Chapter 3

Activities in Organizing Systems

Robert J. Glushko
Erik Wilde
Jess Hemerly
Isabelle Sperano
Robyn Perry

3.1 Introduction

Four activities occur in every organizing system: resource selection, organization, interaction design, and maintenance. While, these activities are iterative, and not entirely separable or sequential, it is helpful to introduce them individually:

Resource Selection - Determining the scope of the organizing system by specifying which resources to include.

Organization - Specifying the principles or rules that determine the arrangement of the resources.

Interaction Design - Designing and implementing the actions, functions or services that make use of the resources.

Maintenance - Managing and adapting the resources, the organization imposed on them, and the interactions with them as needed.

How explicit and formal these activities are depends on the breadth, variety, and scale of resources encompassed by the organizing system. To illustrate this, consider managing your wardrobe. While you select clothes, you are unlikely to write a formal policy of your preferences. You are unlikely to consciously prioritize the ways you expect to search for and locate articles of clothing. You may not explicitly consider the organizing principles you use to arrange them. And while you may put things back in order and discard items you no longer wear, you are unlikely to schedule this as a regular maintenance activity. Even if these activities are neither explicit nor formal, as in a personal wardrobe, they occur in all organizing systems.

In contrast, the organizing systems of institutions like libraries and businesses must be carefully managed. Consider a business looking to develop a data warehouse that will combine data from different sources such as orders, sales, customers, inventory, and finance. The design and operation of the data warehouse will involve the same four high-level activities:

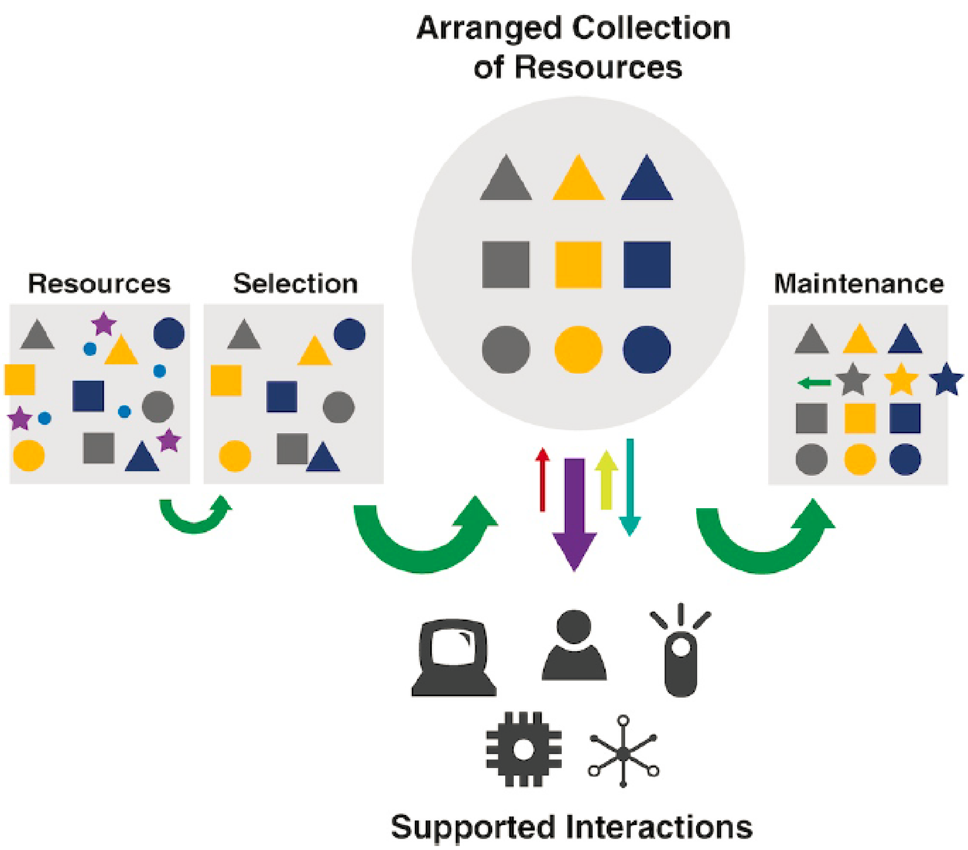
Resource Selection - What data sources should be included? How is their quality assessed? How much of the data is sampled? How are queries composed?

Organization - What data formats and schemas will enable effective processing? Are any necessary data transformations made at load time or query time?

Interaction Design - What are the most important and frequent queries that need to be pre-configured? Does there need to be a graphical user interface for generating queries?

Maintenance - What governance policies and procedures are needed to satisfy retention, compliance, security, and privacy requirements?

Figure 3.1. Four Activities in all Organizing Systems.



Four activities take place in all organizing systems: selection of resources for a collection; intentional organization of the resources; design and implementation of interactions with individual resources or with the collection, and; maintenance of the resources and the interactions over time.

These four activities are deeply ingrained in academic curricula and professional practices. Domain-specific methods and vocabularies have evolved to capture the complex and distinctive sets of experiences and methods of their respective disciplines. Adding a resource to a library collection is called acquisition, but adding to a museum collection is called accessioning. Documenting the contents of library and museum collections to organize them is called cataloging. In libraries, the central interaction is circulation, borrowing and returning the same resource, but in museums, which do not circulate their resources, the primary interactions for users are viewing or visiting the collection. Maintenance activities in libraries and museums are usually described as preservation or curation.

In business organizing systems, selection of resources could involve data generation, capture, sampling, or extraction. Adding resources could involve loading, integration, or insertion. Schema development and data transformation are likely to be important organizing activities. Supported interactions could include querying, reporting, analysis, or visualization. Maintenance activities include deletion, purging, governance, or compliance.

