviors. So while the techniques of data science are new, many of the challenges are not; data scientists need to select resources wisely and decide how best to describe them; they need to understand that resource description and categorization can be biased; they need to understand the tradeoffs and complements between people and computers; and, they need to test the discoveries that algorithms make with controlled experiments.

To make sense of the discussions around data science, one must understand the difference between kind and degree. A hundred years ago, a car's highway travel speed was about forty miles an hour. Today's cars travel twice as fast, but this is just a change in degree. However, an increase in speed to about 17,500 miles an hour achieves an "orbital velocity" that allows us to go into Earth orbit in space, travel that is different in kind.

What about data science? Some data science involves collections of data that are "tall," containing many millions or even billions of records that each have a relatively small number of variables. Being able to analyze "tall" data more rapidly than ever before is primarily a change in degree compared with traditional database techniques. Nevertheless, for collections of data that are "wide," where each record might contain hundreds or thousands of variables, data science techniques might allow us to see patterns that could not be seen at all, or could not be seen affordably and in quantity. Here, data science might be yielding changes in kind.^{6[DS]}

Figure 1.1. An Organizing System.



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

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 an abstract characterization of how some collection of resources is described and arranged to enable human or computational agents to interact with the resources. The Organizing System is an architectural and conceptual view that is distinct from the physical arrangement of resources that might embody it, and also distinct 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 facility is deeply engrained in language and culture.

Our concept of the Organizing System was in part inspired by the concepts proposed in 2000 for bibliographic domains by Elaine Svenonius, in *The Intellectual Foundation of Information Organization*. She recognized that the traditional *information organization* activities of bibliographic description and cataloging were complemented, and partly compensated for, by automated text processing and indexing that were usually treated as part of a separate discipline of *information retrieval*. Svenonius proposed that decisions about organizing information and decisions about retrieving information were inherently linked by a tradeoff principle and thus needed to be viewed as an interconnected system: “The effectiveness of a system for accessing information is a direct function of the intelligence put into organizing it” (p.ix). We celebrate and build upon her insights by beginning each of the sub-parts of **Chapter 2** with a quote from her book.^{7[LIS]}

A systems view of information organization and information retrieval captures and provides structure for the inherent tradeoffs obscured by the silos of traditional disciplinary and category perspectives: the more effort put into organizing information “on the way in” when it is created or added to a collection, the more effectively it can be retrieved, and the more effort 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 that have in mind, we need to reorganize it “on the way in.” This is especially common with digital text or datasets, where previously organized resources or their descriptions might be sorted, translated in format or language, combined, summarized, or otherwise transformed to fit into a new Organizing System. For example, to understand seasonal buying patterns, a retailer might combine shopping data with weather data and calendar data about commonly-watched sporting events (because bad weather and broadcast sports cause people to stay home), and all three datasets would need to describe “time” and “location” in the same way.

A 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 when computers analyze and organize resources. For example, many algorithms for computational classification use *supervised learning* approaches that start with items classified by people.

Finally, a systems view can be applied to Organizing Systems with any kind of resource, enabling more nuanced discussion of how economic, social, and cogni-

tive costs and benefits of organizing are allocated among different stakeholders. Every Organizing System is biased by the perspectives and experiences of the people who create it. Some of these biases are inescapable, a kind of automatic organizing, because they reflect innate human perceptual and cognitive capabilities. Our minds impose structure and find patterns, even when there aren't any, and we are not capable of acting perfectly rationally, so we simplify without realizing it. People are also not very good at thinking about future possibilities and revising their expectations given new evidence, and this mental inertia makes us preserve resources and interactions in Organizing Systems that are no longer needed. Other biases in Organizing Systems reflect more intentional choices that 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. For example, many Organizing Systems arrange people in groups or queues to make interactions more efficient, but when an airline gives boarding priority to customers who paid more for their tickets it might not seem fair to you if are in the last boarding group.

1.3 The Concept of “Resource”

Resource has an ordinary sense of anything of value that can support goal-oriented activity. This definition means that a resource can be a physical thing, a non-physical thing, information about physical things, information about non-physical things, or anything you want to organize. Other words that aim for this broad scope are *entity*, *object*, *item*, and *instance*. *Document* is often used for an information resource in either digital or physical format; *artifact* refers to resources created by people, and *asset* for resources with economic value.

Resource has specialized meaning in Internet architecture. It is conventional to describe web pages, images, videos, and so on as *resources*, and the protocol for accessing them, *Hypertext Transfer Protocol (HTTP)*, uses the *Uniform Resource Identifier (URI)*.^{8[Web]}

Treating as a *primary resource* anything that can be identified is an important generalization of the concept because it enables web-based services, data feeds, objects with RFID tags, sensors or other *smart devices*, or computational agents to be part of Organizing Systems.

Instead of emphasizing the differences between tangible and intangible resources, we consider it essential to determine whether the tangible resource has information content—whether it needs to be treated as being “about” or representing some other resource rather than being treated as a thing in itself. Whether a book is printed or digital, we focus on its information content, what it is about; its tangible properties become secondary. In contrast, the hangars in our closet and the measuring cups in our kitchen are not about anything more than their obvious utilitarian features, which makes their tangible properties

Concert Tickets



*Tickets are physical artifacts that convey event-related **metadata**: including time, place, and seat number; price and terms of admission; and featured performers. For concert goers, tickets offer the promise of all that, and a memory of the ineffable quality of more.*

(Photo by Murray Maloney.)

