Chapter 1. Foundations for "Organizing Systems"

To appear in R. Glushko & C. Borgman (Eds.),
The Discipline of Organizing, 2011

Robert J. Glushko
glushko@berkeley.edu

Contents

1.1 The Discipline of Organizing

We organize things. The local public library organizes printed books, periodicals, maps, CDs, DVDs, and maybe some old record albums. The large university research library also organizes 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. And each of us organizes the many kinds of things in our lives—our books in our homes; financial records in folders and filing cabinets; clothes in dressers and closets; cooking and eating utensils in kitchen drawers and cabinets.

We organize information about things. Library card catalogs 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. And each of us organizes information about things, when we sell our

unwanted stuff on eBay, when we tag our photos on Flickr or Facebook, or when we rate a restaurant on Yelp.

We organize information. The printed cards of the library catalog are now searchable as an online catalog. Large research libraries organize 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. Web portals and aggregation sites organize links to other resources, and different sites might embody substantially different organization because of contrasting goals or because of the degree to which the organization is determined computationally. Web search engines take the computational route and use content and link analysis along with relevance ratings to discriminate among the billions of web pages competing for our attention. Web-based services and information resources can be combined as "mash-ups" or choreographed to carry out information-intensive business models. And each of us organizes personal digital information—email, documents, media, appointments, and contacts—on our computers or smart phones, or in "the cloud" through information services that use Internet protocols.

But let's take a closer look at these three different types or contexts of organizing. Are there clear, principled and useful distinctions between them? An X-ray might be created as digital information, but becomes a thing when it is printed on film, and is information about a thing (the patient getting the X-ray) when being studied by a physician. What is the relationship between a printed document and the digital object created when the document is scanned? Is this the same relationship if the document was first created in digital form and later printed? The differences and relationships between "things" and "information" have long been discussed and debated by philosophers and others; the debate continues because no one has been able to propose distinctions that make sense in every situation.

Certainly there are differences between a printed X-ray and a digital one, or between a book and its bibliographic description in a printed or online catalog. And because the contexts in which we organize differ in so many respects we naturally focus on what is being organized, or on the specific organizing principles, technology, functions or features, individuals or companies involved. So at first glance it might seem that organizing physical resources like books or machine parts has an entirely different character than organizing digital documents or information services.

Nevertheless, if we take a more abstract look at organizing we can bridge the intellectual gulf that separates the many disciplines that share the goal of organizing but that differ in what they organize. Instead of trying to answer philosophical questions, we can think of organizing in ways that don't require us to make distinctions that don't always help. (See the sidebar, "What is Information?").

One reason why these distinctions are elusive is that whether we are organizing things, organizing information about things, or organizing information, we always rely on still another form of organization – the organization of the concepts and words in the language we use when we organize anything. Every culture and its associated language community relies on a rich network of meanings and distinctions about objects, events, settings, mental states, properties, relations and other components of experience that community members use implicitly in their interactions with the world and with each other. And when we say that "Charles Dickens is the

author of A Tale of Two Cities", the meaning of "author" doesn't depend on whether we have a printed copy or a Kindle copy of the book in mind.

This organization of concepts and words is explicit in dictionaries, controlled vocabularies, thesauri, taxonomies, ontologies, and other representations that share the basic goal of specifying how meanings and words are related but differ in how formally and precisely they do it.

We can also recognize that every system of organization involves a collection of resources, a choice of properties or principles used to describe and arrange them, and ways of accessing and interacting with them. 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 lifecycle. We can create a discipline of organizing in a disciplined way.

WHAT IS INFORMATION?

How "things" and "information" contrast, what they have in common, and how to deal with them together are not just interesting philosophical questions. A discipline of organizing has to apply to all kinds of resources, and to do so it must find a way to rise above the debate.

"Information" is an abstract concept, and most of its hundreds of definitions treat it as an idea that swirls around equally hard to define terms like "data," "knowledge," and "communication" (Buckland, 1991; Bates, 2005; Nunberg, 2011). An abstract view of information as an intangible thing is the intellectual foundation for both modern information science and the information economy and society.

But the abstract view of information must coexist with and often conflicts with a much older idea that information is a tangible thing. Five thousand years of human experience with tangible information artifacts has deeply embedded the notion of information as a physical thing in culture, language, and methods of information design and organization, and it always will, so long as we humans inhabit a physical world that contains people and objects that we name, classify, and organize. And even intangible information must be stored and processed by tangible machines, whether massive racks of servers or a pocket-sized smartphone.

Buckland's oft-cited essay "Information as Thing" (Buckland, 1991) rebuts the notion that information is inherently intangible and instead defines it more broadly and provocatively based on function. This makes the objects in museum or personal collections into information-as-thing resources because they can be learned from and serve as evidence. Analogously, when analyzing "What is a Document?" Buckland rejects narrow definitions that emphasize traditional physical forms. Thus an antelope is "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 (Buckland, 1997).

1.2 The "Organizing System" Concept

We propose to unify these many perspectives about organizing and information with the concept of an Organizing System, defined as *an intentionally arranged collection of resources and the interactions they support*. This definition brings together four essential ideas that we will briefly introduce here and develop in detail in subsequent chapters.

- "Resource" has an ordinary sense of "anything of value that can support goal-oriented activity." This broad definition means that a resource can be a physical thing, information about physical things, information about non-physical things, or anything you want to organize. Other words that aim for this comprehensive scope are *entity*, *object*, *item*, *instance*, and *asset*; *document* is often used for any information resource in either digital or physical format; *artifact* has long been used for physical resources created by people and more recently to refer to digital ones in the domains of design or software engineering. The data modeling and information architecture disciplines often use the concept of *information component* to refer to the primitive or smallest units of meaning.

  "Resource" has specialized meaning in Internet architecture. HTTP, the protocol for transferring files across the web, uses "Uniform Resource Identifiers" (URIs). 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 with the URI. Treating as a 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.
- "*Collection*" has deep roots as a concept in libra*ries*, museums and other institutions that select, assemble, *arrange*, and preserve resources. *"Set"* has a similar meaning as a "gathering together of items" but *set* is a foundational and formal concept in mathematics. "*Aggregate*" likewise has a formal sense of *a group or c*ollection of items "that go together" in data *modeling,* but in ordinary usage it typically means an unintentional or accidental collection. 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 *collect*ion) as well as digital ones (my collection of digital music). The sets of resources in scientific or business organizing systems are often described as *datasets* rather than as collections, but it does not overly distort the collection concept because effective scientific and business enterprises businesses are systematic and strategic in deciding what they need to manage and the information they need to do that, just as a library or museum is systematic at developing its collection.
- "*Intentionally Arranged*" emphasizes explicit or implicit acts of organization by people or by computational processes. We can sort resources into different categories, assign descriptive terms to them, apply standard classifications to them, arrange them in hierarchies or networks, or create other kinds of relations and structure. A collection can be arranged in physical locations, in databases or file systems managed by resource descriptions, on the Internet by a pattern of information exchanges or hypertext links, or arranged by computational analysis of these patterns. These arrangements might themselves follow patterns that are influenced by architectural principles or standards for physical environments or web applications. But "intentional arrangement" excludes naturally-occurring patterns created by physical, geological, biological or genetic processes. A pile of debris left after a tornado or tsunami and the strata of the Grand Canyon might exhibit patterns with "consistent or characteristic arrangement" (Bates, 2006, p. 1033-1034), but they aren't Organizing Systems.
- "*Interactions*" include any *activity*, *function*, or *service* supported by or enabled with respect to the resources in a collection or with respect the collection as a whole. The interactions of people with physical resources involve perceptual and psychomotor actions that are

constrained by physical properties and human capabilities. In contrast, machines (and robots) can be designed to interact with physical resources in ways that take advantage of their physical properties. Interactions with information resources can be much more varied because the possible interactions can be any implementable function or service. . Interactions of people with information resources are mostly constrained by their cognitive capabilities rather than their perceptual and psychomotor ones. Interactions with either physical or information resources can also be governed by access policies that specify who can interact with a resource and the terms of interaction.