It is important to recognize that the purpose of using the general terms of selection, organization, interaction design, and maintenance is not to replace more specialized ones. Rather, these terms can facilitate communication and cooperation across different professions and domains that are concerned with organizing. Additionally, this broader perspective helps us identify best practices and patterns across different organizing systems even if it might sound odd to think that the admissions process is how a university selects its student resources, or that deleting information to comply with privacy regulations is maintenance.

3.2 Resource Selection

The most fundamental decision for an organizing system is determining the resource domain: the group or type of resources that are being organized. *Selection* is the process by which resources are identified, evaluated, and then added to a collection in an organizing system. Selection is first shaped by the domain and then by the scope of the organizing system. The scope can be analyzed through six interrelated aspects:

1. the size of the collection,
2. the number and nature of the users,
3. the expected changes to the collection,
4. the lifetime over which the organizing system is expected to operate,
5. the physical or technological environment in which the organizing system is situated or implemented, and
6. the relationship of the organizing system to others that overlap with it in domain or scope.

3.2.1. Selection Criteria

Given the broad definition of resource as "anything of value that can support goal-oriented activity," the selection of resources must involve an assessment to determine whether the resource enables a person or nonhuman agent to perform the intended activities. Selection methods and criteria vary for different types of resources, but their common purpose is to determine this "fitness for use." This highlights the need to have activities in mind before resources are selected to enable them. It also explains why precise selection criteria are harder to define for organizing systems that have diverse sets of users with different goals such as libraries. Beyond "fitness for use," some selection specifications are recorded in laws, regulations or policies that require or prohibit the collection of certain kinds of objects or types of information.

When the resources being considered for a collection are homogeneous and predictable, it is possible to treat them as a class or set. In this situation, selection criteria and organizing principles can be specified in advance, and selection and organizing then become concurrent activities. When the resources are created by automated processes or data entry in business systems, the principles that will be used to organize them can serve as specifications that control resource creation. We can think of database or document schemas (at the implementation tier), or data entry forms or word processor templates (at the user interface tier) as embodiments of the organizing principles of the data records or documents that are then created in conformance with them.

For some types of resources, the specifications that guide selection can be precise and measurable. When a manufacturer of physical goods selects the materials or components that are transformed into its products, it carefully evaluates the candidate resources and their suppliers before making them part of its supply chain. Similarly, employee selection has become highly data-intensive. Employers hire people after assessing the match between their competencies and capabilities and what is needed to do the required activities.

In principle, scientific and business data is selected after its quality and relevance has been assessed. However, this is easy to say but hard to do given the time and monetary costs associated with rigorous quality assurance. That said, it is essential to check individual data items to find data entry problems such as misspellings and duplicate records, or data values that are illegal, statistical outliers, or otherwise suspicious. It is also essential to evaluate the processes by which the data was collected, stored, and managed to ensure it has not been corrupted or otherwise modified. Other important concerns are the format and precision in which the data is stored, and whether the schema governing each instance is sufficiently rigorous.

Even when selection criteria can be measured and evaluated in isolation, they are often incompatible or difficult to satisfy in combination. It would be desirable for data to be timely, accurate, complete, and consistent, but these criteria often trade off against one other. Prioritizing one of these properties can make it impossible to find resources that satisfy other priorities.

Sequential selection decisions can exhibit “path dependence” because the order in which you evaluate candidate resources affects the selections. A familiar human resource selection activity with profound path dependence is the “draft” of college athletes by professional sports teams. The decision that a team makes to select a player in some round of the draft is influenced by the players already on the team, selections made by the team in prior rounds, selections made by other teams, and many other factors.

When your goal is to make inferences about a large population without having to study all of its instances, the selection criteria is random sampling. A good sample for statistical purposes is one in which the selected resources are not different in any significant way from those that were not selected. Sampling is useful when timely decisions are required as well as when large numbers of resources need to be verified to satisfy functional requirements. For example, a manufacturer cannot test every part arriving at the factory, but might randomly test some of them from different shipments.

3.2.2. Looking "Upstream" and "Downstream" to Select Resources

If you can determine where resources come from, you can make better selection decisions by evaluating the people, processes, and organizing systems that create them. Using the analogy of a river, we can follow a resource “upstream” from us until we find the “headwaters.” Physical resources might have their headwaters in a factory, farm, or artist’s studio. Digital resources might have headwaters in a government agency, a scientist’s laboratory, or a website.

Even though finding the headwaters where resources come from is often not easy and sometimes not possible, that is where you are most likely to locate the people best able to answer the questions, described in Chapter 2, that define any organizing system. The resource creators or producers will know the assumptions and tradeoffs they made that influence whether the resources will satisfy your requirements. You should also try to evaluate the processes or algorithms that produce the resources, and then decide if they are capable of yielding resources of acceptable quality.

The best outcome is to find a credible supplier of quality resources. However, if an otherwise desirable supplier does not currently produce resources of sufficient quality, it might be possible to improve the quality by changing the process using instruction or incentives. A clear lesson from the “quality movement” and statistical process control is that investments that fix quality problems at their source are more cost-effective than repeated work to fix preventable problems.

When you cannot obtain resources directly from their source, even if you are confident of their quality, it is important to analyze any evidence or records of their use or interactions as they flow downstream. Physical resources are often paired with printed or digital documents that provide evidence about their origin and authenticity. Barcodes, RFID tags, or other technological mechanisms enable resources to be tracked from their headwaters to where they are used. Tracking is essential for data resources because they can often be added to, derived from, or otherwise changed without leaving visible traces. Just as the water from melted mountain snow becomes less pure as it flows downstream, a data resource can become "dirty" or "noisy" over time. Data often gets dirty when it is merged with other datasets that contain duplicate or seemingly-duplicate information. Data can also become dirty when the hardware or software that stores it changes. A data resource can become inaccurate or obsolete because the world that the data describes has changed with the passage of time.