Concert Ticket

A concert ticket is a vehicle for conveying a package of assertions about an event, so it is a *description resource*, like a card in a library card catalog. A concert ticket is also a *resource* in its own right, with intrinsic value; it can be bought and sold, sometimes for a greater price than its *resource description* specifies. A ticket is a license to use a seat in a venue for a specified purpose at a specified time; after the event, the ticket loses its intrinsic value, but might acquire extrinsic value as an artifact in a collection like this one.

most important. (Of course, there is no sharp boundary here; you can buy “fashion hangers” that make a style statement, and the old measuring cup could be a family memento because it belonged to Grandma).

Many of the resources in Organizing Systems are *description* resources or *surrogate resources* that describe the *pri-*

many resources; library catalog entries or the list of results in web search engines are familiar examples. In museums, information about the production, discovery, or history of ownership of a resource can be more important than the resource; a few shards of pottery are of little value without these associated information resources. Similarly, business or scientific data often cannot be understood or analyzed without additional information about the manner in which they were collected. Most web-based businesses exploit data about how users

interact with resources, such as the log files that record every web search you make, every link you click, and every web page you visit.

Resources that describe, or are associated with other resources are sometimes called *metadata*. However, when we look more broadly at Organizing Systems, it is often difficult to distinguish between the resource being described and any description of it or associated with it. One challenge is that when descriptions are embedded in resources, as metadata often is—in the title page of a book, the masthead of a newspaper, or the source of web pages—deciding which resources are primary is often arbitrary.

A second challenge is that what serves as metadata for one person or process can function as a primary resource or data 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 some 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 refer to people as resources, and we often use that term to avoid specifying the gender or specific role of an employee or worker, as in the management concept of the “human resources” department in a workplace. A business is defined by its intentional arrangement of human resources, and there is both variety and regularity in these arrangements (see the sidebar, *Business Structures* (page 320) in §6.5).^{9[Phil]}

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.^{10[Bus]} However, human beings are uniquely complicated resources, and any Organizing System that uses them must take into account their rights, motivations, and relationships. (See the sidebar, *People as Resources* (page 186).)

1.4 The Concept of “Collection”

A *collection* is a group of resources that have been selected for some purpose. Similar terms are *set* (mathematics), *aggregation* (data modeling), *dataset* (science and business), and *corpus* (linguistics and literary analysis).

We prefer *collection* because it has fewer specialized meanings. *Collection* is typically used to describe personal sets of physical resources (my stamp or record album collection) as well as digital ones (my collection of digital music). We distinguish law libraries from software libraries, knowledge management systems from data warehouses, and personal stamp collections from coin collec-

tions primarily because they contain different kinds of resources. Similarly, we distinguish document collections by resource type, contrasting narrative document types like novels and biographies with transactional ones like catalogs and invoices, with hybrid forms like textbooks and encyclopedias in between.

A collection can contain identifiers for resources along with or instead of the resources themselves, which enables a resource to be part of more than one collection, like songs in playlists.

A *collection* itself is also a *resource*. Like other resources, a collection can have description resources associated with it. An *index* is a *description resource* that contains information about the locations and frequencies of terms in a document *collection* to enable it to be searched efficiently.

Because *collections* are an important and frequently used kind of *resource*, it is important to distinguish them as a separate concept. In particular, the concept of *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 variously organized according to resource type, author, creator, or collector of the resources in the collection, or any number of other principles or properties. In business contexts, the use of “collection” to describe a set of resources is much less common, but businesses organize many types of resources, including their employees, suppliers, customers, products, and the tangible and intangible assets used to create the products and run the business. Indeed, a business itself can sometimes be abstractly described as a collection of resources, especially when the resources are software components or services. (See endnote⁴⁶ [Com].)

A type of resource and its conventional Organizing System are often the focal point of a discipline. Category labels such as library, museum, zoo, and data repository have core meanings and many associated experiences and practices. Specialized concepts and vocabularies often evolve to describe these. The richness that follows from this complex social and cultural construction makes it difficult to define category boundaries precisely.

Libraries can be defined as institutions that “select, collect, organize, conserve, preserve, and provide access to information on behalf of a community of users.” Many Organizing Systems are described as libraries, although they differ from traditional libraries in important respects. (See the sidebar, *What Is a Library?*)

We can always create new categories by stretching the conventional definitions of “library” or other familiar Organizing Systems and adding modifiers, as when Flickr is described as a web-based photo-sharing library. 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

What Is a Library?

Most birds fly, but not all of them do. What characteristics are most important to us when we classify something as a bird? What characteristics are most important when we think of something as a library?

We might treat *circulation*, borrowing and returning the same item, as one of the interactions with resources that defines a library. In that case, an institution that lends items in its collection with the hope that the borrowers return something else that is better hardly seems like a library. But if the resources are the seeds of heirloom plants and the borrowers are expected to return seeds from the plants they grew from the borrowed seeds, perhaps “seed library” is an apt name for this novel Organizing System. Similarly, even though the resources in its collection are encyclopedia articles rather than living species, the Wikipedia open-source encyclopedia resembles the Seed Library by encouraging its users to “return” articles that are improvements of the current ones.