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. An Organizing System embodies the composite impact of the choices made on these design dimensions.

Our concept of Organizing System and the abstract notion of resources lets us confront head on the duality of information as an intangible thing versus information as a tangible thing that is more like the physical objects we organize on shelves and in cabinets and closets. 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, and its tangible properties become secondary. In contrast, the shoes in our closet and the cooking utensils in our kitchen aren't about anything else, which makes their tangible properties more important. .

When an Organizing System deals with tangible resources and their tangibility matters, it follows different principles and must conform to different constraints than when it deals with intangible resources – the most obvious one being that tangible things can only be in one place at a time. Many Organizing Systems—like that in the modern library with online catalogs and physical collections—resolve this constraint by creating digital proxies or surrogates for their tangible resources, thus accommodate both notions of information at the same time. The implications for arranging, finding, using and reusing resources in any Organizing System directly reflect the mix of these two embodiments of information.

An important benefit of the Organizing System concept is that it treats organizing work done by people and organizing work done by computers as having common goals, despite obvious differences in methods. Instead of a view that contrasts information organization as a human activity and information retrieval as a machine one, or information organization as a topic for library and information science and information retrieval as one for computer science, we can acknowledge that computers now assist people in organizing and that people contribute much of the information used by computers to enable retrieval. In this way the Organizing System framework 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, the more effectively it can be retrieved, and the more effort put into retrieving information, the less it needs to be organized first.

We can look at different Organizing Systems to understand how human and computational efforts complement and substitute for each other and to determine the economic, social, and technological contexts in which each can best be employed. And we can determine

how the Organizing System allocates effort and costs among its creators, users, maintainers and other stakeholders.

Another important concept that unifies information organization and information retrieval in the Organizing System framework is *description*. Many of the design dimensions concern the nature and extent of the descriptions of the resources being organized. These descriptions are later compared and combined when we search for, retrieve, and transform resources. This is not a new idea; the process of information retrieval is often characterized as comparing the description of a user's information need with descriptions of the information resources that might satisfy them. But the Organizing System framework highlights the factors that determine the nature and extent of these descriptions and how they determine the capabilities of the processes that locate, compare, combine, or otherwise use them. The Organizing System perspective, which views the traditional concerns of information organization and information retrieval in a more interconnected, systematic, and generative way, can readily adapt to new applications and technologies as they arise.

XML AND THE "SEMANTIC WEB"

The Web's astonishingly rapid adoption as a publishing medium was largely due to the conceptual and technical simplicity of HTML for using tags to mark up pieces of text according to how they should appear. When Tim Berners-Lee invented HTML in 1991, he consciously made the tradeoff for simplicity rather than expressive precision, and HTML's idea that "how you tag is what you see" is so easy to understand that school kids and computer-wary grandmothers can make web pages.

But designing the Web "for eyes" means that it couldn't be understood "by machines" and Berners-Lee and others came to realize that the Web could reach its full potential only it could represent what text meant and not just how it should appear. For this goal HTML was fundamentally inappropriate, because its fixed tag set cannot express anything about the meaning of the content between its tags or assert relationships between pieces of content or other web resources.

In 1998, HTML's limitations inspired the development of XML, the Extensible Markup Language, which enables the creation of markup languages whose tag sets ("element vocabularies" in XML-speak) encode the content distinctions important in particular domains. For example, an XML vocabulary for business transactions might have element types for "price", "item", "unit", "address", "buyer", and "seller". An XML vocabulary for a university application might have element types for "student", "instructor", "course", "semester", and so on.

A few years later, Tim Berners-Lee, James Hendler, and Ora Lassila proposed a more ambitious way to encode content types and relationships in a vision called the "Semantic Web" (Berners-Lee at al, 2001). The essential idea of the Semantic Web is the semantic annotation of resources using one of several dialects of the Resource Description Framework (RDF), a language for making assertions about resources and their relationships. We will discuss RDF and related semantic technologies in Chapter 9, "Describing Relations and Structures".

1.2.1 Traditional Ways to Organize "Organizing Systems"

Classifying Organizing Systems according to the kind of resources they contain is the most obvious and traditional approach. We distinguish law libraries from software libraries, knowledge management systems from data warehouses, and personal stamp collections from coin collections because they contain different kinds of resources.

Organizing Systems are also commonly distinguished by their dominant purposes or the priority of their common purposes. For example, libraries, museums, and archives are often classified as "memory institutions" to emphasize their primary emphasis on resource preservation. In contrast, "management information systems" or "business systems" are categories that include the great variety of software applications that implement the Organizing Systems needed to carry out day-to-day business operations.

Classifying Organizing Systems according to the nature or size of the intended user community is another conventional approach. This size or scope can range from personal Organizing Systems created and used by a single person; to "community-based" Organizing Systems used by informal social groups; to those used by the employees, customers or stakeholders of an enterprise; to those used by an entire community or nation; to global ones potentially used by anyone in the world.

The size of the user community usually correlates with the extent, explicitness, and degree of central control of the organization imposed on the collection of resources and the likelihood that some or all of this organization was determined by computational processes like search engines or machine learning algorithms rather than by people. As a result, Organizing Systems can be classified according to the technologies or system architectures used to implement them, most often in business contexts.

We can get overwhelmed by this proliferation of categories for describing collections of resources, especially when the classification schemes aren't clearly based on just one of these many approaches. For example, the list of "library types" used by the International Federation of Library Associations to organize its activities includes resource-based distinctions (e.g. art libraries, law libraries, social science libraries), purpose-based ones (e.g., academic and research libraries), and user-based distinctions (e.g., public libraries, school libraries, libraries serving persons with print disabilities) (IFLA, 2011). Similarly, large businesses use different Organizing Systems and software applications for inventory management, records management, content management, knowledge management, customer relationship management, data warehousing and business intelligence, e-mail archiving, and other subcategories of collections.

1.2.2 Organizing "Organizing Systems" in a "Design Space"

Each way of thinking about Organizing Systems highlights an important set of design questions and allows us to contrast Organizing Systems that answer the questions differently. But no single classification is sufficient by itself to capture the important contrasts between Organizing Systems in a fixed set of categories.