Other selection processes look "downstream" to select resources using predicted rather than current properties, capability, or suitability. Sports teams often sign promising athletes for their minor league teams. Businesses hire interns, train their employees, and run executive development programs to prepare high-performing low-level managers for executive roles.

3.3 Organization

Organizing systems arrange resources according to many different principles. The choice of organizing principles for a collection of resources depends on the answers to the "why?", "how much?", and "how?" questions posed in Chapter 2. When these answers follow common patterns, the organizing system they create might have a familiar name like "library" or "museum."

The simplest principle for organizing resources is *colocation*—putting all the resources of some type in the same location. However, most organizing systems use principles that are based on specific resource properties or properties derived from the collection as a whole. What properties are significant and how to think about them depends on the number of resources being organized, the purposes for which they are being organized, and on the experiences and biases of the intended users of the organizing system.

Some generic organizing principles like alphabetic or chronological order can apply to almost every type of resource because most resources have names or dates associated with them. Other organizing principles are more specific because they are based on the properties that distinguish one type of resource from another. For example, the periodic table of the elements organizes them into groups and periods based on properties that emerge from their atomic structure.

Property

In this book, we use *property* in a generic and ordinary sense as a synonym for "feature" or "characteristic." Many cognitive and computer scientists are more precise in defining these terms and reserve property for binary predicates (e.g., something is red or not, round or not). If multiple values are possible, the property is called an "attribute," "dimension," or "variable." *Feature* is used in data science and machine learning contexts for both "raw" or observable variables and "latent" ones, extracted or constructed from the original set.

"Subject matter" organization involves the use of a classification system that provides categories and descriptive terms for indicating what a resource is about. The methods for assigning subject classifications are intellectually-intensive as they use derived properties that cannot directly perceived. Many people undertake rigorous training to learn to classify consistently and appropriately.

In libraries, museums, businesses, government agencies and other long-lived institutions, organizing principles are typically documented as cataloging rules, information management policies, or other explicit and systematic procedures so that different people can apply them consistently over time. In contrast, the principles for arranging resources in personal or small-scale organizing systems are usually informal.

The effort required for people to organize large collections of resources is often the motivation for automatically capturing resource properties and applying computational techniques such as machine learning for organizing the resources into categories. However, with such techniques, it is easy to collect an enormous number of properties or features for each resource and, as a result of this complexity, the organizing principles discovered by the computations can be spurious or uninterpretable by people.

3.3.1. Organizing Physical Resources

When the resources being arranged are physical or tangible things any resource can be in only one place at a time. Similarly, when organizing involves recording information in a physical medium, whether in a handwritten note or a book, how this information can be organized is restricted by the intrinsic properties of physical things. The inescapable tangibility of physical resources means that their organizing systems are often strongly influenced by the material or medium of the resources. We usually organize our books in a different location than our CDs and DVDs because they have different material manifestations. Of course, the storage environments for physical resources (shelves, cabinets, closets, and so on) have co-evolved with the physical resources they store.

3.3.1.1. Organizing with Properties of Physical Resources

Physical resources are often organized according to intrinsic physical properties like their size, color, or shape because the human visual system automatically pays a lot of attention to these properties. This "pre-attentive" organizing explains why arranging physical resources using quickly perceived attributes can seem more aesthetic or satisfying than organizing them using properties that take more time to understand.

Physical resources are also commonly organized using intrinsically associated properties such as the place and time they were created or discovered. You might arrange the shirts in your clothes closet by color, by fabric, or style. In addition to, or instead of, these physical properties, you might employ behavioral or usage-based properties to arrange them such as summer and winter or casual and business clothes. Unlike intrinsic properties of resources, which do not change, behavioral or usage-based properties are dynamic: moving to a warmer or colder climate or switching jobs to a more or less formal company could change these properties.

3.3.1.2. Organizing with Descriptions of Physical Resources

To overcome the inherent constraints of organizing physical resources, organizing systems often use resource descriptions of the primary physical ones. The classic example is the library card catalog which consists of printed cards that describe the books. While a specific physical resource might only be in one place, copies of the corresponding resource description can be in many different places at the same time. When the resource descriptions are digital, as when a printed library card catalog is put online, the additional layer of abstraction created enables organizing possibilities that can ignore physical properties of resources and many of the details about how they are stored. For example, a book can be listed in many bibliographies at the same time.

3.3.2. Organizing Places

In the previous two subsections, we treat the place or environment in which we organize resources as given. An alternative perspective assumes that we design the physical environment. These environments could be:

- The land itself, as when we lay out city plans when organizing how people live together and interact in cities.
- A "built environment," a human-made space, particular building, or a set of connected spaces and buildings. A built environment could be a museum, airport, hospital, casino, department store, farm, road system, or any building or space where resources are arranged and people interact with them.

These are not entirely separable contexts, but they are easier to discuss separately.

3.3.2.1. Organizing the Land

Cities naturally emerge in places that can support life and commerce. Almost all major cities are built on coasts or rivers because water enables agriculture, transportation, and power for industrial development. Many ancient cities have crowded and convoluted street plans that do not seem intentionally organized, but some cities in the Middle East were laid out in rough grids over twenty-five hundred years ago.