The photo-sharing website Flickr functions for most of its users as a personal photo archiving site. Flickr’s billions of user-uploaded photos and the choice of many users to share them publicly transform it into a searchable shared collection, and many people also think of Flickr as a photo library. But Flickr lacks the authoritative description and standard classification that typify a library.

A similar categorization challenge arises with the Google Books digitization project. ¹¹[Law]

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).

More generally, a categorical view of Organizing Systems makes it matter greatly which category is used to anchor definitions or comparisons. The Google Books project makes out-of-print and scholarly works vastly more accessible, but when Google co-founder Sergei Brin described it as “a library to last forever” it upset many people with a more traditional sense of what the library category implies. 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.

1.5 The Concept of “Intentional Arrangement”

Intentional arrangement emphasizes explicit or implicit acts of organization by people, or by computational processes acting as proxies for, or 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 always necessary to create an Organizing system. Organization can emerge over time via collective behavior in situations without central control when decisions made by individuals, each acting intentionally, create traces, records, or other information that accumulates over time. Organizing systems that use bottom-up rather than top-down mechanisms are sometimes called self-organizing, because they emerge from the aggregated interactions of actors with resources or with each other. *Self-organizing systems* can change their internal structure or their function in response to feedback or changed circumstances.

This definition is broad enough to include business and biological ecosystems, traffic patterns, and open-source software projects. Another good example of emergent organization involves path systems, where people (as well as ants and other animals) can follow and thereby reinforce the paths taken by their predecessors. When highly orderly and optimal arrangements emerge from local interactions among ants, bees, birds, fish, and other animal species, it is often called “swarm intelligence.” When this happens with human ratings for news stories, YouTube videos, restaurants, and other types of digital and physical resources we call it “crowdsourcing.” What the animal and human situations have in common is that information is being communicated between individuals. Sometimes this communication is direct, as when Amazon shows you the average rating for a book or what books have been bought by people like you. At other times the communication is indirect, achieved when the agents modify their environment (as they do when they create paths) and others can respond to these modifications. Adam Smith's “invisible hand” is another example where individuals collectively generate an outcome they did not directly intend but that arose from their separate self-interested actions as they respond to price signals in the marketplace. 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.^{12[Com] 13[Web] 14[IA]}

The Web as an Organizing System

Today's web barely resembles the system for distributing scientific and technical reports it was designed to be when physicist and computer scientist Tim Berners-Lee devised it in 1990 at the European Organization for Nuclear Research (CERN) lab near Geneva. However, as an *Organizing System* the web still follows the principles that Berners-Lee defined at its creation. These include standard data formats and interaction protocols; no need for centralized control of page creation or linking; remote access over the network from anywhere; and the ability to run on a large variety of computers and operating systems. This architecture makes the web open and extensible, but gives it no built-in mechanisms for authority or trust.^{15[Web]}

Because the web works without any central authority or authorship control, any person or organization can add to it. As a result, even though the web as a whole does not exhibit the centralized intentional arrangement of resources that characterizes many Organizing Systems, we can view it as consisting of millions of Organizing Systems that each embody a separate intentional arrangement of web pages. In addition, we most often interact with the web indirectly by using a search engine, which meets the definition of *Organizing System* because its indexing and retrieval algorithms are principled.

A great many Organizing Systems are implemented as collections of web pages. Some of these collections are created on the web as new pages, some are created by transforming existing collections of resources, and some combine new and existing resources.

The requirement for intentional arrangement excludes naturally occurring patterns created by physical or geological processes from being thought of as Organizing Systems. There is information in the piles of debris left after a tornado or tsunami and the strata of the Grand Canyon. But they are not Organizing Systems because the patterns of arrangement were created by deterministic natural forces rather than by agents following one or more organizing principles. On the other hand, collections of geological data like the measurements of chemical composition from different strata and locations in the Grand Canyon **are** Organizing Systems. Decisions about what to measure, how to combine and analyze the measurements, and any theories that are tested or created, reflect intentional arrangement of the data by the geologist.

Other patterns of resource arrangements are illusions or perceptions that require a particular vantage point. The best examples are patterns of stars as they appear to an observer on Earth. The three precisely aligned stars, often described as "Orion's belt," are hundreds of light years from Earth, and also from each other. The perceived arrangement of the stars is undeniable, but the stars

are not aligned in the universe. Astronomical constellations like Orion are *intentional arrangements* imposed on our perceived locations of the stars, and these perceived arrangements and the explanations for them that constellations provide, form an Organizing System that is deeply embedded in human culture and in the practice of celestial navigation over the seas.

Not an Intentional Arrangement



The composition and arrangement of the rock layers (“strata”) in the Grand Canyon in the Southwest United States have been studied extensively by geologists. The composition of rock suggests the environment in which it was formed, and the absolute and relative arrangement of the rock layers reveals the timing of important geological events.

(Photo by B. Rosen. Creative Commons CC BY-ND 2.0 license.)

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”

The arrangements of resources in an Organizing System follow or embody one or more organizing principles that enable the Organizing System to achieve its purposes. *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. We call this design philosophy “Architectural Thinking” (§11.5.2.)