So instead of a categorical view of Organizing Systems, we can think of them as existing in a multi-dimensional design space in which we can consider many types of collections are at the same time. This gives us a more consistent and comprehensive framework for describing and comparing them. We can describe familiar collection categories like libraries, museums, and

archives as Organizing System design patterns that embody characteristic configurations of design choices. We can then use these patterns to apply knowledge about familiar domains to unfamiliar ones; someone with a business or informatics background can better understand libraries and museums and have intelligent conversations with librarians and museum curators… and vice versa.

But at the same time, a dimensional perspective acknowledges the diversity of instances of these collection types and provides a generative, forward-looking framework for describing hybrid types that don't cleanly fit into the familiar categories and for inventing new or more specialized types of collections and their associated interactions. Thinking of Organizing Systems as points or regions in a multi-dimensional design space suggests new ones and the technological, managerial, or functional innovations that would be required to implement them.

For example, 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 if we treat "borrow and return the same item" as the type of interaction that defines a library. But if the items 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 (Wang, 2010). Even though it differs from traditional libraries on some dimensions, its similarities on other dimensions of its Organizing System give a Seed Library more "family resemblance" to a traditional library than to any other type of collection. And because the dimensional perspective doesn't impose sharp boundaries between categories, we can more readily notice that a Seed Library shares with museums the goal of preserving unique or rare items and with zoos the characteristic that the items in the collection are living species. 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.

Consider photo-sharing website Flickr, which functions for most of its users as a 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 library. But Flickr lacks the authoritative description and standard classification that define traditional libraries. On the other hand, Flickr offers application programming interfaces (APIs) and data feeds that enable clever software developers to invent new types of interactions that surpass those of traditional libraries. Many of these involve searching for photos with particular tags or that were taken at specified places and times with the results integrated into maps and timelines; see the Flickr "App Garden" for many more examples (Flickr, 2011).

A similar categorization challenge arises with the Google Books digitization project (see Sidebar).

THE GOOGLE BOOKS DIGITIZATION PROJECT

In 2004 Google began digitizing millions of books from several major research libraries with the goal of making them available through its search engine. 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 books" 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 didn't 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, which could perhaps be started using the digital copies that the research libraries received in return for giving Google books to scan. This non-proprietary goal might induce the US Congress and other governments to pass legislation that fixes the copyright problems for orphan works.

Google characterized its ambitious project to put tens of millions of books from research libraries online as "a library to last forever" (Brin, 2009). But the Google Books project was widely criticized as not being true to library principles, and we can readily identify design choices that are more characteristic of the Organizing Systems in business domains. First, Google classifies books using the bookstore-oriented BISAC categories rather than using the Library of Congress or Dewey Decimal systems used by academic or public libraries (Nunberg, 2009). Second, Google collects detailed records of user behavior, which conflicts with the cherished library goal to enable anonymous reading (Jones and Janes, 2010). Finally, Google uses these search records to serve targeted ads against book search results, just as it does for regular searches of ordinary web content**.** Legal scholar Pamela Samuelson noted that "anyone aspiring to create a modern equivalent of the Alexandrian library would not have designed it to transform research libraries into shopping malls" (Samuelson, 2009).

A characteristic feature of museums is that they provide highly mediated access to the resources in their collections, whose descriptions and classification are created by curators or professional catalogers. How then should we classify museums and scientific projects that enlist amateur volunteers as curators to describe and classify objects, as herbariaunited.org does for botanical specimens in British museum and university collections and as galaxyzoo.org does for astronomical images (Wright, 2010)?

These examples demonstrate that we can always create new categories by stretching the traditional definitions of "library" and "museum" and adding modifiers, as when we describe Flickr as a web-based photo-sharing library. But in doing so we suggest features that aren't there (like authoritative classification) and omit the features that are distinctive (like tagging by users). The Google Books project makes out-of-print and scholarly works vastly more accessible, but framing it in library terms to suggest it is a public good upsets many people with a more traditional sense of what the library category implies. To them, Google Books is simply another way for Google to make money.

But even though digital technology is radically subdividing the traditional categories of collections by supporting new kinds of specialized information-intensive applications, an opposite and somewhat paradoxical trend has emerged. The common challenges of "going

digital," and the architectural/functional constraints imposed by web implementations, are causing some convergence in the operation of libraries, museums, and archives (Trant, 2009a). Giving every physical resource in a collection a digital surrogate or proxy that is searchable and viewable in a web browser is "erasing the distinctions between custodians of information and custodians of things" (Gilliland-Swetland, 2000).

Taken together, these two trends have one profound implication. If the traditional categories for thinking about collections are splintering in some respects and converging in others, they are less useful in describing innovative collections and their associated interactions. Thus, we need a new concept – the Organizing System – that

- Applies comprehensively and consistently to collections of resources whether they are collections of information, collections of things, collections of information about things, and any combination
- Can embrace the traditional categories where they are appropriate, but not impose them on new types of collections and services where they just don't fit
- Makes it easier to trace the connections between specific requirements or constraints and particular functions or implementation choices.

1.3 The Design Decisions in Organizing Systems

Explicitly or by default, an Organizing System makes many interdependent decisions about the identities of resources; their names, descriptions and other properties; the classes, relations, structures and collections in which they participate; and the people or technologies who create, transform, combine, compare and use them. One important contribution of the idea of the Organizing System is that it moves beyond the debate about the definitions of things, documents, and information with the unifying concept of "resource" while acknowledging that "what is being organized" is just one of the questions or dimensions we need to analyze.

These decisions are deeply intertwined, but it is easier to introduce them as if they were independent. We introduce five groups of design decisions, itemizing the most important dimensions in each group:

- **What is being organized?** What is the mixture of things, information about things, and information in the Organizing System? What is the scope and scale of the domain? Do all non-digital resources have digital proxies or surrogates? Is the collection stable and maintained as it was at some specified time, or are resources continually added or deleted? Are the resources unique, or are they interchangeable members of a class or type? Do the resources change? Do they follow a predictable "life cycle" with a "useful life"—or are they living things?
- **Why it is being organized?** What functions or capabilities will be supported, and for whom? Are the uses and users known or unknown? Are the users primarily people or computational processes? What kind of access and interactions do users have with resources in the Organizing System? Can the resources be viewed, "checked out," or otherwise accessed or used outside of the collection? Does the Organizing System need to satisfy personal, social, or institutional goals?
- **How much is it being organized?** What is the extent or explicitness of description, classification, or relational structure being imposed? Is this description and structure imposed

in a centralized or top-down manner or in a distributed or bottom-up manner? What principles guide and are embodied in the organization? Do the descriptions and classifications conform to standards or are they ad hoc or proprietary?

- **When is it being organized?** Is the organization imposed at design time, or at runtime, just in case, just in time, all the time? Is any of this organizing, governance or curation activity mandated by law or shaped by industry or business practices?
- **How or by whom (or by what computational processes) is it being organized?** Is the organization being performed by individuals, by informal groups, by formal groups, by professionals, by automated methods? Are the organizers also the users? Are there rules or roles that govern the organizing activities of different individuals or groups?