Cities in the United States, and especially those in the western states, are much younger than most cities in Europe and Asia. The natural human tendency to impose order on living spaces had ample room to experiment with city designs in the US. The easiest and most efficient way to organize space is using a coordinate grid, with streets intersecting at perpendicular angles. Salt Lake City, Albuquerque, Phoenix, and Seattle are notable examples of grid cities. Washington DC has a hybrid structure, with radiating diagonal avenues overlaid on a grid.

3.3.2.2. Organizing Built Environments

Built environments influence the expectations, behaviors, and experiences of everyone who enters the space. Employees, visitors, customers, and inhabitants are all subject to the design of the spaces they occupy. These environments can be designed to encourage or discourage interactions between people, to create a sense of freedom or confinement, to reward exploration, or enforce efficiency in the interactions with resources. Sometimes the built environment is designed with a specific collection of resources in mind to enable and reinforce specific interaction goals or policies.

Additionally, most built environments include additional features or descriptions to assist people in wayfinding and orientation. Distinctive architectural elements can create landmarks, and spaces can be differentiated with color, lighting, furnishings, or other means. Signs, room numbers, or directional arrows highlight the way and distance to popular destinations.

We can contrast the built environments of museums, airports, and casinos to see how each of them facilitates or constrains interactions. Many old art museums mimic classical architecture, with grand stairs flanked by tall columns. They have large and dramatic entry halls that invite visitors inside. Modern museums are decidedly less traditional, and some people complain that the architecture of modern art museums can overshadow the art collection because people are induced to pay more attention to the building than to its contents.

Airport design is usually more concerned with efficiency, walkability (maybe with the aid of moving walkways), navigability, and comfort for travelers getting in and out the airport. Wide walkways, multiple staircases, and "people movers" whose doors open in one direction at a time, all encourage people to move in certain directions. This organization is supplemented with many orientation signs and display screens that help passengers find their departure gates, baggage, or ground transportation services.

In contrast, the design goals for casinos are keeping people inside and making it easy for them to lose track of time, so casinos usually have no windows or clocks. Casinos often arrange gambling machines and tables in ways that encourage people to think of them as playgrounds. Additionally, many casinos do not provide orientation and navigation aids because increased confusion makes visitors stay longer and gamble more.

If one accepts the premise that values and bias are at work in decisions about organizing systems, it is difficult not to see it in built environments. Consider queue design in banks, supermarkets, or when boarding airplanes. Assuming that it is desirable to organize people efficiently to minimize wait times and crowding, how should the queue be designed? How many categories of people should there be? What is the basis for the categories?

It may be uncontroversial to include several express lanes in a supermarket checkout because people can choose to buy fewer items if they do not want to wait. Similarly, it seems essential for hospital emergency rooms to have a triage policy that selects patients from the emergency room queue based on their likely benefit from immediate medical attention.

3.3.3. Organizing Digital Resources

Organizing systems that arrange digital resources like digital documents or information services have some intrinsic differences from those that organize physical resources. Digital resources are free from the "one place at a time" limitation because they can be easily copied or interlinked.

An organizing system for digital resources can also use digital description resources that are associated with them. Since the incremental costs of adding processing and storage capacity to digital organizing systems are small, collections of both primary digital resources and description resources can be arbitrarily large. Digital organizing systems can reach a scale that is impossible in organizing systems that are entirely physical, and they can implement services and functions that exploit the exponentially growing processing, storage and communication capabilities available today.

Information resources in either physical or digital form are typically organized using intrinsic properties like author names, creation dates, publisher, or the set of words that they contain. Information resources can also be organized using assigned properties like subject classifications, names, or identifiers. Information resources can also be organized using behavioral or transactional properties collected about individuals or groups of people with similar interaction histories. For example, Amazon and Netflix use browsing and purchasing behavior to make book and movie recommendations.

Just as many laws and regulations restrict the organization of physical resources, many laws and regulations constrain the arrangements of digital ones. Many information systems that generate or collect transactional data are prohibited from sharing any records that identify specific people. Banking, accounting, and legal organizing systems must follow compliance and reporting standards and rules. Similarly, increased regulation of digital resources has made their physical storage location relevant again.

3.3.3.1. Organizing Web-based Resources

The Domain Name System (DNS) is the foundation for organizing web resources. Top-level domains for countries (.us, .jp, .cn) and generic resource categories (.com, .edu. .org, .gov) provide some clues about the resources organized by a website. These clues are most reliable for large established enterprises and publishers; we know what to expect at ibm.com, berkeley.edu, and un.org.

The distributed nature of the web, combined rich hyperlinking, and the fact that the actual storage location of web resources is often unimportant to the end users fundamentally undermine the idea that organizing systems must collect resources and then arrange them under local control to be effective.

3.3.3.2. "Information Architecture" and Organizing Systems

The design field known as information architecture is a specialized approach for designing the information models and their implementations in user interfaces on websites and other information-intensive organizing systems. Abstract patterns of information content and organization are called architectures, so it is straightforward to define the activity of information architecture from the perspective of the discipline of organizing: *Information architecture* is designing an abstract and effective organization of information and then exposing that organization to facilitate navigation and information use. The first part of this definition refers to the intentional arrangement of resources, and the second to the interactions enabled by that arrangement.

Our definition of information architecture implies a methodology for the design of user interfaces and interactions that puts conceptual modeling at the foundation. Best practices in information architecture emphasize the use of systematic principles or design patterns for organizing the resources and interactions in user interfaces. Information architects then translate this logical design into a graphical design and arrange windows, panes, menus, and other user interface components so as to make interactions easy and enjoyable. The logical and graphical organization of a user interface influence how people interact with it and the actions they take or do not take.