Organizing Spices By Cuisine



An alternative to organizing spices alphabetically is to organize them according to cuisines or flavor profiles, which can be defined in terms of ingredients and spices that tend to be used together. Patricia Glushko organizes her spices into three groups: Indian (includes cayenne pepper, coriander, cumin, turmeric), Mediterranean / Middle Eastern (includes basil, dill, oregano, paprika, thyme), and seeds. Each group of spices is in a separate large container, which makes it convenient when cooking.

(Photo by R. Glushko.)

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 those that have influenced the design of Organizing Systems for thousands of years. For example, we organize many collections of resources using the properties that are easiest to perceive, or whose values vary the most among the items in the collection, because these principles make it easy to locate a particular resource. We also group together resources that we often use together, we make resources that we use often more accessible than those we use infrequently, and we put rare or unique resources where we can protect them. Very general and abstract organizing principles are sometimes called design heuristics (e.g., “make things easier to find”). More specific and 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.

Because this book was motivated by the goal of broadening the study of information organization beyond its roots in library and information science, it emphasizes organizing principles with a specific functional purpose like identifying, selecting, retrieving, or preserving resources. However, for thousands of years people have systematically collected things, information about those things, and observations of all kinds, organizing them in an effort to understand how their world works; the Babylonians created inventories and star charts; ancient Egyptians tracked the annual Nile floods; and, Mesoamericans created astronomical calendars. The term *sensemaking* is often used to describe this generic and less specific purpose of organizing to derive meaning from experience by fitting new events or observations into what they already know.^{16[CogSci]}

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). (See the sidebar, *The Three Tiers of Organizing Systems* (page 48).)

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. The essence of a library Organizing System emerges from the resources that it organizes and the interactions with the resources that it enables. Users typically care a lot about the interactions they can perform, like the kinds of searching and sorting allowed by the online library catalog. How the resources and interactions are implemented are typically of little concern.

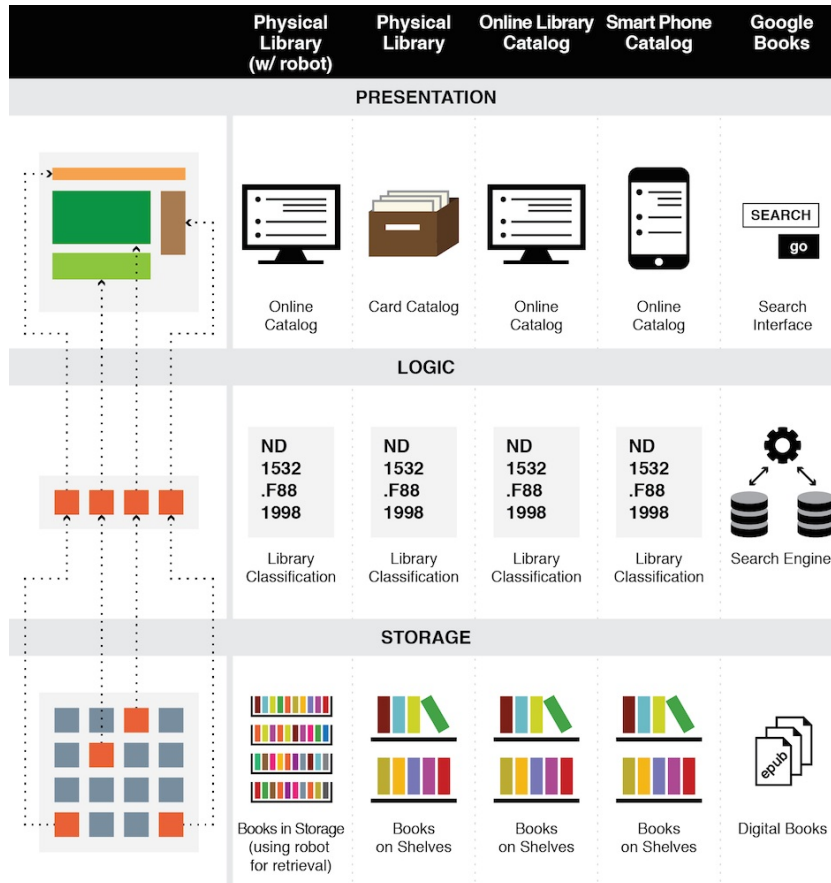
The separation of organizing principles and their implementation is harder to recognize in an Organizing System that only contains physical resources, such as your kitchen or clothes closet, where you appear to have unmediated interactions with resources rather than accessing them through some kind of user interface or “presentation tier” that supports the principles specified in the “middle tier” and realized in the “storage tier.” As a result, people can easily get distracted by presentation-tier concerns. Too often we waste time color-coding file folders and putting labels on storage containers, when it would have better to think more carefully about the logical organization of the folder and container contents. It does not help to use colors and labels to make the logical organization more salient if that is not well designed first.

One place where you can easily appreciate these different tiers for physical resources is 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. 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.

Figure 1.2, Presentation, Logic and Storage Tiers. illustrates the separation of the presentation, logic, and storage tiers for four different types of library Organizing Systems and for Google Books. No two of them are the same in every tier. Note how a library that uses inventory robots to manage the storage of books does not reveal this in its higher tiers. (See the sidebar, *Library Robot* (page 132).)