How well these decisions and related ones coalesce in an Organizing System depends on the requirements and goals of its human and computational users, and on understanding the constraints and tradeoffs that any set of requirements and goals impose. How and when these constraints and tradeoffs are handled can depend on the legal, business and technological contexts in which the Organizing System is designed and deployed; on the relationship between the designers and users of the Organizing System (who may be the same people or different ones); on the economic or emotional or societal purpose of the Organizing System; and on numerous other design, deployment, and use factors.

Our concept of the Organizing System was in part inspired by and generalizes to physical and web-based resource domains the concepts proposed for bibliographic domains by Elaine Svenonius in "The Intellectual Foundation of Information Organization" (Svenonius, 2000). 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—who strongly identifies with a LIS perspective—unequivocally expresses the tradeoff principle in the Organizing 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 following sub-sections with a quote from her book.

1.3.1 What is Being Organized?

"What is difficult to identify is difficult to describe and therefore difficult to organize" (Svenonius, p. 13).

Before we can begin to organize any resource we need to identify it. It might seem straightforward to devise an Organizing System around tangible things, but we must be careful not to beg the question of what a thing is. In different situations the same thing can be treated as a unique item, as one of many equivalent members of a broad category, or as a sub-part of an item rather than as an item on its own. For example, in a museum collection, a handmade carved chess piece might be a separately identified item, identified as part of a set of carved chess pieces, or treated as one of the 33 unidentified sub-parts of an item identified as a chess set (including the board).

When merchants assign stock-keeping units to identify the things they sell, a SKU can be associated with a unique item, to sets of items treated as equivalent for inventory or billing purposes, or to intangible things like warranties. Different merchants or firms might make different decisions about abstraction and granularity when they assign SKUs because of

differences in suppliers, targeted customers, or other business strategies. If you take your car to the repair shop because windshield wiper fluid is leaking, you might be dismayed to find that the broken rubber seal that is causing the leak can't be ordered separately and you have to pay to replace the "wiper fluid reservoir" for which the seal is a minor but vital part. Similarly, when two business applications try to exchange and merge customer information, integration problems will arise if one describes a customer as a single "NAME" component while the other separates the customer's name into "TITLE", "FIRSTNAME," and "LASTNAME."

You probably don't have labels on the cabinets and drawers in your kitchen or clothes closet, but department stores and warehouses have signs in the aisles and on the shelves because of the larger number of things a store needs to organize. As a collection of resources grows, it often becomes necessary to explicitly identify each one; to create surrogates and descriptions like bibliographic records or metadata that distinguish one resource from another; and to create additional organizational mechanisms like shelf labels, store directories, library card catalogs and indexes that facilitate understanding the collection and locating the resources it contains.

Organization mechanisms like aisle signs, store directories and library card catalogs are embedded in the same physical environment as the resources being organized. But when these mechanisms or surrogates are computerized, the new capabilities that they enable create design challenges. This is because such an Organizing System can be designed and operated according to different principles than the Organizing System that only contains physical resources. A single physical resource can only be in one place at a time, and interactions with it are constrained by its size, location, and other properties. In contrast, digital copies and surrogates can exist in many places at once and enable searching, sorting, and other interactions with efficiency and scale impossible for tangible things.

When the resources being organized contain or consist of information content, deciding on the unit of organization is challenging because it might be necessary to look beyond physical properties and consider conceptual or intellectual equivalence. A high school student told to study Shakespeare's play "Macbeth" might treat any printed copy or web version as equivalent, and might even try to outwit the teacher by watching a film adaptation of the play. To the student, all versions of Macbeth seem to be the same resource, and to some degree she has a point: Organizing Systems that follow the rules set forth in the Functional Requirements for Bibliographic Records (Tillett, 2004) treat all the Macbeths as the same "work." However, they also enforce a hierarchical set of distinctions for finer-grained organization. FRBR views books and movies as different "expressions," different print editions as "manifestations," and each distinct physical thing in a collection as an "item." This Organizing System thus encodes the degree of intellectual equivalence while enabling separate identities where the physical form is important.

Preserving documents in their physical or original form is the primary purpose of many Organizing Systems that contain culturally, historically, or economically significant documents that have value as long-term evidence. Archives are a type of collection that focus on resources created by a particular person, organization, or institution, often during a particular time period. Typical examples of archives might be national or government document collections or the specialized Julia Morgan archive at the University of California, Berkeley (Online Archive of California, 2011), which houses documents by the famous architect who designed many of the university's most notable buildings as well as the famous Hearst Castle along the central California coast.

Archival Organizing Systems implement a distinctive answer to the question of what is being organized. Archives are sets of documents that have themselves been previously organized as a result of the processes that created and used them. This "original order" embodies the implicit or explicit Organizing System of the person or entity that created the documents and it is treated as an essential part of the meaning of the collection. As a result, the unit of organization for archival collections is the "fonds"—the original arrangement or grouping—and thus they are not re-organized according to other (perhaps more systematic) classifications. This fundamental principle of archival Organizing Systems was first defined by 19th century French archivists and is often identified as "respect pour les fonds."

The systematic appraisal and curation processes to identify and preserve documents that are characteristic of archives are also essential in the Organizing Systems of information- and knowledge-intensive firms. Businesses and governmental agencies are usually required by law to keep records of financial transactions, decision-making, personnel matters, and other information essential to business continuity, compliance with regulations and legal procedures, and transparency. As with archives, it is sometimes critical that these business knowledge or records management systems can retrieve the original documents, although digital copies that can be authenticated are increasingly being accepted as legally equivalent.

Thus, some Organizing Systems contain legal, business or scientific documents or data that are the digital descendants of paper reports or records of transactions or observations. These Organizing Systems might need to deal with legacy information that still exists in paper form or in electronic formats like image scans that are different from the structural digital format in which more recent information is likely to be preserved. When legacy conversions from printed information artifacts are complete or unnecessary, an Organizing System no longer deals with any of the traditional tangible artifacts. Digital libraries dispense with these artifacts, replacing them with the capability to print copies if needed. This enables libraries of digital documents or data collections to be vastly larger and more accessible across space and time than any library that stores tangible, physical items could ever be.

An increasing number of Organizing Systems handle information that has digital origins. Digital texts can be encoded with explicit markup that captures structural boundaries and content distinctions, which can be used to facilitate organization, retrieval, or both. In contrast, the digital representations of music, photographs, videos, and other non-text content like sensor data are structurally and semantically opaque. As a result, they are generally organized and retrieved using text surrogates or with descriptions extracted by computational processing of the content.

1.3.2 Why is it Being Organized?

"The central purpose of systems for organizing information [is] bringing like things together and differentiating among them" (Svenonius p. xi).

Almost by definition, the essential purpose of any Organizing System is to describe or arrange resources so they can be located and accessed later. How this general purpose is achieved depends on the types of resources or domains being organized, and in the personal, social, or institutional setting in which organization takes place. Chapter 8, "Classification," more fully explains the different purposes for Organizing Systems, the organizing principles they embody, and the methods for assigning items to classifications. Chapter 2, "Design Patterns for Organizing Systems," explains how these purposes and techniques apply in libraries, museums

and archives—the traditional domains of library and information science—and extends the analysis to business and scientific applications of content, knowledge, and data management.