Some information design conventions have become design patterns. Documents use headings, boxes, white space, and horizontal rules to organize information by type and category. Large type signifies more important content than small type, red type indicates an advisory or warning, and italics or bold says "pay attention." Some patterns are general and apply to an entire website, page, or interface genre such as an e-commerce site, blog, social network site, home page, or "about us" page. Other patterns are more specific and affect a part of a site or a single component of a page such as an auto-completion of a text field, breadcrumb menu, or slideshow. All design patterns reflect and reinforce a user's past experiences with content and interface components, and this familiarity reduces the cognitive complexity of interaction and allows users to pay less attention.

Interface designers can take advantage of this familiarity and employ design patterns to manipulate users, control their behaviors, or trick them into taking actions they do not intend. For example, a website may exploit familiar patterns to induce users to click on an ad disguised as a news item, sign up for unwanted e-mails, disclose personal information, or ignore critical terms and conditions because they are buried in tiny text or unusual locations.

Many organizing systems need to support interactions to find, identify, and select resources. Some of these systems contain both physical and digital resources, as in a bookstore with both web and physical channels, and many interactions are implemented across more than one device. Both the cross-channel and multiple-device situations create user expectations that interactions will be consistent across these different contexts. Starting with a conceptual model and separating content and structure from presentation gives organizing systems more implementation alternatives and makes them more robust in the face of technology diversity and change.

3.3.5. Organizing People

Organizing the people who work in a business is often called "human resource management." People are usually organized according to their roles in business processes and strategy. There are myriad ways to organize people that differ in the extent of hierarchical structure, the flow of information up and down the hierarchy, the span of control for managers, and the discretion people have while they perform their assigned work. Like many other professions, human resources is rapidly being transformed into a more data-driven discipline, and some firms use "people analytics" to organize employees to make them more innovative and productive.

3.3.6. Organizing with Multiple Resource Properties

Multiple properties of the resources, the person organizing or intending to use them, and the social and technological environment in which they are organized can collectively shape their organization. For example, the way you organize your home kitchen is influenced by the physical layout of counters, cabinets, and drawers. Other factors are your skills and preferences as a cook, which may affect the number of cookbooks, specialized appliances and tools you own, and the dishes you cook most often. The sizes and shapes of dishes, pots and pans, and food packages also shape how you arrange them.

If multiple resource properties are considered in a fixed order, the resulting arrangement forms a *logical hierarchy*. The top-level categories of resources are created based on the values of the property evaluated first. Each category is further subdivided using other properties until each resource is classified in only a single category. Consider the hierarchical system of folders used by a professor to arrange the digital resources on his computer; the first level distinguishes personal documents from work-related documents; work is then subdivided into teaching and research, teaching is subdivided by year, and year divided by course.

An alternative to hierarchical organization is *faceted classification*, in which the different properties of the resources can be evaluated in any order. For example, you can select wines from the wine.com store catalog by type of grape, cost, or region and consider these property facets in any order. Three people might each end up choosing the same moderately-priced Kendall Jackson California Chardonnay, but one of them might have started the search based on price, one based on the grape varietal, and the third with the region. This kind of interaction in effect generates a different logical hierarchy for every different combination of property values, and each user made his final selection from a different set of wines.

Faceted classification allows a collection of description resources to be dynamically re-organized into as many categories as there are combinations of values on the descriptive facets, depending on the priority or point of view the user applies to the facets. Of course, this only works because the physical resources are not themselves being rearranged, only their digital descriptions.

3.4 Interaction Design

With physical resources, it is easy to distinguish the interactions that are designed into and directly supported by an organizing system because of intentional acts of description or arrangement from those that can take place with resources after they have been accessed. For example, when a book is checked out of a library it might be read, translated, summarized, criticized, or otherwise used—but none of these interactions would be considered a capability of the book that had been designed into the library. In contrast, in organizing systems that contain digital resources the logical boundary between the resources and their interactions is less clear because what you can do with a digital resource is often not apparent.

3.4.1. Affordance and Capability

The concept of *affordance*, introduced by J. J. Gibson, then extended and popularized by Donald Norman, captures the idea that physical resources and their environments have inherent actionable properties that determine, in conjunction with an actor's capabilities and cognition, what can be done with the resource.

When the process of organizing resources involves arranging them using boxes, bins, cabinets, or shelves, the affordances and the implications for access and use can be easily perceived. Resources of a certain size and weight can be picked up and carried away. Books on the lower shelves of bookcases are easy to reach, but those stored ten feet from the ground cannot be easily accessed.

We can analyze the organizing systems with physical resources to identify the affordances and the possible interactions they imply. We can compare the affordances or overall interaction *capability* enabled by different organizing systems for physical resources, and we often do this without thinking about it. The tradeoffs between the amount of work that goes into organizing a collection of resources and the amount of work required to find and use them are inescapable when the resources are physical objects, or information resources that are in physical form. For example, books allow for random access as they can be opened to a particular page, but scrolls do not afford the random access capability.

What and how to count when comparing the capabilities of organizing systems becomes more challenging the further we get from collections of static physical resources. With computers, information systems, and digital resources in general, considerations about affordances and capabilities are not as straightforward.

First, the affordances we can perceive might not be tied to any useful interaction. Every computer screen that you can reach affords touching, but not every screen is touch-sensitive. Second, most of the interactions that are supported by digital resources are not apparent when you first encounter them. You may know from experience what interactions are possible with ".doc" and ".pdf" files, but you probably do not know what interactions are possible with ".xpi" and ".mobi" files.

The *capability* of digital resources and information systems can roughly be assessed by counting the number of associated functions, services, or application program interfaces. However, this very coarse measure does not take into account differences in the capability or generality of a particular interaction. For example, two organizing systems might both have a search function, but differences in the operators they allow, the sophistication of pre-processing of the content to create index terms, or their usability can make them vastly differ in power, precision, and effectiveness.