Because tangible things can only be in one place at a time, many Organizing Systems, like those in the modern library with online catalogs and physical collections, resolve this constraint by creating digital proxies or surrogates to organize their tangible resources, or create parallel digital resources (e.g., digitized books).^{17[Web]} The implications for arranging, finding, using and reusing resources in any Organizing System directly reflect the mix of these two embodiments of information; in this way we can think of the modern library as a digital Organizing System that primarily relies on digital resources to organize a mixture of physical and digital ones.

The Organizing System for a small collection can sometimes use only the minimal or default organizing principle of *colocation*—putting all the resources in the same location: in the same container, on the same shelf, or in the same email in-box. If you do not cook much and have only a small number of spices in your kitchen, you do not need to alphabetize them because it is easy to find the one you want.^{18[Com]}

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.

Some organization emerges implicitly through a *frequency of use* principle. In your kitchen or clothes closet, the resources you use most often migrate to the front because that is the easiest place to return them after using them. But as a collection grows in size, the time to arrange, locate, and retrieve a particular resource becomes more important. The collection must be explicitly organized to make these interactions efficient, and the organization must be preserved after the interaction takes place; i.e., resources are put back in the place they were

The Three Tiers of Organizing Systems

Software architects and designers agree that it is desirable to build applications that separate the storage of data, the business logic or functions that use the data, and the user interface or presentation components through which users or other applications interact with the data. This modular architecture allows each of the three tiers to be upgraded or reimplemented independently to satisfy changed requirements or to take advantage of new technologies. An analogous distinction is that between an algorithm as a logical description of a method for solving a computational problem and its implementation in a particular programming language like Java or Python.

These architectural distinctions are equally important to librarians and information scientists. Our new way of looking at Organizing Systems emphasizes the importance of identifying the desired interactions with resources, determining which organizing principles can enable the interactions, and then deciding how to store and manage the resources according to those principles. Applying architectural thinking to Organizing Systems makes it easier to compare and contrast existing ones and design new ones. Separating the organizing principles in the “middle tier” from their implications in the “data” and “presentation” tiers often makes it possible to implement the same logical Organizing System in different environments that support the same or equivalent interactions with the resources. For example, a new requirement to support searching through a library catalog on a smart phone would only affect the presentation tier.

found. As a result, most Organizing Systems employ organizing principles that make use of properties of the resources being organized (e.g., name, color, shape, date of creation, semantic or biological category), and multiple 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 or style and separate your grilling equipment from the wok and other items you use for making Chinese food.

Unlike those for physical resources, the most useful organizing properties for information resources are those that reflect their content and meaning, and these are not directly apparent when you look at a book, document, or collection of data. Significant intellectual effort or statistical computation is necessary to reveal these properties when assigning subject terms, creating an index, or using them as input features for machine learning and data analysis programs.

Separation Of Organizing Principle From Implementation



Whether spices are organized alphabetically by their names, by cuisines, by season, by frequency of use, or any other principle, this decision is logically distinct from the physical arrangement of the spices. There are many types of spice racks, shelves, circular “lazy susans,” and other devices designed for arranging spices.

(Photo collage created by R. Glushko from various web catalogs.)

The most effective Organizing Systems for information resources often are based on statistical 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 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 re-calculate the probability that words like “beneficiary” or “Viagra” indicate whether a message is “spam” or “not spam” in the collection of messages processed.

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...).^{19[LIS]} We use the more general word, *agent*, for any entity capable of autonomous and intentional organizing effort, because it treats organizing work done by people and organizing work done by computers as having common goals, despite obvious differences in methods.

We can analyze agents in Organizing Systems to understand how human and computational efforts to arrange resources complement and substitute for each other. We can determine the economic, social, and technological contexts in which each type of agent can best be employed. We can determine how the Organizing System allocates effort and costs among its creators, users, maintainers and other stakeholders.

A group of people can be an organizing *agent*, as when a group of people come together in a service club or standards body technical committee in which the members of the group subordinate their own individual agency to achieve a collective good.

We also use the term *agent* when we discuss interactions with Organizing Systems. The entities that most typically access the contents of libraries, museums, or other collections of physical resources are human agents—that is, people. In other Organizing Systems, such as business information systems or data repositories, interactions with resources are carried out by computational processes, robotic devices, or other entities that act autonomously on behalf of a person or group.

In some Organizing Systems, the resources themselves are capable of initiating interactions with other resources or with external agents. This is most obvious with human or other living resources, where a critical part of the design of any Organizing System with them is determining what kinds of interactions they should be encouraged or allowed to initiate. We will return to this issue after we discuss the design of interactions with ordinary resources that are passive, the situation in most Organizing Systems that involve physical resources.

Other resources that can initiate interactions are resources augmented with sensory, computational or communication capabilities that enable them to obtain information from their environment and then do something useful with it. You are probably familiar with RFID tags, which enable the precise identification and location of physical resources as they move through supply chains and stores, and with “smart” devices like Nest thermostats that learn how to program themselves.

1.8 The Concept of “Interactions”

An *interaction* is an action, function, service, or capability that makes use of the resources in a collection or the collection as a whole. 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* any resource in the collection, and discriminating or *selecting* among similar resources.^{20[LIS]}

Some of the interactions with resources in an Organizing System are inherently determined by the characteristics of the resource. Because many museum resources are unique or extremely valuable, visitors are allowed to view them but cannot borrow them, in contrast with most of the resources in libraries. A library might have multiple printed copies of *Moby Dick* but can never lend more of them than it possesses. 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.