"Bringing like things together" is an informal specification of the goals for many Organizing Systems. But because there will likely be a number of more precise requirements or constraints to satisfy, it is only a first approximation even for the Organizing Systems created by individuals with small collections of resources being organized. For example, the way you organize your home kitchen is influenced by the physical layout of counters, cabinets, and drawers; the dishes you cook most often; your skills as a cook, which may influence the number of cookbooks, specialized appliances and tools you own and how you use them; the sizes and shapes of the packages in the pantry and refrigerator; and even your height. But if other people also share the kitchen, you might have to negotiate and compromise on some of the decisions about what goes where because all users' requirements or preferences differ. The more people, companies, or stakeholders are involved in an Organizing System, the more likely this decision making will be explicit, contractual, or even governed by law.

When individuals manage their papers, books, documents, record albums, compact discs, DVDs, and other information sources or information artifacts, their Organizing Systems vary greatly. This is in part because the content of the resources being organized becomes a consideration. Put another way, an information resource has inherently more potential uses than resources like forks or frying pans, so it isn't surprising that the Organizing Systems in offices are even more diverse than those in kitchens.

Many of the Organizing Systems used by individuals are implemented by web applications, and this makes them more accessible because their resources can be accessed from anywhere with a web browser. For example, many people manage their digital photos with Flickr, their home libraries with LibraryThing, and their preferences for dining and shopping with Yelp. It is possible to use these "tagging" sites solely in support of individual goals, as tags like "myfamily," "toread," or "buythis" clearly demonstrate. But maintaining a personal Organizing System with these web applications potentially augments the individual's purpose with social goals like conveying information to others, developing a community, or promoting a reputation. Furthermore, because these community or collaborative applications aggregate and share the tags applied by individuals, they shape the individual Organizing Systems embedded within them when they suggest the most frequent tags for a particular resource. In Chapter 8 on "Classification," section 8.4 discusses this "Social/Distributed Classification" in more detail.

When the scale of the collection or the number of intended users increases, it becomes more important to be explicit about the goals of an Organizing System and in the description of the mechanisms or processes by which these goals are met. We can look back to the invention of mechanized printing in the seventeenth century, which radically increased the number of books and periodicals, as the motivation for libraries to develop systematic methods for cataloging and classifying what they owned and to view themselves as doing more than just preserving a collection. Libraries began progressively more refined efforts to state the functional requirements for their Organizing Systems and to be explicit about how they met those requirements.

Today, the Organizing Systems in a large academic research library must also support many functions and services other than those that directly support search and location of resources in their collections. These include billing systems, interlibrary loan record routing and systems, course reserves, licenses of digital resources from publishers, course material websites, and the library's own web presence. In these respects, the Organizing Systems in non-profit

libraries have much in common with those in corporate information repositories and business applications.

Any firm with an information-driven business model must have processes and technologies in place that govern information creation or capture and then manage its entire life cycle. These processes are diverse and complex: supporting transactions with customers or other firms to carry out business operations, to support research and innovation, and to develop business strategy and tactics in compliance with laws and regulations for accounting, taxes, human resources, data retention, and so on. In large firms these functions are so highly specialized and complex that the different types of Organizing Systems have distinct names: Enterprise Resource Planning (ERP), Enterprise Content Management (ECM), Supply Chain Management, Records Management, Customer Relationship Management (CRM), Business Intelligence (BI), Knowledge Management, and so on. Nevertheless, even though the most important functions in the Organizing Systems of large enterprises are those that manage the information resources needed for its business operation, these firms might also need to maintain corporate libraries and archives.

1.3.3 How Much is it Being Organized?

"It is a general bibliographic truth that not all documents should be accorded the same degree of organization" (Svenonius, p. 24).

Not all resources should be accorded the same degree of organization. In this section we will briefly unpack this notion of degree of organization into its two more important and related dimensions: the amount of description assigned to each resource and the amount of organization of resources into classes or categories. Chapter 5, "Describing Resources," Chapter 6, "Implementing Resource Descriptions," and Chapter 7, "Categories," more thoroughly address these questions about the nature and extent of description in Organizing Systems.

Not all resources in a collection require the same degree of description for the simple reason we discussed in Section 1.3.2: Organizing Systems exist for different purposes and to support different kinds of interactions or functions. Let's contrast two ends of the "degree of description" continuum. Many people use "current events awareness" or "news feed" applications that select news stories whose titles or abstracts contain one or more keywords. This exact match algorithm is easy to implement, but its all-or-none and one-item-at-a-time comparison misses any stories that use synonyms of the keyword; that are written in languages different from that of the keyword; or that are otherwise relevant but don't contain the exact keyword in the limited part of the document that is scanned. However, users with current events awareness goals don't need to see every news story about some event, and this limited amount of description for each story and the simple method of comparing descriptions are sufficient. In addition, tasks like current events awareness can't employ any organizing technique that takes any significant time, especially in domains where "time is money."

On the other hand, this simple Organizing System is inadequate for the purpose of comprehensive retrieval of all documents that relate to some concept, event, or problem. This is a critical task for scholars, scientists, inventors, attorneys and similar professionals who might need to discover every relevant document in some domain. Instead, this type of Organizing System needs rich bibliographic and semantic description of each document, most likely assigned by professional cataloguers, and probably using terms from a controlled vocabulary to enforce consistency in what descriptions mean.

When Panizzi (Panizzi, 1841) published his 91 cataloging rules for the British Library that defined authoritative forms for titles and author names, he ignited a debate about the cost-effectiveness of creating systematic and comprehensive descriptions of the resources in an information collection that continues to the present day (Anderson and Perez-Carballo 2001a, 2001b). The effort to comply with standards for bibliographic description is essential if resources are to be shared between libraries.

An alternative and complement to man-made descriptions for each resource are computer-generated indexes of their textual contents. These indexes typically assign weights to the terms according to calculations that consider the frequency and distribution of the terms in both individual documents and in the collection as a whole to capture more precisely what the documents are about. Naturally, these more sophisticated descriptions of the documents in the collection allow for more sophisticated query processing and comparison operations by the retrieval functions in the Organizing System. For example, query expansion mechanisms or thesauri can automatically add synonyms and related terms to the search. Additionally, retrieved documents can be arranged by relevance, while "citing" and "cited-by" links can be analyzed to find related relevant documents.

A second constraint on the degree of organization comes from the absolute size of the collection within the scope of the Organizing System. Organizing more resources requires more descriptions to distinguish any particular resource from the rest. Similar resources also need to be grouped or classified to emphasize the most important distinctions among the complete set of resources in the collection. A small neighborhood restaurant might have a short wine list with just ten wines, arranged in two categories for "red" and "white" and described only by the wine's name and price. In contrast, a gourmet restaurant might have hundreds of wines in its wine list, which would subdivide its "red" and "white" high-level categories into subcategories for country, region of origin, and grape varietal. The description for each wine might in addition include a specific vineyard from which the grapes were sourced, the vintage year, ratings of the wine, and tasting notes.