We should not assume that supporting more types of interactions necessarily makes a system better or more capable; what matters is how much value is created or invoked in each interaction. A smartphone cluttered with features and apps you never use enables a great many interactions but most add little value.

3.4.2. Interaction and Value Creation

A useful way to distinguish types of interactions is according to the way in which they create value. Apte and Mason propose that interactions can create value through physical manipulation, interpersonal or empathetic contact, and symbolic manipulation or information exchange. The proportions of these three types of value-creating activities can be treated as design parameters. For example, a person-to-person interaction can be transformed into a self-service one by replacing a human service provider like a bank teller with technology like an ATM.

3.4.2.1. Value Creation with Physical Resources

Physical manipulation is often the intrinsic type of interaction with collections of physical resources. The resource might have to be handled or directly perceived to interact with it, and often the experience of interacting with the resource is satisfying or entertaining, making it a goal in its own right. People visit museums, galleries, zoos, or other institutions that contain physical resources because they value the direct, perceptual, or otherwise unmediated interaction that these organizing systems support.

While physical manipulation and interpersonal contact might be required to interact with information resources in physical form such as books in a library, for most people the primary purpose of interacting with a library is to access the information contained in its resources. As books and other documents have become digitized, these incidental physical and interpersonal interactions are no longer needed to access the information contained within them.

In some organizing systems robotic devices, computational processes, or other entities that can act autonomously with no need for a human agent carry out interactions with physical resources. Robots have profoundly increased efficiency in materials management, “picking and packing” in warehouse fulfillment, office mail delivery, and in many other domains where human agents once located, retrieved, and delivered physical resources.

3.4.2.2. Value Creation with Digital Resources

With digital resources, neither physical manipulation nor interpersonal contact is required to exchange information or symbolically manipulate the information contained in the resource. By replacing interactions that involve people and physical resources with symbolic ones, organizing systems can lower costs without reducing user satisfaction. As such, many businesses have automated their information-intensive processes with self-service technology.

Similarly, web search engines eliminate the physical effort required to visit a library and enable users to consult more readily accessible digital resources. A search engine returns a list of the page titles of resources that can be directly accessed with just a click, so it takes little effort to go from the query results to the primary resource.

Museums have aggressively embraced the web to provide access to their collections. While few museum visitors would prefer viewing a digital image over experiencing an original painting, sculpture, or other physical artifact, the alternative is often no access at all. Most museum collections are far larger than the space available to display them, so the web makes it possible to provide access to otherwise hidden resources.

3.4.2.3. Accessibility

The United Nations Convention on the Rights of Persons with Disabilities recognizes accessibility to information and communications technologies as a basic human right. There is also a strong business case for accessibility: studies show that accessible websites are used more often, are easier to maintain, and produce better search results.

Many of the techniques for making a resource accessible involve transforming the resource or its description into a different form so someone who could not perceive it or interact with it in its original form can now do so. The most common operating systems all come with general-purpose accessibility features such as reading text aloud, recognizing speech, magnifying text, increasing cursor size, signaling with flashing lights instead of with sounds, lights to signal keyboard shortcuts for selecting and navigating, and connecting to devices for displaying Braille.

Other techniques are not generic and automatic, and instead require investment by authors or designers to make information accessible. Websites are more accessible when images or other non-text content types have straightforward titles, captions, and “alt text” that describes what they are about. Consistent placement and appearance of navigation controls and interaction widgets is helpful; for example, in a shopping site “My Cart” might always be found at the top right corner of the page.

3.4.3. Access Policies

Different levels of access can apply to different resources and interactions in a collection or to different categories of users. For example, library collections can range from completely open and public, to allowing limited access, to wholly private and restricted. Organizing systems in business domains are more likely to enforce a granular level of access control that distinguishes people according to their roles and the nature of their interactions with resources.

Many of the organizing systems used by individuals are embedded in physical contexts where access controls are applied coarsely. We need a key to get into the house, but we do not need additional permissions or passwords to enter our closets or to take a book from a bookshelf. In our online lives, however, we readily accept and impose more granular access controls. For example, we might allow or block individual "friend" requests on Facebook or mark photos on Flickr as public, private, or viewable only by named groups or individuals.

3.5 Maintenance

Maintenance includes the work to preserve resources, the processes for reevaluating and revising selection criteria, and the removal of unneeded resources from the system. Maintenance and selection are interdependent, and more stringent rules for selecting resources imply a maintenance plan that carefully enforces the same constraints that limit selection. Ideally, maintenance requirements for resources should be anticipated when organizing principles are defined and implemented.

Different domains sometimes use the same terms to describe different maintenance activities and different terms for similar activities. Common maintenance activities are *storage*, *preservation*, *curation*, and *governance*. Storage is most often used when referring to physical or technological aspects of maintaining resources; backup (for short-term storage), archiving (for long-term storage), and migration (moving stored resources from one storage device to another) are similar in this respect. The other three terms refer to activities or methods that more closely overlap in meaning, and we will distinguish them in §3.5.2 through §3.5.4.

3.5.1. Motivations for Maintaining Resources

The concept of *memory institution* broadly applies to a great many organizing systems that share the goal of preserving knowledge and cultural heritage. The primary resources in libraries, museums, data archives or other *memory institutions* are fixed cultural, historic, or scientific artifacts that are maintained because they are unique and original items with future value.

In contrast, in businesses organizing systems, many of the resources that are collected and managed have limited intrinsic value. The motivation for preservation and maintenance is economic; resources are maintained because they are essential in running the business. For example, businesses collect and preserve information about employees, inventory, orders, invoices, etc., because it ensures internal goals of efficiency, revenue generation, and competitive advantage. The same resources (e.g., customer information) are often used by more than one part of the business. Maintaining the accuracy and consistency of changing resources is a major challenge in business organizing systems.