For works not in the public domain, copyright law gives the copyright holder the right to prevent some uses, but at the same time “fair use” and similar copyright doctrines enable certain limited uses even for copyrighted works.^{21[Law]}

Digital resources enable a greater range of interactions than physical ones. Any number of people or processes can request a weather forecast from a web-based weather service because the forecast is not used up by the request 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 *application program interfaces (APIs)* in the Organizing System. In particular, translation, summarization, annotation, and keyword suggestion are highly useful services that are commonly supported by web search engines and other web applications. Similarly, an Organizing System with digital resources can implement a “keep everything up to date” interaction that automatically pushes current content to your browser.

But just as technology can enable interactions, it can prevent or constrain them. If your collection of digital resources (ebooks or music, for example) is not stored on your own computer or device, a continuous Internet connection is a requirement for access. In addition, access control policies and *digital rights management (DRM)* technology can limit the devices that can access the collection and prevent copying, annotation and other actions that might otherwise be enabled by the fair use doctrine.

Interaction design is especially crucial for managing resources that have the capability to initiate interactions with each other or with external agents. Consider the vast differences in how workers behave in businesses organized according to principles of scientific management and those that embody the Kaizen principles of continuous improvement. In the former, work is highly standardized and bureaucratic, giving workers little autonomy. In the latter, work is also standardized, but workers are motivated to analyze and improve work processes whenever possible, and they are given great discretion in how to do that.^{22[Bus]}

Just as with organizing principles, it is useful to think of interactions in an abstract or logical way that does not assume an implementation because it can encourage innovative designs for Organizing Systems.

1.9 The Concept of “Interaction Resource”

Interactions with physical resources sometimes leave traces or other evidence. Many of these traces are unintentional, like fingerprints, a coffee cup stain on a newspaper, or the erosion on a shortcut path across a lawn. Fans of *Sherlock Holmes* and *CSI* know that clever forensic investigators can use these residues of interactions to identify or vindicate suspects. Other interaction traces are intentional, like a student's yellow highlighting or notes in a textbook or spray-painted graffiti on a building. But not every interaction leaves a trace, traces fade over time, and different traces associated with the same resource lack consistency. This means that most 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. Each record of a user choice in accessing, browsing, buying, highlighting, linking, and other interactions then becomes an “interaction resource” that can be analyzed to reorganize the resource collection or otherwise influence subsequent interactions with the primary resources.

Interaction resources are often essential pieces of information that make Organizing Systems function. Most human toll-takers have been replaced by smart “toll tags” that broadcast their identity when the car they are in passes a radio receiver at a tolling location. Each interaction resource created identifies an account and credit card with which to pay the toll; taken together, the collection of these interaction resources can be used as the primary resources in other Organizing Systems that manage traffic congestion, or that 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.

Endnotes for Chapter 1

[1][LIS] (Nunberg 1996, 2011). (Buckland 1991). See also (Bates 2005).

[2][Com] (Glushko and McGrath 2005).

[3][Law] (Buckland 1997). The idea that an antelope could be a document was first proposed in (Briet 1951).

A commercial fisherman in Florida was found with fish in his catch below the legal size limit. An inspector ordered him to return to port and hand the fish over to the authorities; when he dumped them overboard instead, he was charged with violating the Sarbanes-Oxley Act, a law drafted in response to high-profile white-collar crimes such as the Enron scandal. The law imposes harsh penalties for destroying “any record, document, or tangible object” to impede a federal investigation. The fisherman argued that the law should only apply to written documents, but the United States government contended that because the fish were “tangible objects” whose presence on the boat served as the only documentation of the allegedly illegal fishing, there was no practical difference between a fish and a document in this case. The Supreme Court ruled in favor of the fisherman, finding that “tangible object” must be interpreted in the context of “record” and “document” and, as such, only applies to an object “used to record or preserve information.” The fact that a fish is tangible evidence in this case does not make it a document.

(Buckland 1991).

(Liptak 2014). Brief for the United States in Opposition, *Yates v. United States*. SCOTUSblog, March 14, 2014. <http://www.scotusblog.com/case-files/cases/yates-v-united-states/>

For the complete history of the case, see: <http://www.scotusblog.com/case-files/cases/yates-v-united-states/>.

See also the related *Sarbanes-Oxley Act* endnote.¹¹⁸ [Law]

[4][DS] The DIKW hierarchy seems to have been inspired by *The Rock, A Pageant Play* (Eliot 1934) by the poet T S Eliot, whose opening chorus contains these lines:

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

Most people credit Ackoff’s *From Data to Wisdom* (Ackoff 1989) as the first articulation of the hierarchy in an information science and systems context. The hierarchy is mentioned in nearly twenty textbooks, but their close analysis by (Rowley 2007) reveals only partial agreement on the definitions and relation-

ships among the four key concepts. The hierarchy has been criticized as lacking in philosophical rigor (Fricke 2009) and for ignoring the context-specificity of how knowledge is learned and applied (Jennex 2009). (Larose 2014)

[5][LIS] We can continue the debate in the previous paragraphs and the sidebar, **What Is Information?** (page 28) by pointing out that in both common and professional usage, “bibliographic” activities involve describing and organizing information resources of the kinds that might be found in a library. But noted information scientist Patrick Wilson argued for a much broader expanse of the bibliographic universe, suggesting that “it includes manuscripts as well as printed books, bills of lading and street signs as well as personal letters, inscriptions on stone as well as phonograph recordings of speeches, and most notably, memorized texts in human heads and texts stored up in the *memories* of machines” (Wilson 1968, p. 12).