At some point a collection grows so large that it is not economically feasible for people to create bibliographic descriptions or to classify each separate resource. This leaves two approaches that can be done separately or in tandem. The simpler approach is to describe sets of resources or documents as a set or group, which is especially sensible for archives with its emphasis on the fonds (see Section 1.3.1). The second approach is to rely on automated and more general-purpose organizing technologies that organize resources through computational means. Search engines like Google are familiar examples of computational organizing technology, and section 8.5, "Computational Classification," describes other common techniques in machine learning, clustering, and discriminant analysis that can be used to create a system of categories and to assign resources to them.

Chapter 6, "Implementing Resource Descriptions," focuses on the representation and management of descriptions, taking a more technological or implementation perspective. Chapter 10, "Combining Descriptions," and Chapter 11, "Comparing Descriptions," discuss how the nature and extent of descriptions determines the capabilities of the processes that locate, compare, combine, or otherwise use them in information-intensive domains.

1.3.4 When is it Being Organized?

Because bibliographic description, when manually performed, is expensive, it seems likely that the "pre" organizing of information will continue to shift incrementally toward "post" organizing (Svenonius, p. 194-195).

When an author writes a document, he or she gives it some internal organization via title, section headings, typographic conventions, page numbers, and other mechanisms that identify its parts and their significance or relationship to each other. The document could also have some external organization implied by the context of its publication, like the name of its author and publisher, its web address if it is online or has a website, and citations or links to other documents or web pages. These explicit acts of organization by the author are often supplemented by additional organizing information supplied by the publisher or others, such as an ISBN or Library of Congress call number or subject headings. All of these pieces of descriptive data constitute metadata, or information about information.

Digital photos, videos, and documents are always organized to some minimal degree because descriptions are assigned automatically to them by the technology used to create them. At a minimum, these descriptions include the creation time and storage format for the resource, or chronologically by the auto-assigned filename (IMG00001.JPG, IMG00002.JPG, etc.), but often are much more detailed. For example, most digital cameras annotate each photo with metadata about the camera and its settings in the Exchangeable Image File Format (EXIF), and many mobile phones can associate their location along with any digital object they create.

The Organizing System framework recasts the traditional tradeoff between information organization and information retrieval as the decision about *when* the organization is imposed, "on the way in" or "on the way out." This contrast is easiest to see if we compare traditional libraries with Google and other search engines. Before a resource is made available to a user in a library, it is assigned a description according to a controlled vocabulary and embedded in standard classification systems. Not all resources receive the same amount of description and classification, but some amount of human "organizing intelligence" is applied to every resource in the library "on the way in" when it is added to the library's collection. And because libraries have always treated information about a reader's activities and choices as private, they would not alter the description or organization of resources to personalize them "on the way out."

On the other hand, Google applies massive computational power to analyze the contents and associated structures (like links between web pages) to impose organization on resources that have already been published or made available so that they can be retrieved in response to a user's query "on the way out." Google makes use of existing organization within and between information resources when it can, but its unparalleled technological capabilities and scale yield competitive advantage in imposing organization on information that wasn't previously organized digitally. Google has even been criticized for ignoring or undervaluing the descriptive metadata and classifications previously assigned by people and replacing them with computationally assigned descriptors (Nunberg, 2009). Google makes almost all of its money through personalized ad placement, so much of the selection and ranking of search results is determined "on the way out" in the fraction of a second after the user submits a query by using information about the user's search history and current context. Of course, this "on the way out" organization is only possible because of the more generic organization that Google's algorithms have imposed, but that only reminds us of how much the traditional distinction between "information retrieval" and "information organization" is no longer defensible.

However, it is an oversimplification to contrast "on the way in" organization performed by authors or librarians with "on the way out" organization performed by computers because the nature and extent of organization almost always changes over time as the resources governed by the Organizing System are used. The arrangement of resources in a kitchen or in an office changes incrementally as frequently used things end up in the front of the pantry, drawer, shelf or

filing cabinet or on the top of a pile of papers. Printed books or documents acquire margin notes, underlining, turned down pages or coffee cup stains that differentiate the most important or most frequently used parts. Digital documents don't take on coffee cup stains, but when they are edited, their new revision dates put them at the top of directory listings.

The scale of emergent organization of web sites, photos on Flickr, blog posts, and other information resources that can be accessed and used online dwarfs the incremental evolution of individual Organizing Systems. This organization is clearly visible in the pattern of links, tags, or ratings that are explicitly associated with these resources, but search engines and advertisers also exploit the less visible organization created over time by information about which resources were viewed and which links were followed.

This sort of organic or emergent change in Organizing Systems that takes place over time contrasts with the planned and systematic maintenance of Organizing Systems described as curation or governance. These two terms are roughly equivalent, but the former is most often used for libraries, museums, or archives and the latter for enterprise or inter-enterprise contexts. Curation and governance are activities whose goals are to preserve the investments in an Organizing System by anticipating and responding to changes in requirements and opportunities. The Organizing Systems for businesses and industries often change because of the development of de facto or de jure standards, or because of regulations, court decisions, or other events or mandates from entities with the authority to impose them. These topics are discussed in Chapter 3, "The Organizing System Lifecycle".

An important activity in collections of both physical resources and information resources is ensuring their persistence. Physical objects deteriorate over time, usually in a gradual or continuous way that can be managed by regular inspection and maintenance. Digital resources, in contrast, can be lost in a moment, making periodic backup and offsite storage essential to enable business continuity in the event of technology failure. But storage of the digital content alone is insufficient. To ensure meaningful future access, it might be necessary to preserve or continually upgrade any software or computers that implement the Organizing System. The rapid obsolescence of computing and digital storage technology has encouraged the migration of many Organizing Systems to implementation "in the cloud," but this simply shifts the technology concern to the cloud operator and introduces other concerns like data privacy.

A related concern to ensuring that the resources in a collection persist over time is ensuring that they do so in a way that preserves their authenticity. In Organizing Systems like museums and archives that preserve rare or culturally important objects or documents this concern is expressed as the principle of provenance. This is the history of the ownership of a collection or the resources in it. Since ancient Rome notaries have authenticated the creation of important documents, which then must be preserved with an unbroken "chain of custody" to demonstrate that the items have maintained their integrity or value as evidence. A uniquely Chinese technique in Organizing Systems is the imprinting of elaborate red seals on documents, books, and paintings that collectively record the provenance of ownership and the review and approval of the artifact by emperors or important officials.

1.3.5 How (or by Whom) is it Organized?

"The rise of the Internet is affecting the actual work of organizing information by shifting it from a relatively few professional indexers and catalogers to the populace at large. … An important question today is whether the bibliographic universe can be organized both intelligently (that is, to meet the traditional bibliographic objectives) and automatically" (Svenonius, p. 26).

In the preceding quote Svenonius identifies three different ways for the "work of organizing information" to be performed: by professional indexers and catalogers, by the populace at large, and by automated (computerized) processes. Because our notion of the Organizing System is broader than her "bibliographic universe," it is necessary to extend her taxonomy. In particular, we identify authors as a subset of the non-professional user population, and further distinguish users in informal and formal/institutional contexts.