Many business organizing systems preserve information needed to satisfy externally imposed regulatory or compliance policies and serve largely to avoid possible catastrophic costs from penalties and lawsuits. In all these cases, resources are maintained as one of the means employed to preserve the business as an ongoing enterprise, not as an end in itself.

Indefinite preservation is not the central goal of most business organizing systems. These organizing systems mostly manage information needed to carry out day-to-day operations or relatively recent historical information used in decision support and strategic planning. In addition to these internal mandates, businesses have to conform to securities, taxation, and compliance regulations that impose requirements for long-term information preservation.

3.5.2. Preservation

At the most basic level, preservation of resources means maintaining them in conditions that protect them from deterioration. Libraries, museums, and archives aim for stable temperatures and low humidity as they are focused on preventing physical damage to resources. Risk-aware businesses create continuity plans that involve redundant storage of the data and documents needed to stay in business in the event of a natural disaster or other disruption.

3.5.2.1. Digitization and Preserving Resources

Preservation is often a key motive for digitization, but digitization alone is not preservation. Digitization creates preservation challenges because technological obsolescence of computer software and hardware require ongoing efforts to ensure digitized resources can be accessed.

Technological obsolescence is the major challenge in maintaining digital resources. The most visible one is a result of the relentless evolution of the physical media and environments used to store digital information. Computer data began to be stored on magnetic tape and hard disk drives six decades ago, on floppy disks four decades ago, on CDs three decades ago, on DVDs two decades ago, on solid-state drives half a decade ago, and in "cloud-based" or "virtual" storage environments in the last decade.

The second challenge might seem paradoxical. Even though digital storage capacity increases at a staggering pace, the expected useful lifetimes of the physical storage media are measured in years or, at best, in decades. Colloquial terms for this problem are *data rot* or "bit rot." In contrast, books printed on acid-free paper can last for centuries. The contrast is striking; books on library shelves do not disappear if no one uses them, but digital data can be lost if no one wants access to it within a year or two after its creation.

However, these physical lifetime limits are less significant than a third challenge: the software and its associated computing environment used to parse and interpret the resource at the time of preservation might no longer be available when the resource needs to be accessed. Software and services that convert documents from old formats to new ones are widely available, but they are only useful if the old file can be read from its legacy storage medium.

It is easy to say that the solutions to the problems of digital preservation are regular recopying of the digital resources onto new storage media and then migrating them to new formats when significantly better ones come along. In practice, however, how libraries, businesses, government agencies or other enterprises deal with these problems depends on their budgets and their technical sophistication.

3.5.2.2. Preserving the Web

Preservation of web resources is inherently problematic. Unlike libraries, museums, archives, and many other kinds of organizing systems that contain collections of unchanging resources, organizing systems on the web often contain resources that are highly dynamic. Websites change by adding, editing, and removing content.

Web search engines like Google and Bing use crawlers to continually update their indexed collections of web pages, and their search results link to the current version, so preservation of older versions is explicitly not a goal. Furthermore, search engines do not reveal any details about how frequently they update their collections of indexed pages.

The Internet Archive (Archive.org), founded by Brewster Kahle, makes preservation of the web its first and foremost activity. When you enter a URI into its "Wayback Machine" you can see what a site looked like at different moments in time.

3.5.2.3. Preserving Resource Instances

A focus on preserving particular resource instances is most apparent in museums and archives, where collections typically consist of unique and original items. There are many copies and derivative works of the Mona Lisa, but if the original Mona Lisa was destroyed, none would be acceptable as a replacement.

Most business organizing systems, especially those that “run the business” by supporting day-to-day operations, are designed to preserve instances. These include systems for order management, customer relationship management, inventory management, digital asset management, record management, email archiving, and more general-purpose document management. In all of these domains, it is often necessary to retrieve specific information resources to serve customers or to meet compliance or traceability goals.

Recent developments in sensor technology enable extensive data collection about the state and performance of machines, engines, equipment, and other types of physical resources, including human ones. When combined with historical information about maintenance activity, predictive analytics techniques can use this data to determine normal operating ranges and indicators of coming performance degradation or failures. Predictive maintenance can maximize resource lifetimes while minimizing maintenance and inventory costs.

3.5.2.4. Preserving Resource Types and Classes

In contrast to museums, libraries, especially public ones, tend to focus on preserving resource classes rather than individual instances. As libraries are tasked with making their resources available to patrons, most have multiple copies of popular resources. Additionally, the bulk of any given collection is made up of books that have many equivalent copies in other collections. When a library has a copy of *Moby Dick*, it is preserving the abstract *work* rather than the particular physical *instance*—unless the copy of *Moby Dick* is a rare first edition signed by Melville.

While business organizing systems are designed to preserve instances, they are also designed around resource types and classes. These abstractions allow the organizing system to be more efficient while still preserving every specific instance. For example, a database can store a large number of orders because it treats them all the same, but it can still identify any individual instance.

3.5.2.5. Preserving Resource Collections

In some organizing systems, any specific resource might be of little interest or importance in its own right but is valuable because of its membership in a collection of essentially identical items. This is the situation in the data warehouses used by businesses to identify trends in customer or transaction data or in the huge data collections created by scientists. These collections are typically analyzed as complete sets. A scientist does not borrow a single data point when she accesses a data collection; she borrows the complete dataset consisting of millions or billions of data points. This requirement raises difficult questions about what additional software or equipment need to be preserved in an organizing system along with the data to ensure that it can be reanalyzed.