[6][DS] Siegel’s *Predictive Analytics: The Power to Predict who will Click, Buy, Lie or Die* (Siegel 2013) is written for a non-technical audience and enthusiastically describes over 100 applications. *The Master Algorithm* (Domingos 2015) shares Siegel’s enthusiasm but is far more technical; the book attempts to explain and compare the five “tribes” of machine learning: the symbolists, connectionists, evolutionaries, Bayesians, and analogizers. The title of Chris Anderson’s provocative article in Wired Magazine (Anderson, 2008) is self-explanatory: “The end of theory: The data deluge makes the scientific method obsolete.”

“Difference in kind or difference in degree” is an important issue in legal contexts and more generally arises whenever there is a disagreement about whether some difference or change is strict and categorical or whether it is incremental. We introduce it here so that readers can think critically about the socio-business-technical changes that might come about as a result of new methods and technologies for organizing and analyzing data. We believe that data science is on its way to becoming an important part of the organizing tool box. But everyone needs to remember that humans own the tool box, and that they design and build the tools..

[7][LIS] (Svenonius 2000).

[8][Web] The URI identifies a resource as an abstract entity that can have “multiple representations,” which are the “things” that are actually exposed through applications or user interfaces. The HTTP protocol can transfer the representation that best satisfies the content properties specified by a web client, most often a browser. This means that interactions with web resources are always with their representations rather than directly with the resource *per se*. The representation of the resource might seem to be implied by the URI (as when it ends in *.htm* or *.html* to suggest text in *Hypertext Markup Language (HTML)* format), but the URI is not required to indicate anything about the “representation.” A

web resource can be a static web page, but it can also be dynamic content generated at the time of access by a program or service associated with the URI. Some resources like geolocations have “no representations at all;” the resource is simply some point or space and the interaction is “show me how to get there.” The browser and web server can engage in “content negotiation” to determine which “representation” to retrieve, and this is particularly important when that format further requires an external application or “plug-in” in order for it to be rendered properly, as it does when the server returns a Power Point file or an other file format that is not built into the browser.

Internet architecture’s definition of *resource* as a conceptual entity that is never directly interacted with is difficult for most people to apply when those resources are physical or tangible objects, because then it surely seems like we are interacting with something real. So we will most often talk about interactions with resources, and will mention “resource representations” only when it is necessary to align precisely with the narrower Internet architecture sense.

[9][Phil] In addition, groups of people have come together to form “intentional communities” for thousands of years in monasteries, communes, artist colonies, cooperative houses, and religious or ethnic enclaves so they can live with people who share their values and beliefs. A directory of intentional communities organized by type and location is managed by the [Fellowship of Intentional Communities](#).

[10][Bus] The shift from a manufacturing to an information and services economy in the last few decades has resulted in greater emphasis on intellectual resources represented in skills and knowledge rather than on the natural resources of production materials and physical goods.

The intellectual resources of a firm are embodied in a firm’s people, systems, management techniques, history of strategy and design decisions, customer relationships, and intellectual property like patents, copyrights, trademarks, and brands. Some of this knowledge is explicit, tangible, and traceable in the form of documents, databases, organization charts, and policy and procedure manuals. But much of it is tacit: informal and not systematized in tangible form because it is held in the minds and experiences of people; a synonym is “know-how.” A more modern term is *Intellectual Capital*, a concept originated in a 1997 book with that title (Stewart 1997).

[11][Law] In 2004, Google began digitizing millions of books from several major research libraries with the goal of making them available through its search engine (Brin 2009). But many millions of these books are still in copyright, and in 2005 Google was sued for copyright infringement by several publishers and an author’s organization. In 2011 a US District Court judge rejected the proposed settlement the parties had negotiated in 2008 because many others objected to

it, including the US Justice Department, several foreign governments, and numerous individuals (Samuelson 2011).

The major reason for the rejection was that the settlement was a “bridge too far” that went beyond the claims made against Google to address issues that were not in litigation. In particular, the judge objected to the treatment of the so-called “orphan works” that were still under copyright but out of print because money they generated went to the parties in the settlement and not to the rights holders who could not be located (why the books are “orphans”) or to defray the costs of subscriptions to the digital book collection. The judge also was concerned that the settlement did not adequately address the concerns of academic authors—who wrote most of the books scanned from research libraries—who might prefer to make their books freely available rather than seek to maximize profits from them. Other concerns were that the settlement would have entrenched Google’s monopoly in the search market and that there were inadequate controls for protecting the privacy of readers.