Professional organizers undergo extensive training to learn the concepts, controlled descriptive vocabularies, and standard classifications in the particular domains in which they work. They can create and maintain Organizing Systems with consistent high quality, but "any task that requires an organizing intelligence to engage in research is costly" (Svenonius, 2000, p. 27). Expanding the scope of Organizing Systems beyond the bibliographic universe expands the class of professional organizers to include the employees of commercial information services like Westlaw and LexisNexis, who add controlled and, often, proprietary metadata to legal and government documents and other news sources. Scientists and scholars with deep expertise in a domain often function as the professional organizers for data collections, scholarly publications and proceedings, and other specialized information resources in their respective disciplines.

Authors are unlikely to be professional organizers, but presumably the author best understands why something was created and the purposes for which it can be used. To the extent that authors want to help others find a resource, they will assign descriptions or classifications that they expect will be useful to those users.

On the other hand, non-author users in the "populace at large" are most often creating organization for their own benefit. Not only are these ordinary users unlikely to use standard descriptors and classifications, the organization they impose sometimes so closely reflects their own perspective and goals that it isn't useful or accurate for others. Fortunately most users of "Web 2.0" or "community content" applications (O'Reilly, 2005) at least partly recognize that in these applications the organization of resources emerges from the aggregated contributions of all users, which provides incentive to use less egocentric descriptors and classifications. The staggering number of users and resources on the most popular applications inevitably leads to "tag convergence" simply because of the statistics of large sample sizes.

Finally, the vast size of the web and the even greater size of the deep or invisible web—composed of the information stores of business and proprietary information services (He, et al 2007)—makes it impossible to imagine today that it could be organized by anything other than the massive computational power of search engine providers like Google and Microsoft. Nevertheless, in the earliest days of the web, significant human effort was applied to organize it. Most notable is Yahoo!, founded by Jerry Yang and David Filo in 1994 as a directory of favorite web sites. For many years the Yahoo! homepage was the best way to find relevant websites by browsing the extensive system of classification. Today's Yahoo! homepage emphasizes a search engine that makes it appear more like Google or Microsoft Bing, but the Yahoo directory can still be found if you search for it.

WEB 2.0, ENTERPRISE 2.0, LIBRARY 2.0, MUSEUM 2.0, SCIENCE 2.0, GOV 2.0, …

The Web was invented as a publishing and document distribution medium, and later became a platform for business transactions. But after the bursting of the "dot com bubble" in 2000-2001 it was clear that moving a transactional business model to the web was not enough.

In 2004 a retrospective analysis by Tim O'Reilly and Dale Davenport of the web firms that had succeeded proposed the concept of "Web 2.0" for firms whose applications literally get better the more people use them because they "harness the collective intelligence" of their users.

Google, Amazon, eBay, Wikipedia, Facebook, Twitter, and YouTube are familiar examples today of web-based applications and services where value is based on aggregating, interpreting, and responding to enormous amounts of user-generated data and content. Web sites and resources that attract many visitors collect user interactions implicitly and also allow users to annotate, "tag," and evaluate them explicitly. These bottom-up and distributed activities have been called "folksonomies" (Trant, 2009b) and "crowdsourcing" (Howe, 2008).

Tagging, bookmarking, and rating mechanisms are increasingly being adapted for use inside companies as techniques for knowledge management, a trend named "Enterprise 2.0" to emphasize its similarity with "Web 2.0" while pointing out how it differs (McAfee, 2009). Because every user is authenticated to their real identities, and organizational norms and incentives restrict and shape the purposes and nature of user contributions, enterprise 2.0 applications have been successful at capturing expertise and institutional knowledge.

The core Web 2.0 design principle of empowering users to contribute information to help organize some collection of resources is rapidly being generalized to many other domains of Organizing Systems. Some libraries are now discussing how a "Library 2.0" (Maness, 2006) could provide personalized catalogs and information services and enable patrons to interact online with people of similar interests. Similarly, some museums, scientific repositories, and governments are conducting "open access" or "citizen participation" experiments by allowing users access to identify and annotate items, analyze raw data, or create "mashups" or applications that reuse and transform information that formerly was available only in summary form or in finished documents (Simon, 2011; Shneiderman 2008, Drapeau 2010; Robinson et al 2009)

1.4 Organizing this Book

Devising concepts, methods, and technologies for describing and organizing resources have been essential human activities for millennia, evolving both in response to human needs and to enable new ones. Organizing Systems enabled the development of civilization, from agriculture and commerce to government and warfare. Today Organizing Systems are embedded in every domain of purposeful activity, including research, education, law, medicine, business, science, institutional memory, sociocultural memory, governance, public accountability, as well as in the ordinary acts of daily living.

Many of the foundational topics for a discipline of organizing have traditionally been presented from the perspective of the public sector library and taught as "library and information science". These include bibliographic description, classification, naming, authority control, and information standards. We need to update and extend the coverage of these topics to include more private sector and non-bibliographic contexts, multi- and social media, and new information-intensive applications and service systems enabled by mobile, pervasive, and scientific computing. In so doing we can reframe the foundational concepts to make them equally compatible with the disciplinary perspectives of informatics, data and process modeling, and document engineering.

The concept of the Organizing System highlights the design dimensions and decisions that collectively determine the extent and nature of resource organization and the capabilities of the processes that compare, combine, transform and interact with the organized resources. This means that the distinction between "information organization" and "information retrieval" that is often manifested in academic disciplines and curricula is much less important. So this book has few sharp divisions between "information organization" (IO) and "information retrieval" (IR) topics. Instead, it will explain the key concepts and challenges in the design and deployment of Organizing Systems in a way that continuously emphasizes the relationships and tradeoffs between IO and IR.

Chapter 2, "Design Patterns for Organizing Systems". A view that brings together how we organize as individuals with how libraries, museums, governments, scientists, and businesses create Organizing Systems requires that we generalize the organizing concepts and methods from these different domains. Chapter 2 surveys a wide variety of Organizing Systems and describes four activities or functions shared by all of them: selecting resources, arranging them, providing access to them, and maintaining them over time. Chapter 2 also reviews technology trends that shape the design of Organizing Systems and proposes numerous patterns and dimensions that facilitate their analysis and comparison. These patterns provide a framework for innovation by suggesting how new types of Organizing Systems can be created by making a different set of design choices.

Chapter 3, "The Organizing System Lifecycle". Chapter 3 complements the descriptive perspective of Chapter 2 with a more prescriptive one that analyzes the design choices and tradeoffs that must be made in different phases in an Organizing System's lifecycle. System lifecycle models exhibit great variety, but we use a generic four-phase model that distinguishes a domain identification and scoping phase, a requirements phase, a design and implementation phase, and an operational phase.

Chapter 4, "Resources". The design of an Organizing System is most strongly shaped by what is being organized, the first of the five design decisions we introduced earlier in Section 1.3.1. To enable a broad perspective on this fundamental issue we use *resource* to refer to anything being organized, an abstraction that we can apply to physical things, information about physical things, information about non-physical things, or web-based services or objects. Chapter 4 discusses the challenges and methods for identifying the resources in an Organizing System in great detail and emphasizes how these decisions reflect the goals and interactions that must be supported. – the "why" design decisions introduced in Section 1.3.2

Chapter 5, "Describing Resources". The principles by which resources are organized and the kinds of access and interactions that can be supported for them largely depend on the nature and explicitness of the resource descriptions. This "how much description" design question was introduced in Section 1.3.3; Chapter 5 presents a systematic process for creating effective descriptions and analyzes how this general approach can be adapted for different types of Organizing Systems.