3.5.3. Curation

Curation refers to the processes by which a resource in a collection is maintained over time, and it occurs in all organizing systems. This may include actions such as improving access, or restoring or transforming its representation or presentation. Regardless of the specific action, the key to curation is having clear policies for collecting resources and maintaining them over time. Doing so enables people and automated processes to ensure that resources and their descriptions are authoritative, accurate, complete, consistent, and non-redundant.

3.5.3.1. Institutional Curation

Curation is most necessary and explicit in institutional organizing systems where a large number of resources or their heterogeneity requires choices to be made about which ones should be most accessible, how they should be organized to ensure this access, and which ones need most to be preserved to ensure continued accessibility over time. Curation might be thought of as an ongoing or deferred selection activity because curation decisions must often be made on an item-by-item basis.

3.5.3.2. Social and Web Curation

Most websites are not curated systematically, and the decentralized nature of the web and its easy extensibility means that the web as a whole defies curation as traditionally defined. However, this traditional definition has been adapted to refer to the volunteer efforts of individuals that create, maintain, and evaluate web resources. When websites allow users to annotate, “tag,” “like,” or otherwise evaluate its resources, these bottom-up and distributed activities are often referred to as curation by “crowdsourcing.” These continuously aggregated actions and contributions of users result in the development of informal and organic “folksonomies” that create organization and authority through network effects.

Given the web’s decentralized nature, it is often difficult to determine which resources are original, authoritative, or authorized versions. Some take advantage of this by deliberately misclassifying, falsifying, or otherwise maliciously tampering with resources. An entirely new vocabulary has emerged to describe these web resources with bad intent: “spam,” “phishing,” “malware,” “fakeware,” “spyware,” “keyword stuffing,” “spamdexing,” “META tag abuse,” “link farms,” “cybersquatters,” “phantom sites,” “fake news,” and many more. Internet service providers, security software firms, email services, and search engines are engaged in a constant war against these kinds of malicious resources and techniques.

3.5.3.3. Computational Curation

Search engines continuously curate the web because the algorithms they use for computing relevance and ranking determine what resources people are likely to access. At a smaller scale, there are many kinds of tools for managing the quality of a website, such as ensuring that HTML content is valid, that links work, and that the site is being crawled completely. Another familiar example is the spam and content filtering that takes place in our email systems that automatically classifies incoming messages and sorts them into appropriate folders.

3.5.3.4. Discarding, Removing, and Not Keeping

So far, we have discussed maintenance as activities involved in preserving and protecting resources in an organizing system over time. An essential part of maintenance is the phasing out of resources that are damaged, expired, past their effectivity dates, no longer relevant to any interaction, or otherwise unusable.

Many organizations admit to a distinct lack of strategy in the removal aspect of maintenance. A firm with outdated storage technology might have to discard older data simply to make room for new data, and might do so without considering that keeping some summary statistics would be valuable for historical analysis. Other firms might be biased towards keeping information just because they went to the trouble of collecting or acquiring it.

Keeping an organizing system current often involves some amount of elimination of older resources to make space for the new: in fashion retail, the floor is constantly restocked with the latest styles. Software development teams will halt active support and documentation efforts of legacy versions.

Information resources are often discarded to comply with laws about retaining sensitive data. Governments and politicians sometimes destroy documents that might prove damaging or embarrassing if they are discovered through Freedom of Information requests or by opposing political parties.

More positively, the “right to be forgotten” movement and intentional destruction of information records about prior bankruptcy, credit problems, or juvenile arrests after a certain period has passed can be seen as a policy of “social forgetfulness” that gives people a chance to get on with their lives.

3.5.4. Governance

Governance overlaps with *curation* in meaning, but typically has more of policy focus (what should be done), rather than a process focus (how to do it). Governance is also more frequently used to describe curation in business and scientific organizing systems rather than in libraries, archives, and museums. Governance has a broader scope than curation because it extends beyond the resources in a collection and also applies to the software, computing, and networking environments needed to use them. This broader scope also means that governance must specify the rights and responsibilities for the people who might interact with the resources, the circumstances under which that might take place, and the methods they would be allowed to use.

Corporate governance is a common term applied to the ongoing maintenance and management of the relationship between operating practices and long-term strategic goals.

Data governance policies are often shaped by laws, regulations or policies that prohibit the collection of certain kinds of objects or types of information. Privacy laws prohibit the collection or misuse of personally identifiable information about healthcare, education, telecommunications, video rental, and, in some countries, restrict the information collected during web browsing.

3.5.4.1. Governance in Business Organizing Systems

Governance is essential to deal with the frequent changes in business organizing systems and the associated activities of data quality management, access control to ensure security and privacy, compliance, deletion, and archiving. For many of these activities, effective governance involves the design and implementation of standard services to ensure that the activities are performed in an effective and consistent manner.

Today's information-intensive businesses capture and create large amounts of digital data. Because digital data can be easily copied, data governance policies might require that all sensitive data be anonymized or encrypted to reduce the risk of privacy breaches. To identify the source of a data breach or to facilitate the assertion of a copyright infringement claim a digital watermark can be embedded in digital resources.

3.5.4.2. Governance in Scientific Organizing Systems

Scientific data poses special *governance* problems because of its enormous scale, which dwarfs the datasets managed in most business organizing systems. A scientific data collection might contain tens of millions of files and many petabytes of data. Furthermore, because scientific data is often created using specialized equipment or computers and undergoes complex workflows, it can be necessary to curate the technology and processing context along with data to preserve it. An additional barrier to effective scientific data governance is the lack of incentives in scientific culture and publication norms to invest in data retention for reuse by others.