Google’s plan would have dramatically increased access to out of print books, and the rejection of the proposed settlement has heightened calls for an open public digital library (Darnton 2011). A good start toward such a library was the digital copies that the research libraries received in return for giving Google books to scan, which were collected and organized by the Hathi Trust (See the sidebar, *The Hathi Trust Digital Library* (page 142)). In 2010, the Alfred P. Sloan Foundation provided funding to launch the *Digital Public Library of America* (DPLA): <http://dp.la/>. This non-proprietary goal might induce the US Congress and other governments to pass legislation that fixes the copyright problems for orphan works.

[12][Com] Self-organizing is also used to describe phenomena like climate, neural networks, and phase transitions and equilibrium states in physics and chemistry. But when systems involve collections of *inanimate* resources that are very large and open, with complex interactions among the resources, it seems less sensible to attribute intentional arrangement to the outcomes. The resource arrangements that emerge cannot always be interpreted as the result of intentional or deterministic principles and instead are more often described in probabilistic or statistical terms. And even though it involves animate resources, Charles Darwin’s “natural selection” in evolutionary biology is a self-organizing mechanism where intentionality is hard to pinpoint or absent entirely.

The rules governing these local interactions can be simple and yet produce highly complex structures. For example, in flocks of birds or schools of fish the rules are: (1) follow things like you, (2) do not bump into each other, but stay close, and (3) move in the same direction as the rest of the group. With just these three rules computer models can create complex three-dimensional ar-

rangements that can make abrupt changes in shape and density while moving rapidly, just as live things do. (Friederici 2009)

The term “Crowdsourcing” was invented by Jeff Howe in a June 2006 article in *Wired* magazine, and the concept was developed further in a book published two years later (Howe 2006, 2008). “Folksonomy” was coined by Thomas Van der Wal at about the same time in 2004; see <http://vanderwal.net/folksonomy.html> and (Trant 2009).

(Goldstone and Gureckis 2009) present a cognitive science perspective on collective behavior, analyze important themes and controversies, and suggest areas for future research. (Moussaid et al. 2009) analyze self-organizing phenomena in animal swarms and human crowds in terms of information exchange among individuals.

Self-organizing behaviors in ants, bees, bats, cuckoos, fireflies and other animals have been analyzed to identify heuristics that can be applied to difficult optimization problems in network design, cryptography, and other domains where deterministic algorithms are infeasible. (Yang 2010)

(Smith 1776)

[13][Web] (Banzhaf 2009).

[14][IA] The concept of a web page is imprecise because many web pages, especially home pages designed as navigation gateways to an organized collection of pages, are constructed from heterogeneous blocks of content that could have been organized as separate pages.

[15][Web] The “plain web” (Wilde 2008a), whose evolution is managed by the *World Wide Web Consortium (W3C)*, is rigorously standardized, but unfortunately the larger ecosystem of technologies and formats in which the web exists is becoming less so. Web-based Organizing Systems often contain proprietary media formats and players (like Flash) or are implemented as closed environments that are intentionally isolated from the rest of the web (like Facebook or Apple’s iTunes and other smart phone “app stores”).

[16][CogSci] (Weick et al , 2005 p. 410).

[17][Web] Instead of thinking of a digital book as a “parallel resource” to a printed book, we could consider both of them as alternate representations of the same abstract resource that are linked together by an “alternative” relationship, just as we can use the HTML ALT tag to associate text with an image so its content and function can be understood by text-only readers.

[18][Com] For collections of non-trivial size the choice of searching or sorting algorithm in computer programs is a critical design decision because they differ greatly in the time they take to complete and the storage space they require.

For example, if the collection is arranged in an unorganized or random manner (as a “pile”) and every resource must be examined, the time to find a particular item increases linearly with the collection size. If the collection is maintained in an ordered manner, a binary search algorithm can locate any item in a time proportional to the logarithm of the number of items. Analysis of algorithms is a fundamental topic in computer science; a popular textbook is *Introduction to Algorithms* by (Cormen et al. 2009).

[19][LIS] For precise distinctions, see the US Department of Labor, Bureau of Labor Statistics occupational outlook handbooks at <http://www.bls.gov/oco/ocos065.htm> and <http://www.bls.gov/oco/ocos068.htm> and <http://www.michellemach.com/jobtitles/realjobs.html>.

[20][LIS] The four objectives listed in this paragraph as those proposed in 1997 by the *International Federation of Library Associations and Institutions (IFLA)*. The first statement of the objectives for a bibliographic system was made by (Cutter 1876), which (Svenonius 2000) says it is likely the most cited text in the bibliographic literature. Cutter called his three objectives “finding,” “co-locating,” and “choice.”

[21][Law] Copyright law, license or contract agreements, terms of use and so on that shape interactions with resources are part of the Organizing System, but compliance with them might not be directly implemented as part of the system. With digital resources, digital rights management (DRM), passwords, and other security mechanisms can be built into the Organizing System to enforce compliance.

[22][Bus] Frederick Taylor developed “scientific management” to improve industrial efficiency and conducted detailed time and motion studies to devise what he thought were optimal ways to perform work tasks (Taylor 1914). The Kaizen principles of continuous improvement were introduced to Western audiences by Imai Masaaki and by numerous books about their application in the Toyota production system (Masaaki 1986).

Scientific management views a business as a machine, while Kaizen principles treat it as a brain that learns. These metaphors for business organization are among those described by (Morgan 1997) in a classic business textbook. Other metaphors discussed include organisms, cultures, political systems, and psychic prisons.