Chapter 6, "Implementing Resource Descriptions". Chapter 6 complements the conceptual and methodological perspective on the creation of resource descriptions with a technological and implementation one. Stored resource descriptions are often called *metadata*, literally "data about data" but the latter concept implies a narrower range of relationships between descriptions and the resources they describe than the former, which we prefer. Chapter 6 reviews the traditional approaches for describing resources from the library and information

science disciplines as well as the emerging perspectives from the semantic web, linked data, and microformat communities.

Chapter 7, "Categories". Groups or sets of resources with similar or identical descriptions can be treated as equivalent, making them members of an *equivalence class* or *category*. Identifying and using categories are essential human activities that take place automatically for perceptual categories like "red things" or "round things." Categorization is deeply ingrained in language and culture, and we use linguistic and cultural categories without realizing it, but categorization can also be a deeply analytic and cognitive process. Chapter 7 reviews theories of categorization from the point of view of how categories are used in Organizing Systems.

Chapter 8, "Classification". The terms *categorization* and *classification* are often used interchangeably but they are not the sane. **Classification** is applied categorization – the assignment of resources to a system of categories, called classes, using a predetermined set of principles. Chapter 8 describes three different approaches to classification: faceted, social/distributed, and computational. The chapter briefly introduces some of the most commonly used classification systems in libraries and museums as well as new computational approaches for classifying email messages as spam or classifying music by genre.

Chapter 9, "Relationships and Structures". <these were originally separate chapters, but it is difficult to draw a line between the two concepts. We have a good draft of "relationships" and a topical outline for "structures" but we need to revisit them as an integrated whole.>

Chapter 10, "Combining Descriptions". The same domain or type of resources can often have more than one Organizing System because different individuals, groups, firms, or institutions have different experiences, requirements, or capabilities that lead them to describe and arrange resources in different ways. When Organizing Systems overlap, intersect, or are combined (temporarily or permanently), these differences can make it difficult or impossible to locate resources, access them, or otherwise impair their use. Chapter 10 reviews some of the great variety of concepts and techniques that different domains use when combining descriptions – integration, interoperability, data mapping, crosswalks, mashups, and so on.

Chapter 11, "Comparing Descriptions". Processes for information retrieval and locating relevant resources in general are often characterized as comparing the description of a user's needs with descriptions of the resources that might satisfy them. Because description is a foundational concept in the Organizing System perspective, in Chapter 11 we extend and more broadly apply this core idea to describe IR and related applications of natural language processing (NLP) in terms of locating, comparing, and ranking descriptions. The chapter is not intended to replace more conventional treatments and texts on IR and NLP; its goal is to suggest how these applications fit into the broader context of the Organizing System.

Chapter 12, "Epilogue". <not sure yet… Christine's chapter>.

References

Anderson, J., and Perez-Carballo, J. (2001a). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing, *Information Processing & Management*, 37(2): 231-254.

Anderson, J., and Perez-Carballo, J. (2001b). The nature of indexing: how humans and machines analyze messages and texts for retrieval. Part II: Machine indexing, and the allocation of human versus machine effort. *Information Processing & Management*, 37(2): 255-277.

Bates, M. (2005). Information and knowledge: An evolutionary framework for information science. *Information Research*, 10(4):1-35.

Bates, M. (2006). Fundamental Forms of Information, *Journal of the American Society for Information Science*, 57(8):1033-1045.

Berners-Lee, T., Hendler, J., and Lassila, O. (2001). The Semantic Web. *Scientific American*, May 2001.

Brin, S. (2009). A Library to Last Forever. *New York Times*, 8 October 2009.

Buckland, M. (1991). Information as Thing. *Journal of the American Society for Information Science*, 42(5):351-360.

Buckland, M. (1997). What is a "Document"? *Journal of the American Society for Information Science*, 48(9):804-809.

Drapeau. M. (2010), The Three Phases of Government 2.0. O'Reilly Radar, 10 May 2010. http://radar.oreilly.com/2010/05/the-three-phases-of-government.html

Flickr. (2011). The App Garden. (http://www.flickr.com/services/)

Gilliland-Swetland, A. (2000). *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment.* Council on Library and Information Resources, Washington D.C.

He, B., Patel, M., Zhang, Z., and Chang, K. C. 2007. Accessing the deep web. *Communications of the ACM* 50(5): 94-101.

Howe, J. (2008). *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*. Crown Business.

IFLA (2011). Division of Library Types. http://www.ifla.org/en/library-types

Jones, E., and Janes, J. (2010) "Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read," *Policy & Internet*: Vol. 2(4): Article 3. (http://www.psocommons.org/policyandinternet/vol2/iss4/art3)

Maness, J. (2006). Library 2.0 Theory: Web 2.0 and its Implications for Libraries. *Webology*, 3(2), June 2006.

McAfee, A. (2009). *Enterprise 2.0: New Collaborative Tools for Your Organization's Toughest Challenges*. Harvard Business School Press.

Nunberg, G. (2009). Google's Book Search: A Disaster for Scholars. *The Chronicle Review*, (http://chronicle.com/article/Googles-Book-Search-A/48245/)

Nunberg, G. (2011). James Gleick's History of Information. *New York Times*, 18 March 2011.

Online Archive of California. (2011). Julia Morgan Collection, 1893-1980. (http://www.oac.cdlib.org/findaid/ark:/13030/tf7b69n9k9)

O'Reilly, T. (2005). What is Web 2.0? http://oreilly.com/web2/archive/what-is-web-20.html

Panizzi, A. (1841). Rules for the Compilation of the Catalogue.

Robinson, D., Yu, H., Zeller, W., and Felton, E. (2009). Government Data and the Invisible Hand. *Yale Journal of Law & Technology*, 11: 160-175.

Samuelson, P. (2011). Google Book Settlement as Copyright Reform, *Wisconsin Law Review*, 2011(2): 479-562.

Samuelson, P. (2009). Google Books is not a Library. *The Huffington Post*, 13 October 2009. (http://www.huffingtonpost.com/pamela-samuelson/google-books-is-not-a-lib_b_317518.html)

Shneiderman, B. (2008). Science 2.0. *Science* 319(5868):1349-50

Simon, N. (2011). *The Participatory Museum*. http://www.participatorymuseum.org/read/

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. The MIT Press.

Tillett, B. (2004). *What is FRBR? A Conceptual Model for the Bibliographic Universe*. Library of Congress Cataloging Distribution Service.

Trant, J. (2009a). Emerging convergence? Thoughts on museums, archives, libraries, and professional training. *Museum Management and Curatorship*, 24(4): 369-387.

Trant, J. (2009b). Studying social tagging and folksonomy: A review and framework. *Journal of Digital Information*, 10(1). (http://hdl.handle.net/10150/105375)

Wang, J. (2010). A Seed Library for Heirloom Plants Thrives in the Hudson Valley. *New York Times*, 6 October 2010.

Wright, A. (2010). Managing Scientific Inquiry in a Laboratory the Size of the Web. *New York Times*, 28 December 2010.